Bayesian Analysis of Stochastic Process Models
(AMCS xxx)

INSTRUCTOR: Dr. Fabrizio Ruggeri
Email Address: fabrizio@mi.imati.cnr.it
Office Phone: TBA
Office Location: Building 1, Room TAB

Course Web Page: http://www.mi.imati.cnr.it/~fabrizio/kaust13.html

CLASS SCHEDULE: The course will start on Monday March, 25th and end on Saturday April, 20th with 90 minutes classes on Mondays, Wednesdays and Saturdays. Two extra classes, devoted to discussion of the students' projects are planned on April 22nd and April 24th. Room and Building will be announced in due time.

OFFICE HOURS:

- Two hours per week at a time decided by students and instructor during the first class.

Special appointments with the instructor may be arranged by mutual agreement.

ELIGIBILITY:
The course is organized by AMCS KAUST. It is the student's responsibility to check and prove eligibility.

PREREQUISITES:
The course is designed for students who have completed one or more introductory-level courses in statistics and probability. An introductory course on stochastic processes (especially Markov chains and Poisson processes) is not required but it could be helpful. Earlier exposure to Bayesian methods could be helpful as well, but the basics of the Bayesian approach will be presented in the course. Some experience of computer programming and the use of UNIX/LINUX/WINDOWS/MacOS systems or personal computers is assumed. The assigned project will require programming; the use of R (http://www.r-project.org/) is strongly suggested, although other programmes could be used. Students are supposed to know already how to programme in R or other software. Enrolled students will have access to computer facilities with R at KAUST if needed.

TEXT:
The course largely follows the book [1]. Examples with real data sets will be borrowed, mostly, from papers written by the instructor. Additional specific materials will be recommended during the course. The project to be developed during the course will be based on data (real or realistic) provided by the instructor.

COURSE OBJECTIVES:
The student who follows this course will be introduced to Bayesian modeling in

selected, but relevant, stochastic processes and their applications: Markov chains, Poisson processes, reliability and queues. The use of real examples will be helpful in understanding why and how perform a Bayesian analysis, from the elicitation of prior opinion from experts, its use in the choice of the prior distribution and the study of the consequences of such choice to the final estimations, forecasts, and decisions. Examples like gas escapes, train doors' failures, earthquake occurrences, allocation of hospital in beds and few others from engineering and telecommunications will help in going beyond the usual textbook choice of prior, through a critical illustration of the steps taken in analyzing those data. For the same reasons, students will be asked to analyze some data, from the elicitation of priors and modeling to (numerical) computation of estimates and forecasts and interpretation of findings.

COURSE OUTLINE:

- **Bayesian Analysis** (approx. 1 class)
  - Basics (prior, posterior, estimation, forecast, decision)
  - Prior elicitation in practice
  - Sensitivity to the choice of the prior
- **Markov chains** (approx. 3 classes)
  - Inference and prediction for discrete time Markov chains
  - Inference and prediction for continuous time Markov chains
  - Examples (TBA)
- **Poisson processes** (approx. 3 classes)
  - Homogeneous Poisson processes
  - Nonhomogeneous Poisson processes
  - Example (earthquake data)
- **Reliability** (approx. 4 classes)
  - Basic notions
  - Models for repairable and non repairable systems
  - Examples (gas escapes, train doors' failures and others)
- **Queues** (approx. 1 class)
  - Inference for M/M/1 queues
  - Other queues
  - Examples (Bed occupancy in hospital, TBA)

GRADING:

There will be a final examination consisting in the presentation of the data analysis performed, during the April 22nd and 24th classes.

The grading consists of two parts: two short written interim reports on the development of the project, and the final oral presentation of the project. The reports and the presentations are carried out by groups of students. Each group hands in a report for each of the assignments, and makes the final presentation.

Concerning Presentations: Final projects are presented by groups of students according to a certain schedule. Prepare a presentation of at most 30 minutes with overhead material including the formulation of the problem, theoretical analysis, R code to implement the statistical techniques, conclusions, open questions etc. Take the presentation seriously and use it as an opportunity of getting some practical training in the difficult art of oral presentation. Remember that presenting a material in a clear and convincing way requires quite a bit of preparation and training to be successful. We all need practice and positive criticism in this respect, both teachers and students. Practice before presenting the material! Make sure you will fit into the allocated time slot and be ready to answer probing questions!

Concerning Reports: Reports on the development of the project will be asked twice during the course (approximately every two weeks), and will generally involve mainly data analysis and some theoretical developments. The first report will be probably more focused on modeling and the second on computations. Each group should hand in a written report. The report has two purposes: it shows the level of progress in modeling and analyzing the data and gives the opportunity to practice written presentations of solutions. This means that a report with just formulas and/or figures is not acceptable. The report should allow the reader to understand what has been done, the problems addressed and the open ones, and the findings so far of the analysis. The presentation shall be prepared in such a way that fellow students who know nothing about the problem should be able to understand and be satisfied with it. Describe the formulation of the problem, the theoretical background, the results and the conclusions.

Numerical course grades will be based on a composite score:

- Active participation to the course (e.g. comments, extra reading): 10%
- First report: 20%
- Second report: 20%
- Final project presentation: 50%

SPECIAL ACCOMMODATIONS:
If you a have personal activity, a family, or a religious conflict with the course schedule, you may announce it to the instructor. Please contact the instructor during the first week of the course to discuss appropriate accommodations for any conflicts that can be foreseen. For illness-related absences, there are standard procedures to follow.

EXAM POLICY:
No quizzes or tests other than the final exam will normally be given. Acceptable medical excuses must state explicitly that the student should be excused from class.

References

[1] Rios Insua, D., Ruggeri, F., Wiper, M.P. (2012). *Bayesian Analysis of Stochastic Process Models*, Wiley, Chichester, UK.

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/~fabrizio/

COURSE OUTLINE

- **Bayesian Analysis** (1 class)
  - Basics (prior, posterior, estimation, forecast, decision)
  - Prior elicitation in practice
  - Sensitivity to the choice of the prior
- **Markov chains** (approx. 3 classes)
  - Inference and prediction for discrete time Markov chains
  - Inference and prediction for continuous time Markov chains
  - Examples (TBA)
- **Poisson processes** (approx. 3 classes)
  - Homogeneous Poisson processes
  - Nonhomogeneous Poisson processes
  - Example (earthquake data)

COURSE OUTLINE

- **Reliability** (approx. 4 classes)
  - Basic notions
  - Models for repairable and non repairable systems
  - Examples (gas escapes, train doors' failures and others)
- **Queues** (approx. 1 class)
  - Inference for M/M/1 queues
  - Other queues
  - Examples (Bed occupancy in hospital, TBA)

GRADING

Grade based on a composite score:

- Active participation to the course (e.g. comments, extra reading): **10%**
- Data analysis (performed in groups)
  - First short interim (mostly on modeling) written report (due 6/4): **20%**
  - Second short interim (mostly on computations) written report (due 17/4): **20%**
  - Final oral project presentation (22 and 24/4): **50%**
    - 30 minutes presentation with overhead material
    - formulation of the problem
    - modeling
    - R code to implement the statistical techniques
    - conclusions
    - open problems

SCHEDULE - OFFICE HOURS

- Sat. 15.00-16.30
- Mon. 10.30-12.00
- Wed. 15.00-16.30
- Office hours: open to suggestions, possibly same day as classes

TODAY'S CLASS

- Basics on Bayesian statistics
  - prior
  - posterior
  - estimation
  - forecast
  - decision
- Prior elicitation in practice
- Sensitivity to the choice of the prior

ALL BAYESIANS IN DAILY LIFE?

Visit Milano or not?

- Prior knowledge
  - Where is Milano? What is it?
  - Fashion, football, business and nothing else
- Data collection
  - Scuba diving association website
  - City of Milano official website
- Posterior knowledge
  - Probably not a good place for scuba divers
  - Fashion, football, business, art, music, scenic locations nearby (e.g. Lake Como)
- Forecast: Will I enjoy Milano or not?
- Decision: To go or not to go? **GO!**

BAYES THEOREM

- Patient subject to medical diagnostic test (P or N) for a disease D
- *Sensitivity* .95, i.e. $\mathbb{P}(P|D) = .95$
- *Specificity* .9, i.e. $\mathbb{P}(P^C|D^C) = .1$
- Physician's belief on patient having the disease 1%, i.e. $\mathbb{P}(D) = .01$
- Positive test $\Rightarrow \mathbb{P}(D|P)$?

BAYES THEOREM

$$\mathbb{P}(D|P) = \frac{\mathbb{P}(D \bigcap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^C)\mathbb{P}(D^C)}$$

$$= \frac{.95 \cdot .01}{.95 \cdot .01 + .1 \cdot .99} = .0875$$

Positive test updates belief on patient having the disease:
from 1% to 8.75%

Prior opinion updated into *posterior* one

BAYES THEOREM

- Partition $\{A_1, \ldots, A_n\}$ of Ω and $B \subset \Omega : \mathbb{P}(B) > 0$

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)P(A_i)}{\sum_{j=1}^{n} \mathbb{P}(B|A_j)P(A_j)}$$

- X r.v. with density $f(x|\lambda)$, prior $\pi(\lambda)$

$$\Rightarrow \text{posterior } \pi(\lambda|x) = \frac{f(x|\lambda)\pi(\lambda)}{\int f(x|\omega)\pi(\omega)d\omega}$$

BAYESIAN STATISTICS

Bayesian statistics is . . .

- . . . another way to make inference and forecast on population features (*practitioner's view*)
- . . . a way to learn from experience and improve own knowledge (*educated layman's view*)
- . . . a formal tool to combine prior knowledge and experiments (*mathematician's view*)
- . . . cheating (*hardcore frequentist statistician's view*)
- . . .

NOTIONS OF PROBABILITY

- Classical (random choice, equally likely events)

- Frequentist (probability as asymptotic limit of frequency)

- Subjective/Bayesian

- Axiomatic (Kolmogorov), which contains the other three

Bayesian $\Rightarrow$ need to specify subjective P in $(\Omega, \mathcal{F}, P)$

ASSESSMENT OF PRIOR PROBABILITIES

T= person having a tumor in his/her life
I= person having an infarction in his/her life

$\mathbb{P}(T \bigcup I) = .2,\ \mathbb{P}(T) = .3,\ \mathbb{P}(I) = .05,\ \mathbb{P}(T \bigcap I) = .1$

ASSESSMENT OF PRIOR PROBABILITIES

T= person having a tumor in his/her life
I= person having an infarction in his/her life

$\mathbb{P}(T \bigcup I) = .2,\ \mathbb{P}(T) = .3,\ \mathbb{P}(I) = .05,\ \mathbb{P}(T \bigcap I) = .1$

- $\mathbb{P}(T \bigcup I) \geq \mathbb{P}(T)$

- $\mathbb{P}(I) \geq \mathbb{P}(T \bigcap I)$

ASSESSMENT OF PRIOR PROBABILITIES

T= person having a tumor in his/her life
I= person having an infarction in his/her life

$$\mathbb{P}(T \bigcup I) = .3,\ \mathbb{P}(T) = .2,\ \mathbb{P}(I) = .2,\ \mathbb{P}(T \bigcap I) = .15$$

ASSESSMENT OF PRIOR PROBABILITIES

T= person having a tumor in his/her life
I= person having an infarction in his/her life

$\mathbb{P}(T \bigcup I) = .3, \; \mathbb{P}(T) = .2, \; \mathbb{P}(I) = .2, \; \mathbb{P}(T \bigcap I) = .15$

- $.3 = \mathbb{P}(T \bigcup I) = \mathbb{P}(T) + \mathbb{P}(I) - \mathbb{P}(T \bigcap I) = .25$

- $\mathbb{P}(T \bigcup I) = .3, \; \mathbb{P}(T) = .2, \; \mathbb{P}(I) = .2, \; \mathbb{P}(T \bigcap I) = .1$

$\Rightarrow$ assessments should comply with probability rules

ASSESSING DISCRETE DISTRIBUTIONS: BETS

Probability Italy will win next FIFA World Cup

1. I bet $Y = 10\$$ on the Italian victory. How much are you willing to bet with me against the victory? (Say $10\$$ the first time, then $15\$$ and $20\$$)

2. Now let's reverse. You bet $Y = 10\$$ on the victory and you suggest my *fair* bet on the loss (Say $30\$$ the first time, then $25\$$ and $20\$$)

3. Let's repeat 1 and 2 until it is indifferent for you to bet either on the loss or the victory (i.e. $20\$$)

4. Let X be the amount you bet on the loss of Italy

5. Fair bet $\Rightarrow YP(loss) = XP(victory)$

6. $P(victory) = 1 - P(loss) \Rightarrow P(loss) = \dfrac{X}{X+Y} = \dfrac{20}{20+10} = \dfrac{2}{3}$

ASSESSING DISCRETE DISTRIBUTIONS: BETS

Problems

- Many people do not like to bet
- Most people dislike the idea of losing money
- I was talking about a $10\$$ bet, but would you have bet $1000X$ if I had bet $10,000\$$?
- Reaching convergence to a fair bet might be a long process

REFERENCE LOTTERIES

1. Lottery 1
   - Get a trip to Italy if Italy wins
   - Get a trip to Cyprus if Italy looses
2. Lottery 2
   - Get a trip to Italy with probability p, e.g. if a random number generated from a uniform distribution on $[0, 1]$ is $\leq p$
   - Get a trip to Cyprus with probability $1 - p$, e.g. if a random number generated from a uniform distribution on $[0, 1]$ is $> p$
3. Specify p_1. Which lottery do you prefer?
4. If Lottery 1 is preferred offer change p_i to $p_{i+1} > p_i$.
5. If Lottery 2 is preferred offer change p_i to $p_{i+1} < p_i$.
6. When indifference point is reached $\Rightarrow P(victory) = p_i$, else Goto 4.

ASSESSING CONTINUOUS DISTRIBUTIONS

X continuous random variable (e.g. light bulb lifetime)

- Choose $x_1, \ldots, x_n$
- Assess $F(x_i) = P(X \leq x_i), i = 1, n$
- Draw $F(x)$
- Look at $F(x)$ at some points for consistency

or

- Choose probabilities $p_1, \ldots, p_n$
- Find x_i's s.t. $F(x_i) = P(X \leq x_i) = p_1, i = 1, n$
- Draw $F(x)$
- Look at $F(x)$ at some points for consistency

ILLUSTRATIVE EXAMPLE: FREQUENTIST APPROACH

Light bulb lifetime $\Rightarrow X \sim \mathcal{E}(\lambda)$ & $f(x;\lambda) = \lambda e^{-\lambda x} \qquad x, \lambda > 0$

- Sample $\underline{X} = (X_1, \ldots, X_n)$, i.i.d. $\mathcal{E}(\lambda)$

- Likelihood $l_x(\lambda) = \prod_{i=1}^n f(X_i;\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n X_i}$

- MLE: $\hat{\lambda} = n / \sum_{i=1}^n X_i$, C.I., UMVUE, consistency, etc.

What about available prior information on light bulbs behavior?
How can we translate it? $\Rightarrow$ model and **parameter**

ILLUSTRATIVE EXAMPLE: BAYESIAN APPROACH

Light bulb lifetime $\Rightarrow X \sim \mathcal{E}(\lambda)$ & $f(x;\lambda) = \lambda e^{-\lambda x} \qquad x, \lambda > 0$

- Sample $\underline{X} = (X_1, \ldots, X_n)$, i.i.d. $\mathcal{E}(\lambda)$
- Likelihood $l_x(\lambda) = \prod_{i=1}^n f(X_i;\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n X_i}$
- Prior $\lambda \sim \mathcal{G}(\alpha, \beta)$, $\pi(\lambda) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$
- Posterior $\pi(\lambda|\underline{X}) \propto \lambda^n e^{-\lambda \sum_{i=1}^n X_i} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}$

  $\Rightarrow \lambda|\underline{X} \sim \mathcal{G}(\alpha + n, \beta + \sum_{i=1}^n X_i)$

Posterior distribution fundamental in Bayesian analysis

PARAMETER ESTIMATION - DECISION ANALYSIS

- Loss function $L(\lambda, a)$, $a \in \mathcal{A}$ action space

- Minimize $\mathcal{E}^{\pi(\lambda|\underline{X})} L(\lambda, a) = \int L(\lambda, a)\pi(\lambda|\underline{X})d\lambda$ w.r.t. a
  $\Rightarrow \widehat{\lambda}$ Bayesian optimal estimator of λ
  - $\widehat{\lambda}$ posterior median if $L(\lambda, a) = |\lambda - a|$

  - $\widehat{\lambda}$ posterior mean $\mathcal{E}^{\pi(\lambda|\underline{X})}\lambda$ if $L(\lambda, a) = (\lambda - a)^2$

$$
\begin{aligned}
\mathcal{E}^{\pi(\lambda|\underline{X})} L(\lambda, a) &= \int (\lambda - a)^2 \pi(\lambda|\underline{X})d\lambda \\
&= \int \lambda^2 \pi(\lambda|\underline{X})d\lambda - 2a \int \lambda\pi(\lambda|\underline{X})d\lambda + a^2 \cdot 1 \\
&= \int \lambda^2 \pi(\lambda|\underline{X})d\lambda - 2a\mathcal{E}^{\pi(\lambda|\underline{X})}\lambda + a^2
\end{aligned}
$$

PARAMETER ESTIMATION

- Light bulb: posterior mean $\hat{\lambda} = (\alpha + n)/(\beta + \sum_{i=1}^{n} X_i)$
  $\Rightarrow$ compare with
  - prior mean α/β
  - MLE $n/\sum_{i=1}^{n} X_i$
- MAP (Maximum a posteriori)
  $\Rightarrow \hat{\lambda} = \dfrac{\alpha + n - 1}{\beta + \sum X_i}$
- LPM (Largest posterior mode)
  $\Rightarrow$ here it coincides with MAP (unique posterior mode)

PARAMETER ESTIMATION

Prior influence

- Posterior mean: $\mu^* = \dfrac{\alpha + n}{\beta + \sum X_i}$
- Prior mean: $\tilde{\mu} = \dfrac{\alpha}{\beta}$ (and variance $\sigma^2 = \dfrac{\alpha}{\beta^2}$)
- MLE: $\dfrac{n}{\sum X_i}$
- $\alpha_1 = k\alpha$ and $\beta_1 = k\beta \Rightarrow \mu_1 = \mu$ and $\sigma_1^2 = \sigma^2/k$
- $k \to 0 \Rightarrow \mu^* \to$ MLE
- $k \to \infty \Rightarrow \mu^* \to \tilde{\mu}$

PARAMETER ESTIMATION

Prior influence (multinomial data and Dirichlet prior)

$(n_1, \ldots, n_k) \sim \mathcal{MN}(n, p_1, \ldots, p_k)$

$(p_1, \ldots, p_k) \sim \mathcal{D}ir(s\alpha_1, \ldots, s\alpha_k), \ \sum \alpha_i = 1, \ s > 0$

- Posterior mean: $p_i^* = \dfrac{s\alpha_i + n_i}{s + n}$
- Prior mean: $\tilde{p}_i = \alpha_i$
- MLE: $\dfrac{n_i}{n}$
- $s \to 0 \Rightarrow p_i^* \to$ MLE
- $s \to \infty \Rightarrow p_i^* \to \tilde{p}_i$

CREDIBLE INTERVALS

- $\mathcal{P}(\lambda \in A|\underline{X})$, credible (and Highest Posterior Density) intervals
- Compare with confidence intervals
- Light bulb:

$$\mathcal{P}(\lambda \leq z|\underline{X}) = \int_0^z \frac{(\beta + \sum X_i)^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{\alpha+n-1} e^{-(\beta+\sum X_i)\lambda} d\lambda$$

HYPOTHESIS TESTING

- One sided test: $H_0 : \lambda \leq \lambda_0$ vs. $H_1 : \lambda > \lambda_0$

  $\Rightarrow$ Reject H_0 iff $\mathbb{P}(\lambda \leq \lambda_0 | \underline{X}) \leq \alpha$, α significance level

- Two sided test: $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$

  - Do not reject if $\lambda_0 \in A$, A $100(1-\alpha)$% credible interval
  - Consider $\mathbb{P}([\lambda_0 - \epsilon, \lambda_0 + \epsilon] | \underline{X})$
  - Dirac measure: $\mathbb{P}(\lambda_0) > 0$ and consider $\mathbb{P}(\lambda_0 | \underline{X})$

PREDICTION

- Prediction $P(X_{n+1}|\underline{X}) = \int P(X_{n+1}|\lambda)\pi(\lambda|\underline{X})d\lambda$

- Light bulb: $X_{n+1}|\lambda \sim \mathcal{E}(\lambda),\ \lambda|\underline{X} \sim \mathcal{G}(\alpha + n, \beta + \sum X_i)$

- $f_{X_{n+1}}(x|\underline{X}) = (\alpha + n)\dfrac{(\beta + \sum X_i)^{\alpha+n}}{(\beta + \sum X_i + x)^{\alpha+n+1}}$

MODEL SELECTION

Compare $\mathcal{M}_1 = \{f_1(x|\theta_1), \pi(\theta_1)\}$ and $\mathcal{M}_2 = \{f_2(x|\theta_2), \pi(\theta_2)\}$

- Bayes factor

$$\Rightarrow BF = \frac{\int f_1(x|\theta_1)\pi(\theta_1)d\theta_1}{\int f_2(x|\theta_2)\pi(\theta_2)d\theta_2}$$

BF	$2\log_{10} BF$	Evidence in favor of $\mathcal{M}_1$
1 to 3	0 to 2	Hardly worth commenting
3 to 20	2 to 6	Positive
20 to 150	6 to 10	Strong
> 150	> 10	Very strong

- Posterior odds

$$\Rightarrow \frac{P(\mathcal{M}_1|data)}{P(\mathcal{M}_2|data)} = \frac{P(data|\mathcal{M}_1)}{P(data|\mathcal{M}_2)} \cdot \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)} = BF \cdot \frac{P(\mathcal{M}_1)}{P(\mathcal{M}_2)}$$

WHY BAYESIAN? (*A BIASED VIEW*)

- a) $P(Head) = \theta$ vs. b) P(someone passing a given exam)$= \theta$
  - Frequentist interpretation only for a)
  - Subjective opinion on θ in both cases

- Bayesian approach follows from rationality axioms
  - Actions $a \preceq b$ (b *at least as good as* a)
    $\Rightarrow a \preceq b \Leftrightarrow \exists L, \pi : \int L(\theta, b)\pi(\theta)d\theta \leq \int L(\theta, a)\pi(\theta)d\theta$

WHY BAYESIAN? (*A BIASED VIEW*)

- $X \sim \mathcal{B}ern(\theta)$ & sample $X_1 = X_2 = 0$
  $\Rightarrow \widehat{\theta} = 0$ MLE (reasonable?)

- In decision analysis, frequentist procedures average over all possible (unobserved) outcomes, unlike Bayesian ones

- Nuisance parameters, like σ^2 in $\mathcal{N}(\mu, \sigma^2)$, removed by integrating them out

- Predictions: very easy

- Few data and lot of expertise

WHY BAYESIAN? (*A BIASED VIEW*)

- p-value vs. Bayes factor
  $\Rightarrow$ many issues (e.g. p-value depends only on distribution under H_0, unlike Bayes factor), comparisons and attempts to reconcile

- No need for asymptotics, but estimation for any sample size

- MCMC (and its implementation in, e.g., WinBugs) allows for (relatively) straightforward computations in complex models

PRIOR CHOICE

Where to start from?

- $X \sim \mathcal{E}(\lambda)$
- $f(x|\lambda) = \lambda \exp\{-\lambda x\}$
- $P(X \leq x) = F(x) = 1 - S(x) = 1 - \exp\{-\lambda x\}$

$\Rightarrow$ *Physical* properties of λ

- $\mathbf{E}X = 1/\lambda$
- $VarX = 1/\lambda^2$
- $h(x) = \dfrac{f(x)}{S(x)} = \dfrac{\lambda \exp\{-\lambda x\}}{\exp\{-\lambda x\}} = \lambda$ (hazard function)

PRIOR CHOICE

Possible available information

- Exact prior $\pi(\lambda)$ (???)
- Quantiles of X_i, i.e. $P(X_i \leq x_q) = q$
- Quantiles of λ, i.e. $P(\lambda \leq \lambda_q) = q$
- Moments $\mathbb{E}\lambda^k$ of λ, i.e. $\int \lambda^k \pi(\lambda) d\lambda = a_k \Leftrightarrow \int (\lambda^k - a_k)\pi(\lambda) d\lambda = 0$
- Generalised moments of λ, i.e. $\int h(\lambda)\pi(\lambda) d\lambda = 0$
- Most likely value and upper and lower bounds
- ...
- None of them

PRIOR CHOICE

How to get information?

- Results from previous experiments (e.g. 75% of light bulbs had failed after 2 years of operation $\Rightarrow$ 2 years is the 75% quantile of X_i)
- Split of possible values of λ or X_i into equally likely intervals $\Rightarrow$ quantiles
- Most likely value and upper and lower bounds
- *Expected* value of λ and *confidence* on such value (mean and variance)
- ...

PRIOR CHOICE

How to combine information from n experts?

- Individual analyses and comparison a posteriori

- Opinions $(\lambda_1, \ldots, \lambda_n)$ as *sample* from $\pi(\lambda|\omega)$
  - Statements on quantiles G_q of $X \sim \mathcal{E}(\lambda)$
  - $q = 1 - \exp\{-\lambda G_q\}$

  $\Rightarrow$ sample mean and sample variance to fit $\pi(\lambda|\widehat{\omega})$

- Class of priors (n priors or their convex combination)

SEQUENTIAL PRIOR UPDATE

Data arriving altogether or one at the time

- Altogether: $\lambda \sim \mathcal{G}(\alpha, \beta) \rightarrow \mathcal{G}(\alpha + n, \beta + \sum X_i)$

- One at the time:
  $\lambda \sim \mathcal{G}(\alpha, \beta) \rightarrow \mathcal{G}(\alpha + 1, \beta + X_1) \rightarrow \ldots \mathcal{G}(\alpha + n, \beta + \sum X_i)$

SEQUENTIAL UPDATE: POWER PRIOR

Data from past experiments $\Rightarrow$ likelihood $l(\lambda) \propto \lambda^n e^{-\lambda \sum X_i}$

- Prior $\pi(\lambda)(l(\lambda))^\alpha$
- $0 \leq \alpha \leq 1$
- Possible prior on α

PRIOR CHOICE

Which prior?

- $\lambda \sim \mathcal{G}(\alpha, \beta) \Rightarrow f(\lambda|\alpha, \beta) = \beta^{\alpha} \lambda^{\alpha-1} \exp\{-\beta\lambda\}/\Gamma(\alpha)$ (conjugate)
- $\lambda \sim \mathcal{LN}(\mu, \sigma^2) \Rightarrow f(\lambda|\mu, \sigma^2) = \{\lambda\sigma\sqrt{2\Pi}\}^{-1} \exp\{-(\log \lambda - \mu)^2/(2\sigma^2)\}$
- $\lambda \sim \mathcal{GEV}(\mu, \sigma, \theta) \Rightarrow f(\lambda) = \frac{1}{\sigma} \left[1 + \theta \left(\frac{\lambda - \mu}{\sigma}\right)\right]_+^{-1/\theta - 1} \exp\left\{-\left[1 + \theta \left(\frac{\lambda - \mu}{\sigma}\right)\right]_+^{-1/\theta}\right\}$
- $\lambda \sim \mathcal{T}(l, m, u)$ (triangular)
- $\lambda \sim \mathcal{U}(l, u)$
- $\lambda \sim \mathcal{W}(\mu, \alpha, \beta) \Rightarrow f(\lambda) = \frac{\beta}{\alpha} \left(\frac{\lambda - \mu}{\alpha}\right)^{\beta - 1} \exp\{-\left(\frac{\lambda - \mu}{\alpha}\right)^{\beta}\}$
- $\ldots$

PRIOR CHOICE

Choice of a prior

- Defined on suitable set (interval vs. positive real)
- Suitable functional form (monotone/unimodal, heavy/light tails, etc.)
- Mathematical convenience
- *Tradition* (e.g. lognormal for engineers)

PRIOR CHOICE

Gamma prior - choice of hyperparameters

- $X_1, \ldots, X_n \sim \mathcal{E}(\lambda)$
- $f(X_1, \ldots, X_n|\lambda) = \lambda^n \exp\{-\lambda \sum X_i\}$
- $\lambda \sim \mathcal{G}(\alpha, \beta) \Rightarrow f(\lambda|\alpha, \beta) = \beta^\alpha \lambda^{\alpha-1} \exp\{-\beta\lambda\}/\Gamma(\alpha)$
- $\Rightarrow \lambda|X_1, \ldots, X_n \sim \mathcal{G}(\alpha + n, \beta + \sum X_i)$

PRIOR CHOICE

Gamma prior - choice of hyperparameters

- $\mathcal{E}\lambda = \mu = \alpha/\beta$ and $Var\lambda = \sigma^2 = \alpha/\beta^2$
  $\Rightarrow \alpha = \mu^2/\sigma^2$ and $\beta = \mu/\sigma^2$

- Two quantiles $\Rightarrow (\alpha, \beta)$ using, say, Wilson-Hilferty approximation. Third quantile specified to check consistency

- *Hypothetical experiment*: posterior $\mathcal{G}(\alpha + n, \beta + \sum X_i)$
  $\Rightarrow \alpha$ *sample size* and β *sample sum*

PRIOR CHOICE

Using data to choose hyperparameters

- choose a prior $\pi(\lambda|\omega)$ of given functional form and use data to fit ω, i.e. look for $\widehat{\omega} = \arg\max \int f(data|\lambda)\pi(\lambda|\omega)d\lambda$

  (*empirical Bayes*)

Typical example

- i batches of n_i light bulbs each
- light bulbs in same batch with same properties
- light bulbs in different batches with similar properties

PRIOR CHOICE

Hierarchical model

- $X_{ij_i}|\lambda_i \sim \mathcal{E}(\lambda_i), i = 1, n, j_i = 1, n_i$
- $\lambda_i|\underline{\beta} \sim \mathcal{G}(\alpha e^{\underline{Z}_i^T \underline{\beta}}, \alpha)$, α known, s.t. $\mathcal{E}\lambda_i = e^{\underline{Z}_i^T \underline{\beta}}$
- $\pi(\underline{\beta})$
- Improper priors, numerical approximation (Albert, 1988)
- Empirical Bayes
  - $\lambda_i|\underline{\beta}, \underline{d} \sim \mathcal{G}(\alpha e^{\underline{Z}_i^T \underline{\beta}} + n_i, \alpha + \sum x_{ij_i})$, $\lambda_i \perp \lambda_j|\underline{d}$
  - $f(\underline{d}|\underline{\beta}) = \int f(\underline{d}|\underline{\lambda})\pi(\underline{\lambda}|\underline{\beta})d\underline{\lambda}$ maximized by $\underline{\widehat{\beta}}$
    $\Rightarrow \lambda_i|\underline{\widehat{\beta}}, \underline{d} \sim \mathcal{G}(\alpha e^{\underline{Z}_i^T \underline{\widehat{\beta}}} + n_i, \alpha + \sum x_{ij_i})$, $\forall i$
- “Pure” Bayesian approach $\Rightarrow$ prior on $(\alpha, \underline{\beta})$

BAYESIAN SIMULATIONS

Alternative choice: $\lambda \sim \mathcal{LN}(\alpha, \beta)$

- no posterior in closed form $\Rightarrow$ numerical simulation

Markov Chain Monte Carlo (MCMC):

- draw$^{(*)}$ a sample $\lambda^{(1)}, \lambda^{(2)}, \ldots$ (Monte Carlo) ...
- ... from a Markov Chain whose stationary distribution is ...
- ... the posterior $\pi(\lambda|\underline{X})$ and compute ...
- $\mathcal{E}(\lambda|\underline{X}) \approx \sum_{i=m+1}^{n} \lambda^{(i)}/(n-m)$, etc.

(*) For $\lambda = (\theta, \mu) \Rightarrow$ Gibbs sampler:

- draw $\theta^{(i)}$ from $\theta|\mu^{(i-1)}, \underline{X}$
- draw $\mu^{(i)}$ from $\mu|\theta^{(i)}, \underline{X}$
- repeat *until convergence*

MCMC: REGRESSION

- $y = \beta_0 + \beta_1 x + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma^2)$

- $(y_1, x_1), \ldots, (y_n, x_n)$

- Likelihood $\propto (\sigma^2)^{-n/2}$ exp$\{\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2\}$

- Priors: $\beta_0 \sim \mathcal{N}, \beta_1 \sim \mathcal{N}, \sigma^2 \sim \mathcal{IG}$

- Full posterior conditionals:
  $\beta_0 | \beta_1, \sigma^2 \sim \mathcal{N}, \beta_1 | \beta_0, \sigma^2 \sim \mathcal{N}, \sigma^2 | \beta_0, \beta_1 \sim \mathcal{IG}$
  $\Rightarrow$ MCMC

PRIOR CHOICE

Other approaches

- use information to choose parameters of a random distribution on the space of probability measures, e.g. parameter η of Dirichlet process ($P \sim \mathcal{DP}(\eta)$), defined by finite dimensional distributions: for any partition $(A_1, \ldots, A_m)$,
  $\Rightarrow (P(A_1), \ldots, P(A_m)) \sim \mathcal{D}(\eta(A_1), \ldots, \eta(A_m))$

  (*Bayesian nonparametrics*)

- use Jeffreys'/reference/improper priors

  (*objective Bayes*)

- use a class of priors

  (*Bayesian robustness*)

PRIOR CHOICE

Inflows to a reservoir in a given month (Ríos Insua et al,1997)

- Logarithm of inflows: $X_1, \ldots, X_n$ i.i.d. $\mathcal{N}(\theta, \sigma^2)$, with σ^2 known
- Improper prior $\pi(\theta) \propto 1 \Rightarrow$ posterior $\mathcal{N}(\frac{\sum x_i}{n}, \frac{\sigma^2}{n})$
- Proper prior $\mathcal{N}(\mu_0, \sigma_0^2) \Rightarrow$ posterior $\mathcal{N}(\frac{n\bar{x}/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2}, \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)^{-1})$
- Jeffreys prior $\pi(\theta) \propto \sqrt{I(\theta)}$
  - Expected Fisher information $I(\theta) = -E_X\left[\frac{d^2}{d\theta^2} \log f(X|\theta)\right]$
  - Here $\frac{d^2}{d\theta^2} \log f(X|\theta) = -\frac{1}{\sigma^2} \Rightarrow \pi(\theta) \propto 1$

BAYESIAN ROBUSTNESS

Gamma prior $\pi(\lambda)$

- excellent for computations but ...
- ... not matching all and only the prior knowledge

Impossibility of specifying a unique prior $\Rightarrow$ class Γ of priors

ROBUSTNESS: ILLUSTRATIVE EXAMPLE

Influence of prior choice (Berger, 1985)

- $X \sim \mathcal{N}(\theta, 1)$
- Expert's opinion on prior P:
  median at 0, quartiles at ± 1, symmetric and unimodal
- $\Rightarrow$ Possible priors include $\mathcal{C}(0, 1)$ or $\mathcal{N}(0, 2.19)$
- Posterior mean

x	0	1	2	4.5	10
$\mu^C(x)$	0	0.52	1.27	4.09	9.80
$\mu^N(x)$	0	0.69	1.37	3.09	6.87

- Posterior median w.r.t. posterior mean

CONCERNS ON BAYES

Motivations for Bayesian robustness

- Arbitrariness in the choice of $\pi(\theta)$ *et al.*
  $\Rightarrow$ inferences and decisions heavily affected

- Expert unable to provide, in a reasonable time, an *exact* prior reflecting his/her beliefs $\Rightarrow$ huge amount of information (e.g. choice of the functional form of the prior) added by analyst, although not corresponding to actual knowledge

NEED FOR BAYESIAN ROBUSTNESS

- partially specified priors
- conflicting loss functions
- opinions (priors and/or losses) expressed by a group of people instead of one person
- . . .

BAYESIAN ROBUSTNESS

Mathematical tools and *philosophical* approach

- to model uncertainty through classes of priors/models/losses

- to measure uncertainty and its effect

- to avoid arbitrary assumptions

- to favor acceptance of Bayesian approach

BAYESIAN ROBUSTNESS

- An helpful tool to convince agencies (e.g. FDA) to accept Bayesian methods? An old, but still unsolved, problem ...

- Bayesian robustness applied to efficacy of drug: *is the drug efficient for all the priors in a class?*

- Backward Bayesian robustness: *what are the priors leading to state the efficacy of the drug (or its inefficacy)?*

BAYESIAN ROBUSTNESS

A more formal statement about model and prior sensitivity

- $M = \{Q_\theta; \theta \in \Theta\}$, Q_θ probability on $(\mathcal{X}, \mathcal{F}_\mathcal{X})$
- Sample $\underline{x} = (x_1, \ldots, x_n) \Rightarrow$ likelihood $l_x(\theta) \equiv l_x(\theta | x_1, \ldots, x_n)$
- Prior P su $(\Theta, \mathcal{F}) \Rightarrow$ posterior P^*
- Uncertainty about M and/or $P \Rightarrow$ changes in
  - $E_{P^*}[h(\theta)] = \dfrac{\int_\Theta h(\theta) l(\theta) P(d\theta)}{\int_\Theta l(\theta) P(d\theta)}$
  - P^*

Bayesian robustness studies these changes

ROBUST BAYESIAN ANALYSIS

- Need for robustness analysis recognized by many Bayesians but ...
- ... almost nobody considers classes as entertained by robust Bayesians, ...
- ... preferring comparison among few priors (*informal sensitivity*)
- I keep advocating a *proper* robust Bayesian approach

ROBUST BAYESIAN ANALYSIS

Interest in robustness w.r.t. to changes in prior/model/loss but most work concentrated on priors since

- controversial aspect of Bayesian approach
- easier (w.r.t. model) computations
- problems with interpretation of classes of models/likelihood
- often interest in posterior mean (corresponding to optimal Bayesian action under squared loss function) and no need for classes of losses

ROBUST BAYESIAN ANALYSIS

Three major approaches

- *Informal sensitivity*: comparison among few priors

- *Global sensitivity*: study over a class of priors specified by some features

- *Local sensitivity*: infinitesimal changes w.r.t. elicited prior

ROBUST BAYESIAN ANALYSIS

We concentrate mostly on sensitivity to changes in the prior

- Choice of a class Γ of priors

- Computation of a robustness measure, e.g. range $\delta = \overline{\rho} - \underline{\rho}$ $(\overline{\rho} = \sup_{P \in \Gamma} E_{P^*}[h(\theta)]$ and $\underline{\rho} = \inf_{P \in \Gamma} E_{P^*}[h(\theta)])$

  - δ "small" $\Rightarrow$ robustness

  - δ "large", $\Gamma_1 \subset \Gamma$ and/or new data

  - δ "large", Γ and same data

ROBUST BAYESIAN ANALYSIS

Relaxing the unique prior assumption (Berger and O'Hagan, 1988)

- $X \sim \mathcal{N}(\theta, 1)$
- Prior $\theta \sim \mathcal{N}(0, 2)$
- Data $x = 1.5 \Rightarrow$ posterior $\theta | x \sim \mathcal{N}(1, 2/3)$
- Split $\Re$ in intervals with same probability p_i as prior $\mathcal{N}(0, 2)$

ROBUST BAYESIAN ANALYSIS

Refining the class of priors (Berger and O'Hagan, 1988)

I_i	p_i	p_i^*	Γ_Q	Γ_{QU}
(-∞,-2)	0.08	.0001	(0,0.001)	(0,0.0002)
(-2,-1)	0.16	.007	(0.001,0.029)	(0.006,0.011)
(-1,0)	0.26	.103	(0.024,0.272)	(0.095,0.166)
(0,1)	0.26	.390	(0.208,0.600)	(0.322,0.447)
(1,2)	0.16	.390	(0.265,0.625)	(0.353,0.473)
(2,+∞,)	0.08	.110	(0,0.229)	(0,0.156)

- Γ_Q quantile class and Γ_{QU} unimodal quantile class
- Robustness in Γ_{QU}
- Huge reduction of δ from Γ_Q to Γ_{QU}

CLASSES OF PRIORS

Desirable features of classes of priors

- Easy elicitation and interpretation (*e.g. moments, quantiles, symmetry, unimodality)*
- Compatible with prior knowledge (*e.g. quantile class)*
- Simple computations
- Without unreasonable priors (*e.g. unimodal quantile class, ruling out discrete distributions)*

CLASSES OF PRIORS

- $\Gamma_P = \{P : p(\theta; \omega), \omega \in \Omega\}$ *(Parametric class)*
  - $\Gamma_P = \{\mathcal{G}(\alpha, \beta) : \alpha/\beta = \mu\}$
  - $\Gamma_P = \{\mathcal{G}(\alpha, \beta) : l_1 \leq \alpha \leq u_1, l_2 \leq \beta \leq u_2\}$
  - $\Gamma_P = \{\mathcal{G}(\alpha, \beta) : l_1 \leq \alpha/\beta \leq u_1, l_2 \leq \alpha/\beta^2 \leq u_2\}$

CLASSES OF PRIORS

- $\Gamma_Q = \{P : \alpha_i \leq P(I_i) \leq \beta_i, i = 1, \ldots, m\}$ *(Quantile class)*
  - $\Gamma_Q = \{P : \theta_0 \text{ median}\}$
  - $\Gamma_Q = \{P : P(A) = \alpha\}$
  - $\Gamma_Q = \{P : q_1, \ldots, q_n \text{ quantiles of order } \alpha_1, \ldots, \alpha_n\}$
- $\Gamma_{QU} = \{P \in \Gamma_Q, \text{ unimodal}\}$ *(Unimodal quantile class)*
- $\Gamma_{QUS} = \{P \in \Gamma_{QU}, \text{ symmetric}\}$ *(Symmetric, unimodal quantile class)*

CLASSES OF PRIORS

- $\Gamma_{GM} = \{P : \int h_i(\theta) dP(\theta) = \alpha_i, i = 1, \ldots, m\}$ *(Generalized moments class)*
  - $h_i(\theta) = \theta^i$ *(Moments class)*
  - $h_i(\theta) = I_{A_i}(\theta)$ *(Quantile class)*
  - $h(\theta) = \int_{-\infty}^{x} f(t|\theta) dt \Rightarrow \int h(\theta) dP(\theta) = \int_{-\infty}^{x} f(t) dt$ *(Prior predictive distribution)*

CLASSES OF PRIORS

- $\Gamma^{DR} = \{P : L(\theta) \leq \alpha p(\theta) \leq U(\theta), \alpha > 0\}$ *(Density ratio class)*

- $\Gamma^{B} = \{P : L(\theta) \leq p(\theta) \leq U(\theta)\}$ *(Density bounded class)*

- $\Gamma^{DB} = \{F \text{ c}.d.f. : F_l(\theta) \leq F(\theta) \leq F_u(\theta), \forall\theta\}$ *(Distribution bounded class)*

CLASSES OF PRIORS

- Classes introduced so far are defined through some features (e.g. quantiles) ...

- ... whereas now we introduce others (*Neighborhood classes*) which represent perturbations of an elicited prior

CLASSES OF PRIORS

Neighborhood classes

- $\Gamma_\varepsilon = \{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$ *(ε–contaminations)*
  - Proposed by Huber in classical robustness to model outliers
  - $\mathcal{Q}$: all, all symmetric, all symmetric unimodal, generalized moments constraints class, etc.
  - $\epsilon = \epsilon(\theta)$ (need to normalize!)

CLASSES OF PRIORS

Neighborhood classes

- $\Gamma^{DB} = \{F \text{ c.}d.f. : F_0(\theta) - \epsilon \leq F(\theta) \leq F_0(\theta) + \epsilon, \forall\theta\}$ *(Distribution bounded class)*

- $\Gamma^{T}_{\varepsilon} = \{P : \sup_{A \in \mathcal{F}} |P(A) - P_0(A)| \leq \varepsilon\}$ *(Total variation)*

- $K_g = \{P : \varphi_P(x) \geq g(x), \forall x \in [0,1]\}$
  g nondecreasing, continuous, convex: $g(0) = 0$ and $g(1) \leq 1$
  (Concentration function class)

CLASSES OF PRIORS

Some critical issues

- Many classes driven more by mathematical convenience rather than ease of elicitation

- Range easily computed for some *useless* classes (e.g. ϵ-contaminations with all probability measures) but ...

- ... hard to compute for some *meaningful* classes (e.g. unimodal generalized moments constrained class)

NEAR IGNORANCE

- Improper priors
- Uniform distribution on *large* interval (for unbounded Θ)
- Neighborhood of uniform distribution
- Bayesian nonparametrics (e.g. Dirichlet process) centered at a uniform distribution
- Imprecise probabilities
- Frequentist approach

CLASSES OF MODELS

Finite classes (Shyamalkumar, 2000)

- Class $\mathcal{M} = \{\mathcal{N}(\theta, 1), \mathcal{C}(\theta, 0.675)\}$
  (same median and interquartile range)

- $\pi_0(\theta) \sim \mathcal{N}(0, 1)$ baseline prior

- $\Gamma^{A}_{0.1} = \{\pi : \pi = 0.9\pi_0 + 0.1q, q \text{ arbitrary}\}$

- $\Gamma^{SU}_{0.1} = \{\pi : \pi = 0.9\pi_0 + 0.1q, q \text{ symmetric unimodal around zero}\}$

- Interest in $\mathcal{E}(\theta|x)$

CLASSES OF MODELS

Finite classes (Shyamalkumar, 2000)

Data	Likelihood	$\Gamma^{A}_{0.1}$		$\Gamma^{SU}_{0.1}$	
		inf $\mathbf{E}(\theta\|x)$	sup $\mathbf{E}(\theta\|x)$	inf $\mathbf{E}(\theta\|x)$	sup $\mathbf{E}(\theta\|x)$
$x = 2$	Normal	0.93	1.45	0.97	1.12
	Cauchy	0.86	1.38	0.86	1.02
$x = 4$	Normal	1.85	4.48	1.96	3.34
	Cauchy	0.52	3.30	0.57	1.62
$x = 6$	Normal	2.61	8.48	2.87	5.87
	Cauchy	0.20	5.54	0.33	2.88

CLASSES OF MODELS

Parametric models

Box-Tiao, 1962

$$\Lambda_{BT} = \left\{ f(y|\theta,\sigma,\beta) = \frac{\exp\left\{-\frac{1}{2}\left|\frac{y-\theta}{\sigma}\right|^{\frac{2}{1+\beta}}\right\}}{\sigma 2^{(1.5+0.5\beta)}\Gamma(1.5+0.5\beta)} \right\}$$

for any $\theta, \sigma > 0, \beta \in (-1, 1]$

CLASSES OF MODELS

Neighborhood classes

$0 \leq M(\cdot) \leq U(\cdot)$ given and l likelihood

- $\Gamma_\epsilon = \{f : f(x|\theta) = (1-\epsilon)f_0(x|\theta) + (1-\epsilon)g(x|\theta), g \in \mathcal{G}\}$
  (ϵ–contaminations)

- $\Gamma_{DR} = \{f : \exists\alpha \text{ s.t. } M(x-\theta_0) \leq \alpha f(x|\theta_0) \leq U(x-\theta_0)\forall x\}$
  (density ratio class)

- $\Gamma_L = \{l : M(\theta) \leq l(\theta) \leq U(\theta)\}$
  (likelihood neighborhood)

Critical aspects: parameter and probabilistic interpretation

CLASSES OF MODELS

Weighted distribution classes

- $f(x|\theta) \propto \omega(x) f_0(x|\theta)$, $\omega \in \Omega$

- $\Omega_1 = \{\omega : \omega_1(x) \leq \omega(x) \leq \omega_2(x)\}$

- $\Omega_2 = \{\text{nondecreasing } \omega_1(x) \leq \omega(x) \leq \omega_2(x)\}$

Critical aspect: need to normalize $f(x|\theta)$

CLASSES OF LOSSES

Interest in behavior of

- Bayesian estimator
- posterior expected loss

CLASSES OF LOSSES

Parametric classes $\mathcal{L}_\omega = \{L = L_\omega, \omega \in \Omega\}$

$L(\Delta) = \beta(\exp\{\alpha\Delta\} - \alpha\Delta - 1), \alpha \neq 0, \beta > 0$

- $\Delta_1 = (a - \theta) \Rightarrow L(\Delta_1)$ LINEX (Varian, 1975)
  - $\alpha = 1 \Rightarrow L(\Delta_1)$ asymmetric
    (*overestimation worse than underestimation*)
  - $\alpha < 0$
    $\Rightarrow L(\Delta_1) \approx$ exponential for $\Delta_1 < 0$
    $\Rightarrow L(\Delta_1) \approx$ linear for $\Delta_1 > 0$
  - $|\alpha| \approx 0 \Rightarrow L(\Delta_1) \approx \beta\alpha^2\Delta_1^2/2$ (i.e. squared loss)
- $\Delta_2 = (a/\theta - 1)$ (Basu and Ebrahimi, 1991)

CLASSES OF LOSSES

- $\mathcal{L}_U = \{L : L(\theta, a) = L(|\theta - a|), L(\cdot) \text{ any nondecreasing function}\}$

  (Hwang's universal class)

- $\mathcal{L}_\epsilon = \{L : L(\theta, a) = (1 - \epsilon)L_0(\theta, a) + \epsilon M(\theta, a),\ M \in \mathcal{W}\}$

  (ϵ–contamination class)

- $\mathcal{L}_K = \{L : v_{i-1} \leq L(c) \leq v_i,\ \forall c \in C_i,\ i = 1, \ldots, n\}$
  - $(\theta, a) \to c \in \mathcal{C}$ (consequence), e.g. $c = |\theta - a|$
  - $\{C_1, \ldots, C_n\}$ partition of $\mathcal{C}$

  (Partially known class)

$L, L + k \in \mathcal{L}_U$ give same Bayesian estimator minimizing the posterior expected loss, but very different posterior expected loss $\Rightarrow$ robustness calibration

SENSITIVITY MEASURES

Global sensitivity

- Class of priors sharing some features (e.g. quantiles, moments)
- No prior plays a relevant role w.r.t. others

Measures

- Range: $\delta = \overline{\rho} - \underline{\rho}$, with $\overline{\rho} = \sup_{P \in \Gamma} E_{P^*}[h(\theta)]$ and $\underline{\rho} = \inf_{P \in \Gamma} E_{P^*}[h(\theta)]$
  Simple interpretation

- Relative sensitivity $\sup_\pi R_\pi$, with $R_\pi = \frac{(\rho_\pi - \rho_0)^2}{V^\pi}$, $\rho_0 = E_{\Pi_0^*}[h(\theta)]$, $\rho_\pi = E_{\Pi^*}[h(\theta)]$ and $V^\pi = Var_{\Pi^*}[h(\theta)]$
  Scale invariant, decision theoretic interpretation, asymptotic behavior

SENSITIVITY MEASURES

Local sensitivity

- Small changes in one elicited prior
- Most influential x
- Approximating bounds for global sensitivity

Measures

- Derivatives of extrema in $\{K_\varepsilon\}, \varepsilon \geq 0$, neighborhood of $K_0 = \{P_0\}$

$$\overline{E}_\varepsilon(h|x) = \frac{\int h(\theta)l(\theta)P(d\theta)}{\int l(\theta)P(d\theta)} \text{ and } D^*(h) = \left\{\frac{\partial \overline{E}_\varepsilon(h|x)}{\partial \varepsilon}\right\}_{\varepsilon=0}$$

- Gatêaux differential

SENSITIVITY MEASURES

Measures

- Fréchet derivative
  - $\Delta = \{\delta : \delta(\Theta) = 0\}$
  - $\Gamma_\delta = \{\pi : \pi = P + \delta, \delta \in \Delta\}$ and $\Gamma_\varepsilon = \{\pi : \pi = (1-\varepsilon)P + \varepsilon Q\}$
  - $\mathcal{P} = \{\delta \in \Delta : \delta = \varepsilon(Q - P)\} \Rightarrow \Gamma_\varepsilon \subset \Gamma_\delta$
  - $||\delta|| = d(\delta, 0)$
  - $d(P, Q) = \sup_{A \in \mathcal{B}(\Theta)} |P(A) - Q(A)|$
  - $T_h(P + 0) \equiv T_h(P) \equiv \dfrac{\int h(\theta) l(\theta) P(d\theta)}{\int l(\theta) P(d\theta)} = \dfrac{N_P}{D_P}$
  - $\Lambda_h^P(\delta) = T_h(P + \delta) - T_h(P) + o(||\delta||) = \dfrac{D_\delta}{D_P}(T_h(\delta) - T_h(P))$

SENSITIVITY MEASURES

Loss robustness

$\rho_L(\pi, x, a) = \mathcal{E}^{\pi(\cdot|x)} L(\theta, a) = \int L(\theta, a)\pi(\theta|x) d\theta$
posterior expected loss minimized by a_π^L

- $\text{sup}_{L \in \mathcal{L}}\, \rho_L(\pi, x, a) - \text{inf}_{L \in \mathcal{L}}\, \rho_L(\pi, x, a)$

- $\text{sup}_{L \in \mathcal{L}}\, a_\pi^L - \text{inf}_{L \in \mathcal{L}}\, a_\pi^L$

- $\text{sup}_x \left| \frac{\partial}{\partial x} \rho_L(\pi, x, a_\pi^L) \right| - \text{inf}_x \left| \frac{\partial}{\partial x} \rho_L(\pi, x, a_\pi^L) \right|$

COMPUTATIONAL TECHNIQUES

Bayesian inference $\Rightarrow$ complex computations
Robust Bayesian inference $\Rightarrow$ **more** complex computations

$$\sup_P \frac{\int_\Theta f(\theta)P(d\theta)}{\int_\Theta g(\theta)P(d\theta)} = \sup_{\theta\in\Theta} \frac{f(\theta)}{g(\theta)}$$

$$\Rightarrow \overline{\rho} = \sup_{P \in \Gamma} E_{P^*}[h(\theta)]$$

Probability measures as mixture of extremal ones

- $\Gamma_\varepsilon = \{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}_A\} \rightarrow$ Dirac
- $\Gamma_Q = \{P : P(I_i) = p_i, i = 1, \ldots, m\} \rightarrow$ Discrete
- $\Gamma_{SU} = \{P : \text{symmetric, unimodal}\} \rightarrow$ Uniform

COMPUTATIONAL TECHNIQUES

- Linearization technique to compute $\sup_{P \in \Gamma} \frac{\int_\Theta h(\theta) l(\theta) P(d\theta)}{\int_\Theta l(\theta) P(d\theta)}$
  - $\overline{\rho} = \inf\{q | c(q) = 0\}$ where
  - $c(q) = \sup_{P \in \Gamma} \int_\Theta c(\theta, q) P(d\theta)$
  - $c(\theta, q) = l(\theta)\ (h(\theta) - q)$
  - Compute $c(q_i)$, $i = 1, \ldots, m \Rightarrow$ solve $c(q) = 0$
- Discretization of $\Theta \Rightarrow$ Linear programming
- Linear Semi-infinite Programming (for Generalized moments constrained classes)
- Importance sampling

QUEST FOR ROBUSTNESS

Range δ “large” and possible refinement of Γ

- Further elicitation by experts
  - Software (currently unavailable) for interactive sensitivity analysis
  - Ad-hoc tools, e.g. Fréchet derivatives to determine intervals to split in quantile classes
- Acquisition of new data

QUEST FOR ROBUSTNESS

Inherently robust procedures

- Robust priors (e.g. flat-tailed)
- Robust models (e.g. Box-Tiao class)
- Robust estimators
- Hierarchical models
- Bayesian nonparametrics

LACK OF ROBUSTNESS

Range δ “large” and no further possible refinement of Γ

- Choice of a convenient prior in Γ, e.g. a Gaussian in the symmetric, unimodal quantile class, or

- Choice of an estimate of $E_{P^*}[h(\theta)]$ according to an optimality criterion, e.g.

  - Γ–minimax posterior expected loss

  - Γ–minimax posterior regret

- Report the range of $E_{P^*}[h(\theta)]$ besides the entertained value

GAMMA-MINIMAX

$\rho(\pi, a) = E^{\pi^*} L(\theta, a)$ posterior expected loss, minimized by a_π

- $\rho_C = \inf_{a \in \mathcal{A}} \sup_{\pi \in \Gamma} \rho(\pi, a)$
  (Posterior Γ-minimax expected loss)

  Optimal action by interchanging inf and sup for convex losses

- $\rho_R = \inf_{a \in \mathcal{A}} \sup_{\pi \in \Gamma} [\rho(\pi, a) - \rho(\pi, a_\pi)]$
  (Posterior Γ-minimax regret)

  Optimal action: $a_M = \frac{1}{2}(\underline{a} + \bar{a})$, for finite $\underline{a} = \inf_{\pi \in \Gamma} a_{\pi_x}$ and $\bar{a} = \sup_{\pi \in \Gamma} a_{\pi_x}$, $\mathcal{A}$ interval and $L(\theta, a) = (\theta - a)^2$

REFERENCES ON BAYESIAN STATISTICS

Introductory

- *Berry*, Duxbury
- *Lee*, Arnold

General

- *Bernardo and Smith*, Wiley
- *Congdon*, Wiley
- *Robert*, Springer

Nonparametrics

- *Ghosh and Ramamoorthy*, Springer

Robustness

- *Rios Insua and Ruggeri*, Springer

REFERENCES ON BAYESIAN STATISTICS

Applied

- *Gelman, Carlin, Stern and Rubin*, Chapman and Hall
- *Congdon*, Wiley

MCMC

- *Gilks, Richardson and Spiegelhalter*, Chapman and Hall
- *Robert and Casella*, Springer
- *Gamerman and Lopes*, Chapman and Hall/CRC

Stochastic Processes

- *Rios Insua, Ruggeri and Wiper*, Wiley

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/˜fabrizio/

OUTLINE

- Basics on Stochastic Processes
- Discrete Time Markov Chains
- Bayesian Inference for DTMC
- Approximation through DTMC

BASICS ON STOCHASTIC PROCESSES

- A stochastic process $\{X_t, t \in T\}$ is a collection of random variables X_t, indexed by a set T, defined on a sample space Ω, endowed with a σ-algebra $\mathcal{F}$ and a base probability measure $\mathbf{P}$, and taking values in a common measurable space S endowed with an appropriate σ-algebra

- T set of

  - times $\Rightarrow$ temporal stochastic process
    (major focus of the course, with T *space of times*)

  - spatial coordinates $\Rightarrow$ spatial process

  - both time and spatial coordinates $\Rightarrow$ spatio-temporal process

- T discrete $\Rightarrow$ process *in discrete time*, represented through $\{X_n,\ n = 0, 1, 2, ..\}$

- T continuous $\Rightarrow$ process in *continuous time*, with $T = [0, \infty)$ usually in the next

- Values taken by process $\Rightarrow$ *states* of the process, belonging to the *state space* S, which may be either discrete or continuous

BASICS ON STOCHASTIC PROCESSES

- At least two visions of a stochastic process
  - For each $\omega \in \Omega$, $X_t(\omega)$, function of t as realization or sample function of the stochastic process, describes a possible evolution of the process through time
  - For any given t, X_t is a random variable
- Kolmogorov extension theorem $\Rightarrow$ stochastic process completely described providing joint distribution $P\left(X_{t_1} \leq x_1, ..., X_{t_n} \leq x_n\right)$ for any $\{t_1, ..., t_n\}$ with $t_1 < \cdots < t_n$, and for any $\{x_1, \ldots, x_n\}$
- Let $T \subseteq [0, \infty)$. Suppose that, for any $\{t_1, ..., t_n\}$ with $t_1 < \cdots < t_n$,, the random variables $X_{t_1}, ..., X_{t_n}$ satisfy the following consistency conditions:
  1. For all permutations π of $1, \ldots, n$ and $x_1, \ldots, x_n$, we have $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = P(X_{t_{\pi(1)}} \leq x_{\pi(1)}, \ldots, X_{t_{\pi(n)}} \leq x_{\pi(n)})$
  2. For all $x_1, \ldots, x_n$ and $t_{n+1}, \ldots, t_{n+m}$, we have $P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n) = P(X_{t_1} \leq x_1, \ldots, X_{t_n} \leq x_n, X_{t_{n+1}} < \infty, ..., X_{t_{n+m}} < \infty)$

  Then, there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a stochastic process $X_t : T \times \Omega \to \mathbb{R}^n$ having the families $X_{t_1}, ..., X_{t_n}$ as finite-dimensional distributions

BASICS ON STOCHASTIC PROCESSES

$\{X_t, t \in T\}$ stochastic process

- Mean function: $\mu_X(t) = E[X_t]$
- Autocorrelation function: $R_X(t_1, t_2) = E[X_{t_1} X_{t_2}]$
- Autocovariance function: $C_X(t_1, t_2) = E[(X_{t_1} - \mu_X(t_1))(X_{t_2} - \mu_X(t_2))]$
- $\{X_t, t \in T\}$ strictly stationary if $(X_{t_1}, \ldots, X_{t_n})$ has the same distribution as $(X_{t_1+\tau}, \ldots, X_{t_n+\tau})$ for any n, $t_1, t_2, \ldots, t_n$ and τ
  - Stationarity typical feature of a system which stabilized its behavior after running for a long time
- Strict stationarity holds, in particular, for
  - $n = 1 \Rightarrow X_t$'s have the same distribution
  - $n = 2 \Rightarrow$ joint distribution depends on difference between times and not the times themselves, i.e. $F_{X_{t_1}, X_{t_2}}(x_1, x_2) = F_{X_0, X_{t_2 - t_1}}(x_1, x_2)$

BASICS ON STOCHASTIC PROCESSES

- $\{X_t, t \in T\}$ strictly stationary stochastic process
  1. Constant mean function: $\mu_X(t) = \mu_X, \forall t$
  2. Autocorrelation function depends on time differences: $R_X(t_1, t_2) = R(t_2 - t_1)$
  3. Autocovariance function: $C_X(t_1, t_2) = R(t_2 - t_1) - \mu_X^2$ (if all moments exist)
- Process fulfilling 1) and 2) $\Rightarrow$ weakly stationary process
- Weak stationarity $\not\Rightarrow$ strict stationarity, in general
- Strict stationarity $\Rightarrow$ weak stationarity, if first and second moments exist

BASICS ON STOCHASTIC PROCESSES

Transition behavior $\Rightarrow$ forecasting at future times given past observations

- *Conditional transition distribution function* defined, for $t_0 \leq t_1$, by $F(x_0, x_1; t_0, t_1) = P(X_{t_1} \leq x_1 \mid X_{t_0} \leq x_0)$
- Process discrete in time and space $\Rightarrow$ transition probabilities $P_{ij}^{(m,n)} = P(X_n = j \mid X_m = i)$ for $m \leq n$
- Stationary process $\Rightarrow$ transition distribution function depends on time differences $t = t_1 - t_0$, i.e. $F(x_0, x; t_0, t_0 + t) = F(x_0, x; 0, t)\,,\ \forall t_0 \in T$
- For convenience, we will use the notation as $F(x_0, x; t)$ and $P_{ij}^{(n)}$ for stationary processes
- Letting $t \to \infty$, we may consider the long term limiting behavior of the process, typically associated with the stationary distribution. When this distribution exists, computations are usually much simpler than doing short term predictions based on the use of the transition functions.

MARKOVIAN PROCESSES

Except for the case of independence, the simplest dependence form among the random variables in a stochastic process is the Markovian one

- Consider a set of time instants $\{t_0, t_1, \ldots, t_n, t\}$ with $t_0 < t_1 < \cdots < t_n < t$ and $t, t_i \in T$. A stochastic process $\{X_t, t \in T\}$ is Markovian if the distribution of X_t conditional on the values of $X_{t_1}, \ldots, X_{t_n}$ depends only on X_{t_n}, that is, the most recent known value of the process

$$P\left(X_t \leq x \mid X_{t_n} \leq x_n, X_{t_{n-1}} \leq x_{n-1}, ..., X_{t_0} \leq x_0\right) = \\ = P\left(X_t \leq x \mid X_{t_n} \leq x_n\right) = F\left(x_n, x; t_n, t\right)$$

- As a consequence of the previous relation, we have

$$F\left(x_0, x; t_0, t_0 + t\right) = \int_{y \in S} F\left(y, x; \tau, t\right) dF\left(x_0, y; t_0, \tau\right) \tag{1}$$

with $t_0 < \tau < t$.

MARKOVIAN PROCESSES

- If the stochastic process is discrete in both time and space, then, for $n > n_1 > \cdots > n_k$, we have

$$\begin{aligned} P\left(X_n = j \mid X_{n_1} = i_1, X_{n_2} = i_2, ..., X_{n_k} = i_{n_k}\right) &= \\ P\left(X_n = j \mid X_{n_1} = i_1\right) &= p_{i_1 j}^{(n_1, n)}. \end{aligned}$$

- Using this property and taking r such that $m < r < n$, we have

$$\begin{aligned} p_{ij}^{(m,n)} &= P\left(X_n = j \mid X_m = i\right) \\ &= \sum_{k \in S} P\left(X_n = j \mid X_r = k\right) P\left(X_r = k \mid X_m = i\right). \end{aligned} \quad (2)$$

- Equations (1) and (2) are called the *Chapman-Kolmogorov equations* for the continuous and discrete cases, respectively

- We refer to discrete state space Markov processes as Markov chains and will use the term Markov process to refer to processes with continuous state spaces and the Markovian property

DISCRETE TIME MARKOV CHAINS

- One step *transition probability*: $p_{ij}^{(m,m+1)} = P\left(X_{m+1} = j \mid X_m = i\right)$
- $p_{ij}^{(m,m+1)}$ independent of $m \Rightarrow$ stationary process and *time homogeneous* chain
- Using $p_{ij} = P\left(X_{m+1} = j \mid X_m = i\right)$ and $p_{ij}^n = P\left(X_{m+n} = j \mid X_m = i\right), \forall m$
  $\Rightarrow$ Chapman-Kolmogorov equations $p_{ij}^{n+m} = \sum_{k \in S} p_{ik}^n p_{kj}^m \forall n, m \geq$ 0 and i, j
- n*-step transition probability matrix* defined as $\mathbf{P}^{(n)}$, with elements p_{ij}^n
  $\Rightarrow$ Chapman-Kolmogorov equations written $\mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}$
- Matrices $\mathbf{P}^{(n)}$ fully characterize the transition behavior of an homogeneous Markov chain
- When $n = 1$, we refer to the *transition matrix* instead of the one step transition matrix and write $\mathbf{P}$ instead of $\mathbf{P}^{(1)}$

GAMBLER'S RUIN PROBLEM

A gambler with an initial stake, $x_0 \in \mathbb{N}$, plays a coin tossing game where at each turn, if the coin comes up heads, she wins a unit and if the coin comes up tails, she loses a unit. The gambler continues to play until she either is bankrupted or her current holdings reach some fixed amount m. Let X_n represent the amount of money held by the gambler after n steps. Assume that the coin tosses are independent and identically distributed with probability of heads p at each turn. Then, $\{X_n\}$ is a time homogeneous Markov chain with $p_{00} = p_{mm} = 1$, $p_{ii+1} = p$ and $p_{ii-1} = 1 - p$, for $i = 1, \ldots, m-1$ and $p_{ij} = 0$ for $i \in \{1, \ldots, m-1\}$ and $j \neq i-1, i+1$.

DISCRETE TIME MARKOV CHAINS

The analysis of the stationary behavior of an homogeneous Markov chain requires studying the relations among states as follows

- A state j is *reachable* from a state i if $p_{ij}^n > 0$, for some n. We say that two states that are mutually reachable *communicate* and belong to the same *communication class.*

- If all states in a chain communicate among themselves, so that there is just one communication class, we shall say that the Markov chain is irreducible

- In the case of the Gambler's ruin problem, we can see that there are three communication classes: $\{0\}, \{1, \ldots, m-1\}$ and $\{m\}$.

- Given a state i, let p_i be the probability that, starting from state i, the process returns to such state. We say that state i is *recurrent* if $p_i = 1$ and *transient* if $p_i < 1$.

- We may easily see that if state i is recurrent and communicates with another state j, then j is recurrent.

- In the case of Gambler's ruin, only the states $\{0\}$ and $\{m\}$ are recurrent.

DISCRETE TIME MARKOV CHAINS

- A state i has *period* k if $p_{ii}^n = 0$ whenever n is not divisible by k and k is the biggest integer with this property. A state with period 1 is *aperiodic*.

- If i has period k and states i and j communicate, then state $j$has period k. In the Gambler's problem, states $\{0, m\}$ are aperiodic and the others have period 2.

- State i *positive recurrent* if, starting at i, the expected time until return to i is finite

- Positive recurrence is also a class property in the sense that, if i is positively recurrent and states i and j communicate, then state j is also positively recurrent

- In a Markov chain with a finite number of states all recurrent states are positive recurrent

- A positive recurrent, aperiodic state is called ergodic

- Ergodic and irreducible Markov chain $\Rightarrow \pi_j = \lim_{n\to\infty} p_{ij}^n$ independent of i and unique nonnegative solution of $\pi_j = \sum_i \pi_i p_{ij}, j \geq 0$, with $\sum_{i=0}^{\infty} \pi_i = 1$

INFERENCE FOR FINITE, TIME HOMOGENEOUS MARKOV CHAINS

- Transition matrix $\mathbf{P} = (p_{ij})$ where $p_{ij} = P(X_n = j | X_{n-1} = i)$, for states $i, j \in \{1, \dots, K\}$

- If it exists, stationary distribution $\boldsymbol{\pi}$ unique solution of $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, $\pi_i \geq 0, \sum \pi_i = 1$

- Initially, we consider the simple experiment of observing m successive transitions of the Markov chain, say $X_1 = x_1, \dots, X_m = x_m$, given a known initial state $X_0 = x_0$

- Likelihood function $l(\mathbf{P}|\mathbf{x}) = \prod_{i=1}^K \prod_{j=1}^K p_{ij}^{n_{ij}}$ with $n_{ij} \geq 0$ number of observed transitions from state i to state j and $\sum_{i=1}^K \sum_{j=1}^K n_{ij} = m$

- $\widehat{\mathbf{P}}$ MLE for $\mathbf{P}$, with $\widehat{p}_{ij} = \dfrac{n_{ij}}{n_{i\cdot}}$, where $n_{i\cdot} = \sum_{j=1}^K n_{ij}$

- However, especially in chains with large K, there could be some $\widehat{p}_{ij} = 0$

BAYESIAN INFERENCE

- Conjugate prior for $\mathbf{P}$ defined by letting $\mathbf{p}_i = (p_{i1}, \ldots, p_{iK}) \sim$ Dir$(\boldsymbol{\alpha}_i)$, where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iK})$ for $i = 1, \ldots, K$
  $\Rightarrow$ matrix beta prior distribution

- $\Rightarrow$ posterior $\mathbf{p}_i|\mathbf{x} \sim$ Dir$\big(\boldsymbol{\alpha}'_i\big)$ where $\alpha'_{ij} = \alpha_{ij} + n_{ij}$ for $i, j = 1, \ldots, K$

- Scarce prior information $\Rightarrow$ Jeffreys prior, i.e. matrix beta prior with $\boldsymbol{\alpha}_{ij} = 1/2$ for all $i, j = 1, \ldots, K$

- Another improper prior, along lines of Haldane's for binomial data, is $f(\mathbf{p}_i) \propto \prod_{j=1}^{K} \frac{1}{p_{ij}}$, i.e. a matrix beta prior with $\alpha_{ij} \to 0$ for all $i, j = 1, \ldots, K$
  - Posterior distribution $\mathbf{p}_i|\mathbf{x} \sim$ Dir$(n_{i1}, \ldots, n_{ik})$ would imply a posterior mean $E[p_{ij}|\mathbf{x}] = n_{ij}/n_{i\cdot}$ equal to MLE
  - Improper posterior distribution if there are any $n_{ij} = 0$

- A proper Dirichlet prior

SYDNEY BOTANIC GARDENS WEATHER CENTER

Rainfall levels (from `weatherzone.com.au`) illustrate occurrence (2) or non occurrence (1) of rain between February 1st and March 20th 2008. The data are to be read consecutively from left to right. Thus, it rained on February 1st and did not rain on March 20th.

2	2	2	2	2	2	2	2	2	2
1	1	2	1	1	1	1	1	1	1
2	2	1	1	1	1	2	2	2	1
2	1	1	1	1	1	2	1	1	1
1	1	1	1	1	1	1	1	1	

The daily occurrence of rainfall is modeled as a Markov chain with transition matrix

$$\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}.$$

Given a Jeffreys prior, $p_{ii} \sim \text{Be}(1/2, 1/2)$, for $i = 1, 2$, then conditioning on the occurrence of rainfall on February 1st, the posterior distribution is

$$p_{11}|\mathbf{x} \sim \text{Be}(25.5, 5.5) \quad p_{22}|\mathbf{x} \sim \text{Be}(12.5, 6.5).$$

The expectation of the transition matrix is $E[\mathbf{P}|\mathbf{x}] = \begin{pmatrix} 0.823 & 0.177 \\ 0.342 & 0.658 \end{pmatrix}$

DISCRETE TIME MARKOV CHAINS

- Sometimes it may be known that certain transitions are impossible a priori, e.g. remain in a state, so that $p_{ii} = 0$ for $i = 1, \ldots, K$

- $\Rightarrow$ restrict matrix-beta prior to the space of transitions with positive probability and set others to zero $\Rightarrow$ conjugate analysis

- More interesting, but less studied: unknown, a priori, which transitions are possible and which are impossible so that the chain may be periodic or transient

- In this situation, one possibility is to define a hierarchical prior distribution by first setting the probabilities that different transitions are impossible

DISCRETE TIME MARKOV CHAINS

- $P(p_{ij} = 0|q) \propto q$ for $i, j \in 1, \ldots, K$
  - restricted so that $P(\mathbf{p}_i = \mathbf{0}|q) = 0$
  - $\Rightarrow$ row i of transition matrix contains, e.g., exactly r_i zeros at locations $j_1, \ldots, j_{r_i}$ and $K - r_i$ ones at locations $j_{r_i+1}, \ldots, j_K$
  - $\Rightarrow$ prior probability (conditional on q): $\dfrac{q^{r_i}(1-q)^{K-r_i}}{1-q^K}$ for $r = 0, 1, \ldots, K-1$
- $q \sim \text{U}(0, 1)$
- Dirichlet priors for the vectors of nonnull transition probabilities, e.g.

$$(p_{ij_{r_i+1}}, \ldots, p_{ij_K}) \sim \text{Dir}\left(\underbrace{\frac{1}{2}, \ldots, \frac{1}{2}}_{K-r_i}\right).$$

DISCRETE TIME MARKOV CHAINS

- $\mathbf{Z}$ random $K \times K$ matrix s.t. $Z_{ij} = 0$ if $p_{ij} = 0$ and $Z_{ij} = 1$ o.w.
- $\mathbf{z}$ matrix with i-th row of $\mathbf{z}$ with r_i zeros in positions $j_1, \ldots, j_{r_i}$ for $i = 1, \ldots, K$
- Then, the posterior probability that $\mathbf{Z}$ is equal to $\mathbf{z}$ can be evaluated as

$$\begin{aligned}
P(\mathbf{Z} = \mathbf{z}|\mathbf{x}) &\propto f(\mathbf{x}|\mathbf{z})P(\mathbf{z}) \\
&\propto \int f(\mathbf{x}|\mathbf{z}, \mathbf{P}) f(\mathbf{P}|\mathbf{z})\, d\mathbf{P} \int_0^1 P(\mathbf{Z} = \mathbf{z}|q) f(q)\, dq \\
&\propto \frac{1}{\Gamma\left(\frac{1}{2}\right)^{K^2 - K\bar{r}}} \prod_{i=1}^{K} \frac{\Gamma\left(\frac{K - r_i}{2}\right)}{\Gamma\left(\frac{K - r_i}{2} + \sum_{s=r_i+1}^{K} n_{ij_s}\right)} \times \\
&\quad \prod_{s=r_i+1}^{K} \Gamma\left(\frac{1}{2} + n_{ij_s}\right) \int_0^1 \frac{q^{K\bar{r}}(1-q)^{K(K-\bar{r})}}{(1-q^K)^K}\, dq,
\end{aligned}$$

where the probability is positive over the range $n_{ij_1}, \ldots, n_{ij_{r_i}} = 0$ for $i = 1, \ldots, K$

DISCRETE TIME MARKOV CHAINS

- For relatively small dimensional transition matrices, this probability may be evaluated directly, but for Markov chains with a large number of states and many values $n_{ij} = 0$, exact evaluation will be impossible. In such cases, it would be preferable to employ a sampling algorithm over values of $\mathbf{Z}$ with high probability.

- The posterior probability that the chain is periodic could then be evaluated by simply summing those $P(\mathbf{Z} = \mathbf{z}|\mathbf{x})$ where $\mathbf{z}$ is equivalent to a periodic transition matrix.

FORECASTING SHORT TERM BEHAVIOR

- Suppose that we wish to predict future values of the chain
- For example we can predict the next value of the chain, at time $n+1$, using

$$\begin{aligned} P(X_{n+1}=j|\mathbf{x}) &= \int P(X_{n+1}=j|\mathbf{x},\mathbf{P}) f(\mathbf{P}|\mathbf{x})\, d\mathbf{P} \\ &= \int p_{x_n j} f(\mathbf{P}|\mathbf{x})\, d\mathbf{P} = \frac{\alpha_{x_n j}+n_{x_n j}}{\alpha_{x_n \cdot}+n_{x_n \cdot}} \end{aligned}$$

  where $\alpha_{i\cdot} = \sum_{j=1}^{K} \alpha_{ij}$.

- Prediction of state at $t>1$ steps is slightly more complex. For small t, use

$$P(X_{n+t}=j|\mathbf{x}) = \int \left(\mathbf{P}^t\right)_{x_n j} f(\mathbf{P}|\mathbf{x})\, d\mathbf{P}$$

  which gives a sum of Dirichlet expectation terms. However, as t increases, the evaluation of this expression becomes computationally infeasible.

- A simple alternative is to use a Monte Carlo algorithm based on simulating future values of the chain

FORECASTING SHORT TERM BEHAVIOR

- For $s = 1, \ldots, S$:

  Generate $\mathbf{P}^{(s)}$ from $f(\mathbf{P}|\mathbf{x})$.

  Generate $x_{n+1}^{(s)}, \ldots, x_{n+t}^{(s)}$ from the Markov chain with $\mathbf{P}^{(s)}$ and initial state x_n.

- Then, $P(X_{n+t} = j|\mathbf{x}) \approx \frac{1}{S}\sum_{s=1}^{S} I\left(x_{n+t}^{(s)} = j\right)$ where $I(\cdot)$ is an indicator function and $E[X_{n+t}|\mathbf{x}] \approx \frac{1}{S}\sum_{s=1}^{S} x_{n+t}^{(s)}$.

- Assume that it is now wished to predict the Sydney weather on the 21st and 22nd of March. Given that it did not rain on the 20th March, then immediately, we have

$$
\begin{aligned}
P(\text{no rain on 21st March}|\mathbf{x}) &= E[p_{11}|\mathbf{x}] = 0.823, \\
P(\text{no rain on 22nd March}|\mathbf{x}) &= E\left[p_{11}^2 + p_{12}p_{21}|\mathbf{x}\right] = 0.742, \\
P(\text{no rain on both}) &= E[p_{11}^2|\mathbf{x}] = 0.681.
\end{aligned}
$$

FORECASTING STATIONARY BEHAVIOR

- Interest in stationary distribution $\Rightarrow$ straightforward for low dimensional chain where the exact formula for the equilibrium probability distribution can be derived

- Suppose that $K = 2$ and $\mathbf{P} = \begin{pmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{pmatrix}$
  - Equilibrium probability of being in state 1: $\pi_1 = \dfrac{1 - p_{22}}{2 - p_{11} - p_{22}}$
  - Predictive equilibrium distribution: $E[\pi_1|\mathbf{x}] = \displaystyle\int_0^1 \int_0^1 \frac{1 - p_{22}}{2 - p_{11} - p_{22}} f(p_{11}, p_{22}|\mathbf{x})\, dx$

- In the rainfall example $E[\pi_1|\mathbf{x}] = E\left[\dfrac{1 - p_{22}}{2 - p_{11} - p_{22}} \middle| \mathbf{x} \right] = 0.655 \Rightarrow 65\%$ (approx) without rain at the center

FORECASTING STATIONARY BEHAVIOR

- For higher dimensional chains, it is simpler to use a Monte Carlo approach as earlier so that given a Monte Carlo sample $\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(S)}$ from the posterior distribution of $\mathbf{P}$, then the equilibrium distribution can be estimated as

$$E[\boldsymbol{\pi}|\mathbf{x}] \approx \frac{1}{S} \sum_{s=1}^{S} \boldsymbol{\pi}^{(s)}$$

where $\boldsymbol{\pi}^{(s)}$ is the stationary distribution associated with the transition matrix $\mathbf{P}^{(s)}$

MODEL COMPARISON

- Bayes factor to test if data are independent or generated from a Markov chain

- Compare earlier Markov chain model ($\mathcal{M}_1$) with i.i.d. data with distribution $\mathbf{q} = (q_1, \ldots, q_K)$, ($\mathcal{M}_2$) with Dirichlet prior $\mathbf{q} \sim$ Dir$(a_1, \ldots, a_K)$

- Let $n_{i\cdot} = \sum_{j=1}^{K} n_{ij}$ and $\alpha_{i\cdot} = \sum_{j=1}^{K} \alpha_{ij}$, then

$$\begin{aligned} f(\mathbf{x}|\mathcal{M}_1) &= \int f(\mathbf{x}|\mathbf{P}) f(\mathbf{P}|\mathcal{M}_1)\, d\mathbf{P} \\ &= \prod_{i=1}^{K} \frac{\Gamma(\alpha_{i\cdot})}{\Gamma(n_{i\cdot} + \alpha_{i\cdot})} \prod_{j=1}^{K} \frac{\Gamma(\alpha_{ij} + n_{ij})}{\Gamma(\alpha_{ij})} \end{aligned}$$

- Let $a = \sum_{i=1}^{K} a_i$ and $n_{\cdot i}$, then

$$f(\mathbf{x}|\mathcal{M}_2) = \frac{\Gamma(a)}{\Gamma(a+n)} \prod_{i=1}^{K} \frac{\Gamma(a_i + n_{\cdot i})}{\Gamma(a_i)}$$

MODEL COMPARISON

- The Bayes factor can now be calculated as the ratio of the two marginal likelihood functions

- For the Australian rainfall data, assuming that the initial state is known and given the Jeffreys prior for the Markov chain model, the marginal likelihood is

$$f(\mathbf{x}|\mathcal{M}_1) = \left(\frac{\Gamma(1)}{\Gamma(1/2)^2}\right)^2 \frac{\Gamma(25.5)\Gamma(5.5)}{\Gamma(31)} \frac{\Gamma(6.5)\Gamma(12.5)}{\Gamma(19)}$$

  and, taking logs, we have $\log f(\mathbf{x}|\mathcal{M}_1) \approx -28.60$.

- For the independent model, $\mathcal{M}_2$, conditional on the initial state and assuming a beta prior, $q_1 \sim \text{Be}(1/2, 1/2)$, we have

$$f(\mathbf{x}|\mathcal{M}_2) = \frac{\Gamma(1)}{\Gamma(1/2)^2} \frac{\Gamma(31.5)\Gamma(17.5)}{\Gamma(49)}$$

  so that $\log f(\mathbf{x}|\mathcal{M}_2) \approx -33.37$, which implies a strong preference for the Markovian model over the independent model

UNKNOWN INITIAL STATE

- Sometimes X_0 is not fixed in advance $\Rightarrow$ define a prior, e.g. multinomial $P(X_0 = x_0|\boldsymbol{\theta}) = \theta_{x_0}$ where $0 < \theta_k < 1$ and $\sum_{k=1}^{K} \theta_k = 1$

- Dirichlet prior $\boldsymbol{\theta} \sim \text{Dir}(\boldsymbol{\gamma})$ for the multinomial parameters $\Rightarrow$ posterior $\boldsymbol{\theta}|\mathbf{x} \sim \text{Dir}(\boldsymbol{\gamma}')$, with$\gamma'_{x_0} = \gamma_{x_0} + 1$ and $\gamma'_i = \gamma_i$ for $i \neq x_0 \Rightarrow$ inference for $\mathbf{P}$ as before

- As an alternative when the chain has been running for some time before the start of the experiment, suppose the initial state generated from the equilibrium distribution, $\boldsymbol{\pi}$, of the Markov chain

- Then, making the dependence of $\boldsymbol{\pi}$ on $\mathbf{P}$ obvious, the likelihood function becomes

$$l(\mathbf{P}|\mathbf{x}) = \pi(x_0|\mathbf{P}) \prod_{i=1}^{K} \prod_{j=1}^{K} p_{ij}^{n_{ij}}$$

UNKNOWN INITIAL STATE

- In this case, simple conjugate inference is impossible but, given the same prior distribution for $\mathbf{P}$ as above, it is straightforward to generate a Monte Carlo sample of size S from the posterior distribution of $\mathbf{P}$ using, for example, a rejection sampling algorithm as follows:

- For $s = 1, \ldots, S$:

  For $i = 1, \ldots, K$, generate $\tilde{\mathbf{p}}_i \sim$ Dir($\boldsymbol{\alpha}'$) with $\boldsymbol{\alpha}'$ with $\alpha'_{ij} = \alpha_{ij} + n_{ij}$

  Set $\tilde{\mathbf{P}}$ to be the transition probability matrix with rows $\tilde{p}_1, \ldots, \tilde{p}_K$.

  Calculate the stationary probability function $\tilde{\boldsymbol{\pi}}$ satisfying $\tilde{\boldsymbol{\pi}} = \tilde{\boldsymbol{\pi}}\tilde{\mathbf{P}}$.

  Generate $u \sim$ U$(0, 1)$. If $u < \tilde{\pi}(x_0)$, set $\mathbf{P}^{(s)} = \tilde{\mathbf{P}}$. O.w. repeat from step 1

- In the rainfall example, assume the weather on February 1st generated from equilibrium distribution. Monte Carlo sample of size 10000 $\Rightarrow$
  $E[\mathbf{P}|\mathbf{x}] \approx \begin{pmatrix} 0.806 & 0.194 \\ 0.321 & 0.679 \end{pmatrix}$ and $E[\pi_1|\mathbf{x}] \approx 0.618$, close to previous results

PARTIALLY OBSERVED DATA

- Assume now that the Markov chain is only observed at a number of finite time points.
- Suppose, for example, that x_0 is a known initial state and that we observe $\mathbf{x}_o = (x_{n_1}, \ldots, x_{n_m})$, where $n_1 < \ldots < n_m \in \mathbb{N}$.
- Likelihood function $l(\mathbf{P}|\mathbf{x}_o) = \prod_{i=1}^{m} p_{n_{i-1}n_i}^{(t_i - t_{i-1})}$

  where $p_{ij}^{(t)}$ represents the (i, j)-th element of the t step transition matrix,
- Computation of this likelihood often complex $\Rightarrow$ preferable to consider inference based on the reconstruction of missing observations
- Let $\mathbf{x}_m$ represent the unobserved states at times $1, \ldots, t_1 - 1, t_1 + 1, \ldots, t_{n-1} - 1, t_{n-1} + 1, \ldots, t_n$ and let $\mathbf{x}$ represent the full data sequence.
- Then, given a matrix beta prior, we have that $\mathbf{P}|\mathbf{x}$ is also matrix beta.

PARTIALLY OBSERVED DATA

- Furthermore, it is immediate that

$$P(\mathbf{x}_m|\mathbf{x}_o,\mathbf{P}) = \frac{P(\mathbf{x}|\mathbf{P})}{P(\mathbf{x}_o|\mathbf{P})} \propto P(\mathbf{x}|\mathbf{P}) \quad (3)$$

  which is easy to compute for given $\mathbf{P}, \mathbf{x}_m$.

- One possibility would be to set up a Metropolis within Gibbs sampling algorithm to sample from the posterior distribution of $\mathbf{P}$.

- Such an approach is reasonable if the amount of missing data is relatively small. However, if there is much missing data, it will be very difficult to define an appropriate algorithm to generate data from $P(\mathbf{x}_m|\mathbf{x}_o,\mathbf{P})$ in (3). In such cases, one possibility is to generate the elements of $\mathbf{x}_m$ one by one, using individual Gibbs steps. Thus, if t is a time point amongst the times associated with the missing observations, then we can generate a state x_t using

$$P(x_t|\mathbf{x}_{-t},\mathbf{P}) \propto p_{x_{t-1}x_t}p_{x_tx_{t+1}}$$

  where $\mathbf{x}_{-t}$ represents the complete sequence of states except for the state at time t.

PARTIALLY OBSERVED DATA

- One disadvantage of such approaches is that with large amounts of missing data, the Gibbs algorithms are likely to converge slowly as they will depend on the reconstruction of large quantities of latent variables.

- For the Sydney rainfall example, total rainfall was observed for the 21st and 22nd of March. From these data, it can be assumed that it rained on at least one of these two days. In this case, the likelihood function, including this data, becomes

$$l(\mathbf{P}|\mathbf{x}) = p_{11}^{25} p_{12}^{5} p_{21}^{6} p_{22}^{12} (p_{11}p_{12} + p_{12}p_{21} + p_{12}p_{22}) = p_{11}^{25} p_{12}^{6} p_{21}^{6} p_{22}^{12} (p_{11} + 1)$$

so that $p_{22}|\mathbf{x} \sim \text{Be}(12.5, 6.5)$ as earlier and p_{11} has a mixture posterior distribution

$$p_{11}|\mathbf{x} \sim 0.44\ \text{Be}(26.5, 6.5) + 0.56\ \text{Be}(25.5, 6.5).$$

The posterior mean is

$$E[\mathbf{P}|\mathbf{x}] = \begin{pmatrix} 0.800 & 0.200 \\ 0.342 & 0.658 \end{pmatrix}.$$

The predictive equilibrium probability is $E[\pi_1|\mathbf{x}] = 0.627$.

HIGHER ORDER CHAINS AND MIXTURES

- $\{X_n\}$ Markov chain of order r if $P(X_n = x_n | X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-r} = x_{n-r}, \ldots, X_{n-1} = x_{n-1})$

- Chain of order r represented as first order one combining states
  - Second order, homogeneous Markov chain $\{X_n\}$ with 2 possible states (1, 2)
  - $p_{ijl} = P(X_n = l | X_{n-1} = j, X_{n-2} = i)$ for $i, j, l = 1, 2$
  - first order transition matrix is

$$
\begin{array}{c} \\ (1,1) \\ (1,2) \\ (2,1) \\ (2,2) \end{array}
\begin{array}{c}
\begin{array}{cccc} (1,1) & (1,2) & (2,1) & (2,2) \end{array} \\
\begin{pmatrix}
p_{111} & p_{112} & 0 & 0 \\
0 & 0 & p_{121} & p_{122} \\
p_{211} & p_{212} & 0 & 0 \\
0 & 0 & p_{221} & p_{222}
\end{pmatrix}
\end{array}
$$

HIGHER ORDER CHAINS AND MIXTURES

- Reduction of higher order Markov chains to first order ones $\Rightarrow$ large number of states
- X_n taking values in $\{1, \ldots, K\}$ $\Rightarrow$ K^r states needed to define an r-th order chain
- Mixture transition distribution (MTD) model of Raftery (1985) one of the most popular approaches to modeling r-th order dependence
  - $P(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_{n-r} = x_{n-r}) = \sum_{i=1}^{r} w_i p_{x_{n-i} x_n}$, with $\sum_{i=1}^{r} w_i = 1$ and $\mathbf{P} = (p_{ij})$ transition matrix
  - $\Rightarrow$ more parsimonious modeling: in previous example 3 parameters (w_1, p_{11}, p_{22}) instead of 4 ($p_{111}, p_{121}, p_{211}, p_{221}$)

HIGHER ORDER CHAINS AND MIXTURES

Bayesian inference for full r-th order Markov chain model carried out, in principle, as for first order model, expanding the number of states appropriately

- Markov chains of orders $r = 2$ and 3 considered for Australian rainfall example
- Be$(1/2, 1/2)$ priors used for first non zero element of each row of the transition matrix
- Initial r states known, supposedly generated from equilibrium distribution
- Predictive equilibrium probabilities of different states under each model:

		States							
r	2	$(1,1)$	$(1,2)$	$(2,1)$	$(2,2)$				
	π	0.5521	0.1198	0.1198	0.2084				
r	3	(111)	(112)	(121)	(122)	(211)	(212)	(221)	(222)
	π	0.4567	0.0964	0.0731	0.0550	0.0964	0.0317	0.0550	0.1357

- Log marginal likelihoods: -30.7876 (2nd order) and -32.1915 (3rd order) $\Rightarrow$ first order preferred (-28.60)

HIGHER ORDER CHAINS AND MIXTURES

Straightforward Bayesian inference for the mixture transition distribution (MTD) model

- Assume known order r of the Markov chain mixture
- Define indicator variable Z_n s.t. $P(Z_n = z|\mathbf{w}) = w_z$, $z = 1, r$
- Consider $P(X_n = x_n|X_{n-1} = x_{n-1}, \ldots, X_{n-r} = x_{n-r}, Z_n = z, \mathbf{P}) = p_{x_{n-z}x_n}$
- $\Rightarrow P(X_n = x_n|X_{n-1} = x_{n-1}, \ldots, X_{n-r} = x_{n-r}) = \sum_{i=1}^{r} P(X_n = x_n|X_{n-1} = x_{n-1}, \ldots, X_{n-r} = x_{n-r}, Z_n = z, \mathbf{P})P(Z_n = z|\mathbf{w}) = \sum_{i=1}^{r} w_i p_{x_{n-i}x_n}$
- Posterior: $P(Z_n = z|X_n = x_n, \ldots, X_{n-r} = x_{n-r}, \mathbf{P}) = \frac{w_z p_{x_{n-z}x_n}}{\sum_{j=1}^{r} w_j p_{x_{n-j}x_n}}$
- ”Usual” matrix beta prior for $\mathbf{P}$
- Dirichlet prior for $\mathbf{w}$: $\mathbf{w} \sim$ Dir$(\beta_1, \ldots, \beta_r)$
- Known initial states of the chain (could take a probability model $P(x_0, \ldots, x_{r-1})$)

HIGHER ORDER CHAINS AND MIXTURES

- Given sequence of data $\mathbf{x} = (x_0, \ldots, x_n)$ and independent indicator variables $\mathbf{z} = (z_r, \ldots, z_n) \Rightarrow$

$$
\begin{aligned}
f(\mathbf{P}|\mathbf{x}, \mathbf{z}, \mathbf{w}) &\propto \prod_{t=r}^{n} p_{x_{t-z_t} x_t} f(\mathbf{P}) \\
f(\mathbf{w}|\mathbf{z}) &\propto \prod_{t=r}^{n} w_{z_t} f(\mathbf{w})
\end{aligned}
$$

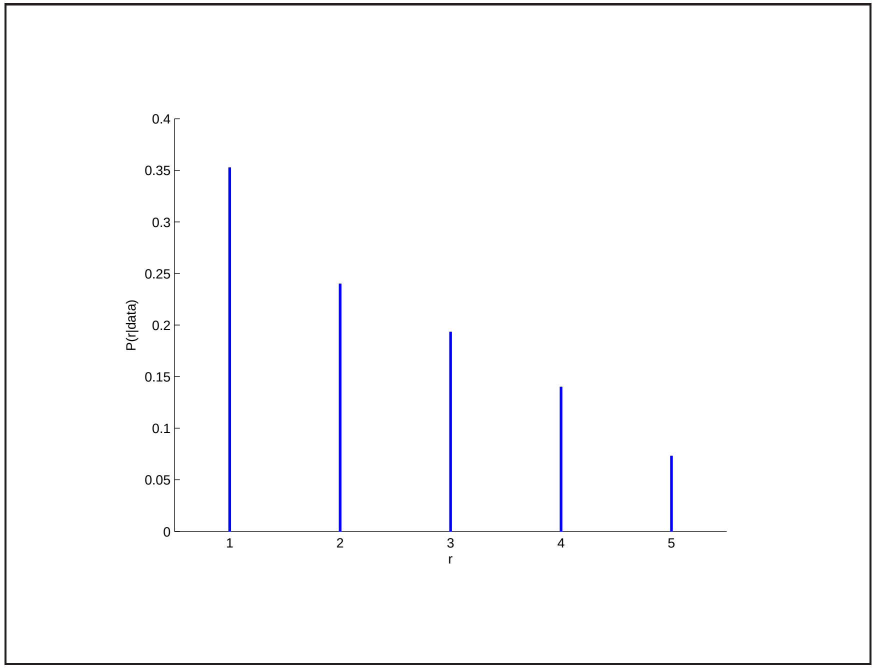
- Available full conditionals $\Rightarrow$ Gibbs sampling to get posterior distribution of $\mathbf{w}, \mathbf{P}$
  - $\mathbf{z}|\texttt{data}, \mathbf{w}, \mathbf{P}$ in closed form
  - $\mathbf{P}|\texttt{data}, \mathbf{w}, \mathbf{z}$ matrix beta distribution
  - $\mathbf{w}|\texttt{data}, \mathbf{z}, \mathbf{P}$ Dirichlet distribution

HIGHER ORDER CHAINS AND MIXTURES

- Two possible approaches possible with unknown chain order
  - Fit models with different, fixed order and compare them via Bayes factor
  - Specify a prior distribution on the order and use Reversible Jump Markov chain Monte Carlo (RJMCMC) to estimate it
- Mixture transition models of orders up to 5 for Australian rainfall data
- Known first five data
- Discrete uniform prior $r \sim \text{DU}[1,5]$ prior distribution on the order
- Dirichlet prior distributions $\mathbf{w}|r \sim \text{Dir}\left(\underbrace{\frac{1}{2},\ldots,\frac{1}{2}}_{r}\right)$
- 200000 reversible jump MCMC iterations

HIGHER ORDER CHAINS AND MIXTURES

Estimated posterior distribution, via RJMCMC, of Markov chain order



Simple Markov chain most likely model $\Rightarrow$ previous results confirmed

DTM PROCESSES WITH CONTINUOUS STATE SPACE

- Markov processes at discrete times can be defined with both discrete and continuous state spaces

- Equilibrium distribution exists for a Markov chain with discrete state space for
  - aperiodic chain
  - all positive recurrent states

- Although the condition of positive recurrence cannot be sensibly applied to chains with continuous state space, a similar condition known as Harris recurrence applies to chains with continuous state space, which essentially means that the chain can get close to any point in the future

- Harris recurrent, aperiodic chains possess an equilibrium distribution, so that if the conditional probability distribution of the chain is $P(X_n|X_{n-1})$, then the equilibrium density π satisfies $\pi(x) = \int P(x|y)\pi(y)\, dy$

- As with Markov chains with discrete state space, a sufficient condition for a process to possess an equilibrium distribution is to be reversible

DTM PROCESSES WITH CONTINUOUS STATE SPACE

- Autoregressive (AR) models: simple examples of continuous space Markov chains

- AR(1): $X_n = \phi_0 + \phi_1 X_{n-1} + \epsilon_n$
  - ϵ_n sequence of i.i.d. Gaussian r.v.'s with zero mean and variance σ^2
  - Weakly, but not strictly, stationary if $|\phi_1| < 1$
  - Non stationary process if $|\phi_1| \geq 1$
  - $\mu_X = \phi_0 + \phi_1 \mu_X \Rightarrow \mu_X = \frac{\phi_0}{1-\phi_1}$

- AR(k): $X_n = \phi_0 + \sum_{i=1}^{k} \phi_i X_{n-i} + \epsilon_n$

- AR(k) (weakly) stationary if all roots z_i of $\phi_0 z^k - \sum_{i=1}^{k} \phi_i z^{k-i}$ satisfy $|z_i| < 1$

INFERENCE FOR AR(k) MODELS

- $X_1 = x_1, \ldots, X_k = x_k$ known initial values
- Sample of n data: $X_{k+1} = x_{k+1}, \ldots, X_{k+n} = x_{k+n}$
- Prior distributions
  - $\frac{1}{\sigma^2} \sim \text{Ga}\left(\frac{a}{2}, \frac{b}{2}\right)$
  - $\boldsymbol{\beta} \sim \text{N}(\mathbf{m}, \mathbf{V})$
- Full conditional posterior distributions
  - $\boldsymbol{\phi}|\sigma^2, \mathbf{x} \sim \text{N}\left(\left(\mathbf{V}^{-1} + \frac{1}{\sigma^2}\mathbf{Z}\mathbf{Z}^T\right)^{-1}\left(\mathbf{V}^{-1}\mathbf{m} + \frac{1}{\sigma^2}\mathbf{Z}^T\mathbf{x}\right), \left(\mathbf{V}^{-1} + \frac{1}{\sigma^2}\mathbf{Z}\mathbf{Z}^T\right)^{-1}\right)$
  - $\frac{1}{\sigma^2}|\mathbf{x} \sim \text{Ga}\left(\frac{a+n}{2}, \frac{b+(\mathbf{x}-\mathbf{Z}\boldsymbol{\phi})^T(\mathbf{x}-\mathbf{Z}\boldsymbol{\phi})}{2}\right)$
  - where $\mathbf{x} = (x_{k+1}, \ldots, x_{k+n})^T$, $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^T$ and $\mathbf{z}_t = (1, x_{t+k-1}, \ldots, x_t)^T$
- Gibbs sampling to get a sample from posterior distribution

INFERENCE FOR AR(k): STATIONARITY

Straightforward extension this to incorporate stationarity

- MCMC sample generated from posterior distribution of $\boldsymbol{\beta}, \sigma^2$
- Reject sampled values with roots ≥ 1
- $\Rightarrow$ Sample reduced to a sample from the posterior distribution based on a normal gamma prior distribution truncated onto the region where the parameters satisfy the stationarity condition

INFERENCE FOR AR(k): MODEL SELECTION

Selection of appropriate value of k through

- Prior on k and RJMCMC
- Bayes factor
- DIC (Deviance Information Criterion)
  - Model $\mathcal{M}$ parameterized by $\boldsymbol{\theta}$
  - Sample $\mathbf{x}$
  - $DIC_{\mathcal{M}} = -4E\left[\log f(\mathbf{x}|\boldsymbol{\theta})|\mathbf{x}\right] + 2\log f(\mathbf{x}|E[\boldsymbol{\theta}|\mathbf{x}])$
  - Lower values of the DIC $\Rightarrow$ more plausible models

INFERENCE FOR AR(k)

- Quarterly data on seasonally adjusted gross national product of the USA between 1947 and 1991

- AR(3) model chosen by MLE (Tsay, 2005)

- Consider AR models with 0 up to 4 lags and use DIC to choose the model

- First 4 data known

- Independent prior distributions $\frac{1}{\sigma^2} \sim \text{Ga}(0.001, 0.001)$ and $\beta_i \sim \text{N}(0, 0.0001)$ for $i = 0, \ldots, k$

- WinBUGS used to run Gibbs sampler with 100000 iterations to burn in and 100000 iterations in equilibrium in each case

INFERENCE FOR AR(k)

- Values of the DIC for each model

Lags	DIC
0	-1065.2
1	-1090.3
2	-1099.1
3	**-1102.7**
4	-1092.3

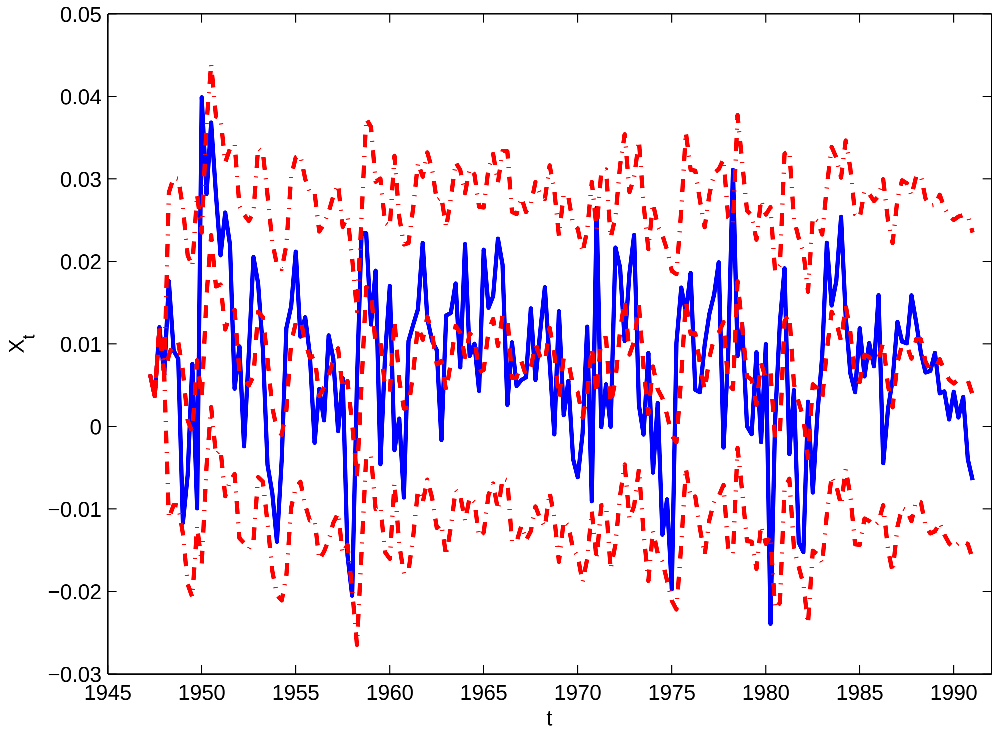
- DIC $\Rightarrow$ AR(3) model

- Tsay: $X_n = 0.0047 + 0.35X_{n-1} + 0.18X_{n-2} - 0.14X_{n-3} + \epsilon_n$, with estimated standard deviation of error term $\widehat{\sigma} = 0.0098$

- Ours: posterior mean predictor $0.0047+0.3516X_{n-1}+0.1798X_{n-2}-0.1445X_{n-3}$ and 0.0100 as posterior mean of σ

INFERENCE FOR AR(k)

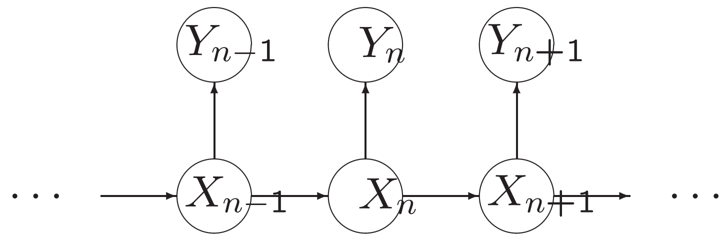
Actual data and (Bayesian) fitted AR(3) model with 95% predictive intervals



HIDDEN MARKOV MODELS

Hidden Markov models (HMMs) widely applied to analysis of weakly dependent data in diverse areas such as econometrics, ecology and signal processing

- Observations Y_n for $n = 0, 1, 2, \ldots$ generated from a conditional distribution $f(y_n|X_n)$ with parameters dependent on unobserved or hidden state $X_n \in \{1, 2, \ldots K\}$
- Hidden states follow a Markov chain with transition matrix $\mathbf{P}$ and an initial distribution, usually assumed to be the equilibrium distribution, $\pi(\cdot|\mathbf{P})$, of the underlying Markov chain
- Influence diagram representing the dependence structure of a HMM



HIDDEN MARKOV MODELS

Extension to HMM with continuous state space: simple example $\Rightarrow$ Dynamic Linear Model (DLM)

- DLM with univariate observations X_n characterized by $\{F_n, G_n, V_n, W_n\}$
  - F_n known vector of dimension $m \times 1$, for each n
  - G_n is known $m \times m$ matrix, for each n
  - V_n known variance, for each n
  - W_n known $m \times m$ variance matrix, for each n
- $X_n = F_n\theta_n + v_n, v_n \sim \mathrm{N}(0, V_n)$
- $\theta_n = G_n\theta_{n-1} + w_n, w_n \sim \mathrm{N}(0, W_n)$
- Information D_n defined recursively as $D_n = D_{n-1} \cup \{x_n\}$

HIDDEN MARKOV MODELS

- Sample data $\mathbf{y} = (y_0, \ldots, y_n)$
- Likelihood $l(\boldsymbol{\theta}, \mathbf{P}|\mathbf{y}) = \sum_{x_0,\ldots,x_n} \pi(x_0) f(y_0|\boldsymbol{\theta}_{x_0}) \prod_{j=1}^{n} p_{x_{j-1}x_j} f(y_j|\boldsymbol{\theta}_{x_j})$
- Likelihood contains K^{n+1} terms $\Rightarrow$ in practice usually impossible to compute directly
- Other approaches taken

HIDDEN MARKOV MODELS

- Suppose states $\mathbf{x}$ of hidden Markov chain known
- $\Rightarrow$ likelihood simplifies to
$l(\boldsymbol{\theta}, \mathbf{P}|\mathbf{x}, \mathbf{y}) = \pi(x_0) \prod_{j=1}^{n} p_{x_{j-1}x_j} \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta}_{x_i}) = l_1(\mathbf{P}|\mathbf{x}) l_2(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$, with
$l_1(\mathbf{P}|\mathbf{x}) = \pi(x_0|\mathbf{P}) \prod_{j=1}^{n} p_{x_{j-1}x_j}$ and $l_2(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y}) = \prod_{i=0}^{n} f(y_i|\boldsymbol{\theta}_{x_i})$
- ”Usual” matrix beta prior distribution for $\mathbf{P}$ $\Rightarrow$ simple rejection algorithm used to sample from $f(\mathbf{P}|\mathbf{x})$
- Sampling from each $\boldsymbol{\theta}_i|\mathbf{x}, \mathbf{y}$ (straightforward if $Y|\boldsymbol{\theta}$ standard exponential family distribution and conjugate prior for $\boldsymbol{\theta}$)
- Let $\mathbf{x}_{-t} = (x_0, x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n) \Rightarrow$, for $i = 1, \ldots, K$,

$$\begin{aligned} P(x_0 = i|\mathbf{x}_{-0}, \mathbf{y}) &\propto \pi(i) p_{ix_1} f(y_1|\boldsymbol{\theta}_i) \\ P(x_t = i|\mathbf{x}_{-t}, \mathbf{y}) &\propto p_{x_{t-1}i} p_{ix_{t+1}} f(y_t|\theta_i) \quad \text{for } 1 < t < n \\ P(x_n = i|\mathbf{x}_{-n}, \mathbf{y}) &\propto p_{x_{n-1}i} f(y_n|\boldsymbol{\theta}_i) \end{aligned}$$

- Gibbs sampling

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it   *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/~fabrizio/

CONTINUOUS TIME MARKOV CHAINS

- CTMCs continuous time stochastic processes with discrete state space
- Here homogeneous CTMCs with finite state space
- System remains an exponential time at each state and, when leaving such state, it evolves according to probabilities which depend only on the leaving state
- Simple proof of *Markov property* and *Homogeneity* $\Rightarrow$ *Exponential distribution*
  - Initial value $X_0 = i$
  - $\overline{F}(t) = P(\tau > t | X_0 = i)$, $t > 0$
  - τ waiting time for a change of state from state i

$$\begin{aligned} P(\tau > s + t | X_0 = i) &= P(\tau > s + t | X_0 = i, \tau > s) P(\tau > s | X_0 = i) \\ &= P(\tau > s + t | X_s = i) P(\tau > s | X_0 = i) \end{aligned}$$

  - $\Rightarrow \overline{F}(s + t) = \overline{F}(s)\overline{F}(t)$
  - $\Leftrightarrow \overline{F}(t) = e^{-\lambda t}, t > 0, \lambda > 0$

CONTINUOUS TIME MARKOV CHAINS

- $\{X_t\}_{t \in T}$ continuous time stochastic process with discrete state space $E = \{1, 2, \ldots, K\}$
- When the process enters into state i, it remains there for an exponentially distributed time period with mean $1/\nu_i$
- At the end of this time period, the process will move to a different state $j \neq i$ with probability p_{ij}, s.t. $\sum_{j=1}^{K} p_{ij} = 1, \ \forall i$, and $p_{ii} = 0$
- For physical or logical reasons, some additional p_{ij}'s could also be zero
- $\mathbf{P} = (p_{ij})$ transition probability matrix
- Transition points $\{X_n\}$ and transition matrix define an embedded (discrete time) Markov chain
- $\{X_t\}$ will be designated a CTMC with parameters $\mathbf{P}$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_K)^T$

BIRTH-DEATH PROCESS

- Birth-death process particular example of a CTMC with state space $\{0, 1, 2, \ldots, K\}$ denoting population size
- Transitions occur either as single births, with rate λ_i or single deaths, with rate μ_i, for $i = 0, \ldots, K$, where $\mu_0 = \lambda_K = 0$
- Transition probabilities: $p_{i,i+1} = \lambda_i/(\lambda_i + \mu_i)$, $p_{i,i-1} = \mu_i/(\lambda_i + \mu_i)$ and $p_{ij} = 0$ for $i = 0, \ldots, K$ and $j \notin \{i-1, i+1\}$ (properties of exponentials)
- Times between transitions Ex(ν_i), $\nu_i = \lambda_i + \mu_i$ (minimum of two exponentials)
- Markovian queueing system: birth-death process with
  - i people in the system (*alive*)
  - arrivals (*births*) with rate λ_i
  - service completed (*deaths*) with rate μ_i
- Poisson process: pure birth process with infinite state space $\{0, 1, 2, \ldots\}$, $\mu_i = 0$ and $\lambda_i = \lambda$ for all i

CONTINUOUS TIME MARKOV CHAINS

- $r_{ij} = \nu_i p_{ij}$: jumping intensities (from state i into state j)
- Set $r_{ii} = -\sum_{j \neq i} r_{ij} = -\nu_i,\ i \in \{1, \dots, K\}$ ($\Rightarrow \sum_j r_{ij} = 0$)
- r_{ij} interpreted as flow rates (see Resnick p. 385-6):
  - $\nu_i t \approx P$(system leaves i before t)
  - $\nu_i p_{ij} t \approx P$(system leaves i before t and goes to j)
- $\mathbf{\Lambda} = (r_{ij})$: intensity matrix (or infinitesimal generator of the process)

CONTINUOUS TIME MARKOV CHAINS

- Short term behavior of CTMC described through forward Kolmogorov system of differential equations

- $P_{ij}(t) = P(X_{t+s} = j|X_s = i) = P(X_t = j|X_0 = i)$ transition probability functions

- Suitable regularity conditions (see e.g. Ross, 2009)
$\Rightarrow P'_{ij}(t) = \sum_{k \neq j} r_{kj} P_{ik}(t) - \nu_j P_{ij}(t) = \sum_k r_{kj} P_{ik}(t)$

- $\Rightarrow \mathbf{P}'(t) = \mathbf{\Lambda P}(t)$ and $\mathbf{P}(0) = \mathbf{I}$,
with $\mathbf{P}(t) = (P_{ij}(t))$ matrix of transition probability functions and $\mathbf{I}$ identity matrix

- $\mathbf{P}(t) = \exp(\mathbf{\Lambda} t)$ analytic solution of the system to be solved, for given t, using matrix exponentiation

CONTINUOUS TIME MARKOV CHAINS

- Simplest case: $\mathbf{\Lambda}$ diagonalizable (i.e. $\mathbf{D} = \mathbf{S}^{-1}\mathbf{\Lambda}\mathbf{S}$) with different eigenvalues
  - $\mathbf{D}$ diagonal matrix with distinct eigenvalues $\lambda_1, \ldots, \lambda_K$ of $\mathbf{\Lambda}$ as entries
  - $\mathbf{S}$ invertible matrix made of eigenvectors corresponding to eigenvalues in $\mathbf{\Lambda}$
  - Eigenvalues and eigenvectors from $\mathbf{\Lambda}\mathbf{S} = \mathbf{S}\mathbf{D}$
- Decompose $\mathbf{\Lambda} = \mathbf{S}\mathbf{D}\mathbf{S}^{-1}$

$$\begin{aligned}
\exp(\mathbf{\Lambda}t) &= \sum_{i=0}^{\infty} \frac{(\mathbf{\Lambda}t)^i}{i!} = \sum_{i=0}^{\infty} \frac{(\mathbf{S}\mathbf{D}\mathbf{S}^{-1})^i t^i}{i!} = \mathbf{S}\left[\sum_{i=0}^{\infty} \frac{(\mathbf{D}t)^i}{i!}\right]\mathbf{S}^{-1} \\
&= \mathbf{S}\begin{pmatrix} \exp(\lambda_1 t) & 0 & \ldots & 0 \\ 0 & \exp(\lambda_2 t) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \exp(\lambda_K t) \end{pmatrix}\mathbf{S}^{-1}
\end{aligned}$$

CONTINUOUS TIME MARKOV CHAINS

- As with the discrete time case, forecasting the long term behavior of a CTMC means that we need to consider the equilibrium distribution

- Under suitable conditions (see e.g. Ross, 2009), for given $\mathbf{P}$ and $\boldsymbol{\nu}$, the equilibrium distribution $\{\pi_j\}_{j=1}^{K}$, if it exists, is obtained through the solution of the system

$$\nu_j \pi_j = \sum_{i \neq j} r_{ij} \pi_i, \qquad \forall j \in \{1, \ldots, K\},$$

$$\sum_j \pi_j = 1; \qquad \pi_j \geq 0$$

INFERENCE FOR CTMC

- $\mathbf{P}$ and transition rates $\boldsymbol{\nu}$ unknown and unrelated (i.e. elements of $\mathbf{P}$ not known functions of $\boldsymbol{\nu}$)

- Observe initial state x_0, times, t_i, and states, x_i, $i = 1, \ldots, n$, of the first n transitions of the chain

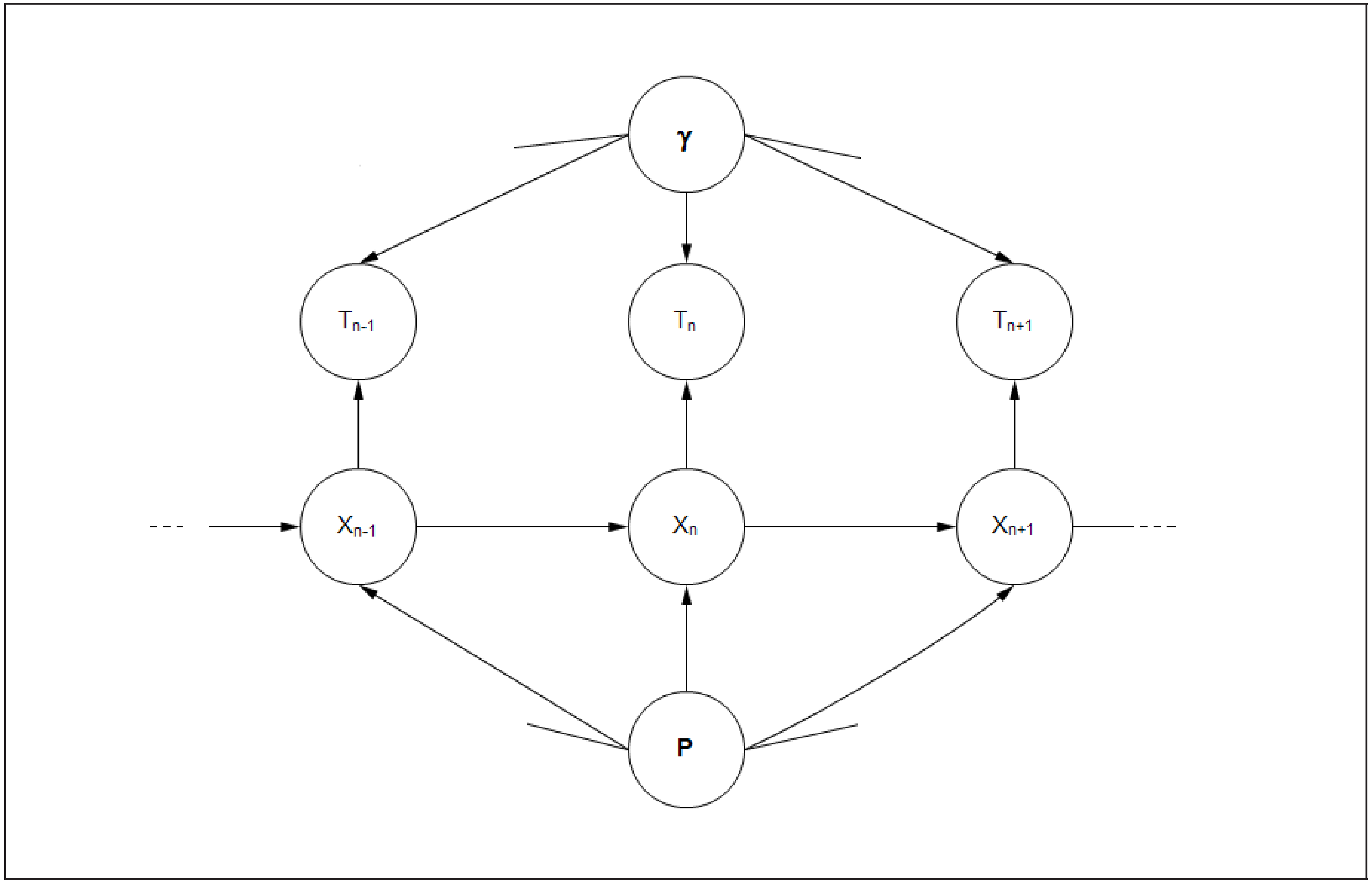
- $\Rightarrow$ Likelihood

  $l(\mathbf{P}, \boldsymbol{\nu}|\text{data}) = \prod_{i=1}^{n} \nu_{x_{i-1}} \exp(-\nu_{x_{i-1}}(t_i - t_{i-1}))p_{x_{i-1}x_i} \propto \prod_{i=1}^{K} \nu_i^{n_i} \exp(-\nu_i T_i) \prod_{j=1}^{K} p_{ij}^{n_{ij}},$

  with n_{ij} number of observed transitions from i to j, T_i total time spent in state i and $n_i = \sum_{j=1}^{K} n_{ij}$ total number of transitions out of state i, for $i, j \in \{1, \ldots, K\}$

- Note that many alternative experiments have likelihood functions of the same form (e.g. the same T_i can be obtained with different individual stays)

INFLUENCE DIAGRAM FOR A CTMC



INFERENCE FOR CTMC

- Likelihood: $l(\mathbf{P}, \boldsymbol{\nu}|\text{data}) = l_1(\mathbf{P}|\text{data}) l_2(\boldsymbol{\nu}|\text{data})$,

  with $l_1(\mathbf{P}|\text{data}) = \prod_{i=1}^K \prod_{j=1}^K p_{ij}^{n_{ij}}$ and $l_2(\boldsymbol{\nu}|\text{data}) = \prod_{i=1}^K \nu_i^{n_i} \exp(-\nu_i T_i)$

- Independent priors for $\mathbf{P}$ and $\boldsymbol{\nu}$ $\Rightarrow$ independent posteriors $\Rightarrow$ separate inference for $\mathbf{P}$ and $\boldsymbol{\nu}$

- Known initial state (o.w. as before)

- Conjugate matrix beta prior for $\mathbf{P}$

- Priors $\nu_i \sim \text{Ga}(a_i, b_i) \Rightarrow$ posteriors $\nu_i|\text{data} \sim \text{Ga}(a_i + n_i, b_i + T_i)$, for $i = 1, \ldots, K$

INFERENCE FOR CTMC

Estimation of intensity matrix $\mathbf{\Lambda} = (r_{ij})$

- Posterior distributions of ν_i and p_{ij}, $\forall i, j$, sufficiently concentrated $\Rightarrow$ use posterior modes $\widehat{\nu}_i$ and $\widehat{p}_{ij}$ to estimate $\widehat{r}_{ij}$, $\forall i, j$
  - $\widehat{r}_{ij} = \widehat{\nu}_i \widehat{p}_{ij}, \ i \neq j$
  - $\widehat{r}_{ii} = -\widehat{\nu}_i, \ i = 1, \ldots, m$
  - For the Dirichlet-multinomial model (on P and X's) we get

$$\widehat{\nu}_i = \frac{\alpha_i + n_i - 1}{\beta_i + \sum_{j=1}^{n_i} t_{ij}}; \ \widehat{p}_{ij} = \frac{\delta_{ij} + n_{ij} - 1}{\sum_{l \neq i}(n_{il} + \delta_{il}) - k + 1}; \ \widehat{r}_{ij} = \widehat{\nu}_i \widehat{p}_{ij}, \widehat{r}_{ii} = -\widehat{\nu}_i$$

- Posterior samples $\{\nu^\eta\}_{\eta=1}^N$ and $\{\mathbf{P}^\eta\}_{\eta=1}^N \Rightarrow$ samples from posterior of $\{r_{ij}^\eta\}$, $\forall i, j$
  - $\{r_{ij}^\eta = \nu_i^\eta p_{ij}^\eta\}_{\eta=1}^N, \ i \neq j$
  - $\{r_{ii}^\eta = -\nu_i^\eta\}_{\eta=1}^N, \ i = 1, \ldots, k$
  - Summarize all samples through, e.g., sample means $\frac{1}{N} \sum_{\eta=1}^N r_{ij}^\eta, \forall i, j$

FORECASTING SHORT TERM BEHAVIOR

- Forecast can be based on the solution of the system of differential equations characterizing short term behavior,

  $\mathbf{P}'(t) = \mathbf{\Lambda}\mathbf{P}(t)$ and $\mathbf{P}(0) = \mathbf{I}$,

  when parameters $\mathbf{P}$ and $\boldsymbol{\nu}$ are fixed

- Need to take into account the uncertainty about parameters to estimate the predictive matrix of transition probabilities $\mathbf{P}(t)|$data

- Various options can be considered

FORECASTING SHORT TERM BEHAVIOR

- Posteriors of $\mathbf{P}$ and $\boldsymbol{\nu}$ sufficiently concentrated $\Rightarrow$ summarize them through posterior modes $\widehat{\boldsymbol{\nu}}$ and $\widehat{\mathbf{P}}$

- Assume $\boldsymbol{\Lambda}(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}})$ diagonalizable with K different eigenvalues

- $\Rightarrow$ estimate $\mathbf{P}(t)|$data through

$$\mathbf{S}(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}}) \begin{pmatrix} \exp(\lambda_1(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}})t) & 0 & ... & 0 \\ 0 & \exp(\lambda_2(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}})t) & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \exp(\lambda_k(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}})t) \end{pmatrix} \mathbf{S}(\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}})^{-1}.$$

FORECASTING SHORT TERM BEHAVIOR

- Obtain Monte Carlo samples, $\boldsymbol{\nu}^{(s)}, \mathbf{P}^{(s)}$, for $s = 1, \ldots, S$
- $\Rightarrow$ solve the corresponding decomposition for each s
- $\Rightarrow$ sample $\mathbf{P}(t)^{(s)}$, for $s = 1, \ldots, S$
- Summarize through sample mean $\frac{1}{S} \sum_s \mathbf{P}(t)^{(s)}$
- Procedure easily implemented for K relatively small
- For large K matrix exponentiation might be too computationally intensive to be used within a Monte Carlo type scenario
- Use, for example, Reduced Order Model (ROM)

REDUCED ORDER MODEL (GRIGORIU, 2009)

- Method possibly useful for predictive computations with stochastic process models
- Posterior predictive computation of an event A dependent on $\boldsymbol{\theta}$: $P(A|\mathbf{x}) = \int P(A|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$
- Monte Carlo approximation based on large sample $\{\theta_i\}_{i=1}^N$ from the posterior

$$P_{MC}(A|\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} P(A|\boldsymbol{\theta}_i)$$

- Approach infeasible when finding each $P(A|\boldsymbol{\theta}_i)$ computationally demanding $\Rightarrow$ unable to use large enough samples to apply standard MC approximations
- Assume posterior $f(\boldsymbol{\theta}|\mathbf{x})$ can be approximated by simple probability distribution with m support points $\{\boldsymbol{\theta}_i\}_{i=1}^m$, each with probability p_i, $i = 1, ..., m$, with m small enough so that m predictive computations $P(A|\boldsymbol{\theta}_i)$ are actually amenable
- Aim at approximating quantity of interest through $\tilde{P}(A) = \sum_{i=1}^m P(A|\boldsymbol{\theta}_i) p_i$, satisfactorily under appropriate conditions

REDUCED ORDER MODEL (GRIGORIU, 2009)

- Determine the order $m \geq 1$ of ROM, based only on maximum number m of *acceptable* $P(A|\boldsymbol{\theta})$ computations

- Determine the range $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$ of $\tilde{\boldsymbol{\Theta}}$ for the selected m

- Compute probabilities $\{p_1, \ldots, p_m\}$ of $\{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m\}$

- Use ROM approximation (details in Grigoriu, 2009)

FORECASTING SHORT TERM BEHAVIOR

1. For $s = 1, \dots, S$, sample $\boldsymbol{\nu}^{(s)}, \mathbf{P}^{(s)}$ from the relevant posteriors

2. Cluster the sampled values into m clusters and spread the centroids to obtain the ROM range $(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)})$ for $i = 1, \dots, m$

3. Compute the optimal ROM probabilities by solving $\min_{q_1,\dots,q_m} e(q_1, \dots, q_m)$ s.t. $\sum_{r=1}^{m} q_r = 1, \quad q_r \geq 0, \; r = 1, \dots, m$

4. For $i = 1$ to m

   (a) Compute $\boldsymbol{\Lambda}(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)})$

   (b) Decompose $\boldsymbol{\Lambda}(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)}) = \mathbf{S}(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)})\mathbf{D}(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)})\mathbf{S}^{-1}(\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)})$

   (c) Compute $\mathbf{P}(t)|\nu^{(i)}, \mathbf{P}^{(i)}$ through

$$\mathbf{S}\left(\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)}\right) \begin{pmatrix} \exp(\lambda_1(\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)})t) & 0 & \dots & 0 \\ 0 & \exp(\lambda_2(\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)})t) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \exp(\lambda_m(\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)})t) \end{pmatrix} \mathbf{S}\left(\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)}\right)^{-1}$$

5. Approximate $\mathbf{P}(t)|$data through $\sum_{i=1}^{m} q_i \; \mathbf{P}(t)|\boldsymbol{\nu}^{(i)}, \mathbf{P}^{(i)}$

FORECASTING LONG TERM BEHAVIOR

- Long term forecasting of CTMC behavior undertaken in different ways depending on concentration of posterior and available computational budget

- Sufficiently concentrated posteriors $\Rightarrow$ replace parameters with posterior modes
  - Solve system: $\widehat{\nu}_j \pi_j = \sum_{i \neq j} \widehat{r}_{ij} \pi_i, \forall j \in \{1, \ldots, K\}$, $\quad \sum_j \pi_j = 1; \pi_j \geq 0$
  - $\Rightarrow$ Obtain approximate point summary of predictive equilibrium distribution $\{\widehat{\pi}_i\}_{i=1}^m$
  - Drawback: approach does not give a measure of uncertainty

- Obtain samples from the posteriors, $\boldsymbol{\nu}^{(s)}, \mathbf{P}^{(s)}$, for $s = 1, \ldots, S$
  - obtain sampled probabilities, $\pi_i^{(s)}$, for $s = 1, \ldots, S$, from the predictive equilibrium distribution through repeated solution of previous system
  - summarize probabilities through means $\widehat{\pi}_i = \frac{1}{S} \sum_{s=1}^{S} \pi_i^{(s)}$ for $i = 1, \ldots, K$

- For large K, use ROM for computationally expensive solution of system of equations

PREDICTING TIMES BETWEEN TRANSITIONS

- Given current state i, straightforward prediction of the time to the next transition
- Exponential model Ex(ν_i) for T, time to the next transition
- Gamma posterior $\nu_i|$data $\sim$ Ga$(a_i + n_i, b_i + T_i)$
- $\Rightarrow P(T \leq t|\text{data}) = \left(\frac{b_i + T_i}{b_i + T_i + t}\right)^{a_i+n_i}$
- Predictions of times up to more than one transition using Monte Carlo approaches as earlier

HARDWARE AVAILABILITY THROUGH CTMC

- Recent increased interest in reliability, availability and maintainability (RAM) analyses of hardware (HW) systems and, in particular, safety critical systems

- Concerned with availability of hardware systems, modeled through CTMCs evolving through a discrete set of states, some of which correspond to ON configurations and the rest to OFF configurations

- Transition from an ON to an OFF state entails a system failure, whereas a transition from an OFF to an ON system implies a repair

- Here emphasize availability, key performance parameter for Information Technology systems

- There are many hardware configurations aimed at attaining very high system availability, e.g. 99.999% of time, through transfer of workload when one, or more, system components fail, or intermediate failure states with automated recovery

- CTMC with states $\{1, 2, \ldots, l\}$ corresponding to operational (ON) configurations, whereas states $\{l+1, \ldots, K\}$ correspond to OFF configurations

HARDWARE AVAILABILITY THROUGH CTMC

- A classical approach to availability estimation of CTMC HW systems would calculate MLE for CTMC parameters, compute equilibrium distribution given these and estimate the long term fraction of time that the system remains in ON configurations

- A shortfall of this approach is that it does not account for parameter uncertainty, whereas the fully Bayesian framework automatically incorporates this uncertainty

- We provide short term and long term forecasting

HARDWARE AVAILABILITY THROUGH CTMC

- Consider steady state prediction of the system
- $\Rightarrow$ (unconditional) availability as sum of equilibrium probabilities for ON states
  - Conditional (upon $\boldsymbol{\nu}$ and $\mathbf{P}$) availability: $A|\nu, P = \sum_{i=1}^{l} \pi_i | \boldsymbol{\nu}, \mathbf{P}$
  - Unconditional availability given by
    1. $\widehat{A}|data \simeq \sum_{i=1}^{l} \widehat{\pi}_i$, using approximate equilibrium distribution if posteriors sufficiently concentrated
    2. $\frac{1}{S}\sum_{s=1}^{S} A^{(s)}$, using sample $\{A^{(s)} = \sum_{i=1}^{l} \pi_i^{(s)}\}_{s=1}^{S}$
    3. $\sum_{i=1}^{m} q_i \left(\sum_{j=1}^{l} \pi_j^{(i)} | \nu^{(i)}, \mathbf{P}^{(i)}\right)$, based on ROM equilibrium distribution if computational budget only allows for m equilibrium distribution computations

HARDWARE AVAILABILITY THROUGH CTMC

- Interest in (short term) interval availability

- Define random variables Y_t and A_t

$$Y_t|\boldsymbol{\nu},\mathbf{P} = \begin{cases} 1, & \text{if } X_t|\boldsymbol{\nu},\mathbf{P} \in \{1,2,\ldots,l\}, \\ 0, & \text{otherwise} \end{cases}$$

$$A_t|\boldsymbol{\nu},\mathbf{P} = \frac{1}{t}\int_0^t (Y_u|\boldsymbol{\nu},\mathbf{P})\ du$$

- Interval availability

$$I_t|\boldsymbol{\nu},\mathbf{P} = E[A_t|\boldsymbol{\nu},\mathbf{P}] = \frac{1}{t}\sum_{j=1}^{l}\int_0^t \pi_j(u|\boldsymbol{\nu},\mathbf{P})\,du,$$

  where $\pi_j(u|\boldsymbol{\nu},\mathbf{P}) = P(X_u = j|\boldsymbol{\nu},\mathbf{P})$

- Interval availability approximated with a one dimensional integration method, like Simpson's rule

HARDWARE AVAILABILITY THROUGH CTMC

- Key computation $\pi_j(t|\boldsymbol{\nu},\mathbf{P})$ terms, $j = 1,...,K$ through solution of Chapman-Kolmogorov system of differential equations (see e.g. Ross, 2009)

$$\pi'(t|\nu,P) = (\Lambda|\boldsymbol{\nu},\mathbf{P})\cdot\pi(t|\boldsymbol{\nu},\mathbf{P}); \quad t\in[0,T)$$

$$\pi(0|\boldsymbol{\nu},\mathbf{P}) = \pi^{(0)}, where$$

  - $\pi(t|\boldsymbol{\nu},\mathbf{P}) = (\pi_1(t|\boldsymbol{\nu},\mathbf{P}),\ldots,\pi_K(t|\boldsymbol{\nu},\mathbf{P}))$
  - $\boldsymbol{\pi}^{(0)} = (\pi_1^{(0)},\pi_2^{(0)},\ldots,\pi_K^{(0)})$ initial state probability vector
  - $\boldsymbol{\Lambda}|\boldsymbol{\nu},\mathbf{P}$ intensity matrix, conditional on $\boldsymbol{\nu},\mathbf{P}$

- $\Rightarrow$ Analytic solution, with matrix exponentiation as key operation

$$\pi(t|\boldsymbol{\nu},\mathbf{P}) = \pi^{(0)}\exp(\boldsymbol{\Lambda}t|\boldsymbol{\nu},\mathbf{P})$$

HARDWARE AVAILABILITY THROUGH CTMC

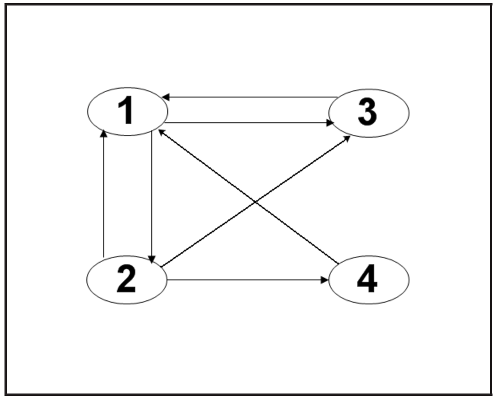
- Posterior interval availability

$$I_t|data = \int\int E[A_t|\boldsymbol{\nu}, \mathbf{P}]\pi(\boldsymbol{\nu}, \mathbf{P}|\text{data})\, d\mathbf{P}\, d\boldsymbol{\nu}$$

- As discussed earlier, at least three approaches to get predictive availability

1. $E[A_t|\widehat{\mathbf{P}}, \widehat{\boldsymbol{\nu}}]$, with $\widehat{\mathbf{P}}$ and $\widehat{\boldsymbol{\nu}}$ posterior modes for posteriors sufficiently concentrated
2. $\frac{1}{S}\sum_{s=1}^{S} E\left[A_t|\mathbf{P}^{(s)}, \boldsymbol{\nu}^{(s)}\right]$ for appropriate posterior samples $\{\mathbf{P}^{(s)},\ \boldsymbol{\nu}^{(s)}\}_{s=1}^{S}$, o.w.
3. $\sum_{i=1}^{m} q_i E\left[A_t|\mathbf{P}^{(i)}, \boldsymbol{\nu}^{(i)}\right]$, based on ROM, if computational budget allows only for m availability computations

DUAL-DUPLEX MODEL



- System described by dual-duplex model
- Dual-duplex system designed to detect a fault using a hardware comparator that switches to a hot standby redundancy
- Dual-duplex system designed in double modular redundancy to improve reliability and safety
- Dual-duplex system has high reliability, availability and safety $\Rightarrow$ applied in embedded control systems like airplanes
- two ON states $\{1, 2\}$ and two OFF states $\{3, 4\}$

DUAL-DUPLEX MODEL

- Transition probability matrix

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \left(\begin{array}{cccc} 0 & p_{12} & p_{13} & 0 \\ p_{21} & 0 & p_{23} & p_{24} \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{array} \right) \end{array},$$

- Permanence rates: ν_1, ν_2, ν_3 and ν_4

DUAL-DUPLEX MODEL

- Dir$(0, 1, 1, 0)$ and Dir$(1, 0, 1, 1)$ priors for rows 1 and 2, respectively
- Data counts $n_{12} = 10$, $n_{13} = 4$, $n_{21} = 7$, $n_{23} = 1$, $n_{24} = 2$
- $\Rightarrow$ Posterior

$$\begin{array}{rcl}(p_{11}, p_{12}, p_{13}, p_{14})|data & \sim & \text{Dir}(0, 1 + 10, 1 + 4, 0) \\ (p_{21}, p_{22}, p_{23}, p_{24})|data & \sim & \text{Dir}(1 + 7, 0, 1 + 1, 1 + 2)\end{array}$$

- Posterior means:

$$\widehat{p}_{11} = 0, \widehat{p}_{12} = 0.69, \widehat{p}_{13} = 0.31, \widehat{p}_{14} = 0$$

$$\widehat{p}_{21} = 0.62, \widehat{p}_{22} = 0, \widehat{p}_{23} = 0.15, \widehat{p}_{24} = 0.23$$

DUAL-DUPLEX MODEL

- Relatively sure about 1 failure every 10 hours for states with ν_1 and ν_2 $\Rightarrow$ priors $\nu_1 \sim \text{Ga}(0.1, 1)$ and $\nu_2 \sim \text{Ga}(0.1, 1)$

- Less sure about states with $\nu_3,\ \nu_4$, expecting around 5 repairs per hour $\Rightarrow$ priors $\text{Ga}(10, 2)$ and $\text{Ga}(10, 2)$ for ν_3, ν_4

- Available data: for state 1, 14 times which add up 127.42; 10 with sum 86.81, for state 2; 5 with sum 1.09, for state 3; and, 2 with sum 0.27, for state 4)

- $\Rightarrow$ Posterior parameters of permanence rates, posterior means and standard deviation

...	α	β	Mean	Std. Dev.
ν_1	0.1+14	1+127.42	0.11	0.03
ν_2	0.1+10	1+86.81	0.12	0.04
ν_3	10+5	2+1.09	4.85	1.25
ν_4	10+2	2+0.27	5.29	1.53

DUAL-DUPLEX MODEL

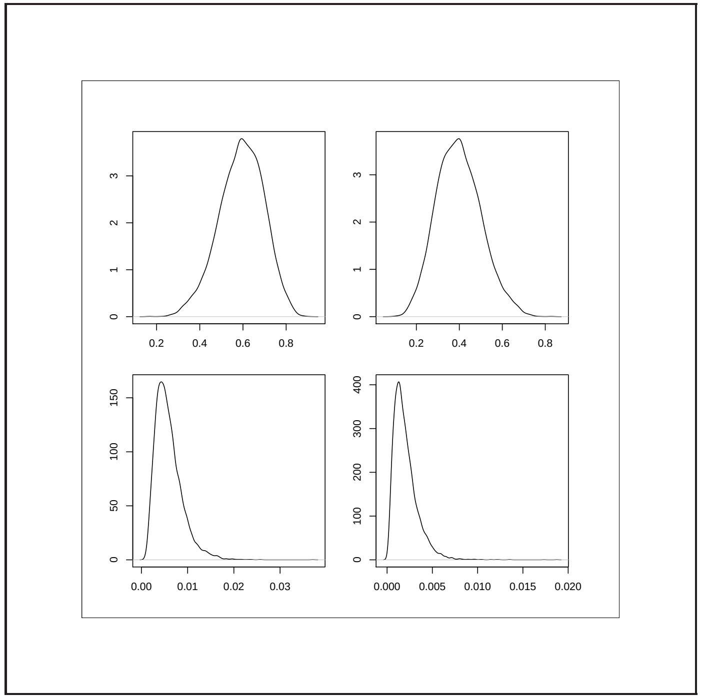
- For fixed P and ν equilibrium solution given by

$$\begin{cases} \nu_1\pi_1 = r_{21}\pi_2 + r_{31}\pi_3 + r_{41}\pi_4, \\ \nu_2\pi_2 = r_{12}\pi_1, \\ \nu_3\pi_3 = r_{13}\pi_1 + r_{23}\pi_2, \\ \nu_4\pi_4 = r_{24}\pi_2, \\ \pi_1 + \pi_2 + \pi_3 + \pi_4 = 1, \\ \pi_i \geq 0, \end{cases}$$

- Solutions, with $\Delta = \nu_2\nu_3\nu_4 + r_{12}\nu_3\nu_4 + r_{13}\nu_2\nu_4 + r_{12}r_{23}\nu_4 + r_{12}r_{24}\nu_3$
  - $\pi_1 = \dfrac{\nu_2\nu_3\nu_4}{\Delta}$
  - $\pi_2 = \dfrac{r_{12}\nu_3\nu_4}{\Delta}$
  - $\pi_3 = 1 - (\pi_1 + \pi_2 + \pi_3)$
  - $\pi_4 = \dfrac{r_{12}r_{24}\nu_3}{\Delta}$

DUAL-DUPLEX MODEL

Density plots for posterior equilibrium distribution



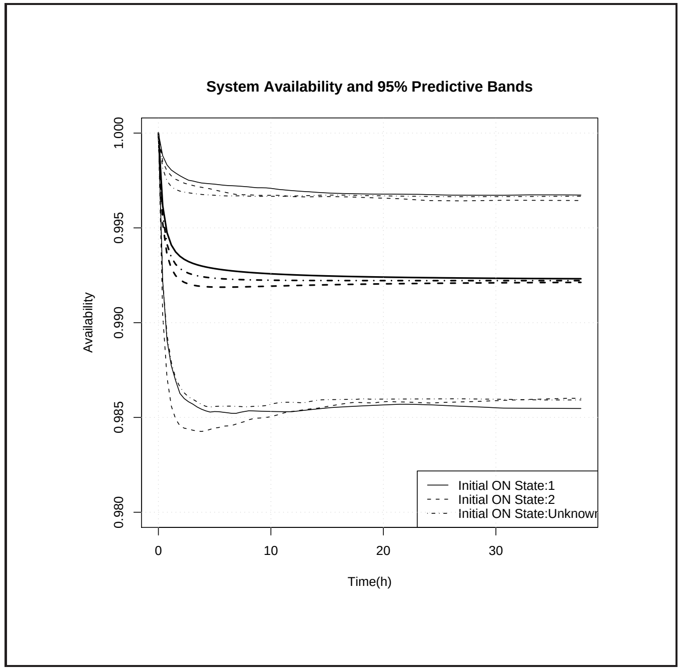
Mean probabilities

$$\widehat{\pi}_1 = 0.5931, \quad \widehat{\pi}_2 = 0.3990, \quad \widehat{\pi}_3 = 0.0059, \quad \widehat{\pi}_4 = 0.0020.$$

DUAL-DUPLEX MODEL

- Split interval $[0, t)$ into 200 subintervals
- Compute state probability vector $\pi(t)|\boldsymbol{\nu}, \mathbf{P}$ at each 200 endpoints
- Estimate system availability using Simpson's rule
- Plot of system availability for initial state 1, 2 or unknown, with 95% predictive bands around the central values

DUAL-DUPLEX MODEL



- $\Rightarrow$ Uncertainty negligible, with relative errors less than 1%
- In fact, dual-duplex system designed as a high-availability device

SEMI-MARKOVIAN PROCESSES

- Semi-Markovian processes (SMP) generalize CTMCs assuming that the times between transitions are not necessarily exponential

- $\{X_t\}_{t \in T}$ CTMC evolving within a finite state space $E = \{1, 2, \ldots, K\}$

- When the process enters state i, it remains there for a random time T_i, with parameter $\boldsymbol{\nu}_i$, positive with probability 1

- $f_i(.|\boldsymbol{\nu}_i)$ and $F_i(.|\boldsymbol{\nu}_i)$ density and distribution functions of T_i, respectively

- Define $\mu_i = E[T_i|\boldsymbol{\nu}_i]$

- When leaving i, the process moves to j with probability p_{ij}, s.t. $\sum_j p_{ij} = 1, \ \forall i$, and $p_{ii} = 0$

- As for CTMCs, the transition probability matrix, $\mathbf{P} = (p_{ij})$, defines an embedded DTMC

- Parameters for SMP: $\mathbf{P} = (p_{ij})$ and $\boldsymbol{\nu} = (\boldsymbol{\nu}_i)$

SEMI-MARKOVIAN PROCESSES

- Suppose states at transitions and times between transitions are observed $\Rightarrow$ inference about $\mathbf{P}$ as before
- Assume posterior distribution of $\boldsymbol{\nu}$ available, possibly through a sample
- Interest in long term forecasting of the proportion of time π_j the system spends in each state j
- For fixed $\mathbf{P}$ and $\boldsymbol{\nu}$, use the following procedure:
  1. Compute, if it exists, the equilibrium distribution $\bar{\boldsymbol{\pi}}$ of the embedded Markov chain, with transition matrix $\mathbf{P}$: $\bar{\boldsymbol{\pi}} = \bar{\boldsymbol{\pi}}\mathbf{P}$ and $\sum_{i=1}^{K} \bar{\pi}_i = 1, \bar{\pi}_i \geq 0$
  2. Compute, if they exist, the expected holding times at each state, $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_K)$
  3. Compute $\pi = \sum_{i=1}^{K} \bar{\pi}_i \mu_i$ (expected holding time in any state)
  4. Compute the equilibrium distribution $\boldsymbol{\pi}$ where $\pi_i = \dfrac{\bar{\pi}_i \mu_i}{\pi}$

SEMI-MARKOVIAN PROCESSES

- Uncertainty $(\boldsymbol{\nu}, \mathbf{P})$ can be incorporated into the forecasts, through their posterior predictive distributions

- As discussed before, various approaches can be considered
  - For concentrated posteriors, estimates of $\widehat{\mathbf{P}}, \widehat{\boldsymbol{\mu}}$ plugged in the previous procedure to estimate predictive equilibrium distribution $\widehat{\boldsymbol{\pi}}$
  - Under greater uncertainty
    * Monte Carlo sample, $\boldsymbol{\nu}^{(s)}, \mathbf{P}^{(s)}$, $s = 1, \ldots, S$, from their posterior distributions
    * $\Rightarrow$ sample $\boldsymbol{\pi}^{(s)}$, $s = 1, \ldots, S$, for repeated solutions of previous procedure
    * predictive equilibrium distribution estimated from posterior mean $\widehat{\boldsymbol{\pi}}_i = \frac{1}{S} \sum_{s=1}^{S} \boldsymbol{\pi}^{(s)}$
  - If previous procedure too costly computationally $\Rightarrow$ use ROM to reduce the computational load

SEMI-MARKOVIAN PROCESSES

- Conditional on parameter values, short term forecasting for a SMP involves complex numerical procedures based on Laplace-Stieltjes transforms

- $P_{ij}(t) = P(X_{t+s} = j|X_s = i) = P(X_t = j|X_0 = i)$ transition probability functions

- $f_i(.|\boldsymbol{\nu}_i)$ and $F_i(.|\boldsymbol{\nu}_i)$ density and distribution functions of T_i, respectively

- Evolution of $P_{ij}(t)$ described by forward Kolmogorov equations, for $i, j = 1, \ldots, K$

$$P_{ii}(t) = (1 - F_i(t)) + \int \sum_{k=1}^{K} p_{ik} f_i(u) P_{ki}(t-u) du$$

$$P_{ij}(t) = \int \sum_{k=1}^{K} p_{ik} f_i(u) P_{kj}(t-u) du, i \neq j$$

SEMI-MARKOVIAN PROCESSES

- Equations in matrix form

$$\mathbf{P}(t) = \mathbf{W}(t) + \int_0^t (\mathbf{PF}(u))\mathbf{P}(t-u)\,du$$

  - $\mathbf{W}(t)$ diagonal matrix with $1 - F_i(t)$ in the i-th diagonal position of the matrix
  - $\mathbf{F}(u)$ matrix with all elements in its i-th row equal $f_i(u)$

- Matrix Laplace-Stieltjes transform $\mathbf{M}^*(s)$ of the matrix function $\mathbf{M}(t)$:
$\mathbf{M}^*(s) = \int_0^\infty \mathbf{M}(t)\,\text{exp}(-st)\,dt$, with $m^*_{ij}(s) = \int_0^\infty m_{ij}(t)\,\text{exp}(-st)\,dt$

- Basic Laplace-Stieltjes transform properties $\Rightarrow \mathbf{P}^*(s) = \mathbf{W}^*(s) + (\mathbf{PF}^*(s))\mathbf{P}^*(s)$

- Simple matrix operations lead to $\mathbf{P}^*(s) = (\mathbf{I} - \mathbf{PF}^*(s))^{-1}\mathbf{W}^*(s)$

- $\mathbf{W}^*(s)$ diagonal matrix with elements $\frac{1}{s}(1 - f_i^*(s))$, because of properties of the Laplace-Stieltjes transform

- Need to find the inverse transform of $\mathbf{P}^*(s)$ to obtain $\mathbf{P}(t)$

SEMI-MARKOVIAN PROCESSES

As earlier, account for uncertainty about $(\boldsymbol{\nu}, \mathbf{P})$ and appeal to various approaches

- For sufficiently concentrated posteriors, uncertainty about $(\boldsymbol{\nu}, \mathbf{P})$ summarized through posterior modes $\widehat{\boldsymbol{\nu}}$ and $\widehat{\mathbf{P}}$
  - Obtain predictive Laplace-Stieltjes transform $(\mathbf{I} - \widehat{\mathbf{P}}\mathbf{F}^*(s|\widehat{\boldsymbol{\nu}}))^{-1}\mathbf{W}^*(s|\widehat{\nu})$
  - Invert it numerically to obtain an approximation of $\mathbf{P}(t|\text{data})$ based on $\mathbf{P}(t|\widehat{\boldsymbol{\nu}}, \widehat{\mathbf{P}})$

- For not concentrated posteriors
  - Obtain samples from the posteriors: $\{\boldsymbol{\nu}^\eta\}_{\eta=1}^N, \ \{\mathbf{P}^\eta\}_{\eta=1}^N$
  - Find the corresponding Laplace-Stieltjes transform for each sampled value

$$P^L(s|P^\eta, \nu^\eta) = (I - P^\eta F^L(s|\nu^\eta))^{-1} W^L(s|\nu^\eta)$$

  - Get the Monte Carlo approximation to the Laplace-Stieltjes transform

$$P^L_{MC}(s) = \frac{1}{N} \sum_{\eta=1}^{N} P^L(s|P^\eta, \nu^\eta)$$

SEMI-MARKOVIAN PROCESSES

- For not concentrated posteriors
  - Invert Monte Carlo approximation to the Laplace-Stieltjes transform to approximate $P(t|data)$
  - Alternatively, we could have adopted the more expensive computationally, but typically more precise procedure which consists of inverting the Laplace-Stieltjes transform at each sampled parameter and, then, form the Monte Carlo sum of inverses as an approximation to $P(t|data)$

- Too computationally intensive Laplace-Stieltjes transform computation and inversion involved in the above procedure $\Rightarrow$ use ROM, as for CTMCs
  - approximate $P^L(s)$ through $\sum_{i=1}^{m} q^\eta P^L(s|P^\eta, \nu^\eta)$
  - invert it to approximate $P(t|data)$

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it   *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/~fabrizio/

POISSON PROCESS

- One of the simplest and most applied stochastic processes
- Used to model occurrences (and counts) of rare events in time and/or space, when they are not affected by past history
- Applied to describe and forecast incoming telephone calls at a switchboard, arrival of customers for service at a counter, occurrence of accidents at a given place, visits to a website, earthquake occurrences and machine failures, to name but a few applications
- Special case of CTMCs with jumps possible only to the next higher state and pure birth processes, as well as model for arrival process in $M/G/c$ queueing systems
- Simple mathematical formulation and relatively straightforward statistical analysis $\Rightarrow$ very practical, if approximate, model for describing and forecasting many random events

POISSON PROCESS

- Counting process $N(t)$, $t \geq 0$: stochastic process counting number of events occurred up to time t

- $N(s,t]$, $s < t$: number of events occurred in time interval $(s,t]$

- Poisson process with intensity function $\lambda(t)$: counting process $N(t), t \geq 0$, s.t.

  1. $N(0) = 0$
  2. Independent number of events in non-overlapping intervals
  3. $P\left(N(t, t+\Delta t] = 1\right) = \lambda(t)\Delta t + o(\Delta t)$, as $\Delta t \to 0$
  4. $P\left(N(t, t+\Delta t] \geq 2\right) = o(\Delta t)$, as $\Delta t \to 0$

- Definition $\Rightarrow P(N(s,t] = n) = \dfrac{\left(\int_s^t \lambda(x)dx\right)^n}{n!} e^{-\int_s^t \lambda(x)dx}$, for $n \in \mathbb{Z}^+$

  $\Rightarrow N(s,t] \sim \text{Po}\left(\int_s^t \lambda(x)dx\right)$

POISSON PROCESS

- Intensity function: $\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t, t + \Delta t] \geq 1)}{\Delta t}$
  - HPP (homogeneous Poisson process): constant $\lambda(t) = \lambda$, $\forall t$
  - NHPP (nonhomogeneous Poisson process): o.w.
- HPP with rate λ
  - $N(s, t] \sim \text{Po}(\lambda(t - s))$
  - Stationary increments (distribution dependent only on interval length)

POISSON PROCESS

- Mean value function $m(t) = E[N(t)], t \geq 0$
- $m(s,t] = m(t) - m(s)$ expected number of events in $(s,t]$
- If $m(t)$ differentiable, $\mu(t) = m'(t)$, $t \geq 0$, Rate of Occurrence of Failures (ROCOF)
- $P\left(N(t, t + \Delta t] \geq 2\right) = o(\Delta t)$, as $\Delta t \to 0$
  $\Rightarrow$ orderly process
  $\Rightarrow \lambda(t) = \mu(t)$ a.e.
- $\Rightarrow m(t) = \int_0^t \lambda(x)dx$ and $m(s,t] = \int_s^t \lambda(x)dx$
- $\Rightarrow m(t) = \lambda t$ and $m(s,t] = \lambda(t-s)$ for HPP with rate λ

POISSON PROCESS

- Arrival times $\{T_n, n \in \mathbb{Z}^+\}$:

$$T_n := \begin{cases} \min\{t : N(t) \geq n\} & n > 0 \\ 0 & n = 0 \end{cases}$$

- $\{T_n, n \in \mathbb{N}^+\}$ stochastic process, sort of dual of $N(t)$

- Interarrival times $\{X_n, n \in \mathbb{Z}^+\}$:

$$X_n := \begin{cases} T_n - T_{n-1} & n > 0 \\ 0 & n = 0 \end{cases}$$

- Interarrival and arrival times related through $T_n = \sum_{i=1}^{n} X_i$

POISSON PROCESS

- In a Poisson process $N(t)$ with intensity $\lambda(t)$
  - T_n has density $g_n(t) = \dfrac{\lambda(t)[m(t)]^{n-1}}{\Gamma(n)} e^{-m(t)}$
  - X_n has distribution function, conditional upon the occurrence of the $(n-1)$-st event at T_{n-1}, given by $F_n(x) = \dfrac{F(T_{n-1}+x) - F(T_{n-1})}{1 - F(T_{n-1})}$, with $F(x) = 1 - e^{-m(x)}$
  - $\Rightarrow$ distribution function of T_1 (and X_1) given by $F_1(t) = 1 - e^{-m(t)}$, and density function $g_1(t) = \lambda(t)e^{-m(t)}$
- For HPP with rate λ
  - Interarrival times, and first arrival time, have an exponential distribution Ex(λ) ($\Rightarrow$ HPP renewal process)
  - n-th arrival time, T_n, has a gamma distribution Ga(n, λ), for each $n \geq 1$
  - Link between HPP and exponential distribution

POISSON PROCESS

Poisson process $N(t)$ with intensity function $\lambda(t)$ and mean value function $m(t)$

- $T_1 < \ldots < T_n$: n arrival times in $(0, T] \Rightarrow P(T_1, \ldots, T_n) = \prod_{i=1}^{n} \lambda(T_i) \cdot e^{-m(T)}$

  $\Rightarrow$ likelihood

- $\Rightarrow P(T_1, \ldots, T_n) = \lambda^n e^{-\lambda T}$ for HPP with rate λ

- n events occur up to time $t_0 \Rightarrow$ distributed as order statistics from cdf $m(t)/m(t_0)$, for $0 \leq t \leq t_0$ (uniform distribution for HPP)

POISSON PROCESS

- Under suitable conditions, Poisson processes can be merged or split to obtain new Poisson processes (see Kingman, p. 14 and 53, 1993)

- Useful in applications, e.g.
  - merging gas escapes from pipelines installed in different periods
  - splitting earthquake occurrences into minor and major ones

- **Superposition Theorem**
  - n independent Poisson processes $N_i(t)$, with intensity function $\lambda_i(t)$ and mean value function $m_i(t)$, $i = 1, \ldots, n$
  - $\Rightarrow N(t) = \sum_{i=1}^{n} N_i(t)$, for $t \geq 0$, Poisson process with intensity function $\lambda(t) = \sum_{i=1}^{n} \lambda_i(t)$ and mean value function $m(t) = \sum_{i=1}^{n} m_i(t)$

POISSON PROCESS

- **Coloring Theorem**
  - $N(t)$ be a Poisson process with intensity function $\lambda(t)$
  - Multinomial random variable Y, independent from the process, taking values $1, \ldots, n$ with probabilities $p_1, \ldots, p_n$
  - Each event assigned to classes (*colors*) $A_1, \ldots, A_n$ according to $Y \; Rightarrown$ independent Poisson processes $N_1(t), \ldots, N_n(t)$ with intensity functions $\lambda_i(t) = p_i\lambda(t)$, $i = 1, \ldots, n$

- Coloring Theorem extended to the case of time dependent probabilities $p(t)$, defined on $(0, \infty)$
  - As an example, for an HPP with rate λ, if events at any time t are kept with probability $p(t) \Rightarrow$ Poisson process with intensity function $\lambda p(t)$

POISSON PROCESS: INFERENCE

- $N(t)$ HPP with parameter λ
- n events observed in the interval $(0, T]$
- Likelihood for two possible experiments
  - Times $T_1, \ldots, T_n$ available

    Theorem on Poisson processes $\Rightarrow l(\lambda|data) = (\lambda T)^n e^{-\lambda T}$
  - Only number n available

    Properties of Po$(\lambda T) \Rightarrow l(\lambda|data) = \dfrac{(\lambda T)^n}{n!} e^{-\lambda T}$
- Proportional likelihoods $\Rightarrow$ same inferences (Likelihood Principle, Berger and Wolpert, 1988)
- In both cases, likelihood not dependent on the actual occurrence times but only on their number

POISSON PROCESS: INFERENCE

- Gamma priors conjugate w.r.t. λ in the HPP
- Prior Ga(α, β)
- $\Rightarrow f(\lambda|n, T) \propto (\lambda T)^n e^{-\lambda T} \cdot \lambda^{\alpha-1} e^{-\beta\lambda}$
- $\Rightarrow$ posterior Ga$(\alpha + n, \beta + T)$
- Posterior mean $\widehat{\lambda} = \dfrac{\alpha + n}{\beta + T}$
- Posterior mean combination of
  - Prior mean $\widehat{\lambda}_P = \dfrac{\alpha}{\beta}$
  - MLE $\widehat{\lambda}_M = \dfrac{n}{T}$

ACCIDENTS IN THE CONSTRUCTION SECTOR

Ríos Insua et al (1999)

- Interest in number of accidents in some companies in the Spanish construction sector
- 75 accidents and an average number of workers of 364 in 1987 for one company
- Number of workers constant during the year
- Times of all accidents of each worker are recorded
- Accidents occur *randomly* $\Rightarrow$ HPP model justified
- Each worker has the same propensity to have accidents $\Rightarrow$
  - HPP with same λ for all of them
  - If one year corresponds to $T = 1 \Rightarrow$ number of accidents for each worker follows the same Poisson distribution Po(λ)

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Accidents of different workers are independent
  - Apply Superposition Theorem
  - $\Rightarrow$ Number of accidents for all workers given by an HPP with rate 364λ
- Gamma prior Ga$(1, 1)$ on λ
  - Likelihood $l(\lambda|data) = (364\lambda)^{75}e^{-364\lambda}$
  - Posterior gamma Ga$(76, 365)$
  - Posterior mean $76/365 = 0.208$
  - Prior mean 1
  - MLE $75/364 = 0.206$
  - Posterior mean closer to MLE $\Rightarrow$ think of the hypothetical experiment (just 1 hypothetical sample w.r.t. 75 actual ones)

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Prior Ga$(1, 1) \Rightarrow$ mean 1 and variance 1
  - *large* variance in this experiment
  - $\Rightarrow$ scarce confidence on the prior assessment of mean equal to 1
- Prior Ga$(1000, 1000) \Rightarrow$ mean 1 and variance 0.001
  - *Small* variance in this experiment
  - $\Rightarrow$ strong confidence on the prior assessment of mean equal to 1
- $\Rightarrow$ Posterior Ga$(1075, 1364)$
- Posterior mean $1075/1364 = 0.79$
- Prior mean 1
- MLE $75/364 = 0.21$
- Posterior mean $10075/10364 = 0.97$ for a Ga$(10000, 10000)$ prior

POISSON PROCESS: INFERENCE

- Computation of quantities of interest
  - analytically (e.g. posterior mean and mode)
  - using basic statistical software (e.g. posterior median and credible intervals)

- Accidents in the construction sector
  - Gamma prior Ga$(100, 100)$ for the rate λ
  - Posterior mean: $175/464 = 0.377$
  - Posterior mode: $174/464 = 0.375$
  - Posterior median: 0.376
  - $[0.323, 0.435]$: 95% credible interval $\Rightarrow$ quite concentrated distribution
  - Posterior probability of interval $[0.3, 0.4]$: 0.789

NON CONJUGATE ANALYSIS

- Improper priors
  - Controversial, although rather common, choice, which might reflect lack of knowledge
  - Possible choices
    * $f(\lambda) \propto 1$: Uniform prior
      $\Rightarrow$ posterior Ga$(n+1, T)$
    * $f(\lambda) \propto 1/\lambda$: Jeffreys prior given the experiment of observing times between events
      $\Rightarrow$ posterior Ga$(n/2, T)$
    * $f(\lambda) \propto 1/\sqrt{\lambda}$: Jeffreys prior given the experiment of observing the number of events in a fixed period
      $\Rightarrow$ posterior Ga$(n+1/2, T)$

NON CONJUGATE ANALYSIS

- Lognormal prior LN$\left(\mu, \sigma^2\right)$
- $\Rightarrow$ posterior $f(\lambda|n, T) \propto \lambda^n e^{-\lambda T} \cdot \lambda^{-1} e^{-(\log\lambda-\mu)/(2\sigma^2)}$
- Normalizing constant C and other quantities of interest (e.g., the posterior mean $\hat{\lambda}$) computed numerically, using, e.g., Monte Carlo simulation

1. `Set` $C = 0$, $D = 0$ `and` $\hat{\lambda} = 0$. $i = 1$.
2. `While` $i \leq M$, `iterate through`
   - . `Generate` λ_i `from a lognormal distribution LN`$\left(\mu, \sigma^2\right)$
   - . `Compute` $C = C + \lambda_i^n e^{-\lambda_i T}$
   - . `Compute` $D = D + \lambda_i^{n+1} e^{-\lambda_i T}$
   - . $i = i + 1$
3. `Compute` $\hat{\lambda} = \frac{C}{D}$.

NON CONJUGATE ANALYSIS

- Given the meaning of λ (expected number of events in unit time interval or inverse of mean interarrival time), it may often be considered that λ is bounded

- $\Rightarrow$ Prior on a bounded set

- Uniform prior on the interval $(0, L]$

- $\Rightarrow$ Posterior $f(\lambda|n,T) \propto \lambda^n e^{-\lambda T} I_{(0,L]}(\lambda)$

- Normalizing constant $\gamma(n+1, LT)/T^{n+1}$, with $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ lower incomplete gamma function

- Posterior mean $\hat{\lambda} = \dfrac{1}{T}\dfrac{\gamma(n+2, LT)}{\gamma(n+1, LT)}$

FORECASTING

- n events observed in the interval $(0, T]$
- Interest in forecasting number of events in subsequent intervals: $P(N(T, T+s] = m)$,
- For $s > 0$ and integer m

$$\begin{aligned} P(N(T,T+s]=m) &= \int_0^\infty P(N(T,T+s]=m|\lambda)\, f(\lambda|n,T)\, d\lambda \\ &= \int_0^\infty \frac{(\lambda s)^m}{m!} e^{-\lambda s}\, f(\lambda|n,T)\, d\lambda \end{aligned}$$

Posterior Ga$(\alpha+n, \beta+T)$

$$\Rightarrow P(N(T,T+s]=m) = \frac{s^m}{m!} \frac{(\beta+T)^{\alpha+n}}{(\beta+T+s)^{\alpha+n+m}} \frac{\Gamma(\alpha+n+m)}{\Gamma(\alpha+n)}$$

FORECASTING

- Expected number of events in the subsequent interval

$$\begin{aligned} E[N(T, T+s]] &= \int_0^\infty E[N(T, T+s]|\lambda] f(\lambda|n, T) d\lambda \\ &= \int_0^\infty \lambda s \, f(\lambda|n, T) d\lambda \end{aligned}$$

Posterior Ga$(\alpha + n, \beta + T)$

$\Rightarrow E[N(T, T+s]] = s\dfrac{\alpha + n}{\beta + T}$

ACCIDENTS IN THE CONSTRUCTION SECTOR

- Gamma prior Ga$(100, 100)$ for the rate λ
- Posterior gamma Ga$(175, 464)$, having observed 75 accidents with 346 workers in 1987
- Interest in number of accidents during the first six months of 1988 (i.e. $s = 0.5$), when the number of workers has increased to 400 (i.e. $m = 400$)
- T_{1987} denotes December, 31st, 1987
- $\Rightarrow N(T_{1987}, T_{1987} + 0.5] \sim$ Po$(400\lambda \cdot 0.5)$
- $E[N(T_{1987}, T_{1987} + 0.5]] = 400 \cdot 0.5 \dfrac{175}{464} = 75.431$
- Interested in probability of 100 accidents in the six months:
$$P(N(T_{1987}, T_{1987} + 0.5] = 100) = \frac{200^{100}}{100!} \frac{464^{175}}{664^{275}} \frac{\Gamma(275)}{\Gamma(175)} = 0.003$$
- Probability of no accidents in the six months: $(464/664)^{175} \approx 0$

CONCOMITANT POISSON PROCESSES

- k Poisson processes $N_i(t)$, with parameter λ_i, $i = 1, \ldots, k$
- Observe n_i events over an interval $(0, t_i]$ for each process $N_i(t)$
- Processes could be related at different extent
- Typical example, as in Cagno et al (1999), provided by gas escapes in a network of pipelines which might differ in, e.g., location and environment
- Based on such features, pipelines could be split into subnetworks and an HPP for gas escapes in each of them is considered
- Some possible mathematical relations among the HPPs are presented, using the gas escape example for illustrative purposes

CONCOMITANT POISSON PROCESSES

- **Independence**
  - Processes correspond to completely different phenomena, e.g., gas escapes in completely different pipelines, for material, location, environment, etc.
  - Completely different λ_i, with no relation
  - Conjugate gamma priors Ga(α_i, β_i) for each process $N_i(t)$, $i = 1, \ldots, k$
  - $\Rightarrow$ Posterior distribution Ga$(\alpha_i + n_i, \beta_i + t_i)$
- **Complete similarity**
  - Identical processes with same λ, e.g., the gas pipelines are identical for material, laying procedure, environment and operation
  - Likelihood $l(\lambda|data) \propto \prod_{i=1}^{k} (\lambda t_i)^{n_i} e^{-\lambda t_i} \propto \lambda^{\sum_{i=1}^{k} n_i} e^{-\lambda \sum_{i=1}^{k} t_i}$
  - Gamma prior Ga(α, β)
  - $\Rightarrow$ Gamma posterior Ga$\left(\alpha + \sum_{i=1}^{k} n_i, \beta + \sum_{i=1}^{k} t_i\right)$

CONCOMITANT POISSON PROCESSES

Partial similarity (exchangeability):

- Processes with similar λ_i, i.e. different but from same prior, corresponding e.g. to similar, but not identical, conditions for the gas pipelines

- Hierarchical model:

$$\begin{aligned} N_i(t_i)|\lambda_i &\sim \text{Po}(\lambda_i t_i),\ i = 1, \ldots, k \\ \lambda_i|\alpha, \beta &\sim \text{Ga}(\alpha, \beta),\ i = 1, \ldots, k \\ f(\alpha, \beta) & \end{aligned}$$

- Experiment corresponding to observe k HPPs $N_i(t)$, $i = 1, \ldots, k$, with parameter λ_i, until time t_i

- Different choices for prior $f(\alpha, \beta)$ proposed in literature

CONCOMITANT POISSON PROCESSES

Different choices for prior $f(\alpha, \beta)$

- Albert (1985)
  - Model reparametrization
    - $\mu = \alpha/\beta$
    - $\gamma_i = \beta/(t_i + \beta), i = 1, \ldots, k$
  - Noninformative priors $f(\mu) = 1/\mu$ and $f(\gamma) = \gamma^{-1}(1-\gamma)^{-1}$
  - Approximations to estimate mean and variance of λ_i's

- George et al (1993): failures of ten power plants
  - Exponential priors Ex(1) or Ex(0.01) for α
  - Gamma priors Ga$(0.1, 1)$ and Ga$(0.1, 0.01)$ for β
  - Informal sensitivity analysis, with *small* and *large* values of α and β

CONCOMITANT POISSON PROCESSES

Masini et al (2006)

- Priors

$$\begin{aligned} f(\alpha) &\propto \Gamma(\alpha+1)^k/\Gamma(k\alpha+a),\ k \text{ integer } \geq 2, a > 0 \\ \beta &\sim \text{Ga}(a,b) \end{aligned}$$

- Proper prior distribution on α
  - Shape depends on the parameter a, appearing also in the prior on β
  - For $a > 1$, decreasing density for all positive α
  - For $a \leq 1$, increasing density up to its mode, $(1-a)/(k-1)$, and then decreasing
  - Numerical experiments showed wide range of behaviors representing possible different beliefs
  - Mathematical convenience: both gamma functions in the prior cancel when integrating out α and β and computing the posterior distribution of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$

CONCOMITANT POISSON PROCESSES

- Posterior distribution of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$ given by

$$f(\boldsymbol{\lambda}|data) \propto \frac{\prod_{i=1}^{k} \lambda_i^{N_i(t_i)-1} \exp\{-\lambda_i t_i\}}{(\sum_{i=1}^{k} \lambda_i + b)^a (-\log H(\boldsymbol{\lambda}))^{k+1}},$$

with $H(\boldsymbol{\lambda}) = \prod_{i=1}^{k} \lambda_i (\sum_{i=1}^{k} \lambda_i + b)^{-k}$

- Normalizing constant C computed numerically, using, e.g., Monte Carlo simulation

1. Set $C = 0$. $i = 1$.
2. Until convergence is detected, iterate through
   - For $j = 1, \ldots, k$ generate $\lambda_j^{(i)}$ from Ga$(N_j(t_j), t_j)$
   - Compute $H^{(i)}(\boldsymbol{\lambda}) = \prod_{m=1}^{k} \lambda_m^{(i)} (\sum_{n=1}^{k} \lambda_n^{(i)} + b)^{-k}$
   - Compute $C = \sum_{l=1}^{i} \frac{\prod_{j=1}^{k} \Gamma(N_j(t_j)) t_j^{-N_j(t_j)}}{(\sum_{m=1}^{k} \lambda_m^{(i)} + b)^a (-\log H^{(i)}(\boldsymbol{\lambda}))^{k+1}}$
   - $i = i + 1$

COVARIATES

- Masini et al (2006): two models considering covariates
  1. directly in the parameters
  2. in the prior distributions of the parameters
- Key idea: find relations among processes through their covariates
- Two different gas subnetworks could differ on the pipe diameter (small vs. large) but they might share the location $\Rightarrow$ data from all subnetworks used to determine contribution of the covariate (diameter) in inducing gas escapes
- m covariates taking, for simplicity, values 0 or 1 $\Rightarrow$ 2^m possible combinations
- For each combination $j, j = 1, \ldots, 2^m$
  - Covariate values $(X_{j1}, \ldots, X_{jm})$
  - $\Rightarrow$ Poisson process $N_j(t)$ with parameter $\lambda \prod_{i=1}^m \mu_i^{X_{ji}}$

    $\Rightarrow$ All null covariates $\Rightarrow$ HPP with parameter λ

COVARIATES

- Consider only one covariate ($m = 1$)
  - Only two possible combinations (e.g., small vs. large diameter in the gas pipelines)
  - $\Rightarrow$ Two HPPs $N_1(t)$ and $N_2(t)$ with rates λ and $\lambda\mu$, respectively
- Experiment
  - n_0 events observed in $(0, t_0]$ for 0-valued covariate
  - n_1 events observed in $(0, t_1]$ for 1-valued covariate

  $\Rightarrow$ likelihood $l(\lambda, \mu | data) \propto (\lambda t_0)^{n_0} e^{-\lambda t_0} \cdot (\lambda \mu t_1)^{n_1} e^{-\lambda \mu t_1}$
- Gamma priors Ga(α, β) and Ga(γ, δ) for λ and μ, respectively
- $\Rightarrow$ full conditional posteriors

$$\begin{aligned} \lambda | \mathbf{n}, \mathbf{t}, \mu &\sim \text{Ga}(\alpha + n_0 + n_1, \beta + t_0 + \mu t_1) \\ \mu | \mathbf{n}, \mathbf{t}, \lambda &\sim \text{Ga}(\gamma + n_1, \delta + \lambda t_1), \end{aligned}$$

with $\mathbf{n} = (n_0, n_1)$ and $\mathbf{t} = (t_0, t_1)$

COVARIATES

- No closed forms available for posteriors

- Sample easily obtained through Gibbs sampling

  1. `Choose initial values` λ^0, μ^0. $i = 1$.
  2. `Until convergence is detected, iterate through`
     . `Generate` $\lambda^i | \mathbf{n}, \mathbf{t}, \mu^{i-1} \sim \texttt{Ga}\big(\alpha + n_0 + n_1, \beta + t_0 + \mu^{i-1} t_1\big)$
     . `Generate` $\mu^i | \mathbf{n}, \mathbf{t}, \lambda^i \sim \texttt{Ga}\big(\gamma + n_1, \delta + \lambda^i t_1\big)$
     . $i = i + 1$

- Straightforward extension to more than one covariate
  - Independent gamma priors
  - $\Rightarrow$ Full gamma conditional posteriors
  - $\Rightarrow$ Gibbs sampling

COVARIATES

- k Poisson processes $N_i(t)$ with covariates $\mathbf{X}_i' = (X_{i1}, \ldots, X_{im})$

- Covariates introduced in previous hierarchical model

$$\begin{aligned} N_i(t_i)|\lambda_i &\sim \text{Po}(\lambda_i t_i),\ i = 1, \ldots, k \\ \lambda_i|\alpha, \boldsymbol{\beta} &\sim \text{Ga}\big(\alpha \exp\{\mathbf{X}_i'\boldsymbol{\beta}\}, \alpha\big),\ i = 1, \ldots, k \\ f(\alpha, \boldsymbol{\beta}) & \end{aligned}$$

- $\Rightarrow$ $\exp\{\mathbf{X}_i'\boldsymbol{\beta}\}$ prior mean of each λ_i

- Proper priors chosen for both α and $\boldsymbol{\beta}$

- $\Rightarrow$ posterior sampled using MCMC

- Possible alternative: $\lambda_i|\alpha, \boldsymbol{\beta} \sim \text{Ga}\big(\alpha \exp\{2\mathbf{X}_i'\boldsymbol{\beta}\}, \alpha \exp\{\mathbf{X}_i'\boldsymbol{\beta}\}\big)$

- $\Rightarrow$ prior mean $\exp\{\mathbf{X}_i'\boldsymbol{\beta}\}$ and variance $1/\alpha$ (not dependent on covariates)

COVARIATES

Empirical Bayes alternative

- Fixed α
  - Perform sensitivity analysis w.r.t. α
- Estimate $\boldsymbol{\beta}$ following an empirical Bayes approach
  - Find $\widehat{\boldsymbol{\beta}}$ maximizing $P(N_1(t_1) = n_1, \ldots, N_k(t_k) = n_k | \boldsymbol{\beta}) =$

    $$= \int P(N_1(t_1) = n_1, \ldots, N_k(t_k) = n_k | \lambda_1, \ldots, \lambda_k) f(\lambda_1, \ldots, \lambda_k | \boldsymbol{\beta}) d\lambda_1 \ldots d\lambda_k$$

- $\Rightarrow$ Independent gamma posterior distributions
  $\lambda_i | n_i, t_i \sim \text{Ga}\Big(\alpha \exp\{\mathbf{X}_i' \widehat{\boldsymbol{\beta}}\} + n_i, \alpha + t_i\Big), i = 1, \ldots, k$
- $\Rightarrow$ posterior mean $\dfrac{\alpha \exp\{\mathbf{X}_i' \widehat{\boldsymbol{\beta}}\} + n_i}{\alpha + t_i}, i = 1, \ldots, k$

NONHOMOGENEOUS POISSON PROCESS

- NHPPs characterized by intensity function $\lambda(t)$ varying over time
- $\Rightarrow$ NHPPs useful to describe (*rare*) events whose rate of occurrence evolves over time (e.g. gas escapes in steel pipelines)
  - Life cycle of a new product
    - $*$ initial elevated number of failures (*infant mortality*)
    - $*$ almost steady rate of failures (*useful life*)
    - $*$ increasing number of failures (*obsolescence*)

    $\Rightarrow$ NHPP with a *bathtub* intensity function
- NHPP has no stationary increments unlike the HPP
- Superposition and Coloring Theorems can be applied to NHPPs
- Elicitation of priors raises similar issues as before

INTENSITY FUNCTIONS

Many intensity functions $\lambda(t)$ proposed in literature (see McCollin (ESQR, 2007))

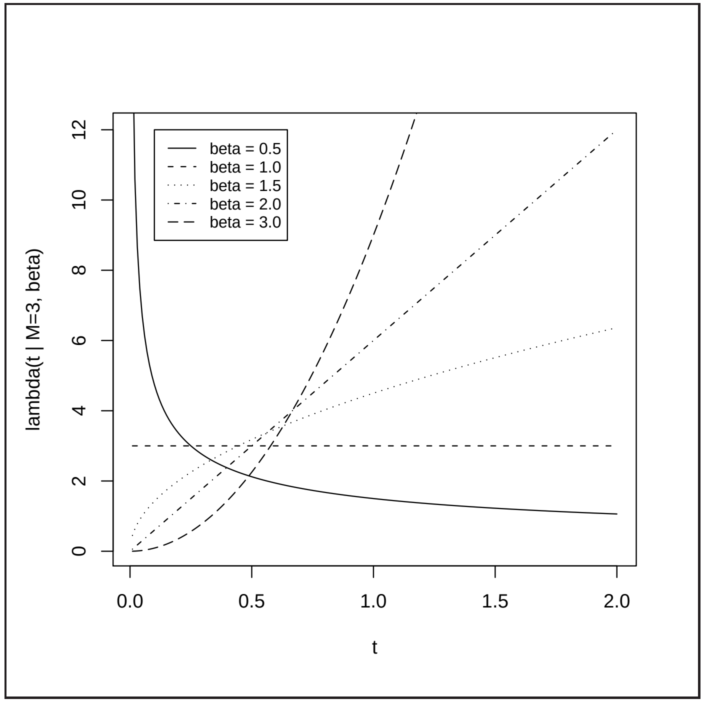
- Different origins
  - Polynomial transformations of HPP constant rate
    - $\lambda(t) = \alpha t + \beta$ (linear ROCOF model)
    - $\lambda(t) = \alpha t^2 + \beta t + \gamma$ (quadratic ROCOF model)
  - Actuarial studies (from hazard rates)
    - $\lambda(t) = \alpha\beta^t$ (Gompertz)
    - $\lambda(t) = \alpha\beta^t + \gamma t + \delta$
    - $\lambda(t) = e^{\alpha+\beta t} + e^{\gamma+\delta t}$
  - Reliability studies
    - $\lambda(t) = \alpha + \beta t + \frac{\gamma}{t+\delta}$ (quite close to *bathtub* for adequate values)
    - $\lambda(t) = \alpha\beta(\alpha t)^{\beta-1}\,\text{exp}\{\alpha t^\beta\}$ (Weibull software model)

INTENSITY FUNCTIONS

- Different origins
  - Logarithmic transformations
    - $\lambda(t) = \frac{\alpha}{t}$ ($\Rightarrow$ logarithmic $m(t)$)
    - $\lambda(t) = \alpha \log t + \alpha + \beta$
    - $\lambda(t) = \alpha \log(1 + \beta t) + \gamma$
    - $\lambda(t) = \frac{\alpha \log(1 + \beta t)}{1 + \beta t}$ (Pievatolo et al, underground train failures)
  - Associated to distribution functions
    - $\lambda(t) = \alpha f(t; \beta)$, with $f(\cdot)$ density function

NONHOMOGENEOUS POISSON PROCESS

- Different mathematical properties
  - Increasing, decreasing, convex or concave
    - $\lambda(t) = M\beta t^{\beta-1}$, $M, \beta > 0$ (Power Law Process)
    - Different behavior for different βs



NONHOMOGENEOUS POISSON PROCESS

- Different mathematical properties
  - Periodicity (Lewis)
    - $\lambda(t) = \alpha \exp\{\rho \cos(\omega t + \varphi)\}$
    - Earthquake occurrences (Vere-Jones and Ozaki, 1982)
    - Train doors' failures (Pievatolo et al., 2003)
  - Unimodal, starting at 0 and decreasing to 0 when t goes to infinity
    - Ratio-logarithmic intensity
    - $\lambda(t) = \dfrac{\alpha \log(1 + \beta t)}{1 + \beta t}$
    - Train doors' failures (Pievatolo et al., 2003)

NONHOMOGENEOUS POISSON PROCESS

- Properties of the system under consideration
  - Processes subject to faster and faster (slower and slower) occurrence of events $\Rightarrow$ increasing (decreasing) $\lambda(t)$
  - Failures of doors in subway trains, with no initial problems, then subject to an increasing sequence of failures, which later became more rare, possibly because of an intervention by the manufacturer $\Rightarrow$ ratio-logarithmic $\lambda(t)$ (Pievatolo et al., 2003)
  - New product $\Rightarrow$ life cycle described by *bathtub* intensity
  - Finite number of bugs to be detected during software testing $\Rightarrow$ $m(t)$ finite over an infinite horizon
  - Unlimited number of death in a population $\Rightarrow$ $m(t)$ infinite over an infinite horizon (as a good approximation)

CLASSES OF NHPPs

Classes of NHPPs

- Defined through density $f(t)$, with cdf $F(t)$
  - $\lambda(t) = \theta f(t)$ and $m(t) = \theta F(t)$
  - $\Rightarrow \theta$ interpreted as (finite) expected number of events over an infinite horizon
  - Number of bugs in software (Ravishanker et al.)
- Separable intensity $\lambda(t|M, \beta) = Mg(t, \beta)$
  - $M, \beta > 0$
  - g nonnegative function on $[0, \infty)$
  - Very popular intensities:
    * $g(t, \beta) = \beta t^{\beta-1}$ (Power law process)
    * $g(t, \beta) = e^{-\beta t}$ (Cox-Lewis process)
    * $g(t, \beta) = 1/(t + \beta)$ (Musa-Okumoto process)

CLASSES OF NHPPs

- *Musa and Okumoto*: $\lambda(t) \left(= [m(t)]^{'}\right) = \lambda e^{-\theta m(t)}$
  $\Rightarrow m(t) = \dfrac{1}{\theta} \log(\lambda\theta t + 1)$ for $m(0) = 0$
- *PLP*: $\lambda(t) = M\beta t^{\beta-1} \Rightarrow [m(t)]^{'} = \dfrac{\beta m(t)}{t}$
- $\lambda(t) = a(e^{bt} - 1) \Rightarrow [m(t)]^{'} = b[m(t) + at]$
- $\lambda(t) = a \log(1 + bt) \Rightarrow [m(t)]^{'} = \dfrac{b[m(t) + at]}{1 + bt}$
- $\Rightarrow \; [m(t)]^{'} = \dfrac{\alpha m(t) + \beta t}{\gamma + \delta t}$
- $y^{'} = \dfrac{\alpha y + \beta x}{\gamma + \delta x}$
- $\Rightarrow y = e^{\int \alpha/(\gamma+\delta x)dx} \left\{ \int \dfrac{\beta x}{\gamma + \delta x} e^{-\int \alpha/(\gamma+\delta x)dx} dx + c \right\}$

CLASSES OF NHPPs

$m(t)$	$\lambda(t)$
$\frac{t}{\delta}$	$\frac{1}{\delta}$
$\frac{t^2}{2\gamma}$	$\frac{t}{\gamma}$
$\frac{t}{\delta} - \frac{\gamma}{\delta^2} \log\left(1 + \frac{\delta}{\gamma} t\right)$	$\frac{t}{\gamma + \delta t}$
$\lvert c \rvert t^{\alpha/\delta}$	$\lvert c \rvert \frac{\alpha}{\delta} t^{\alpha/\delta - 1}$
$\beta\gamma\left(e^{t/\gamma} - \frac{t}{\gamma} - 1\right)$	$\beta\left(e^{t/\gamma} - 1\right)$
$\frac{\beta}{\delta - 1}\left\{t + \gamma\left[1 - \left(1 + \frac{\delta}{\gamma} t\right)^{1/\delta}\right]\right\}$	$\frac{\beta}{\delta - 1}\left\{1 - \left(1 + \frac{\delta}{\gamma} t\right)^{1/\delta - 1}\right\}$
$\beta\gamma\left(1 + \frac{t}{\gamma}\right)\log\left(1 + \frac{t}{\gamma}\right) - \beta t$	$\beta \log\left(1 + \frac{t}{\gamma}\right)$

CLASSES OF NHPPs

- *Small changes in parameters may imply significative changes in the mathematical expression for* $m(t)$
  - $\alpha = 0, \beta = 1, \gamma = 1, \delta = 2$
    $\Rightarrow m_0(t) = t/2 - (1/4)\log(1 + 2t)$
  - $\alpha > 0, \beta = 1, \gamma = 1, \delta = 2$
    $\Rightarrow m_\alpha(t) = \{t + (1/\alpha)(1 + 2t)^{\alpha/2}\}/(2 - \alpha)$
  - $\Rightarrow \lim_{\alpha \downarrow 0} m_\alpha(t) = m_0(t), \forall t$
- *Open problems*
  - Interpretation
  - Properties of the models (e.g. continuity)
  - Sensitivity and model selection

INFERENCE

- Intensity function of $N(t)$ denoted $\lambda(t|\boldsymbol{\theta})$, $\boldsymbol{\theta}$ parameter
- Events observed at times $T_1 < \ldots < T_n$ in $(0, T]$
- Likelihood $l(\boldsymbol{\theta}|T_1, \ldots, T_n) = \prod_{i=1}^{n} \lambda(T_i|\boldsymbol{\theta}) \cdot e^{-m(T)}$
- Class of NHPPs with $\lambda(t) = \theta f(t|\omega)$
- $\Rightarrow l(\theta, \omega|T_1, \ldots, T_n) = \theta^n \prod_{i=1}^{n} f(T_i|\omega) \cdot e^{-\theta F(T|\omega)}$
  - Exponential distribution Ex(ω): $f(t|\omega) = \omega e^{-\omega t}$ and $F(t|\omega) = 1 - e^{-\omega t}$
  - $\Rightarrow l(\theta, \omega|T_1, \ldots, T_n) = \theta^n \omega^n e^{-\omega \sum_{i=1}^{n} T_i - \theta(1 - \exp\{-\omega T\})}$

INFERENCE

- Likelihood $l(\theta, \omega | T_1, \ldots, T_n) = \theta^n \prod_{i=1}^{n} f(T_i|\omega) \cdot e^{-\theta F(T|\omega)}$
- Independent priors $\theta \sim \text{Ga}(\alpha, \delta)$ and $f(\omega)$
- $\Rightarrow$ posterior conditionals

$$\begin{aligned} \theta | T_1, \ldots, T_n, \omega &\sim \text{Ga}(\alpha + n, \delta + F(T|\omega)) \\ \omega | T_1, \ldots, T_n, \theta &\propto \prod_{i=1}^{n} f(T_i|\omega) e^{-\theta F(T|\omega)} f(\omega) \end{aligned}$$

- Sample from posterior applying Metropolis step within Gibbs sampler

INFERENCE

- Class of NHPPs with $\lambda(t|M,\beta) = Mg(t,\beta)$
- Likelihood $l(M,\beta|T_1,\ldots,T_n) = M^n \prod_{i=1}^{n} g(T_i,\beta) \cdot e^{-MG(T,\beta)}$
- $G(t,\beta) = \int_0^t g(u,\beta)du$
- Independent priors $M \sim$ Ga(α,δ) and $f(\beta)$
- $\Rightarrow$ posterior conditionals

$$
\begin{aligned}
M|T_1,\ldots,T_n,\beta &\sim \text{Ga}(\alpha+n,\delta+G(T,\beta)) \\
\beta|T_1,\ldots,T_n,M &\propto \prod_{i=1}^{n} g(T_i,\beta)e^{-MG(T,\beta)}f(\beta)
\end{aligned}
$$

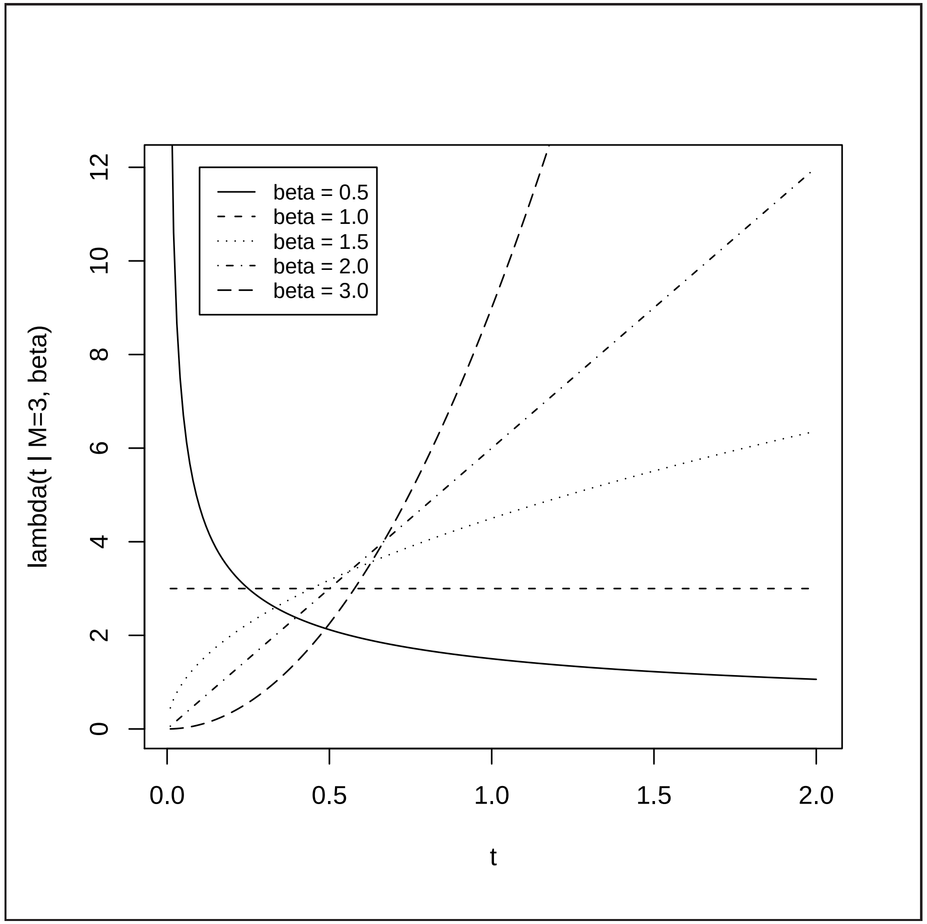
- Sample from posterior applying Metropolis step within Gibbs sampler

NONHOMOGENEOUS POISSON PROCESS

$N(t)$ Power Law process (PLP) (or Weibull process)

- Two parameterizations:
  - $\lambda(t|\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$ and $m(t;\alpha,\beta) = \left(\frac{t}{\alpha}\right)^{\beta}$, $\alpha,\beta,t > 0$
  - $\lambda(t;M,\beta) = M\beta t^{\beta-1}$ and $m(t;M,\beta) = Mt^{\beta}$, $M,\beta > 0$
  - Link: $\alpha^{-\beta} = M$
- Parameters interpretation
  - $\beta > 1 \Rightarrow$ reliability decay
  - $\beta < 1 \Rightarrow$ reliability growth
  - $\beta = 1 \Rightarrow$ constant reliability
  - $M = m(1)$ expected number of events up to time 1

POWER LAW PROCESS



FREQUENTIST ANALYSIS

Failures $\underline{T} = (T_1, \ldots, T_n) \Rightarrow$ likelihood

$$l(\alpha, \beta \mid \underline{T}) = (\beta/\alpha)^n \prod_{i=1}^{n} (T_i/\alpha)^{\beta-1}\, e^{-(y/\alpha)^\beta}$$

- *Failure truncation* $\Rightarrow y = T_n$

  MLE: $\widehat{\beta} = n / \sum_{i=1}^{n-1} \log(T_n/T_i)$ and $\widehat{\alpha} = T_n / n^{1/\widehat{\beta}}$

  C.I. for $\beta : \left(\widehat{\beta}\chi^2_{\gamma/2}(2n-2)/(2n), \widehat{\beta}\chi^2_{1-\gamma/2}(2n-2))/(2n)\right)$

- *Time truncation* $\Rightarrow y = T$

  MLE: $\widehat{\beta} = n / \sum_{i=1}^{n} \log(T/T_i)$ and $\widehat{\alpha} = T / n^{1/\widehat{\beta}}$

  C.I. for $\beta : \left(\widehat{\beta}\chi^2_{\gamma/2}(2n)/(2n), \widehat{\beta}\chi^2_{1-\gamma/2}(2n)/(2n)\right)$

Unbiased estimators, $\widehat{\lambda}(t)$, approx. C.I., hypothesis testing, goodness-of-fit, etc.

BAYESIAN ANALYSIS

Failure truncation $\equiv$ Time truncation

$$l(\alpha, \beta \mid \underline{T}) = (\beta/\alpha)^n \prod_{i=1}^{n} (T_i/\alpha)^{\beta-1}\, e^{-(y/\alpha)^\beta}$$

- $\pi(\alpha, \beta) \propto (\alpha\beta^\gamma)^{-1} \quad \alpha > 0, \beta > 0, \gamma = 0, 1 \Rightarrow \beta|\underline{T} \sim \widehat{\beta}\chi^2_{2(n-\gamma)}/(2n)$
  - Posterior exists, except for $\gamma = 0$ and $n = 1$
  - $\widehat{\beta} = n/\sum_{i=1}^{n} \log(T/T_i)$
  - Posterior mean $\tilde{\beta} = (n - \gamma)/\sum_{i=1}^{n} \log(T/T_i)$
  - Credible intervals easily obtained with standard statistical software
- $\pi(\alpha) \propto \alpha^{-1}$ and $\beta \sim \mathrm{U}(\beta_1, \beta_2) \Rightarrow \pi(\beta|\underline{T}) \propto \beta^{n-1} \prod_{i=1}^{n} T_i^\beta I_{[\beta_1,\beta_2]}(\beta)$
- $\pi(\alpha|\beta) \propto \beta s^{a\beta}\alpha^{-a\beta-1}e^{-b(s/\alpha)^\beta} \quad a, b, s > 0$ and $\beta \sim \mathrm{U}(\beta_1, \beta_2)$

$$\Rightarrow \pi(\beta|\underline{T}) \propto \beta^n \prod_{i=1}^{n} \left(\frac{T_i}{s}\right)^\beta \left[(\frac{T_n}{s})^\beta + b\right]^{-n-a} I_{[\beta_1,\beta_2]}(\beta)$$

- In all case $\alpha|\underline{T}$ by simulation (but $\alpha|\beta, \underline{T}$ inverse of a Weibull)

BAYESIAN ANALYSIS

Other parametrization

- $l(M, \beta \mid T_1, \ldots, T_n) = M^n \beta^n \prod_{i=1}^{n} T_i^{\beta-1} e^{-MT^\beta}$

- Independent priors $M \sim \text{Ga}(\alpha, \delta)$ and $\beta \sim \text{Ga}(\mu, \nu)$

- Possible dependent prior: $M|\beta \sim \text{Ga}\big(\alpha, \delta^\beta\big)$

- $\Rightarrow$ posterior conditionals (in red changes for dependent prior)

$$\begin{aligned} M|T_1, \ldots, T_n\beta &\sim \text{Ga}\big(\alpha + n, \delta^{\beta} + T^\beta\big) \\ \beta|T_1, \ldots, T_n M &\propto \beta^{\mu+n-1} \exp\{\beta(\sum_{i=1}^{n} \log T_i - \nu) - MT^\beta - M\delta^\beta\} \end{aligned}$$

- Sample from posterior applying Metropolis step within Gibbs sampler

Interest in posterior $\mathcal{E}\beta, \mathcal{P}\{\beta < 1\}$, modes, C.I.'s, $\mathcal{E}M$ (for $\lambda(t) = M\beta t^{\beta-1}$)

BAYESIAN ROBUSTNESS

- Interruption dates for a 115 kV transmission line (Rigdon and Basu, 1989)
- 13 failure dates available, from July 15, 1963 to December 19, 1971
  - 12 failure times, assuming first failure date as time 0
  - Data time truncated on December 31, 1971
- Re-scale failure times so that $[15/7/1963, 31/12/1971]$ becomes $[0, 1]$
- $\Rightarrow$ Factor e^{-MT^β} in likelihood becomes e^{-M}

  (Simplifying assumption for illustrative purposes)
- Likelihood $l(\beta, M|t_1, \ldots, t_{12}) = \beta^{12}u^{\beta-1}M^{12}e^{-M}$, with

  (re-scaled) failure times t_i's and their product u

BAYESIAN ROBUSTNESS

- Independent priors on M and $\beta \Rightarrow$ independent posterior $\Rightarrow$ focus only on β
- Prior P_0: $\beta \sim \text{Ga}(2, 1.678) \Rightarrow$ prior median $1 \Rightarrow$ reliability growth and decay are equally likely, a priori
- $\Rightarrow 0.722$ posterior mean of β
- ϵ–contamination class of priors around P_0
  - $\Gamma_\epsilon = \{P : P = (1-\epsilon)P_0 + \epsilon Q, Q \in \mathcal{Q}\}$
  - $\epsilon = 0.1$
  - $\mathcal{Q}$ class of all probability measures
- $\sup_{\Gamma_\epsilon} E[\beta|data] = 0.757$ (for Dirac measure at 0.953)
- $\inf_{\Gamma_\epsilon} E[\beta|data] = 0.680$ (for Dirac measure at 0.509)
- Very robust estimates $\Rightarrow$ reliability growth

CHANGE POINTS

- *Bathtub* shaped intensity function $\lambda(t)$ describes life cycle of a new product with an initial decreasing part, a constant part and a final increasing one

- $\lambda(t)$ could be described by the intensity functions of three distinct PLPs
  - First part: $\beta_1 < 1$, M_1
  - Second part: $\beta_2 = 1$, M_2
  - Third part: $\beta_3 > 1$, M_3

- Need to estimate
  - $\beta_1, \beta_2, \beta_3$
  - M_1, M_2, M_3
  - Change points t_1 and t_2

CHANGE POINTS

PLP but valid for previous general class

Changes at each failure time

- Hierarchical Model
  - β_i i.i.d. LN$\big(\phi, \sigma^2\big), i = 0, \ldots, n$
  - $\phi \sim$ N$\big(\mu, \tau^2\big)$
  - $\sigma^2 \sim$ IGa(ρ, γ)
- Gamma prior for M
- Conditional posteriors
  - Gamma for M
  - Inverse Gamma for σ^2
  - Normal for ϕ
  - Known (apart from a constant) for β_i's

$\Rightarrow$ Metropolis-Hastings and Gibbs sampling

CHANGE POINTS

Changes at each failure time

- Dynamic Model
  - $\log \beta_i = \log a + \log \beta_{i-1} + \epsilon_i, i = 1, \ldots, n$
  - $\epsilon_i \sim \mathrm{N}\big(0, \sigma^2\big)$
- Priors
  - Gamma for M
  - Inverse Gamma for σ^2
  - Lognormal for $a \mid \sigma^2$
  - Lognormal for $\beta_0 \mid \sigma^2$
- Conditional posteriors
  - Gamma for M
  - Inverse Gamma for σ^2
  - Lognormal for a
  - Known (apart from a constant) for β_i's

$\Rightarrow$ Metropolis-Hastings and Gibbs sampling

CHANGE POINTS

Changes at a random number of failures

- Dynamic model as before
- Bernoulli r.v.'s for change/no change
- Beta priors on Bernoulli parameter

Changes at a random number of points
$\Rightarrow$ Reversible jump MCMC with steps:

- change of M and β at a randomly chosen change point
- change to the location of a randomly chosen change point
- "birth" of a new change point at a randomly chosen location in $(0, y]$
- "death" of a randomly chosen change point

COAL-MINING DISASTERS

Ruggeri and Sivaganesan (2005)

- Dates of British serious coal-mining disasters, between 1851 and 1962: a well-known data set for change point analysis
- RJMCMC to find change points
- Posterior probabilities for number of change points

k	0	1	2	3
Prob.	0.01	0.85	0.14	0.09

- Strong evidence in favor of one point (like in Raftery and Akman, 1986) but some weak evidence for 2 points
- March 1892: posterior median of the change point (conditional on having only a single change point)
- April 1886 - June 1896: 95% equal tail credible interval

COMPOUND POISSON PROCESSES

- In Poisson processes events occur individually $\Rightarrow$ sometimes a limitation
- Batch arrivals
  - passengers exiting a bus at a bus stop
  - arrival of multiple claims to an insurance company
- $\Rightarrow$ Compound Poisson process (see, e.g., Snyder and Miller, 1991), generalizes the Poisson process to allow for multiple arrivals
  - $N(t)$ Poisson process with intensity $\lambda(t)$ (*center of clusters*)
  - Sequence of i.i.d. random variables $\{Y_i\}$, independent of $N(t)$ (*size of jump*)
  - $\Rightarrow$ Compound Poisson process: counting process (*jump process*) defined by
    * $S(t) = \sum_{i=1}^{N(t)} Y_i$,
    * with $S(t) = 0$ when $N(t) = 0$

COMPOUND POISSON PROCESSES

- $S(t)$ compound Poisson process $\Rightarrow$ for $t < \infty$
  1. $E[S(t)] = E\left[E\left[\sum_{i=1}^{n} Y_i | N(t) = n\right]\right] = m(t)E[Y_i]$,
  2. $V[S(t)] = E\left[V\left[\sum_{i=1}^{n} Y_i | N(t) = n\right]\right] + V\left[E\left[\sum_{i=1}^{n} Y_i | N(t) = n\right]\right]$
     $= E[N(t)]V[Y_i] + V[N(t)](E[Y_i])^2 = m(t)E\left[Y_i^2\right]$.

- Inference for compound Poisson processes $\Rightarrow$ complex task, as shown by a simple example
  - $N(t)$: HPP with rate λ
  - $\{Y_i\}$: i.i.d. exponential Ex(μ) r.v.'s

COMPOUND POISSON PROCESSES

- Suppose $S(t) = s$ observed, with $t, s > 0$
- Likelihood from

$$
\begin{aligned}
f(S(t) = s) &= \sum_{n=1}^{\infty} f\left(\sum_{i=1}^{n} Y_i = s | N(t) = n\right) P(N(t) = n) \\
&= \sum_{n=1}^{\infty} \frac{\mu^n}{n-1!} s^{n-1} e^{-\mu s} \cdot \frac{(\lambda t)^n}{n!} e^{-\lambda t},
\end{aligned}
$$

- since sum of n i.i.d. exponential Ex$(\mu) \Rightarrow$ Ga(n, μ)
- Gamma priors Ga(α, β) and Ga(γ, δ) for μ and λ, respectively
- $\Rightarrow$ posterior

$$
f(\mu, \lambda | S(t) = s) \propto \sum_{n=1}^{\infty} \frac{s^{n-1}}{n-1!} \frac{\beta^{\alpha}}{\Gamma(\alpha)} \mu^{\alpha+n-1} e^{-(\beta+s)\mu} \cdot \frac{t^n}{n!} \frac{\delta^{\gamma}}{\Gamma(\gamma)} \lambda^{\gamma+n-1} e^{-(\delta+t)\lambda},
$$

COMPOUND POISSON PROCESSES

- Integrating w.r.t. $\mu \Rightarrow$ posterior on λ

$$f(\lambda|S(t)=s) \propto \sum_{n=1}^{\infty} \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)} \frac{\beta^{\alpha}}{(\beta+s)^{\alpha+n}} \frac{s^{n-1}}{n-1!} \cdot \frac{\Gamma(\gamma+n)}{\Gamma(\gamma)} \frac{\delta^{\gamma}}{(\delta+t)^{\gamma+n}} \frac{t^n}{n!} g(\gamma+n, \delta+t),$$

  with $g(a,b)$ density function of a gamma Ga(a,b) random variable

- Similar for μ, integrating w.r.t. λ

- Multiple observations $\Rightarrow$ very cumbersome computations

- $E[S(T)] = \int m(t|\lambda) E[Y_i|\mu] f(\mu, \lambda|S(t)=s) d\mu d\lambda$

COMPOUND POISSON PROCESSES

- Widely used to model insurance claims
- $X(t) = X(0) + ct - \sum_{i=1}^{N(t)} Y_i$
  - $X(0)$ initial insurer's reserve
  - $X(t)$ insurer's reserve at time t
  - c constant premium (paid to insurer) rate
  - $N(t)$ number of claims up to time t
  - Y_i, $i = 1, \ldots, N(t)$, amount of i-th claim
- Interest in ruin probability

MODULATED POISSON PROCESS

- Simple extension of Poisson process: introduction of covariates in $\lambda(t)$
  - Masini et al (2006): rate λ of HPP multiplied by factors $\mu_i^{X_{ji}}$ depending on covariates X_{ji} taking values 0 or 1

- Events occur according to a modulated Poisson process if
  - $\lambda_i(t) = \lambda_0(t)e^{\mathbf{X}_i'\boldsymbol{\beta}}$ for $i = 1, \ldots, n$, with
  - $\lambda_0(t)$ baseline intensity function
  - $\mathbf{X}_i = (X_{i1}, \ldots, X_{im})$, $i = 1, \ldots, n$ different combinations of covariates
  - m-variate parameter $\boldsymbol{\beta}$

INFERENCE

- Each combination of covariates produces a Poisson process $N_i(t)$
- Superposition Theorem $\Rightarrow$ unique Poisson process $N(t)$
- Bayesian inference for general modulated Poisson processes very similar to one for HPP, with the addendum of a distribution over the parameter $\boldsymbol{\beta}$
- Call center arrival data (Soyer and Tarimcilar, 2008, and Landon et al, 2010)
  - Calls typically linked to individual advertisements
  - Interest in evaluating the efficiency of advertisements
  - X_i vector of covariates describing characteristics of i-th advertisement
    - media expenditure (in $'s)
    - venue type (monthly magazine, daily newspaper, etc.)
    - ad format (full page, half page, color, etc.)
    - offer type (free shipment, payment schedule, etc.)
    - seasonal indicators

SELF-EXCITING PROCESSES

- In a Poisson process occurrence of events does not affect intensity function at later times
- Such property not always realistic
  - Sequence of aftershocks after major earthquake
  - Introduction of new bugs during software testing and debugging
- $\Rightarrow$ Self-exciting process (SEP) introduced to describe phenomena in which occurrences affect next ones (Snyder and Miller, 1991, Ch. 6, and Hawkes and Oakes, 1974)
- Deterministic intensity function of NHPP

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{P(N(t, t + \Delta t] \geq 1)}{\Delta t}$$

SELF-EXCITING PROCESSES

- Intensity process associated to SEP, dependent on history

$$\lambda(t) = \mu(t) + \sum_{j=1}^{N(t^-)} g_j(t - T_j),$$

  - $\mu(t)$ deterministic function
  - T_j's occurrence times
  - g_j's nonnegative functions expressing influence of past observations on the intensity process

- Likelihood function formally similar to NHPP's

  - $P(T_1, \ldots, T_n) = \prod_{i=1}^{n} \lambda(T_i) \cdot e^{-\int_0^T \lambda(t)dt}$, with
  - $T_1 < \ldots < T_n$ arrival times in $(0, T]$

SELF-EXCITING PROCESSES

- New tyre of a bicycle goes flat according to an HPP with rate λ
  - No degradation over $[0, T]$, if not punctured
  - Punctures occur *randomly*
- Flat tyre occurs at $T_1 < \ldots < T_n$ over $[0, T]$
  - Tyre repaired at each T_i but more prone to new failures
  - $\Rightarrow$ Add μ_i to previous rate of HPP
  - $\Rightarrow$ Stepwise HPP, very simple example of SEP, with $\lambda(t) = \lambda + \sum_{i=1}^{N(t^-)} \mu_i I_{(T_i,\infty)}(t)$
- Likelihood $\prod_{i=1}^{n} \left(\lambda + \sum_{j=1}^{i-1} \mu_j \right) e^{-\lambda T - \sum_{i=1}^{n} \mu_i (T - T_i)}$
- Gamma priors on λ and μ_i's
- $\Rightarrow$ Mixtures of gamma distributions as full conditional posteriors

DOUBLY STOCHASTIC POISSON PROCESSES

- Intensity function of a self-exciting process $\Rightarrow$ (random) intensity process
  - dependent on $N(t)$ itself
  - paths known when observing events in $N(t)$

- Doubly stochastic Poisson process or Cox process (Cox, 1955)
  - Extension of Poisson processes, allowing for unknown paths of the intensity process, when only $N(t)$ is given
  - Two step randomization procedure:
    process $\Lambda(t)$ used to generate another process $N^*(t)$ acting as its intensity
    * $N(t)$ is a Poisson process on $[0, \infty)$
    * $\Lambda(t)$ stochastic process, independent from $N(t)$, with nondecreasing paths, s.t. $\Lambda(0) \geq 0$
    * $\Rightarrow N^*(t) = N(\Lambda(t))$ doubly stochastic Poisson process

DOUBLY STOCHASTIC POISSON PROCESSES

- Definition $\Rightarrow N(t)$ Poisson process conditional on sample path $\lambda(t)$ of process $\Lambda(t)$
- $\lambda(t)$ deterministic $\Rightarrow N(t)$ Poisson process
- $\Lambda(t) \equiv \Lambda$ r.v. $\Rightarrow$ mixed Poisson process
- Very few papers on Bayesian analysis of doubly stochastic Poisson processes
  - Gutiérrez-Peña and Nieto-Barajas (2003) modelled $\Lambda(t)$ with a gamma process
  - Varini and Ogata (forthcoming) on seismic applications
- Problem: repeated observations (paths) of the process needed to estimate intensity process and avoid indistinguishability from a NHPP based on a single path

MARKED POISSON PROCESSES

- Points of Poisson process might be labeled with some extra information
- Observations become pairs (T_i, m_i)
  - T_i occurrence time
  - m_i (*mark*) outcome of an associated random variable
- Thinning (or Coloring Theorem)
  - Equivalent to introducing a mark m valued $\{1, \ldots, n\}$ and
  - assigning the event to the class A_m in the family of mutually exclusive and exhaustive classes $\{A_1, \ldots, A_n\}$

EARTHQUAKE OCCURRENCES

- Earthquakes as point events subject to randomness

- Earthquakes occurrences often modeled as realizations of a point process, since Vere-Jones (1970) (see Vere-Jones, 2011, for an account of the history of stochastic models used in analyzing seismic activities)

- Some stochastic processes
  - NHPP with $\lambda(t) = \alpha \exp\{\rho \cos(\omega t + \varphi)\}$ (Vere-Jones and Ozaki, 1982)
  - Marked Poisson processes used to jointly model occurrence and magnitude of earthquakes (Rotondi and Varini, 2003)
  - Stress release model to analyze data in the Italian Sannio-Matese-Ofanto-Irpinia region (considered later) (Rotondi and Varini, 2007, but model introduced by Vere-Jones, 1978)
    * Justified by Reid's physical theory: stress in a region accumulates slowly over time, until it exceeds the strength of the medium, and, then, it is suddenly released and an earthquake occurs

EARTHQUAKE OCCURRENCES

- Data from Sannio Matese, area in southern Italy subject to a consistent, sometimes very disruptive, seismic activity
  - 6.89 magnitude earthquake on 23/11/1980 in Irpinia caused many casualties and considerable damage
- Shocks in Sannio Matese since 1120 catalogued exhaustively in Postpischl (1985), using current and historical data such as church records
- For each earthquake, the catalogue contains many data
  - Occurrence time (often up to the precise second)
  - Latitude and longitude
  - Intensity (strength of shaking at location as determined from effects on people and environment)
  - Magnitude (energy released at the source of the earthquake, measured by seismographs or, earlier, computed from intensity)
  - Name of the place of occurrence

EARTHQUAKE OCCURRENCES

- Exploratory data analysis identified three different behaviors of the occurrence time process since 1120 (Ladelli et al, 1992)
- More formal analysis $\Rightarrow$ change-point model as before
- Consider earthquakes in the third period, i.e. from 1860 up to 1980
- Presence of foreshocks and aftershocks, sometimes hardly recognized $\Rightarrow$ consider one earthquake, the strongest, as the main shock in any sequence lasting one week
- Sannio Matese divided into three sub-regions, relatively homogeneous from a geophysical viewpoint

EARTHQUAKE OCCURRENCES

- Marked Poisson model (occurrence time and magnitude)

- X interoccurrence times (in years) of a major earthquake (i.e. with magnitude not smaller than 5)
  - First interoccurrence time given by elapsed time between first and second earthquake

- Y' magnitude of a major earthquake

- Z number of minor earthquakes occurred in a given area since previous major one

- Earthquakes occur according to an HPP

- Each earthquake has probability p of being a major earthquake (and $1 - p$ of being a minor one)

EARTHQUAKE OCCURRENCES

- Coloring Theorem $\Rightarrow$ decompose the Poisson process into two independent processes with respective rates λp (major) and $\lambda(1-p)$ (minor)

- $\Rightarrow$ Interoccurrence times $X \sim$ Ex(λp)

- $\Rightarrow$ Conditionally on time x (realization of X), $Z \sim$ Po$(\lambda(1-p)x)$

- $\Rightarrow Z \sim$ Ge(p) since, for $z \in \mathbb{N}$,

$$\begin{aligned} P(Z=z) &= \int_0^\infty P(Z=z|X=x)f(x)dx \\ &= \int_0^\infty e^{-\lambda(1-p)x}\frac{[\lambda(1-p)x]^z}{z!}\cdot \lambda p e^{-\lambda p x}dx = p(1-p)^z \end{aligned}$$

EARTHQUAKE OCCURRENCES

- Magnitude $Y^{'}$ independent on X and Z
- Although continuous, model $Y^{'}$ as a discrete r.v. which gets the values $(5, 5.1, 5.2, \ldots)$ (one decimal, in general, in data recorded in the earthquakes catalogue)
- Actually consider $Y = 10(Y^{'} - 5) \sim$ Ge(μ)
- Approximation works well for $\mu \simeq 1$
  - $\Rightarrow \sum_{k=K}^{\infty} \mu(1-\mu)^k = (1-\mu)^K \simeq 0$, even for small K
  - $\Rightarrow$ Quite small probability of Y being large
- Joint density of (X, Y, Z) given by

$$f(x, y, z) = f(x)\, P(Y = y)\, P(Z = z|X = x) = \lambda p e^{-\lambda x} \frac{[\lambda(1-p)x]^z}{z!} \mu(1-\mu)^y$$

- $\Rightarrow$ Likelihood, for n observations $\{X_i, Y_i, Z_i\}$

$$l(p, \lambda, \mu|data) = \lambda^{n+\sum Z_i} e^{-\lambda \sum X_i} p^n (1-p)^{\sum Z_i} \mu^n (1-\mu)^{\sum Y_i} \prod_{i=1}^{n} \frac{X_i^{Z_i}}{Z_i!}$$

EARTHQUAKE OCCURRENCES

- Independent priors
  - $p \sim \text{Be}(\alpha_1, \beta_1)$
  - $\lambda \sim \text{Ga}(\alpha_2, \beta_2)$
  - $\mu \sim \text{Be}(\alpha_3, \beta_3)$
- $\Rightarrow$ Independent posterior distributions
  - $p|\text{data} \sim \text{Be}(n + \alpha_1, \sum Z_i + \beta_1)$
  - $\lambda|\text{data} \sim \text{Ga}(n + \sum Z_i + \alpha_2, \sum X_i + \beta_2)$
  - $\mu|\text{data} \sim \text{Be}(n + \alpha_3, \sum Y_i + \beta_3)$

EARTHQUAKE OCCURRENCES

- Posterior mean
  - $E[p|\text{data}] = \dfrac{n+\alpha_1}{n+\alpha_1+\sum Z_i+\beta_1}$
  - $E[\lambda|\text{data}] = \dfrac{n+\sum Z_i+\alpha_2}{\sum X_i+\beta_2}$
  - $E[\mu|\text{data}] = \dfrac{n+\alpha_3}{n+\alpha_3+\sum Y_i+\beta_3}$

EARTHQUAKE OCCURRENCES

- Predictive densities for X_{n+1}, Y_{n+1} and Z_{n+1}

$$f(x_{n+1}|\text{data}) = \frac{(n+\sum Z_i+\alpha_2)(\sum X_i+\beta_2)^{n+\sum Z_i+\alpha_2}\Gamma(n+\alpha_1+\sum Z_i+\beta_1)}{\Gamma(n+\alpha_1)\Gamma(\sum Z_i+\beta_1)}\times$$

$$\times\int_0^1 \frac{p^{n+\alpha_1}(1-p)^{\sum Z_i+\beta_1-1}}{(px_{n+1}+\sum X_i+\beta_2)^{n+\sum Z_i+\alpha_2+1}}dp,$$

$$P(Y_{n+1}=y_{n+1}|\text{data}) = \frac{(n+\alpha_3)\Gamma(n+\alpha_3+\sum Y_i+\beta_3)\Gamma(\sum Y_i+\beta_3+y_{n+1})}{\Gamma(\sum Y_i+\beta_3)\Gamma(n+\alpha_3+\sum Y_i+\beta_3+y_{n+1}+1)},$$

$$P(Z_{n+1}=z_{n+1}|\text{data}) = \frac{(n+\alpha_1)\Gamma(n+\alpha_1+\sum Z_i+\beta_1)\Gamma(\sum Z_i+\beta_1+z_{n+1})}{\Gamma(\sum Z_i+\beta_1)\Gamma(n+\alpha_1+\sum Z_i+\beta_1+z_{n+1}+1)},$$

for $x_{n+1}\geq 0$ and $y_{n+1}, z_{n+1}\in\mathbb{N}$

EARTHQUAKE OCCURRENCES

- $p \sim$ Be(α_1, β_1), $\lambda \sim$ Ga(α_2, β_2) and $\mu \sim$ Be(α_3, β_3) independent a priori

- p: probability that, given an earthquake occurs, it is a major one
  - Major earthquakes very unlikely w.r.t. minor ones $\Rightarrow$ assume $E[p]$ close to 0
  - $\alpha_1 = 2$ and $\beta_1 = 8 \Rightarrow E[p] = 1/5$

- Major earthquake occurs, on average, every ten years
  - $\Rightarrow E[X] = 10 = 1/(\lambda p)$
  - $\alpha_2 = 2$ and $\beta_2 = 4 \Rightarrow E[\lambda] = 1/2 \Rightarrow E[\lambda]E[p] = 1/10$

- As discussed earlier, μ very close to 1
  - $\alpha_3 = 8$ and $\beta_3 = 2 \Rightarrow E[\mu] = 4/5$

- Prior variances denote strong beliefs:
  $V[p] = 4/275$, $V[\lambda] = 1/8$ and $V[\mu] = 4/275$

EARTHQUAKE OCCURRENCES

Number and sums of observations in three areas in Sannio Matese

	n	$\sum x_i$	$\sum y_i$	$\sum z_i$
zone 1	3	50.1306	9	713
zone 2	16	81.6832	53	1034
zone 3	14	118.9500	100	812

Parameters of posterior distributions

	$p \sim$ Be	$\lambda \sim$ Ga	$\mu \sim$ Be
zone 1	(5,721)	(718,54.1306)	(11,11)
zone 2	(18,1042)	(1052,85.6832)	(24,55)
zone 3	(16,820)	(828,122.9500)	(22,102)

EARTHQUAKE OCCURRENCES

Posterior expectations (and standard deviations)

	$E[p\|\text{data}]$	$E[\lambda\|\text{data}]$	$E[\mu\|\text{data}]$
zone 1	.0069	13.2642	.5000
	(.0031)	(.4950)	(.1043)
zone 2	.0170	12.2778	.3038
	(.0040)	(.3785)	(.0514)
zone 3	.0191	6.7344	.1774
	(.0047)	(.2340)	(.0342)

95% Credible intervals

	$p\|\text{data}$	$\lambda\|\text{data}$	$\mu\|\text{data}$
zone 1	(.0022,.0141)	(12.3116,14.2518)	(.2978,.7022)
zone 2	(.0101,.0256)	(11.5470,13.0307)	(.2081,.4089)
zone 3	(.0110,.0295)	(6.2835,7.2008)	(.0976,.2128)

EARTHQUAKE OCCURRENCES

- Percentage of major earthquakes very small and not significantly different for the three zones, although relatively few major earthquakes occur in zone 1 (smallest p and $p\lambda$, besides a small number of shocks)
  - Posterior density of each p very concentrated around its mean ($\simeq .01$)
  - For all p's $\Rightarrow$ probability larger than 95% of being in the interval $(.0022, .0295)$
- Quite different occurrence rate in the three zones, as λ has very different means and credible intervals: highest rates in zones 1 and 2
- Few minor earthquakes in zone 3, but major earthquakes characterized by larger magnitudes (small μ) $\Rightarrow$ most disruptive earthquakes occur mainly in zone 3
- Zone 2 and 3 with similar number of major earthquakes, but longer interoccurrence times, smaller number of minor earthquakes and larger magnitudes in zone 3
  - Does greater elapsed times between shocks imply greater magnitudes?

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

- Motivated by problems of vibration in commercial aircrafts that cause fatigue in materials, Birnbaum and Saunders (1969) introduced a probability distribution (BS) that describes lifetimes of specimens exposed to fatigue due to cyclic stress

- BS model is based on a physical argument of cumulative damage that produces fatigue in materials, considering the number of cycles under stress needed to force a crack extension due to fatigue to grow beyond a threshold, provoking the failure of the material

- Density and cdf of BS distribution given by

$$f(x) = \frac{\sqrt{\frac{x}{\beta}} + \sqrt{\frac{\beta}{x}}}{2\alpha x}\phi\left(\frac{\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}}{\alpha}\right) \text{ and } F(x) = \Phi\left(\frac{\sqrt{\frac{x}{\beta}} - \sqrt{\frac{\beta}{x}}}{\alpha}\right).$$

  with $x > 0$, $\alpha, \beta > 0$, and ϕ and Φ standard Gaussian pdf and cdf

- Several extensions and generalizations of the BS distribution

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

- NHPP $N(t)$ number of cycles occurring in $[0, t]$
  - Intensity function $\lambda(t)$, with $\lambda(t) > 0, \forall t$
  - Mean value function $\Lambda(t)$ strictly increasing

- $\{\xi_k;\ k \in \mathbb{N}\}$ sequence of i.i.d. r.v.'s denoting crack extension due to kth stress cycle
  - Mean μ and variance σ^2

- $\Rightarrow$ compound Poisson $W_t = \sum_{k=1}^{N_t} \xi_k$: total crack extension produced during $[0, t]$

- Interest in $W_t > \omega^*$ since failure occurs for crack length exceeding a critical value ω^*

- $T = \inf\{t > 0\colon W_t > \omega^*\}$ lifetime until failure occurs

- $\Rightarrow$ Interest in deriving the distribution of T

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

Theorem

- $\{R_t; t \geq 0\}$ process defined as

$$R_t = \frac{\sum_{k=0}^{N_t} \xi_k - \mu \Lambda(t)}{[\{\mu^2 + \sigma^2\} \Lambda(t)]^{1/2}}$$

- $\Lambda(t) \underset{t \to \infty}{\longrightarrow} \infty$

- $\Rightarrow R_t \xrightarrow[t \to \infty]{\mathcal{D}} R$
  - $R \sim \mathrm{N}(0, 1)$
  - $\mathcal{D}$ convergence in distribution

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

- T and W_t related:

$$\{T \leq t\} = \{W_t \geq \omega^*\} = \left\{ \frac{\sum_{k=1}^{N_t} \xi_k - \mu\Lambda(t)}{\left[\{\mu^2 + \sigma^2\}\Lambda(t)\right]^{1/2}} \geq \frac{\omega^* - \mu\Lambda(t)}{\left[\{\mu^2 + \sigma^2\}\Lambda(t)\right]^{1/2}} \right\}$$

- Combined with Central Limit Theorem $\Rightarrow$ approximate, for t *large enough*

$$\mathbb{P}(W_t \geq \omega^*) = \mathbb{P}(T \leq t) \approx \Phi\left(\frac{\mu\sqrt{\Lambda(t)}}{[\mu^2 + \sigma^2]^{1/2}} - \frac{\omega^*}{[\mu^2 + \sigma^2]^{1/2}\sqrt{\Lambda(t)}} \right)$$

- As in Birnbaum and Saunders, approximation treated as exact

- $\Rightarrow$ Λ-BS distribution $\Lambda\text{-BS}(\alpha, \beta, \Lambda)$, with cdf

$$F_T(t) = \mathbb{P}(T \leq t) = \Phi\left(\left[\sqrt{\Lambda(t)/\beta_\Lambda} - \sqrt{\beta_\Lambda/\Lambda(t)}\right]/\alpha\right),\ t > 0,$$

with $\alpha = \sqrt{[\mu^2 + \sigma^2]/[\omega^*\mu]}$ and $\beta_\Lambda = \omega^*/\mu$

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

- Link with a Gaussian r.v. Z

$$Z = \frac{1}{\alpha}\left[\sqrt{\Lambda(T)/\beta_\Lambda} - \sqrt{\beta_\Lambda/\Lambda(T)}\right] \sim \mathrm{N}(0,1)$$

$$\Leftrightarrow T = \Lambda^{-1}\left(\beta_\Lambda\left[\alpha Z/2 + \sqrt{\{\alpha Z/2\}^2 + 1}\right]^2\right) \sim \Lambda - \mathrm{BS}(\alpha, \beta, \Lambda)$$

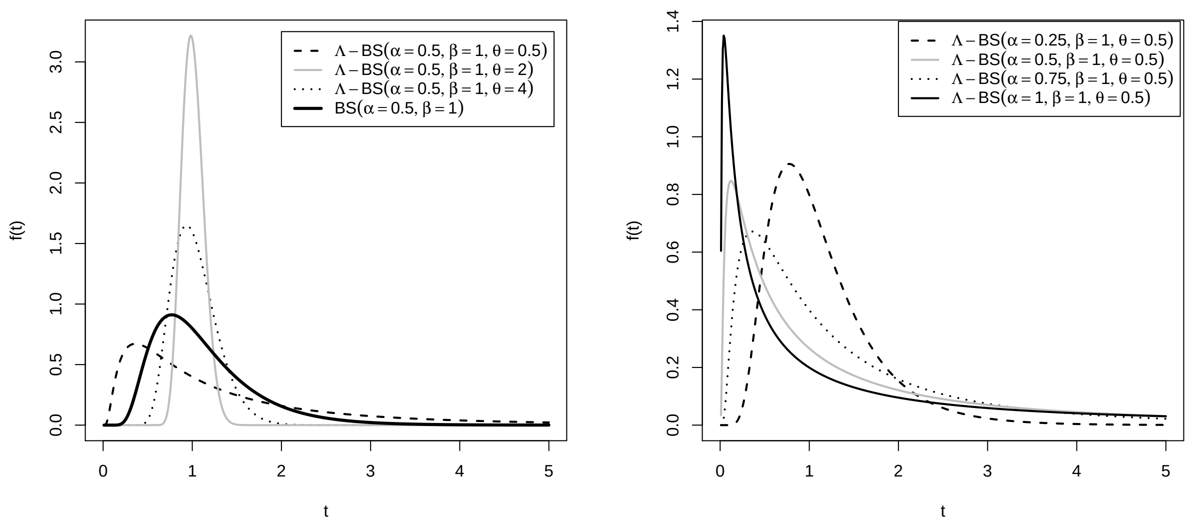
- Pdf of $T \sim \Lambda\text{-BS}(\alpha, \beta, \Lambda)$ given by $f_T(t) = \phi\left(a_t\right) A_t$, for $t > 0$, $\alpha > 0$ and $\beta > 0$, where

$$a_t = a_t(\alpha, \beta, \Lambda) = \frac{1}{\alpha}\left[\sqrt{\frac{\Lambda(t)}{\beta_\Lambda}} - \sqrt{\frac{\beta_\Lambda}{\Lambda(t)}}\right], \quad A_t = \frac{\mathrm{d}}{\mathrm{d}t} a_t = \frac{\Lambda'(t)\left[\Lambda(t) + \beta_\Lambda\right]}{2\alpha\sqrt{\beta_\Lambda}\Lambda(t)^{3/2}}$$

- BS distribution obtained for HPP with $\lambda = 1$

A NEW BIRNBAUM-SAUNDERS DISTRIBUTION

- PLP: $\lambda(t) = M\theta t^{\theta-1}$
- $\Rightarrow$ pdf $f_T(t) = \dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{1}{2\alpha^2}\left[\left\{\dfrac{t}{\beta}\right\}^{\theta} + \left\{\dfrac{\beta}{t}\right\}^{\theta} - 2\right]\right) \dfrac{\theta}{2\alpha\, t}\left[\left\{\dfrac{t}{\beta}\right\}^{\theta/2} + \left\{\dfrac{\beta}{t}\right\}^{\theta/2}\right]$
- Pdf does not depend on M
- Plots of Λ-BS densities for some α, β and θ



OTHER ISSUES

- Spatio-temporal models, especially spatial point processes, including Poisson ones, are getting more and more popular, mostly stemming from environmental and epidemiological problems (Banerjee et al, 2004)
- (Extended) gamma process conjugate prior on the intensity function for data coming from replicates of a Poisson process (Lo, 1982)
- Intensity function of a spatial NHPP modeled with a Bayesian nonparametric mixture (Kottas and Sansò, 2007)
- Under the Bayesian nonparametric approach, intensity function seen as a realization from a process $\Rightarrow$ data viewed as arising from a doubly stochastic Poisson process
- de Miranda and Morettin (2011) used wavelet expansions to model the intensity function in a classical framework $\Rightarrow$ possible Bayesian approach

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/~fabrizio/

WHAT IS RELIABILITY?

- Probability that a system operates correctly, under specified conditions, for a given time

- Quality over time (Condra, 1993)

- $P(T \geq t)$ (reliability function), with T failure time, nonnegative r.v.

SOME ISSUES IN RELIABILITY

- *System* performance
- Monetary costs
- Social costs
- Warranty (length, cost, forecast)
- Inventory of spare parts
- Maintenance and replacement policy
- Product testing
- Degradation up to failure
- Safety and security

REFERENCES ON RELIABILITY

Encyclopedia

- *Ruggeri, Kenett and Faltin*, Wiley

Probability

- *Barlow and Proschan*, SIAM

Statistics

- *Meeker and Escobar*, Wiley
- *Rigdon and Basu*, Wiley

Bayesian Statistics

- *Martz and Waller*, Wiley
- *Hamada, Wilson, Reese and Martz*, Springer

BASIC DEFINITIONS

- Failure time T, with pdf $f(t)$ and cdf $F(t)$, $t \geq 0$
- Reliability function: $S(t) = \mathbb{P}(T > t) = \int_t^\infty f(x)dx$
- Mean time to failure $MTTF = \int_0^\infty tf(t)dt = \int_0^\infty S(t)dt$
- Mean time between failures $MTBF$
- Hazard function (hazard rate, failure rate):

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}$$

  - $h(t)$ not a density but $h(t)\Delta t \approx \mathbb{P}(t \leq T < t + \Delta t | T \geq t)$
  - $h(t) = \frac{f(t)}{S(t)}$
  - $f(t) = h(t)S(t) = h(t)e^{-\int_0^t h(x)dx}$

REPAIRABLE SYSTEMS

Failure of a water pump in a car

- Water pump $\Rightarrow$ non-repairable system
- Car $\Rightarrow$ repairable system

Most common models for repairable systems:

- Renewal Process *("Good as new")*
  - sequence of i.i.d. r.v.'s denoting time between two failures
- Non-homogeneous Poisson Process (NHPP) *("Bad as old")*

Both models have drawbacks:

Repair $\Rightarrow$ reliability growth but not "as new"

Different models used in disjoint time intervals

RENEWAL PROCESS

- Sequence of failure times $X_0 = 0 \leq X_1 \leq X_2 \leq \cdots$
- Interfailure times $T_i = X_i - X_{i-1}$ for $i = 1, 2, \ldots$
- If $T_1, T_2, \ldots$ sequence of i.i.d. random variables
- $\Rightarrow \{T_i\}$ stochastic process called *renewal process*
- HPP renewal process since interfailure times are i.i.d. exponential random variables
- Two simple examples of Bayesian estimation via conjugate priors

RENEWAL PROCESS

- Number of cycles (daily beard cuts of the same individual) run by an electric razor before battery exhaustion and replacement
- $\Rightarrow$ Sequence of integer-valued random variables $N_1, \ldots, N_n$
- Assume N_i's, $i = 1, \ldots, n$, are i.i.d. Poisson Po(λ) random variables
- $\Rightarrow$ likelihood given by $l(\lambda|data) = \dfrac{\lambda^{\sum_{i=1}^n N_i}}{\prod_{i=1}^n N_i!} e^{-n\lambda}$
- Conjugate gamma Ga(α, β) prior for λ
- $\Rightarrow$ Posterior Ga$\big(\alpha + \sum_{i=1}^n N_i, \beta + n\big)$
- Posterior mean $\dfrac{\alpha + \sum_{i=1}^n N_i}{\beta + n}$

RENEWAL PROCESS

- Posterior predictive distribution of $(n+k)$'th failure time X_{n+k}, for any integer $k > 0$
  - Sum of k i.i.d. Po(λ) random variables $\Rightarrow$ Po$(k\lambda)$
  - for $m \geq \sum_{i=1}^{n} N_i$,

$$P(X_{n+k} = m | N_1, \ldots, N_n) = \int P(X_{n+k} = m|\lambda) f(\lambda | N_1, \ldots, N_n) d\lambda$$

$$= \int \frac{(k\lambda)^{m-\sum_{i=1}^{n} N_i}}{(m - \sum_{i=1}^{n} N_i)!} e^{-k\lambda} \cdot \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} N_i}}{\Gamma(\alpha + \sum_{i=1}^{n} N_i)} \lambda^{\alpha+\sum_{i=1}^{n} N_i - 1} e^{-(\beta+n)\lambda} d\lambda$$

$$= \frac{k^{m-\sum_{i=1}^{n} N_i}}{(m - \sum_{i=1}^{n} N_i)!} \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} N_i}}{(\beta+n+k)^{\alpha+m}} \frac{\Gamma(\alpha+m)}{\Gamma(\alpha + \sum_{i=1}^{n} N_i)}.$$

  - As a consequence, for $r = 0, 1, \ldots$,

$$P(N_{n+1} = r | N_1, \ldots, N_n) = \frac{1}{r!} \frac{(\beta+n)^{\alpha+\sum_{i=1}^{n} N_i}}{(\beta+n+1)^{\alpha+r+\sum_{i=1}^{n} N_i}} \frac{\Gamma(\alpha + r + \sum_{i=1}^{n} N_i)}{\Gamma(\alpha + \sum_{i=1}^{n} N_i)}$$

RENEWAL PROCESS

- Red traffic light, whose bulbs are substituted upon failure
- Replacement time negligible w.r.t. bulb lifetime
- All the bulbs have the same characteristics and operate under identical conditions
- $\Rightarrow$ assume interfailure times, $T_1, \ldots, T_n$, a sequence of i.i.d. exponential Ex(λ) random variables
- Likelihood $l(\lambda|data) = \lambda^n e^{-\lambda \sum_{i=1}^n T_i}$
- Conjugate gamma Ga(α, β) prior chosen for λ
- $\Rightarrow$ Posterior gamma Ga$\left(\alpha + n, \beta + \sum_{i=1}^n T_i\right)$
- Posterior mean $\dfrac{\alpha + n}{\beta + \sum_{i=1}^n T_i}$

RENEWAL PROCESS

- Posterior predictive density $f_{n+k}(x)$ of the $(n+k)$-th failure time $X_{n+k} = \sum_{i=1}^{n+k} T_i$, for any integer $k > 0$
  - Sum of k i.i.d. Ex(λ) random variables $\Rightarrow$ Ga(k, λ)
  - for $x > 0$,

$$f_{n+k}(x + \sum_{i=1}^{n} T_i | T_1, \ldots, T_n) = \int f_k(x|\lambda) f(\lambda|T_1, \ldots, T_n) d\lambda$$

$$\begin{aligned} &= \int \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x} \cdot \frac{(\beta + \sum_{i=1}^{n} T_i)^{\alpha+n}}{\Gamma(\alpha+n)} \lambda^{\alpha+n-1} e^{-(\beta+\sum_{i=1}^{n} T_i)\lambda} d\lambda \\ &= x^{k-1} \frac{\Gamma(\alpha+n+k)}{\Gamma(k)\Gamma(\alpha+n)} \frac{(\beta + \sum_{i=1}^{n} T_i)^{\alpha+n}}{(\beta + x + \sum_{i=1}^{n} T_i)^{\alpha+n+k}}. \end{aligned}$$

  - One-step-ahead predictive distribution is given by

$$f_{n+1}(x|T_1, \ldots, T_n) = (\alpha+n) \frac{(\beta + \sum_{i=1}^{n} T_i)^{\alpha+n}}{(\beta + x + \sum_{i=1}^{n} T_i)^{\alpha+n+1}}$$

FEATURES OF A NHPP

- NHPP used to model reliability growth/decay
- NHPP good for
  - prototype testing
  - repair of small components in complex systems
- Repair strategies in a NHPP:
  - instantaneous
  - minimal repair ($\Rightarrow$ back to previous reliability)

Repairs could worsen the reliability

NHPP

- Issues already presented earlier:
  - Definition, properties and theorems
  - Choice of intensity function
  - Elicitation of priors
  - Estimation and forecasting
- Other issues to be presented in case studies
- Here:
  - Further on prior elicitation
  - Reliability measures

EXPERT ELICITATION

Campodonico and Singpurwalla (1995)

- Analyst A chooses process, i.e. $\lambda(t)$
- A asks Expert E questions on $\mu_1 = E[N(T_1)]$ and $\mu_2 = E[N(T_2)]$, $T_1 < T_2$
- Expert E
  - does not know $\lambda(t)$
  - chooses a functional form for the μ_i's
  - visualizes his/her own cdf's for μ_i with location and scale m_i and s_i, $i = 1, 2$
  - E gives m_i and s_i, $i = 1, 2$ to A

EXPERT ELICITATION

- A reparameterizes the model, e.g. PLP
  - $\mu_i = MT_i^\beta, i = 1, 2$
    - $\Rightarrow \beta = \dfrac{\log \mu_2/\mu_1}{\log T_2/T_1}$
    - $\Rightarrow M = \dfrac{\log T_2 \log \mu_1 - \log T_1 \log \mu_2}{\log T_2 - \log T_1}$
  - $\Rightarrow$ PLP with parameters μ_1, μ_2

- A builds the prior on $(\mu_1, \mu_2 | m_1, s_1, m_2, s_2)$
  - *likelihood* $f(m_1, s_1, m_2, s_2 | \mu_1, \mu_2)$ based on his/her opinion on E and assumptions (e.g. independence and truncated Gaussian distributions) on the quantities
  - *flat prior* for values of (μ_1, μ_2) over the support of the *likelihood*

EXPERT ELICITATION

Betrò and Guglielmi (1996)

- PLP: $\lambda(t; M, \beta) = M\beta t^{\beta-1}$
- T_1 first failure time: $P[T_1 > s|M, \beta] = e^{-Ms^\beta}$
  - Independent priors on M and β
  - Gamma prior $M \sim$ Ga(a, b)
  - Expert asked about lower and upper bounds, l_i and u_i, on $P[T_1 > s_i]$, $i = 1, \ldots, n$
  - $\pi(\beta) \in \Gamma$, generalized moments class:

$$l_i \leq P[T_1 > s_i] = \int_0^\infty H_i(\beta)\pi(\beta)d\beta \leq u_i,$$

with $H_i = \left(\frac{b}{s_i{}^\beta + b}\right)^a$, $i = 1, \ldots, n$

- $E[N(T)|M, \beta] = MT^\beta$, similarly

RELIABILITY MEASURES

- System reliability (for a PLP)
  - Data on the same system (observed up to y):

    $$R((y,s]) = P\left(N(y,s) = 0 | M, \beta\right) = e^{-M\left(s^\beta - y^\beta\right)}$$

  - Data on equivalent system:

    $$R(s) = P\left(N(s) = 0 | M, \beta\right) = e^{-Ms^\beta}$$

- Expected number of failures in future intervals
  - Same system: $E[N(y,s]|M,\beta] = M\left(s^\beta - y^\beta\right)$
  - Equivalent system: $E[N(s)|M,\beta] = Ms^\beta$

- Intensity function at y:

  Reliability growth models without further improvements $\Rightarrow$ constant intensity $\lambda(y)$

RELIABILITY MEASURES

- Posterior on (M, β) given data $\mathbf{T}$
- Estimates:
  - $R\widehat{((y, s]})= E[R((y, s]|M, \beta)] = \int e^{-M(s^{\beta}-y^{\beta})} f(M, \beta|\mathbf{T}) dM d\beta$
  - $E[\widehat{N(y, s}]] = E[E[N(y, s]|M, \beta]] = \int M(s^{\beta} - y^{\beta}) f(M, \beta|\mathbf{T}) dM d\beta$
- If MCMC sample available from posterior
  - Monte Carlo computation of the integrals

FAILURE COUNT DATA

- Experiments introduced so far for NHPPs assume observation of failure times
- Here we assume known only the number of failures in an interval
- Failures of vehicle from brand ABC
  - Under warranty, cars taken to ABC at any failure

    $\Rightarrow$ (Random) failure times t_i observed for each system (individual vehicle)
  - After warranty expiration, cars taken at any mechanics and ABC contacts, every once in a while, k buyers about their car failures

    $\Rightarrow$ (Random) number n_i of failures in a (random) period t_i for k systems
- $\Rightarrow k$ independent PLPs with the same $\lambda(t)$ (assuming systems are operated under same conditions)

FAILURE COUNT DATA

Calabria, Guida and Pulcini (1994)

- k systems operated under same conditions
- $\Rightarrow k$ independent PLPs with the same $\lambda(t)$
- Only failure counts in interval available, not actual failure times
- Maximum likelihood estimation of β, α, $\lambda(t)$ and $M(t)$ (none in closed form)
- Study (through Monte Carlo) of "relative bias", i.e. ratio of bias of MLEs to true values
- Bootstrap confidence interval (based on small samples)

FAILURE COUNT DATA

Calabria, Guida and Pulcini (1994)

- Time truncated experiment about 10 systems
- Data generated from PLP with $\lambda(t) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1}$
  - $\alpha = 100$ (hours) and $\beta = 1.5$
- Censoring times T_i drawn from U$(0, 1200)$
- Censoring times T_i and number of failures n_i

System	1	2	3	4	5	6	7	8	9	10
T_i	1042	932	997	1087	900	849	764	202	141	479
n_i	36	26	29	28	32	29	22	6	2	7

- Consider PLP with $M = (1/\alpha)^{\beta}$ parameter $\Rightarrow M = 0.001$

FAILURE COUNT DATA

- Likelihood obtained as product of densities

$$f(n_i, T_i|\beta, M) = f(n_i|T_i, \beta, M) f(T_i|\beta, M), \text{ for } i = 1, \ldots, k,$$

  - Censoring times independent of the failure process (and β and M)
  - $\Rightarrow$ only $f(n_i|T_i, \beta, M)$ considered

- $\Rightarrow$ Likelihood

$$l(\beta, M \mid \mathbf{d}) \propto p^\beta M^s \exp\left[-M \sum_{i=1}^{k} T_i^\beta\right], \text{ with}$$

  - $s = \sum_{i=1}^{k} n_i$
  - $p = \prod_{i=1}^{k} T_i^{n_i}$
  - $\mathbf{T} = (T_1, \ldots, T_k)$
  - $\mathbf{n} = (n_1, \ldots, n_k)$
  - $\mathbf{d} = (\mathbf{T}, \mathbf{n})$

FAILURE COUNT DATA

- Joint prior $f(\beta, M) = f(M|\beta) f(\beta)$
  - $M|\beta \sim \text{Ga}\big(\rho, \sigma^\beta\big)$
  - $\beta \sim \text{Ga}(\nu, \mu)$
  - Mazzali (1996) considered independent priors ($M \sim \text{Ga}(\rho, \sigma)$) as well, with no significant difference in posterior inference

- Joint posterior

$$f(M, \beta \mid \mathbf{d}) \propto (p\sigma^\rho)^\beta M^{s+\rho-1} \beta^{\nu-1} \exp(-\mu\beta) \exp[-M(\sigma^\beta + \sum_{i=1}^{k} T_i^\beta)]$$

- Posterior conditional distributions

$$M \mid \beta, \mathbf{d} \sim \text{Ga}\Big(s + \rho, \sigma^\beta + \textstyle\sum_{i=1}^k T_i^\beta\Big)$$

$$\beta \mid M, \mathbf{d} \propto (p\sigma^\rho)^\beta \beta^{\nu-1} e^{-\mu\beta - M\sigma^\beta}$$

- Marginal posteriors easily sampled via a Metropolis within Gibbs algorithm

FAILURE COUNT DATA

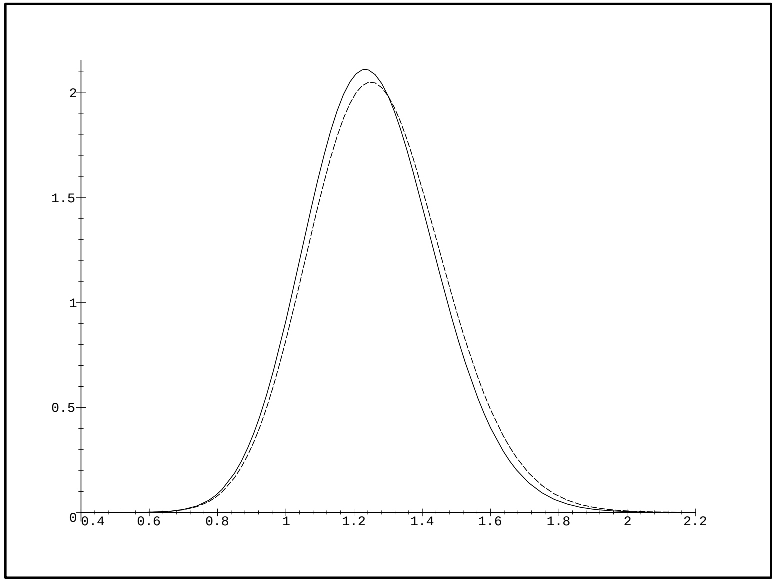
- $M|\beta \sim$ Ga$\big(15, 0.55^{\beta}\big)$
- $\beta \sim$ Ga$(2, 1)$
- $\tilde{\beta} = 1.274$ posterior mean of β, very close to 1.275 (MLE)
- Unlike the frequentist approach, the Bayesian one allows for a direct, straightforward assessment on β and the system reliability
- Posterior cdf of β for selected δ's

δ	0.8	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.8
$P(\beta \leq \delta \mid \mathbf{d})$	.005	.09	0.21	0.40	0.60	0.78	0.89	0.95	0.99

- Systems very likely under reliability decay
  - β has large probability of being greater than 1
  - 90% highest posterior density interval given by $(0.940, 1.564)$

FAILURE COUNT DATA

- $\mu = 1$, $\nu = 2$, $\sigma = 1.1$ and $\rho = 30$
- Posterior of β under
  - dependent prior: dotted line
  - independent prior: solid line
- No significant difference



FAILURE COUNT DATA

Estimation of mean value function $m(t)$

- $m(t|M,\beta) = Mt^\beta$
- $E[m(t) \mid \mathbf{d}] = E[t^\beta E[M|\beta,\mathbf{d}]] = E[t^\beta (s+\rho)/(\sigma^\beta + \sum T_i^\beta)]$
- Latter expected value taken w.r.t. posterior of β
  - Integrating joint posterior w.r.t. M
  - $\Rightarrow$ posterior of β

$$f(\beta \mid \mathbf{d}) \propto (p\sigma^\rho)^\beta \frac{\Gamma(s+\rho)}{(\sigma^\beta + \sum_{i=1}^k T_i^\beta)^{s+\rho}} \beta^{\nu-1} e^{-\mu\beta}.$$

- $\Rightarrow$ Posterior mean of $m(t)$

$$E[m(t) \mid \mathbf{d}] = \frac{\int_0^{+\infty} (tp\sigma^\rho)^\beta \beta^{\nu-1} \exp(-\mu\beta)(s+\rho)(\sigma^\beta + \sum T_i^\beta)^{-s-\rho-1}\, d\beta}{\int_0^{+\infty} (p\sigma^\rho)^\beta \beta^{\nu-1} \exp(-\mu\beta)(\sigma^\beta + \sum T_i^\beta)^{-s-\rho}\, d\beta}$$

FAILURE COUNT DATA

- Estimate of intensity function $\lambda(t)$

$$E[\lambda(t)|\mathbf{d}] = \frac{1}{t}\frac{\int_0^{+\infty}(tp\sigma^\rho)^\beta\beta^\nu \exp(-\mu\beta)(s+\rho)(\sigma^\beta+\sum T_i^\beta)^{-s-\rho-1}\,d\beta}{\int_0^{+\infty}(p\sigma^\rho)^\beta\beta^{\nu-1}\exp(-\mu\beta)(\sigma^\beta+\sum T_i^\beta)^{-s-\rho}\,d\beta}$$

- Interest, as a performance measure, in the distribution of $N(0,t]$, i.e. the number of failures in $(0,t]$
- Conditional on M and β

$$P(N(0,t]=r|\beta,M) = \frac{1}{r!}\left(Mt^\beta\right)^r \exp\left(-Mt^\beta\right)$$

- Posterior (unconditional) predictive distribution

$$\begin{aligned} P(N(0,t]=r|\mathbf{d}) &= \int_0^{+\infty}\int_0^{+\infty} P(N(0,t]=r|\beta,M) f(\beta,M\mid\mathbf{d})\,d\beta\,dM \\ &= \frac{1}{r!}\frac{\int_0^{+\infty}(t^r p\sigma^\rho)^\beta\beta^{\nu-1}\exp(-\mu\beta)\frac{\Gamma(s+\rho+r)}{\left(t^\beta+\sigma^\beta+\sum T_i^\beta\right)^{s+\rho+r}}\,d\beta}{\int_0^{+\infty}(p\sigma^\rho)^\beta\beta^{\nu-1}\exp(-\mu\beta)\frac{\Gamma(s+\rho)}{\left(\sigma^\beta\sum T_i^\beta\right)^{s+\rho}}\,d\beta} \end{aligned}$$

FAILURE COUNT DATA

- Interest in predicting the behavior of another, $k+1$-th, identical system, given data on k identical systems

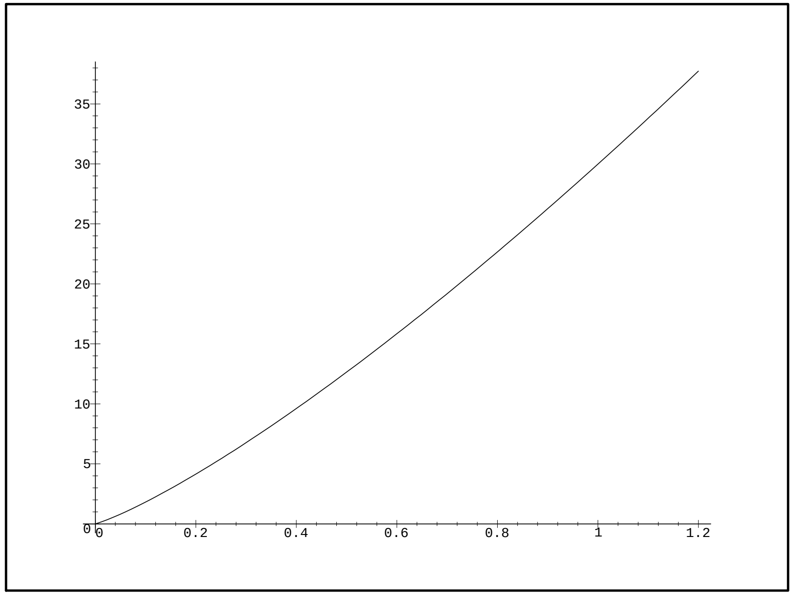
- $\Rightarrow$ Interest for a given t or $0 < t \leq T$
  - Expected number of failures $N(0,t]$
  - Intensity function $\lambda(t)$
  - System reliability $R(t) = P(N(0,t])$

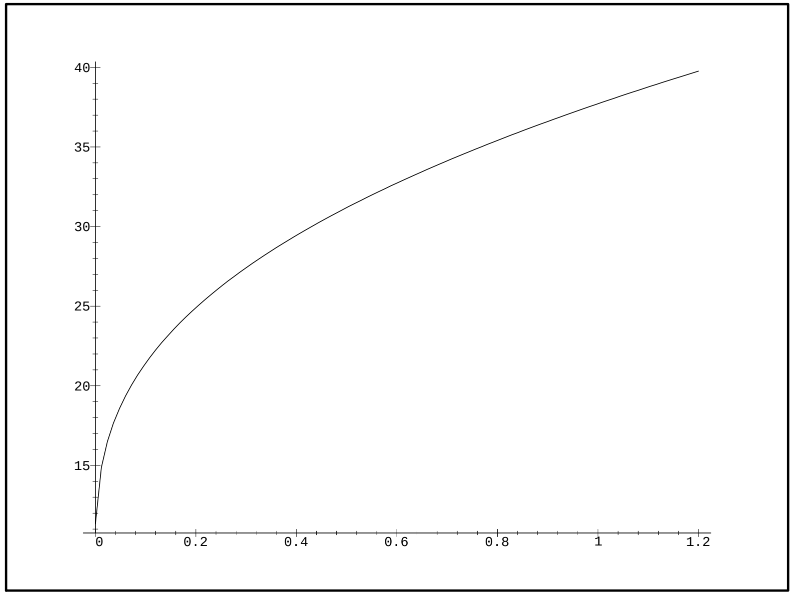
- In our example, interest in expected number of failures up to time $t = 1200$
  - Posterior mean $\widehat{m}(1200) = 38.81$
  - Compare with MLE at 38.70

- $m(1) = M \Rightarrow$ posterior estimate $\widehat{M} = 0.009$

FAILURE COUNT DATA

Plots of $E[m(t) \mid \mathbf{d}]$ (left) and $E[\lambda(t) \mid \mathbf{d}]$ (right) for $t \in [0, 1.2]$ (in thousands of hours)

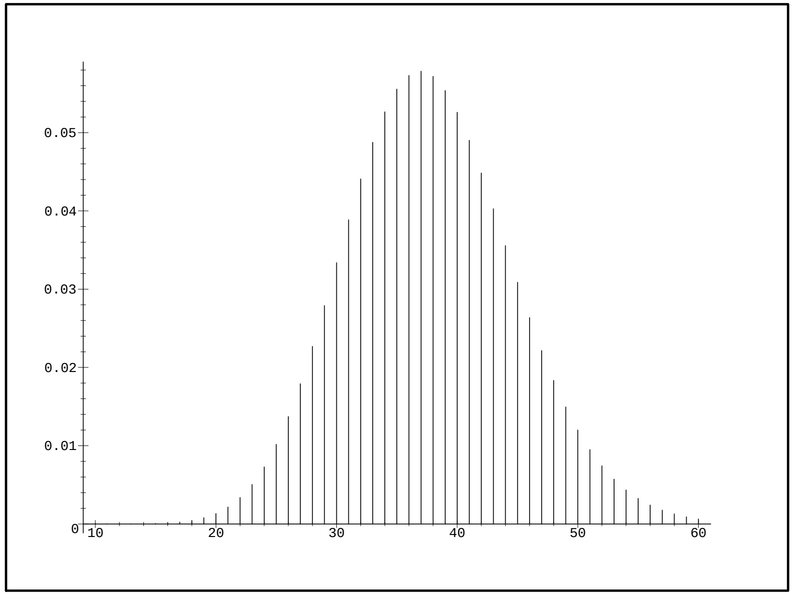


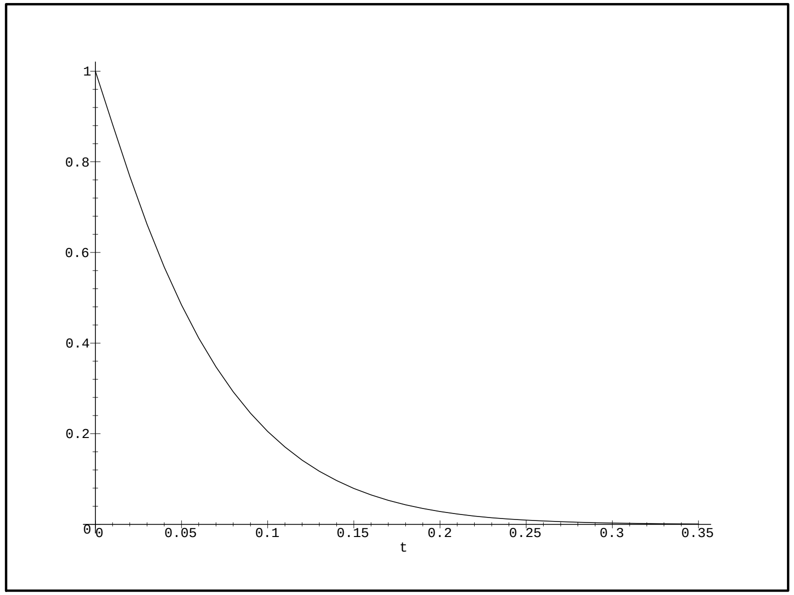


- Estimate of $m(t)$ denotes a quick increase at early stages, getting steadier later
- Estimate of $\lambda(t)$ denotes reliability decay, with a very steep increase in the failure rate right after the initial time (seen also in $m(t)$)

FAILURE COUNT DATA

Posterior predictive distribution of $N(0,1200]$ (left) and $P(N(0,t]=0)$ as a function of t (right)





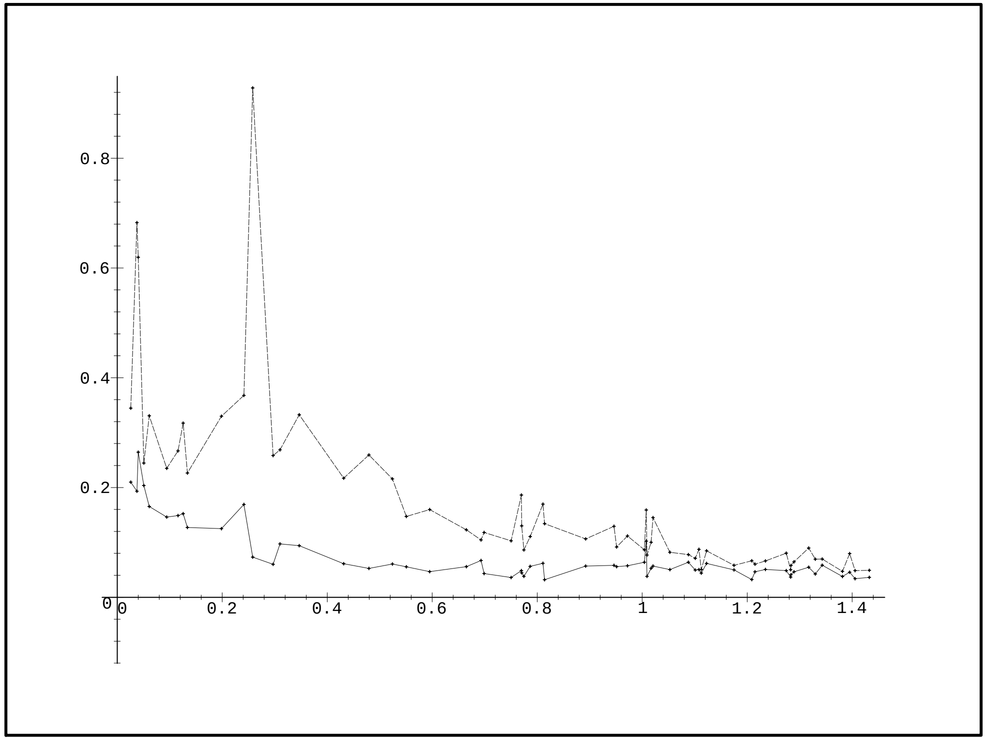
FAILURE COUNT DATA

Simulation study to compare MLEs and posterior means

- 60 β's drawn from U$(0, 1.5)$
- For each β, repeat 100 times:
  - 10 systems
  - Censoring times T_i from U$(0, 1.2)$ (in thousands of hours)
  - Data generated from PLP ($\alpha = 0.1, \beta$)
  - Find MLE $\widehat{\beta}$
  - Find $\mathcal{E}[\beta|\underline{T}]$ for $\mu = 1$, $\nu = 2$, $\sigma = 1.1$ and $\rho = 30$
- Compute $ASE = \sum_{i=1}^{100} (\beta - \beta^*)^2/100$ with $\beta^* =$ MLE or Bayes
- Compare Averaged Squared Errors for each β's

FAILURE COUNT DATA

Averaged Squared Errors for each β's



- Dotted line: MLE (*the one above*)
- Solid line: Posterior mean (*the one below*)

FAILURE COUNT DATA

Sensitivity analysis

- $M|\beta \sim \text{Ga}\big(15, .55^{\beta}\big)$
- $\beta \sim \text{Ga}(\nu, \mu)$, $.5 \leq \mu, \nu \leq 2 \Rightarrow .25 \leq \mathcal{E}\beta \leq 4$
  - Parametric class of priors
- Posterior ranges and MLEs

	MLE	Posterior expectation
β	1.275	$[1.198, 1.292]$
$\lambda(1200)$	41.13	$[38.11, 42.47]$
$m(1200)$	38.70	$[37.77, 39.05]$

- $\Rightarrow$ Strong support to (weak) reliability decay

CASE STUDY: GAS ESCAPES

- Company responsible for a large metropolitan gas distribution network developed in the last century
- Distribution network characterized by very non-homogeneous technical and environmental features (material, diameter of pipes, laying location, etc.)
- Distribution network consists of several thousand kilometers of pipelines providing gas at low, medium and high pressure
- Most of the network is at low-pressure (20 mbar over atmospheric pressure) and attention will be concentrated on it

GAS ESCAPES: RISKS AND COSTS

- Possible explosions: casualties and destructions

- Cost of installation of pipelines

- Cost of maintenance and replacement of riskier pipelines

- Labor cost of emergency and inspection squads

GAS ESCAPES: ATTENUATION OF RISKS

- Gas smells by law (introduction of smelling chemicals) to favor gas escape detections
- Pipelines installation and pipes in house according to the law
- Material chosen according to characteristics of installation area (traffic, ground moisture, residential area or not, etc.)
- Efficient calling center to report possible escapes
- Sufficient size and training of emergency and inspection squads
- Identification of risk factors (kind of junction, material, laying conditions, etc.)
  $\Rightarrow$ *Statistics and Decision Analysis*

GAS ESCAPES: REPLACEMENT POLICY

- Setting up an efficient replacement policy in an urban gas distribution network
- i.e. change of a certain kind of pipelines with a safer one
- assessment of the propensity of failure of the different kinds of pipelines
- assessment of the **probability** of failure of the different kinds of pipelines
- change of pipelines with highest propensity/probability of failure

GAS ESCAPES: DEMERIT-POINT-CARDS

- Propensity to failure determined by demerit-point-cards in many companies (see, e.g., technical report by British Gas Corporation)

- Influence of various quantitative and qualitative factors (diameter, laying depth, etc.) is quantified by assigning a score to each of them (e.g. if laying depth is between 0.9 and 1.5 m, the score is 20)

- Positive aspects
  - easy to specify
  - highlighting the critical factors strongly correlated with pipeline failures

GAS ESCAPES: DEMERIT-POINT-CARDS

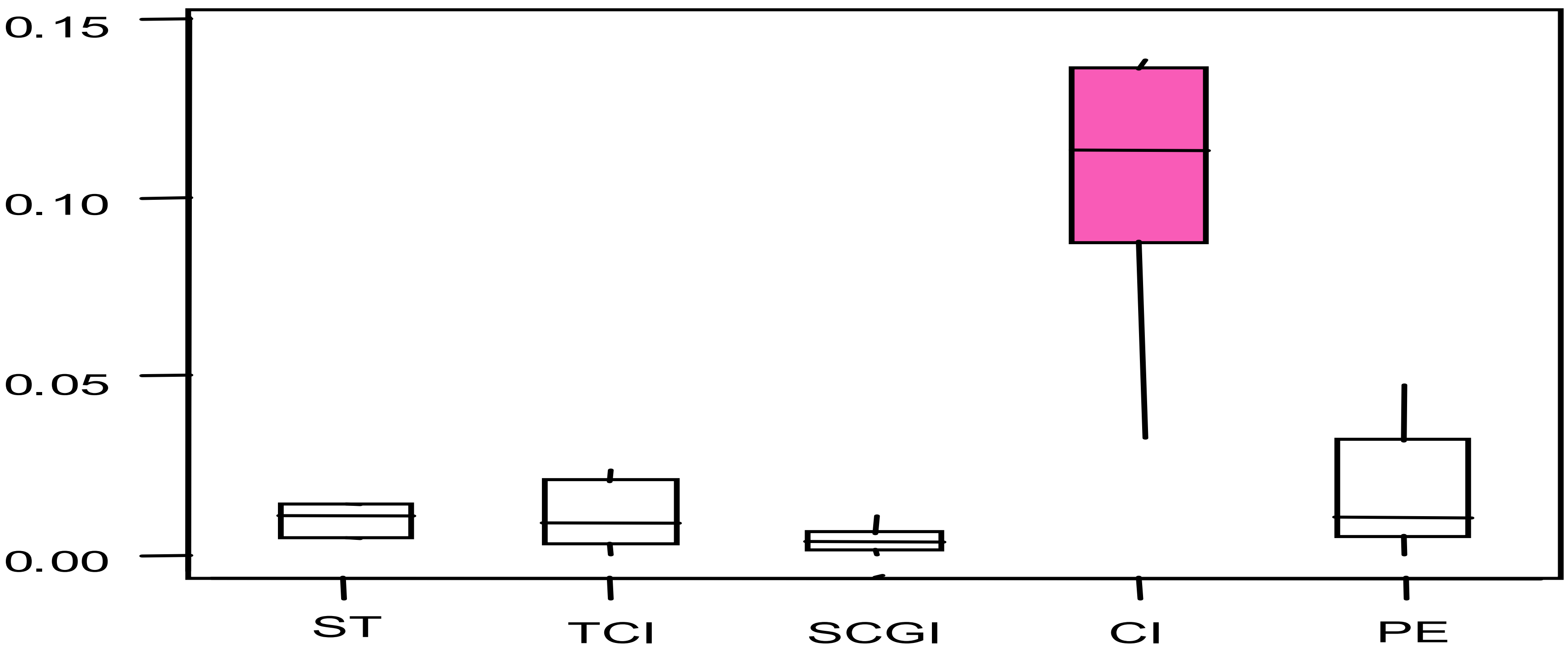
- Negative aspects
  - Setting of the scores, in our context, for the considered factors (and the choice of the factors) would be based only on previous empirical experience in other cities without any adjustment for the current context
  - An aggregate demerit score, obtained by adding the individual scores of the different factors, often hides possible interactions between the considered factors (e.g. diameter and laying depth), so losing other important information and worsening an already critical situation
  - Propensity-to-failure score of a given section of gas pipeline is considered independently of the length of the section and the planning period
  - *Stochastic nature of the phenomenon is ignored*

GAS ESCAPES: MATERIALS

Several materials for low-pressure pipelines

- traditional cast iron (CI)
- treated cast iron (TCI)
- spheroidal graphite cast iron (SGCI)
- steel (ST)
- polyethylene (PE)

GAS ESCAPES: YEARLY FAILURE RATES



GAS ESCAPES: INTEREST ON CI

- higher failure rate than other materials (even by one order of magnitude)

- covers more than a quarter of the whole network

- about 6000 different pipe sections with homogeneous characteristics, ranging in length from 3 to 250 meters for a total of 312 kilometers

GAS ESCAPES: FACTORS

Three groups of factors identified based upon

- studies in other companies
- reports in literature
- discussions with company's experts

GAS ESCAPES: FACTORS

- Intrinsic features of the pipeline-section
  - thickness
  - diameter
  - age
- factors concerning the laying of the pipeline-section
  - depth
  - location
  - ground characteristics
  - type and state of the pavement
  - laying techniques
- environmental parameters of the pipeline-section
  - traffic characteristics
  - intensity of underground services
  - external temperature and moisture

GAS ESCAPES: FACTORS

Reports on those factors rarely available and useful

- companies hardly disclose data on failures/escapes
- companies are in general responsible for a single city network, making very difficult any comparison between different situations and data re-utilisation/sharing
- data registration methods change over the long period of operation of the pipelines
- data are poorly registered or recorded for other purposes which are not sufficiently coherent with the requirements of a correct safety and reliability analysis

GAS ESCAPES: MANOVA ON FACTORS

- Multivariate analysis of variance (MANOVA) to consider mean differences on two or more dependent variables (i.e. failure factors) simultaneously

- CI distribution network (and the corresponding number of failures) over a significantly long period (10 years) was divided into different classes based on the previous factors

- For each factor, two levels were identified, where the notations ”high” and ”low” were qualitative rather than quantitative

GAS ESCAPES: MANOVA ON FACTORS

- The study identified diameter, laying depth and location as the most significant factors

- The other factors given above were not particularly important, as they turned out

  - either to be homogeneous for the analyzed distribution network (e.g. the soil used during installation works always had the same chemical and mechanical characteristics, while external temperature and moisture actually do not have a different effect on different pipeline sections, as the ground has a strong insulating capacity, even with shallow laying)

  - or to be strongly correlated to the above factors (e.g. the thickness of a pipe is fixed for a given diameter).

- Those considerations were fully shared and validated by company experts

GAS ESCAPES: FIRST FINDINGS

Following this preliminary data analysis, the most important conclusions were

- This industrial sector is characterized by a remarkable *shortage of data* because pipeline failures are rare and available information is often inadequate. This scarcity of data, together with some underlying ”noise” stemming from imprecise recording of data, suggested that information obtained from the company archives should be improved with expert judgements using a *Bayesian approach*

GAS ESCAPES: FIRST FINDINGS

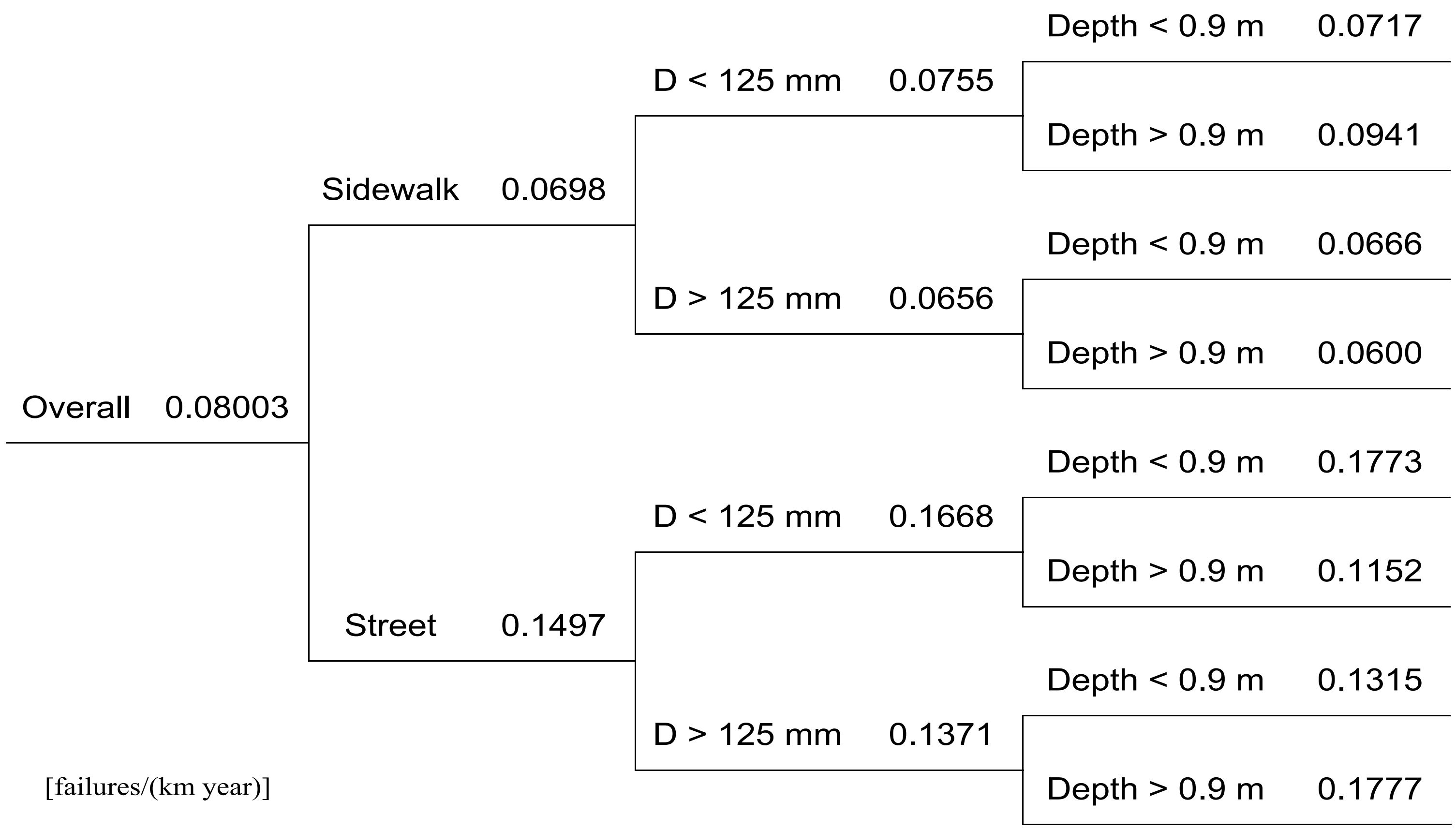
- CI pipeline failure rate seems to be *scarcely sensitive to wear* or proximity to a previous failure (or leak) in the same pipeline section, but it is mostly influenced by accidental stress, even if the useful life phase may be considered longer than 50 years. So propensity-to-failure in a unit time period or unit length does not vary significantly with time and space. Since failures are rare events (an assumption confirmed by available data), it was felt appropriate to model them with an *homogeneous (in time and space) Poisson process*

- The evident importance of interactions between factors (e.g. diameter is significant with shallow but not with deep laying) led to the abandonment of the additive approach, typical of the demerit-point-cards, which sums the effects of the factors, in favor of the determination of pipeline classes derived from the combination of the levels of the most significant factors. This proposed data organization also facilitates the *expression of experts judgements*

FAILURES IN CAST-IRON PIPES

- Some materials (e.g. steel) subject to aging $\Rightarrow$ NHPP
- Cast-iron is not aging $\Rightarrow$ HPP in space and time
- HPP with parameter λ (unit failure rate in time and space)
- n failures in $[0,T] \times \mathcal{S}, \Rightarrow L(\lambda|n,T,\mathcal{S}) = (\lambda s T)^n e^{-\lambda s T}$, with $s = meas(\mathcal{S})$
- Data: $n = 150$ failures in $T = 6$ years on a net $\approx s = 312$ Km long $\Rightarrow L(\lambda|n,T,\mathcal{S}) = (1872\lambda)^{150} e^{-1872\lambda}$ (if considering all failures together)
- MLE $\widehat{\lambda} = n/(sT) = 150/1872 = 0.080$
- Consider the 8 classes determined by the two levels of the relevant covariates: diameter, location and depth

FAILURES IN CAST-IRON PIPE



CHOICE OF THE PRIOR

HPP with parameter λ and n failures in $[0, T] \times \mathcal{S} \Rightarrow L(\lambda|n, T, \mathcal{S}) = (\lambda s T)^n e^{-\lambda s T}$

- $\lambda \sim \mathcal{G}(\alpha, \beta) \Rightarrow \lambda|n, T, \mathcal{S} \sim \mathcal{G}(\alpha + n, \beta + sT)$ (Conjugate prior)
- $$\begin{aligned} \pi(\lambda|n, T, \mathcal{S}) &\propto L(\lambda|n, T, \mathcal{S})\pi(\lambda) \\ &\propto \lambda^n e^{-\lambda s T} \cdot \lambda^{\alpha-1} e^{-\beta\lambda} \end{aligned}$$
- α and β chosen to match, e.g.
  - mean α/β and variance α/β^2 (or mode)
  - *Ideal* experiment with $\alpha =$ number of failures in the observation time β
  - two quantiles, with a third given for consistency

ELICITATION OF EXPERTS' OPINIONS

- The expert judgements were collected by an ad hoc questionnaire and integrated with historical data by means of Bayesian inference

- Experts from three areas within the company were selected to be interviewed:
  - pipeline design: responsible for designing the network structure (4 experts)
  - emergency squad: responsible for locating network failures (8 experts)
  - operations: responsible for the repair of broken pipelines (14 experts)

ELICITATION OF EXPERTS' OPINIONS

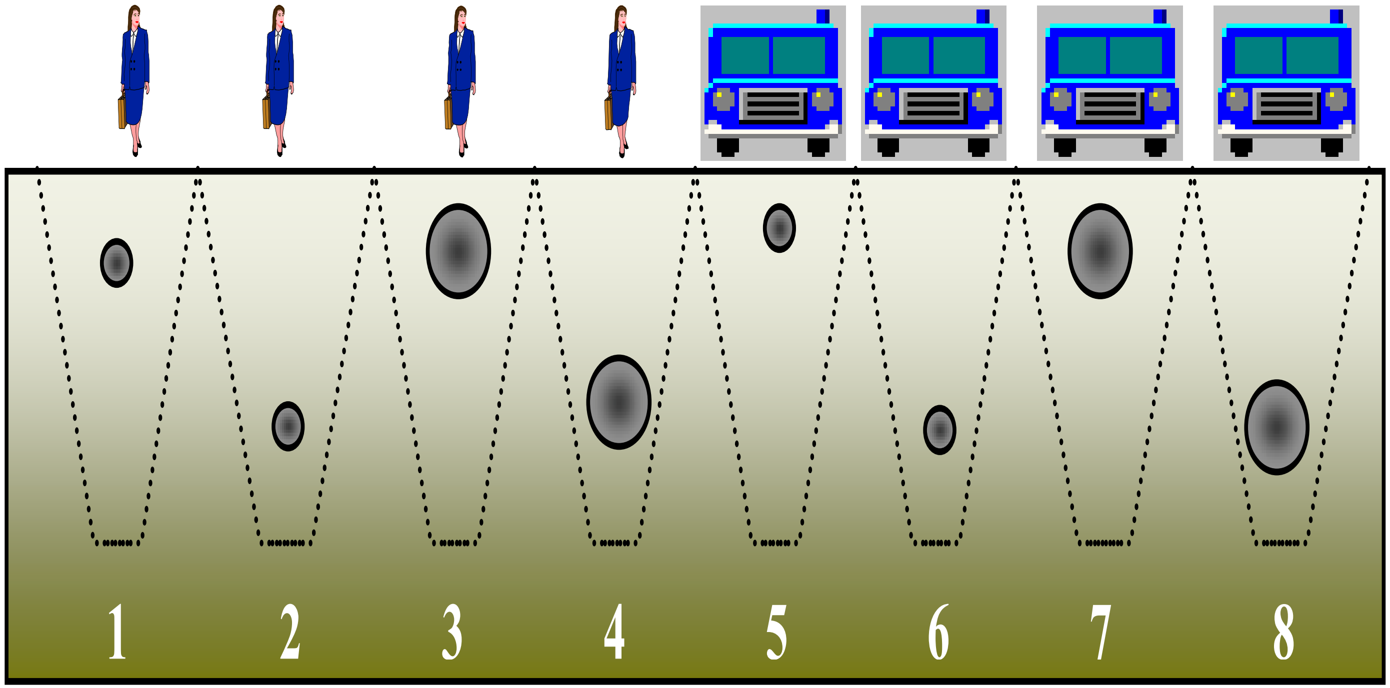
- Interviewees were actually not able to say how many failures they expected to see on a kilometer of a given kind of pipe in a year (the situation became even more untenable when they were asked to express the corresponding standard deviation or upper and lower bounds)

- The experts had great difficulty in saying how and how much a factor influenced the failure and expressing opinions directly on the model parameters while they were able to compare the performance against failure of different pipeline classes

- To obtain such a propensity-to-failure index, each expert was asked to compare the pipeline classes pairwise. In a pairwise comparison the judgement is the expression of the relation between two elements that is given, for greater simplicity, in a linguistic shape

- The linguistic judgement scale is referred to a numerical scale (Saaty's proposal: Analytic Hierarchy Process) and the numerical judgements can be reported in a single matrix of pairwise comparisons

ELICITATION OF EXPERTS' OPINIONS



ANALYTIC HIERARCHY PROCESS

Two alternatives A and B

B	"equally likely as"	A $\to$ 1
B	"a little more likely than"	A $\to$ 3
B	"much more likely than"	A $\to$ 5
B	"clearly more likely than"	A $\to$ 7
B	"definitely more likely than"	A $\to$ 9

Pairwise comparison for alternatives $A_1, \ldots, A_n$

$\Rightarrow$ square matrix of size n

$\Rightarrow$ eigenvector associated with the largest eigenvalue

$\Rightarrow (P(A_1), \ldots, P(A_n))$

ANALYTIC HIERARCHY PROCESS

An expert's opinion on propensity to failure of cast-iron pipes

Class	1	2	3	4	5	6	7	8
1	1	3	3	3	1/6	1	1/6	3
2	1/3	1	1/4	2	1/6	1/2	1/5	1
3	1/3	4	1	1	1/4	1	1/6	2
4	1/3	1/2	1	1	1/5	1	1/5	1
5	6	6	4	5	1	4	4	5
6	1	2	1	1	1/4	1	1/6	1
7	6	5	6	5	1/4	6	1	4
8	1/3	1	1/2	1	1/5	1	1/4	1

MATHEMATICS OF AHP

- $A = \{a_{ij}\}$ matrix from pairwise comparisons in AHP

- A strongly consistent if $a_{ij} = a_{ik}a_{kj}$, for all i, j, k
  $\Rightarrow A$ represented by normalized weights $(w_1, \ldots, w_n)$ s.t.

$$A = \begin{pmatrix} w_1/w_1 & w_1/w_2 & w_1/w_3 & \ldots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & w_2/w_3 & \ldots & w_2/w_n \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \ldots & w_3/w_n \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ w_n/w_1 & w_n/w_2 & w_n/w_3 & \ldots & w_n/w_n \end{pmatrix}$$

  $\Rightarrow a_{ij} = w_i/w_j = (w_i/w_k) \cdot (w_k/w_j) = a_{ik}a_{kj}$, for all i, j, k

- Unfortunately, human judgements are not in general consistent

- $\Rightarrow$ Need to find a consistent matrix and a measure of inconsistency

MATHEMATICS OF AHP

- A consistent $\Rightarrow$ Find weights w_i's as solution of

$$\begin{bmatrix} w_1/w_1 & w_1/w_2 & w_1/w_3 & \dots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & w_2/w_3 & \dots & w_2/w_n \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \dots & w_3/w_n \\ \dots & \dots & \dots & \dots & \dots \\ w_n/w_1 & w_n/w_2 & w_n/w_3 & \dots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \dots \\ w_n \end{bmatrix}$$

- $A\mathbf{w} = n\mathbf{w}$ or $(A - nI)\mathbf{w} = \mathbf{O}$ system of homogeneous linear equations, with nontrivial solution iff $det(A - nI) = 0 \Rightarrow n$ eigenvalue of A, unique since
  - $\{$number of nonnull eigenvalues = rank of A $= 1\}$, since each row is a linear combination of the others
  - sum of eigenvalues equals the trace of the matrix, i.e. sum of its diagonal elements, and here $tr(A) = n$
- The eigenvector $\mathbf{w}$ has positive entries and is unique up to a constant $\Rightarrow$ normalized dividing entries by their sum
- A consistent $\Rightarrow$ weights given by normalized eigenvector

MATHEMATICS OF AHP

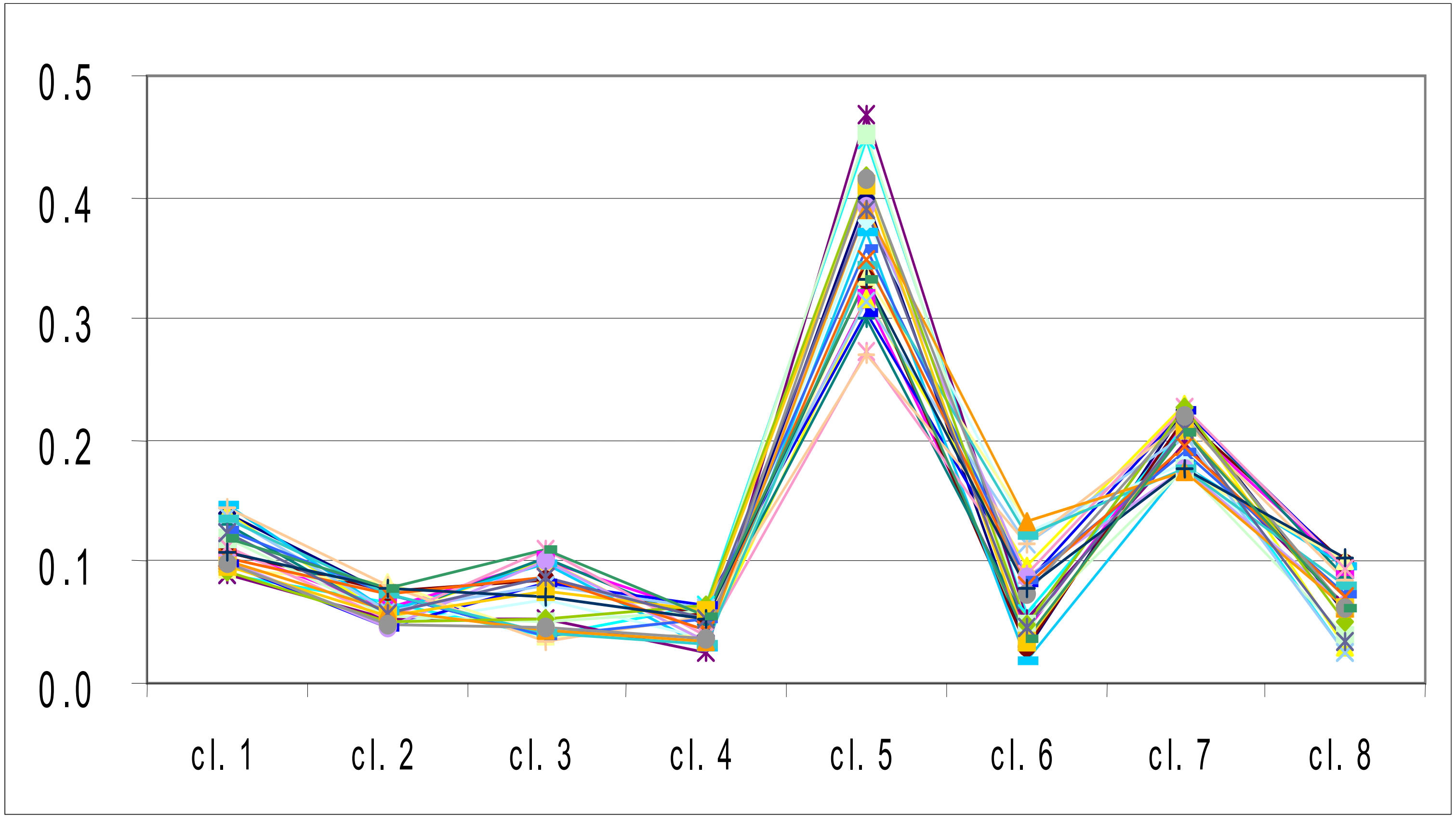
- A not consistent $\Rightarrow$ take eigenvector of $A\mathbf{w} = \lambda_{max}\mathbf{w}$, with λ_{max} largest eigenvalue (always $\lambda_{max} \geq n$ for positive reciprocal matrices and $\lambda_{max} = n$ for consistent ones)

- $\dfrac{\lambda_{max} - n}{n - 1}$ measure of inconsistency (difference divided by the number of the other eigenvalues)

- In order to derive a meaningful interpretation of either the difference or the consistency index, Saaty simulated random pairwise comparisons for different size matrices, calculating the consistency indices, and arriving at an average consistency index for random judgments for each size matrix. He then defined the consistency ratio as the ratio of the consistency index for a particular set of judgments, to the average consistency index for random comparisons for a matrix of the same size (*quoted from Forman and Selly)*

ELICITATION OF EXPERTS' OPINIONS

Values elicited by experts ⇒ similar opinions



MODELS FOR CAST-IRON PIPES

Independent classes $A_i, i = 1, 8$, given by 3 covariates (diameter, location and depth) $\Rightarrow$ find the "most risky" class

- Failures in the network occur at rate $\lambda \Rightarrow$ failures in class A_i occur at rate $\lambda_i = \lambda P(A_i)$
- $P(A_i)$ given by AHP for any expert
- Choice of $\lambda \Rightarrow$ *critical*
  - Proper way to proceed:
    - $*$ Use experts' opinions through AHP to get a Dirichlet prior on $p_i = P(A_i)$
    - $*$ Ask the experts about the global unit rate for gas escapes, λ, and get a gamma prior on it
  - What we did
    - $*$ Estimate λ by MLE $\widehat{\lambda}$ with a unique HPP for the network
    - $*$ Use experts' opinions through AHP to get a prior on $\lambda_i = \lambda P(A_i)$

MODELS FOR CAST-IRON PIPES

- Choice of priors
  - Gamma vs. Lognormal
  - For each expert, eigenvector from AHP multiplied by $\widehat{\lambda} \Rightarrow$ *sample* about $(\lambda_1, \ldots, \lambda_8)$
  - Mean and variance of priors on λ_i's estimated from the *sample* of size 26 (number of experts)
- Posterior mean of failure rate λ_i for each class
- Classes ranked according to posterior means (largest $\Rightarrow$ most keen to gas escapes)
- Sensitivity
  - Classes of Gamma priors with mean and/or variance in intervals
  - Classes of Gamma priors with λ in an interval
- Non-dominated actions under classes of priors/losses

NON-DOMINATED ACTIONS

- $L(\lambda, a)$ loss function with action $a \in \mathcal{A}$ (action space)
- Γ class of priors
- Could consider classes of priors and/or losses
- $E_\pi L(\lambda, a)$ posterior expected loss for any $\pi \in \Gamma$ and $a \in \mathcal{A}$
- $a \in \mathcal{A}$ is a non-dominated alternative if there is no other alternative $b \in \mathcal{A}$ s.t.
  - $E_\pi L(\lambda, b) \leq E_\pi L(\lambda, a), \ \forall \pi \in \Gamma$
  - there is $\pi_0 \in \Gamma$ s.t. $E_{\pi_0} L(\lambda, b) < E_{\pi_0} L(\lambda, a)$

MODELS FOR CAST-IRON PIPES

Hierarchical model

- $Y_i|\lambda_i \sim \mathcal{P}(\lambda_i t_i), i = 1, 8$ $\quad t_i$ known time length

- $\lambda_i|\underline{\beta} \sim \mathcal{G}(\alpha e^{\underline{X}_i^T \underline{\beta}}, \alpha)$, α known, s.t. $\mathcal{E}\lambda_i = e^{\underline{X}_i^T \underline{\beta}}$

- $\pi(\beta)$

- Improper priors, numerical approximation (Albert, 1988)

- Empirical Bayes
  - $\lambda_i|\underline{\beta}, \underline{d} \sim \mathcal{G}(\alpha e^{\underline{X}_i^T \underline{\beta}} + y_i, \alpha + t_i)$, $\lambda_i \perp \lambda_j|\underline{d}$
  - $f(\underline{d}|\underline{\beta}) = \int f(\underline{d}|\underline{\lambda})\pi(\underline{\lambda}|\underline{\beta})d\underline{\lambda}$ $\quad$ maximised by $\underline{\widehat{\beta}}$
    $\Rightarrow \lambda_i|\underline{\widehat{\beta}}, \underline{d} \sim \mathcal{G}(\alpha e^{\underline{X}_i^T \underline{\widehat{\beta}}} + y_i, \alpha + t_i)$, $\forall i$

- "Pure" Bayesian approach $\Rightarrow$ prior on $(\alpha, \underline{\beta})$

ESTIMATES' COMPARISON

- Location: **W** (under walkway) or **T** (under traffic)
- Diameter: **S** (small, $<$ 125 mm) or **L** (large, $\geq$ 125 mm)
- Depth: **N** (not deep, $<$ 0.9 m) or **D** (deep , $\geq$ 0.9 m)

Class	MLE	Bayes ($\mathcal{LN}$)	Bayes ($\mathcal{G}$)	Hierarchical
TSN	*.177*	**.217**	**.231**	**.170**
TSD	*.115*	*.102*	*.104*	*.160*
TLN	*.131*	*.158*	*.143*	*.136*
TLD	**.178**	*.092*	*.094*	*.142*
WSN	.072	.074	.075	.074
WSD	.094	.082	.081	.085
WLN	.066	.069	.066	.066
WLD	.060	.049	.051	.064

Highest value; *2^{nd}-4^{th} values*

- Location is the most relevant covariate
- TLD: 3 failures along 2.8 Km but quite unlikely to fail according to the experts
- $\mathcal{LN}$ and $\mathcal{G}$ $\Rightarrow$ similar answers

NON-DOMINATED ACTIONS

- Range of prior opinions on λ_i's, $i = 1, 8$, by 14 experts vs.

- Range of non-dominated actions based on a quantile class

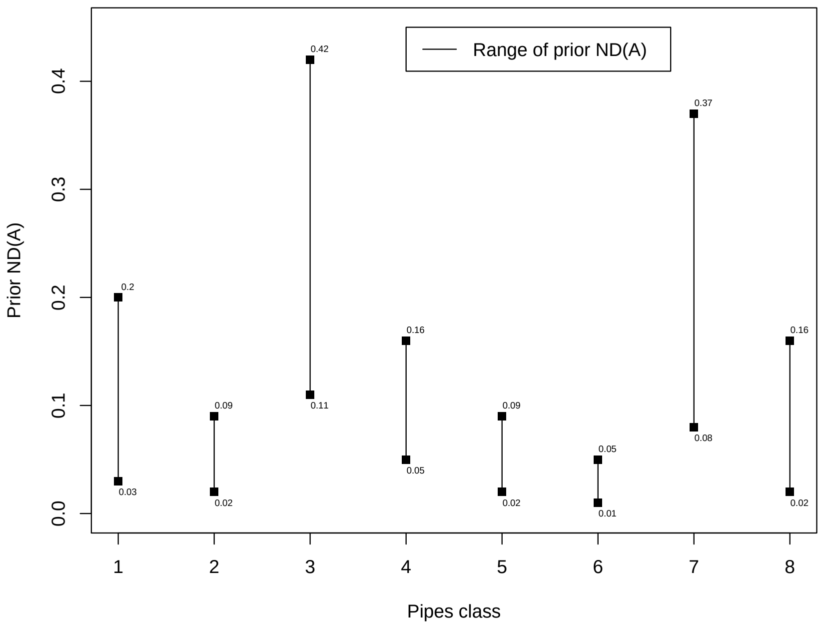
$$\Gamma = \{\Pi : \Pi(q_j) = p_j, j = 1, k\}, \text{ with}$$

  - Π used both for probability measure and cdf
  - p_j's fixed set of probabilities
  - q_j's empirical quantiles based on experts' opinions on λ_i's
  - $k = 3$ or 7 number of quantiles

- *Misalignment in class numbering*

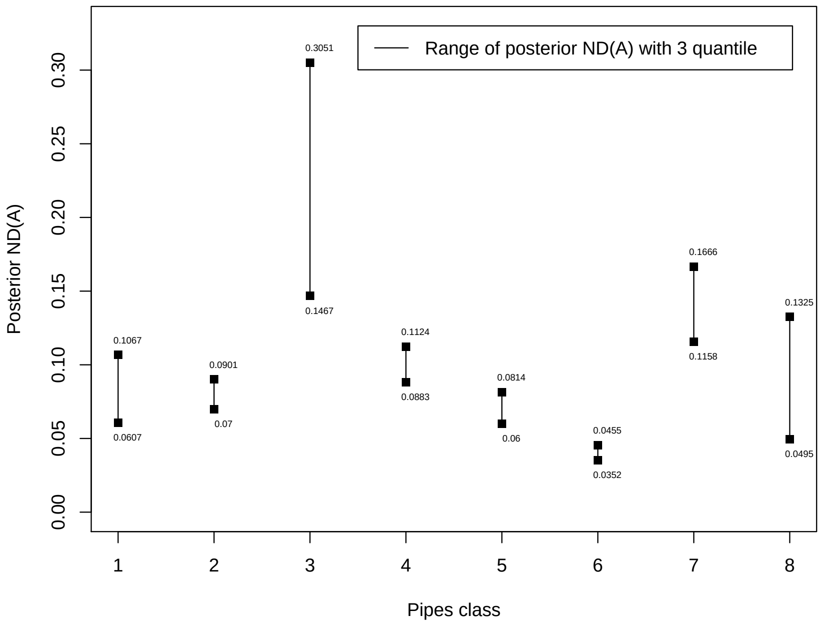
NON-DOMINATED ACTIONS

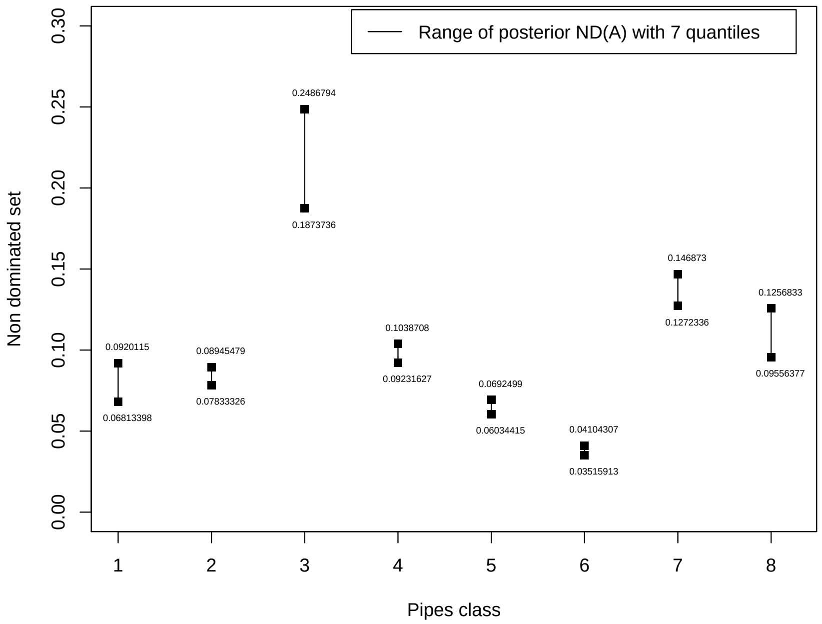
Range of prior opinions



NON-DOMINATED ACTIONS

Range of non-dominated actions for quantile class: 3 quantiles (left) and 7 (right)





GAS ESCAPES: MATERIALS

- traditional cast iron (CI)
  - resistent to corrosion and usage
  - 70-80 years of useful life
  - very fragile with respect to random shocks
- steel (ST)
  - subject to corrosion, despite of cover, and affected by usage
  - 30-40 years of useful life
  - very robust with respect to random shocks
- polyethylene (PE)
  - cheap, currently used in most replacements
  - resistent to corrosion
  - very fragile with respect to ground digging

GAS ESCAPES: MATERIALS

- Evolution over time from CI to ST to PE

- Decision based upon conflicting aspects
  - costs (material, placement, replacement, etc.)
  - reliability (corrosion, frailty, etc.)
  - external conditions (stray currents, traffic, digging, etc.)

- Small example of *Multicriteria Decision Making*

FAILURES IN STEEL PIPES

- Steel pipelines are subject to corrosion, leading to reduction of wall thickness; it can be reduced using
  - bitumen cover
  - cathodic protection (via electric current), working especially when bitumen cover is imperfect
- Most of the low pressure network is without cathodic protection to avoid electrical interference with other metal structures
- Cathodic protected areas can be tested (using electricity) once or twice a year to check for cover status; other areas cannot $\Rightarrow$ need to identify riskier cases for their preventive inspection $\Rightarrow$ *Statistics*
- Different causes (e.g. digging) destroy bitumen cover and start corrosion process

FAILURES IN STEEL PIPES

- Data: 53 failures in 30 years on an expanding net, ≈ 380 Km long (year 2000)
- Three major factors related to failures
  - Age
    - continuous electrolytic process reducing wall thickness
  - Type of corrosion
    - natural corrosion
    - galvanic corrosion
    - corrosion by interference (or stray currents)
  - Lay location
    - near streetcar substations or train stations
    - o.w.

FAILURES IN STEEL PIPES: AGE

Physical aspects

- Different installation dates of different sections
  - same physical properties?
  - same installation procedure?
- Unknown date for start of corrosion process
- Different operating conditions
  - diameter, location, depth
  - electricity in the ground
  - kind of pipes (e.g. junction)

FAILURES IN STEEL PIPES: AGE

Mathematical aspects

- Different installation dates of different sections
  - A unique process or as many as (say) the installation years?
  - If many processes, one for each installation date or one starting (say) on July 1st every year?
  - Parameters for each process?
    - equal
    - similar (exchangeable)
    - completely independent
- Unknown date for start of corrosion process
  - model or ignore it?
- Different operating conditions
  - reasonable or feasible discriminating among them?

FAILURES IN STEEL PIPES

Bad data quality

- Diameter, location, depth and installation data of broken pipes sometimes unavailable
  $\Rightarrow$ *impossible to perform analysis similar to the one on cast-iron pipes (there data collected in 1991-96 and here on a larger period because of their scarceness and age dependence)*

- evolution of the net exactly known for the last 10 years and approximately since first installation (1930)
  $\Rightarrow$ *we assume it known, after interviewing company's experts, and performing linear interpolation, adjusted for the years of WWII*

- unknown installation date of few failed pipes
  $\Rightarrow$ *imputation with statistical methods or looking at nearby failed pipes*

- (probably) improperly recorded escapes (e.g. six in 24 hours in different parts of the city, without any physical explanation, e.g. earthquake)
  $\Rightarrow$ *impossible to examine what happened so that we kept them as they were and statistical analysis had been badly affected*

FAILURES IN STEEL PIPES

- Network split into subnetworks based upon year of installation, *as if* pipes were installed on July, 1st each year

- Removals and replacements *not relevant* for the network reliability $\Rightarrow$ *repairable system*

- Independent PLP's for each subnetwork

- *Superposition Theorem*: Sum of independent NHPPs with intensity functions $\lambda_i(t)$ is still a NHPP with intensity function $\lambda(t) = \sum \lambda_i(t)$

- The same parameters vs. exchangeable ones in each PLP

- Statistical analysis
  - parameter estimation
  - prediction of future escape times
  - computation of reliability measures

FAILURES IN STEEL PIPES

- Experts asked about time up to first failure
  - Choice of section of the network (e.g. length l)
  - Choice of time intervals in a list (e.g. $[T_0, T_1]$)
  - Degree of belief on each interval (choice among 95%, 85% and 75%); e.g. for PLP (M, β)
    $\Rightarrow P([T_0, T_1]) = \exp\{-lMT_0^\beta\} - \exp\{-lMT_1^\beta\} = 0.95$
  - Check for consistency, e.g. $A \subset B \not\Rightarrow P(A) > P(B)$

- Pooling of experts' opinions
  - $\Rightarrow$ *sample* from priors
  - $\Rightarrow$ hyperparameters in priors, matching moments

MODELS FOR STEEL PIPES

PLP's with exchangeable M_s and β_s; known installation dates

- r number of years
- $\underline{M} = (M_1, \ldots, M_r)$ and $\underline{\beta} = (\beta_1, \ldots, \beta_r)$
- l_s length of network installed in year $s = 1, \ldots, r$
- $\Rightarrow r$ independent PLP's, nonhomogeneous in time but homogeneous in space: $\lambda_s(t) = l_s M_s \beta_s t^{\beta_s - 1}$, $s = 1, \ldots, r$
- $[T_0, T_1]$: interval when recording failures
- δ_k: installation date of k-th failed pipe
- $|I_s|$ number of failures of pipes installed in year $s = 1, \ldots, r$

MODELS FOR STEEL PIPES

Likelihood function $L(\underline{M}, \underline{\beta}; \underline{t}, \underline{\delta})$

$$\prod_{k=1}^{n} \beta_{\delta_k} l_{\delta_k} M_{\delta_k} (t_k - \delta_k)^{\beta_{\delta_k}-1} e^{-\sum_{s=1}^{r} l_s M_s [(T_1 - s)^{\beta_s} - (s \vee T_0 - s)^{\beta_s}]}$$

- $M_s \sim \mathcal{E}(\theta_M)$ and $\beta_s \sim \mathcal{E}(\theta_\beta)$, $s = 1, \ldots, r$
- $\theta_M \sim \mathcal{E}(\tau_M)$ and $\theta_\beta \sim \mathcal{E}(\tau_\beta)$
- M_s's $\perp$ β_s's but exchangeable among themselves
  - $\pi(M_1, \ldots, M_r) = \int \prod_{i=1}^{r} \theta_M e^{-\theta_M M_i} \pi(\theta_M) d\theta_M$
  - $\pi(\beta_1, \ldots, \beta_r) = \int \prod_{i=1}^{r} \theta_\beta e^{-\theta_\beta \beta_i} \pi(\theta_\beta) d\theta_\beta$

MODELS FOR STEEL PIPES

- Posterior $\pi(\underline{M}, \underline{\beta}|\underline{t}, \underline{\delta})$ obtained integrating out θ_M and θ_β

- $\Rightarrow \pi(\underline{M}, \underline{\beta}|\underline{t}, \underline{\delta}) \propto$

$$\left(\prod_{s=1}^{r}(l_s M_s \beta_s)^{|I_s|}\right)\left(\prod_{k=1}^{n}(t_k-\delta_k)^{\beta_{\delta_k}-1}\right) e^{-\sum_{s=1}^{r} l_s M_s [(T_1-s)^{\beta_s}-(s\vee T_0-s)^{\beta_s}]}.$$

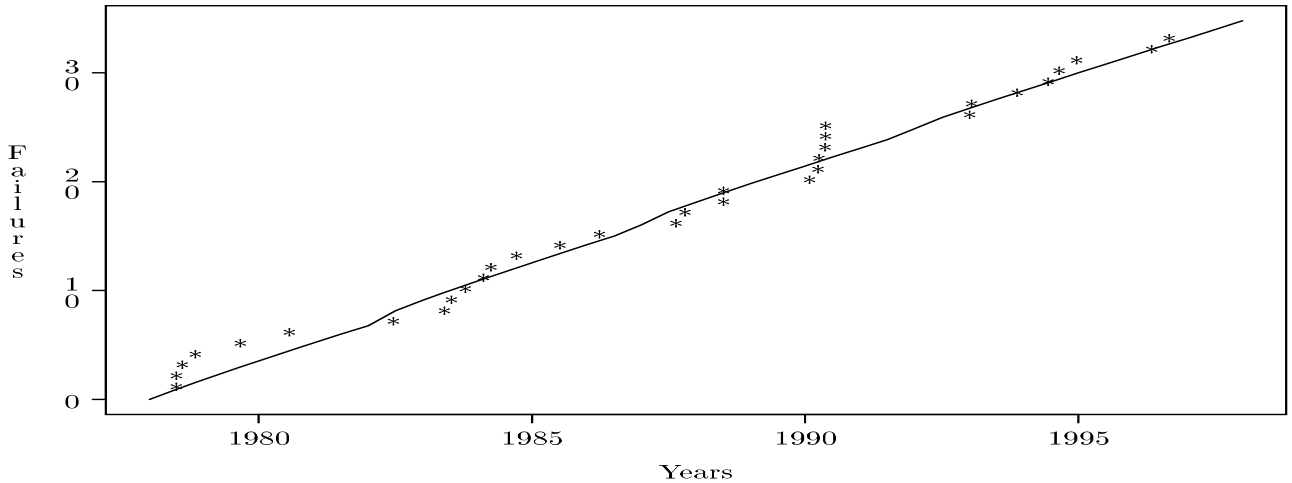
$$\cdot \tau_M \tau_\beta \frac{r!}{\left[\sum_{s=1}^{r}(M_s+\tau_M/r)\right]^{r+1}} \frac{r!}{\left[\sum_{s=1}^{r}(\beta_s+\tau_\beta/r)\right]^{r+1}}$$

MODELS FOR STEEL PIPES

Exchangeable M and β; known installation dates

95% credible intervals for reliability measures:

- System reliability over 5 years: $P\{N(1998, 2002) = 0\} \Rightarrow [0.0000964, 0.01]$
- Expected number of failures in 5 years: $EN(1998, 2002) \Rightarrow [4.59, 9.25]$
- Mean value function (solid) vs. cumulative $\#$ failures (points)



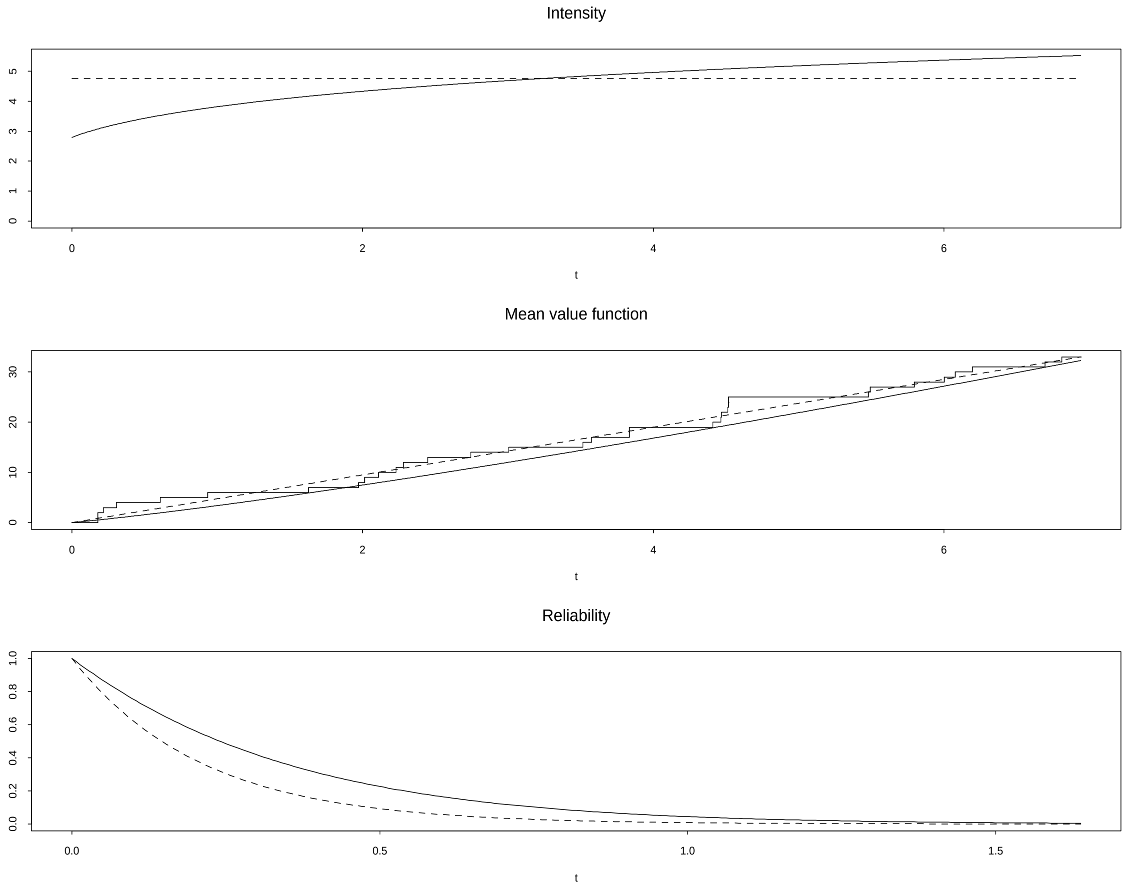
MODELS FOR STEEL PIPES

Other models considered by Pievatolo and Ruggeri (2004)

- Same M and/or β
- Unknown installation dates
  - Prior distribution on them
- Censored data
  - Known number (but not times) of gas escapes before observation period

A VERY SIMPLE NHPP MODEL

MLE (dashed) vs. Bayes (solid) for $\lambda_\theta(t) = a\ln(1 + bt) + c$



TYPES OF CORROSION

- Natural corrosion
  - due to ground properties, e.g. very wet ground is a good conductor easing development of the electrolytic phenomenon
- Galvanic corrosion
  - network made of different materials
  - contact of two different materials with imperfect insulation
  - corrosion started by potential difference between two different materials
- Corrosion by interference (or stray currents)
  - presence of stray currents in the ground coming from other electrical plants badly insulated (e.g. streetcar substations or train stations)
  - when discharging on steel pipe they increase the corrosion rate by various orders of magnitude

LAYING LOCATION

- Areas near streetcar substations or train stations (Zone A)
  - Streetcar substations generate current, which goes through the aerial line and is transformed into power by the streetcar; then it goes back to the substation through the steel streetcar tracks and the trunk of negative electric cables hidden underground (which are the cause of stray currents due to bad insulation)
  - Near railway stations, the stray currents derive not only by the bad insulation of the tracks, but also by the strong electrical field coming from the passage of the train
- Other areas (Zone B)

DATA: ZONE AND CORROSION

Failure rate (failures) by zone and type of corrosion

	Natural (N)	Galvanic (G)	By interference (I)
Zone A (12 km^2)	0.583 (7)	0.083 (1)	0.500 (6)
Zone B (88 km^2)	0.068 (6)	0.057 (5)	0.091 (8)

- Different failure rates for natural corrosion

  $\Rightarrow$ suspects on right reporting by repairing squads

EXPERTS' OPINIONS

- Experts
  - 2 technicians assessing pipes conditions after excavation
  - 2 engineers expert of technical and management aspects

- Analytic Hierarchy Process (AHP) as before
  - Qualitative pairwise comparisons with answers: *equally likely*, *a little more likely*, *much more likely*, *clearly more likely*, *definitely more likely*
    $\Rightarrow$ quantitative judgements

- Questions
  - *In your opinion is a failure more likely to happen in zone A or in zone B? How much more likely*? $\Rightarrow P(failure\ in\ A) = P(A)$ and $P(B)$
  - Pairwise comparisons like: *In an area with (without) streetcar substations or railways stations is it more likely to have natural or galvanic corrosion? How much more likely?* $\Rightarrow P(N|A)$, $P(G|A)$, $P(I|A)$, $P(N|B)$, $P(G|B)$, and $P(I|B)$

EXPERTS' OPINIONS

- $P(A)$ and $P(B)$ known and
- $P(N|A)$, $P(G|A)$, $P(I|A)$, $P(N|B)$, $P(G|B)$, and $P(I|B)$ known
- $\Rightarrow P(N) = P(N|A)P(A) + P(N|B)P(B)$
- $\Rightarrow P(A|N) = \frac{P(N|A)P(A)}{P(N)}$
- The same for $P(G), P(I), P(A|G), P(A|I)$
- Probabilities obtained for all experts and pooled

	Mean	St. dev.
$P(A)$	0.7938	0.1962
$P(B)$	0.2063	0.1962
$P(A\|N)$	0.6133	0.2114
$P(A\|G)$	0.6221	0.2168
$P(A\|I)$	0.9581	0.0574
$P(N)$	0.1636	0.0403
$P(G)$	0.2767	0.1298

POSTERIOR PROBABILITIES

- $P(A) = p$ probability that a failure occurs in zone A

- Conditional upon observing n total failures, the number n_A of failures in A is a Binomial r.v.
$\Rightarrow p(n_A|n,p) \propto \binom{n}{n_A} p^{n_A}(1-p)^{n-n_A}$

- Prior on p: $\mathcal{B}e(a,b)$ conjugate w.r.t. Binomial model

- $\Rightarrow$ posterior: $\mathcal{B}e(a+n_A, b+n-n_A)$

- Bayes estimator of p: posterior mean $\dfrac{a+n_A}{a+b+n}$

		Historical (MLE)	Prior	Posterior
p	(zone A, 12 km^2)	0.4528	0.7938	0.4790
$1-p$	(zone B, 88 km^2)	0.5472	0.2062	0.5210

POSTERIOR PROBABILITIES

	Historical (MLE)	Prior		Posterior	
		Mean	St. Dev.	Mean	St. Dev.
$P(A\|N)$	0.5385	0.6133	0.2114	0.5662	0.1065
$P(A\|G)$	0.1667	0.6221	0.2168	0.4125	0.0351
$P(A\|I)$	0.4286	0.9581	0.0574	0.6700	0.0909

	Historical (MLE)	Prior		Posterior	
		Mean	St. Dev.	Mean	St. Dev.
$P(N)$	0.3940	0.1636	0.0403	0.2290	0.0388
$P(G)$	0.1818	0.2767	0.1298	0.2498	0.0400
$P(I)$	0.4242	0.5597	0.1565	0.5212	0.0461

MODEL SELECTION

- Gas escapes caused by corrosion: natural, galvanic and by stray currents
- $\lambda(t) = \beta$ (HPP) vs. $\lambda(t) = \beta t/(\gamma + t)$ (NHPP)
- Number of failures in $[0, T]$
  - HPP: $\mathcal{P}(\beta T)$
  - NHPP: $\mathcal{P}(\int_0^T \beta t/(\gamma + t) dt)$
- Bayes factor $BF = \dfrac{\int L(\beta, 0)\Pi(d\beta)}{\int L(\beta, \gamma)\Pi(d\beta)\Pi(d\gamma)}$

UNCERTAINTY ON PRIOR DISTRIBUTION

- So far we have assumed there exists a unique prior but it is very questionable
  - impossibility of specifying a distribution exactly based upon experts' opinions
  - group of people with different opinions
- Specify class of priors, compatible with prior knowledge
- Compute upper and lower bounds on quantity of interest and check if they are close $\Rightarrow$ robustness or not
- $\beta \sim \mathcal{G}(a, b)$ and $\pi(\gamma) \in \Gamma = \{\pi : median\ at\ 1\}$
- Quantity of interest here: Bayes factor

MODEL SELECTION

Corrosion	BF	$E\beta \mid d$	$E\gamma \mid d$
Galvanic	$(0.68, 0.82)$	$(0.59, 1.10)$	$(0.59, 8.08)$
Natural	$(0.25, 0.54)$	$(0.87, 2.40)$	$(0.71, 22.64)$
Stray Currents	$(2.00, 13968.02)$	$(0.82, 1.00)$	$(0.00, 0.16)$

- $\lambda(t) = \beta$ (HPP) vs. $\lambda(t) = \beta t/(\gamma + t)$ (NHPP)
- Bayes factor $BF = \dfrac{\int L(\beta, 0)\Pi(d\beta)}{\int L(\beta,\gamma)\Pi(d\beta)\Pi(d\gamma)}$
- Upper and lower bounds on $BF \Rightarrow$ HPP better for stray currents and worse o.w.

NONPARAMETRIC APPROACH

events in $[T_0, T_1] \sim \mathcal{P}(\Lambda[T_0, T_1])$, with $\Lambda[T_0, T_1] = \Lambda(T_1) - \Lambda(T_0)$

Parametric case: $\Lambda[T_0, T_1] = \int_{T_0}^{T_1} \lambda(t) dt$

Nonparametric case: $\Lambda[T_0, T_1] \sim \mathcal{G}(\cdot, \cdot) \Rightarrow \Lambda$ d.f. of the random measure M

Notation: $\mu B := \mu(B)$

Definition 1 *Let α be a finite, σ-additive measure on $(\mathbb{S}, \mathcal{S})$. The random measure μ follows a* **Standard Gamma** *distribution with* shape α *(denoted by $\mu \sim \mathcal{GG}(\alpha, 1)$) if, for any family $\{S_j, j = 1, \ldots, k\}$ of disjoint, measurable subsets of $\mathbb{S}$, the random variables μS_j are independent and such that $\mu S_j \sim \mathcal{G}(\alpha S_j, 1)$, for $j = 1, \ldots, k$.*

Definition 2 *Let β be an α-integrable function and $\mu \sim \mathcal{GG}(\alpha, 1)$. The random measure $M = \beta\mu$, s.t. $\beta\mu(A) = \int_A \beta(x)\mu(dx), \forall A \in \mathcal{S}$, follows a* **Generalised Gamma** *distribution, with* shape α *and* scale β *(denoted by $M \sim \mathcal{GG}(\alpha, \beta)$).*

NONPARAMETRIC APPROACH

Consequences:

- $\mu \sim \mathcal{P}_{\alpha,1}, \mathcal{P}_{\alpha,1}$ unique p.m. on $(\Omega, \mathcal{M})$, space of finite measures on $(\mathbb{S}, \mathcal{S})$, with these finite dimensional distributions

- $M \sim \mathcal{P}_{\alpha,\beta}$, **weighted random measure**, with $\mathcal{P}_{\alpha,\beta}$ p.m. induced by $\mathcal{P}_{\alpha,1}$

- $EM = \beta\alpha$, i.e. $\int_\Omega M(A)\mathcal{P}_{\alpha,\beta}(dM) = \int_A \beta(x)\alpha(dx), \forall A \in \mathcal{S}$

Theorem 1 *Let* $\underline{\xi} = (\xi_1, \ldots, \xi_n)$ *be* n *Poisson processes with intensity measure* M. *If* $M \sim \mathcal{GG}(\alpha, \beta)$ *a priori, then* $M \sim \mathcal{GG}(\alpha + \sum_{i=1}^n \xi_i, \beta/(1 + n\beta))$ *a posteriori.*

NONPARAMETRIC APPROACH

Data: $\{y_{ij}, i = 1 \dots k_j\}_{j=1}^n$ from $\underline{\xi} = (\xi_1, \dots, \xi_n)$

Bayesian estimator of M: measure $\widetilde{M}$ s.t., $\forall S \in \mathcal{S}$,

$$\widetilde{M}S = \int_S \frac{\beta(x)}{1+n\beta(x)} \alpha(dx) + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \frac{\beta(y_{ij})}{1+n\beta(y_{ij})} \mathbb{I}_S(y_{ij})$$

Constant $\beta \Rightarrow \widetilde{M}S = \frac{\beta}{1+n\beta}[\alpha S + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \mathbb{I}_S(y_{ij})]$

Bayesian estimator of reliability R, $RS = P(\xi S = 0)$, $S \in \mathcal{S}$:

$$\widetilde{R}S = \exp\left\{ -\int_S \ln(1 + \frac{\beta(x)}{1+n\beta(x)}) \alpha(dx) - \sum_{j=1}^{n} \sum_{i=1}^{k_j} \ln(1 + \frac{\beta(y_{ij})\mathbb{I}_S(y_{ij})}{1+n\beta(y_{ij})}) \right\}$$

Constant $\beta \Rightarrow \widetilde{R}S = \left(1 + \frac{\beta}{1+n\beta}\right)^{-(\alpha S + \sum_{j=1}^{n} \xi_j S)}$

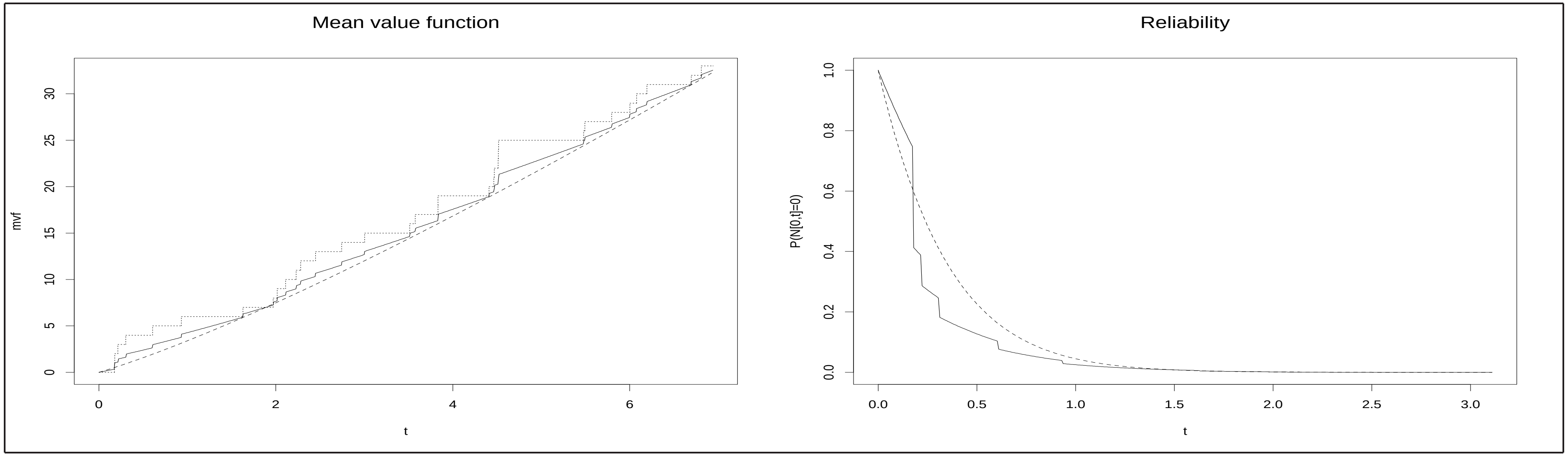
STEEL PIPES

Parametric NHPP: $\widetilde{\Lambda_\theta}(t) = \int_0^t [\tilde{a} \log(1 + \tilde{b}t)]dt + \hat{c}t$

Nonparametric model: $M \sim \mathcal{P}_{\alpha,\beta} : \alpha(ds) := \widetilde{\Lambda_\theta}(s)/\sigma\, ds$, $\beta(s) := \sigma$

$\Rightarrow \mathcal{E}MS = \widetilde{\Lambda_\theta}S$ and $VarMS = \sigma\widetilde{\Lambda_\theta}S$

$\Rightarrow MS$ "centered" at parametric estimator $\widetilde{\Lambda_\theta}S$ and closeness given by σ



Nonparametric (solid) and parametric (dashed) estimators and cumulative $N[0,t]$ (dotted).

PARAMETRIC VS. NONPARAMETRIC

$[0,T]$ split into n disjoint I_j, $j = 1, \ldots, n$

Data: $\underline{k} = (k_1, \ldots, k_n)$, with $k_j = \{\#\text{obs. in } I_j\} \Rightarrow f(\underline{k} \mid \Lambda) = e^{-\Lambda(T)} \prod_{j=1}^{n} \frac{(\Lambda I_j)^{k_j}}{k_j!}$

Parametric: $P(\underline{k} \mid H_P) = \int_{\mathbb{R}_+^3} e^{-\Lambda_\theta(T)} \prod_{j=1}^{n} \frac{[\Lambda_\theta I_j]^{k_j}}{k_j!} \pi(\theta)\, d\theta$

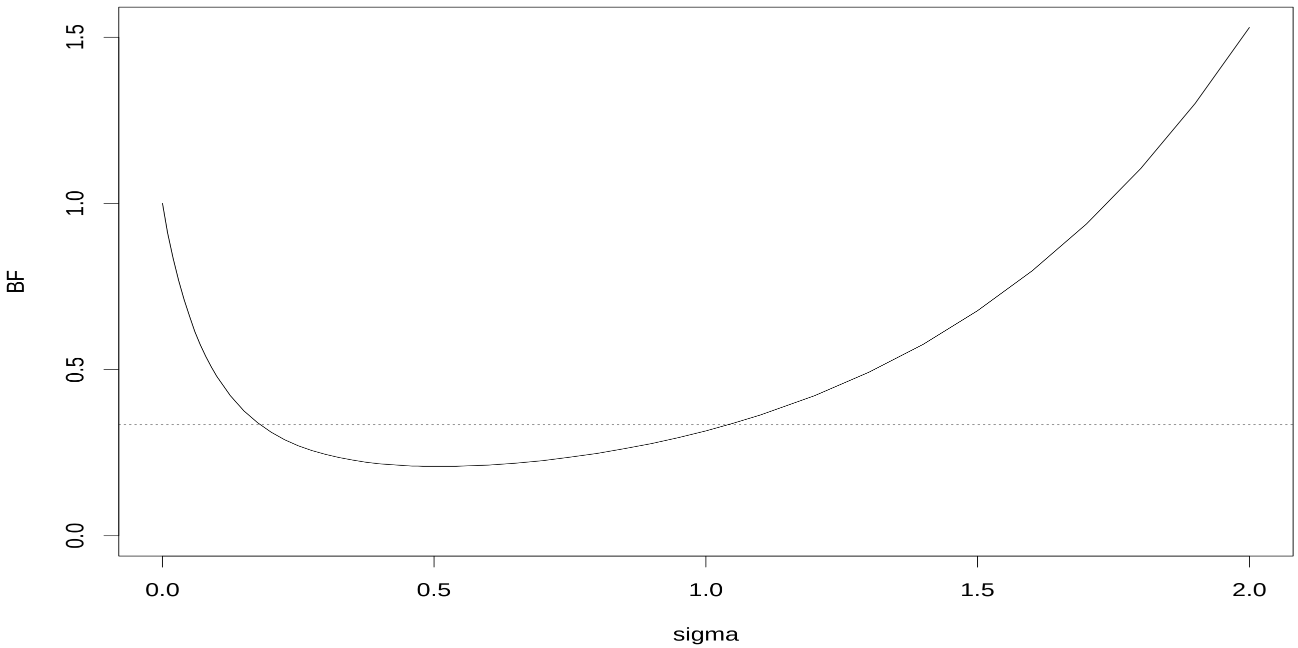
Nonparametric: $\underline{k} \mid M, \theta \sim f(\underline{k} \mid M_\theta)$, $M \mid \theta \sim \mathcal{GG}(\Lambda_\theta/\sigma, \sigma)$ and $\theta \sim \pi$:

$$P(\underline{k} \mid H_N) = \int_{\mathbb{R}_+^3} \prod_{j=1}^{n} \left[\frac{\prod_{i=0}^{k_j-1} (\Lambda_\theta I_j + i\sigma)}{k_j! \exp\left[\left(\frac{\Lambda_\theta I_j}{\sigma} + k_j\right) \ln(1+\sigma)\right]} \right] \pi(\theta)\, d\theta$$

Bayes Factor: $BF_{PN} = \dfrac{P(\underline{k} \mid H_P)}{P(\underline{k} \mid H_N)} = \dfrac{\int_{\mathbb{R}_+^3} e^{-\Lambda_\theta(T)} \prod_{j=1}^{n} (\Lambda_\theta I_j)^{k_j} \pi(\theta)\, d\theta}{\int_{\mathbb{R}_+^3} \prod_{j=1}^{n} \left[(1+\sigma)^{-(\Lambda_\theta I_j/\sigma + k_j)} \prod_{i=0}^{k_j-1} (\Lambda_\theta I_j + i\sigma)\right] \pi(\theta)\, d\theta}$

PARAMETRIC VS. NONPARAMETRIC

Bayes factor BF_{PN} as a function of σ



CASE STUDY: SUBWAY TRAINS' FAILURES

- New underground line opened in a major Italian city on 3/5/1990 just before the FIFA World Cup
- 40 trains delivered to the transportation company between 11/1989 and 3/1991 and put on service between 4/1990 and 7/1992
- Transportation company viewpoint: interest about delivered trains for
  - financial costs (LCC: Life Cycle Cost)
    - forecast of costs over all useful life
  - service quality (RAM: Reliability, Availability, Maintainability)
    - forecast of quality over all useful life
- *Main goal: check of fulfillment of contract specifics before warranty expiration*

RAM: RELIABILITY, AVAILABILITY, MAINTAINABILITY

- EU laws require specification of precise performance values (instead of generic descriptions) in competitions to assign contracts in public services

- RAM is an important performance index (sometimes RAMS, ”Safety”)

- RAM(S) parameters important for
  - manufacturer
    - design, components’ choice, experiments, production and control
  - transportation company
    - purchase cost, maintenance cost, spare parts, service unavailability, etc.

- RAM(S) parameters strongly correlated and depending on *failures*

RAM: RELIABILITY

- Main features
  - failure rate $\lambda = \lambda(t)$
  - mean time to failure $MTTF = \int_0^\infty t f(t) dt = \int_0^\infty S(t) dt$
  - mean time between failures $MTBF$
  - mean distance between failures
  - mean time between unscheduled maintenance actions
- Main reliability indices in the company
  - number of failures for 1 million Km's
    - with delays exceeding 30 minutes
    - with delays exceeding 5 minutes
    - with delays below 5 minutes
  - number of failures of an item (e.g. door) in 100,000 Km's

R**A**M: AVAILABILITY

Probability that a system operates correctly, under specified conditions, *when required* (compare with *for a given time*)

- $MTBF/(MTBF + MTTR)$ ($MTTR$ mean time to repair)
- N_a/N_t (N_a trains available at rush hour out of a total of N_t)
- $\sum_i (OH_i - PMH_i - CMH_i) / \sum_i OH_i$
  - i index of interested items (e.g. all trains)
  - OH_i operation hours for item i
  - PMH_i preventive maintenance hours for item i
  - CMH_i corrective maintenance hours for item i

RA**M**: MAINTAINABILITY

Probability of performing proper maintenance within a given time

- Corrective maintenance cost: $CM = 1000 \sum_i (WCC_i \cdot WTC_i + MCC_i)/K$
  - i index of the n requested corrective maintenance interventions
  - WCC_i unit cost of workforce for intervention i
  - WTC_i time of workforce for intervention i
  - MCC_i total cost of materials for intervention i
  - K km's run during observation period
- Preventive maintenance cost: $PM = 1000 \sum_i (WCP_i \cdot WTP_i + MCP_i)/K$
- Global index $GM = CM + PM$ to be minimized

LIFE CYCLE COST

Total cost of ownership of machinery and equipment, including its cost of acquisition, operation, maintenance, conversion, and/or decommission

- Need to monitor costs during all useful life (lower purchase price could lead to higher costs in future)

- Main categories of costs incurred by transportation company upon introduction of new trains
  - organization in general (e.g. administration)
  - organization strictly related to trains (e.g. training of drivers and technicians, maintenance procedures)
  - batch of train (e.g. size, deposit)
  - individual train (e.g.failures)

(SUBSET OF) COSTS

- Trains purchase
- Subway line (e.g. tracks, platform, stations) costruction/modification
- Deposit (e.g. for maintenance, night stops)
- Personnel costs (e.g. training, salary)
- Operating costs (e.g. energy, SW and HW, cleaning)
- Maintenance (e.g. training and size of squad, site and size for spare parts)
- Technical changes
- Missed gains and image damages
- Environmental costs
- End-of-life dismissal of trains

STATEMENT OF THE PROBLEM

- Improving RAM and LCC is a huge task
- Both RAM and LCC are strongly affected by reliability
- $\Rightarrow$ interest in failures of some components, identified by EDA
  - doors (major cause of failures)
  - engine wheels
  - two converters
- Doors failure data collected between 1/4/1990 and 31/12/1998
- Data collected by B.Sc. student at Politecnico di Milano for his dissertation

STATEMENT OF THE PROBLEM

- Interest in process modelling and estimation
  - all doors failure regardless of the cause
  - only individual, major failure causes
- Interest in reliability check before warranty expiration
  - reliability standards set by the contract between manufacturer and transportation company
  - structural failures occurred during the warranty time are responsibility of the manufacturer (and the company's after warranty expiration)
  - transportation company can ask for manufacturer's intervention on trains if there is evidence of poor reliability during warranty time (but not later)

REPAIRABLE SYSTEMS

Theorem (due to Grigelionis, see Thompson, 1988, p.69) states that, *under suitable conditions*, superposition of *many* failure processes, one for each failure mode, is *approximately* a Poisson process

- Doors are complex systems made of many components, subject to different failure causes

- Upon failures, repairs are immediate (i.e. done in a negligible time w.r.t. doors lifetime) and minimal (i.e. just the cause of the failure is fixed)

- Repairs bring reliability back to its status just before failures (*bad as old* property)

$\Rightarrow$ Non-homogeneous Poisson process (NHPP)

FAILURE MODELS

- All door failures together regardless of cause or separated by cause
- A NHPP used directly to model failures or ...
- ... NHPP for each *independent* cause and apply **Superposition Theorem**
  - consider n independent Poisson processes $N_t^{(i)}, t \geq 0$, with intensity function $\lambda^{(i)}(t)$, $i = 1, \ldots, n$
  - consider the sum process $N_t = \sum_{i=1}^{n} N_t^{(i)}$, $t \geq 0$
  - $\Rightarrow N_t$ is a Poisson process with intensity function $\lambda(t) = \sum_{i=1}^{n} \lambda^{(i)}(t)$
- *In any case*, a NHPP for the failures

FAILURE MODELS

- Trains could be considered as **a**) *different* systems or **b)** the *same, unique* system
  - **a**) likelihood as product of likelihoods based upon individual intensities
  - **b)** likelihood based upon sum of intensities (as a consequence of the Superposition Theorem)

- Behaviour of different trains could be ...
  - **i)** *equal* $\Rightarrow$ same parameters
  - **ii)** *similar* $\Rightarrow$ different parameters from same distribution (*exchangeability*)
  - **iii)** *different* $\Rightarrow$ different parameters from different distributions
  - **iv)** *"almost" different* $\Rightarrow$ some common parameters and other different ones from different distributions

- EDA lead to **iii)** in Part I, and further physical considerations lead to **iv)** in Part II

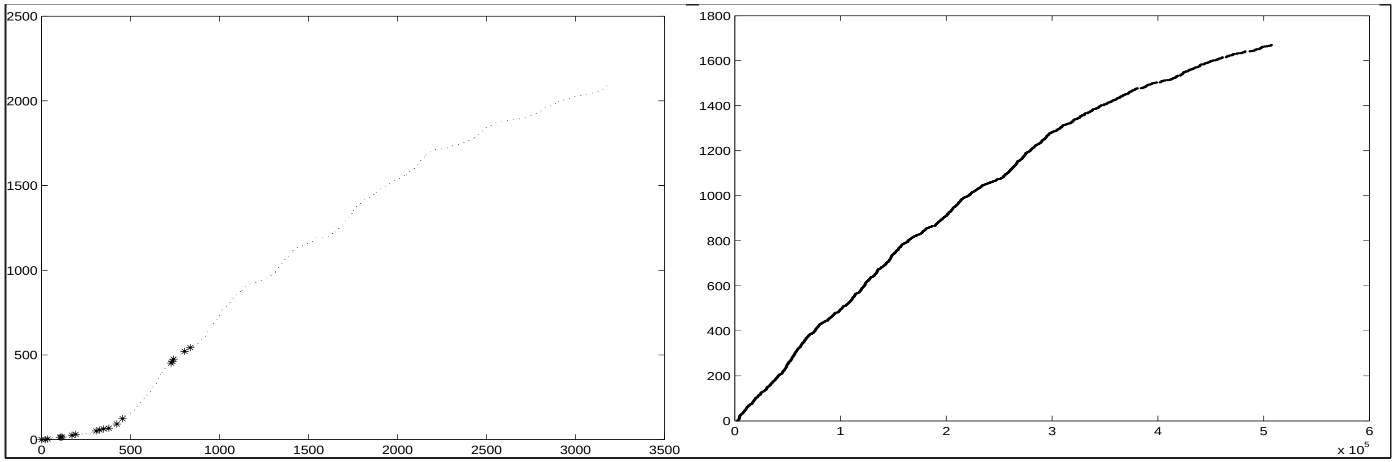
FAILURE MODELS

- Likelihood based on failures in one train
  - Estimation of its intensity function
  - Comparison between its cumulative and estimated expectation of number of failures
  - Predictive distribution and expectation of its future failures
- Likelihood based on failures in all train
  - Comparison between cumulative and estimated expectation of number of failures among all trains
  - Predictive distribution and expectation of future failures of a new train

DOUBLE SCALE DATA

Data: more than 2000 door failures of 40 trains, put on service from 1/4/1990 to 20/7/1992, observed up to 31/12/1998

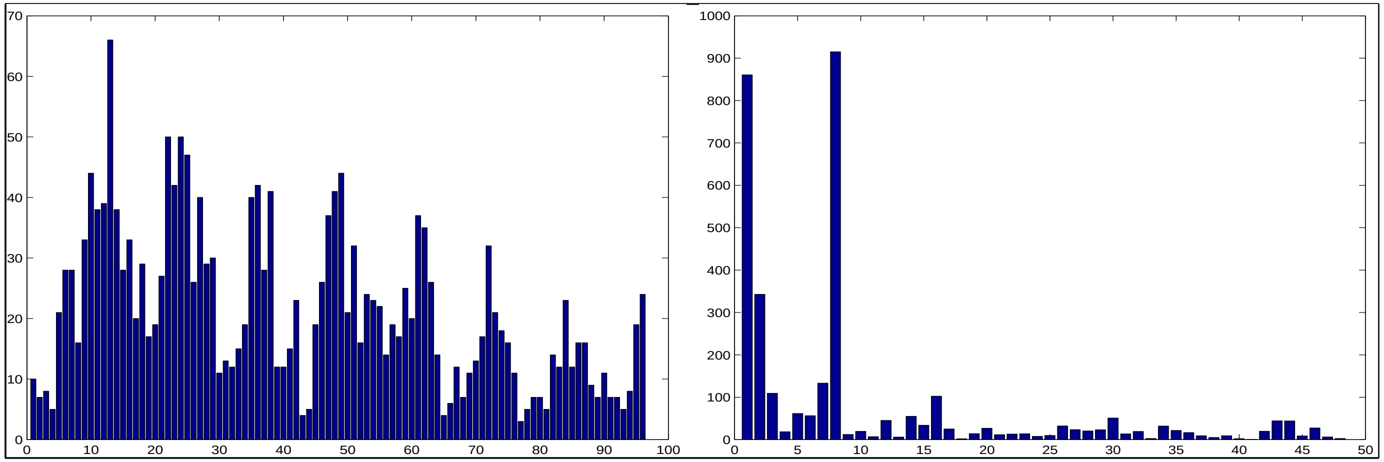
Goal: checking components reliability before warranty's expiration



Failures vs. days (left) and failures vs. kilometers (right)

- Concavity denotes improvement over time
- Oscillations
- Transient behaviour during first 500 days

SEASONALITY



Left: Monthly no. of failures for the 40 trains starting January 1991

Right: Spectrum of the time series of the monthly number of failures from 1991 to 1998

- Decreasing trend
- Periodicity (estimated at 12 months by the spectrum)
- NHPP: $\lambda(t) = \exp\{\alpha + \rho \sin(\omega t + \theta)\}$

MODEL FOR DOORS FAILURES

Marked Poisson process on time scale

$\lambda(t;\theta_1,\theta_2) = \mu(g(t);\theta_1)s(t;\theta_2)$

- $\mu(k;\theta_1) = \beta_0 \dfrac{\log(1+\beta_1 k)}{(1+\beta_1 k)}$
  - $\mu(0;\theta_1) = 0$, maximum at $(e-1)/b_1$ and $\lim_{k\to\infty} u(k;\theta_1) = 0$
  - m.v.f. $\Lambda(k) = \beta_0 \log^2(1+\beta_1 k)/(2\beta_1)$

    suitable for actual cumulative number of failures
- $s(t;\theta_2) = \exp\{\rho\cos(\omega t+\varphi)\}$(periodic component)
- From EDA we could take $k = g(t) = at + bt^2$ and substitute above
- We actually took kilometers $k|t \sim \mathcal{N}(g(t),\sigma^2)$

MODEL FOR DOORS FAILURES

- j-th train monitored in $[0, T_j]$
- Failures at times $(t_1, \ldots, t_{n_j}) = \mathbf{t_j}$ and kilometers $(k_1, \ldots, k_{n_j}) = \mathbf{k_j}$
- Likelihood for j-th train

  $L_j(\theta_1, \theta_2) = \prod_{i=1}^{n_j} \mu(g(t_i); \theta_1) s(t_i; \theta_2) \exp \left[- \int_0^{T_j} \mu(g(t); \theta_1) s(t; \theta_2) dt \right]$
- Non-Bayesian analysis

Parameter	MLE	C.I.	Parameter	MLE	C.I.
$a \times 10^{-2}$	1.209	[1.171, 1.247]	$b \times 10^2$	2.025	[1.862, 2.188]
$\sigma^2 \times 10^{-7}$	5.809	[4.214, 8.345]	$\rho \times 10$	3.234	[0.000, 6.779]
$\beta_0 \times 10^2$	7.358	[5.640, 9.076]	$\beta_1 \times 10^5$	2.239	[1.938, 2.540]

DIAGNOSTIC PLOTS FOR ONE TRAIN

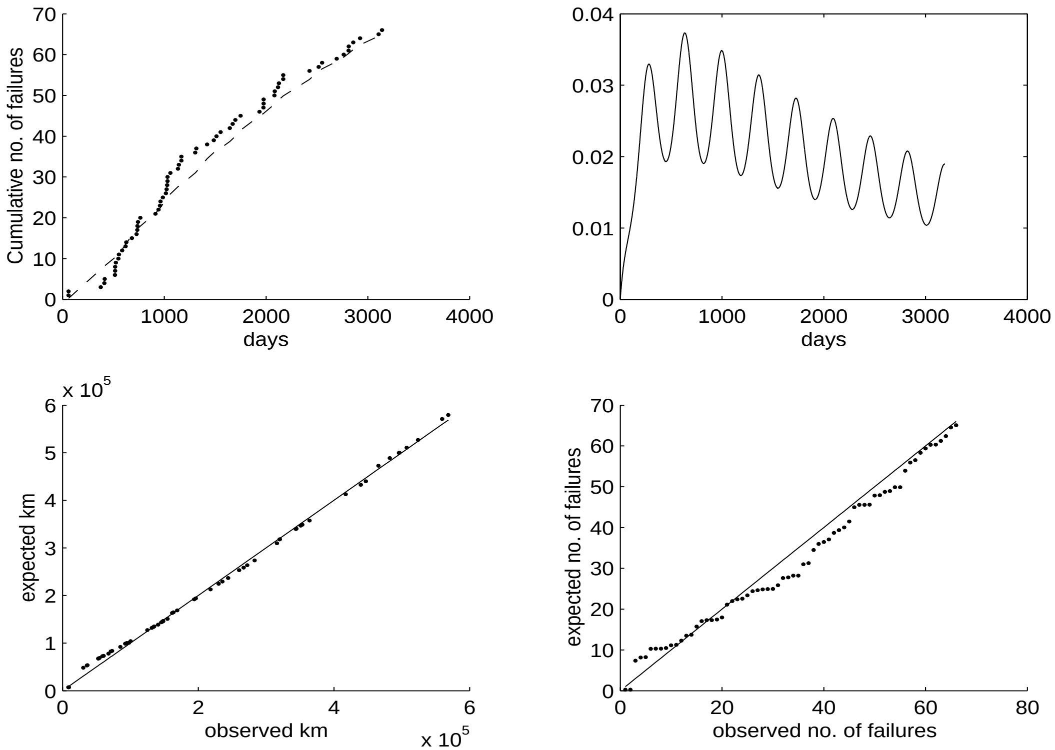
Given $\widehat{\theta} = (\widehat{a}, \widehat{b}, \widehat{\sigma^2}, \widehat{\rho}, \widehat{\beta}_0, \widehat{\beta}_1)$

$\Rightarrow$ need to check if the model is *good*

$\Rightarrow$ plots and tests

- Estimated m.v.f. $\Lambda(\cdot; \widehat{\theta})$ vs. cumulative number of observed failures
- Estimated intensity function $\lambda(\cdot; \widehat{\theta})$
- Expected ($\widehat{a}t_i + \widehat{b}t_i^2$) vs. observed odometer readings at failure times t_i
- Expected ($\Lambda(t_i; \widehat{\theta})$) vs. observed ($i$) number of failures

DIAGNOSTIC PLOTS FOR ONE TRAIN



Estimated m.v.f. vs. observed failures (top left), estimated intensity function (top right), expected vs. observed odometer readings at failure times (bottom left) and expected vs. observed number of failures (bottom right)

DIAGNOSTIC FOR ONE TRAIN

Theorem 1 *Let* $\Lambda(t)$ *be a continuous nondecreasing function. Then* $T_1, T_2, \ldots$ *are arrival times in a Poisson process* N_t *with m.v.f.* $\Lambda(t)$ *if and only if* $\Lambda(T_1), \Lambda(T_2), \ldots$ *are arrival times in an HPP* H_t *with failure rate one.*

- $\widehat{\Lambda}(t)$ estimated from data $T_1, T_2, \ldots$
- *Suppose* $T_1, T_2, \ldots$ from NHPP with m.v.f. $\widehat{\Lambda}(t)$
- $Y_1 = \widehat{\Lambda}(T_1), Y_2 = \widehat{\Lambda}(T_2), \ldots$ *data* from HPP with rate 1
- Interarrival times $X_i = Y_i - Y_{i-1}$ i.i.d. $\mathcal{E}(1)$
- $U_i = \exp\{-X_i\}$ i.i.d. $\mathcal{U}[0,1]$
- Should $\widehat{\Lambda}(t)$ be the right model, then U_i's should be uniform r.v.'s
- Kolmogorov-Smirnov test to check if data are coming from uniform distribution
- **Unsatisfactory results**

A BAYESIAN MODEL

- Interest in
  - checking if trains fulfill reliability requirements before warranty expiration
  - mathematical model able to predict failures based upon current failure data and knowledge
- $\Rightarrow$ a (more complex) Bayesian model
  - first 2 years of data used to estimate parameters
  - number of failures predicted in the following 1, 2, 3, 4, 5 years (for which observed data are available)
  - compute $\mathcal{E}(N(2, 2+i)|N(0, 2)) = \int \Lambda((2, 2+i)|\theta)\pi(\theta|N(0, 2))d\theta$, with 95% credible interval (from simulations), for $i = 1, 5$
  - comparison between predicted and actual observed failure data (cumulative number)
  - *good forecast*

HIERARCHICAL MODEL

- Hierarchical model with $g(t)$ realisation of a Gamma process

$$
\begin{aligned}
g(t) &\sim \mathcal{G}(at, b) \\
\theta &\sim \pi(\theta) \\
[\mathbf{t} \mid \mathbf{g}, \theta] &= NHPP\{\mu(g(t); \theta_1) s(t; \theta_2)\} \\
[\mathbf{k} \mid \mathbf{t}, \mathbf{g}] &= \prod_{i=1}^{n} \delta_{g(t_i)}(\cdot)
\end{aligned}
$$

- MCMC algorithm (Gibbs sampling with Metropolis steps within)

GENERATION OF g

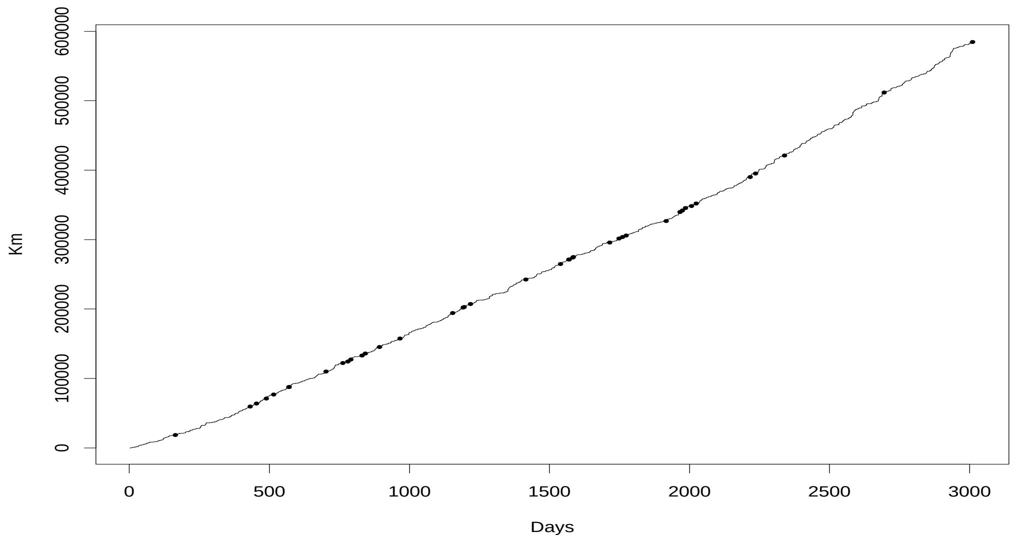
- g updated with an acceptance/rejection step

- g needs to go through observed failure data $k_i = g(t_i)$

- link between Dirichlet and Gamma distributions

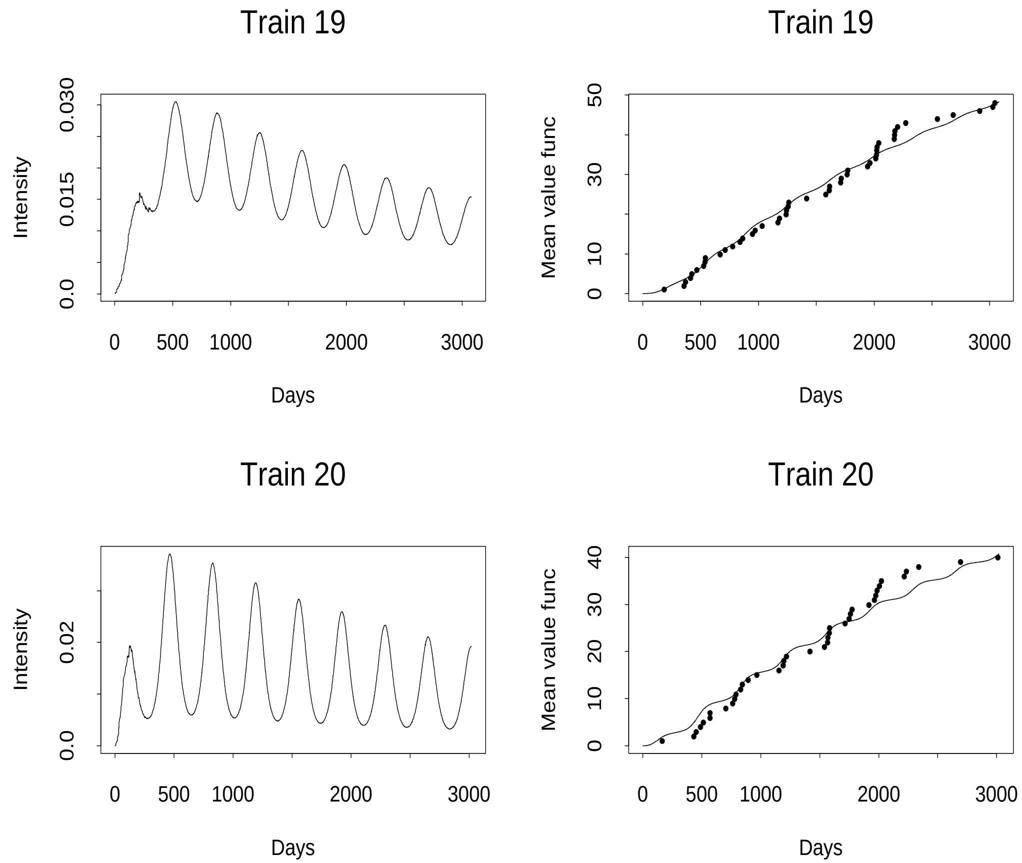
- $g(t)$ points drawn from the cumulative distribution of a Dirichlet process, multiplied by $g(t_i) - g(t_{i-1})$ and shifted above by $g(t_{i-1})$

GENERATION OF g

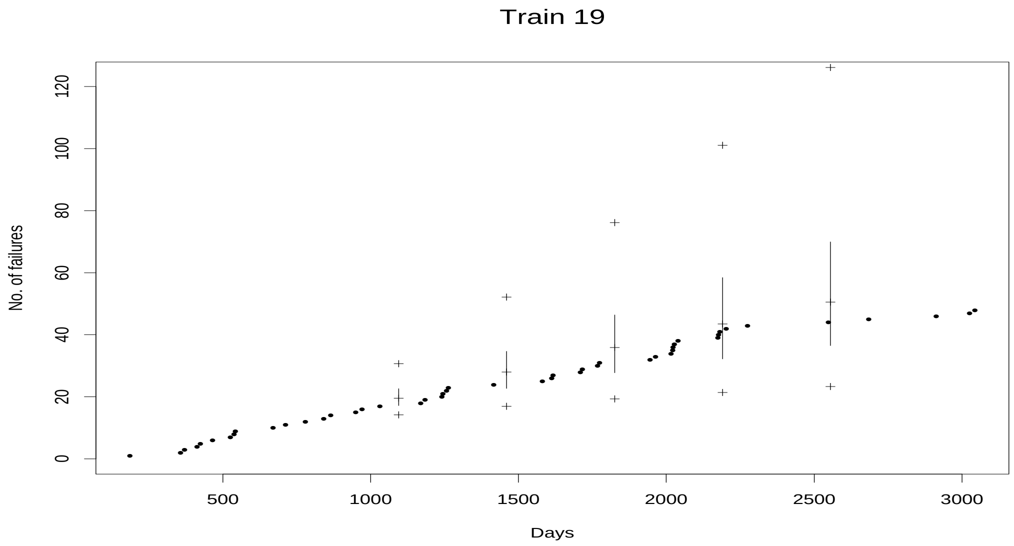


An example of g during the MCMC run

INTENSITY AND MEAN VALUE FUNCTION ESTIMATION



FORECAST



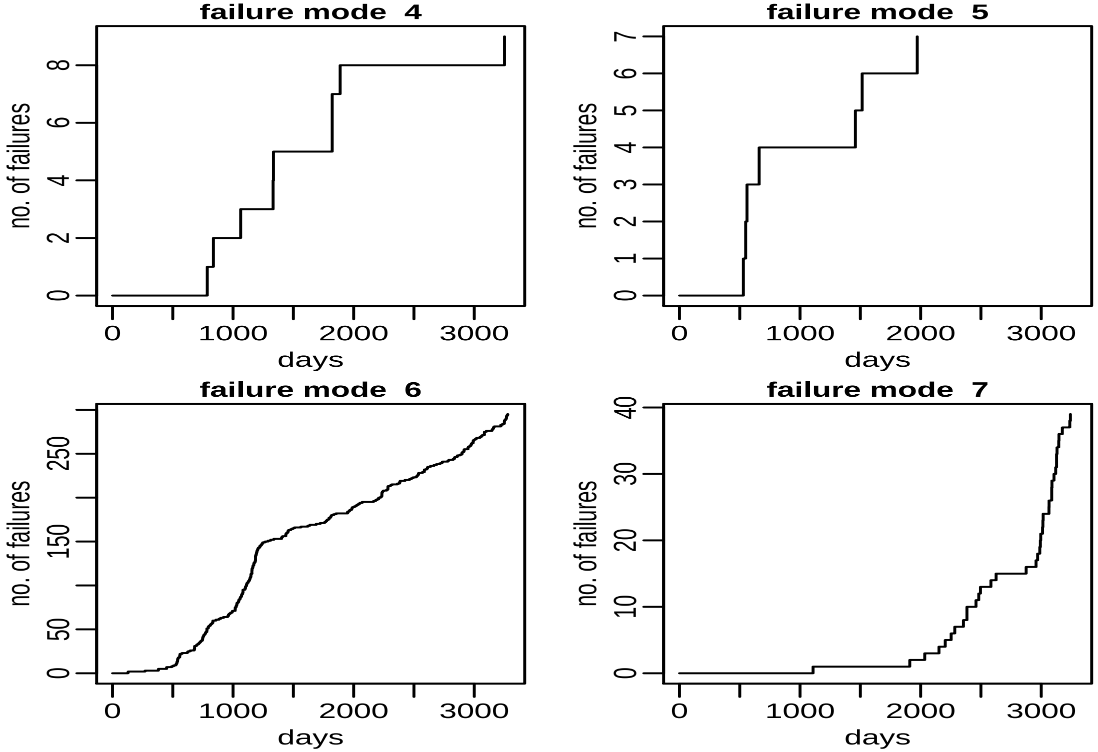
Prediction intervals of the number of failures for train 19 using 730 days (2 years) of observations, up to 5 years ahead. The vertical lines are the interquartile intervals with the posterior median; the plus signs are the extremes of 95% posterior probability intervals

DIFFERENT FAILURE MODES MODEL

Code	Subsystem	No. of parts	Total failures
1	opening commands (electrical)	14	530
2	cables and clamps	4	33
3	mechanical parts	67	1182
4	electrical protections	12	9
5	power supply circuit	2	7
6	pneumatic gear	31	295
7	electro-valves	8	39

Classification of failure modes and total failures per mode for all trains in nine years

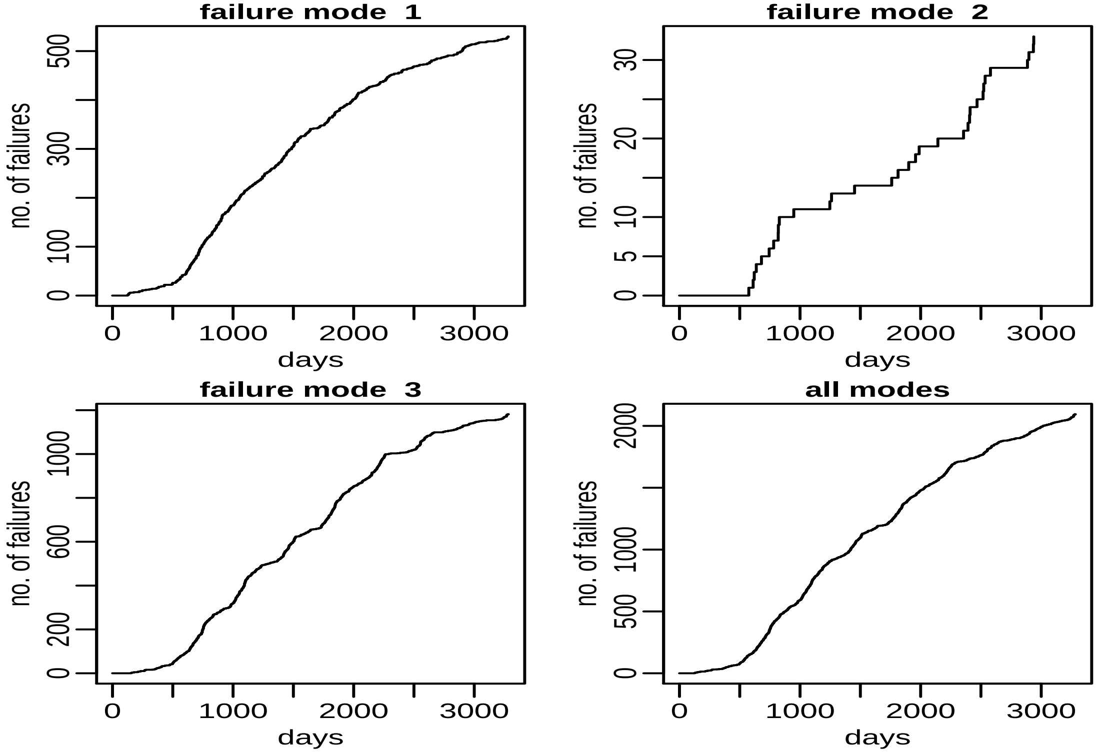
CUMULATIVE NUMBER OF FAILURES



- Failure modes 4 and 5 very rare ⇒ not enough information for fitting a stochastic process model

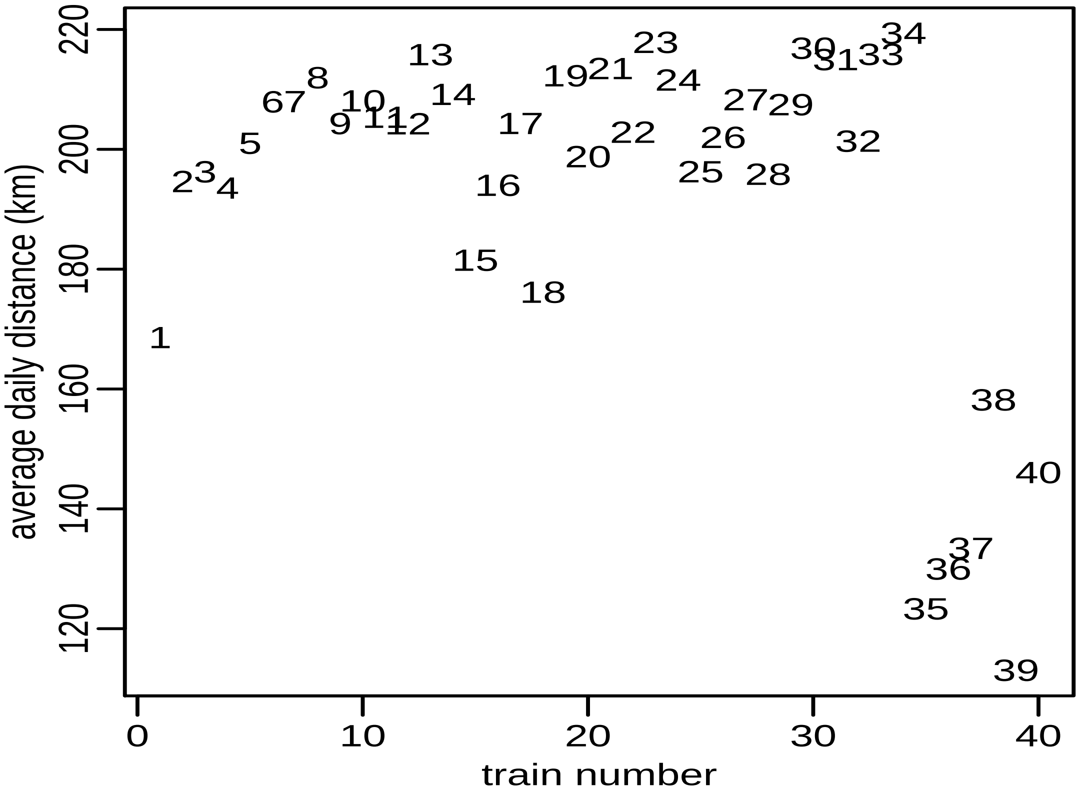
- Failure modes 6 and 7 show change-points

CUMULATIVE NUMBER OF FAILURES



- Failure modes 1, 2 and 3 display a more regular pattern
- Mode 2 failures are only 0.11 per train and per year
- $\Rightarrow$ concentrate on failure modes 1 and 3

AVERAGE DAILY DISTANCE



- Different average daily distance
- More recent trains are used less daily

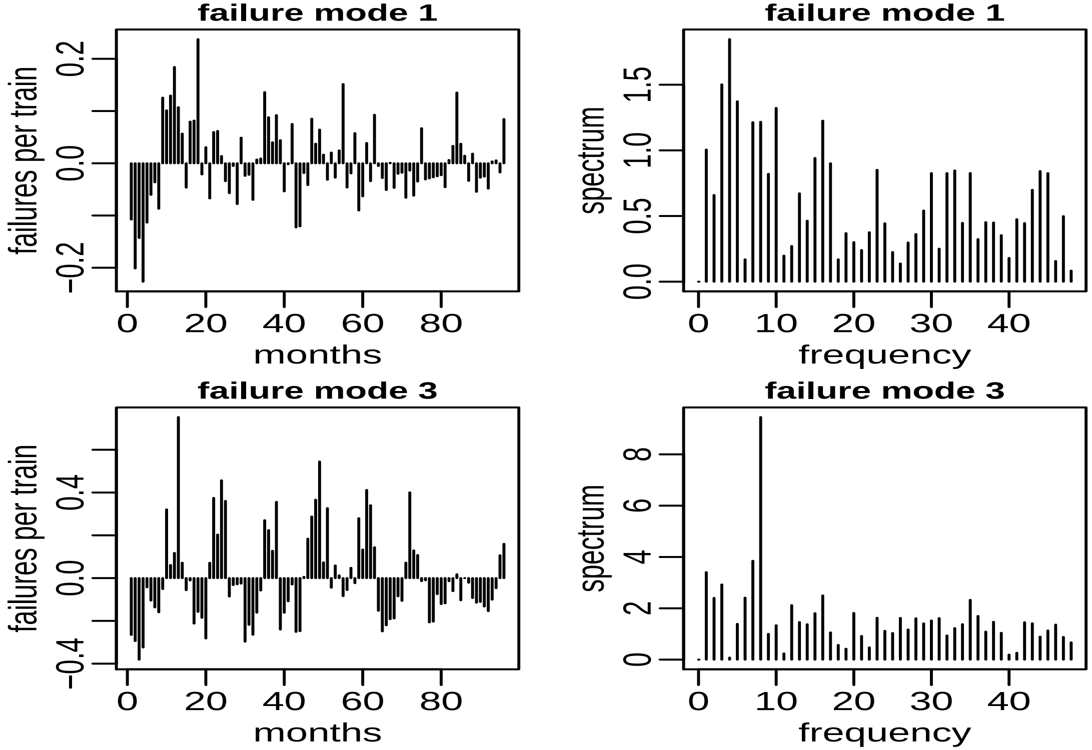
BIVARIATE INTENSITY FUNCTION

For each train i

$$\begin{aligned}\lambda_i(t,s) &= \mu \exp\left\{-\gamma(s-a_i-c_i(t-t_{0i}))^2 w(t-t_{0i})\right\} \cdot \\ &\cdot \exp\left\{A\cos(\omega(t-d))\right\} \lambda_0(t-t_{0i})\end{aligned}$$

- t_{0i} starting operation date
- $a_i + c_i(t-t_{0i})$ expected distance after $(t-t_{0i})$ days in service ((a_i, c_i) different for every train, as seen before)
- $w(\cdot)$ positive weight function, rather close to 0 for $(t-t_{0i}) \approx 0$ and to 1 for $(t-t_{0i})$ large (initial relation between distance and time not linear)
  e.g. $w(z) = \dfrac{\sqrt{1+z}}{1+\sqrt{1+z}}$, bounded between 0.5 and 1
- $\lambda_0(\cdot)$ is a baseline intensity function (depending on time since first ride), common to all trains except for starting point
- exponentiated cosine is a periodic component with phase d (depending on calendar time), common to all trains

PERIODIC COMPONENT



- Periodogram of monthly time series of failure modes 1 and 3 (after detrending)
- No clear frequency for failure mode 1 $\Rightarrow$ omit periodic component in intensity
- 12-month cycle evident for failure mode 3

BASELINE INTENSITY

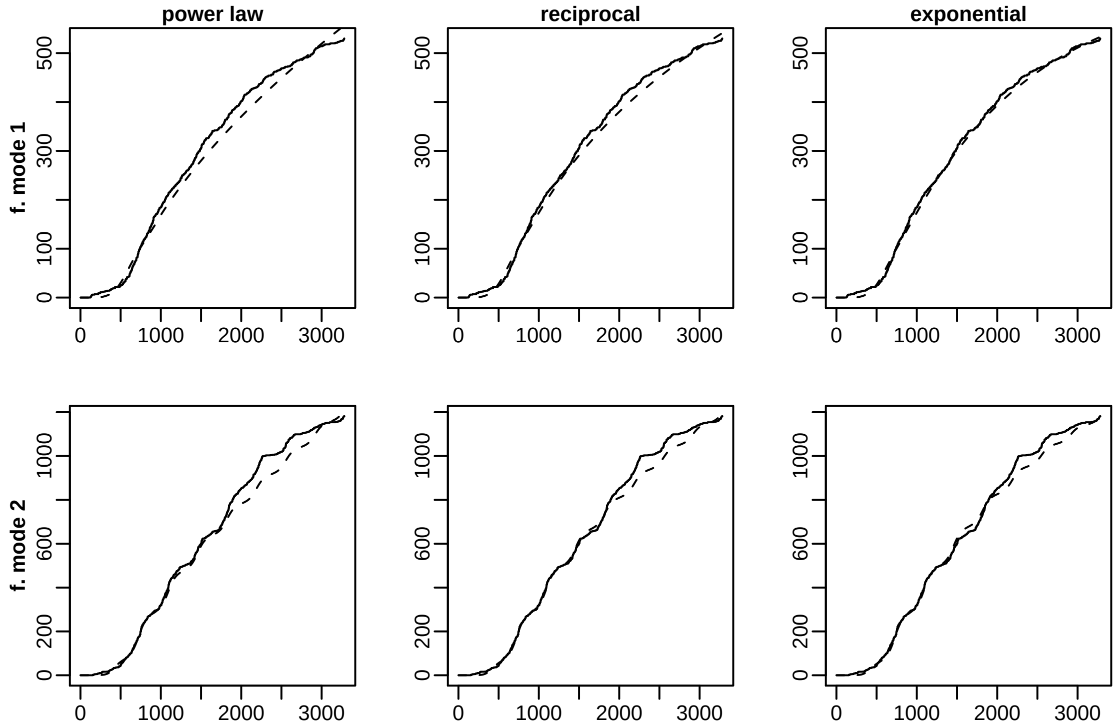
- $\Lambda_0(u) = Mu^b$ (Power Law process)
- $\Lambda_0(u) = \ln(1 + bu)$ (Reciprocal)
- $\Lambda_0(u) = (1 - e^{-bt})/b$ (Exponential)

We omit writing likelihood, priors, posterior conditionals and MCMC implementation

ESTIMATE OF MEAN VALUE FUNCTION

- Posterior mean of $\Lambda(t;\theta)$
  - correct one
  - requires numerical integration of $\lambda(t;\theta)$ at each MCMC step
- Plot of $\Lambda(t;\widehat{\theta}) = \sum_{i=1}^{40} \int_{t_{0i}}^{t} \lambda_i(u;\widehat{\theta})\,du, \qquad t = 1,\ldots,3287$
  - $\widehat{\theta}$ estimate of θ from MCMC run
  - $\lambda_i(t) = \mu\sqrt{\frac{\pi}{\gamma w(t-t_{0i})}}\,\Phi\left\{(a_i + c_i(t-t_{0i}))\sqrt{2\gamma w(t-t_{0i})}\right\}\,\cdot$
    $\cdot\,\exp\{A\cos(\omega\,(t-d))\}\;\lambda_0(t-t_{0i})$
    (marginal of $\lambda_i(t,s)$)
  - not optimal but useful

ESTIMATE OF MEAN VALUE FUNCTION



- Cumulative number of failures for all trains and estimated mean value function (dashed)
- Row 1: failure mode 1; Row 2: failure mode 3
- Each column is for a different baseline (exponential in third column is the best)

FORECAST OF FUTURE FAILURES OF GIVEN MODE

- D_{T_0} data available at day T_0

- $\pi(\cdot \mid D_{T_0})$ posterior density of θ

Predictive distribution

$$P(N_{T_0+u} - N_{T_0} = x \mid D_{T_0}) = \int e^{-\{\Lambda(T_0+u;\theta)-\Lambda(T_0;\theta)\}} \frac{\{\Lambda(T_0+u;\theta) - \Lambda(T_0;\theta)\}^x}{x!} \pi(\theta \mid D_{T_0})\, d\theta$$

Expected value

$$\mathsf{E}(N_{T_0+u} - N_{T_0} \mid D_{T_0}) = \int \{\Lambda(T_0+u;\theta) - \Lambda(T_0;\theta)\}\, \pi(\theta \mid D_{T_0})\, d\theta$$

FORECAST OF FUTURE FAILURES OF MODE 1

end of recording period	forecasting horizon (years)	95% credibility interval	true value	posterior mean
1992	1	(86, 143)	83	114
	2	(79, 140)	72	109
	3	(71, 138)	62	105
1993	1	(69, 124)	72	97
	2	(59, 121)	62	90
	3	(50, 119)	42	85
1994	1	(50, 100)	62	74
	2	(41, 95)	42	66
	3	(32, 91)	35	59
1995	1	(38, 81)	42	59
	2	(30, 74)	35	51
	3	(24, 68)	23	44
1996	1	(27, 60)	35	43
	2	(20, 52)	23	35
1997	1	(19, 46)	23	39

FAILURE FORECAST OF NEW TRAIN

- $N_H(t)$ failure Poisson process for new train
- $\lambda_H(t;\theta)$ intensity function and $\Lambda_H(t;\theta)$ mean value function
- D_t failure data up to time t
- $T_0 = 2$ years

$$\Pr(N_H(T_0) > x_U \mid D_t) = 1 - \int \sum_{x=0}^{x_U} e^{-\Lambda_H(T_0;\theta)} \frac{[\Lambda_H(T_0;\theta)]^x}{x!} \pi(\theta \mid D_t)\, d\theta$$

f. mode 1	x_U	3	4	5	6	7	8	9	10	11	12	13
	prob.	0.82	0.68	0.52	0.36	0.23	0.14	0.07	0.04	0.02	0.01	0.00
f. mode 3	x_U	12	13	14	15	16	17	18	19	20	21	22
	prob.	0.47	0.36	0.26	0.18	0.12	0.08	0.05	0.03	0.01	0.01	0.00

CASE STUDY: SOFTWARE RELIABILITY

Software reliability can be defined as *the probability of failure-free operation of a computer code for a specified mission time in a specified input environment*

According to the Software Engineering Institute, even experienced programmers inject about one defect into every 10 lines of code. A laudable standard which is aspired to by many modern manufacturers of commercial software is what is referred to as the **Five 9s** - software which works 99.999% of the time

SOFTWARE RELIABILITY: ENVIRONMENTS

Techniques to achieve reliable software systems, aimed at

- Fault prevention
- Fault removal
- Fault tolerance (i.e. providing service despite of faults)
- Fault forecasting *(room for statistics ...)*

Stages, including

- Testing (globally and/or parts of codes)
- Operation
- Debugging

SOFTWARE RELIABILITY: ENVIRONMENTS

Software

- Desktop computing
- Client/server computing
- Web-deployed applications
- .net enterprise (e.g. banking on line)

Intervention

- Perfect repair
- Imperfect repair
- **Bugs introduction**

Different environments, different models

SOFTWARE RELIABILITY: MODELS

Seminal paper by Jelinski and Miranda (1972)

More than 100 models after it (Philip Boland, *MMR2002*)

Many models clustered in few classes

Search for unifying models (e.g. Self-exciting process, Chen and Singpurwalla, 1997)

SOFTWARE RELIABILITY: MODELS

Most software reliability models fall into one of two categories (Singpurwalla and Wilson, 1994)

- [Type I]: models on times between successive failures based on
  - [Type I-1] failure rates (e.g. Jelinski-Moranda)
  - [Type I-2] inter-failure times as function of previous inter-failure times (e.g. random coefficient autoregressive model, Singpurwalla and Soyer, 1985)

- [Type II] models (counting processes) on observed number of failures up to time t (e.g. NHPP)

SOFTWARE RELIABILITY: MODELS

Failures at $T_1, T_2, \ldots, T_n$

Inter-failure times $T_i - T_{i-1} \sim \mathcal{E}(\lambda_i)$, independent, $i = 1, \ldots, n$

- $\lambda_i = \phi(N - i + 1)$, $\phi \in \mathbb{R}^+, N \in \mathbb{N}$, *(Jelinski-Moranda, 1972)*
  - Program contains an initial number of bugs N
  - Each bug contributes the same amount to the failure rate
  - After each observed failure, a bug is detected and corrected

Straightforward Bayesian inference with priors $N \sim \mathcal{P}(\nu)$ and $\phi \sim \mathcal{G}(\alpha, \beta)$

SOFTWARE RELIABILITY: MODELS

- $\lambda_i = \phi(N - p(i-1))$, $\phi \in \mathbb{R}^+$, $N \in \mathbb{N}$, $p \in [0,1]$, *(Goel and Okumoto, 1978)*
  - Imperfect debugging
- $\lambda_i = \phi\delta^i$, $\phi \in \mathbb{R}^+$, $\delta \in (0,1)$, *(Moranda, 1975)*
  - Failure rate (geometrically) decreasing

Failure rate constant between failures; different from

- $h(t_i) = \dfrac{\alpha}{\beta_0 + \beta_1 i + t_i}$ *(Littlewood and Verall, 1973)*
- $h(t_i) = (N - i + 1)\phi t_i$ *(Schick and Wolverton, 1973)*

FAILURE RATES

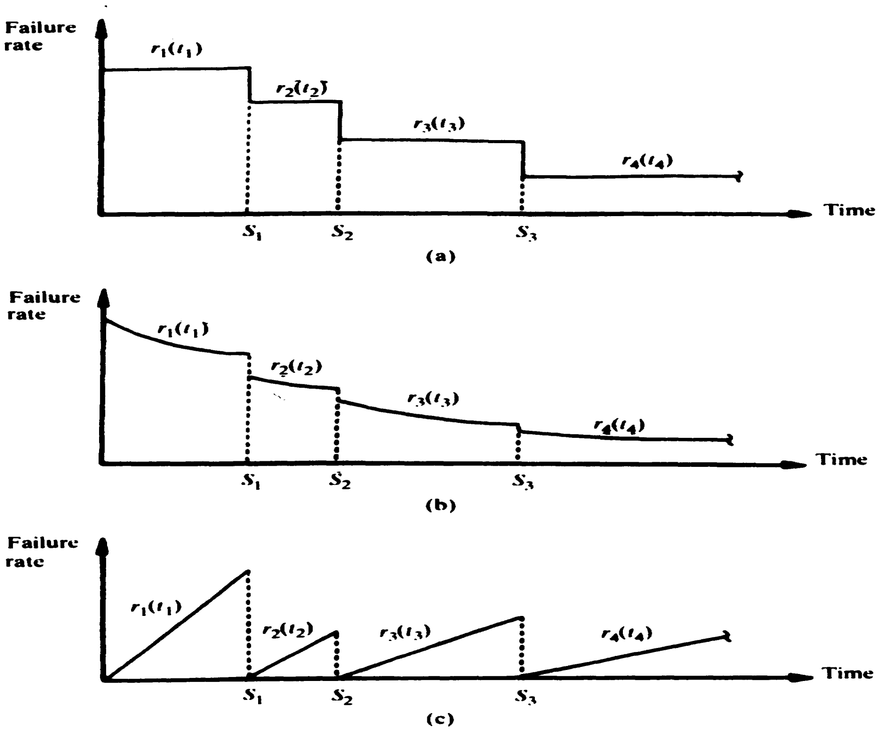


Figure 2.1. (a) The failure rates of the model of Jelinski and Moranda. (b) The failure rates of the model of Littlewood and Verall. (c) The failure rates of the model of Schick and Wolverton

Chen and Singpurwalla, Adv. Appl. Prob., 1997

SOFTWARE RELIABILITY: MODELS

Random coefficient autoregressive model
(Singpurwalla and Soyer, 1985)

- T_i interfailure times and $Y_i = \log T_i$, $i = 1, n$
- $Y_i = \theta_i Y_{i-1} + \epsilon_i, i = 1, n$
- $\epsilon_1, \ldots, \epsilon_n \sim \mathcal{N}(0, \sigma_1^2)$, i.i.d.
- $\theta_1, \ldots, \theta_n \sim \mathcal{N}(\lambda, \sigma_2^2)$, i.i.d.
- $\lambda \sim \mathcal{N}(\mu, \sigma_3^2)$

SOFTWARE RELIABILITY: MODELS

Martingale processes
(Basu and Ebrahimi, 2003)

Interfailure times $T_i \sim \mathcal{E}(\lambda_i)$, $i = 1, n$, conditionally independent given $\lambda_1, \ldots, \lambda_n$

- $\lambda_1 \sim \mathcal{G}(\alpha, \beta_1)$ and $\lambda_i | \lambda_{(-1)} \sim \mathcal{G}(\alpha, \alpha/\lambda_{i-1}), i > 1,$
  $\Rightarrow E(\lambda_i | \lambda_{i-1}) = \lambda_{i-1}$

- $\lambda_1 \sim \mathcal{G}(\alpha_1, \beta_1)$ and $\lambda_i | \lambda_{(-i)} \sim \mathcal{G}(\tau\lambda_{i-1}^2, \tau\lambda_{i-1}), i > 1,$
  $\Rightarrow E(\lambda_i | \lambda_{i-1}) = \lambda_{i-1}$

SOFTWARE RELIABILITY

- Bugs in software induce failures
- Fixing current bugs sometimes implies introduction of new bugs
- Lack of knowledge about effects of bugs fixing
- $\Rightarrow$ need for models allowing for possible, unobserved introduction of new bugs in a context aimed to reduce bugs
- Software affects our life at a larger extent and its malfunctioning could be very harmful
- Goal: Detecting bad fixing of bugs and reliability level

BUGS INTRODUCTION: MODELS

Failures at $T_1, T_2, \ldots, T_n$

Inter-failure times $T_i - T_{i-1} \sim \mathcal{E}(\lambda_i)$, independent, $i = 1, \ldots, n$

- $\lambda_{i+1} = \lambda_i e^{-\theta_i}$, $\quad \lambda_i, \theta_i \in \mathbb{R}^+$, independent
  (Gaudoin, Lavergne and Soler, 1994)
  - $\theta_i = 0 \Rightarrow$ no debugging effect
  - $\theta_i > 0 \Rightarrow$ good quality debugging
  - $\theta_i < 0 \Rightarrow$ **bad quality debugging**

BUGS INTRODUCTION: MODELS

- $\lambda_{i+1} = (1 - \alpha_i - \beta_i)\lambda_i + \mu\beta_i, \quad \lambda_i, \mu \in \mathbb{R}^+$, *(Gaudoin, 1999)*
  - Imperfect debugging
  - α_i good debugging rate
  - β_i **bad debugging rate**

BUGS INTRODUCTION: MODELS

Birth-death process *(Kremer, 1983)*

- $p_n(t) = \mathcal{P}r\{X(t) = n\}$
- $\nu(t)$ **birth rate**
- $\mu(t)$ death rate
- a initial population

$$p_n^{'}(t) = (n-1)\nu(t)p_{n-1}(t) - n[\nu(t)+\mu(t)]p_n(t) + (n+1)\mu(t)p_{n+1}(t), n \geq 0$$

with $p_{-1} \equiv 0$ and $p_n(0) = \mathbb{1}_{\{n=a\}}$

HIDDEN MARKOV MODEL

- Failure times $t_1 < t_2 < \ldots < t_n$ in $(0, y]$
- Y_t latent process describing *reliability status* of software at time t (e.g. growing, decreasing and constant)
- Y_t changing only after a failure $\Rightarrow Y_t = Y_m$ for $t \in (t_{m-1}, t_m]$, $m = 1, \ldots, n+1$, with $t_0 = 0$, $t_{n+1} = y$ and $Y_{t_0} = Y_0$
- $\{Y_n\}_{n \in \mathbb{N}}$ Markov chain with
  - discrete state space E
  - transition matrix $\mathbb{P}$ with rows $\mathbb{P}_i = (P_{i1}, \ldots, P_{ik})$, $i = 1, \ldots, k$

HIDDEN MARKOV MODEL

- Interarrival times of m-th failure $X_m|Y_m = i \sim \mathcal{E}(\lambda(i)), i = 1, \ldots, k, m = 1, \ldots, n$
- X_m's independent given $Y \Rightarrow f(X_1, \ldots, X_n|Y) = \prod_{m=1}^{n} f(X_m|Y)$
- $\mathbb{P}_i \sim \mathcal{D}ir(\alpha_{i1}, \ldots, \alpha_{ik}), \forall i \in E$, i.e. $\pi(\mathbb{P}_i) \propto \prod_{j=1}^{k} P_{ij}^{\alpha_{ij}-1}$
- Independent $\lambda(i) \sim \mathcal{G}(a(i), b(i)), \forall i \in E$
- Interest in posterior distribution of $\Theta = (\lambda^{(k)}, \mathbb{P}, Y^{(n)})$
  - $\lambda^{(k)} = (\lambda(1), \ldots, \lambda(k))$
  - $Y^{(n)} = (Y_1, \ldots, Y_n)$

ORDERING OF STATES

- Independent λ's $\Rightarrow$ no ordering among states

- No 0 in transition matrix $\Rightarrow$ jumps possible from any state to any state

$\Rightarrow$ difficult ranking of states in terms of reliability

- prior on ordered λ's $\Rightarrow$ identification of different levels of reliability

- Bi-(or tri-) diagonal transition matrix allowing only jumps into the nearest best (nearest best and worst) state

EXTENSION: PRIOR ON ORDERED λ'S

- $X \sim \mathcal{G}(\alpha, \beta) \perp (Y_1, \ldots, Y_m) \sim \mathcal{D}ir(a_1, \ldots, a_m) : \sum_{i=1}^m a_i = \alpha$

- Take $(\lambda_1, \ldots, \lambda_m) : \lambda_m = X$ & $\lambda_j = X \sum_{i=1}^j Y_j, j = 1, m-1$

- $\Rightarrow X = \lambda_m$ & $Y_j = \frac{\lambda_j - \lambda_{j-1}}{\lambda_m}, j = 1, m-1$ (with $\lambda_0 = 0$)

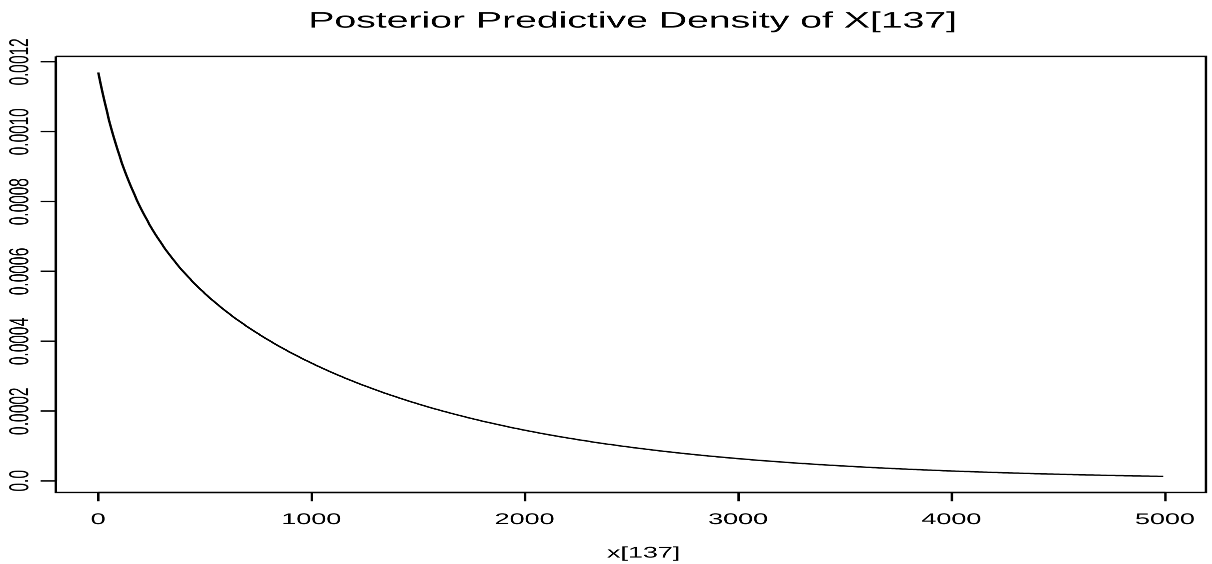
- $f(\lambda_1, \ldots, \lambda_m) = \beta^\alpha e^{-\beta\lambda_m} \prod_{j=1}^m \frac{(\lambda_j - \lambda_{j-1})^{a_j - 1}}{\Gamma(a_j)} I_{\{\lambda_1 < \lambda_2 < \ldots < \lambda_m\}}$

- $\Rightarrow \lambda_j | \lambda_{(-j)} \sim \mathcal{B}e(a_j, a_{j+1})$ on $(\lambda_{j-1}, \lambda_{j+1}), j < m$

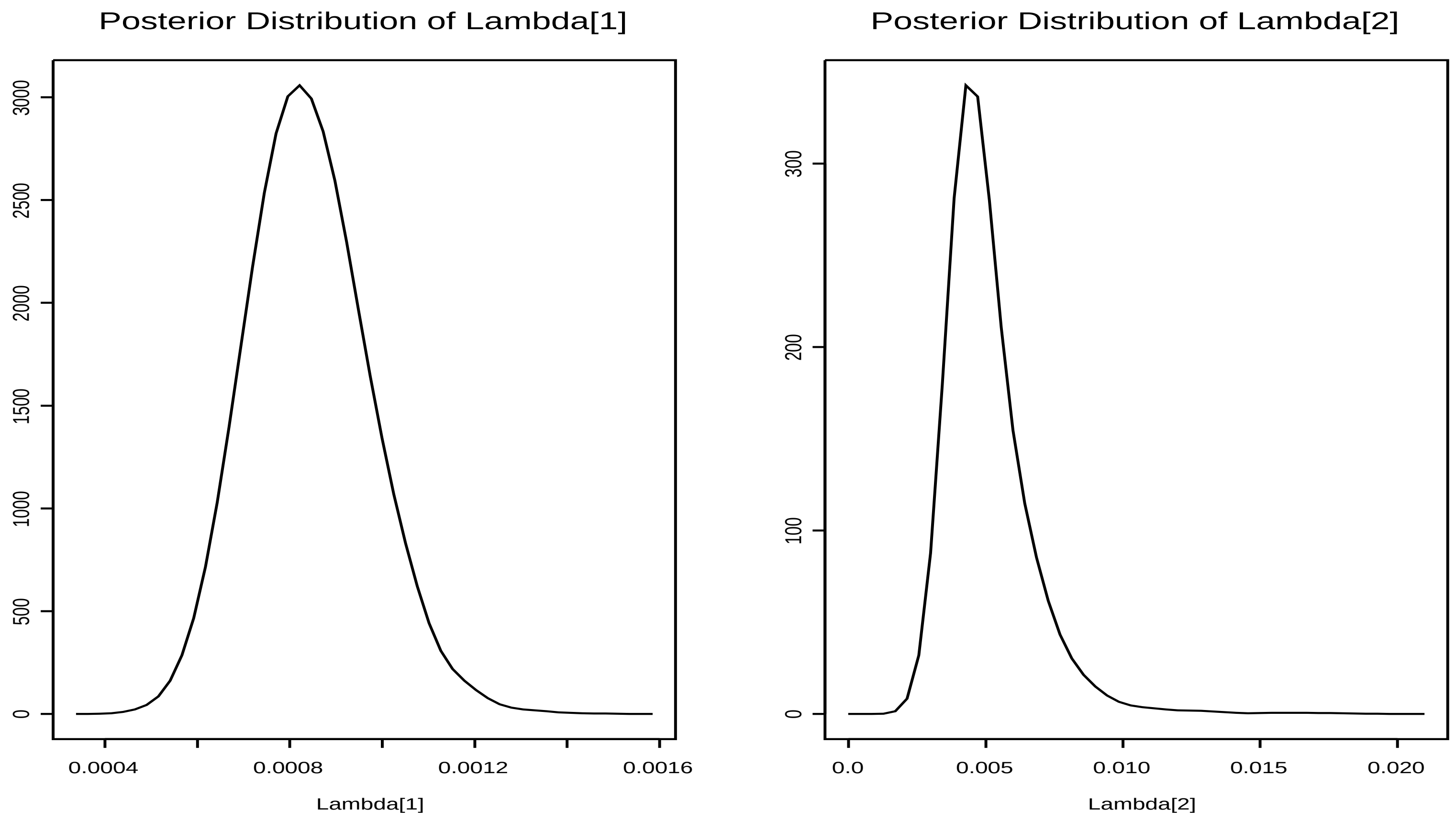
SOFTWARE RELIABILITY - MUSA DATA

- Musa System 1 data: 136 software failure times

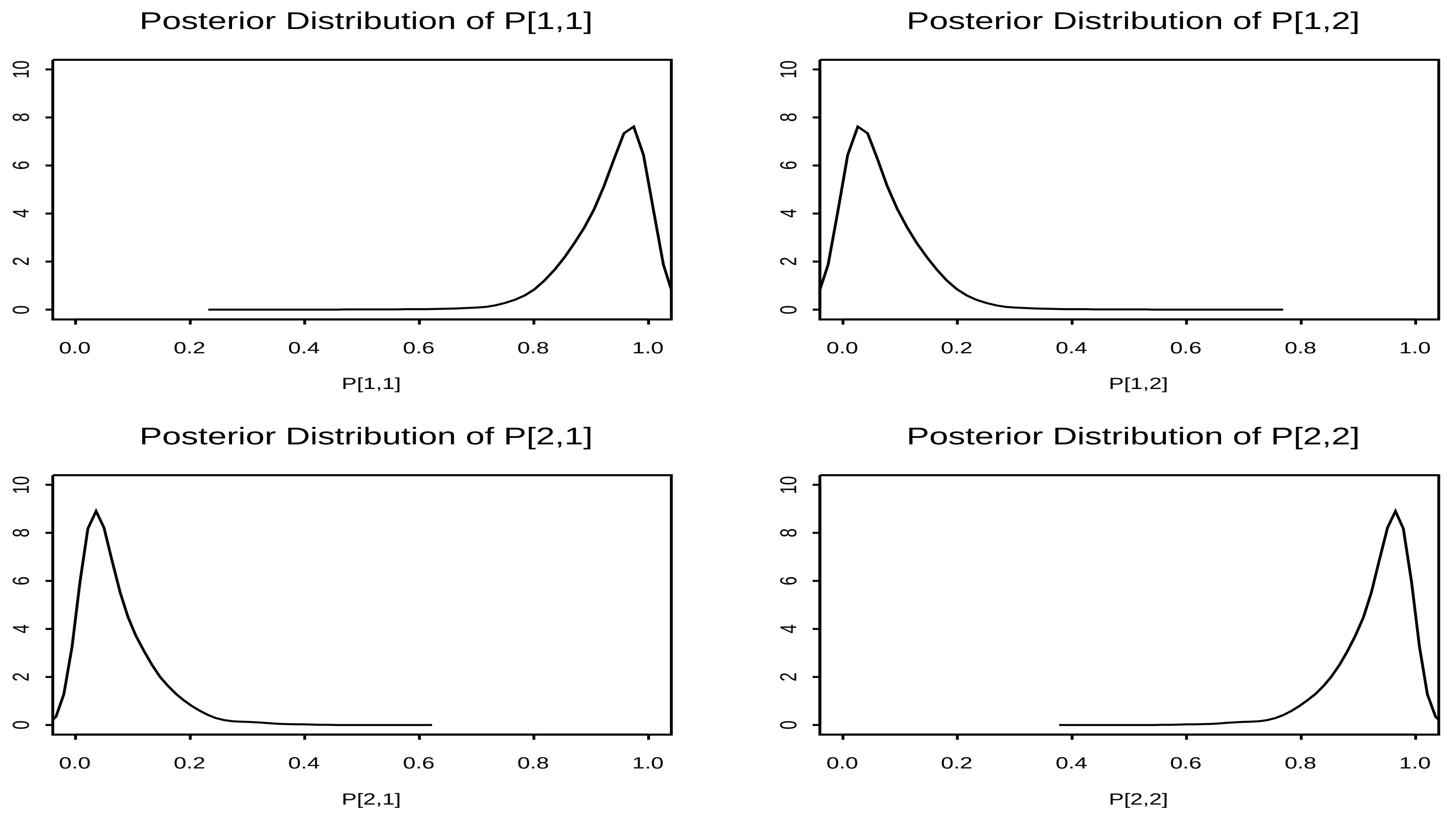
- Hidden Markov model with 2 unknown states



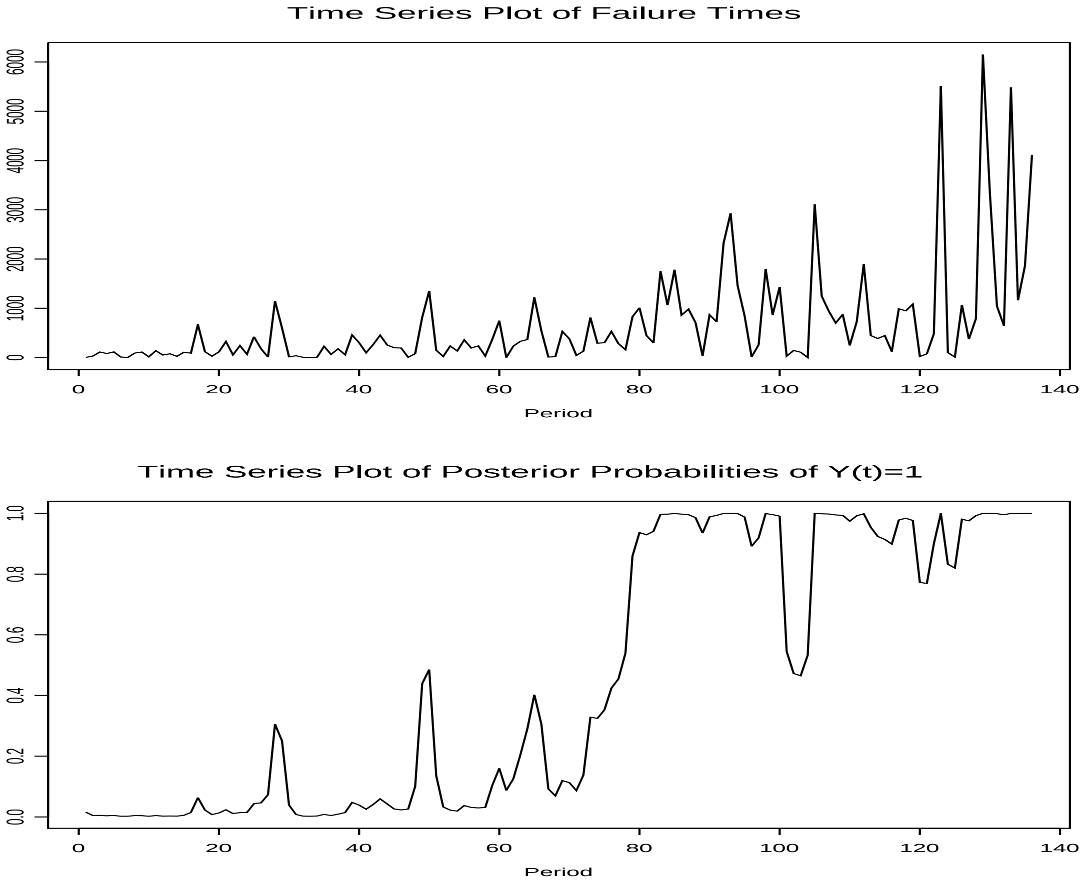
SOFTWARE RELIABILITY - MUSA DATA



SOFTWARE RELIABILITY - MUSA DATA



SOFTWARE RELIABILITY - MUSA DATA



Longer failure times $\Rightarrow$ higher Bayes estimator of probability of ”good” state

SELF-EXCITING PROCESS WITH LATENT VARIABLES

- NHPPs widely used in (software) reliability, characterised by intensity function $\mu(t)$
- Self-exciting processes (SEPs) add extra terms $g(t - t_i)$ to the intensity as a consequence of events at t_i (e.g. introduction of new bugs)
- Binary latent variables modelling the introduction of new bugs
- Interest in
  - detecting possible bugs introduction
  - finding optimal stopping time t^* either as $\{\min t : \lambda(t) \leq \lambda^*\}$ or minimising $\mathcal{E}\{C_T(t) + C_O(N(t, T])\}$
  - $C_T(t) = tC_T$ testing cost
  - $C_O(N(t, T]) = C_O \cdot N(t, T]$ operational failure costs during useful life $(t, T]$

SOFTWARE RELIABILITY

- $\lambda(t) = \mu(t) + \sum_{j=1}^{N(t^-)} Z_j g_j(t - t_j)$
- $\mu(t)$: nonincreasing intensity (*bugs removal)*
- $N(t^-)$: $\#$ failures before t
- $Z_j = 1, 0$ (*introduction of new bug or not*)
- $g_j(u) \geq 0$ for $u > 0$ and 0 for $u \leq 0$
- Likelihood $L(\theta; \underline{t}, \underline{Z}) = f(\underline{t}|\underline{Z}, \theta) f(\underline{Z}|\theta)$
  - $f(\underline{t}|\underline{Z}, \theta) = \prod_{i=1}^{n} \lambda(t_i) e^{-\int_0^T \lambda(t) dt}$
  - Bernoulli *priors* on Z_j's updated by data
- $P(Z_j = 1|\underline{t})$, etc.

CASE STUDY 4: CYLINDER LINERS WEAR

- Grimaldi Group is one of the largest ship companies in Italy (and, probably, in Europe), involved in both freight and passengers transportation and operating more than 60 ships, mostly in Europe but also overseas
- Interest in preventive and optimal maintenance of ships
- Earlier models studied by engineers and a physicist in Napoli
- Researchers from Milano involved recently on (more sophisticated?) mathematical models from a Bayesian perspective

PROBLEM DESCRIPTION

- Marine diesel engines are obliged to have high levels of reliability and availability to meet operating requirements

- A costly maintenance programme (including inspection, repair and replacement) is required to satisfy these requirements

- ”Condition based” maintenance, based on identifiable warnings of the onset of a failure, is one of the most effective policies

- In this context, preventive replacement of components is carried out on the basis of their current conditions, rather than on accumulated operating time, i.e. only when actually required, reducing costs and increasing system availability

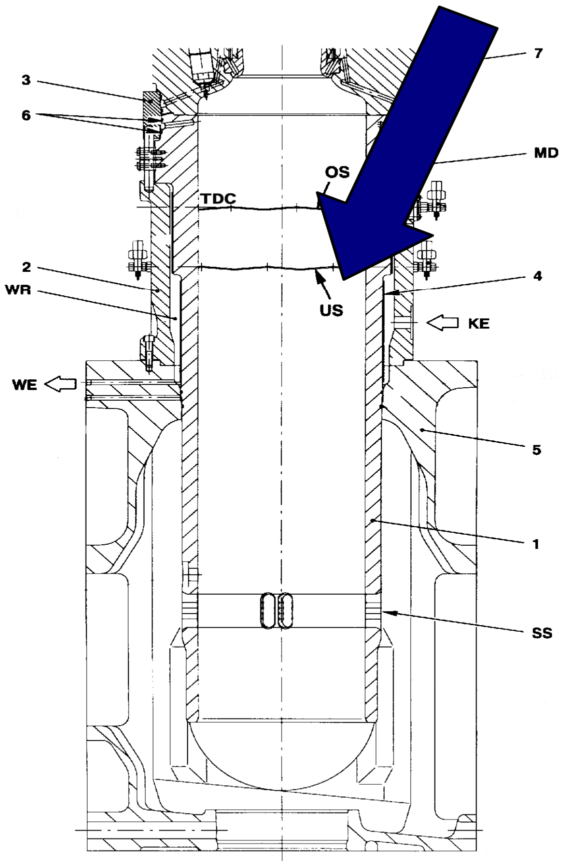
PROBLEM DESCRIPTION

- Wearing of cylinder liners is a major factor in causing failures in heavy-duty diesel engines

- Liner replacement is generally carried out when the maximum wear of the internal surface approaches a threshold (4mm in our case) imposed by warranty clauses; note that the liner walls are 100mm thick

- The largest wear is usually achieved at the Top Dead Center of the liner, which is subject to high thermomechanical and tribological (i.e. related to interactions between surfaces in relative motion) stresses which produce relevant early local damages

PROBLEM DESCRIPTION



1	Cylinder Liner
2	Supporting ring
3	Water guide jacket
4	Insulating band
5	Cylinder jacket
6	Cylinder cover
7	Cylinder head
KE	Cooling water inlet
KB	Cooling bores in the cylinder liner
OS	Upper row of lubricating grooves
US	Lower row of lubricating grooves
SS	Scavenge ports
WR	Water space
WE	Water space drain
TDC	Top Dead Center
MD	Metallic sealing

PROBLEM DESCRIPTION

- Major wear is due to the high quantity of abrasive particles on the piston surface, produced from the combustion of heavy fuels and oil degradation (soot)

- If the lubricant film along the walls of the liner is less thick than the soot particle, then the soot can cause abrasion on the liner metal surface (soot particles are harder than engine parts)

- Besides abrasive wearing, there is a corrosive wearing due to sulphuric acid, nitrons/nitric acids and water

GOALS OF THE RESEARCH

The company's viewpoint

- Forecast the behaviour of the wear process

- Develop a parsimonious, but efficient, inspection policy which increases ships availability and prevents exceedance of the maximum allowed wear (costs due to any failure occurred with a wear under such threshold are charged to the liner' constructor, whereas the naval company pays o.w.)

GOALS OF THE RESEARCH

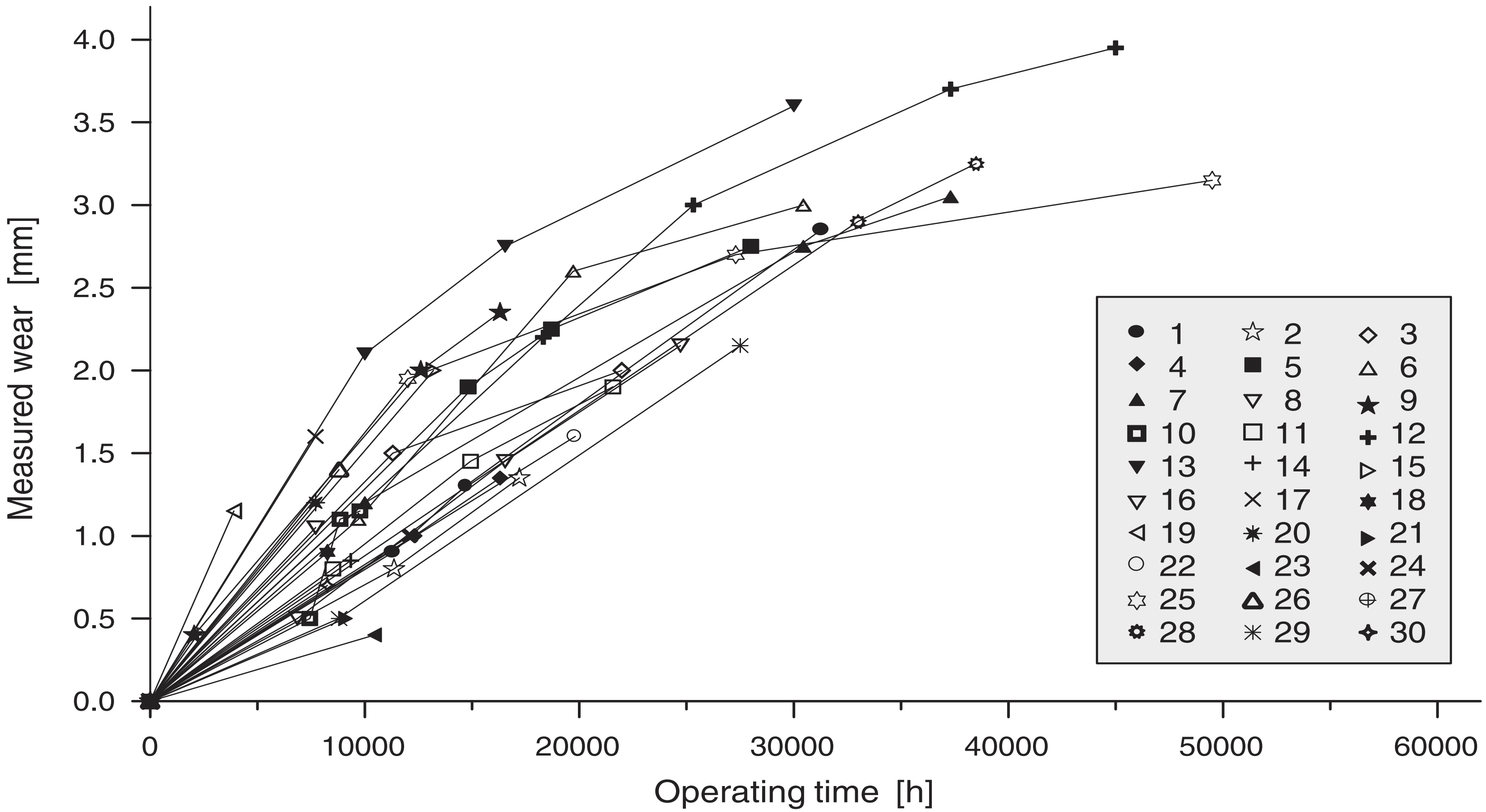
The researchers' viewpoint

- Introduce models capable of describing the wear process and forecasting its behaviour

- Study of optimal maintenance policies, e.g. computing the probability of exceeding the maximum allowed wear before the next inspection is performed

CASE STUDY

- Data about wear in 30 cylinder liners of 8 cylinder SULZER RTA 58 engines, equipping twin ships of the Grimaldi Group
- Data collected after each liner was mounted (dates after 1/1/1996) until 30/12/2004
- Liners were periodically inspected and the accumulated wear was measured by positioning a micrometer inside a planned hole in correspondence of the Top Dead Center (i.e. the point of maximum wear)
- Micrometer accuracy is 0.05mm so that all measures are approximated to the nearest multiple of 0.05mm
- Failed liners are replaced by new ones
- Wear should increase more slowly as accumulated wear increases (see next figure)



Measured wear [mm]
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10000
20000
30000
40000
50000
60000
Operating time [h]
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

COMPETING RISK MODEL

Early work (Bocchetti, Giorgio, Guida and Pulcini, 2006)

- Two possible causes of failure: wear (as before) and thermal crack
- The thermal crack of the liner is due to fatigue cracking caused by repeated occurrences of thermal shocks
- Thermal shocks are caused by changes in the temperature of the cooling fluid on the external surface of the liner, often aggravated by corrosion caused by inadequate chemical treatment of the cooling water
- The changes occur mostly during the maneuver operations when it is difficult to keep the temperature constant
- As a consequence, a small crack can arise in the external surface of the liner and then propagate towards the inside

COMPETING RISK MODEL

- Failure mechanisms are independent since
  - wear is internal to the liner and is not influenced by the thermal shock which is external
  - the point in which maximum wear and crack could occur are quite far apart
  - external temperature is quite uninfluential on the process inside
- Current research concentrates only on wear, which is deemed as the most relevant phenomenon

COMPETING RISK MODEL

- Wearing events, modelled by NHPP $N(t)$, induce an accumulated wear $W(t) = cN(t)$, with c constant to be estimated

- $\lambda(t) = a\,\text{exp}\{-bt\}$, $a, b > 0$, intensity function of $N(t)$, preferred to the Power Law process (PLP) with decreasing $\lambda(t) = M\beta t^{\beta-1}$, $M > 0, 0 < \beta < 1$, to avoid unlimited growth of the wear process

- Weibull distribution used to describe occurrence of thermal cracks (justified by plotting the Kaplan-Meier estimator on Weibull paper)

- Thermal crack dominating cause of failure in the first 5000 hours and then wear dominates

GOALS OF OUR RESEARCH

Our viewpoint

- Describe the wear process and forecast its behaviour
- Introduce models (e.g. diffusion processes) capable of better describing the dynamics of the wear process
- Use available information (in a situation of few data) through a Bayesian approach

PHYSICAL ASSUMPTIONS

- wear increment decreases as a function of wear
- background activity of *tiny* particles and corrosion lead to rather negligible wear increments
- *large* soot particles responsible for the most relevant wear increments

MATHEMATICAL ASSUMPTIONS

- wear $W(t)$ evolves over time and influences wear increment $dW(t)$
  $\Rightarrow$ stochastic differential equation (SDE) $dW(t) = f(W(t))$
- additive accumulation of wear as a sequence of small normally distributed effects (e.g. *tiny* particles and corrosion)
  $\Rightarrow$ possibly a Wiener process with drift
- wear increment jumps because of *large* particles
  $\Rightarrow$ possibly a jump process in the SDE

OUR FIRST MODEL

$$dW(t) = W(t^-)\left\{\mu dt + \sigma dB(t) + dJ(t)\right\}$$

- μ drift, $\mu > 0$, and σ volatility, $\sigma > 0$
- $B(t)$ Wiener process at time t
- $N(t)$ homogeneous Poisson process (HPP) with rate λ, $\lambda > 0$, denoting number of collisions of *large* particles up to time t
- τ_j, $j = 1, \ldots, N(t)$, (unobserved) collision times up to time t
- $Y_j = \frac{W(\tau_j) - W(\tau_j^-)}{W(\tau_j^-)}$, $j = 1, \ldots, N(t)$, jump sizes ($Y_j > 0 \Rightarrow$ upward jumps)
- $J(t) = \sum_{j=1}^{N(t)} Y_j$ jump process

OUR FIRST MODEL

$$dW(t) = W(t^-)\,\{\mu dt + \sigma dB(t) + dJ(t)\}$$

Positive aspects

- closed-form solution

$$W(t) = W(0)\exp\left\{(\mu - \sigma^2/2)t + \sigma B(t)\right\}\prod_{j=1}^{N(t)}(Y_j + 1)$$

- explicit expression for conditional likelihood (upon the number of collisions in each interval between inspection times)

- feasible Bayesian inference, based on MCMC methods

OUR FIRST MODEL

$$dW(t) = W(t^-)\left\{\mu dt + \sigma dB(t) + dJ(t)\right\}$$

Negative aspects

- wear increment proportional to wear, but proper choice of jump process can contrast the direct proportionality (e.g. by considering a decreasing λ over time, with λ the parameter of the HPP determining the number of collisions)

- cannot start with $W(0) = 0$ since the solution is

$$W(t) = W(0)\exp\left\{(\mu - \sigma^2/2)t + \sigma B(t)\right\}\prod_{j=1}^{N(t)}(Y_j + 1)$$

- Y_j are not observed $\Rightarrow$ need to sample them during MCMC

MODEL ON THICKNESS

$$dT(t) = -T(t^-)\,\{\mu dt + \sigma dB(t) + dJ(t)\}$$

- $T(t)$ thickness of the wall
- $T_0 = T(0)$ initial thickness
- T^* threshold (minimum allowed) thickness
- μ drift, $\mu > 0$, and σ volatility, $\sigma > 0$
- $B(t)$ Wiener process at time t
- $Y_j = \frac{T(\tau_j^-) - T(\tau_j)}{T(\tau_j^-)}$, $j = 1, \ldots, N(t)$, jump sizes
- $J(t) = \sum_{j=1}^{N(t)} Y_j$ jump process

MODEL ON THICKNESS

$$dT(t) = -T(t^-)\,\{\mu dt + \sigma dB(t) + dJ(t)\}$$

Positive aspects

- thickness decrement proportional to thickness

- closed-form solution

$$T(t) = T(0)\,\text{exp}\left\{-(\mu + \sigma^2/2)t - \sigma B(t)\right\} \prod_{j=1}^{N(t)} (1 - Y_j)$$

MODEL ON THICKNESS

$$dT(t) = -T(t^-)\{\mu dt + \sigma dB(t) + dJ(t)\}$$

Negative aspects

- not monotone but we expect very minor oscillations
- $(1 - Y_j) \in (0, 1)$ since $Y_j \in (0, 1)$ but we use a lognormal model for $(1 - Y_j)$, considering a Gaussian distribution which assigns negligible probability to the positive values of $\log(1 - Y_j)$
- Y_j are not observed $\Rightarrow$ need to sample them during MCMC

SOLUTION OF SDE

$$dT(t) = -T(t^-)\{\mu dt + \sigma dB(t) + dJ(t)\}$$

Solution of SDE (Runggaldier, 2003)

$$T(t) = T(0)\exp\left\{-(\mu + \sigma^2/2)t - \sigma B(t)\right\}\prod_{j=1}^{N(t)}(1 - Y_j)$$

Conditional upon $N(t) = n$ and taking $t_0 = 0$

$$\begin{aligned} T(t) &\sim T(0)\cdot\mathcal{LN}\left(-(\mu + \sigma^2)t, \sigma^2 t\right)\cdot\mathcal{LN}\left(an, b^2 n\right) \\ &\sim \mathcal{LN}\left(\log T(0) - (\mu + \sigma^2/2)t + an, \sigma^2 t + b^2 n\right) \end{aligned}$$

Starting at s, with $T(s) > 0$, and conditional on $N(s,t) = n$

$$T(t) \sim \mathcal{LN}\left(\log T(s) - (\mu + \sigma^2/2)(t - s) + an, \sigma^2(t - s) + b^2 n\right)$$

MOMENTS

$$T(t) \sim \mathcal{LN}\left(\log T(0) - (\mu + \sigma^2/2)t + an, \sigma^2 t + b^2 n\right)$$

$U \sim \mathcal{LN}(p, q^2)$

- $\mathcal{E}U = \exp\{p + q^2/2\}$
- $Var(U) = \exp\{2p + q^2\}(\exp(q^2) - 1)$

$$\mathcal{E}(T(t)|N(t)) = \exp\{g + hN(t)\} \Rightarrow \mathcal{E}(T(t)) = \exp\{g - \lambda t + \lambda t e^h\}$$

- $g = \log T(0) - \mu t$
- $h = a + b^2/2$

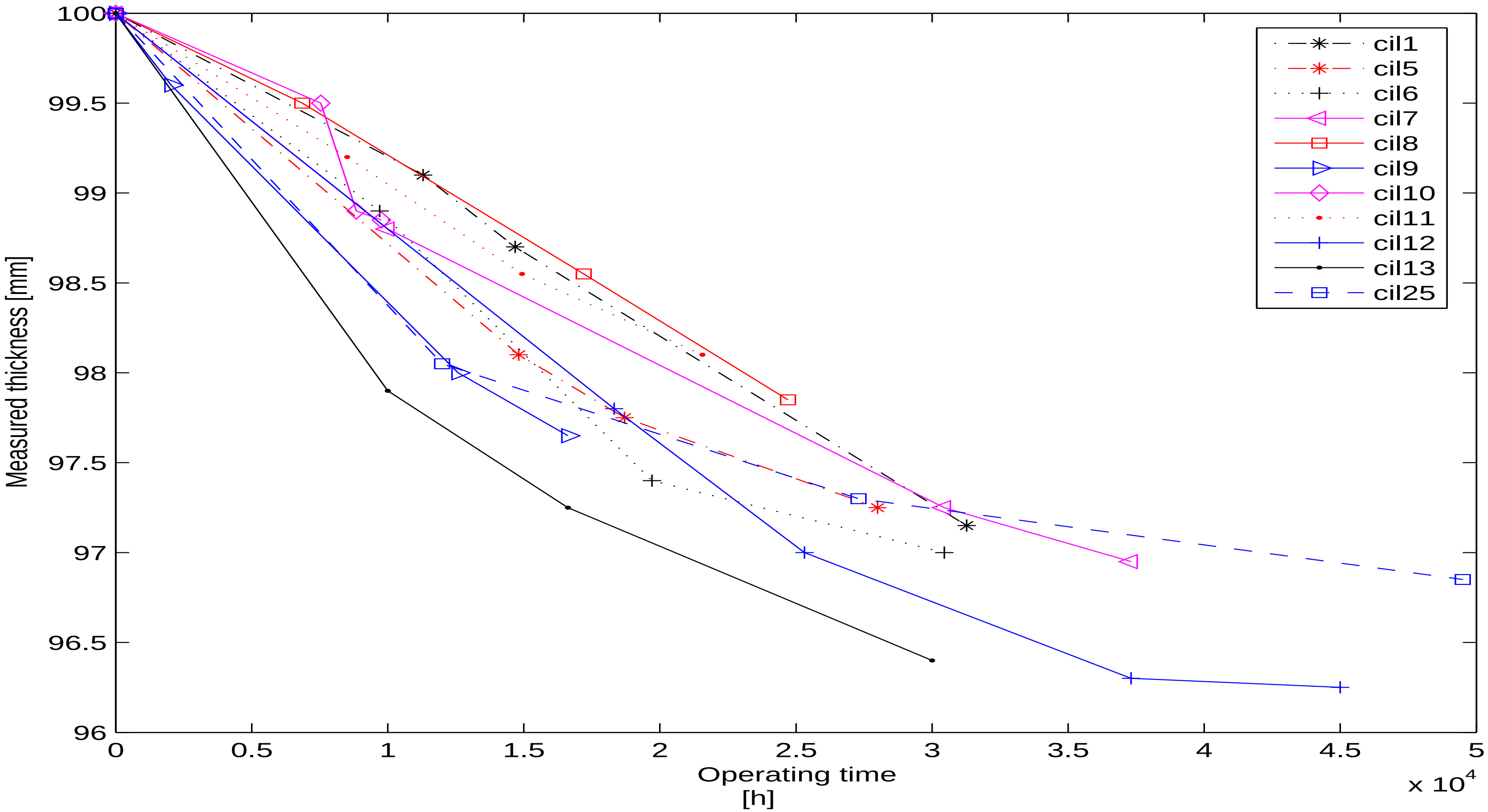
$Var(T(t))$ similar

DATA

- $i = 1, \ldots, 30$ labels of 30 cylinder liners
- $T(t_{i0}) = 100$mm initial liner thickness
- $W_{max} = 4$mm maximum allowed wear
- $T_{max} = 96$mm minimum allowed thickness
- n_i number of inspections performed on the i-th liner;
- $t_{i,j}$ age (in operating hours) of i-th liner at time of j-th inspection;
- $T_{i,j}$ thickness (in mm) measured for i-th liner at time of j-th inspection.

DATA

i	n_i	$t_{i,1}$	$T_{i,1}$	$t_{i,2}$	$T_{i,2}$	$t_{i,3}$	$T_{i,3}$	$t_{i,4}$	$T_{i,4}$
1	3	11300	99.10	14680	98.70	31270	97.15		
2	2	11360	99.20	17200	98.65				
3	2	11300	98.50	21970	98.00				
4	2	12300	99.00	16300	98.65				
5	3	14810	98.10	18700	97.75	28000	97.25		
6	3	9700	98.90	19710	97.40	30450	97.00		
7	3	10000	98.80	30450	97.25	37310	96.95		
8	3	6860	99.50	17200	98.55	24710	97.85		
9	3	2040	99.60	12580	98.00	16620	97.65		
10	3	7540	99.50	8840	98.90	9770	98.85		
11	3	8510	99.20	14930	98.55	21560	98.10		
12	4	18320	97.80	25310	97.00	37310	96.30	45000	96.25
...	...	...	...	...	...				
30	1	8250	99.30						



Measured thickness [mm]
100
99.5
99
98.5
98
97.5
97
96.5
96
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Operating time
[h]
x 10^4
cil1
cil5
cil6
cil7
cil8
cil9
cil10
cil11
cil12
cil13
cil25

LIKELIHOOD

We consider just one cylinder liner

Different models for joint analysis of all liners

- completely different liners $\Rightarrow$ independent models with different parameters $\Rightarrow$ separated analyses
- similar (*exchangeable*) liners $\Rightarrow$ conditionally independent models with parameters from same prior $\Rightarrow$ hierarchical model
- same kind of liners operated under same conditions $\Rightarrow$ product of likelihoods with same parameters and prior
- different operating conditions $\Rightarrow$ models with covariates (e.g. the ship they are in)

LIKELIHOOD

Independent increments (Øksendal, 1998)

$$f(T(t_n) = w_n, \ldots, T(t_1) = w_1) =$$

$$\begin{aligned}
&= f(T(t_n) - T(t_{n-1}) = w_n - w_{n-1}, \ldots, T(t_1) - T(t_0) = w_1 - w_0) \\
&= \prod_{i=1}^{n} f(T(t_i) - T(t_{i-1}) = w_i - w_{i-1}) \\
&= \sum_{n_1,\ldots,n_n} \prod_{i=1}^{n} f(T(t_i) - T(t_{i-1}) = w_i - w_{i-1} | N(t_{i-1}, t_i) = n_i) f(N(t_{i-1}, t_i) = n_i) \quad (1)
\end{aligned}$$

$\Rightarrow$ likelihood based on (1) too difficult to deal with

PARTIAL LIKELIHOOD

- partial likelihood based on $N(t_{i-1}, t_i) = n_i$, $i = 1, \ldots, n$

- treat $N(t_{i-1}, t_i) = n_i$, $i = 1, \ldots, n$, as *parameters* and draw from their (conditional) posterior distribution in the MCMC sampling

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sqrt{\sigma^2(t_i - t_{i-1}) + b^2 n_i}\, w_i} \exp\{-\frac{[\log w_i - \log w_{i-1} + (\mu + \sigma^2/2)(t_i - t_{i-1}) - a n_i]^2}{2[\sigma^2(t_i - t_{i-1}) + b^2 n_i]}\}$$

PRIORS

- $\mu \sim \mathcal{N}(\rho, \tau^2)$
- $a \sim \mathcal{N}(\delta, \epsilon^2)$
- $\sigma^2 \sim \mathcal{IG}(\alpha, \beta)$
- $b^2 \sim \mathcal{IG}(\gamma, \eta)$
- $N_i \sim \mathcal{P}(\lambda(t_i - t_{i-1})), i = 1, \ldots, n$
- $\lambda \sim \mathcal{G}(\phi, \psi)$

PARAMETERS ELICITATION

$\mathcal{E}(T(t)) = \mathcal{E}(T(t))(\theta)$ and $Var(T(t)) = Var(T(t))(\theta)$
$\Rightarrow$ elicitation of a prior on θ

- Fix times $u_1, \ldots, u_k$
- Ask the experts about expected thickness at times $u_1, \ldots, u_k$ and the ranges in which they are almost sure (99%) the thickness will be
- Use the given information to match $\mathcal{E}(T(t))$ and $Var(T(t))$ for each $u_1, \ldots, u_k$ $\Rightarrow 2k$ equations
- Problems
  - transform equations into priors
  - consistency checks

POSTERIOR CONDITIONALS

Set

- $\Delta_i = t_i - t_{i-1}$
- $\Theta_i = \sigma^2(t_i - t_{i-1}) + b^2 n_i$
- $\Sigma_i = \log w_i - \log w_{i-1} + \sigma^2(t_i - t_{i-1})/2 - a n_i$
- $\Gamma_i = \log w_i - \log w_{i-1} + (\mu + \sigma^2/2)(t_i - t_{i-1})$
- $\Phi_i = \log w_i - \log w_{i-1} + (\mu + \sigma^2/2)(t_i - t_{i-1}) - a n_i$
- $\Omega_i = \log w_i - \log w_{i-1} + \mu(t_i - t_{i-1}) - a n_i$

POSTERIOR CONDITIONALS

- $\mu \sim \mathcal{N}(\frac{\rho/\tau^2 - \sum_{i=1}^{n}(\Delta_i \Sigma_i/\Theta_i)}{1/\tau^2 + \sum_{i=1}^{n} \Delta_i^2/\Theta_i}, \left\{1/\tau^2 + \sum_{i=1}^{n} \Delta_i^2/\Theta_i\right\}^{-1})$

- $a \sim \mathcal{N}(\frac{\delta/\epsilon^2 + \sum_{i=1}^{n}(n_i \Gamma_i/\Theta_i)}{1/\epsilon^2 + \sum_{i=1}^{n} n_i^2/\Theta_i}, \left\{1/\epsilon^2 + \sum_{i=1}^{n} n_i^2/\Theta_i\right\}^{-1})$

- $\sigma^2 \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \prod_{i=1}^{n} \{\sigma^2 \Delta_i + b^2 n_i\}^{-1/2} e^{\{-(\Omega_i + \sigma^2 \Delta_i/2)^2/[2(\sigma^2 \Delta_i + b^2 n_i)]\}}$

- $b^2 \propto (b^2)^{-(\gamma+1)} \exp\{-\eta/b^2\} \prod_{i=1}^{n} \{\sigma^2 \Delta_i + b^2 n_i\}^{-1/2} \exp\{-\Phi_i^2/[2(\sigma^2 \Delta_i + b^2 n_i)]\}$

- $N_i \propto \{\sigma^2 \Delta_i + b^2 n_i\}^{-1/2} w_i^{-1} \exp\{-(\Gamma_i - a n_i)^2/[2(\sigma^2 \Delta_i + b^2 n_i)]\} \frac{(\lambda \Delta_i)^{n_i}}{n_i!} \exp\{-\lambda \Delta_i\}$, $i = 1, \ldots, n$

- $\lambda \sim \mathcal{G}(\phi + \sum_{i=1}^{n} n_i, \psi + t_n - t_0)$

DATA AUGMENTATION

Only 1 to 4 observations for each liner
$\Rightarrow$ hard to get *reliable* estimates
$\Rightarrow$ data augmentation

(Elerian et al, 2001, Eraker, 2001, Golightly and Wilkinson, 2005, and Gilioli, Pasquali and FR, *Bulletin of Mathematical Biology*, 2007)

DATA AUGMENTATION

Steps within the MCMC algorithm

1. Choose a number M of points to generate between two inspection points t_{i-1} and t_i, with thickness w_{i-1} and w_i, respectively

2. Get M pairs time/thickness $(y_j^{(0)}, z_j^{(0)})$ by interpolating between (t_{i-1}, w_{i-1}) and (t_i, w_i)

3. Based on Euler's approximation, add the Gaussian contribution of the *new* observations $(y_j^{(m)}, z_j^{(m)})$ to the likelihood

4. Compute the conditional distribution of each *new* observation (considered as *parameter*), given all the other parameters and *new* observations

5. Draw another value $(y_j^{(m)}, z_j^{(m)})$ of the *new* observation (Metropolis-Hastings step)

6. Back to Step 3 until an adequate sample is obtained

PRACTICAL PROBLEM . . .

- Inverse gamma as proposal distribution in the Metropolis steps for σ^2 and b^2

- Very very low acceptance rate under
  - independent proposal (the same at each step)
  - random walk proposal (with mean equal to last drawn values)

. . . AND SOLUTION

- $\sigma^2 \propto (\sigma^2)^{-(\alpha+1)} e^{-\beta/\sigma^2} \prod_{i=1}^{n} \{\sigma^2 \Delta_i + b^2 n_i\}^{-1/2} e^{\{-(\Omega_i + \sigma^2 \Delta_i/2)^2/[2(\sigma^2 \Delta_i + b^2 n_i)]\}}$

- $\Rightarrow \sigma^2 \propto \mathcal{IG}(\alpha, \beta) \cdot h(\sigma^2)$

- Two steps procedure

  - Estimate $\mathcal{E}\sigma^2$ (given other parameters) by importance sampling, i.e.

    $$\frac{\sum \sigma_i^2 h(\sigma_i^2)}{\sum h(\sigma_i^2)},$$

    with $\mathcal{IG}(\alpha, \beta)$ as sampling distribution

  - Perform the Metropolis step with $\mathcal{IG}$ with mean $\mathcal{E}\sigma^2$ and given variance

COMMENTS

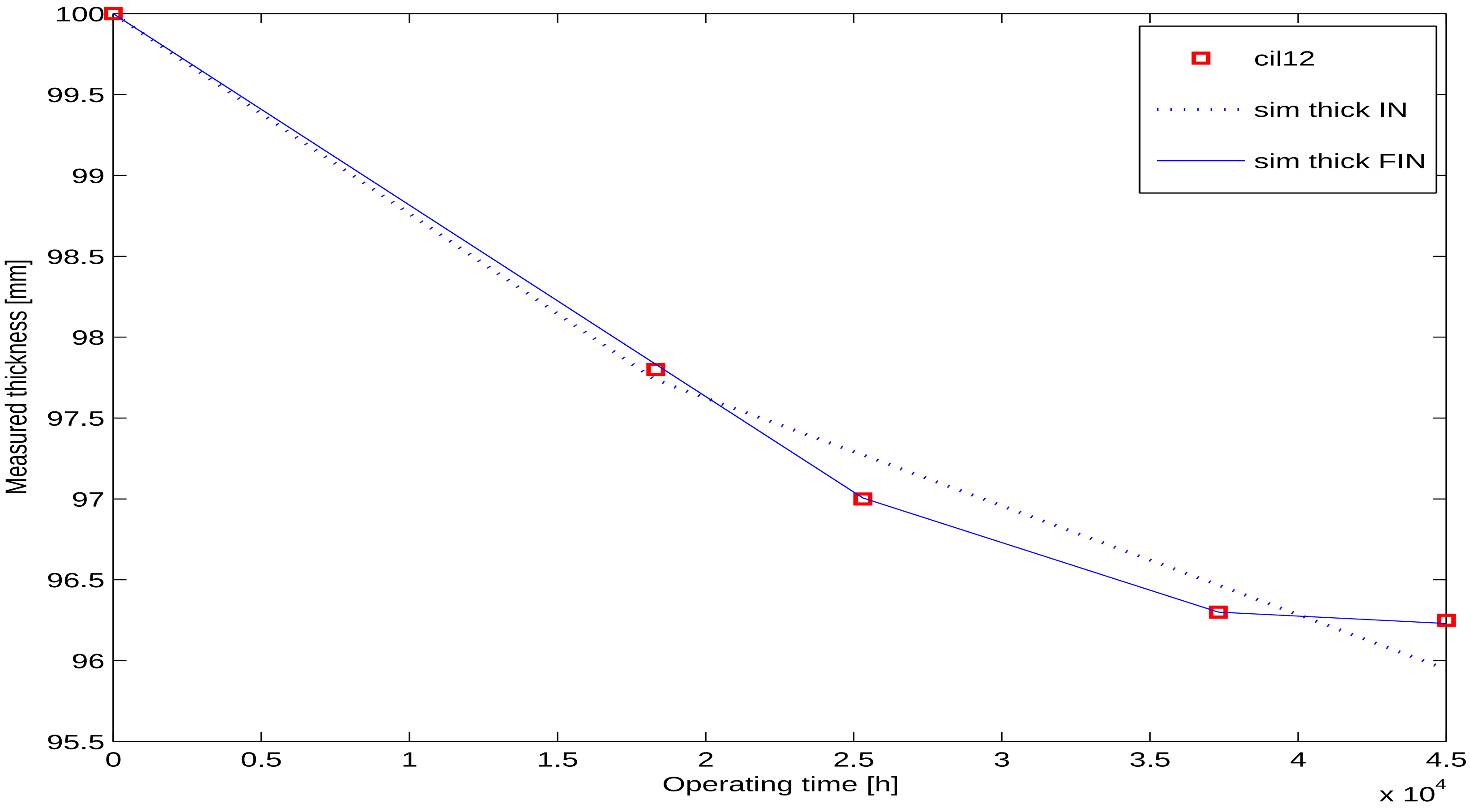
- Two step procedure and data augmentation improve results but worsen execution time

- λ (the parameter of the HPP related to jumps) is rather uninfluential

- Robustness w.r.t. changes in the priors for σ^2 and b^2

- Significant change in expected number of jumps in each interval from prior to posterior

Trajectories with estimated parameters (a priori and posteriori)



CASE STUDY 5: COMPETING RISKS

Many causes of *failures*: here maintenance and actual failure of pumps

Nr.								
1-8	(14,0)	(30,0)	(48,0)	(7,0)	(3,0)	(2,0)	(7,0)	(6,0)
10-16	(11,0)	(7,0)	(8,0)	(24,0)	(21,0)	(5,0)	(1,0)	(7,0)
17-24	(3,0)	(92,0)	(13,0)	(10,0)	(4,0)	(10,0)	(49,0)	(89,0)
25-32	(48,1)	(12,1)	(8,1)	(3,1)	(3,0)	(3,1)	(3,1)	(28,0)
33-34	(23,0)	(22,0)						

Pump data: 0 = failure, 1 = maintenance

Interest in

- determining the goodness of the company's maintenance policy
- proposing an optimal maintenance policy

RANDOM SIGN CENSORING MODEL

Definition 3 *Given a random variable X, consider $Y = X - W\delta$, where W is a random variable with $0 < W < X$ and $\delta = \{-1, 1\}$ is also a random variable independent of X. The variable $Z \equiv [\min(X, Y), 1(Y < X)]$, with $1(\cdot)$ denoting the set function, is called a random sign censoring of X by Y.*

- X failure, Y maintenance and Z observed times
- We want $Y < X$ (i.e. $\delta = -1$) and W as close as possible to 0
- C_R repair and C_M maintenance costs s.t. $C_M < C_R$
- Current expected cost (after n observed failures):
  $\mathcal{E}_C = C_R P(\delta_{n+1} = -1 | \underline{t}, \underline{\delta}) + C_M P(\delta_{n+1} = 1 | \underline{t}, \underline{\delta})$
- Expected cost at time t under *our* model (based upon predictive distribution after n observed failures):
  $\mathcal{E}_t = C_R P(X_{n+1} \leq t | \underline{t}, \underline{\delta}) + C_M P(X_{n+1} > t | \underline{t}, \underline{\delta})$
- Looking for critical t^c s.t. $\mathcal{E}_C = \mathcal{E}_{t^c}$
- Choose $t^* = t^c - \epsilon$, $\epsilon > 0$

Bayesian Analysis of Stochastic Process Models

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it *Fabrizio.Ruggeri@kaust.edu.sa*
www.mi.imati.cnr.it/~fabrizio/

BASIC QUEUEING CONCEPTS

- Queueing system: structure in which *clients* arrive according to some arrival process and wait, if necessary, before receiving service from one or more *servers*

- *Client* attended upon arrival if there are free servers; o.w. it will leave the system immediately or wait for some time until it can wait no longer or until a server becomes available

- Queueing system often summarized by six characteristics $A/S/c/K/M/R$, using Kendall's (1953) notation
  - A and S: forms of arrival and service processes, respectively
  - c number of servers
  - K finite or infinite capacity of the system
  - M finite or infinite customer population
  - R service discipline

BASIC QUEUEING CONCEPTS

- As an example, $M/D/2/10/\infty/\mathrm{FIFO}$
  - M: Markovian arrival process, with exponential i.i.d. inter-arrival times
  - D: Fixed or deterministic service times
  - 2: number of servers
  - 10: capacity of the queue
  - ∞: size of the population
  - FIFO: service policy (*First In First Out)*
- Often systems summarized as in $M/G/c$, shorthand for
  - Markovian arrival process
  - General service time distribution
  - c servers
  - Infinite capacity
  - Infinite population
  - FIFO service policy

QUANTITIES OF INTEREST

- Arrival and service processes
- Client's perspective
  - Waiting time in queue
  - Size of the waiting queue
- Server's perspective
  - Busy period
  - Idle time between services
- $\Rightarrow$ Introduction of performance measures

QUANTITIES OF INTEREST

$$
\begin{aligned}
N_q(t) &= \text{number of clients waiting in queue at time } t \\
N_b(t) &= \text{number of busy servers at time } t \\
N(t) &= N_q(t) + N_b(t) = \text{number of clients in system at time } t \\
W_q(t) &= \text{time spent waiting in queue by a client arriving at time } t \\
W(t) &= \text{time spent in system by a client arriving at instant } t \\
&= W_q(t) + S, \quad \text{with } S \text{ service time of the client}
\end{aligned}
$$

- Quantities of interest
  - typically stochastic
  - exact distributions difficult to obtain for most queueing systems
- Assuming system reaches equilibrium as time increases
- $\Rightarrow$ more straightforward analysis of long term behavior

STABILITY

- Stability of queueing system depends on traffic intensity ρ
- $G/G/c$ system
  - General inter-arrival time distribution G
  - General service distribution G
  - c servers
  - Infinite population
  - Infinite capacity
  - FIFO service discipline
- $\Rightarrow \rho = \lambda E[S]/c$, with
  - λ mean arrival rate, i.e. mean number of arrivals per unit time
  - $E[S]$ mean service time

STABILITY

- For $\rho > 1$, then, on average,
- $\Rightarrow$ arrivals occur at a faster rate than can be handled by the servers
- $\Rightarrow$ queue size will tend to grow over time
- For $\rho = 1$, equilibrium cannot be reached except for both arrival and service deterministic distributions
- For $\rho < 1$, proved (see, e.g., Wolfson, 1986)
  - distributions of $N(t), N_q(t), N_b(t), W(t)$ and $W_q(t)$ approach equilibrium distributions
- N: equilibrium queue size
  - $P(N = n) = \lim_{t \to \infty} P(N(t) = n)$

STABILITY

- W: equilibrium time spent by an arriving customer in the system
  - $F_W(w) = P(W \leq w) = \lim_{t \to \infty} P(W(t) \leq w)$
- Similar definitions for N_b, N_q and W_q
- Little's (1961) laws relating mean numbers of clients in the system and queue to average waiting or queueing time
  - $E[N] = \lambda E[W]$
  - $E[N_b] = \lambda E[W_q]$
- Variables of interest related with total work of servers
  - B length of a server's busy period, i.e. time between arrival of a client in an unoccupied server system and first instant in which the server is empty again
  - I length of a server's idle period, i.e. length of time that a server is unoccupied

MAIN QUEUEING MODELS

- Probabilistic properties
  - $M/M/1$ and Related Systems
  - $GI/M/1$ and $GI/M/c$ Systems
  - $M/G/1$ System
  - $GI/G/1$ Systems
- Use $P(N = n)$ instead of $P(N = n|\lambda, \mu)$

$M/M/1$ AND RELATED SYSTEMS

- $M/M/1$ system
  - Arrivals according to HPP with parameter λ
  - Service time i.i.d. exponential with parameter μ
  - One server
- Birth-death process
  - Birth as arrival in the queue
  - Death as service completion
- Traffic intensity $\rho = \lambda E[S]/c = \lambda/\mu$
- $\Rightarrow$ system stable if $\rho < 1$

$M/M/1$ AND RELATED SYSTEMS

- Stable system $\Rightarrow$ equilibrium distributions available (Gross and Harris, 1998)
- Number of clients in the system
  - $N \sim \text{Ge}(1-\rho), \quad \text{with} \quad E[N] = \dfrac{\rho}{1-\rho}$ ($\uparrow \infty$ for $\rho \uparrow 1$)
- Number of clients in the queue waiting to be served
  - $P(N_q = n) = \begin{cases} P(N=0) + P(N=1) & \text{if } n = 0 \\ P(N = n+1) & \text{for } n \geq 1 \end{cases}$
- Time spent by an arriving customer in the system
  - $W \sim \text{Ex}(\mu - \lambda), \quad \text{with} \quad E[W] = \dfrac{1}{\mu - \lambda} = \dfrac{1}{\mu(1-\rho)}$ ($\uparrow \infty$ for $\rho \uparrow 1$)
- Time spent queueing
  - $P(W_q \leq t) = 1 - \rho e^{-(\mu-\lambda)t}$ for $t \geq 0$ ($\Rightarrow P(W_q = 0) = 1 - \rho$)

$M/M/1$ AND RELATED SYSTEMS

- $M/M/1$ system: one of the few with analytical expression for short term distribution of number of clients $N(t)$ (Clarke, 1953)
- Suppose n_0 clients in the system at time 0

$$
\begin{aligned}
P(N(t)=n) &= e^{-(\lambda+\mu)t}\Big[\rho^{(n-n_0)/2} I_{n-n_0}\left(2\sqrt{\lambda\mu}t\right) \\
&\quad +\rho^{(n-n_0-1)/2} I_{n+n_0+1}\left(2\sqrt{\lambda\mu}t\right) \\
&\quad +(1-\rho)\,\rho^{n/2}\sum_{j=n+n_0+1}^{\infty}\rho^{-j/2} I_j\left(2\sqrt{\lambda\mu}t\right)\Big], \text{with}
\end{aligned}
$$

$I_j(c)$ modified Bessel function of the first kind, i.e.

$$I_j(c)=\sum_{k=0}^{\infty}\frac{(c/2)^{j+2k}}{k!(j+k)!}$$

$M/M/1$ AND RELATED SYSTEMS

- Density function of the duration, B, of a busy period

$$f_B(t) = \frac{\sqrt{\mu/\lambda}\, e^{-(\mu+\lambda)t}\, I_1\left(2\sqrt{\lambda\mu}t\right)}{t}, \quad \text{for } t \geq 0$$

- Density function of the length of a server's idle period I

$$f_I(t) = \lambda e^{-\lambda t}, \quad \text{for } t \geq 0$$

- Many results available for other Markovian systems
  - multiple or infinite servers
  - finite capacity
  - bulk arrivals
- See, e.g., Gross and Harris (1998) or Nelson (1995)

$GI/M/1$ AND $GI/M/c$ SYSTEMS

- $GI/M/1$ system
  - Independent, generally distributed interarrival times
  - Service time i.i.d. exponential with parameter μ
  - One server
- For a stable $GI/M/1$ system $\Rightarrow$ stationary distributions
  - Number of clients in the system *found by an arriving customer*

$$P(N_a = n) = (1-\eta)\eta^n \quad \text{for } n = 0, 1, 2, \ldots, \text{with}$$

    * η smallest positive root of $f_A^*(\mu(1-s)) = s$
    * $f_A^*(s)$ Laplace-Stieltjes transform of the inter-arrival time distribution
    * Laplace-Stieltjes transform of a continuous variable, X, with cdf $F_X(\cdot)$

$$f_X^*(s) = \int_{-\infty}^{\infty} e^{-sx} dF(x)$$

for $s \in \mathbb{C}$, wherever this integral exists

$GI/M/1$ AND $GI/M/c$ SYSTEMS

- For a stable $GI/M/1$ system $\Rightarrow$ stationary distributions
  - Number of clients in the system

$$P(N = n) = \begin{cases} 1 - \rho & \text{if } n = 0 \\ \rho P(N_a = n - 1) & \text{for } n > 0 \end{cases} \text{, with}$$

    ρ traffic intensity of the system
  - Time spent in the system

$$P(W \leq t) = 1 - e^{-\mu(1-\eta)t} \quad \text{for } t \geq 0$$

- Extension from $GI/M/1$ system to multi-server $GI/M/c$ systems
  - Number c of channels taken in account in finding η, smallest positive root of $f_A^*(c\mu(1-s)) = s$
  - Given the root $\Rightarrow$ formulae for waiting time and queue size distributions (see, e.g., Allen, 1990, or Gross and Harris, 1998)

$M/G/1$ SYSTEM

- $M/G/1$ system
  - Arrivals according to HPP with parameter λ
  - Independent, general service times
  - One server

- Stable system $\Rightarrow$ equilibrium distributions available (Gross and Harris, 1998)
  - Distribution of number of clients in the system found recursively through

$$P(N=n) = P(N=0)P(Y=n) + \sum_{j=1}^{n+1} P(N=j)P(Y=n-j+1), \text{with}$$

    * $P(N=0) = 1-\rho$
    * Y represents the number of arrivals during a service time, i.e.

$$P(Y=y) = \int_0^\infty \frac{(\lambda t)^y e^{-\lambda t}}{y!} f_S(t)\, dt$$

    * $f_S(t)$ service time density

$M/G/1$ SYSTEM

- Stable system $\Rightarrow$ equilibrium distributions available (Gross and Harris, 1998)
  - Distributions of waiting time and other variables of interest derived, in general, only in terms of Laplace-Stieltjes transforms
    - Laplace-Stieltjes transform of time spent queueing, W_q, given by

$$f^*_{W_q}(s) = \frac{(1-\rho)s}{s - \lambda[1 - f^*_S(s)]}, \text{with}$$

$f^*_S(s)$ Laplace-Stieltjes transform of service time density

  - Pollaczek-Khintchine formula: general result for $M/G/1$ systems expressing mean queueing time in terms of average arrival rate λ and service rate μ, and the variance σ_s^2 of service time distribution

$$[W_q] = \frac{\lambda\left(\sigma_s^2 + \frac{1}{\mu^2}\right)}{(1-\rho)}$$

$GI/G/1$ SYSTEMS

- Except for some very specific systems, there are very few exact results known concerning the distributions of queue size, busy period etc. for $GI/G/1$ queueing systems

- When exact results are unavailable, one way of estimating the distributions of these variables is to use Discrete Event Simulation techniques (see, e.g., Ch.9 of DRI, FR, MPW, 2012)
  - Interarrival and service times simulated over a sufficiently large time period $T \Rightarrow$ system assumed in equilibrium
  - Computation of corresponding performance measures based on a sufficiently large sample size
  - $\Rightarrow$ Reasonable approximations for small traffic intensity
  - $\Rightarrow$ Inefficient with traffic intensity close to 1 since a very large time period T is needed before assuming equilibrium

BAYESIAN INFERENCE

- Queueing theory dates back to Erlang (1909)
- Inference for queueing systems more recent (Clarke, 1957)
- First Bayesian approaches to inference for Markovian systems in early 1970's (Bagchi and Cunningham, 1972, Muddapur, 1972 and Reynolds, 1973)
- Advantages of Bayesian approach in Armero and Bayarri (1999 et at.)
- Rios Insua, FR and Wiper (and his coauthors)

ADVANTAGES OF BAYESIAN APPROACH

- Uncertainty about system stability easily quantified
  - $P(\rho < 1|\text{data})$: probability of stable (single server) queueing system
  - From a classical viewpoint
    - Point and interval estimators for ρ but ...
    - ... not clear how to measure uncertainty about whether or not a queueing system is stable
- Restrictions in parameter space easily handled
  - Often queueing system assumed stable
  - $\Rightarrow$ take prior distribution for traffic intensity on $[0, 1)$
  - For heavy traffic, stable system $\Rightarrow$ possible MLE $\widehat{\rho} \geq 1$
  - $\Rightarrow$ how to compute equilibrium probabilities of queue size, etc. assuming equilibrium?

ADVANTAGES OF BAYESIAN APPROACH

- Straightforward prediction
  - E.g. System size at a given time

$$P(N(t) = n|\text{data}) = \int_{\boldsymbol{\theta}} P(N(t) = n|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\text{data})\, d\boldsymbol{\theta}$$

  - Equilibrium incorporated by conditioning on system stability
  - Standard classical approach of using plug in estimates can fail to produce sensible predictions for equilibrium probabilities specially under conditions of heavy traffic (see, e.g., Schruben and Kulkarni, 1982)

- Straightforward design
  - Bayesian decision making techniques used to find optimal decision about, e.g.,
    - number of servers
    - capacity of the system
  - $\Rightarrow$ to meet some specified cost or utility condition

PROBLEMS WITH BAYESIAN APPROACH

- Main practical difficulty concerns the experiment to be carried out
- In practice, easy to observe aspects of a queueing system, like
  - queue sizes at given times, customer waiting times and busy period lengths
- **BUT** their distributions
  - usually unknown or available as Laplace-Stieltjes transforms
- $\Rightarrow$ Likelihood function hard, if not impossible, to derive
- $\Rightarrow$ Inferential techniques which do not depend on the likelihood may be needed
- $\Rightarrow$ Most Bayesian papers assume arrival and service processes observed separately
  - $\Rightarrow$ (Usually) straightforward likelihood
  - $\Rightarrow$ Separate observation of the processes usually more expensive and time consuming in practice than simply observing, e.g., lengths of busy periods

INFERENCE FOR $M/M/1$ SYSTEMS

- Most Bayesian work (e.g. McGrath and Singpurwalla, 1987, and Armero and Bayarri, 1996) on $M/M/1$ or related Markovian systems

- $M/M/1$ with unknown arrival and service rates, λ and μ respectively

- Simple experiment of observing first n_a and n_s interarrival and service times, respectively
  - Fixed number of arrivals, decided a priori
  - Observed number of arrivals in a given time period, decided a priori

- Both cases $\Rightarrow$ similar likelihood

INFERENCE FOR $M/M/1$: MLE

- t_a: total time taken for the first n_a arrivals
- t_s: total time taken for the first n_s service completions
- Sum of n i.i.d. exponential random variables has $\Rightarrow$ Erlang distribution
- $\Rightarrow$ Likelihood

$$l(\lambda, \mu|\text{data}) = \frac{\lambda^{n_a}}{\Gamma(n_a)} t_a^{n_a - 1} e^{-\lambda_a t_a} \frac{\mu^{n_s}}{\Gamma(n_s)} t_s^{n_s - 1} e^{-\mu_s t_s} \propto \lambda^{n_a} e^{-\lambda_a t_a} \mu^{n_s} e^{-\mu_s t_s}$$

- $\Rightarrow$ MLE for λ and μ
  - $\widehat{\lambda} = 1/\bar{t}_a$, with mean inter-arrival $\bar{t}_a = t_a/n_a$
  - $\widehat{\mu} = 1/\bar{t}_s$, with mean service times $\bar{t}_s = t_s/n_s$
- $\Rightarrow \widehat{\rho} = \widehat{\lambda}/\widehat{\mu}$ MLE of traffic intensity ρ
- If $\widehat{\rho} < 1 \Rightarrow \widehat{\rho}$ can replace ρ in the formulae for predictive distributions of queue size, waiting time etc. in stable systems

INFERENCE FOR $M/M/1$: BAYES

- Likelihood

$$l(\lambda, \mu|\text{data}) = \frac{\lambda^{n_a}}{\Gamma(n_a)} t_a^{n_a - 1} e^{-\lambda_a t_a} \frac{\mu^{n_s}}{\Gamma(n_s)} t_s^{n_s - 1} e^{-\mu_s t_s} \propto \lambda^{n_a} e^{-\lambda_a t_a} \mu^{n_s} e^{-\mu_s t_s}$$

- $\Rightarrow$ Gamma conjugate prior distributions for λ and μ

$$\lambda \sim \text{Ga}(\alpha_a, \beta_a) \quad \text{and} \quad \mu \sim \text{Ga}(\alpha_s, \beta_s),$$

  with $\alpha_a, \beta_a, \alpha_s, \beta_s > 0$

- Methods to elicit $\alpha_a, \beta_a, \alpha_s, \beta_s$ seen earlier

- Jeffreys priors as alternatives, under little prior information

$$f(\lambda, \mu) \propto \frac{1}{\lambda \mu}$$

- $\Rightarrow$ limiting case of gamma priors when α_a, β_a, α_s and β_s approach zero

INFERENCE FOR $M/M/1$: BAYES

- Given likelihood and gamma priors $\Rightarrow \lambda$ and μ independent a posteriori

$$\lambda|\text{data} \sim \text{Ga}(\alpha_a^*, \beta_a^*) \quad \text{and} \quad \mu|\text{data} \sim \text{Ga}(\alpha_s^*, \beta_s^*),$$

  with $\alpha_a^* = \alpha_a + n_a$, $\beta_a^* = \beta_a + t_a$, $\alpha_s^* = \alpha_s + n_s$ and $\beta_s^* = \beta_s + t_s$

- Jeffreys prior $\Rightarrow \lambda|\text{data} \sim \text{Ga}(n_a, t_a)$ and $\mu|\text{data} \sim \text{Ga}(n_s, t_s)$

- Posterior distribution of ρ (Armero, 1985)
  - From $2\beta_a^*\lambda|\text{data} \sim \chi^2_{2\alpha_a^*}$, $2\beta_s^*\mu|\text{data} \sim \chi^2_{2\alpha_s^*}$ and
  - ratio of two χ^2 distributions divided by their degrees of freedom $\Rightarrow$ F distributed
  - $\Rightarrow \left.\dfrac{\alpha_s^*\beta_a^*}{\alpha_a^*\beta_s^*}\rho\right|\text{data} \sim \text{F}^{2\alpha_s^*}_{2\alpha_a^*}$
  - Jeffreys prior $\Rightarrow \left.\dfrac{n_s t_a}{n_a t_s}\rho\right|\text{data} \sim \text{F}^{2n_s}_{2n_a}$
  - Jeffreys prior $\Rightarrow$ Bayesian credible intervals for ρ coincide with frequentist ones

INFERENCE FOR $M/M/1$: BAYES

- Posterior mean of ρ for gamma prior

$$E[\rho|\text{data}] = \frac{\alpha_a^* \beta_s^*}{(\alpha_s^* - 1)\beta_a^*}$$

- Posterior mean of ρ for Jeffreys prior

$$E[\rho|\text{data}] = \frac{t_s n_a}{t_a(n_s - 1)} = \frac{n_s}{n_s - 1}\widehat{\rho}$$

  different from MLE $\widehat{\rho}$

- Posterior probability of stable system

$$P(\rho < 1|\text{data}) = \frac{(\beta_a^*/\beta_s^*)^{\alpha_a^*}}{\alpha_a^* B(\alpha_a^*, \alpha_s^*)} {}_2F_1\left(\alpha_a^* + \alpha_s^*, \alpha_a^*; \alpha_a^* + 1; -\frac{\beta_a^*}{\beta_s^*}\right),$$

  where ${}_2F_1(a, b; c; d)$ is the Gauss hypergeometric function

$${}_2F_1(a, b; c; d) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 x^{b-1}(1-x)^{c-b-1}(1-dx)^{-1}\, dx$$

TESTING FOR STABILITY

- Posterior on $\rho \Rightarrow$ assess stability of queueing system (Armero and Bayarri, 1994)
- Formal test for stationarity: $H_0 : \rho < 1$ vs. $H_1 : \rho \geq 1$
- p_0 posterior probability for H_0
- $p_1 = 1 - p_0$ posterior probability for H_1
- Two types of error
  - Assume H_0 is true when H_1 is true
  - Assume H_1 is true when H_0
- First error associated with loss l_{01} and second with loss l_{10}
- Losses determined by gravity of each error
- Minimizing the expected loss, H_0 optimal decision if $l_{01}p_1 < l_{10}p_0 \quad \text{or} \quad \frac{p_0}{p_1} > \frac{l_{01}}{l_{10}}$

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Suppose a stable system
- From $\left.\frac{\alpha_s^*\beta_a^*}{\alpha_a^*\beta_s^*}\rho\right|$ data $\sim \mathrm{F}_{2\alpha_a^*}^{2\alpha_s^*}$ and
- $P(\rho < 1|\text{data}) = \frac{(\beta_a^*/\beta_s^*)^{\alpha_a^*}}{\alpha_a^* B(\alpha_a^*,\alpha_s^*)} {}_2F_1\left(\alpha_a^* + \alpha_s^*, \alpha_a^*; \alpha_a^* + 1; -\frac{\beta_a^*}{\beta_s^*}\right)$
- $\Rightarrow$ density of traffic intensity conditional on stability condition, $0 < \rho < 1$

$$f(\rho|\text{data}, \rho < 1) = \frac{\alpha_a^*}{{}_2F_1\left(\alpha_a^* + \alpha_s^*, \alpha_a^*; \alpha_a^* + 1; -\frac{\beta_a^*}{\beta_s^*}\right)} \rho^{\alpha_a^* - 1} \left(1 + \frac{\beta_a^*}{\beta_s^*}\rho\right)^{-(\alpha_a^* + \alpha_s^*)}$$

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Predictive distribution of number of clients in a system
  - From $N \sim \text{Ge}(1-\rho), \quad \text{with} \quad E[N] = \frac{\rho}{1-\rho} \Rightarrow$

$$P(N=n|\text{data}, \rho<1) = \int_0^1 (1-\rho)\rho^n f(\rho|\text{data}, \rho<1)\, d\rho$$

$$= \frac{\alpha_a^* \Gamma(\alpha_a^*+n)}{\Gamma(\alpha_a^*+n+2)} \frac{{}_2F_1\left(\alpha_a^*+\alpha_s^*, \alpha_a^*+n; \alpha_a^*+n+2; -\frac{\beta_a^*}{\beta_s^*}\right)}{{}_2F_1\left(\alpha_a^*+\alpha_s^*, \alpha_a^*; \alpha_a^*+1; -\frac{\beta_a^*}{\beta_s^*}\right)}$$

- Predictive distribution of number of clients queueing in equilibrium, obtained from
  - $P(N=n|\text{data}, \rho<1)$
  - $P(N_q=n) = \begin{cases} P(N=0)+P(N=1) & \text{if } n=0 \\ P(N=n+1) & \text{for } n \geq 1 \end{cases}$

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Nonexistence of mean of the predictive distribution of N
  - From $N \sim \text{Ge}(1-\rho), \quad \text{with} \quad E[N] = \frac{\rho}{1-\rho}$

$$
\begin{aligned}
E[N|\text{data}, \rho < 1] &= E\left[\frac{\rho}{1-\rho}\middle| \text{data}, \rho < 1\right] \\
&= \frac{1}{P(\rho < 1|\text{data})} \int_0^1 \frac{\rho}{1-\rho} f(\rho|\text{data})\, d\rho,
\end{aligned}
$$

  divergent integral as $f(\rho|\text{data})$ does not approach 0 when ρ tends to 1

- Predictive means of the equilibrium waiting time W and queueing time W_q distributions do not exist either, applying Little's laws
  - $E[N] = \lambda E[W]$ and $E[N_b] = \lambda E[W_q]$

- Nonexistence of these moments not only in Bayesian approach: similar result for MLE (see Schruben and Kulkarni, 1982)

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Alternative approaches:
  - $\rho < 1 - \epsilon$ in a Bayesian approach (Lehoczky, 1990)
    - finite moments
    - very sensitive to choice of ϵ (Rios Insua, FR, Wiper)
  - More later on
    - assumption of equilibrium
    - prior distributions going to zero as ρ approaches unity

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Explicit forms for limiting waiting time and queueing time (but not busy period or transient queue size) densities and distribution functions available in terms of complex functions (Armero and Bayarri, 1994)

- Monte Carlo sampling: simple alternative to estimate these distributions
  - Random sample of size R, $((\lambda_1, \mu_1), \ldots, (\lambda_R, \mu_R))$ drawn from posterior
  - Predictive distributions estimated through sample averages
  - Full Monte Carlo sample used for, e.g., duration of a busy period or size of queue at a fixed time in future, when no assumption on equilibrium is needed
    * Predictive, busy period density function

$$f_B(t) = \frac{\sqrt{\mu/\lambda}\, e^{-(\mu+\lambda)t}\, I_1\left(2\sqrt{\lambda\mu}t\right)}{t}, \quad \text{for } t \geq 0$$

estimated through

$$f_B(t|\text{data}) \approx \frac{1}{R} \sum_{r=1}^{R} \frac{\sqrt{\mu_r/\lambda_r} e^{-(\mu_r+\lambda_r)t} I_1\left(2\sqrt{\lambda_r\mu_r}t\right)}{t} \quad \text{for } t \geq 0$$

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- Predictive, busy period density function

$$f_B(t) = \frac{\sqrt{\mu/\lambda}\, e^{-(\mu+\lambda)t}\, I_1\left(2\sqrt{\lambda\mu}t\right)}{t}, \quad \text{for } t \geq 0$$

estimated through

$$f_B(t|\text{data}) \approx \frac{1}{R}\sum_{r=1}^{R} \frac{\sqrt{\mu_r/\lambda_r} e^{-(\mu_r+\lambda_r)t} I_1\left(2\sqrt{\lambda_r\mu_r}t\right)}{t} \quad \text{for } t \geq 0$$

- Using this method, all predictive density and distribution functions can be estimated

- Equilibrium condition taken into account considering only pairs λ_r, μ_r s.t. $\lambda_r < \mu_r$

PREDICTIVE DISTRIBUTIONS GIVEN EQUILIBRIUM

- No need for Monte Carlo sampling for predicting idle time distribution
  - Given posterior on λ
  - $\Rightarrow$ Predictive distribution of length of a server's idle period, I, easily evaluated
    - $f_I(t) = \lambda e^{-\lambda t}, \quad \text{for } t \geq 0$
    - $\Rightarrow$ A posteriori, for $t > 0$,

$$f_I(t|\text{data}) = \int_0^\infty \lambda e^{-\lambda t} \frac{\beta_a^{*\alpha_a^*}}{\Gamma(\alpha_a^*)} \lambda^{\alpha_a^*-1} e^{-\beta_a^* \lambda} \, d\lambda = \frac{\alpha_a^* \beta_a^{*\alpha_a^*}}{(\beta_a^* + t)^{\alpha_a^*+1}},$$

    - $\Rightarrow I + \beta_a^* | \texttt{data} \sim \text{Pa}(\alpha_a^*, \beta_a^*)$, shifted Pareto distribution

AUTOMATIC TELLER MACHINE

- Inter-arrival and service time data for 98 users of an automatic teller machine in Berkeley, California (Hall, 1991)
- $M/M/1$ system reasonable choice, because of Poisson arrivals and exponential services (Hall, 1991)
- Sufficient statistics: $n_a = n_s = 98$, $t_a = 119.71$ and $t_s = 81.35$ minutes
- Jeffreys' prior
  - Posterior probability of stable system: 0.9965

$$P(\rho < 1|\text{data}) = \frac{(\beta_a^*/\beta_s^*)^{\alpha_a^*}}{\alpha_a^* B(\alpha_a^*, \alpha_s^*)} {}_2F_1\left(\alpha_a^* + \alpha_s^*, \alpha_a^*; \alpha_a^* + 1; -\frac{\beta_a^*}{\beta_s^*}\right)$$

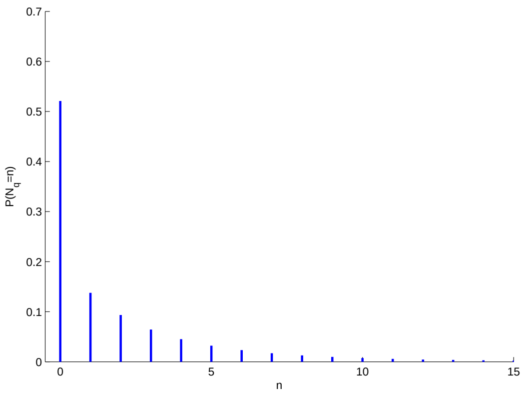
  - Posterior mean of ρ: 0.668

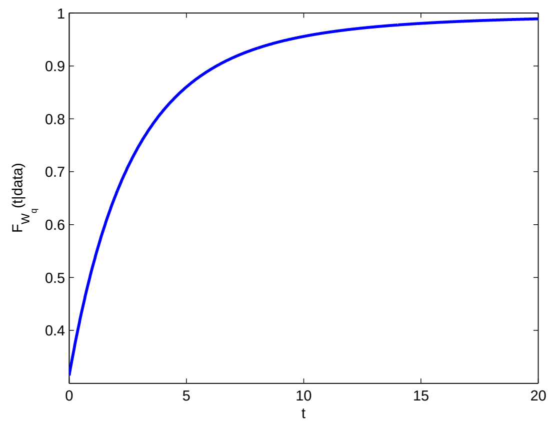
$$E[\rho|\text{data}] = \frac{\alpha_a^* \beta_s^*}{(\alpha_s^* - 1)\beta_a^*}$$

  - High probability of stable system $\Rightarrow$ estimation of system properties under equilibrium

AUTOMATIC TELLER MACHINE

- Stable system
- N_q number of clients queueing for service in equilibrium
- W_q time spent queueing for service by an arriving customer

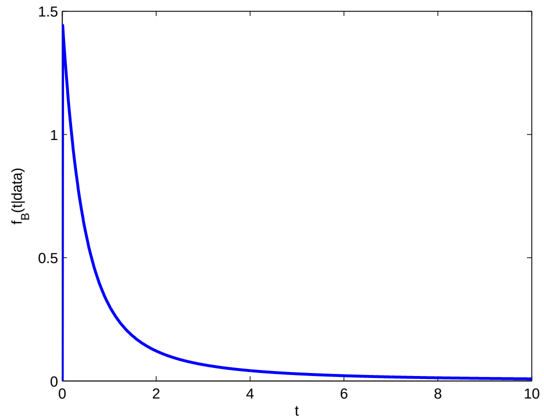




Predictive probability function of N_q (left hand side) and cumulative distribution function of W_q (right hand side)

AUTOMATIC TELLER MACHINE

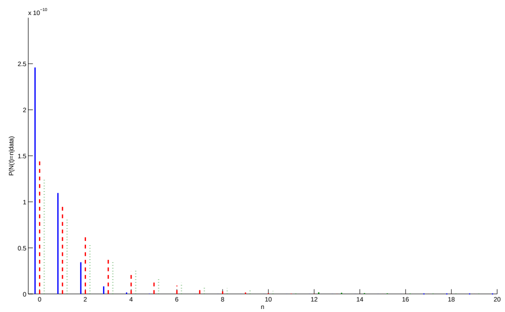
- Estimation of busy period and transient distributions
- $\Rightarrow$ Monte Carlo sample of 1000 data generated from posteriors of λ and μ



Predictive density function of duration of a busy period

AUTOMATIC TELLER MACHINE

- System initially empty
- Transient probability function appears to converge to a limit over time
- In fact, probabilities after 50 minutes very close to predictive equilibrium probabilities



Predictive density function of numbers of clients in system after 1, (solid line) 10 (dashed line) and 50 (dot dash line) minutes

ALTERNATIVE PRIOR FORMULATIONS

- Assume *a priori* stable queueing system
- $\Rightarrow$ Choose a prior s.t. traffic intensity $\rho < 1$
- Reparameterization using ρ and μ instead of λ and μ (Armero and Bayarri, 1994)

$$\begin{aligned} \mu|\rho &\sim \text{Ga}(\alpha, \beta + \gamma\rho) \\ \rho &\sim \text{GH}(\delta, \epsilon, \alpha, \nu), \end{aligned}$$

i.e. a Gauss Hypergeometric distribution: $X \sim \text{GH}(\alpha, \beta, \gamma, \nu)$ if

$$f(x) = \frac{1}{B(\alpha, \beta)_2F_1(\gamma, \alpha; \alpha + \beta; -\nu)} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{(1+\nu x)^{\gamma}} \quad x > 0$$

for $\alpha, \beta, \gamma, \nu > 0$, where ${}_2F_1(a, b; c; d)$ is the Gauss hypergeometric function

$${}_2F_1(a, b; c; d) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 x^{b-1}(1-x)^{c-b-1}(1-dx)^{-1}\,dx$$

- $\Rightarrow$ Reparameterized likelihood

$$l(\mu, \rho|\text{data}) \propto \rho^{n_a}\mu^{n_a+n_s}e^{-(t_s+t_a\rho)\mu}.$$

ALTERNATIVE PRIOR FORMULATIONS

- $\Rightarrow$ posterior distribution of same form

$$\begin{array}{rcl} \mu|\rho, \text{data} & \sim & \text{Ga}(\alpha^*, \beta^* + \gamma^* \rho) \\ \rho|\text{data} & \sim & \text{GH}(\delta^*, \epsilon^*, \alpha^*, \nu^*) \end{array}$$

where $\alpha^* = \alpha + n_a + n_s$, $\beta^* = \beta + t_s$, $\gamma^* = \gamma + t_a$, $\delta^* = \delta + n_a$, $\epsilon^* = \epsilon$ and $\nu^* = \frac{\gamma^*}{\beta^*}$

- Predictive probability function of number of clients in the system

$$\begin{array}{rcl} P(N = n|\text{data}) & = & E\left[(1-\rho)\rho^n \mid \text{data}\right] \\ & = & \dfrac{B\left(\delta^* + n, \epsilon^* + 1\right)}{B\left(\delta^*, \epsilon^*\right)} \dfrac{{}_2F_1\left(\alpha^*, \delta^* + n; \delta^* + \epsilon^* + n + 1; -\nu^*\right)}{{}_2F_1\left(\alpha^*, \delta^*; \delta^* + \epsilon^*; -\nu^*\right)}. \end{array}$$

- Predictive mean number of clients in the system in equilibrium exists for $\epsilon > 1$

$$E[N|\text{data}] = \frac{\delta^*}{\epsilon^* - 1} \frac{{}_2F_1\left(\alpha^*, \delta^* + 1; \delta^* + \epsilon^*; -\nu^*\right)}{{}_2F_1\left(\alpha^*, \delta^*; \delta^* + \epsilon^*; -\nu^*\right)}$$

- k-th moment of N exists $\Leftrightarrow \epsilon > k$

ALTERNATIVE PRIOR FORMULATIONS

- Similarly (see Armero and Bayarri, 1994)
  - Expressions for waiting time and queueing time distributions
  - k-th moments of these distributions also exist $\Leftrightarrow \epsilon > k$
- Choice of ϵ in Gauss hypergeometric prior for ρ: very important
  - ϵ not updated given the experimental data
  - Default *non-informative* prior (Armero and Bayarri, 1994)

$$f(\mu, \rho) \propto \frac{1}{\mu}\frac{(1-\rho)^2}{\rho}$$

    corresponding to $\alpha = \beta = \gamma = \delta = \nu = 0$ and $\epsilon = 3$ in the conjugate prior formulation
  - $\Rightarrow$ Existence, a posteriori, of predictive means and variances of system size, waiting times, etc.

AUTOMATIC TELLER MACHINE

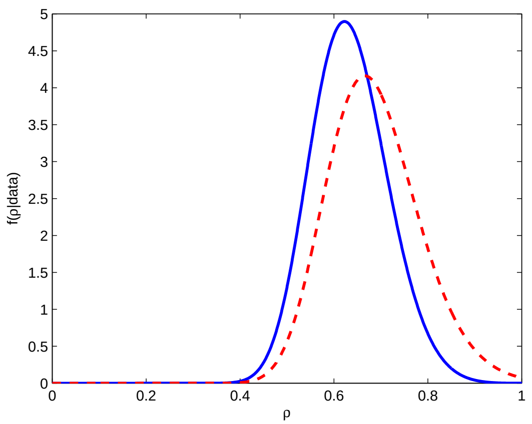
- Inter-arrival and service time data for 98 users of an automatic teller machine in Berkeley, California (Hall, 1991)

- Posterior density of ρ based on prior $f(\mu, \rho) \propto \dfrac{1}{\mu}\dfrac{(1-\rho)^2}{\rho}$ vs

- Posterior, truncated F density of ρ conditional on equilibrium based on Jeffreys prior $f(\lambda, \mu) \propto \dfrac{1}{\lambda\mu}$ for λ and μ

AUTOMATIC TELLER MACHINE

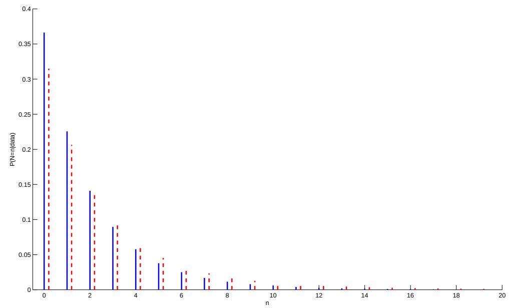
Posterior densities of ρ given the Gauss hypergeometric (solid line) and Jeffreys (dashed line) priors



- $\Rightarrow$ GH posterior concentrated on slightly lower values of ρ than truncated F posterior

AUTOMATIC TELLER MACHINE

Posterior predictive density of N given the Gauss hypergeometric (solid line) and conjugate (dashed line) priors



- $\Rightarrow$ Distribution of N shorter tailed for GH prior (expected from previous plot with posterior shifted towards lower values of ρ)

AUTOMATIC TELLER MACHINE

- Sensitivity to prior parameters

- $\Rightarrow$ values of expected posterior values of N for priors of the form

$$f(\mu,\rho) \propto \frac{1}{\mu}\frac{(1-\rho)^{\epsilon-1}}{\rho}$$

for different values of ϵ

ϵ	1.0	1.7	1.8	1.9	2.0	3.0	4.0
$E[N\|\text{data}]$	∞	168.6	7.3	2.3	2.1	1.9	1.7

- $\Rightarrow$ High sensitivity to changes in ϵ

- Sensitivity w.r.t. ϵ, power of $(1-\rho) \Rightarrow$ major drawback in using Gauss hypergeometric prior (FR, Rios Insua, Wiper, 1996)

ALTERNATIVE EXPERIMENTS

- Observing arrival and service processes separately $\Rightarrow$ straightforward inference
- **BUT**, in practice, often easier to observe number of clients in the system at given times, or waiting times of clients or durations of busy periods
- $\Rightarrow$ Drawback of such experiments: likelihood with complicated form, like in
  - Empty system at time $t_0 = 0$
  - Numbers of clients in system, $(n(t_1), \ldots, n(t_m))$, observed at m time periods, $(t_1, \ldots, t_m)$
  - Short term distribution of number of clients $N(t)$ (Clarke, 1953)

$$
\begin{aligned}
P(N(t)=n) \quad = \quad & e^{-(\lambda+\mu)t} \Big[\rho^{(n-n_0)/2} I_{n-n_0} \left(2\sqrt{\lambda\mu t}\right) \\
& + \rho^{(n-n_0-1)/2} I_{n+n_0+1} \left(2\sqrt{\lambda\mu t}\right) \\
& + (1-\rho)\, \rho^{n/2} \sum_{j=n+n_0+1}^{\infty} \rho^{-j/2} I_j \left(2\sqrt{\lambda\mu t}\right) \Big]
\end{aligned}
$$

ALTERNATIVE EXPERIMENTS

- Numbers of clients in system
  - $\Rightarrow$ Likelihood

$$
\begin{aligned}
l(\lambda, \mu|\text{data}) \quad = \quad & \prod_{i=1}^{m} e^{-(\lambda+\mu)(t_i - t_{i-1})} \left[\rho^{\frac{n(t_i)-n(t_{i-1})}{2}} I_{n(t_i)-n(t_{i-1})} \left(2\sqrt{\lambda\mu(t_i - t_{i-1})} \right) + \right. \\
& \rho^{\frac{n(t_i)-n(t_{i-1})-1}{2}} I_{n(t_i)-n(t_{i-1})-1} \left(2\sqrt{\lambda\mu(t_i - t_{i-1})} \right) + \\
& \left. (1-\rho)\, \rho^{n(t_i)-n(t_{i-1})} \sum_{j=n(t_i)+n(t_{i-1})+2}^{\infty} \rho^{-j/2} I_j \left(2\sqrt{\lambda\mu(t_i - t_{i-1})} \right) \right] \text{, with}
\end{aligned}
$$

$I_j(c)$ modified Bessel function of first kind, i.e. $I_j(c) = \sum_{k=0}^{\infty} \frac{(c/2)^{j+2k}}{k!(j+k)!}$

  - Usual prior distributions
  - $\Rightarrow$ approximate methods such as numerical integration or MCMC necessary to evaluate the posterior distribution

ALTERNATIVE EXPERIMENTS

- Stable system and observed data sufficiently spaced in time
- $\Rightarrow$ data assumed to be generated approximately independently from equilibrium distribution $N \sim \text{Ge}(1-\rho)$
- $\Rightarrow$ likelihood approximated through

$$
\begin{aligned}
l(\lambda,\mu|\text{data}) &\approx \prod_{i=1}^{m}\left(1-\frac{\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^{n(t_i)} \\
&= \left(1-\frac{\lambda}{\mu}\right)^{m}\left(\frac{\lambda}{\mu}\right)^{\sum_{i=1}^{m} n(t_i)} \\
&= (1-\rho)^{m}\rho^{\sum_{i=1}^{m} n(t_i)}.
\end{aligned}
$$

- Conjugate beta prior $\rho \sim \text{Be}(a,b)$
- $\Rightarrow$ Posterior $\rho|\text{data} \sim \text{Be}\left(a+\sum_{i=1}^{m} n(t_i), b+m\right)$

ALTERNATIVE EXPERIMENTS

- $\Rightarrow$ Predictive distribution of N easy to evaluate
- Problems with this approach
  - Likelihood approximation poor for values of ρ close to 1
  - Approximate likelihood as function only of ρ
    - $\Rightarrow$ no information about arrival and service rates individually
    - $\Rightarrow$ strong prior knowledge of at least one of these parameters necessary for inferences on waiting times or transient distributions dependent on both parameters

INFERENCE FOR NON MARKOVIAN SYSTEMS

- Systems where either the arrival or service process is not Markovian
- $\Rightarrow$ Start with $GI/M/1$ and $GI/M/c$ systems
- Basic experiment
  - n_a interarrival times with sums t_a and n_s service times with sums t_s
- Gamma prior $\mu \sim$ Ga(α_s, β_s) for service rate
- $\Rightarrow$ Gamma posterior $\mu|$data $\sim$ Ga(α_s^*, β_s^*), with $\alpha_s^* = \alpha_s + n_s$ and $\beta_s^* = \beta_s + t_s$
- Parametric model for the interarrival times $\Rightarrow$ straightforward inferences
  - Generation of posterior distributions
  - Monte Carlo samples taken from these distributions
  - For each set of sampled parameters, computation of distribution of queue size, waiting time etc. using earlier formulae
  - Predictive distributions estimated by averaging over all samples

INFERENCE FOR $Er/M/1$ SYSTEMS

- Erlang interarrival times (Wiper, 1998)
- Arrivals occur in ν i.i.d exponential stages with rate λ/ν
- $\Rightarrow$ Interarrival time $X|\nu,\lambda \sim \mathrm{Er}(\nu,\lambda)$ with $E[X|\nu,\lambda] = 1/\lambda$
- Single exponential server with rate μ
  - Traffic intensity $\rho = \lambda/\mu$
  - Stable system if $\rho < 1$

INFERENCE FOR $Er/M/1$ SYSTEMS

- For a $GI/M/1$ stable system
  - $\Rightarrow$ Number of clients in the system *found by an arriving customer* $N_a \sim$ Ge(η)
    - η smallest positive root of $f_A^*(\mu(1-s)) = s$
    - $f_A^*(s)$ Laplace-Stieltjes transform of the inter-arrival time distribution
  - $\Rightarrow$ Number N of clients in the system

$$P(N = n) = \begin{cases} 1-\rho & \text{if } n = 0 \\ \rho P(N_a = n-1) & \text{for } n > 0 \end{cases}$$

- For an $Er/M/1$ stable system
  - $\Rightarrow$ geometric stationary distribution of N with parameter η, where

$$\eta \left(1 - \frac{(\eta - 1)}{\rho\nu}\right)^{\nu} = 1$$

  - $\Rightarrow$ Time spent in the system by an arriving customer $W|\mu, \eta \sim$ Ex($\mu(1-\eta)$)

INFERENCE FOR $Er/M/1$ SYSTEMS

- Conjugate prior for (ν, λ), for $\lambda > 0$, $\nu = 1, 2, \ldots$

$$f(\nu, \lambda) \propto \frac{\theta_a^{\nu-1} \nu (\lambda\nu)^{\alpha_a \nu - 1} e^{-\beta_a \nu \lambda}}{((\nu-1)!)^{\alpha_a}}$$

  - $\Rightarrow$ Conditional distribution $\lambda|\nu \sim \text{Ga}(\nu\alpha_a, \beta_a\nu)$
  - $\Rightarrow$ Marginal $P(\nu) \propto \dfrac{\Gamma(\alpha_a \nu)}{((\nu-1)!)^{\alpha_a}} \left(\dfrac{\theta_a}{\beta_a^{\alpha_a}}\right)^{\nu-1}$

- Default prior, setting $\theta_a = 1$ and $\alpha_a = \beta_a = 0$ in conjugate $\Rightarrow f(\lambda, \nu) \propto \dfrac{1}{\lambda}$

- Conjugate prior $\Rightarrow$ posterior

$$\begin{aligned} \lambda|\nu, \text{data} &\sim \text{Ga}(\nu\alpha_a^*, \beta_a^*\nu), \\ P(\nu|\text{data}) &\propto \frac{\Gamma(\alpha_a^* \nu)}{((\nu-1)!)^{\alpha_a^*}} \left(\frac{\theta_a^*}{\beta_a^{*\alpha_a^*}}\right)^{\nu-1}, \end{aligned}$$

$\alpha_a^* = \alpha_a + n_a$, $\beta_a^* = \beta_a + t_a$, $\theta_a^* = \theta_a T_a$, T_a product of observed inter-arrival times

INFERENCE FOR $Er/M/1$ SYSTEMS

- Posterior expected traffic intensity

$$E[\rho|\text{data}] = \frac{\alpha_a^* \beta_s^*}{\beta_a^*(\alpha_s^* - 1)}$$

- Posterior predictive probability of stable system

$$P(\rho < 1|\text{data}) = \sum_{\nu} P(\nu|\text{data}) \frac{(\beta_a^* \nu / \beta_s^*)^{\alpha_a^* \nu}}{\alpha_a^* \nu B(\alpha_a^* \nu, \alpha_s^*)} {}_2F_1 \left(\alpha_a^* \nu + \alpha_s^*, \alpha_a^* \nu; \alpha_a^* \nu + 1; -\frac{\beta_a^* \nu}{\beta_s^*} \right)$$

- Impossible to derive exact expressions for predictive distributions of numbers of clients in the system or a clients waiting time, even under stability

INFERENCE FOR $Er/M/1$ SYSTEMS

- Alternative approach
  - For each (integer) ν (up to some fixed maximum) use Monte Carlo to sample the joint posterior distribution of λ, μ given ν
  - For each ν get a sample (of size M sufficiently large) $\lambda_i^{(\nu)}, \mu_i^{(\nu)}$ s.t. $\lambda_i^{(\nu)} < \mu_i^{(\nu)}$ for $i = 1, \ldots, M$
  - $\Rightarrow$ Distribution of N estimated through

$$P(N|\text{data}) \approx \frac{1}{M} \sum_{\nu=1} P(\nu|\text{data}) \left[\sum_{i=1}^{M} P(N|\lambda_i^{(\nu)}, \mu_i^{(\nu)}) \right]$$

  - Similarly for other quantities of interest

$Er/M/1$: SIMULATED DATA

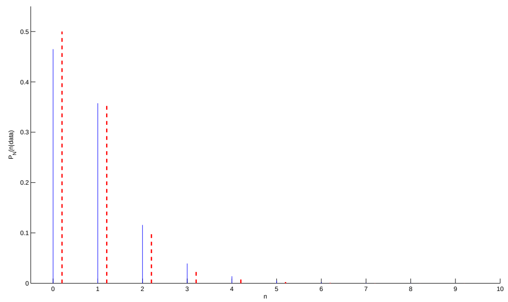
- 100 inter-arrival times simulated from $\text{Er}(5, 0.5)$
  - Sufficient statistics $n_a = 100, t_a = 191.93, \log T_a = 55.34$
- 100 service data simulated from $\text{Ex}(1)$
  - Sufficient statistics $n_s = 100$, $t_s = 101.70$
- Non-informative priors: $f(\nu, \lambda) \propto \frac{1}{\lambda}$ and $f(\mu) \propto \frac{1}{\mu}$
- $\Rightarrow$ Posterior for μ: $\mu|\text{data} \sim \text{Ga}(100, 101.70)$
- Posterior distribution of λ, ν straightforward from

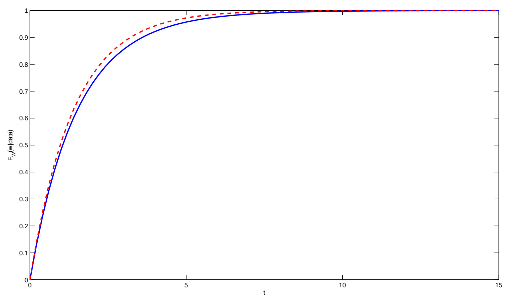
$$\begin{aligned}\lambda|\nu, \text{data} &\sim \text{Ga}(\nu\alpha_a^*, \beta_a^*\nu), \\ P(\nu|\text{data}) &\propto \frac{\Gamma(\alpha_a^*\nu)}{((\nu-1)!)^{\alpha_a^*}} \left(\frac{\theta_a^*}{\beta_a^{*\alpha_a^*}}\right)^{\nu-1},\end{aligned}$$

- Expected value of ρ: 0.535 (true value: 0.5)
- $P(\rho < 1|\text{data}) = 0.9999$

$Er/M/1$: SIMULATED DATA

Predictive (solid line) and true (dashed line) distributions of N and W





- In both cases, true and predictive distributions are similar but predictive distribution has longer tail (expected since estimated ρ larger than true one)

NON-EXISTENCE OF PREDICTIVE MOMENTS

- Theorem (Wiper, 1998)

  Consider a $G/M/1$ queue with arrival rate λ and service rate μ; if a prior density $f(\lambda, \mu)$ such that $f(\lambda, \mu = \lambda) > 0$ is used, then the expected queueing time in equilibrium does not exist

- Using Little's theorems, the result can be generalized to N, W and so on

Bayesian estimation in stochastic predator-prey models

Fabrizio Ruggeri, Sara Pasquali[1] and Gianni Gilioli[2]

[1]Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/~fabrizio/

[2]Dipartimento di Scienze Biomediche e Biotecnologie
Università di Brescia - Italy

OUTLINE

- Description of the problem: *Phytoseiulus persimilis* (predator) vs. *Tetranychus urticae* (prey) in a strawberry field in Sicily
- Interest in *functional response*, i.e. outcome of stochastic search and capture processes
- Stochastic model with linear functional response: Lotka-Volterra
- Stochastic model with nonlinear functional response: Ivlev
- Current research

INTEGRATED PEST MANAGEMENT

- IPM programs implemented to
  - minimise losses due to plant pests
  - control insect vectors of important plant, animal, and human diseases
  - reduce impact of control techniques on environment and human health
- Biological control techniques, based on manipulation of inter-specific relationships (use of predators, parasites, diseases and plant resistance to suppress pest populations)
- Population dynamics models can provide insight into the complexity of interacting populations systems under the effect of control operations and contribute to the evaluation and improvement of IPM tactics and strategies

BIOLOGICAL CONSIDERATIONS

- **Goal:** Study of functional response (i.e. response of predator to prey abundance [consumption rate of a single predator]) in an acarine predator-prey system
- Functional response is system specific (i.e. depends on predator, prey and environment)
- Analysis of functional response usually performed in laboratory or in simple arenas
- Do lab experiments represent the behaviour of functional response in natural systems?
- Extension to natural systems should consider
  - scale problems: different animal behaviour in small environments w.r.t. large ones
  - characteristics of environment organisation (plants layout might affect predator-prey interaction)
  - degree of artificiality introduced by the experimental setup
- Increasing interest on field observations and analysis of time series

ASSUMPTIONS

- closed system (no immigration and/or emigration)
  $\Rightarrow$ only local dynamics considered

- system protected against interference of factors not represented in the model (e.g. presence of other alternative preys or competitive predators)

- functional response unaffected by abiotic variations (temperature, humidity, etc.)

- interest in a single cycle of the population and not in the long term behaviour (i.e. in the prompt control of the prey by the predator and not in the persistence of the predator)

- knowledge of all biodemographic parameters characterising prey and predator populations

CHOICE OF THE MODEL

Selection of an appropriate model for the functional response is critical

- Models should mimic the qualitative behaviour of predator-prey populations system, i.e. admitting stable solutions with prey extinction or an equilibrium
- Models may be prey dependent or depend on both prey and predator densities (e.g. their ratio, like in Arditi and Ginzburg, 1989), and take into account different behavioural and physiological aspects
- Parameter estimation for the functional response is the most problematic aspect because of the complexity of behavioural and physiological responses related to the predation process occurring in a heterogeneous environment
- Plant dynamics not taken into account by the model that only focuses on interaction between prey and predator
- Behavioural, demographic and environmental stochasticities introduced in the model through noise terms affecting each population as well as their interaction

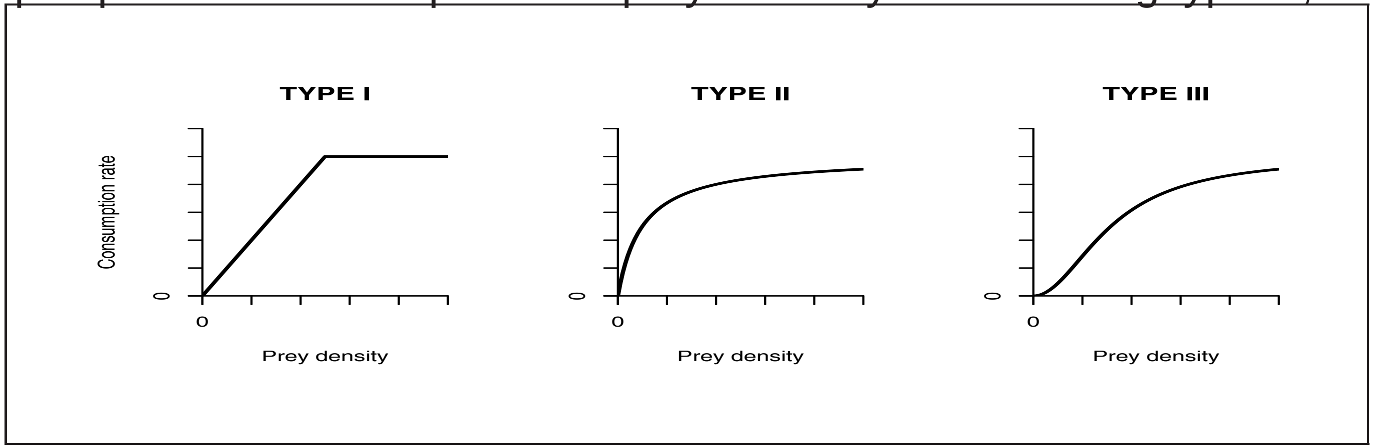
CHOICE OF THE MODEL

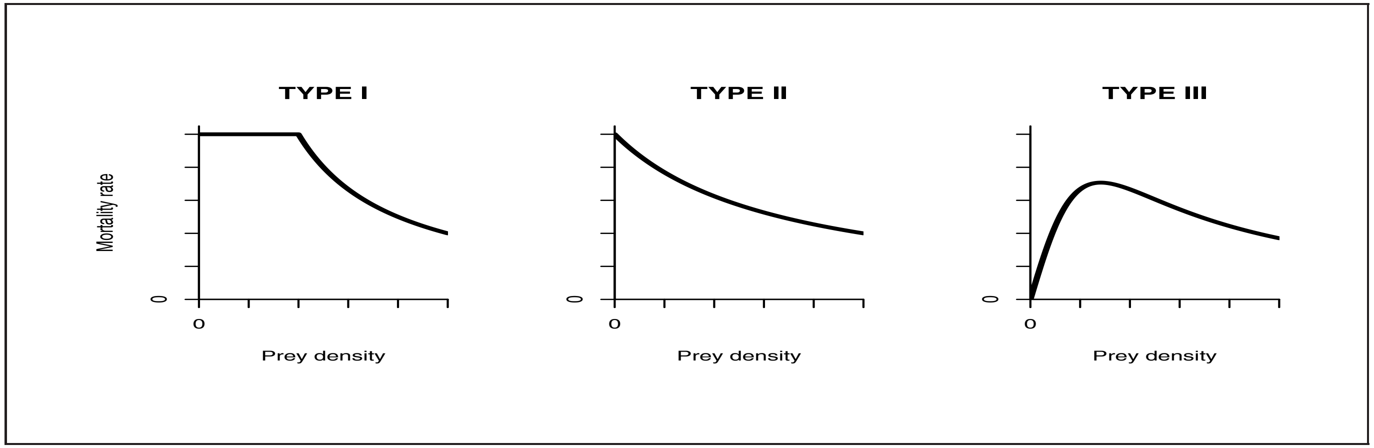
Factors affecting the functional response

- Prey density
- Prey behaviour (more defensive at high density)
- Predator density (conflict among predators at high density)
- Spatial distribution of prey and predator
- Limited consumption capacity for each predator
- Environmental factors (e.g. weather, interactions with humans and/or other animals, predators of the predator)

CHOICE OF THE MODEL

Per capita predator consumption and prey mortality rates: Holling types I, II and III





Pictures from Asseburg (2005)

CHOICE OF THE MODEL

We are not going to consider spatial features explicitly, unlike Nachman (1987) who performed many computer simulations for a non-stochastic model, using a mix of data about an acarine predator-prey system obtained from a laboratory or collected in three glasshouses over a 6 months period. He found out that the population dynamics of an acarine predator-prey system is influenced by spatial features such as

- the number of plants
- short and long distance dispersal (i.e. movement away from an existing population)
- degree of spatial association between the two mite populations within plants

At the local level, he proved that long-term balance between extinctions and colonisations can be achieved if

- the predator efficiency is not too high
- the dispersal rates of both prey and predators are low or moderate
- only a small fraction of the migrants perform long-distance dispersal (depending on, e.g. the frequency of plant/human contacts)

DETERMINISTIC MODEL

$$\begin{cases} dx_t = [r x_t G(x_t) - y_t F(x_t, y_t; q)]\, dt & x(0) = x_0 \\ dy_t = [c y_t F(x_t, y_t; q) - u y_t]\, dt & y(0) = y_0 \end{cases}$$

- x_t normalised biomass of prey
- y_t normalised biomass of predator
- r = specific growth rate of the prey
- c = specific production rate of the predator
- u = specific loss rate of predator
- q = efficiency of the predation process
- $G(x)$ = growth of the prey in absence of predators
- $F(x, y; q)$ = functional response of the predator to the prey abundance

LOTKA-VOLTERRA SYSTEM

- $G(x) = 1 - x$: overcrowding penalises prey growth in absence of predators
- $F(x, y; q) = qx$: efficacy of predation proportional to predation efficiency and prey abundance (but not dependent on predator abundance, as suggested by some authors)

$\Rightarrow$ Deterministic system becomes

$$\begin{cases} dx_t = [rx_t (1 - x_t) - qx_t y_t]\, dt \\ dy_t = [cqx_t y_t - uy_t]\, dt \end{cases}$$

- Main limit: no saturation of the predator when the ingested prey increases
- Main advantage: the model is simple and limits the number of parameters to be taken into account

LOTKA-VOLTERRA SYSTEM

$$\begin{cases} dx_t = [rx_t(1 - x_t) - qx_ty_t]\,dt \\ dy_t = [cqx_ty_t - uy_t]\,dt \end{cases}$$

Three steady state solutions (Buffoni et al., 1995)

- the null state $E_0 = (0, 0)$, always unstable (small changes in E_0 may imply large changes in the system)
- the noncoexistence state $E_1 = (1, 0)$, stable if $q < \frac{u}{c}$
- the coexistence state $E^* = (\frac{u}{qc}, \frac{r}{q}(1 - \frac{u}{qc}))$ exists and is stable iff $q > \frac{u}{c}$

(c production rate of predator, u loss rate of predator, q predation efficiency)

STOCHASTIC MODEL

q subject to noise and dependent on time $\Rightarrow q_t = q_0 + \sigma\xi_t$

- σ positive constant
- q_0 unknown parameter **to be estimated**
- ξ_t Gaussian white noise process

$\Rightarrow$ (early) stochastic model

$$\begin{cases} dx_t = [rx_t(1-x_t) - q_0 x_t y_t]\,dt - \sigma x_t y_t dw_t^{(1)} \\ dy_t = [cq_0 x_t y_t - u y_t]\,dt + c\sigma x_t y_t dw_t^{(1)} \end{cases}$$

$w_t^{(1)}$: Wiener process affecting the prey-predator interaction $x_t y_t$ in the system

$\Rightarrow$ demographic stochasticity (i.e. variability in population growth rates due to differences among individuals in survival and reproduction)

STOCHASTIC MODEL

Environmental stochasticity affects prey and predator (i.e. different birth and death rate in different period because of weather, diseases, etc.) and sampling error affecting population abundance estimates

- $w_t^{(2)}$: Wiener process independent of $w_t^{(1)}$

- ε and ρ positive parameters

$\Rightarrow$ stochastic model

$$\begin{cases} dx_t = [rx_t(1-x_t) - q_0 x_t y_t]dt - \sigma x_t y_t dw_t^{(1)} + \varepsilon x_t dw_t^{(2)} \\ dy_t = [cq_0 x_t y_t - u y_t]dt + c\sigma x_t y_t dw_t^{(1)} + \rho y_t dw_t^{(2)} \end{cases}$$

Solutions not necessarily in the compact $[0,1] \times [0,1]$

STOCHASTIC MODEL

Function $\chi(z)$

- continuously differentiable and Lipschitz
- equal to 1 in the compact $[\eta, 1-\eta]$
- decreasing in $(1-\eta, +\infty)$
- increasing in $(-\infty, \eta)$
- $\lim_{z\to-\infty} \chi(z) = \lim_{z\to+\infty} \chi(z) = 0$
- $\chi(0) = \chi(1) = \eta$

$\Rightarrow$ (final) stochastic model

$$\begin{cases} dx_t &= [rx_t(1-x_t) - q_0 x_t y_t]\chi(x_t)dt - \sigma x_t y_t \chi(x_t) dw_t^{(1)} + \varepsilon x_t \chi(x_t) dw_t^{(2)} \\ dy_t &= [cq_0 x_t y_t - u y_t]\chi(y_t)dt + c\sigma x_t y_t \chi(y_t) dw_t^{(1)} + \rho y_t \chi(y_t) dw_t^{(2)} \end{cases}$$

STOCHASTIC MODEL

(Bivariate) diffusion process

$$dX_t = \mu\left(X_t, q_0\right) dt + \beta\left(X_t\right) dW_t,\ X_0 = x_0, t \geq 0$$

- $X_t = [x_t, y_t]^T$
- Drift coefficient: $\mu\left(X_t, q_0\right) = \begin{bmatrix} \left[rx_t(1-x_t) - q_0 x_t y_t\right]\chi(x_t) \\ \left[cq_0 x_t y_t - uy_t\right]\chi(y_t) \end{bmatrix}$
- Diffusion coefficient: $\beta\left(X_t\right) = \begin{bmatrix} -\sigma x_t y_t \chi(x_t) & \varepsilon x_t \chi(x_t) \\ c\sigma x_t y_t \chi(y_t) & \rho y_t \chi(y_t). \end{bmatrix}$

Coefficients μ and β

- bounded and continuously differentiable w.r.t. X_t and q_0
- satisfy conditions for existence and uniqueness of a strong solution of a stochastic differential equation (see e.g. Øksendal, 1998)

LIKELIHOOD

- Log-likelihood $\log L(q_0) =$

  $-\int_0^T \mu^T(X_t, q_0)\left[\beta(X_t)\beta^T(X_t)\right]^{-1} dX_t + \frac{1}{2}\int_0^T \mu^T(X_t, q_0)\left[\beta(X_t)\beta^T(X_t)\right]^{-1} \mu(X_t, q_0)\, dt$

- q_0 linear in the model $\Rightarrow$ μ decomposed into $\mu(X_t; q_0) = a(X_t)q_0 + b(X_t)$, with

  $a(X_t) = \begin{bmatrix} -x_t y_t \chi(x_t) \\ c x_t y_t \chi(y_t) \end{bmatrix} \quad ; \quad b(X_t) = \begin{bmatrix} r x_t(1 - x_t)\chi(x_t) \\ -u y_t \chi(y_t) \end{bmatrix}$

- Score function $S(q_0)$(i.e. derivative of $\log L(q_0)$ w.r.t. q_0) $=$

  $-\int_0^T a^T(X_t)\left[\beta(X_t)\beta^T(X_t)\right]^{-1} dX_t + \int_0^T a^T(X_t)\left[\beta(X_t)\beta^T(X_t)\right]^{-1} \mu(X_t; q_0)\, dt$

- $\hat{X} = (X_0, X_1, ..., X_p)$ observations at times $t_0, t_1, ..., t_p$

- Discretised score function $S_N(q_0)$

  $= -\sum_{i=1}^{p} a^T(X_{i-1})\left[\beta(X_{i-1})\beta^T(X_{i-1})\right]^{-1}\left[X_i - X_{i-1} - \mu(X_{i-1})\Delta_i\right]$

  approximates well the continuous score function for small intervals $\Delta_i = t_i - t_{i-1}$

ESTIMATION

Given $\widehat{X} = (X_0, X_1, ..., X_p)$ at times $t_0, t_1, ..., t_p \Rightarrow$ replace the system

$$dX_t = \mu\left(X_t, q_0\right) dt + \beta\left(X_t\right) dW_t,\ X_0 = x_0, t \geq 0$$

with the Euler-Maruyama approximation, given by

$$X_{t_{i+1}} = X_{t_i} + \mu\left(X_{t_i}, q_0\right) \Delta_{t_i} + \beta\left(X_{t_i}\right)\left(W_{t_{i+1}} - W_{t_i}\right),$$

where $\Delta_{t_i} = t_{i+1} - t_i$.

The approximation is much better when Δ_{t_i}'s are *small*.

ESTIMATION

Maximum likelihood estimator (for discrete sampled data)

$$\widehat{q}_{0,p} = \frac{1}{T}\sum_{i=1}^{p} \frac{1}{\rho y_{i-1} + c\varepsilon x_{i-1}} \left[-\frac{\rho\,(x_i - x_{i-1})}{x_{i-1}\chi(x_{i-1})} + r\rho(1 - x_{i-1})\Delta_i + \frac{\varepsilon(y_i - y_{i-1})}{y_{i-1}\chi(y_{i-1})} + \varepsilon u \Delta_i \right]$$

Drawback: MLE consistent and asymptotically Gaussian only if time between 2 observations small w.r.t. total observation time (here very few observations!)

Bayesian estimator (posterior mean or median)

Posterior $\pi\left(q_0 | \widehat{X}\right) \propto \pi(q_0) \prod_{i=1}^{p} f\left(X_i | X_{i-1}, q_0\right)$

- prior distribution $\pi(q_0)$
- $f\left(X_i | X_{i-1}, q_0\right) \propto \left| \left[\beta\left(X_{i-1}\right)\beta^T\left(X_{i-1}\right)\right]^{-1} \right|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2}\left[X_i - X_{i-1} - \mu(X_{i-1}, q_0)\Delta_i\right]^T \cdot \right.$

  $\left. \cdot \left[\Delta_i\, \beta\left(X_{i-1}\right)\beta^T\left(X_{i-1}\right)\right]^{-1} \left[X_i - X_{i-1} - \mu(X_{i-1}, q_0)\Delta_i\right] \right\}$

BAYESIAN ESTIMATION

Priors on q_0

- Prior $\mathcal{N}\left(\mu_I, \sigma_I^2\right)$

  $\Rightarrow$ posterior $\mathcal{N}\left(\frac{\sigma_I^2\mu_X+\mu_I\sigma_X^2}{\sigma_I^2+\sigma_X^2}, \frac{\sigma_X^2\sigma_I^2}{\sigma_I^2+\sigma_X^2}\right)$, with $\mu_X = \widehat{q}_{0,p}$ and $\sigma_X^2 = \frac{\sigma^2}{T}$

  $\Rightarrow$ Bayesian estimator $\frac{\sigma_I^2\mu_X+\mu_I\sigma_X^2}{\sigma_I^2+\sigma_X^2}$, i.e. weighted average of MLE and prior mean

- Improper prior $\pi\left(q_0\right) \propto I_{(0,\infty)}(q_0)$

  $\Rightarrow$ posterior $\frac{1}{\sqrt{2\pi\sigma_X^2}\left[1-\Phi\left(-\frac{\mu_X}{\sigma_X}\right)\right]} e^{-\frac{1}{2\sigma_X^2}(q_0-\mu_X)^2} I_{(0,\infty)}(q_0)$

- Gamma prior (more later)

MCMC AND DATA AUGMENTATION

Problem: discrete observations and Euler approximation of SDE $\Rightarrow$ use sufficiently large number of (real and latent) data to ensure arbitrarily small discretization bias (see Elerian et al., 2001, for a discussion)

M latent data generated (using Euler-Maruyama discretisation) between two consecutive observations

Matrix of all data

$$Y = \begin{pmatrix} x(t_0) & x^*(t_1) & \ldots & x^*(t_M) & x(t_{M+1}) & \ldots & x^*(t_{n-1}) & x(t_n) \\ y(t_0) & y^*(t_1) & \ldots & y^*(t_M) & y(t_{M+1}) & \ldots & y^*(t_{n-1}) & y(t_n) \end{pmatrix}$$

$n = (p-1)M + p$ total number of observations

- $X_i = (x(t_i), y(t_i))$ the real datum at time t_i
- $X_i^* = (x^*(t_i), y^*(t_i))$ the latent datum at t_i
- Y_i both a real and a latent datum at time t_i

GENERATION OF LATENT DATA

(Minor) novelty: random block size (as in Elerian et al., 2001) for multivariate models (as in Golightly and Wilkinson, 2005)

1st step: we generate latent data by means of a linear interpolation

s**-th iteration:** we generate the latent datum X_i^* from the conditional distribution $\pi(X_i^*|Y_{i-1}, Y_{i+1}; q_0)$ where Y_{i-1} is obtained at iteration s and Y_{i+1} at iteration $s-1$.

$$\pi\left(X_i^*|Y_{i-1}, Y_{i+1}; q_0\right) \propto$$

$$exp\left\{-\frac{1}{2}\left[X_i^* - Y_{i-1} - \mu(Y_{i-1}; q_0)\Delta t\right]^T \left(\Delta t\beta(Y_{i-1})\beta^T(Y_{i-1})\right)^{-1}\right.$$

$$\left.\left[X_i^* - Y_{i-1} - \mu(Y_{i-1}; q_0)\Delta t\right]\right\}$$

$$exp\left\{-\frac{1}{2}\left[Y_{i+1} - X_i^* - \mu(X_i; q_0)\Delta t\right]^T \left(\Delta t\beta(X_i^*)\beta^T(X_i^*)\right)^{-1}\right.$$

$$\left.\left[Y_{i+1} - X_i^* - \mu(X_i^*; q_0)\Delta t\right]\right\}.$$

BLOCK SIZE SAMPLING

Fixed or random block size m from $M-1 \sim Poisson(\lambda)$

Update latent observations in block of length m (Elerian et al., 2001)

$X^*_{(k,m)} = \left(X^*_k, X^*_{k+1}, ..., X^*_{k+m-1}\right)$

$$f\left(X^*_{(k,m)}|Y_{k-1}, Y_{k+m}; q_0\right) \propto \prod_{j=k}^{k+m-1} \pi\left(X^*_j|Y_{j-1}, Y_{k+m}; q_0\right)$$

Simulation from $f\left(X^*_{(k,m)}|Y_{j-1}, Y_{k+m}; q_0\right)$, i.e. depending on 2 observations: one before the datum and other after the block. M-H algorithm with proposal density proportional to

$$\prod_{j=k}^{k+m-1} \mathcal{N}^t\left(\frac{1}{2}\left(Y_{j-1} + Y_{k+m}\right), \frac{1}{2}\Delta t \cdot \beta\beta^T\left(Y_{j-1}\right)\right)$$

$\mathcal{N}^t$: truncated normal distribution proposed by Eraker (2001) and used also by Golightly and Wilkinson (2005)

SIMULATION OF q_0

- Gamma prior on q_0
- At each step of the MCMC simulation
  - after simulation of $Y \Rightarrow$ simulation of q_0
  - M-H to generate q_0 from $\pi(q_0|Y) \propto \pi(q_0) \prod_{i=1}^{n} \pi(Y_i|Y_{i-1}, q_0)$
    $\Rightarrow$ gamma proposal
- *Usual* checks
  - Burn-in
  - Decorrelation, acceptance probability, mixing properties
  - Diagnostic tests for convergence

APPLICATION

Interaction between prey *Tetranychus urticae* and predator *Phytoseiulus persimilis*

$$\begin{cases} dx_t &= [rx_t(1-x_t) - q_0 x_t y_t]\chi(x_t)dt - \sigma x_t y_t \chi(x_t) dw_t^{(1)} + \varepsilon x_t \chi(x_t) dw_t^{(2)} \\ dy_t &= [cq_0 x_t y_t - u y_t]\chi(y_t)dt + c\sigma x_t y_t \chi(y_t) dw_t^{(1)} + \rho y_t \chi(y_t) dw_t^{(2)} \end{cases}$$

$$x(0) = x_0 \quad y(0) = y_0$$

- $r = 0.11 \quad c = 0.35 \quad u = 0.09$ (Buffoni and Gilioli, 2003)

- $\hat{\sigma} = 0.321, \hat{\varepsilon} = 0.079, \hat{\rho} = 0.106, \hat{q}_0 = 2.4767$ (estimated through least squares after extensive survey of 8 separated local predator-prey dynamics)

- Gamma prior on q_0: mean = 2.4767 and variance = 2

- Block size

  – (initial): $M - 1$ Poisson with $\lambda = 3$, but rejected because of slow convergence and poor mixing

  – (final) M equal to number of latent data between points

APPLICATION

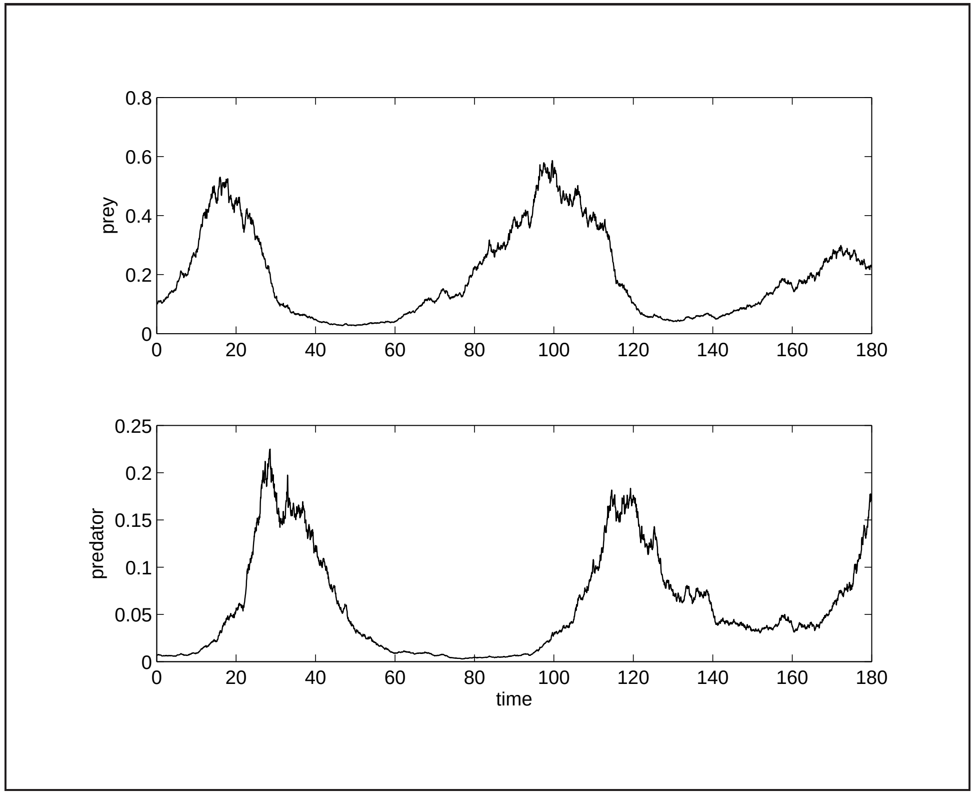
Two cases

- simulated data (to test method performance)
- field data

SIMULATED DATA

- Initial conditions: $x_0 = 0.1 \quad y_0 = 0.007$
- Total time: $T = 180$ days
- $q_0 = 1.5$
- 10 data simulated from the system
- MLE compared with posterior median (preferred to posterior mean)
- MLE: 1.7176
- Posterior median without latent data: 1.7174 (very close to MLE)
- Posterior median for 2 latent data: 1.5656
- (Visual) inspection of trajectories $\Rightarrow$ MLE underestimates maximum values of prey and predators and display anticipated cycles
- Improper prior leads to similar posterior

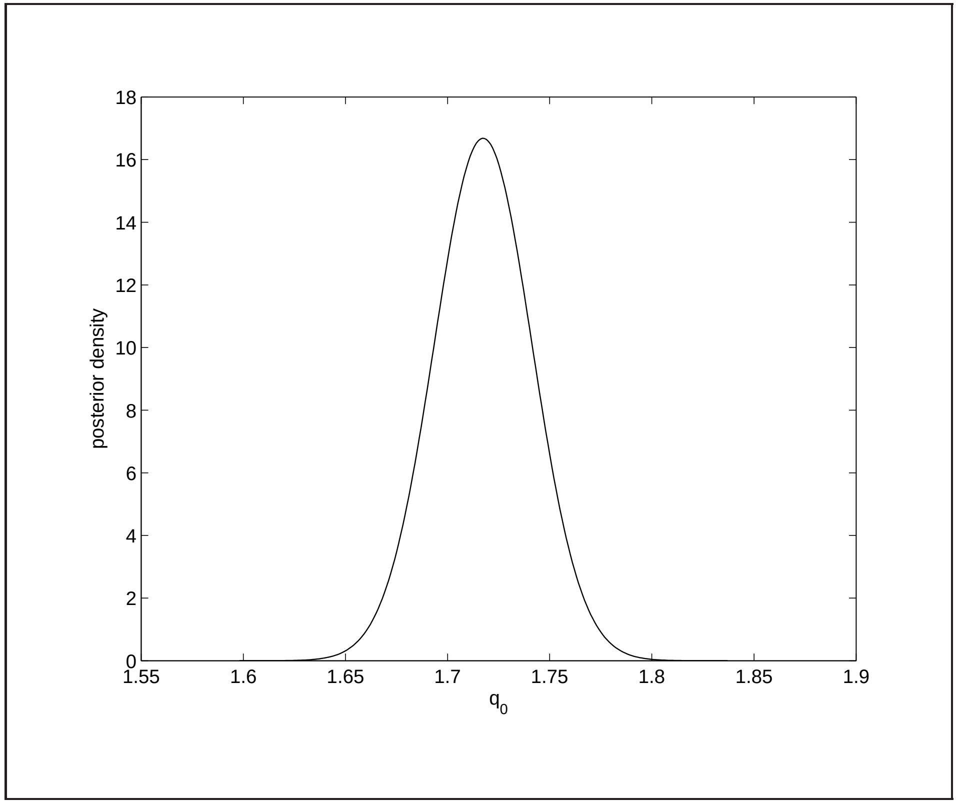
DYNAMICS OF THE SYSTEM



Prey and predator biomass as function of time

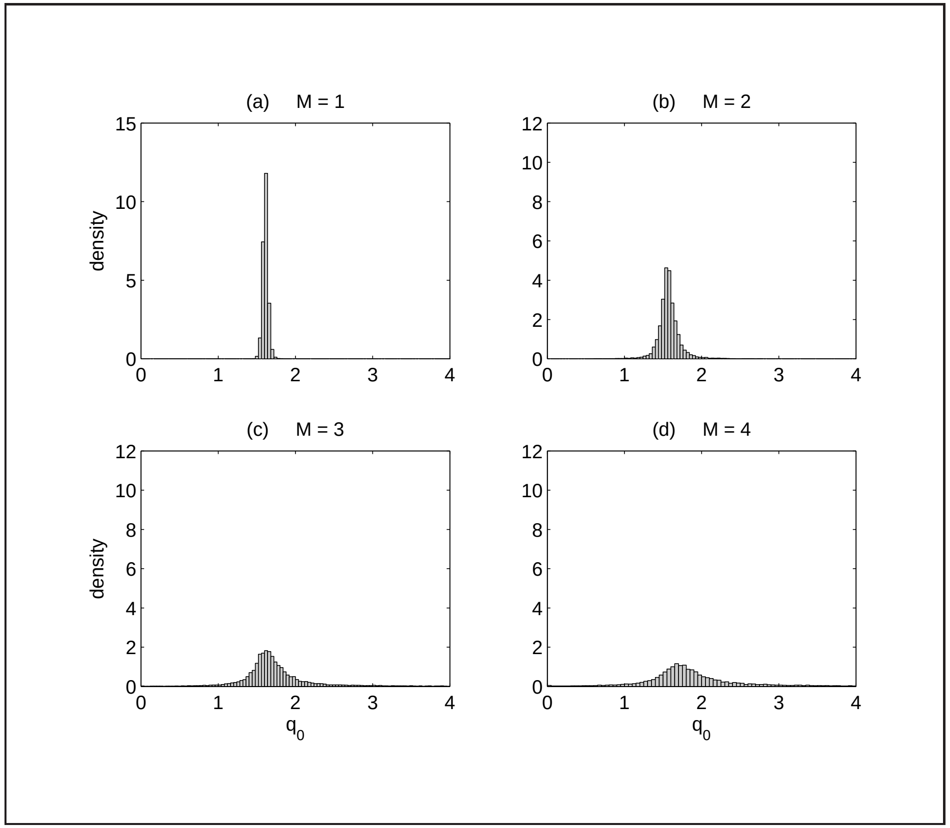
($x_0 = 0.1, y_0 = 0.007$, $q_0 = 1.5$, $T = 180$)

NO LATENT DATA



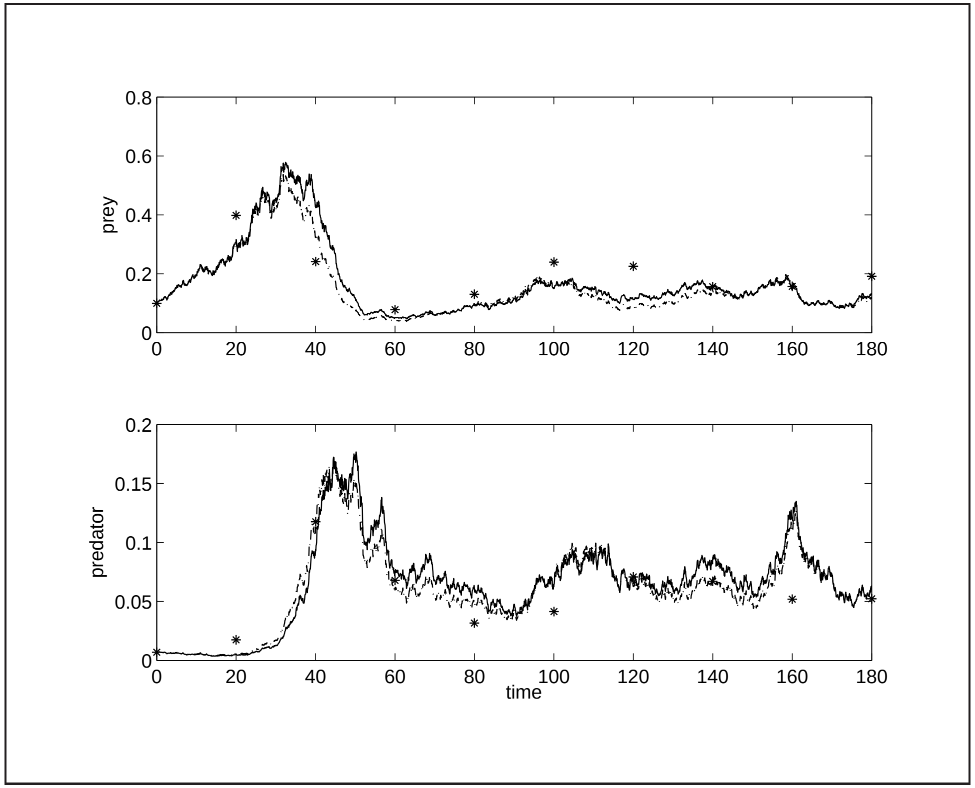
Posterior density with median 1.7174, obtained using a gamma prior

LATENT DATA



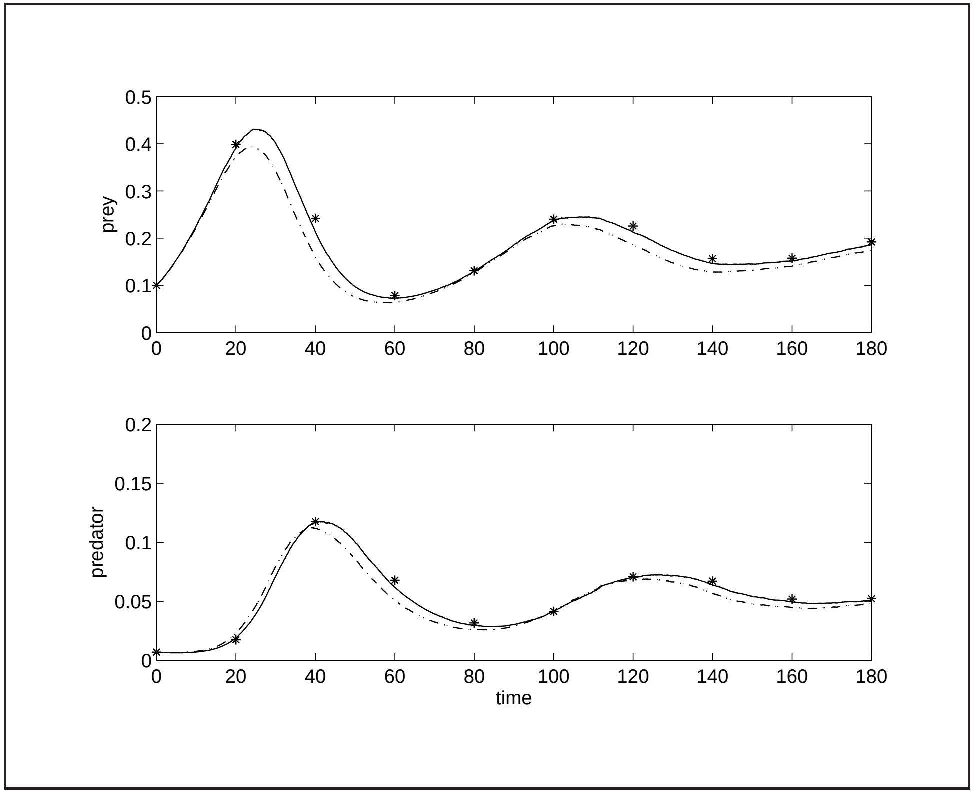
Gamma prior and MCMC with 100000 simulations. Latent data: (a) 1 with median 1.6114; (b) 2 with median 1.5656; (c) 3 with median 1.6737; (d) 4 with median 1.8070

LATENT DATA



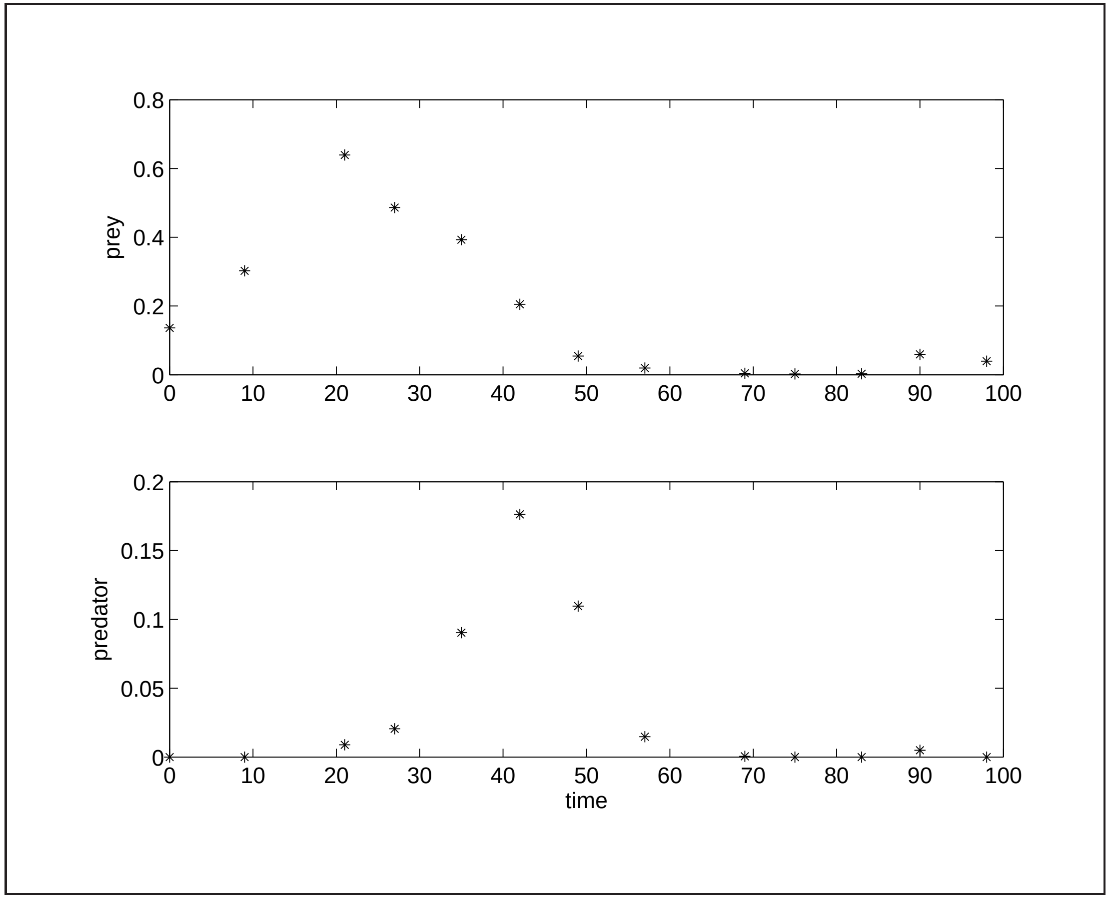
Simulated trajectory for prey and predator biomass as function of time, for initial conditions $x_0 = 0.1, y_0 = 0.007$, obtained with MLE $q_0 = 1.7176$ (dashed) and Bayesian estimate $q_0 = 1.5656$ (continuous). Asterisks denote simulated data

LATENT DATA



Mean over 1000 trajectories of prey and predator biomass as function of time, for initial conditions $x_0 = 0.1, y_0 = 0.007$, obtained with MLE $q_0 = 1.7176$ (dashed) and Bayesian estimate $q_0 = 1.5656$ (continuous). Asterisks denote simulated data

FIELD DATA

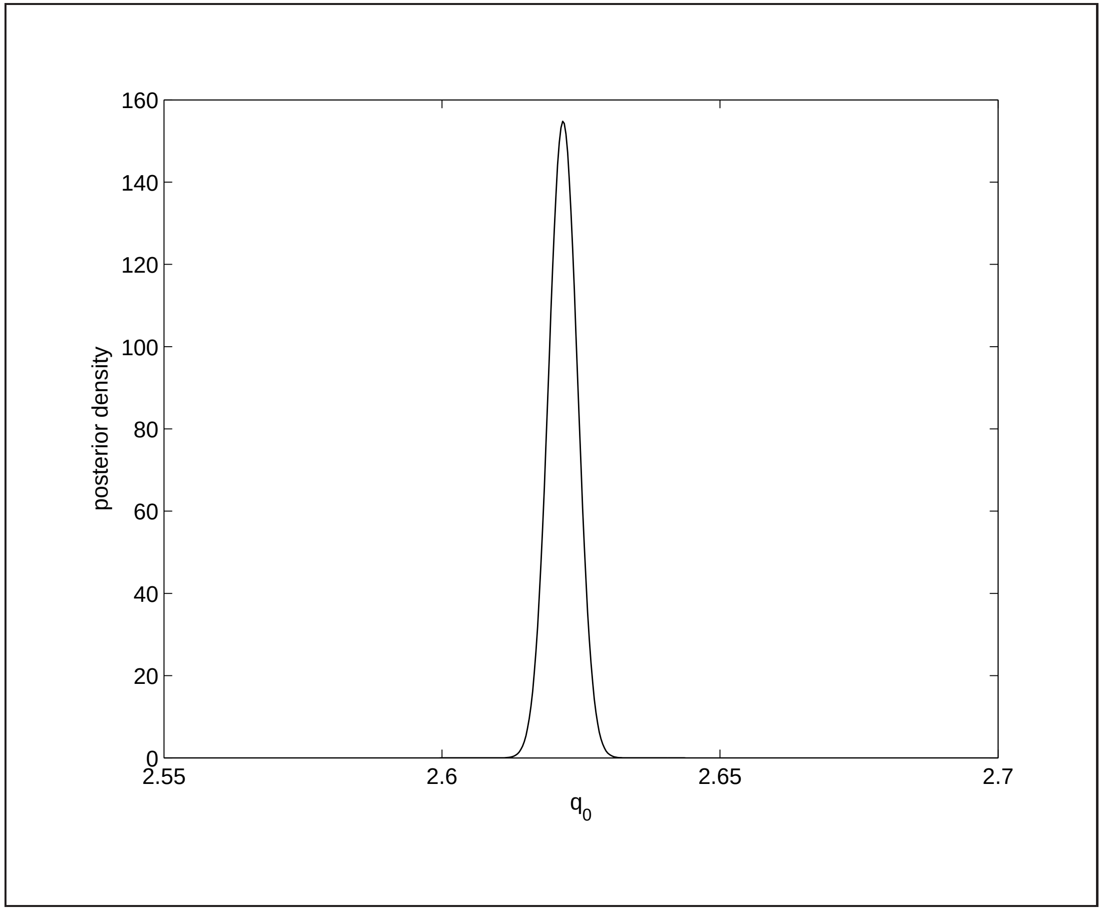


13 data collected in a one-hectare strawberry crop in Ispica (Ragusa, Italy)

FIELD DATA

- Data from intensive sampling of 12 separated local predator-prey dynamics
- To minimise the sampling efforts, only adult mites were collected to avoid leaves collection and laboratory observation of larvae
- Only first cycle of the prey (48 *days*) $\Rightarrow$ no interest in predator persistence $\Rightarrow$ reject data 8-13
- Rarity conditions strongly affecting functional response $\Rightarrow$ reject data 1
- MLE: 2.6218, very close to posterior median with no latent data (and gamma prior with mean 2.4767 and variance 2 from a different extensive survey of 8 separated local predator-prey dynamics)

NO LATENT DATA



Posterior density with gamma prior

INDEXES

No knowledge about *true* q_0
$\Rightarrow$ comparison between estimates using ad hoc indexes on prey population dynamics

- M: maximum size of the population
  $\Rightarrow$ related to impact (damage) of the prey population on the plants

- T_{max}: time to reach maximum size
  $\Rightarrow$ related to capability of the predator to restrain prey population exponential growth

- $T_{0.5}$ (or $T_{0.1}$): time to halve (or reduce to one tenth) maximum size
  $\Rightarrow$ related to capability of the predator to cause a rapid decrease of the prey population

- I: integral of the population up to $T_{0.1}$
  $\Rightarrow$ measurement of population pressure on the resource

Similar definitions for predator population dynamics

M and I most important indexes for biologists

INDEXES: PREY

Lat. data	post. median	M	T_{max} (dd)	$M_{0.5}$	$T_{0.5}$ (dd)	$M_{0.1}$	$T_{0.1}$ (dd)	I
1	2.03	.610	22.93	.305	30.67	.0610	39.26	14.09
2	1.94	.613	23.72	.306	31.23	.0613	40.47	14.53
3	1.90	.615	23.72	.307	31.49	.0614	41.03	14.73
4	1.86	.616	23.72	.308	31.78	.0616	41.68	14.95
5	1.83	.617	23.72	.309	32.01	.0617	42.30	15.15
6	1.82	.618	24.07	.309	32.11	.0618	42.56	15.24
7	1.80	.618	24.07	.309	32.21	.0618	42.76	15.31
8	1.80	.618	24.07	.309	32.21	.0618	42.69	15.28
9	1.81	.618	24.07	.309	32.14	.0618	42.43	15.26
10	1.82	.617	24.07	.309	32.08	.0618	42.47	15.20
11	1.83	.617	23.72	.309	32.01	.0617	42.30	15.15
12	1.83	.617	23.72	.309	32.98	.0617	42.27	15.14
13	1.84	.617	23.72	.308	32.88	.0617	41.98	15.05
15	1.85	.616	23.72	.308	32.85	.0616	41.85	15.01
20	1.87	.615	23.72	.308	32.69	.0616	41.49	14.88
MLE	2.62	.601	21.89	.300	28.03	.0601	33.87	12.09

INDEXES: PREDATOR

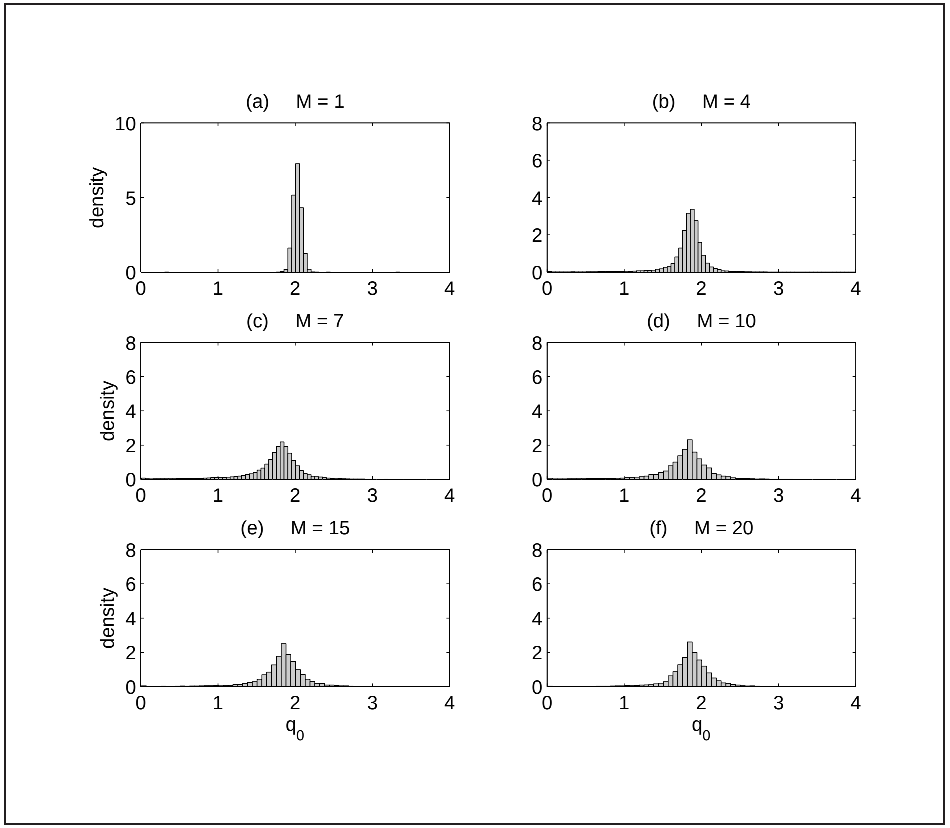
Lat. data	post. median	M	T_{max} (dd)	$M_{0.5}$	$T_{0.5}$ (dd)	$M_{0.1}$	$T_{0.1}$ (dd)	I
1	2.03	.168	34.69	.0842	47.33	.0168	70.23	3.47
2	1.94	.168	34.69	.0839	48.25	.0168	72.55	3.51
3	1.90	.167	35.31	.0837	48.71	.0167	73.73	3.53
4	1.86	.167	35.57	.0836	49.20	.0167	75.13	3.54
5	1.83	.167	36.00	.0835	49.62	.0167	76.34	3.56
6	1.81	.167	36.00	.0834	49.78	.0167	77.09	3.56
7	1.80	.167	36.42	.0834	49.95	.0167	77.78	3.57
8	1.80	.167	36.42	.0834	49.91	.0167	77.58	3.57
9	1.81	.167	36.42	.0834	49.85	.0167	77.32	3.57
10	1.82	.167	36.00	.0834	49.72	.0167	76.73	3.56
11	1.83	.167	36.00	.0835	49.62	.0167	76.31	3.56
12	1.83	.167	36.00	.0835	49.59	.0167	76.31	3.56
13	1.84	.167	35.57	.0835	49.39	.0167	75.75	3.55
15	1.85	.167	35.57	.0835	49.29	.0167	75.43	3.55
20	1.87	.167	35.57	.0836	49.03	.0167	74.74	3.54
MLE	2.62	.174	31.36	.0869	43.02	.0174	62.72	3.23

INDEXES

Findings (compared with empirical data):

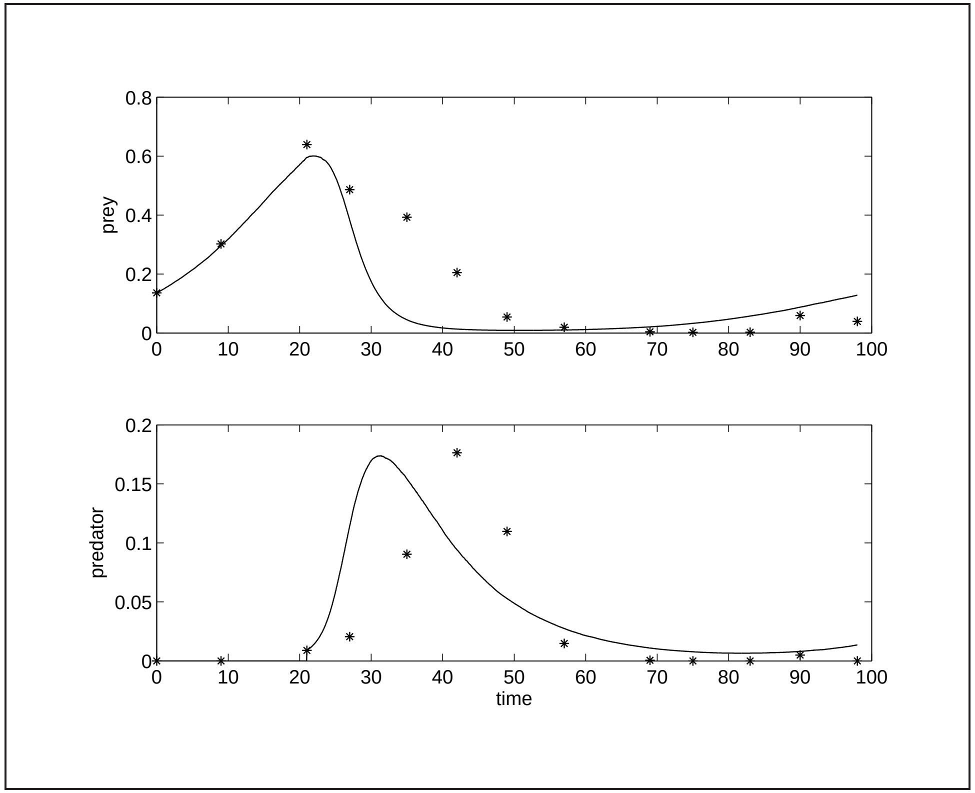
- Best I (integral of population) and M (maximum size of population) for MCMC with 7 latent data
- Differences between 5 to 12 latent data are negligible
- I quite similar for all latent data sizes
- MLE and Bayesian with no latent data good at approximating T (one day delay) but underestimate I and M
- I for predator best for MLE
- From the trajectories: T_{max} approximated well for prey but anticipated for predator (although M is fine)
- Sensitivity study: small variations in q_0 do not affect trajectories

LATENT DATA



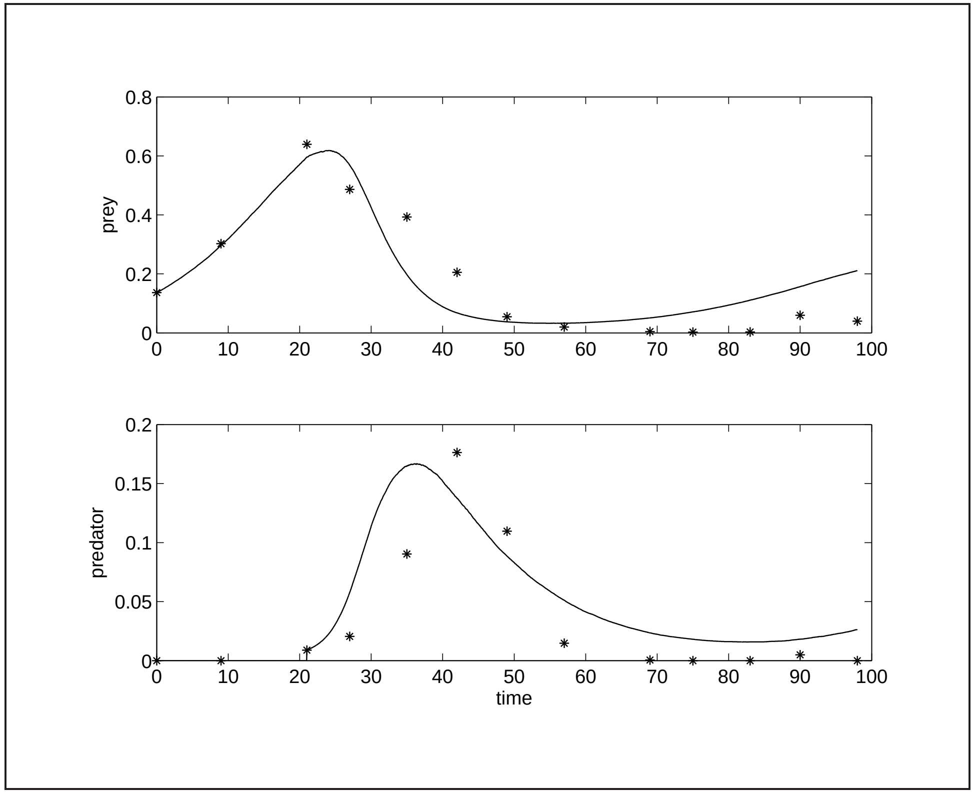
Gamma prior for q_0 and 100000 simulations in MCMC. (number of latent data, median):
(a) (1, 2.03) (b) (4, 1.86) (c) (7, 1.80) (d) (10, 1.82) (e) (15, 1.85) (f) (20, 1.8733)

LATENT DATA



Classical estimate. Continuous line: mean of 1000 trajectories of prey and predator for $q_0 = 2.6218$. Asterisks: field observations

LATENT DATA



MCMC estimate with 7 latent data between two consecutive observations. Continuous line: mean of 1000 trajectories of prey and predator for $q_0 = 1.8008$. Asterisks: field observations

LOTKA-VOLTERRA VS. IVLEV

- Lotka-Volterra provides the advantage of linearity at the cost of the simplified biological assumption of an unlimited predator per-capita consumption rate

- Ivlev overcomes the previous drawback and keeps just one parameter

- Ivlev displays qualitative properties that well interpret the behaviour of the predator-prey system of interest for biological control

- Like Lotka-Volterra, Ivlev presents three equilibrium points, characterised by extinction of both prey and predator, extinction of the predator, coexistence of prey and predator (under suitable condition for the parameters - see, e.g., Buffoni and Gilioli, 2003)

IVLEV MODEL

$$\begin{cases} dx_t = \left[rx_t(1-x_t) - by_t\left(1-e^{-q_0 x_t}\right)\right]dt + b\sigma y_t e^{-x_t} dw_t^{(1)} + \varepsilon x_t dw_t^{(2)} \\ dy_t = y_t\left[b'\left(1-e^{-q_0 x_t}\right) - u\right]dt - b'\sigma y_t e^{-x_t} dw_t^{(1)} + \rho y_t dw_t^{(2)} \end{cases}$$

- Initial conditions: $(x(0), y(0)) = (x_0, y_0)$
- x_t and y_t: biomass of prey and predator at time t per habitat unit (plant) normalised w.r.t. prey carrying capacity per habitat unit
- r: maximum specific growth rate of the prey
- b: maximum specific predation rate
- b': maximum specific predator production rate
- u: specific predator loss rate due to natural mortality
- q_0: measure of the efficiency of the predation process

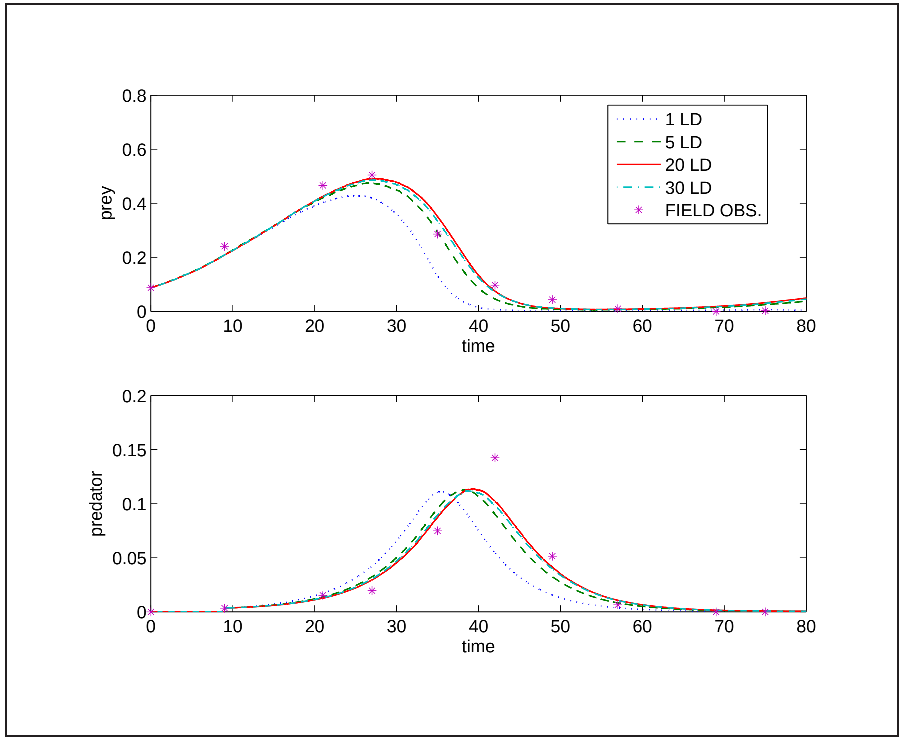
IVLEV MODEL

- Functional response $\left(1 - e^{-q_0 x_t}\right)$ of the deterministic term is subject to random fluctuations
- q_0 only affects the slope of the curve but not its shape. All the possible functional responses, obtained for different values of q_0, are similar for large values of x_t
- Consider an effect of random fluctuations more pronounced for small values of x_t and then decreasing for large values of x_t
- $\left(1 - e^{-q_0 x_t}\right) \Rightarrow \left(1 - e^{-q_0 x_t}\right) + \sigma e^{-\lambda x_t} \xi_t$
- Parameter λ allows to set the threshold at which the fluctuations become negligible (we put $\lambda = 1$)
- Here we estimate 4 parameters: $\theta = (q_0, \sigma, \varepsilon, \rho)$

MCMC FOR IVLEV MODEL

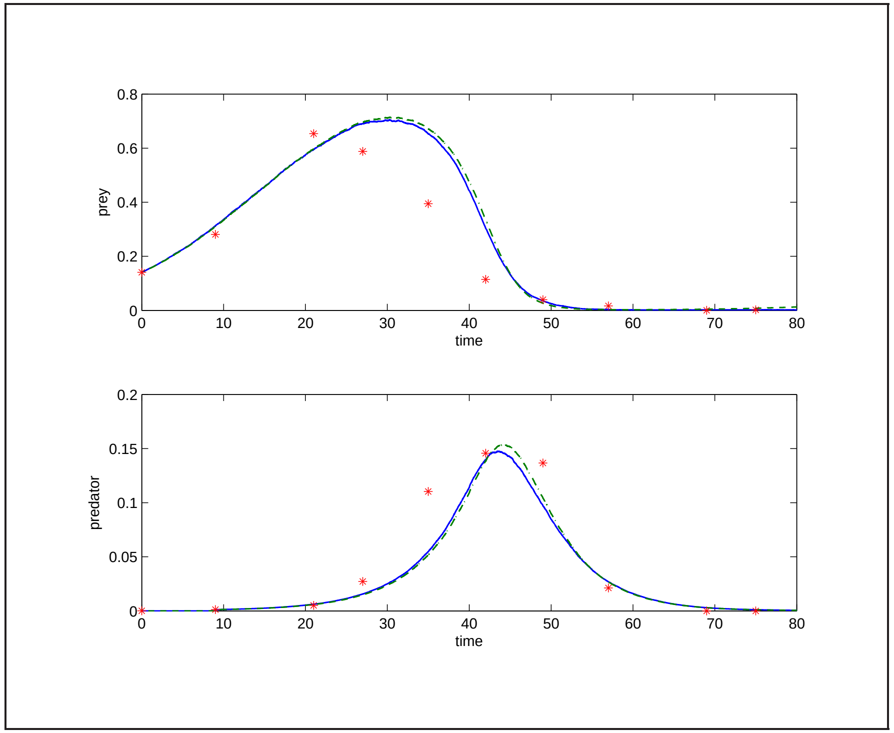
- MCMC with minor differences w.r.t. Lotka-Volterra about data augmentation (a different proposal distribution for latent data)
- Estimation of 3 further parameters induces various M-H steps
- As before, forecasting property of performed numerical experiments evaluated comparing simulated trajectories and observed biomass
- Data collected in 3 similar fields (A, B, and C)
- Parameters estimated from data in field A and posterior used to forecast trajectories in field B (same area, crop and agronomic conditions but slightly different initial conditions)
- Comparison of trajectories over field B obtained directly from data in B and from data in A

FORECASTS FOR FIELD A



Prey and predator biomass as function of time in field A, for different values of latent data. Trajectories as mean over 100 simulations and 50 values of q_0, σ, ε and ρ from posterior distributions. Asterisks denote field data

FORECASTS FOR FIELD B



Prey and predator biomass in field B. Trajectories as mean over 100 simulations and 50 values of q_0, σ, ε and ρ from posterior distributions for 20 latent data. $\star$ denotes field data. Continuous line: direct estimate on field B. Dashed line: predicted trajectories, starting from posterior distribution on A.

OTHER WORKS

- Generalist predators, consuming more than one prey and switching among them, induce multi-species functional responses; Negative binomial distribution for consumption of prey i at location j (see Asseburg, 2005)

- Bayesian nonparametrics (see Gaussian process for functional response in Patil, 2007)

- Boys, Wilkinson and Kirkwood (2008) proposed a Markov jump process s.t. in the interval $(t, t + dt)$ jumps are possible from (x_t, y_t) to $(x_t + 1, y_t)$, $(x_t - 1, y_t + 1)$ and $(x_t, y_t - 1)$, con probability linearly dependent on parameters θ_1, θ_2 and θ_3 for which gamma priors are taken. When data are discretely observed, they have to generate latent data as well (using MCMC, block updating and Poisson process approximation)

WHAT WE WOULD LIKE TO DO

- Consider other forms of functional response (e.g. Holling models)
- Introduction of a third equation (e.g. resources, i.e. evolution of plants)
- **More data to be collected in new planned experiments**
- Similar problems with covariates (e.g. olive tree fly at different stages and altitudes/temperature/humidity)

WHAT WE ARE DOING

DeAngelis-Beddington functional response $F(x, y; q) = \dfrac{qxy}{1 + k_1 x + k_2 y}$

$$\begin{cases} dx_t = \left[r x_t \left(1 - x_t\right) - \frac{q x_t y_t}{1 + k_1 x_t + k_2 y_t} \right] dt \\ dy_t = \left[c \frac{q x_t y_t}{1 + k_1 x_t + k_2 y_t} - u y_t \right] dt \end{cases}$$

- r = specific growth rate of the prey
- c = specific production rate of the predator
- u = specific loss rate of predator
- q = efficiency of the predation process
- k_1 = effect of handling time on feeding rate
- k_2 = magnitude of interference among predators

Joint work with Sara Pasquali, Laura Martin Fernandez and Ettore Lanzarone

WHAT WE ARE DOING

Heat equation

$$\rho \frac{\partial T(x,t)}{\partial t} - \lambda \frac{\partial^2 T(x,t)}{\partial x^2} = \dot{q}(x,t)$$

- Interest in thermal conductivity λ
- Approximation of second derivative with difference of first derivatives
- λ affected by noise
- Data from experiment (working on it ...) on a polymer slab

Joint work with Sara Pasquali and Ettore Lanzarone

BIBLIOGRAPHY

R. Arditi, L.R. Ginzburg (1999), Coupling in predator-prey dynamics: ratio-dependence, *Journal of Theoretical Biology*, **139**, 311-326.

C. Asseburg (2005), A Bayesian approach to modelling field data on multi-species predator-prey interactions, *Ph.D. Thesis*, University of St. Andrews, UK.

R.J. Boys, D.J.Wilkinson, T.B.L. Kirkwood (2008), Bayesian inference for a discretely observed stochastic kinetic model, *Statistics and Computing*, **18**, 125-135.

G. Buffoni, G. Di Cola, J. Baumgartener, V. Maurer (1995), A mathematical model of trophic interactions in an acarine predator-prey system, *Journal of Biological System*, **3**, 303-312.

G. Buffoni, G. Gilioli (2003), A lumped parameter model for acarine predator-prey population interactions, *Ecological Modelling*, **170**, 155-171.

O. Elerian, S. Chib, N. Shephard (2001), Likelihood inference for discretely observed nonlinear diffusions, *Econometrica*, **69**, 959-993.

BIBLIOGRAPHY

B. Eraker (2001), MCMC analysis of diffusion models with application to finance, *Journal of Business & Economic Statistics*, **19**, 177-191.

G. Gilioli, S. Pasquali, F. Ruggeri (2008), Bayesian inference for functional response in a stochastic predator-prey system, *Bulletin of Mathematical Biology*, **70**, 358-381.

G. Gilioli, S. Pasquali, F. Ruggeri (2012). Nonlinear functional response parameter estimation in a stochastic predator-prey model. *Mathematical Biosciences and Engineering*,**9**, 75-96.

A. Golightly, D.J. Wilkinson (2005), On Bayesian inference for stochastic kinetic models using diffusion approximations, *Biometrics*, **61**, 781788.

G. Nachman (1987), Systems analysis of acarine predator-prey interactions. II. The role of spatial processes in system stability, *Journal of Animal Ecology*, **56**, 267-281.

A. Patil (2007), Bayesian nonparametrics for inference of ecological dynamics, *Ph.D. Thesis*, University of California at Santa Cruz, USA.

Modelling bugs introduction during software testing

Fabrizio Ruggeri, Antonio Pievatolo[1] and Refik Soyer[2]

[1]Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/˜fabrizio/

[2]George Washington University
Washington D.C., USA

OUTLINE

- Models review
- Problem: possible introduction of new bugs
- Hidden Markov model
- Current research: Self-exciting process with latent variables

SOFTWARE RELIABILITY: MODELS

- Software reliability can be defined as *the probability of failure-free operation of a computer code for a specified mission time in a specified input environment*
- Seminal paper by Jelinski and Moranda (1972)
- More than 100 models after it (Philip Boland, *MMR2002*)
- Many models clustered in few classes
- Search for unifying models (e.g. Self-exciting process, Chen and Singpurwalla, 1997)

SOFTWARE RELIABILITY: MODELS

Most software reliability models fall into one of two categories (Singpurwalla and Wilson, 1994)

- [Type I]: models on times between successive failures based on
  - [Type I-1] failure rates (e.g. Jelinski-Moranda)
  - [Type I-2] inter-failure times as function of previous inter-failure times (e.g. random coefficient autoregressive model, Singpurwalla and Soyer, 1985)
- [Type II] models (counting processes) on observed number of failures up to time t (e.g. NHPP)

SOFTWARE RELIABILITY: MODELS

Failures at $T_1, T_2, \ldots, T_n$

Inter-failure times $T_i - T_{i-1} \sim \mathcal{E}(\lambda_i)$, independent, $i = 1, \ldots, n$

- $\lambda_i = \phi(N - i + 1)$, $\phi \in \mathbb{R}^+, N \in \mathbb{N}$, *(Jelinski-Moranda, 1972)*
  - Program contains an initial number of bugs N
  - Each bug contributes the same amount to the failure rate
  - After each observed failure, a bug is detected and corrected

Straightforward Bayesian inference with priors $N \sim \mathcal{P}(\nu)$ and $\phi \sim \mathcal{G}(\alpha, \beta)$

SOFTWARE RELIABILITY: MODELS

- $\lambda_i = \phi(N - p(i-1))$, $\phi \in \mathbb{R}^+, N \in \mathbb{N}$, $p \in [0, 1]$,
  (Goel and Okumoto, 1978)
  - Imperfect debugging

- $\lambda_i = \phi\delta^i$, $\phi \in \mathbb{R}^+$, $\delta \in (0, 1)$, *(Moranda, 1975)*
  - Failure rate (geometrically) decreasing

Failure rate constant between failures; different from

- $h(t) = \dfrac{\alpha}{\beta_0 + \beta_1 i + (t - t_{i-1})}$, $t \in (t_{i-1}, t_i]$ *(Littlewood and Verall, 1973)*

- $h(t) = (N - i + 1)\phi(t - t_{i-1})$, $t \in (t_{i-1}, t_i]$ *(Schick and Wolverton, 1973)*

based on hazard rate $h(t)$ (with t_i's failure times)

FAILURE RATES

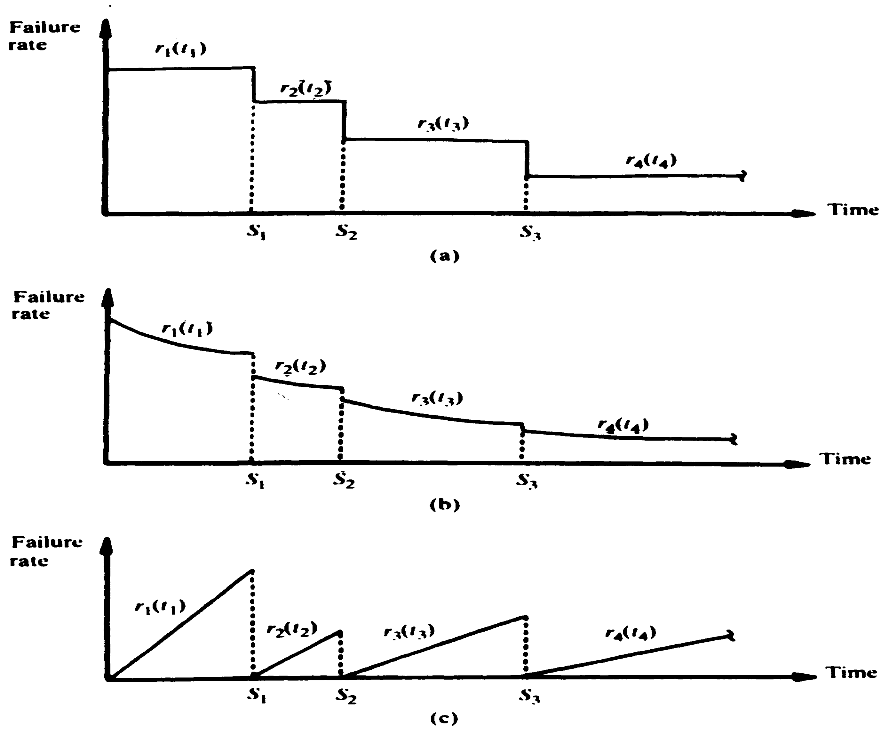


Figure 2.1. (a) The failure rates of the model of Jelinski and Moranda. (b) The failure rates of the model of Littlewood and Verall. (c) The failure rates of the model of Schick and Wolverton

Chen and Singpurwalla, Adv. Appl. Prob., 1997

SOFTWARE RELIABILITY: MODELS

Random coefficient autoregressive model
(Singpurwalla and Soyer, 1985)

- T_i interfailure times and $Y_i = \log T_i$, $i = 1, n$
- $Y_i = \theta_i Y_{i-1} + \epsilon_i, i = 1, n$
- $\epsilon_1, \ldots, \epsilon_n \sim \mathcal{N}(0, \sigma_1^2)$, i.i.d.
- $\theta_1, \ldots, \theta_n \sim \mathcal{N}(\lambda, \sigma_2^2)$, i.i.d.
- $\lambda \sim \mathcal{N}(\mu, \sigma_3^2)$

SOFTWARE RELIABILITY: MODELS

Martingale processes
(Basu and Ebrahimi, 2003)

Interfailure times $T_i \sim \mathcal{E}(\lambda_i)$, $i = 1, n$, conditionally independent given $\lambda_1, \ldots, \lambda_n$

- $\lambda_1 \sim \mathcal{G}(\alpha, \beta_1)$ and $\lambda_i | \lambda_{(-i)} \sim \mathcal{G}(\alpha, \alpha/\lambda_{i-1}), i > 1,$
  $\Rightarrow E(\lambda_i | \lambda_{i-1}) = \lambda_{i-1}$

- $\lambda_1 \sim \mathcal{G}(\alpha_1, \beta_1)$ and $\lambda_i | \lambda_{(-i)} \sim \mathcal{G}(\tau\lambda_{i-1}^2, \tau\lambda_{i-1}), i > 1,$
  $\Rightarrow E(\lambda_i | \lambda_{i-1}) = \lambda_{i-1}$

BAYESIAN SOFTWARE RELIABILITY

Kuo, Handbook of Statistics, 2005

Review paper

- Models
- Bayesian inference
- Model selection
- Optimal release policy

Limited number of faults $\Rightarrow$ room for Bayesian analysis

BAYESIAN SOFTWARE RELIABILITY

Ravishanker, Liu and Ray, 2008

- Count of number of failures in intervals $[t_0, t_1), \ldots, [t_{T-1}, t_T)$, with $t_0 = 0$ and $t_T = T$ (possibly divided into categories with different gravities)

- NHPP with m.v.f $M(t_{i-1}, t_i) = \theta \{F(t_i) - F(t_{i-1})\}$, $i = 1, T$
  - θ unknown expected number of failures over infinite horizon
  - F c.d.f., here Weibull

- Different parameters of Poisson r.v.'s in different intervals
  - $\theta \left\{ \prod_{j=1}^{i-1} (1 - p_{t_j}) \right\} p_{t_i}$, for interval $[t_{i-1}, t_i)$ as $i = 1, \ldots, T$
  - $p_{t_i} = 1 - \exp\{\beta_{S_{t_i}} (t_i - t_{i-1})^\alpha\}$

BAYESIAN SOFTWARE RELIABILITY

Ravishanker, Liu and Ray, 2008

- Evolution between L, possibly ordered, parameters through a transition matrix $\beta_{S_{t_i}} = \beta_j$ if $S_{t_i} = j$
- $\Rightarrow$ NHPP model with Markov switching mean value function
- Estimation and prediction via MCMC
- Model selection about L via BIC or other criterion

Major differences w.r.t. our work

- Count data w.r.t. actual failure times
- Easier interpretation of the parameter measuring reliability changes (here β, later λ)
- No need for us to estimate θ

STATEMENT OF THE PROBLEM

Bugs in software induce failures

Fixing current bugs sometimes implies introduction of new bugs

Lack of knowledge about effects of bugs fixing

$\Rightarrow$ need for models allowing for possible, unobserved introduction of new bugs in a context aimed to reduce bugs

BUGS INTRODUCTION: MODELS

Failures at $T_1, T_2, \ldots, T_n$

Inter-failure times $T_i - T_{i-1} \sim \mathcal{E}(\lambda_i)$, independent, $i = 1, \ldots, n$

- $\lambda_{i+1} = \lambda_i e^{-\theta_i}$, $\quad \lambda_i, \theta_i \in \mathbb{R}^+$, independent
  (Gaudoin, Lavergne and Soler, 1994)
  - $\theta_i = 0 \Rightarrow$ no debugging effect
  - $\theta_i > 0 \Rightarrow$ good quality debugging
  - $\theta_i < 0 \Rightarrow$ **bad quality debugging**

BUGS INTRODUCTION: MODELS

- $\lambda_{i+1} = (1 - \alpha_i - \beta_i)\lambda_i + \mu\beta_i, \quad \lambda_i, \mu \in \mathbb{R}^+$, *(Gaudoin, 1999)*
  - Imperfect debugging
  - α_i good debugging rate
  - β_i **bad debugging rate**

BUGS INTRODUCTION: MODELS

Birth-death process *(Kremer, 1983)*

- $p_n(t) = \mathcal{P}r\{X(t) = n\}$
- $\nu(t)$ **birth rate**
- $\mu(t)$ death rate
- a initial population

$$p_n^{'}(t) = (n-1)\nu(t)p_{n-1}(t) - n[\nu(t)+\mu(t)]p_n(t) + (n+1)\mu(t)p_{n+1}(t), n \geq 0$$

with $p_{-1} \equiv 0$ and $p_n(0) = \mathbb{1}_{\{n=a\}}$

HIDDEN MARKOV MODEL

Failure times $t_1 < t_2 < \ldots < t_n$ in $(0, y]$

Y_t latent process describing *reliability status* of software at time t (e.g. growing, decreasing and constant)

Y_t may change only after a failure
$\Rightarrow Y_t = Y_m$ for $t \in (t_{m-1}, t_m]$, $m = 1, \ldots, n+1$
with $t_0 = 0$, $t_{n+1} = y$ and $Y_{t_0} = Y_0$ *given for now*
$\Rightarrow$ consider $\{Y_n\}_{n \in \mathbb{N}}$ Markov chain with discrete state space E

X_m interarrival time of m-th failure, $m = 1, \ldots, n$

HIDDEN MARKOV MODEL

Markov chain $Y = \{Y_n\}_{n \in \mathbb{N}}$

- E discrete state space *(card*$(E) = k < \infty$*)*

- $\mathbb{P}$ transition matrix with rows $\mathbb{P}_i = (P_{i1}, \ldots, P_{ik})$, $i = 1, \ldots, k$

Interarrival times $X_m | Y_m = i \sim \mathcal{E}(\lambda(i)), i = 1, \ldots, k, m = 1, \ldots, n$

$\mathbb{P}$ and $\lambda(i)$ unknown

HIDDEN MARKOV MODEL

(Durand and Gaudoin, 2005)

Parameter estimation

- Data partially observed (only X_m but not Y_m)
  $\Rightarrow$ difficult parameter estimation by maximum likelihood

- $\Rightarrow$ EM algorithm for likelihood maximisation in the context of missing values *(McLachlan and Krishnan, 1997)*

- $\Rightarrow$ sequence of values converging to the consistent solution of the likelihood equation, **provided** the starting point is close to the optimal point

- $\Rightarrow$ start from many initial values

HIDDEN MARKOV MODEL

(Durand and Gaudoin, 2005)

Hidden states number estimation

- Take any $k \in [K_{\mathsf{min}}, K_{\mathsf{max}}]$
- For each k compute MLE via EM algorithm
- Choose k with lowest BIC
- Selection possibly affected by starting point of EM algorithm

HIDDEN MARKOV MODEL

(Durand and Gaudoin, 2005)

Selection of transition matrix via BIC

E.g. ordered $\lambda(1) > \lambda(2) > \ldots > \lambda(k)$

- Upper triangular matrix $\Rightarrow$ failure rates can only decrease
- Tridiagonal matrix $\Rightarrow$ only *small* increase and decrease in failure rate

HIDDEN MARKOV MODEL

X_m's independent given $Y \Rightarrow f(X_1, \ldots, X_n|Y) = \prod_{m=1}^{n} f(X_m|Y)$

$\mathbb{P}_i \sim \mathcal{D}ir(\alpha_{i1}, \ldots, \alpha_{ik}), \forall i \in E$, i.e. $\pi(\mathbb{P}_i) \propto \prod_{j=1}^{k} P_{ij}^{\alpha_{ij}-1}$

Independent $\lambda(i) \sim \mathcal{G}(a(i), b(i)), \forall i \in E$

Interest in posterior distribution of $\Theta = (\lambda^{(k)}, \mathbb{P}, Y^{(n)})$

- $\lambda^{(k)} = (\lambda(1), \ldots, \lambda(k))$
- $Y^{(n)} = (Y_1, \ldots, Y_n)$

LIKELIHOOD

For observed Y, joint density given by

$$\begin{aligned} f(X_1,\dots,X_n,Y_1,\dots,Y_n) &= f(X_1,\dots,X_n|Y_1,\dots,Y_n)f(Y_1,\dots,Y_n) \\ &= \prod_{m=1}^{n} P_{Y_{m-1}Y_m}\lambda(Y_m)e^{-\lambda(Y_m)X_m} \end{aligned}$$

Here unobserved Y treated as *parameter*

$\Rightarrow L(\Theta) = \prod_{m=1}^{n} \lambda(Y_m)e^{-\lambda(Y_m)X_m}$, with $\Theta = (\lambda^{(k)}, \mathbb{P}, Y^{(n)})$

Posterior distribution $\pi(\Theta|X_1,\dots,X_n)$ proportional to

$$\prod_{m=1}^{n}\left[P_{Y_{m-1}Y_m}\lambda(Y_m)e^{-\lambda(Y_m)X_m}\right]\cdot\prod_{i=1}^{k}\left[[\lambda(i)]^{a(i)-1}e^{-b(i)\lambda(i)}\prod_{j=1}^{k}P_{ij}^{\alpha_{ij}-1}\right]$$

FULL CONDITIONAL POSTERIORS

- $\mathbb{P}_i | Y^{(n)} \sim \mathcal{D}ir(\alpha_{ij} + \sum_{m=1}^{n} \mathbf{1}_{\{Y_{m-1}=i, Y_m=j\}}; j \in E), \forall i \in E$

- $\lambda(i) | Y^{(n)}, X^{(n)} \sim \mathcal{G}(a^*(i), b^*(i)), \forall i \in E$
  - ⋆ $a^*(i) = a(i) + \sum_{m=1}^{n} \mathbf{1}_{\{Y_m=i\}}$ & $b^*(i) = b(i) + \sum_{m=1}^{n} \mathbf{1}_{\{Y_m=i\}} X_m$
  - ⋆ $X^{(n)} = (X_1, \ldots, X_n)$

- $\pi(Y_m | Y^{(-m)}, \lambda(Y_m), X^{(n)}, \mathbb{P}) \propto P_{Y_{m-1}, Y_m} \lambda(Y_m) e^{-\lambda(Y_m) X_m} P_{Y_m, Y_{m+1}}$
  - ⋆ $\sum_{j \in E} P_{Y_{m-1}, j} \lambda(j) e^{-\lambda(j) X_m} P_{j, Y_{m+1}}$ normalizing constant
  - ⋆ $Y^{(-m)} = (Y_1, \ldots, Y_{m-1}, Y_{m+1}, \ldots, Y_n)$

POSTERIOR SAMPLE AND QUANTITIES

Gibbs sampling: posterior sample from $\pi(\Theta|X^{(n)})$ by iteratively drawing from the given full conditional posterior distributions

Posterior predictive distribution of X_{n+1} after observing $X^{(n)}$

$$\pi(X_{n+1}|X^{(n)}) = \sum_{j\in E}\int \pi(X_{n+1}|\lambda(j))P_{Y_n,j}\pi(\Theta|X^{(n)})d\Theta,$$

approximated as a Monte Carlo integral via

$$\pi(X_{n+1}|X^{(n)}) \approx \frac{1}{G}\sum_{g=1}^{G} \pi(X_{n+1}|\lambda^g(Y^g_{n+1}))$$

with Y^g_{n+1} sampled, given the posterior sample Y^g_n, using the Dirichlet posterior on $\mathbb{P}_{Y^g_n}$

ORDERING OF STATES

- Independent λ's $\Rightarrow$ no ordering among states

- No 0 in transition matrix $\Rightarrow$ jumps possible from any state to any state

$\Rightarrow$ difficult ranking of states in terms of reliability

- prior on ordered λ's $\Rightarrow$ identification of different levels of reliability

- Bi-(or tri-) diagonal transition matrix allowing only jumps into the nearest best (nearest best and worst) state

EXTENSION: PRIOR ON ORDERED λ'S

- $X \sim \mathcal{G}(\alpha, \beta) \perp (Y_1, \ldots, Y_m) \sim \mathcal{D}ir(a_1, \ldots, a_m) : \sum_{i=1}^{m} a_i = \alpha$

- Take $(\lambda_1, \ldots, \lambda_m) : \lambda_m = X$ & $\lambda_j = X \sum_{i=1}^{j} Y_i, j = 1, m-1$

- $\Rightarrow X = \lambda_m$ & $Y_j = \frac{\lambda_j - \lambda_{j-1}}{\lambda_m}, j = 1, m-1$ (with $\lambda_0 = 0$)

- $f(\lambda_1, \ldots, \lambda_m) = \beta^{\alpha} e^{-\beta\lambda_m} \prod_{j=1}^{m} \frac{(\lambda_j - \lambda_{j-1})^{a_j - 1}}{\Gamma(a_j)} I_{\{\lambda_1 < \lambda_2 < \ldots < \lambda_m\}}$

- $\Rightarrow \lambda_j | \lambda_{(-j)} \sim \mathcal{B}e(a_j, a_{j+1})$ on $(\lambda_{j-1}, \lambda_{j+1}), j < m$

EXTENSION: UNKNOWN NUMBER OF STATES

Selection of number of hidden states via Reversible Jump MCMC *(Green, 1995)* $\Rightarrow$ allows for simulation of posterior distributions in parameter spaces of variable size

Ordered $\lambda(1) > \lambda(2) > \ldots > \lambda(k)$

RJMCMC with steps

- [Move] λ_i changed to another value in $(\lambda_{i-1}, \lambda_{i+1})$
- [Death] Merge λ_i and λ_{i+1} into λ_i^* and rearrange indices
- [Birth] Split λ_i into $\lambda_{i,1}$ and $\lambda_{i,2}$ and rearrange indices

EXTENSION: UNKNOWN NUMBER OF STATES

- Models M_k, with k hidden states, $k = 1, K$

- Choice among models via Bayes factor $\frac{f(t|M_i)}{f(t|M_j)}$, with data t

- Basic marginal identity (Chib 1995):
  $\log f(t|M) = \log f(t|\theta^*, M) + \log f(\theta^*|M) - \log(\theta^*|t, M)$

- $\Rightarrow$ need to find adequate θ^* from MCMC output

ACTUAL IMPLEMENTATION

- Simulations with actual and simulated data evidenced some drawbacks
  - Prior with ordered λ's was leading to parameters too close each other
  - RJMCMC was hardly moving from $k = 2$
- $\Rightarrow$ prior with unordered λ's and different models with different state space sizes, despite
  - computation of Bayes factors
  - label switching about λ's

ACTUAL IMPLEMENTATION

- Different state space size $\Rightarrow$ different models
- Pairwise comparison by Bayes factor $\frac{p(D|i)}{p(D|j)}$ for models i, j and data D
- $D = x^{(n)} = (x_1, x_2, \ldots, x_n)$
- $p(D)$ *(dependence on i omitted)* not available in analytical form and nontrivial evaluation using posterior Monte Carlo samples
- $p(D) = \frac{p(D|\Theta)p(\Theta)}{p(\Theta|D)}$ holds for any Θ, say Θ^*
- log marginal likelihood estimated using Monte Carlo samples

  $\ln \widehat{p(D)} = ln\, \widehat{p(D|\Theta^*)} + \ln \widehat{p(\Theta^*)} - \ln \widehat{p(\Theta^*|D)}$
- $p(D|\Theta^*)$ and $p(\Theta^*)$ can be evaluated at Θ^* whereas $p(\Theta^*|D)$ not immediately available

ACTUAL IMPLEMENTATION

- $p(D) = p(x^{(n)}) = \dfrac{p(x^{(n)}|\boldsymbol{\lambda}^{(k)}, \boldsymbol{Y}^{(n)})p(\boldsymbol{\lambda}^{(k)})p(\boldsymbol{Y}^{(n)}|\boldsymbol{P})\,p(\boldsymbol{P})}{p(\boldsymbol{\lambda}^{(k)}, \boldsymbol{P}, \boldsymbol{Y}^{(n)}|x^{(n)})}$

  evaluated at posterior modes of $(\boldsymbol{\lambda}^{(k)}, \boldsymbol{P}, \boldsymbol{Y}^{(n)}) = (\boldsymbol{\lambda}^{*(k)}, \boldsymbol{P}^{*}, \boldsymbol{Y}^{*(n)})$

- Numerator evaluated very easily (densities and parameters are known)

- Chain's rule for the denominator

  $p(\boldsymbol{\lambda}^{*(k)}, \boldsymbol{P}^{*}, \boldsymbol{Y}^{*(n)}|x^{(n)}) = p(\boldsymbol{\lambda}^{*(k)}|\boldsymbol{Y}^{*(n)}, x^{(n)})p(\boldsymbol{P}^{*}|\boldsymbol{Y}^{*(n)})p(\boldsymbol{Y}^{*(n)}|x^{(n)})$

  – $p(\boldsymbol{\lambda}^{*(k)}|\boldsymbol{Y}^{(n)}, x^{(n)})$ product of independent gamma densities

  – $p(\boldsymbol{P}^{*}|\boldsymbol{Y}^{*(n)})$ product of independent Dirichlet densities

  – Still to evaluate $p(\boldsymbol{Y}^{*(n)}|x^{(n)})$

ACTUAL IMPLEMENTATION

- Set $\boldsymbol{Y}^{(t-1)} = (Y_1, \ldots, Y_{t-1})$ and $\boldsymbol{Y}^{(s>t)} = (Y_{t+1}, \ldots, Y_n)$
- Chain's rule
  $p(\boldsymbol{Y}^{*(n)}|x^{(n)}) = p(Y_1^*|x^{(n)})\, p(Y_2^*|Y_1^*, x^{(n)}) \cdots p(Y_t^*|\boldsymbol{Y}^{*(t-1)}, x^{(n)}) \cdots p(Y_n^*|\boldsymbol{Y}^{*(n-1)}, x^{(n)})$
- Use MCMC draws to estimate ($t = 1$ or $t > 1$)
  - $p(Y_1^*|x^{(n)}) \approx \frac{1}{G}\sum_{g=1}^{G} p(Y_1^*|(\boldsymbol{\lambda}^{(k)})^{(g)}, (\boldsymbol{Y}^{-1})^{(g)}, \boldsymbol{P}^{(g)}, x^{(n)})$

    for any $(\boldsymbol{\lambda}^{(k)})^{(g)}$, $(\boldsymbol{Y}^{-1})^{(g)}$, $\boldsymbol{P}^{(g)}$ $\Rightarrow$ *known conditional density of* ***fixed*** Y_1^*
  - $p(Y_t^*|\boldsymbol{Y}^{*(t-1)},x^{(n)}) \approx \frac{1}{G'}\sum_{g=1}^{G'} p(Y_t^*|(\boldsymbol{\lambda}^{(k)})^{(g)},\boldsymbol{P}^{(g)},(\boldsymbol{Y}^{(s>t)})^{(g)},\boldsymbol{Y}^{*(t-1)},x^{(n)})$
    - $p(Y_t^*|\boldsymbol{Y}^{*(t-1)}, x^{(n)}) = \int p(Y_t^*|\boldsymbol{\lambda}^{(k)}, \boldsymbol{P}, \boldsymbol{Y}^{(s>t)}, \boldsymbol{Y}^{*(t-1)}, x^{(n)})\cdot$
      $\cdot p(\boldsymbol{\lambda}^{(k)}, \boldsymbol{P}, \boldsymbol{Y}^{(s>t)}|\, \boldsymbol{Y}^{*(t-1)}, x^{(n)}) d\boldsymbol{\lambda}^{(k)}\, d\boldsymbol{P}\, d\boldsymbol{Y}^{(s>t)}$
    - chain's rule to split second density in r.h.s.
    - additional sampling from $p(\boldsymbol{\lambda}^{(k)}, \boldsymbol{P}, \boldsymbol{Y}^{(s>t)}|\, \boldsymbol{Y}^{*(t-1)}, x^{(n)})$ with further split (*skipped here*) of densities

JELINSKI-MORANDA DATA

34 software failure times

2 states for Y_m

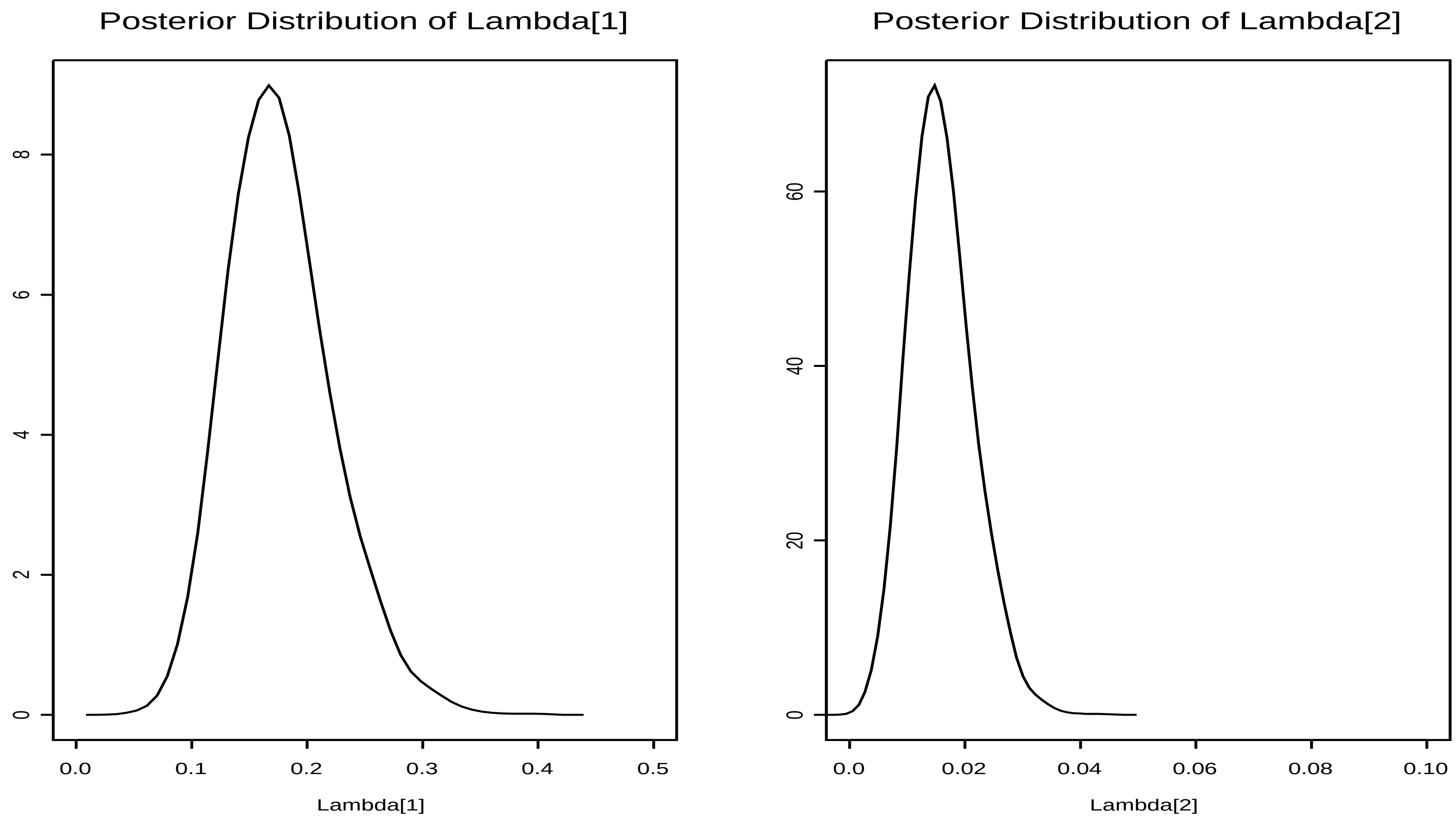
$\mathbb{P}_i \sim \mathcal{B}eta(1,1), i = 1,2$ *(uniform)*

$\lambda(i) \sim \mathcal{G}(0.01, 0.01), i = 1,2$ *(diffuse)*

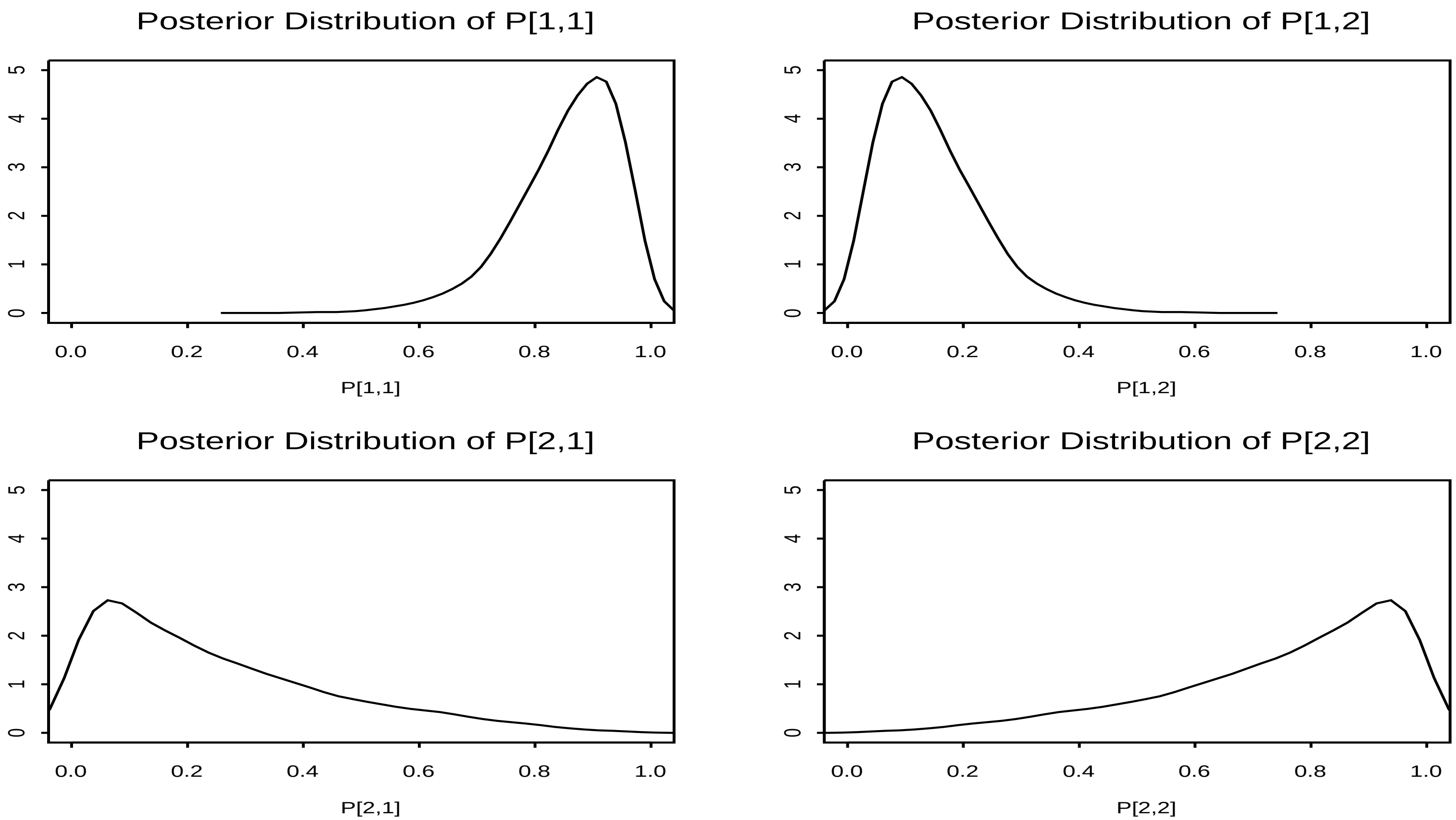
5000 iterations

Convergence of Gibbs sampler pretty good

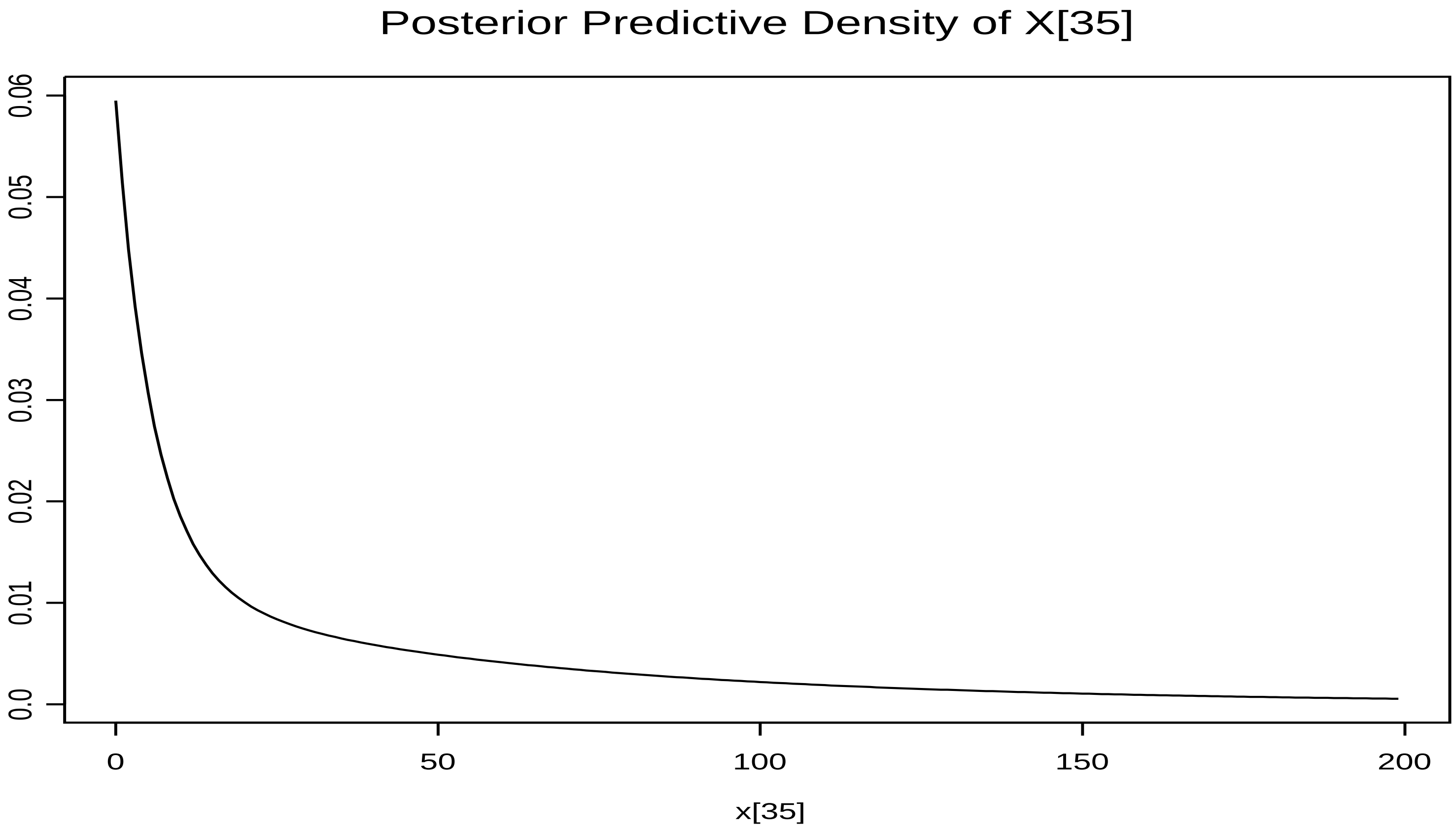
JELINSKI-MORANDA DATA



JELINSKI-MORANDA DATA



JELINSKI-MORANDA DATA



JELINSKI-MORANDA DATA

Posterior Probabilities of State 1 over Time

m	X_m	$P(Y_m = 1\|D)$	m	X_m	$P(Y_m = 1\|D)$	m	X_m	$P(Y_m = 1\|D)$
1	9	0.8486	2	12	0.8846	3	11	0.9272
4	4	0.9740	5	7	0.9792	6	2	0.9874
7	5	0.9810	8	8	0.9706	9	5	0.9790
10	7	0.9790	11	1	0.9868	12	6	0.9812
13	1	0.9872	14	9	0.9696	15	4	0.9850
16	1	0.9900	17	3	0.9886	18	3	0.9858
19	6	0.9714	20	1	0.9584	21	11	0.7100
22	33	0.2036	23	7	0.3318	24	91	0.0018
25	2	0.6012	26	1	0.6104	27	87	0.0020
28	47	0.0202	29	12	0.2788	30	9	0.2994
31	135	0.0006	32	258	0.0002	33	16	0.1464
34	35	0.0794						

Expected posterior probability of the ”bad” state decreases as we observe longer failure times

JELINSKI-MORANDA DATA

- $p(D|k) = \prod_{i=1}^{k} \frac{\Gamma(a^*(i))}{\Gamma(a(i))} \frac{b(i)^{a(i)}}{b^*(i)^{a^*(i)}} \times \frac{\Gamma(k)^k}{k} \prod_{i=1}^{k} \frac{\prod_{j=1}^{k} \Gamma(1+m_{ij})}{\Gamma(k+m_i)} \times \frac{1}{p(\boldsymbol{Y}^{*(n)}|D)}$

  with $m_{ij} = \sum_t \mathbf{1}(Y_t = i, Y_{t+1} = j)$ and $m_i = \sum_j m_{ij}$

- $\log p(D|k) = -148.92, -139.81, -142.43, -144.63$ for $k = 1, 2, 3, 4$

  $\Rightarrow$ choose $k = 2$

JELINSKI-MORANDA DATA

- Estimation of state Y_t at time t affected by *label switching*, i.e. non-preserved ordering of λ's (e.g. $Y_t = 1 \not\Rightarrow \lambda(1)$ best, i.e. smallest)
  - $k = 2 \Rightarrow Y_t^*$ most frequent value in sequence $\{Y_t^g\}_{g\geq 1}$ (label switching never occurs during Gibbs sampling)
  - Label switching occurs for $k > 2$
    - keep information consistent throughout iterations, i.e. rank of $\lambda^g(Y_t^g)$ within the vector of sorted rates $(\lambda^g((1)), \ldots, \lambda^g((k)))$
    - build table of frequencies for sequence of ranks (1 to k) of $\{\lambda^g(Y_t^g)\}_{g\geq 1}$
    - relative frequency of, say, rank 2 is the sample average estimate of the posterior probability that the environment is the second best at epoch t
    - $\Rightarrow Y_t^*$ rank with highest frequency

MUSA SYSTEM 1 DATA

136 software failure times

2 states for Y_m

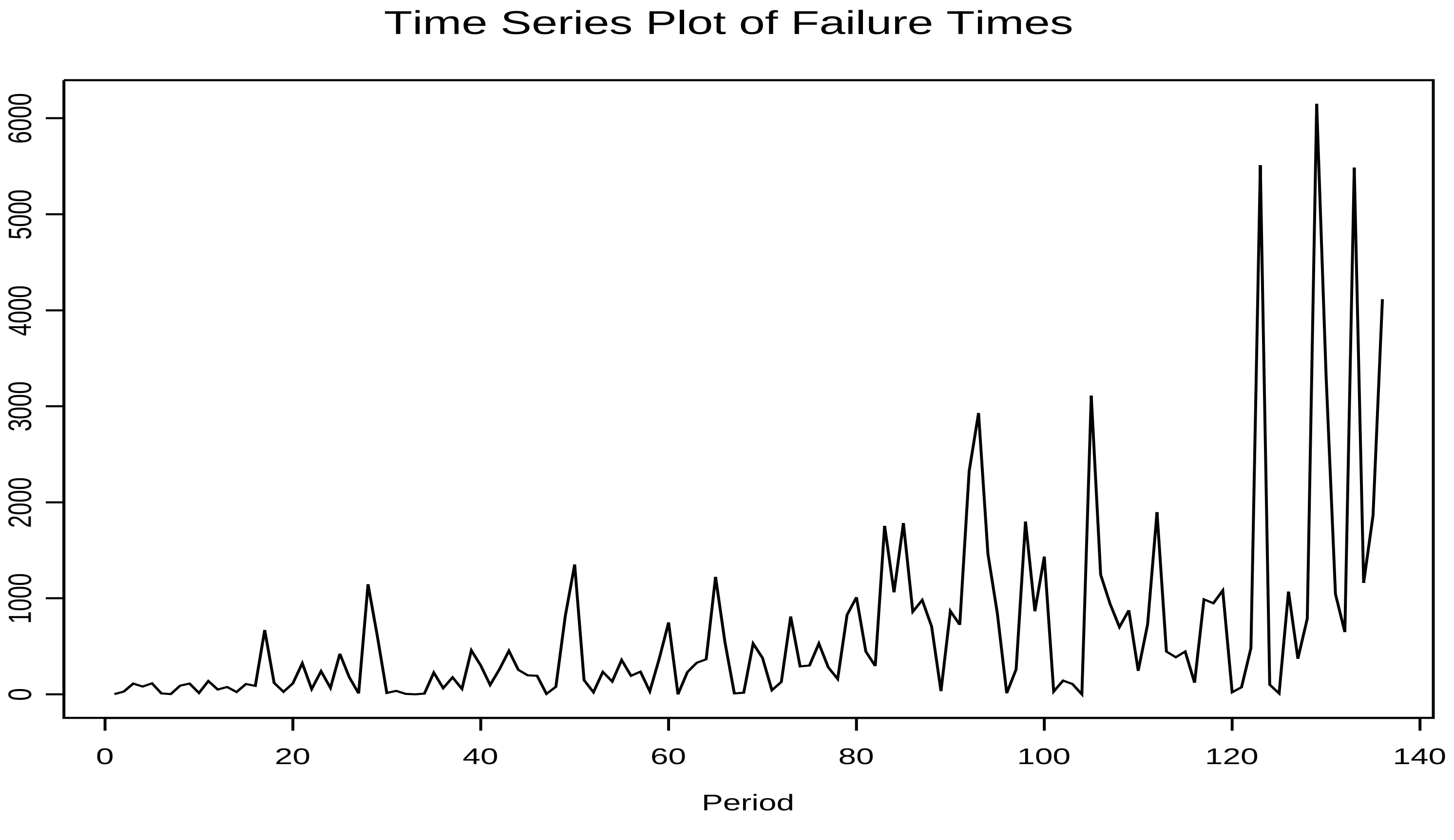
$\mathbb{P}_i \sim \mathcal{B}eta(1,1), i = 1,2$ *(uniform)*

$\lambda(i) \sim \mathcal{G}(0.01, 0.01), i = 1,2$ *(diffuse)*

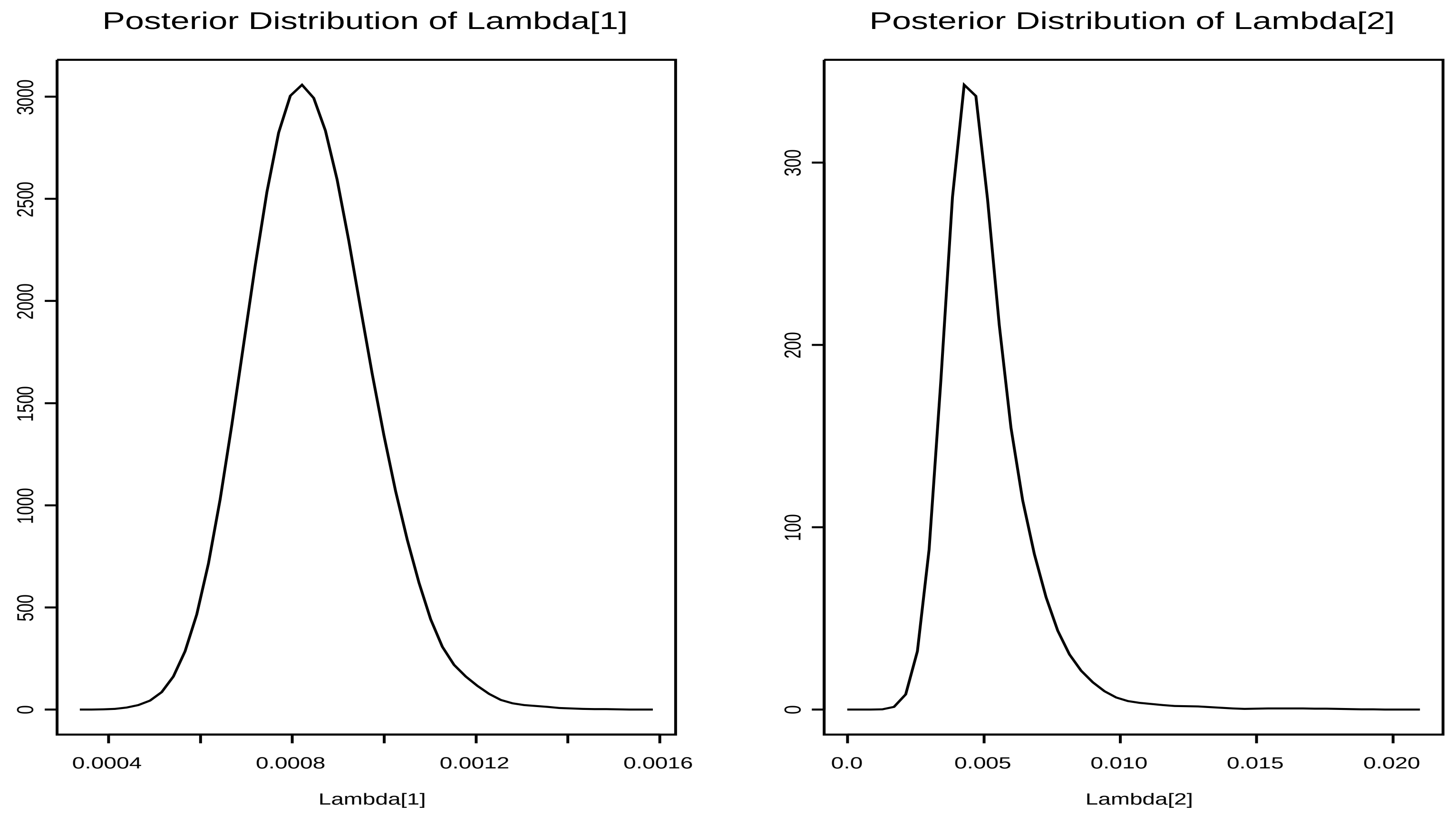
5000 iterations

Convergence of Gibbs sampler pretty good

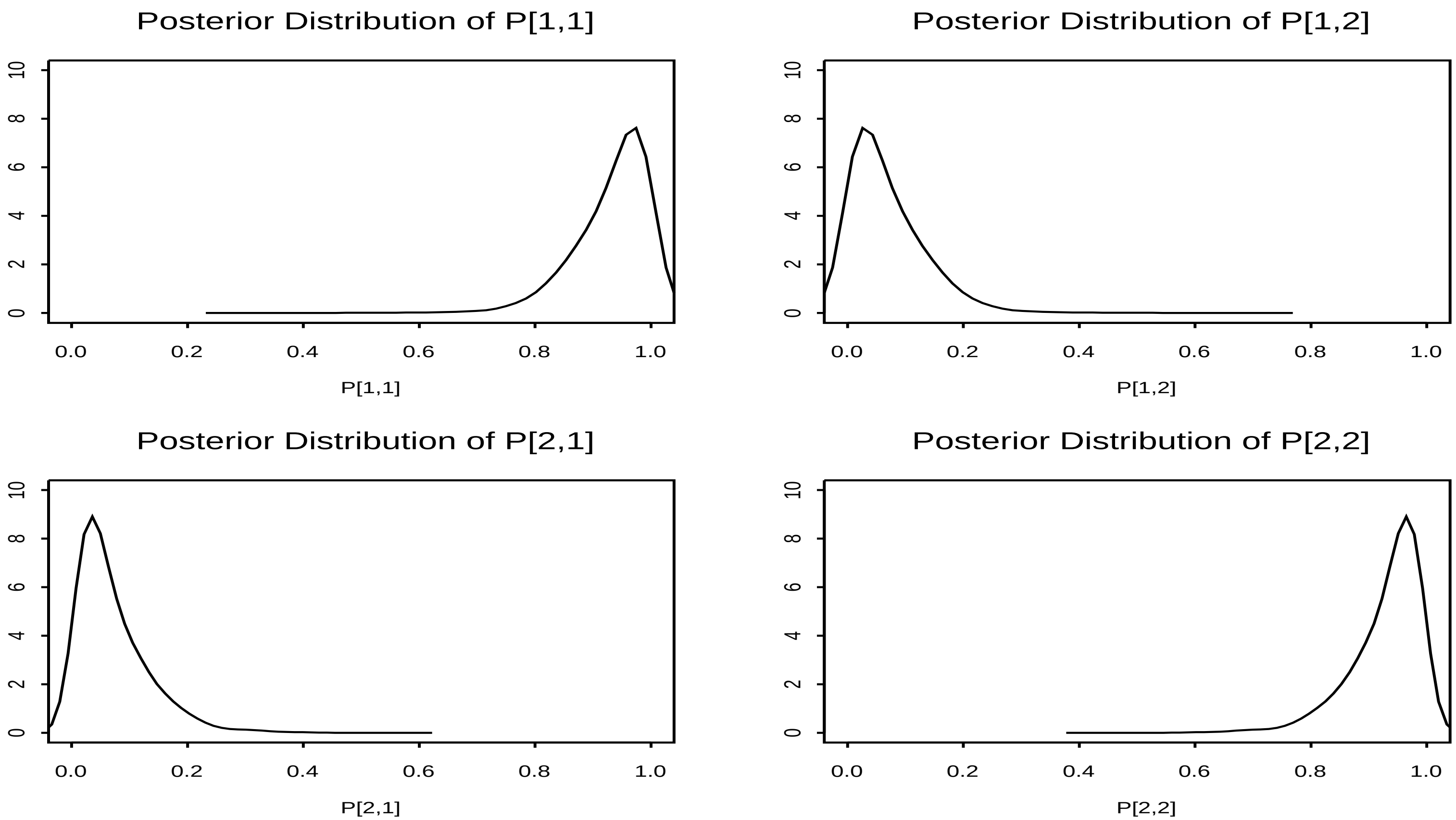
MUSA SYSTEM 1 DATA



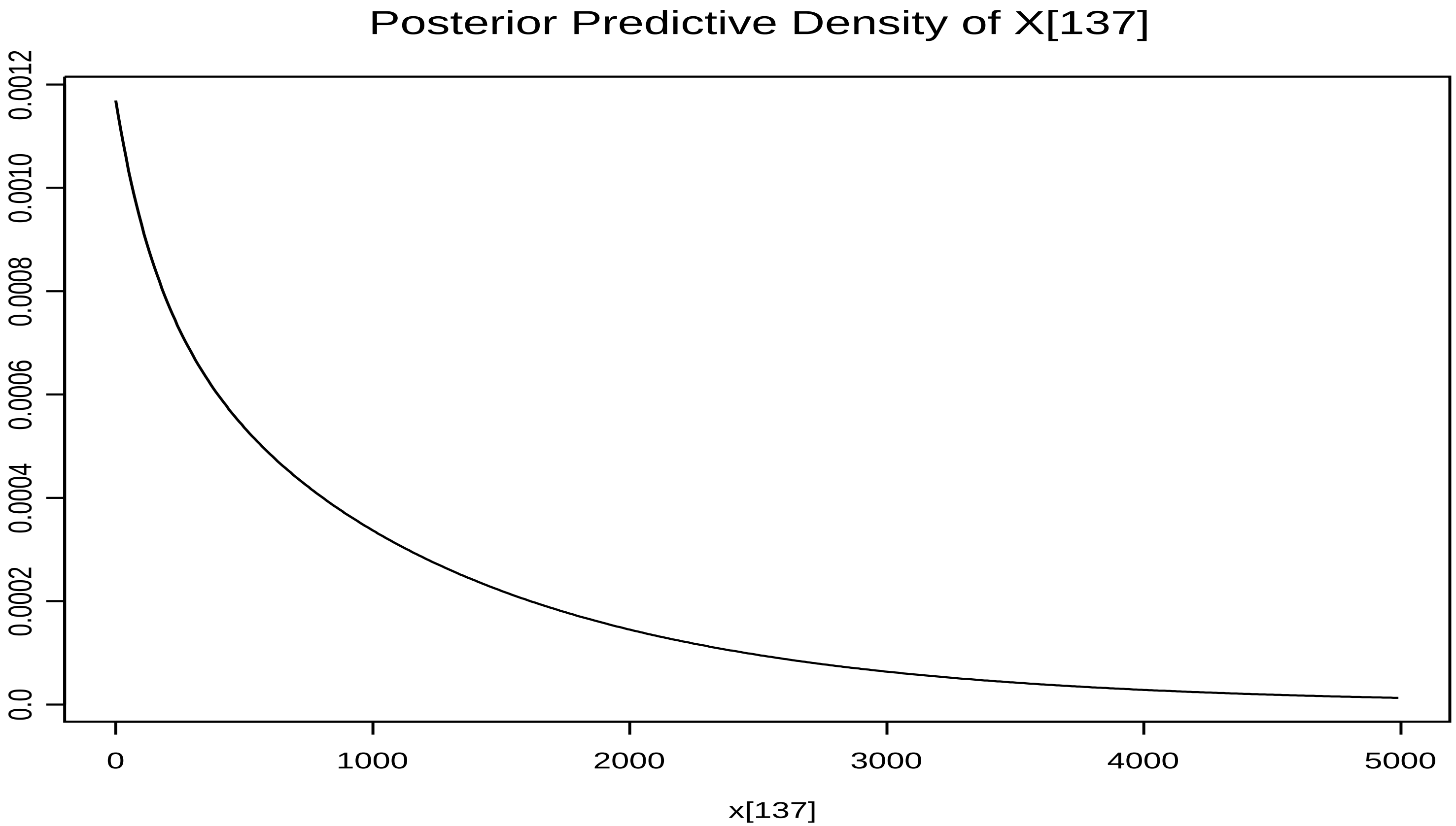
MUSA SYSTEM 1 DATA



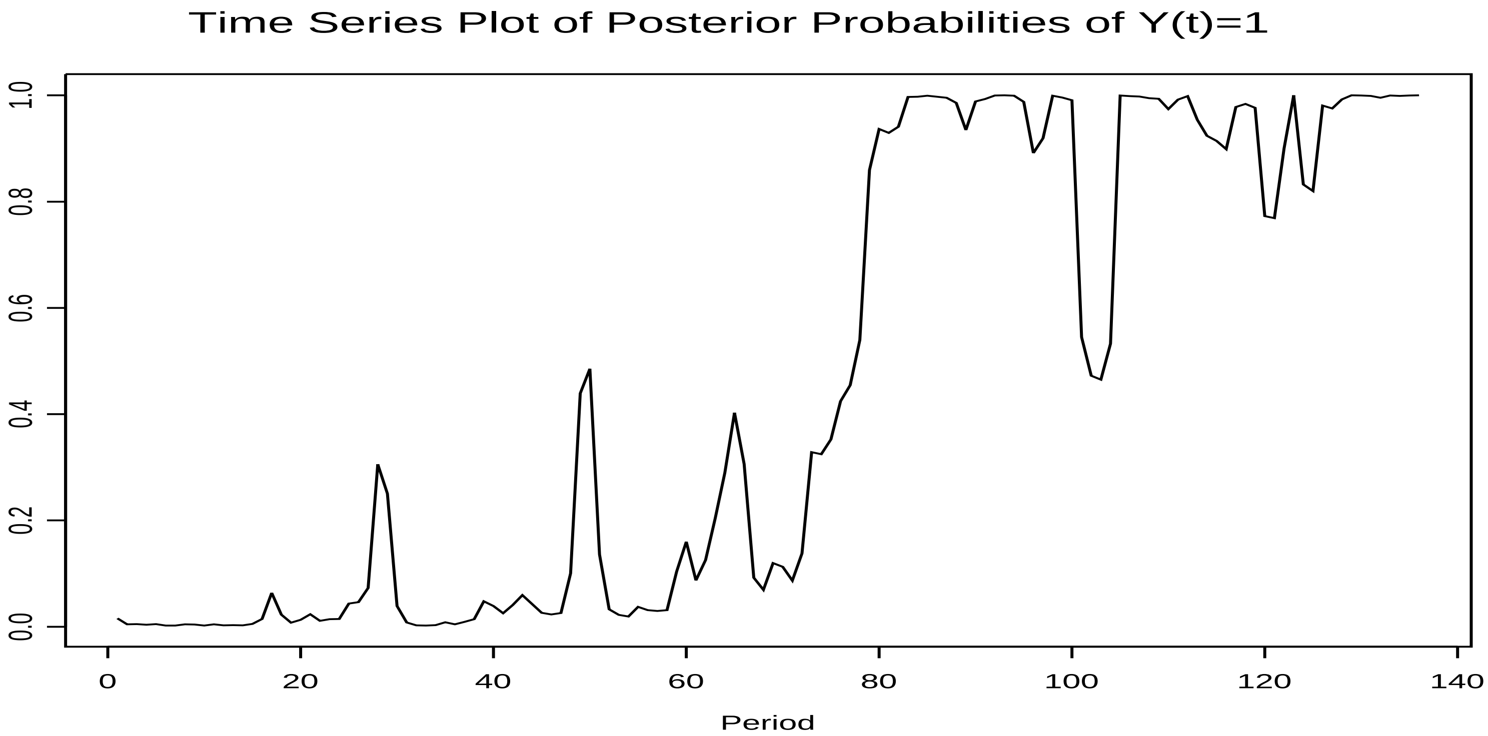
MUSA SYSTEM 1 DATA



MUSA SYSTEM 1 DATA



MUSA SYSTEM 1 DATA



Expected posterior probability of the ”good” state increases as we observe longer failure times

MUSA SYSTEM 1 DATA

Marginal likelihoods

k	avg. of $\log \widehat{p(D\|k)}$	no. of runs	st. dev.
1	-1023.549	0 (exact)	0
2	-997.9537	2	0.23
3	-988.5743	5	0.64
4	-990.1471	5	0.39
5	-992.4615	7	1.54

- repeated computations of denominator of $P(D|i)$ to improve accuracy of estimates
- $k = 3$ best model but ...
  - identical posterior distributions of two smallest λ
  - largest lambda due to three failure times very very close to 0
- $\Rightarrow$ choose more parsimonious $k = 2$

ACTUAL IMPLEMENTATION

Comparison between conditional and unconditional prior on λ's

- Conditional prior avoids in principle the label switching problem but, in practice, unconditional one works well, for both our simulated and actual data, with the *tricks* we used about ranking
- Similar computational efforts: easier simulation from unconditional prior but, then, need to look at rankings at each epoch to address label switching problem
- Conditional prior tends to favour merging of parameters: need to further check with other data
- No significant difference in computing Bayes factors

CURRENT WORK

- Introduction of covariates (namely software metrics) in HMM
  Pievatolo, F.R., Soyer and Wiper

- Self-exciting process with latent variables
  Landon, F.R. and Soyer

SELF-EXCITING PROCESS WITH LATENT VARIABLES

NHPPs widely used in (software) reliability, characterised by an intensity function $\mu(t)$

Self-exciting processes (SEPs) add extra terms $g(t-t_i)$ to the intensity as a consequence of events at t_i (e.g. introduction of new bugs)

Binary latent variables modelling the introduction of new bugs

SELF-EXCITING PROCESS WITH LATENT VARIABLES

$\Rightarrow$ SEP with intensity $\lambda(t) = \mu(t) + \sum_{j=1}^{N(t^-)} Z_j g_j(t - t_j)$

- $\mu(t)$ intensity of process w/o introduction of new bugs
- $N(t^-)$ number of failures right before t
- $t_1 < t_2 < \ldots < t_n$ failures in $(0, T]$
- $Z_j = 1$ if bug introduced after j–th failure and $Z_j = 0$ o.w.
- $g_j(u) \geq 0$ for $u > 0$ and $= 0$ o.w.

SELF-EXCITING PROCESS WITH LATENT VARIABLES

$\underline{t} = (t_1, \ldots, t_n)$ failures in $(0, T]$

$\underline{Z} = (Z_1, \ldots, Z_n)$ latent variables at $\underline{t} = (t_1, \ldots, t_n)$

Likelihood $L(\theta; \underline{t}, \underline{Z}) = f(\underline{t}|\underline{Z}, \theta) f(\underline{Z}|\theta)$

$$\begin{aligned} f(\underline{t}|\underline{Z}, \theta) &= \prod_{i=1}^{n} \lambda(t_i) e^{-\int_0^T \lambda(t)dt} \\ &= \prod_{i=1}^{n} \left[\mu(t_i) + \sum_{j=1}^{i-1} Z_j g(t_i - t_j) \right] e^{-\int_0^T \mu(t)dt - \sum_{j=1}^{N(T^-)} Z_j \int_0^{T-t_j} g_j(t)dt} \end{aligned}$$

[θ omitted]

Posterior distribution on $(\theta, \underline{Z})$ through MCMC

Bayesian Analysis of Call Center Arrival Data

Fabrizio Ruggeri[1], Refik Soyer[2] and Murat Tarimcilar[2]

[1]Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it
www.mi.imati.cnr.it/~fabrizio/

[2]George Washington University
Washington D.C., USA

OUTLINE OF THE TALK

- General issues on call centers data
- Some models for call centers data
- Efficacy of advertising campaigns
- Bayesian models
- Example
- Future research

CALL CENTER

- Centralised hub aimed to make or get calls to/from (prospective) customers

- (Often) primary point of contact between customers and businesses

- Major investment for many organisations

- 2.86 million operator positions in over 50,000 call centers in the US, with some locations employing over 1000 agents

- Not only businesses, but also governments, etc.

Some citations from `Weinberg, Brown and Stroud, 2006`

CALL CENTER OPERATIONS

- Different functions
  - Only inbound calls (e.g. requests for assistance)
  - Only outbound calls (e.g. promotions)
  - Both of them (possibly, outbound calls when idle from inbound ones, including answers to previous inbound calls)

- Different structures
  - All issues handled by equally trained agents
  - Various levels, e.g.
    * automated answers to simple issues
    * contact with lowly trained agents for ordinary issues
    * contact with highly trained agents (e.g. supervisors) for the most complex issues

CALL CENTER OPERATIONS

- Different technologies
  - Direct phone calls to agents
  - Phone calls to a computerised system, routing calls, e.g., to automated answers and different levels of agents
  - Computer-phone integration, allowing for identification of customers and immediate availability of his/her data (personal and past communications)
  - *Contact center*: Computer supporting agents via a full range of media (e-mail, fax, web pages, chat)
- Different policies
  - No customer must be lost
  - No premium customer must be lost
  - Trade between losses due to customer losses and center staffing

CALL CENTER QUALITY

- Qualitative measures, e.g.
  - Customer satisfaction about
    - user-friendly system
    - length of hold-in-line
    - effectiveness of answers
  - Company *image* affecting future businesses
- Quantitative measures, e.g.
  - *Abandonment*, as fraction of customers leaving the queue before service
  - *Retrial*, as average number of calls needed to solve a problem
  - *Waiting*, as its average or some percentiles of the waiting-time distribution
  - **Profit**

CALL CENTER DATA CLASSIFICATION

- Operational data
  - Typically, aggregated data over some periods (minutes/daily/weekly/yearly) from history of each call, e.g.
    - total number of calls served or abandoned
    - average waiting time
    - agents' utilisation level
- Marketing data
  - Combination of phone data with customer's profile and past history
  - Euro-figures for past sales and future marketing targets
- Psychological data
  - Surveys of customers, agents and managers about subjective perception of service level and working environment

Classification taken from `Koole and Mandelbaum, 2001`

CALL CENTER DATA USE

- Customer level
  - Future marketing target based on past transactions of existing customers
  - Identification of possible (and profitable) new customers
- Company level
  - Quality of service (lost calls, waiting times, etc.)
  - Staffing of call center
  - Monitoring of agents' performance (*might be unlawful*)
  - Training of agents

CALL CENTER

(Statistical) interest in

- Forecasting demand in different periods (e.g. from hourly to yearly)

- Customer loss

- Optimal number of agents (possibly time-dependent)

- Customers' (both current and perspective) profiles

- etc.

MODELS FOR CALL CENTER ARRIVAL DATA

Customer level

- Customer profile, using CRM (Customer Relations Management) and Data Mining

- Degree of satisfaction, mostly via surveys

MODELS FOR CALL CENTER ARRIVAL DATA

Company level

- Time series (especially ARIMA processes)
  - arrivals over periods of time
  - detection of different patterns in different periods (e.g. before and after Xmas)
  - normalised (w.r.t. total daily calls) arrivals during each day
- Queueing models
  - arrival time
  - system availability
  - (optimal) number of channels
  - service policy (e.g. premium customers)

MODELS FOR CALL CENTER ARRIVAL DATA

Poisson processes with time dependent arrival rates

- Doubly stochastic Poisson models

- Nonhomogeneous Poisson processes (NHPPs)

Focus on forecasting models for optimal scheduling and staffing of telephone operators in a call center, using aggregate arrival data

BACKGROUND

- Consumer Electronics Producer
  - Limited variety of products
  - Long life cycle
  - Aging products, sporadic upgrades
  - Targeted advertising
- Sales
  - Average sale is around $ 500.00
  - Almost all the sales are through the sales call center
  - Call return is very low if first time the customer is denied the call
- Advertisement
  - Many media venues used, although print media is majority
  - Each ad is targeted and urging customers to place calls

OBJECTIVES

- Evaluation of different advertisement policies from a marketing viewpoint through call center arrival data

- Analysis of effects of relevant covariates, e.g.
  - Media expense (in $'s)
  - Venue type (monthly magazine, daily newspaper, etc.)
  - Ad format (full page, half page, colour, etc.)
  - Offer type (free shipment, payment schedule, etc.)
  - Seasonal indicators

- Prediction of calls volume generated by specific ad over any interval of interest

- Arrival data are not aggregated

DATA

Typical call arrival data for an ad

Time interval (in days)	Number of calls
$(0, 1]$	6
$(1, 2]$	5
$(2, 3]$	1
$(3, 4]$	3
$(4, 5]$	2
$(5, 6]$	2
$(6, 7]$	3
$(7, 8]$	0
$(8, 9]$	2
$(9, 10]$	2
$(10, 18]$	0

- Number of calls for ad in unit intervals
- Features of the ad (media, cost, etc.)
- Corresponding ad known for each call in earlier work

GOALS OF THE ONGOING RESEARCH

- Handling calls unassigned to any ad
- Different models for covariates
- Bayesian nonparametrics

NOTATIONS

- $C_1, \ldots, C_m$: m campaigns (ad)
- $T_1, \ldots, T_m$: campaigns starting times, $T_1 \leq T_2 \leq \ldots \leq T_m$
- Calls recorded as number of arrivals in $I_j = (t_{j-1}, t_j]$, $j = 1, n$
  (*starting times coincide with one of the endpoints of the intervals* I_j)
- n_{ij}: number of calls in I_j related to campaign C_i, $i = 1, m$, $j = 1, n$
- u_j: number of unallocated calls in I_j (0 in earlier work)
  (*in each* I_j *consider only campaigns* C_i *with* $T_i \leq t_{j-1}$)

NONHOMOGENEOUS POISSON PROCESS

- $N_t, t \geq 0$ # events by time t
- $N(y, s)$ # events in $(y, s]$
- $\Lambda(t) = \mathcal{E} N_t$ mean value function
- $\Lambda(y, s) = \Lambda(s) - \Lambda(y)$ expected # events in $(y, s]$

$N_t, t \geq 0$, NHPP with intensity function $\lambda(t)$ *iff*

1. $N_0 = 0$
2. independent increments
3. $\mathcal{P}\{\#$ events in $(t, t+h) \geq 2\} = o(h)$
4. $\mathcal{P}\{\#$ events in $(t, t+h) = 1\} = \lambda(t)h + o(h)$

$$\Rightarrow \mathcal{P}\{N(y, s) = k\} = \frac{\Lambda(y, s)^k}{k!} e^{-\Lambda(y,s)}, \forall k \in \mathcal{N}$$

NONHOMOGENEOUS POISSON PROCESS

$\lambda(t) \equiv \lambda \; \forall t \Rightarrow$ HPP

- $\lambda(t)$: intensity function of N_t

- $\lambda(t) := \lim_{\Delta \to 0} \dfrac{\mathcal{P}\{N(t, t+\Delta] \geq 1\}}{\Delta}, \; \forall t \geq 0$

- $\mu(t) := \frac{d\Lambda(t)}{dt}$: Rocof (rate of occurrence of failures)

Property 3. $\Rightarrow \mu(t) = \lambda(t)$ a.e. $\Rightarrow \Lambda(y, s) = \int_y^s \lambda(t)dt$

MODULATED NHPP MODEL

- NHPP's with intensity dependent on covariates

- $\lambda_i(t, Z_i) = \lambda_0(t)\,\text{exp}\{\gamma' Z_i\}$

- $\lambda_0(t)$ baseline intensity and γ parameter

- $\frac{\lambda_i(t,Z_i)}{\lambda_j(t,Z_j)} = \text{exp}\{\gamma'(Z_i - Z_j)\}$
  $\Rightarrow$ Proportional Intensities Model

Cox, 1972

MODEL

Call center arrival data

- modeled as NHPP
- dependent on the campaign (i.e. covariates Z_i)
- dependent on the starting point T_i

$\Rightarrow$ for campaign C_i

- $\lambda_i(t) = \lambda_0(t - T_i)\,\exp\{\gamma' Z_i\} I_{[T_i,\infty)}(t)$
- $\Lambda_i(t) = \Lambda_0(t - T_i)\,\exp\{\gamma' Z_i\} I_{[T_i,\infty)}(t)$

MODEL

Superposition Theorem

Sum of independent NHPP with intensity functions $\lambda_i(t)$ *is still a NHPP with intensity function* $\lambda(t) = \sum \lambda_i(t)$

MODEL

Number of calls decreasing to zero $\Rightarrow$ Power Law Process (PLP)

- $\lambda(t) = M\beta t^{\beta-1}$

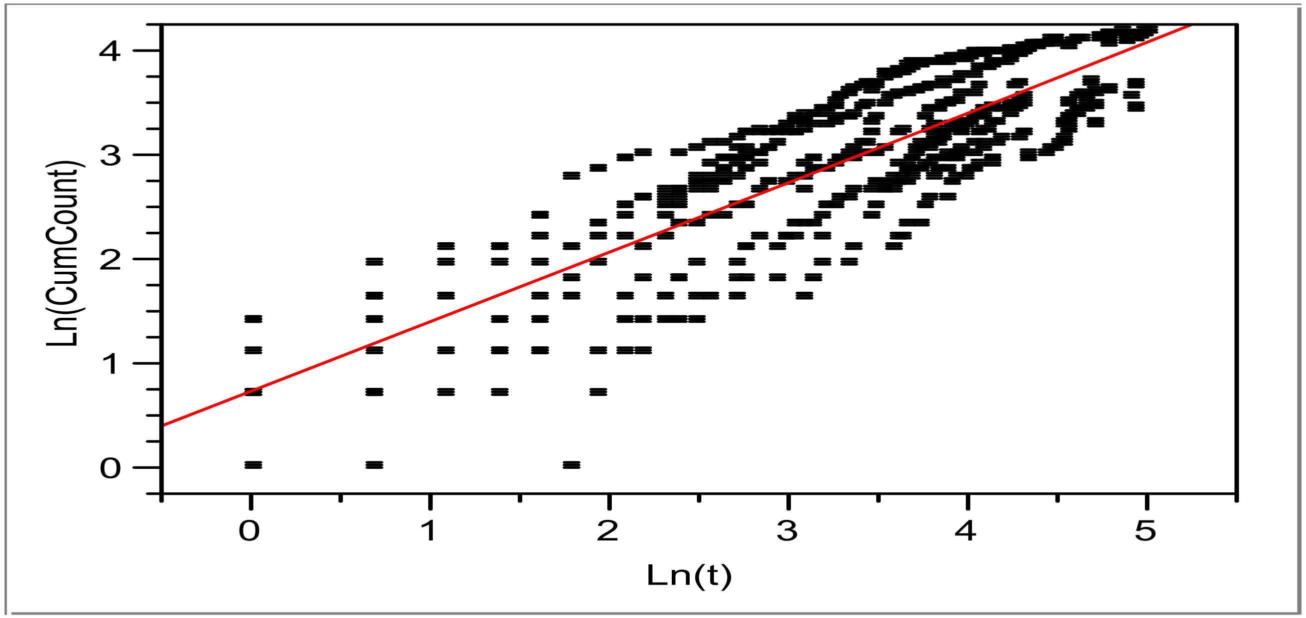
- $\Lambda(t) = Mt^{\beta}$

$M, \beta, t > 0$ but here $0 < \beta < 1 \Rightarrow \lambda(t) \downarrow 0$ for $t \to \infty$

Alternative: $\lambda(t) = \beta_0 \frac{\log(1+\beta_1 t)}{(1+\beta_1 t)}$
(increasing from 0 and then decreasing to 0)

MODEL

- Cumulative number of arrivals approximates Λ
- PLP $\Rightarrow \log \Lambda(t) = \log M + \beta \log t$
- $\Rightarrow$ plot of $\log \Lambda(t)$ v.s. $\log t$ to check appropriateness of PLP



MODEL WITH PERFECT LINKAGE OF CALLS

- No further details on the previous model to avoid repetitions with the model with unallocated calls

- Random effects model, accounting for differences unexplained by covariates

  – Model similar to the previous one, but

  – NHPPs with $\Lambda_i(t) = M_i t^\beta$

  – $\log M_i = \theta + \phi_i$, ϕ_i random effect terms

  – ϕ_i's conditional independent $\mathcal{N}(0, 1/\tau)$ and $\tau \sim \mathcal{G}(a_\tau, b_\tau)$

DIGRESSION: MODEL FOR UNASSIGNED CALLS

- Unassigned calls define a new process with $\Lambda_u(t) = \delta_u M t^\beta$
- δ_u r.v. rescaling the baseline mean value function
- Model describes arrivals of unallocated calls but not their assignment to campaigns and, consequently, these data are less useful to measure effectiveness of campaigns

A POSSIBLE (BUT INCONVENIENT) MODEL

- Arrivals in the interval I_j, $j = 1, n$:
  - n_{ij} related to campaign C_i, $i = 1, m$
  - u_j unallocated
- Interest in $\mathcal{P}(n_{1j} \in C_1, \ldots, n_{mj} \in C_m, u_j \in I_j)$
- All possible allocations: $\underline{u}_j = \{u_{1j} \in C_1, \ldots, u_{mj} \in C_m\}$, with $\sum_{l=1}^{m} u_{lj} = u_j$
- Specify a distribution $\mathcal{P}(\underline{u}_j)$ (e.g. multinomial) on the allocation
- Compute

  $\mathcal{P}(n_{1j} \in C_1, \ldots, n_{mj} \in C_m, u_j \in I_j) =$
  $\sum_{\text{all } \underline{u}_j} \mathcal{P}(n_{1j} + u_{1j} \in C_1, \ldots, n_{mj} + u_{mj} \in C_m | \underline{u}_j)\mathcal{P}(\underline{u}_j)$

ASSUMPTIONS

- Latent variables $\underline{Y}_j = (Y_{1j}, \ldots, Y_{mj}) \sim Mult(u_j, p_{1j}, \ldots, p_{mj})$, for each I_j
  $\Rightarrow$ **unknown number of unallocated calls assigned to each campaign**

- Drop the notations $\in C_i$ from the probabilities

- $\Delta_{ij} = \Lambda_i(t_j) - \Lambda_i(t_{j-1})$, for each i, j

- For each interval I_j, $\mathcal{P}(n_{1j} + Y_{ij}, \ldots, n_{mj} + Y_{mj}, \underline{Y}_j, u_j) =$

$$
\begin{aligned}
&= \mathcal{P}(n_{1j} + Y_{ij}, \ldots, n_{mj} + Y_{mj} | \underline{Y}_j, u_j) \mathcal{P}(\underline{Y}_j | u_j) \mathcal{P}(u_j) \\
&= \left\{ \prod_{i=1}^{m} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{n_{ij} + Y_{ij}!} e^{-\Delta_{ij}} \right\} \left\{ \binom{u_j}{Y_{1j}, \ldots, Y_{mj}} \prod_{i=1}^{m} p_{ij}^{Y_{ij}} \right\} \mathcal{P}(u_j)
\end{aligned}
$$

- No interest in $\mathcal{P}(u_j) \Rightarrow$ partial likelihood

PARTIAL LIKELIHOOD

For any NHPP

$$\prod_{j=1}^{n} \binom{u_j}{Y_{1j},\ldots,Y_{mj}} \left\{ \prod_{i=1}^{m} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{n_{ij}+Y_{ij}!} e^{-\Delta_{ij}} p_{ij}^{Y_{ij}} \right\} = e^{-\sum_{i=1}^{m} \Lambda(t_n)} \prod_{j=1}^{n} \binom{u_j}{Y_{1j},\ldots,Y_{mj}} \left\{ \prod_{i=1}^{m} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{n_{ij}+Y_{ij}!} p_{ij}^{Y_{ij}} \right\}$$

For a PLP

- $\sum_{i=1}^{m} n_{ij} = n_j \ (j = 1, n)$, $\sum_{j=1}^{n} n_{ij} = N_i^*$ and $\sum_{j=1}^{n} Y_{ij} = U_i^* \ (i = 1, m)$
- $\sum_{j=1}^{n} n_j = N$ and $\sum_{j=1}^{n} Y_j = U$

$$M^{N+U} e^{-M \sum_{i=1}^{m} (t_n - T_i)^{\beta} e^{\gamma' Z_i}} \prod_{i=1}^{m} e^{\gamma' Z_i (N_i^* + U_i^*)} \prod_{j=1}^{n} \binom{u_j}{Y_{1j}, \ldots, Y_{mj}} \cdot$$

$$\cdot \left\{ \prod_{i=1}^{m} \frac{\left\{ [(t_j - T_i)^{\beta} - (t_{j-1} - T_i)^{\beta}] \right\}^{n_{ij}+Y_{ij}}}{n_{ij} + Y_{ij}!} p_{ij}^{Y_{ij}} \right\}$$

PRIORS

- Intervals $I_j, j = 1, n \Rightarrow$ Independent $\underline{p}_j \sim Dir(\alpha_{1j}, \ldots, \alpha_{mj})$
- $\alpha_{ij} = \alpha e^{-\delta(t_j - T_i)} I_{(T_i,\infty)}(t_j)$
  - $\alpha_{ij} = 0$ for unstarted campaigns ($\Rightarrow$ degenerate Dirichlet)
  - decreasing α_{ij} from an interval to next ones (calls unlikely from older ads: $\mathcal{E} p_{ij} = \frac{\alpha_{ij}}{\sum_{l=1}^{m} \alpha_{lm}}$)
  - α, δ: either known or prior
- $M \sim \mathcal{G}(a, b)$
- any $\pi(\beta)$ (nothing convenient for simulations)
- any $\pi(\gamma)$ (nothing convenient for simulations)

POSTERIORS

- $\underline{n}^{(j)} = (n_{1j}, \ldots, n_{mj})$, for interval I_j, $j = 1, n$

- $\underline{n} = (\underline{n}^{(1)}, \ldots, \underline{n}^{(n)})$

- $\underline{p}_j | \underline{Y}_j, \underline{n} \sim Dir(\alpha_{1j} + Y_{1j}, \ldots, \alpha_{mj} + Y_{mj})$, $j = 1, n$

- $\gamma | M, \underline{n}, \beta, \underline{Y}_j$: $\pi(\gamma) e^{-\sum_{i=1}^{m} \{M(t_n - T_i)^{\beta} - (N_i^* + U_i^*)\} e^{\gamma' Z_i}}$

POSTERIORS

- $M|\beta, \underline{n}, \gamma \sim \mathcal{G}(a + N + U, b + \sum_{i=1}^{m}(t_n - T_i)^{\beta} e^{\gamma' Z_i})$

- $\beta|M, \underline{n}, \gamma, \underline{Y}_j \propto$
$\pi(\beta) e^{-M \sum_{i=1}^{m}(t_n - T_i)^{\beta} e^{\gamma' Z_i}} \prod_{j=1}^{n} \prod_{i=1}^{m} [(t_j - T_i)^{\beta} - (t_{j-1} - T_i)^{\beta}]^{n_{ij} + Y_{ij}}$

- $\mathcal{P}(\underline{Y}_j|\beta, M, \gamma, \underline{n}) \propto \binom{u_j}{Y_{1j}, \ldots, Y_{mj}} \prod_{i=1}^{n} \frac{\Delta_{ij}^{n_{ij}+Y_{ij}}}{(n_{ij}+Y_{ij})!} p_{ij}^{\Delta_{ij} Y_{ij}}, \; j = 1, n$

$\Rightarrow$ MCMC simulation (Gibbs with Metropolis steps within)

EXAMPLE

- Weekly time intervals
- 10 campaigns
- Cost of the campaign (in $1000) as unique covariate
- Actual data considered and missing links randomly assigned
- Interest in interval 4 with 3 active campaigns and 4 missing links

EXAMPLE

Posterior probabilities

Y_{ij}	Ad1	Ad2	Ad3
0	0.1362	0.5070	0.2766
1	0.2732	0.3414	0.2654
2	0.2756	0.1246	0.2406
3	0.2208	0.0228	0.1556
4	0.0942	0.0042	0.0618

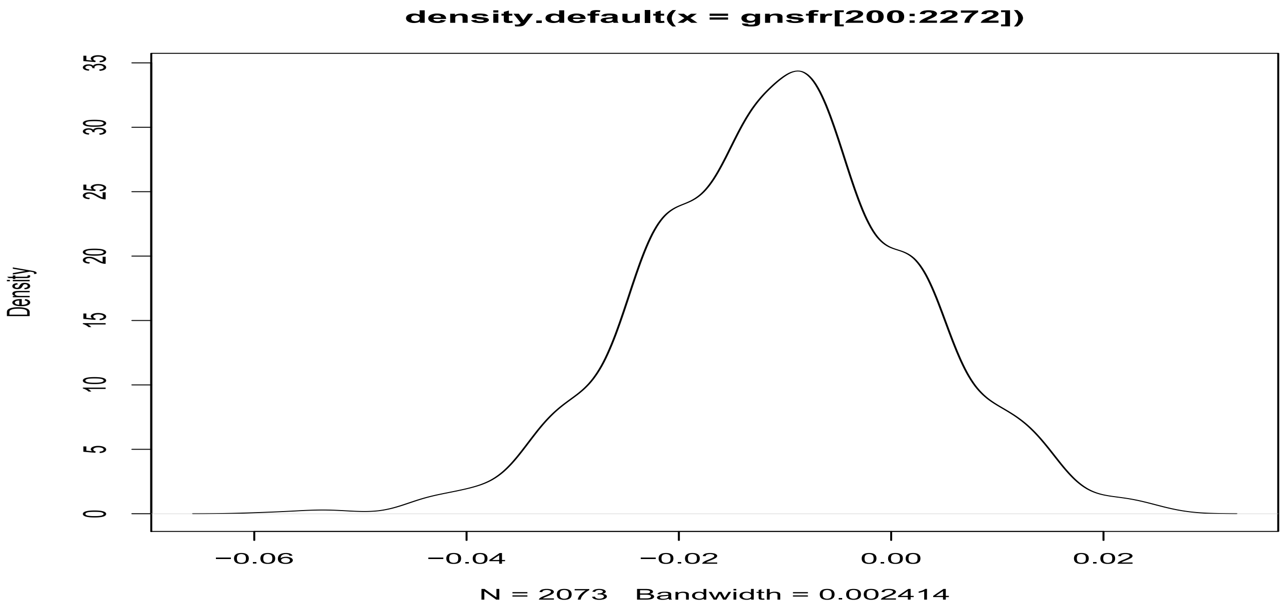
Posterior means and actual values

	Ad1	Ad2	Ad3
Mean	1.86	0.67	1.50
Actual	1	0	3

Predictive distribution of calls in future intervals

EXAMPLE

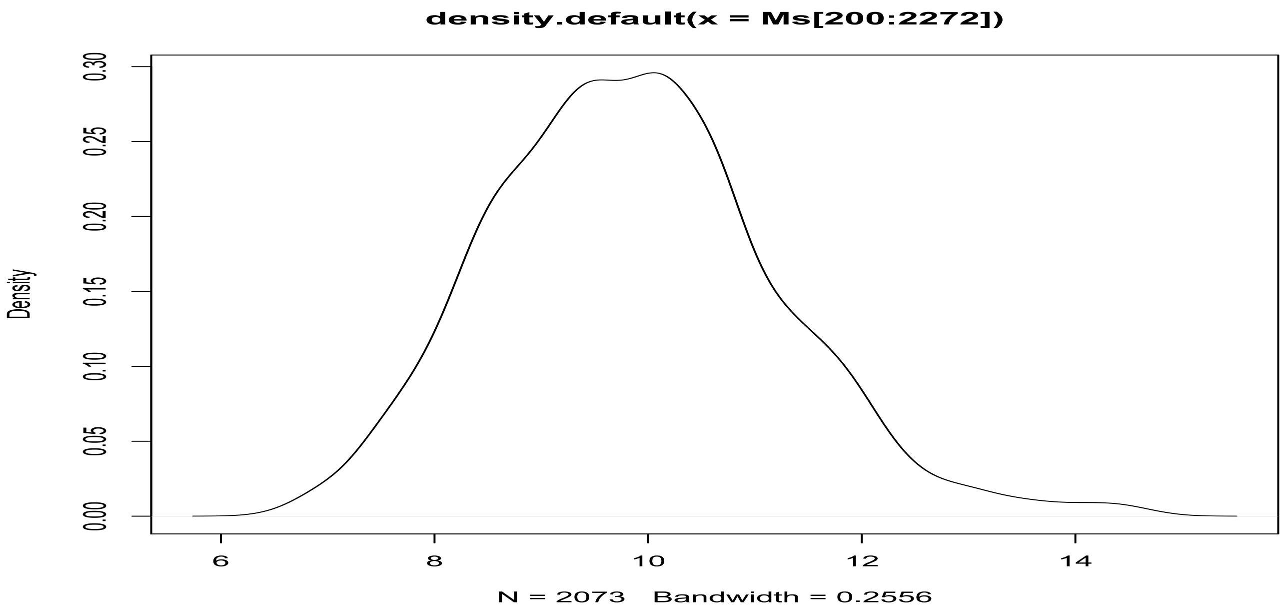
Posterior density of γ (covariate coefficient)



- $Z = 0$ for campaign cost above a threshold and $Z = 1$ under it
- High probability of negative γ
- $\Rightarrow$ higher $\lambda(t)$ for expensive campaign $\Rightarrow$ more calls for it

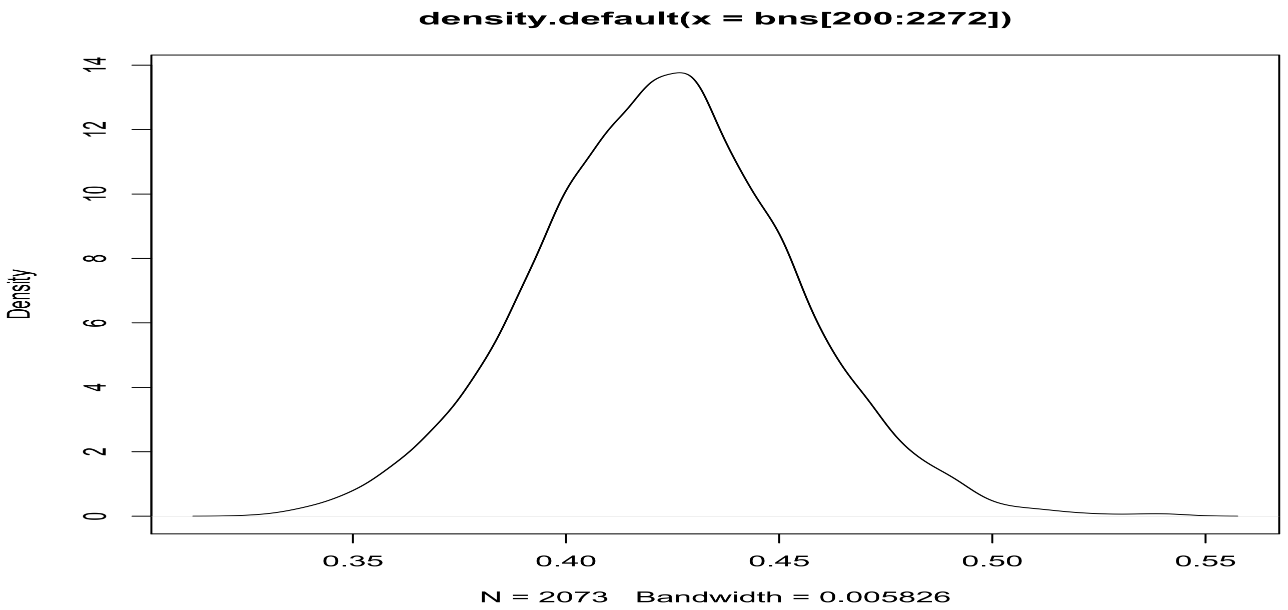
EXAMPLE

Posterior density of M (PLP parameter: $\lambda(t) = Mt^{\beta}$)



EXAMPLE

Posterior density of β (PLP parameter: $\lambda(t) = Mt^{\beta}$)



FUTURE RESEARCH

- More detailed data analysis
- Other models including covariates
- Nonparametric approach

HIERARCHICAL NHPP MODELS

- m campaigns $C_i \Rightarrow m$ PLPs with parameters (M_i, β_i), $i = 1, m$

- Possible priors

  - $\beta_i | \rho, \underline{\delta} \sim \mathcal{G}(\rho \, \mathsf{exp}\{\underline{X}_i^T \underline{\delta}\}, \rho)$

  - $M_i | \nu, \underline{\sigma} \sim \mathcal{G}(\nu \, \mathsf{exp}\{\underline{X}_i^T \underline{\sigma}\}, \nu)$

- $\rho, \nu, \underline{\delta}, \underline{\sigma}$

  - Priors

  - Empirical Bayes

NONPARAMETRIC APPROACH

events in $[T_0, T_1] \sim \mathcal{P}(\Lambda[T_0, T_1]), \Lambda[T_0, T_1] = \Lambda(T_1) - \Lambda(T_0)$

Parametric case: $\Lambda[T_0, T_1] = \int_{T_0}^{T_1} \lambda(t)dt$

Nonparametric case: $\Lambda[T_0, T_1] \sim \mathcal{G}(\cdot, \cdot)$

$\Longrightarrow \Lambda$ d.f. of the random measure M

Notation: $\mu B := \mu(B)$

Definition 1 *Let α be a finite, σ-additive measure on $(\mathbb{S}, \mathcal{S})$. The random measure μ follows a* **Standard Gamma** *distribution with* shape α *(denoted by $\mu \sim \mathcal{GG}(\alpha, 1)$) if, for any family $\{S_j, j = 1, \ldots, k\}$ of disjoint, measurable subsets of $\mathbb{S}$, the random variables μS_j are independent and such that $\mu S_j \sim \mathcal{G}(\alpha S_j, 1)$, for $j = 1, \ldots, k$.*

Definition 2 *Let β be an α-integrable function and $\mu \sim \mathcal{GG}(\alpha, 1)$. The random measure $M = \beta\mu$, s.t. $\beta\mu(A) = \int_A \beta(x)\mu(dx), \forall A \in \mathcal{S}$, follows a* **Generalised Gamma** *distribution, with* shape α *and* scale β *(denoted by $M \sim \mathcal{GG}(\alpha, \beta)$).*

NONPARAMETRIC APPROACH

Consequences:

- $\mu \sim \mathcal{P}_{\alpha,1}, \mathcal{P}_{\alpha,1}$ unique p.m. on $(\Omega, \mathcal{M})$, space of finite measures on $(\mathbb{S}, \mathcal{S})$, with these finite dimensional distributions

- $M \sim \mathcal{P}_{\alpha,\beta}$, **weighted random measure**, with $\mathcal{P}_{\alpha,\beta}$ p.m. induced by $\mathcal{P}_{\alpha,1}$

- $EM = \beta\alpha$, i.e. $\int_\Omega M(A)\mathcal{P}_{\alpha,\beta}(dM) = \int_A \beta(x)\alpha(dx), \forall A \in \mathcal{S}$

NONPARAMETRIC APPROACH

Theorem 1 *Let* $\underline{\xi} = (\xi_1, \ldots, \xi_n)$ *be* n *Poisson processes with intensity measure* M*. If* $M \sim \mathcal{GG}(\alpha, \beta)$ *a priori, then* $M \sim \mathcal{GG}(\alpha + \sum_{i=1}^{n} \xi_i, \beta/(1 + n\beta))$ *a posteriori.*

Data: $\{y_{ij}, i = 1 \ldots k_j\}_{j=1}^{n}$ from $\underline{\xi} = (\xi_1, \ldots, \xi_n)$

Bayesian estimator of M: measure $\widetilde{M}$ s.t., $\forall S \in \mathcal{S}$,

$$\widetilde{M}S = \int_S \frac{\beta(x)}{1 + n\beta(x)} \alpha(dx) + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \frac{\beta(y_{ij})}{1 + n\beta(y_{ij})} \mathbb{I}_S(y_{ij})$$

Constant $\beta \Longrightarrow \widetilde{M}S = \dfrac{\beta}{1 + n\beta} [\alpha S + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \mathbb{I}_S(y_{ij})]$

NONPARAMETRIC APPROACH

Data (calls) recorded as number of events in disjoint intervals

- Comparison between parametric and nonparametric models

- Update of the Gamma process

A Bayesian multi-fractal model with application to analysis and simulation of disk usage

Fabrizio Ruggeri

Consiglio Nazionale delle Ricerche

Istituto di Matematica Applicata e Tecnologie Informatiche

Via Bassini 15 - Milano - Italy

fabrizio@mi.imati.cnr.it

www.mi.imati.cnr.it/˜fabrizio/

Bruno Sansó

Department of Applied Mathematics and Statistics

University of California at Santa Cruz

1156 High Street - Santa Cruz, CA, 95064 - USA

bruno@ams.ucsc.edu

www.ams.ucsc.edu/ bruno

OUTLINE

- Problem description
- Features of data
- (Haar) wavelets
- Model
- Inference (MLE and Bayesian)
- Forecasting
- Future research

PROBLEM DESCRIPTION

- Evaluation of the performance of storage systems
  $\Rightarrow$ design and implementation of computers with heavy I/O workloads
  $\Rightarrow$ optimal size of storage systems in a company
  - cost of storage systems and handling
  - cost of failed I/O operations
- Actual measurements of disk usage (e.g. length of transferred packets (in bytes), `read/write` flag) are difficult and expensive
- Very difficult tasks:
  - model to describe I/O
  - parameter estimation
  - generation of traces (i.e. disk usage history)

FEATURES OF THE DATA

- empirical studies show a slow power-law decrease of the autocorrelation function
  $\Rightarrow$ Long range dependence

- Similar behaviour at different scales
  $\Rightarrow$ Self-similarity ?(possibly attenuated)

- High, abrupt jumps
  $\Rightarrow$ Burstiness

FEATURES OF THE DATA

- Data dependent not only on the last jump
  $\Rightarrow$ no Markov models

- High jumps quite likely
  $\Rightarrow$ no Poisson models

- Data not strictly self-similar and nonnegative
  $\Rightarrow$ no Fractional Brownian motion

  unique Gaussian process with

  - stationary increments
  - statistically self-similar:
    $B(at) \equiv a^H B(t)$ (in the sense of finite distributions)
  - H: Hurst parameter ($1/2 \leq H \leq 1$)

- H changes over time
  $\Rightarrow$ multifractals?

METHOD

H_t varies erratically $\Rightarrow$ Multifractals

We consider an example of multifractal: Multiplicative cascade based on Haar wavelet transform

Work originated by Ribeiro et al (Proceedings ACM Sigmetrics, 1999) and by Hang and Madhyastha (Proceedings MSST 2005), aimed at

- modeling of the process
- statistical estimation of the parameters in a Bayesian fashion
- forecast of other data (i.e. generation of ”similar” data)

INTRODUCTION TO WAVELETS

$\psi \in L^2(\mathbb{R})$:

- $\int_{\mathbb{R}} \psi(x)dx = 0$
- $\int_{\mathbb{R}} \psi^2(x)dx = 1$

$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k), j, k \in \mathbb{Z}$

ψ *wavelet* iff

$\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ orthonormal basis in $L^2(\mathbb{R})$, i.e.

$\forall f \in L^2(\mathbb{R}) \Rightarrow f(x) = \sum_{j,k\in\mathbb{Z}} d_{j,k}\psi_{j,k}(x)$

with $d_{j,k} =< f, \psi_{j,k} >= \int f(x)\psi_{j,k}(x)dx$

INTRODUCTION TO WAVELETS

$\phi(x)$ *scaling function* s.t.

- $\int_{\mathbb{R}} \phi(x)dx = 1$
- $\phi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}} h_k \phi(2x-k)$

h_k filter coefficients ($\exists$ many)

$\Rightarrow \psi(x)$ *wavelet* s.t.

$\psi(x) = \sqrt{2}\sum_{k\in\mathbb{Z}}(-1)^k h_{1-k}\phi(2x-k)$

$\Rightarrow$ wavelet system

$\phi_{j,k}(x) = 2^{j/2}\phi(2^j x-k)$ and

$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x-k), j,k \in \mathbb{Z}$

MULTIRESOLUTION ANALYSIS

$V_j, j \in \mathbb{Z}$, closed subspaces of $L^2(\mathbb{R})$:

- $\ldots V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$
- $\bigcap_j V_j = \emptyset$ and $\bigcup_j V_j = L^2(\mathbb{R})$
- $f(x) \in V_j \Leftrightarrow f(2x) \in V_{j+1}$

$V_0 = \{f \in L^2(\mathbb{R}) : f(x) = \sum_k \alpha_k \phi(x-k)\}$
from scaling function ϕ

$\Rightarrow \{\phi_{j,k}\}_{j,k}$ orthonormal basis for V_j

$\Rightarrow P_j f = \sum_k < f, \phi_{j,k} > \phi_{j,k}$

V_j-approximation (orthogonal projection) of $f \in L^2(\mathbb{R})$

$\Rightarrow W_j$ orthogonal complement of V_j in V_{j+1}, spanned by *details* $\psi_{j,k}, k \in \mathbb{Z}$

$\Rightarrow f(x) = \sum_{k \in \mathbb{Z}} c_{j_0,k} \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x)$

Finite approximations

INTRODUCTION TO WAVELETS

[Haar (1910)]: $\forall f \in \mathcal{C}([0,1]) \Longrightarrow$

- $f(x) \approx f_n(x) =< \xi_0, f > \xi_0(x) +$
  $+ < \xi_1, f > \xi_1(x) + \ldots + < \xi_n, f > \xi_n(x)$,

  with $< \xi_i, f >= \int \xi_i(x) f(x) dx, \forall i$
- f_n converges uniformly to f, as $n \to \infty$

$\xi_0(x) = \mathbf{1}(0 \leq x \leq 1)$

$\xi_1(x) = \mathbf{1}(0 \leq x \leq 1/2) - \mathbf{1}(1/2 \leq x \leq 1)$

$\xi_2(x) = \sqrt{2}[\mathbf{1}(0 \leq x \leq 1/4) - \mathbf{1}(1/4 \leq x \leq 1/2)]$

$\ldots$

$$\begin{aligned} \xi_n(x) &= 2^{j/2}[\mathbf{1}(k2^{-j} \leq x \leq (k+1/2)2^{-j}) - \\ &\quad -\mathbf{1}((k+1/2)2^{-j} \leq x \leq (k+1)2^{-j})] \end{aligned}$$

$n = 2^j + k, j \geq 0, 0 \leq k \leq 2^j - 1$

Note:

$\xi_n(x) = \xi_{jk}(x) = 2^{j/2}\xi_1(2^j x - k), n = 2^j + k$
$\xi_0 \to$ ”average” and $\xi_n(x), n > 0 \to$ ”details”
ξ_0 *scaling function* and $\xi_1(x)$ *wavelet*

Therefore we can describe our data (in our case bytes in I/O in some time units) at different scales

Suppose we have data observed in 1024 intervals of length 1ms, we can consider 512 intervals of length 2 ms, and so on.

Our goals will be:

- describe the process that splits the bytes arrived in 1024 ms in two groups (those arrived in the first 512 ms and those in the last ones) and so on

- generate ”similar” data for intervals of size ”$\ldots, 512, 1024, 2048, \ldots$” ms

CASCADE MODEL

- Haar scaling coefficient u_{jk}:
  local mean of signal at different scales and shifts

- Scaling and wavelet coefficients by recursion . . .
  - $u_{j-1,k} = 2^{-1/2}(u_{j,2k} + u_{j,2k+1})$
  - $w_{j-1,k} = 2^{-1/2}(u_{j,2k} - u_{j,2k+1})$

- . . . which becomes
  - $u_{j,2k} = 2^{-1/2}(u_{j-1,k} + w_{j-1,k})$
  - $u_{j,2k+1} = 2^{-1/2}(u_{j-1,k} - w_{j-1,k})$

- Signal $X(t)$ non negative $\Leftrightarrow u_{j,k} \geq 0$, for all j, k

  $\Rightarrow X(t) \geq 0 \Leftrightarrow |w_{j,k}| \leq u_{j,k}, \forall j, k$

CASCADE MODEL

Wavelet coefficients: realizations of r.v.'s W and U and recursively computed by

$$W_{j-1,k} = A_{j-1,k} U_{j-1,k}$$

$A_{j-1,k}$ r.v. on $[-1, 1] \Rightarrow |W_{j,k}| \leq U_{j,k}$

- $U_{j,2k} = 2^{-1/2}(1 + A_{j-1,k})U_{j-1,k}$

- $U_{j,2k+1} = 2^{-1/2}(1 - A_{j-1,k})U_{j-1,k}$

Ribeiro et al. (1999):

- symmetric Beta distributions for $A_{j-1,k}$ with parameters determined heuristically, e.g. to match the signal's theoretical wavelet-domain energy decay

- coefficients simulated using such distributions

- real data compared with simulated ones

Main difference: we perform a standard *statistical analysis*

CASCADE MODEL

$B_{j-1,k} = (1 + A_{j-1,k})/2$, r.v.'s on $[0, 1]$ for all j, k

$\Rightarrow W_{j-1,k} = (2B_{j-1,k} - 1)U_{j-1,k}$

- $U_{j,2k} = \sqrt{2}B_{j-1,k}U_{j-1,k}$
- $U_{j,2k+1} = \sqrt{2}(1 - B_{j-1,k})U_{j-1,k}$

Correspondence between Haar scaling coefficients $U_{j,k}$ and number of packets $V_{j,k}$ transferred in the intervals $[k2^{-j}, (k+1)2^{-j})$, $j \geq 0, 0 \leq k \leq 2^j - 1$:

- $U_{0,0} = V_{0,0}$
- $U_{j,k} = 2^{j/2}V_{j,k}$

CASCADE MODEL

- $V_{j,2k} = B_{j-1,k} V_{j-1,k}$
- $V_{j,2k+1} = (1 - B_{j-1,k}) V_{j-1,k}$

$$\Rightarrow B_{j-1,k} = \frac{V_{j,2k}}{V_{j,2k} + V_{j,2k+1}}$$

Therefore, we start with a total number of packets $V_{0,0}$ (V_0, from now on) in the interval $[0, 2^T)$, and we split their number in the intervals $[0, 2^{T-1})$ and $[2^{T-1}, 2^T)$ according to $B_{0,0}$, and we keep splitting intervals in halves with $B_{j,k}$ determining the proportion of packets assigned to each interval. Therefore, we model the event history, i.e. the number of transferred packets, using a cascade algorithm.

LIKELIHOOD

$\underline{V}_j = (V_{j,0}, \ldots, V_{j,2^j-1}), j \geq 0$

$$\begin{aligned} P(\underline{V}_n, \ldots, \underline{V}_0) &= P(\underline{V}_n | \underline{V}_{n-1}) \ldots P(\underline{V}_1 | \underline{V}_0) P(\underline{V}_0) \\ &= \prod_{j=1}^{n} \prod_{k=0}^{2^{j-1}-1} P(V_{j,2k}, V_{j,2k+1} | V_{j-1,k}) \cdot P(\underline{V}_0) \\ &= \prod_{j=1}^{n} \prod_{k=0}^{2^{j-1}-1} P(B_{j-1,k}) \cdot P(\underline{V}_0) \end{aligned}$$

- $V_0 \sim \mathcal{P}(\lambda)$, with $\lambda > 0$

  (alternative: $\mathcal{G}eom$)

- $B_{j,k} \sim \pi_{0j}\delta_0 + \pi_{1j}\delta_1 + (1 - \pi_{0j} - \pi_{1j})\mathcal{B}e(p_j, p_j)$

  $0 \leq k \leq 2^j - 1$, $\pi_{0j} \geq 0$, $\pi_{1j} \geq 0$ and $\pi_{0j} + \pi_{1j} \leq 1$

  $\Rightarrow$ possibility of all mass either to left or right

LIKELIHOOD

$$\prod_{k=0}^{2^j-1} P(B_{j,k}) = 1^{N_{2j}} \cdot \pi_{0j}^{N_{0j}} \cdot \pi_{1j}^{N_{1j}} \cdot (1 - \pi_{0j} - \pi_{1j})^{N_{3j}}$$

$$\cdot \quad \prod_{k \in \mathcal{N}_j} \frac{[B_{j,k}^{p_j-1}(1 - B_{j,k})^{p_j-1}]}{B(p_j, p_j)}$$

$$\begin{aligned}
N_{0j} &= \{\#k : V_{j+1,2k} = 0, V_{j+1,2k+1} \neq 0, V_{j,k} \neq 0\} \\
N_{1j} &= \{\#k : V_{j+1,2k} \neq 0, V_{j+1,2k+1} = 0, V_{j,k} \neq 0\} \\
N_{2j} &= \{\#k : V_{j+1,2k} = 0, V_{j+1,2k+1} = 0, V_{j,k} = 0\} \\
N_{3j} &= 2^j - N_{0j} - N_{1j} - N_{2j} \\
\mathcal{N}_j &= \{k : V_{j+1,2k} \neq 0, V_{j+1,2k+1} \neq 0, V_{j,k} \neq 0\}
\end{aligned}$$

Approximation: $V_{j,k}$ should be integer but Beta distributions $\Rightarrow$ noninteger values

MAXIMUM LIKELIHOOD ESTIMATION

- $\widehat{\pi}_{0j} = N_{0j}/(N_{0j} + N_{1j} + N_{3j})$

- $\widehat{\pi}_{1j} = N_{1j}/(N_{0j} + N_{1j} + N_{3j})$

- $\widehat{p_j}$ numerical solution of
  $$\sum_{k \in \mathcal{N}_j} \log[B_{j,k}(1 - B_{j,k})] - 2N_{3j}[\Psi(p_j) - \Psi(2p_j)] =$$
  with Ψ the digamma function

- $\widehat{\lambda} = V_0$

BAYESIAN ESTIMATION

Prior distributions

- $(\pi_{0j}, \pi_{1j}) \sim \mathcal{D}ir(\alpha_{0j}, \alpha_{1j}, \alpha_{2j})$

- $p_j \sim \mathcal{G}(\rho_j, \mu_j)$

  e.g. $\rho_j = c\theta^{-2j}$ and $\mu_j = c\theta^{-j}$

  mean θ^{-j} and variance $1/c$

  $\Rightarrow$ link between different levels of the cascade

- $\lambda \sim \mathcal{G}(\alpha, \beta)$

BAYESIAN ESTIMATION

Posterior distributions

- $(\pi_{0j}, \pi_{1j}) \sim \mathcal{D}ir(\alpha_{0j} + N_{0j}, \alpha_{1j} + N_{1j}, \alpha_{2j} + N_{3j})$

- $\pi(p_j) \propto \frac{p_j^{\rho_j - 1} e^{-\{\mu_j - \sum_{k \in \mathcal{N}_j} \log[B_{j,k}(1 - B_{j,k})]\} p_j}}{B(p_j, p_j)^{N_{3j}}}$

- $\lambda \sim \mathcal{G}(\alpha + V_0, \beta + 1)$

BAYESIAN ESTIMATION

Bayesian estimators, under squared losses

- $\widehat{\lambda} = \frac{\alpha + V_0}{\beta + 1}$

- $\widehat{\pi}_{0j} = \frac{\alpha_{0j} + N_{0j}}{\alpha_{0j} + N_{0j} + \alpha_{1j} + N_{1j} + \alpha_{2j} + N_{3j}}$

- $\widehat{\pi}_{1j} = \frac{\alpha_{1j} + N_{1j}}{\alpha_{0j} + N_{0j} + \alpha_{1j} + N_{1j} + \alpha_{2j} + N_{3j}}$

- $\widehat{p_j}$ can be only computed numerically

FORECASTING AND SIMULATION

Real traces (i.e. data on disk usage) are expensive
$\Rightarrow$ interest in simulating data

- at finer (than observed) levels
- in future intervals
- for ”similar” traces

Forecasting of packets in future intervals and simulation of similar traces, without going to finer than observed levels imply:

- Simulation of total number of packets
- Use of same models, i.e. $B_{j,k}$, parameters and posteriors as for observed data (under suitable condition), e.g.
  - same interval in ”similar” traces
  - in, say, $(0, 512]$ and $(1024, 1536]$, when $(0, 1024]$ is observed and $(1024, 2048]$ is the future interval

Details only on simulation at finer levels

FORECASTING AND SIMULATION

Simulation of total number of packets

$V_0 \sim \mathcal{P}(\lambda)$ and $\lambda \sim \mathcal{G}(\alpha, \beta)$
$\Rightarrow \lambda | V_0 = v_0 \sim \mathcal{G}(\alpha + v_0, \beta + 1)$

Posterior predictive distribution, $v \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}(V_0 = v) &= \int_0^\infty \mathbb{P}(V_0 = v|\lambda)\pi(\lambda|\alpha, \beta, v_0)d\lambda \\
&= \int_0^\infty \frac{\lambda^v}{v!} e^{-\lambda} \frac{\lambda^{\alpha+v_0-1}(\beta+1)^{\alpha+v_0} e^{-(\beta+1)\lambda}}{\Gamma(\alpha+v_0)} d\lambda \\
&= \frac{1}{v!} \frac{\Gamma(\alpha+v_0+v)}{\Gamma(\alpha+v_0)} \frac{(\beta+1)^{\alpha+v_0}}{(\beta+2)^{\alpha+v_0+v}}
\end{aligned}$$

FORECASTING AND SIMULATION

Simulation at finer, unobserved levels

- Observed data at levels $j = 0, \ldots, J$
  $\Rightarrow$ models for $B_{j,k}$ and inference on parameters of their distributions

- Goal: split data in smaller intervals at unobserved level $J + 1$

- Model for r.v.'s $B_{J+1,k}$, $0 \leq k \leq 2^{J+1} - 1$:
  $\pi_{0,J+1}\delta_0 + \pi_{1,J+1}\delta_1 + (1 - \pi_{0,J+1} - \pi_{1,J+1})\mathcal{B}e(p_{J+1}, p_{J+1})$

- Parameter $\omega_{J+1} = (\pi_{0,J+1}, \pi_{1,J+1}, p_{J+1})$

- Data at levels $j = 0, \ldots, J$
  $\Rightarrow f(\omega_{J+1}|data)$ (different choices)

- Simulation of $B_{J+1,k}$:
  $f(B_{J+1,k}|data) = \int f(B_{J+1,k}|\omega_{J+1}) f(\omega_{J+1}|data) d\omega_{J+1}$

FORECASTING AND SIMULATION

Consider p_{J+1} and predictive $f(p_{J+1}|data)$
(similar for $\pi_{0,J+1}$ and $\pi_{1,J+1}$)

Exchangeable p_j's

- Same distribution $f(p_j|\rho, \mu)$ for all p_j's, $j \geq 0$

- Prior on (ρ, μ)

- Data at levels $j = 0, \ldots, J \Rightarrow$ posterior $\pi(\rho, \mu|data)$

- Simulation of p_{J+1}:
  $f(p_{J+1}|data) = \int f(p_{J+1}|\rho, \mu)\pi(\rho, \mu|data)d\rho d\mu$

FORECASTING AND SIMULATION

Markovian p_j's

(More complex Markovian structures can be considered as well)

- p_0 given

- $p_j = p_{j-1}\lambda$, $j \geq 1$

- Prior on λ, updated by data up to level J

- Simulation of p_{J+1}:
  $f(p_{J+1}|data) = \int f(p_{J+1}|\lambda)\pi(\lambda|data)d\lambda$

FORECASTING AND SIMULATION

Empirical Bayes

- Prior $p_j \sim \mathcal{G}(c\theta^{-2j}, c\theta^{-j})$ (as before)
- $\mathbf{p}$ and π: all $\pi_{0,j}, \pi_{1,j}$ and p_j at levels $j = 0, \ldots, J$
- Look for $\widehat{\theta}$ maximising
  $f(data) = \int f(data|\mathbf{p}, \pi) f(\mathbf{p}|\theta) f(\pi) d\mathbf{p} d\pi$
- Consider the prior $p_{J+1} \sim \mathcal{G}(c\widehat{\theta}^{-2j}, c\widehat{\theta}^{-j})$

FUTURE RESEARCH

- Application to actual I/O data (Bellcore and UCSC)
  - exploratory data analysis to check assumptions (non-Markov, non-Poisson, etc.)
  - estimation
  - forecast
- Possible extension to other data: monthly rainfall in different Venezuelan areas
  - forecast future monthly rainfall in the same areas
  - estimation of monthly rainfall at finer level (smaller areas) $\Rightarrow$ rely on expert opinion on the choice of prior parameters!

FUTURE RESEARCH

- Links with Polya Trees, a nonparametric model to assign probability measures on the space of probability measures, through a sequence of Beta r.v.'s on a dyadic expansion of $[0, 1]$
  - established theory for Beta r.v.'s (Lavine, 1992, 1994)
  - mixture of Beta and two Dirac instead of Beta r.v.'s $\Rightarrow$ changes in the PT theory?
  - Data recorded in intervals and not pointwise
  - Random total mass
- Discrete distribution for $B_{j,k}$ instead of mixture of Dirac and Beta
  - Poisson distribution?
$\Rightarrow$ link with Gamma Process (Lo, 1982)?

FUTURE RESEARCH

- Similarity between actual and simulated data?
  $\Rightarrow$ look at different aspects, e.g.

  - distribution of queue length, if possible

  - divergence measures (e.g. Kullback-Leibler), Lorenz curve and concentration function

  - autocovariance function

  - spectrum estimation, related to multi-fractality and occurrence of values of H_t

- Decision problem, through expected utility maximisation, e.g. optimal size of a storage system

  - probability of number of concurrent packets exceeding the system size

  - cost for failed I/O operations and system size

APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2004; **20**:305–321
Published online 16 July 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.542

Long-range dependence and performance in telecom networks

Pier Luigi Conti[1,‡], Antonio Lijoi[2,*†] and Fabrizio Ruggeri[3,§]

[1]*Dipartimento di Statistica, Probabilità e Statistica Applicate, Università di Roma "La Sapienza", Italy*
[2]*Dipartimento di Economia Politica e Methodi Quantitativi, Università di Pavia, Italy*
[3]*Istituto di Matematica Applicata e Tecnologie Informatiche, C.N.R., Milano, Italy*

SUMMARY

Telecommunications systems have recently undergone significant innovations. These call for suitable statistical models that can properly describe the behaviour of the input traffic in a network. Here we use fractional Brownian motion (FBM) to model cumulative traffic network, thus taking into account the possible presence of long-range dependence in the data. A Bayesian approach is devised in such a way that we are able to: (a) estimate the Hurst parameter H of the FBM; (b) estimate the overflow probability which is a parameter measuring the quality of service of a network: (c) develop a test for comparing the null hypothesis of long-range dependence in the data versus the alternative of short-range dependence. In order to achieve these inferential results, we elaborate an MCMC sampling scheme whose output enables us to obtain an approximation of the quantities of interest. An application to three real datasets, corresponding to three different levels of traffic, is finally considered. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: fractional Brownian motion; long-range dependence; overflow probability; teletraffic data

1. INTRODUCTORY ASPECTS AND MOTIVATIONS

The introduction of new technologies for telecommunications, based on packet switched networks, has led to new teletraffic problems. Due to the exponential growth of the Internet, the study and analysis of telecommunications is of considerable and increasing importance. In Internet communications, the transmission of data files, e-mail, video signals, etc. generates an information stream. The user generating such a stream is commonly referred to as a traffic source. The information stream produced by a traffic source is segmented into variable size packets called *datagrams*, according to the Internet Protocol (IP, for short); see Reference [1]. A datagram is composed by a header area and a data area. The header area essentially contains routing information, i.e. the source and destination IP addresses, as well as the information to interpret the data area. IP datagrams are routed through IP routers, which interconnect input links with output links. Output links are equipped with buffers to store and schedule IP datagrams for transmission.

*Correspondence to: Antonio Lijoi, Dipartimento di Economia Politica e Methodi Quantitativi, Università di Pavia, Italy.
†E-mail: lijoi@unipv.it
‡E-mail: pierluigi.conti@uniroma1.it
§E-mail: fabrizio@mi.imati.cnr.it

Received 8 November 2002
Revised 22 October 2003
Accepted 15 March 2004

Apart from the case of dedicated telephone lines, IP packets are transmitted using standard commercial telephone lines. The common transport tool in telephone lines is the asynchronous transfer mode (ATM) technique; see, e.g. Reference [2]. In ATM, the information stream produced by a user (traffic source) is split into fixed-length packets. To identify both the source and the destination, a fixed-length label is added to each information packet, to form an ATM *cell*. According to their labels, cells are routed through ATM nodes connecting input links with output links. The typical behaviour of a traffic source consists of alternating activity (ON) and silence (OFF) periods. The transmission rate is usually constant during each ON period. Because of the presence of several traffic sources simultaneously connected, it is far from being unusual that different cells simultaneously require the same output link. To overcome the competition among these cells, a buffer is used, and cells that cannot be immediately transmitted are stored in it. This means that an ATM element is characterized by a *queue* of cells at the output link. Since the buffer size is finite, a cell entering the system when the buffer is already full cannot be either transmitted or stored in the buffer and, then, it is *lost*.

New standards have been recently defined in order to transmit IP datagrams by ATM technique *via* 'Cell switch routers' (see References [1,3,4]). The basic idea consists in interconnecting IP routers by ATM links. IP datagrams are first fragmented into ATM cells, then transmitted by an ATM link and finally reassembled. If one ATM cell originated by the fragmentation of an IP datagram is lost in the output buffer, then all the other ATM cells belonging to the same IP datagram are lost, and the IP datagram must be retransmitted.

Packet switching networks are essentially networks of queues. The evaluation of the performance of a single server queue, the ATM multiplexer, composed by an ATM link with buffer, is a fundamental step in assessing the performance of ATM networks and, because of the use of ATM switching fabrics in IP routers, of IP networks. This last point is particularly relevant, because of the recent growth of applications using Internet protocol. IP technology does not eliminate the need to deploy ATM networks, because ATM offers a standard set of traffic management mechanisms that can inter-operate among different providers to allow efficient support for different types of services and effectively guarantee a good quality of service (QoS) to the connections.

Telecommunication networks are characterized by QoS requirements. A fundamental QoS parameter is the *cell loss probability*, as suggested in Reference [5] and the ITU-T Recommendation I.356 Reference [6]. The cell loss probability is defined as the 'long term fraction of lost cells'. As a convenient approximation of it, the *overflow probability* is commonly used; see, e.g. Reference [5]. It is defined as the probability that the number of cells in an infinite queue exceeds the buffer size. Cell loss probability is the most important QoS parameter in ATM and/or IP networks. As already stressed, in applications using the IP protocol, if an ATM cell is lost, then the corresponding IP datagram is completely lost, and must be retransmitted. This may clearly cause the congestion of communications networks and delay in the transmission. The higher the cell loss probability, the stronger the phenomena of congestion and transmission delay.

In order to guarantee an acceptable QoS, it is of primary importance to have at least an idea of the corresponding cell loss probability (or better, of the overflow probability). In ATM networks, the traffic is controlled by the connection admission control (CAC) function. Before establishing a new connection, the corresponding source is asked to declare some standard intensity traffic parameters. On the basis of such parameters, the CAC function computes the cell loss probability, and then checks if the system has enough resources to accept the new connection without infringing QoS requirements. On the other hand, in IP networks there is no preventive traffic control. Sources do not declare any intensity traffic parameters, so that the cell loss probability cannot be computed. This motivated the need of estimating the cells loss probability on the basis of observed data. In fact, the evaluation of the cell loss probability allows one to answer some basic problems, such as determining the utilization level of a link such that the QoS requirements are met. Therefore the statistical estimation of the buffer overflow probability is unavoidable if one wants to assess the performance of ATM and/or IP networks.

The main contribution of the present paper is a Bayesian approach to the estimation of the overflow probability. In detail, Section 2 contains a description of the model and the related and related statistical problems. In Section 3, a Bayesian technique to estimate the overflow probability is developed. Finally, in Section 4 an application to real data is considered. Data come from measurements made in Italy by Telecom Italia, in the framework of the European ATM Pilot Project. The applications considered are videoconference, teleteaching, and transport of routing information between IP network routers. All applications use the Internet Protocol over ATM, as described above.

2. THE MODEL AND ITS MOTIVATIONS

Stochastic models for packed switched traffic traditionally fall into one of two categories: burst scale models, and cell scale models; see Reference [5, pp. 309, 310, 389] for a good description. Burst scale models are based on the fluid flow approximation for the packet stream produced by a source. Cell level models are primarily useful for ATM traffic data; cf. e.g. Reference [7]. They are essentially based on the idea that all transmission systems work in discrete time. In fact, there exists an elementary time unit, the *time slot*, such that no more than one cell per time slot can be transmitted. The relative merits of these two different approaches to modelling teletraffic phenomena is briefly discussed, for instance, in Reference [5, Chapters 16, 17]. We adopt here the burst scale approach, which proves useful especially when studying the characteristic of the aggregated traffic produced by several users simultaneously connected.

Suppose that N sources are simultaneously connected to a traffic node, and let $A_i(t)$ be the amount of traffic generated by the ith source during the time interval $(0, t]$, $t > 0$, $i = 1, \dots, N$. Furthermore, let

$$X(t) = \sum_{i=1}^{N} A_i(t) \tag{1}$$

be the global amount of traffic generated by the N sources up to time t. In the sequel, the stochastic process $(X(t);\ t > 0)$ will be referred to as the 'cumulative arrival process'. Suppose that the service time (i.e. the channel capacity) is constant, and equal to c, and let V be the unfinished work of the system. It is not difficult to show (cf. Reference [8]) that the following equality in distribution holds true:

$$V \overset{\mathrm{d}}{=} \sup_{t \geqslant 0} (X(t) - ct) \tag{2}$$

Relationship (2) is of basic importance in studying the performance of telecommunication systems. Let u be the buffer size. Then, the *overflow probability*, closely related to the loss probability, is equal to

$$Q(u) = P(V > u) = P\left(\sup_{t \geqslant 0} (X(t) - ct) > u \right) \tag{3}$$

As mentioned before, the overflow probability is the most important measure of performance for telecommunication systems, since it is a good approximation of the loss probability for buffered systems. Formula (3) shows that the problem of evaluating the overflow probability essentially consists of studying the distribution of the supremum of a stochastic process.

Hence, some assumptions on the process ($X(t)$; $t>0$) are in order. A common hypothesis is that it is a Gaussian process with stationary increments. Such an assumption essentially rests on (1) and the functional central limit theorem. The assumptions on the covariance function of ($X(t)$; $t\geqslant 0$) are more delicate. In the sequel, we will suppose that the process can be expressed in the form

$$X(t) = \mu t + \sigma Z(t) \tag{4}$$

where ($Z(t)$; $t\geqslant 0$) is a fractional Brownian motion (FBM, for short). It is characterized by the following properties:

(i) $Z(t)$ is a Gaussian process with stationary increments.
(ii) $Z(0) = 0$ a.s.
(iii) $E[Z(t)]=0$ and $E[Z(t)^2]=t^{2H}$ for all positive t.

The parameter H is the *Hurst parameter*: it takes values in the interval $(\frac{1}{2}, 1)$ and, if $H=\frac{1}{2}$, $Z(t)$ reduces to the standard Brownian motion.

Model (4), with $Z(t)$ FBM, was first proposed as a realistic model for aggregated traffic by Norros [9,10]. Its most important feature is that it is a *self-similar* process: $Z(\alpha t) =^{d} \alpha^H Z(t)$ for every positive α. Clearly, the process $(X(t)-\mu t)$ possesses the same property. The self-similar nature of Ethernet traffic was first shown by statistical (frequentist) analysis of Bellcore traffic data. See References [11,12]. A good bibliographical guide for the subject is in Reference [13].

As a consequence, the increments of $Z(t)$ (and those of $X(t)$, as well) are stationary with *long-range dependence* whenever $H>\frac{1}{2}$. To be precise, let $Y_i = X(i)-X(i-1)$, $i=1,\ldots,N$. From the well-known formula (see Reference [14], pp. 52, 56)

$$E[Z(t)Z(s)] = \tfrac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}) \qquad \forall t,\ s\geqslant 0$$

it is seen that the correlation coefficient between the increments Y_i and Y_{i+k} is equal to

$$\rho(k) = \tfrac{1}{2}((k+1)^{2H} - 2k^{2H} + (k-1)^{2H}) \qquad \forall k\geqslant 1 \tag{5}$$

The most important property of (5) is that $\rho(k)$ tends to zero very slowly as k tends to infinity. In fact, by a Taylor expansion it is easy to see that $\rho(k) \sim H(2H-1)k^{2(H-1)}$ as $k\to\infty$, for every $\frac{1}{2}<H<1$.

Long-range dependence is essentially generated by ON and/or OFF periods with infinite variance; see Reference [5]. From a practical point of view, this means that when the traffic sources generate traffic with high variability, where the ON periods can be very long (cf. Reference [15]), one should expect that the aggregated traffic stream entering the transmission system, $X(t)$, is characterized by the presence of long-range dependence. This is important not only from a theoretical point of view. In fact, long-range dependence could potentially have a great influence on the performance of telecommunication systems, since it could considerably increase the overflow probability; see Reference [16]. Furthermore, stochastic models that do not allow for long-range dependence could severely underestimate the loss probability. These two facts provide the most important justifications to the use of model (4).

Unfortunately, even for model (4) the loss probability (3) cannot be written in a closed form. However, using a result by Hüsler and Piterbarg [17] it can be shown that, under the stability condition $\mu < c$, the following holds:

$$q_u(H, \mu, \sigma^2) := Q(u) \sim \frac{\sqrt{\pi}(c-\mu)^{1-H}(1-H)^{3/2-H-1/H} H_{2H}}{H^{3/2-H} 2^{(H-1)/(2H)} \sigma^{1/H-1}} u^{(1-H)^2/H} \Psi(A\sigma^{-1}u^{1-H}) \tag{6}$$

as u tends to infinity, where $\Psi(\cdot)$ is the survival function of normal standard distribution,

$$A = \frac{(c-\mu)^H}{H^H(1-H)^{1-H}}$$

and

$$H_{2H} = \lim_{t\to\infty} (1/t) E\left[\exp\left(\max_{0\leqslant s\leqslant t} (-s^{2H} + \sqrt{2}Z(s))\right)\right] \tag{7}$$

$Z(t)$ being a FBM. Since the usual buffer size is $u = 500$, or $u = 1000$, the asymptotic approximation (6) is satisfactory.

The exact value of constant (7) is not known. Luckily enough, in Reference [18] it is shown that

$$0.12 \leqslant H_{2H} \leqslant 3.1 \tag{8}$$

Estimate (8) will be used in the sequel.

Remark 1

The most important part of relationships (8) is the upper bound 3.1, at least from a practical point of view. In fact, as already outlined in the Introduction, communication providers should guarantee a loss probability smaller than a given threshold. Hence, the upper bound in (8) is much more important than the lower bound.

Remark 2

The constant H_{2H} is given an equivalent definition in Reference [19]. It suggests, as the author himself points out, the possibility of evaluating H_{2H} numerically for some values of $2H$.

Observe that the unfinished work V is infinite a.s. whenever $\mu \geqslant c$, so that in our setting the loss probability turns out to be equal to

$$Q(u) \sim \frac{\sqrt{\pi}(c-\mu)^{1-H}(1-H)^{3/2-H-1/H} H_{2H}}{H^{3/2-H} 2^{(H-1)/(2H)} \sigma^{1/H-1}} u^{(1-H)^2/H}\, \Psi(A\sigma^{-1}u^{1-H}) I_{\mu<c} + I_{\mu\geqslant c} \tag{9}$$

where $I_{\mu<c}$ is 1 if $\mu < c$ and is zero otherwise (similarly one defines $I_{\mu\geqslant c}$).

Since in applications the parameters of model (4) are unknown, they must be estimated by the observed data. The goal of the present paper is to propose a Bayesian approach to such an estimation problem. More specifically, in Section 3, the priors for the unknown parameters are introduced, and updated on the basis of the sample data. Since the posteriors cannot be expressed in a closed form, a computational scheme based on MCMC is adopted. As a by-product, a Bayesian approach to the problem of testing for independence ($H = \frac{1}{2}$) against long-range dependence ($H > \frac{1}{2}$) is obtained. Finally, in Section 4 an application to real data is provided.

3. BAYESIAN ANALYSIS

According to guidelines provided in the previous sections, we now proceed to illustrate the Bayesian set-up for our statistical analysis.

Let n be the sample size and let $t_1, \ldots, t_n$ be fixed time instants. Correspondingly, n observations of the process $\{X(t):\ t \geqslant 0\}$ are made and they are denoted by $X(t_1) = x_1, \ldots, X(t_n) = x_n$. For the sake of brevity, in the sequel, we will use the vector notation

$$\boldsymbol{x} = (x_1, \ldots, x_n)' \quad \boldsymbol{t} = (t_1, \ldots, t_n)'$$

Moreover, Ω_H denotes (apart from the constant σ^2) the covariance matrix of $(X(t_1), \ldots, X(t_n))$, whose (i, j)th element is

$$\omega_{i,j}(H) = \tfrac{1}{2}(t_i^{2H} + t_j^{2H} - |t_i - t_j|^{2H})$$

Since the process $(X(t);\ t \geqslant 0)$ is assumed to be a FBM, the likelihood function coincides with

$$f(H, \mu, \sigma^2; \boldsymbol{x}) = \frac{|\Omega_H|^{-1/2}}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x} - \mu\boldsymbol{t})'\Omega_H^{-1}(\boldsymbol{x} - \mu\boldsymbol{t})\right\}$$

with $|\Omega_H|$ denoting the determinant of matrix Ω_H. As far as prior specification for the vector of parameters $(H, \mu, \sigma^2) \in [\frac{1}{2}, 1) \times \mathbb{R} \times \mathbb{R}^+$ we set

$$\pi_0(\mathrm{d}H) = \varepsilon\delta_{1/2}(\mathrm{d}H) + (1 - \varepsilon)\pi_0^*(H)\boldsymbol{I}_{1/2<H<1}\,\mathrm{d}H$$

$$\pi_1(\mu|\sigma^2) = \frac{1}{\sigma\sqrt{2\pi w}} \exp\left(-\frac{(\mu - \mu_c)^2}{2\sigma^2 w}\right)$$

$$\pi_2(\sigma^2) = \frac{\lambda^v}{\Gamma(v)}\left(\frac{1}{\sigma^2}\right)^{v+1} \exp\left(-\frac{\lambda}{\sigma^2}\right)$$

for some $\varepsilon \in [0,1]$, and $\delta_x(\cdot)$ is the Dirac function at x. The prior specification we are adopting deserves some further explanation. The prior π_0 for the Hurst parameter H is essentially motivated by the following important fact: the (FBM) input process does possess completely different characteristics according to the values of H. In particular, if $H = \frac{1}{2}$, it reduces to a standard Brownian motion, which is an independent increments process (with short-range dependence behaviour and strong Markov, as well). If $H \in (\frac{1}{2}, 1)$ then the (FBM) input process possesses increments with long-range dependence and it turns out to be non-Markov. The prior π_0 takes into account this basic fact, and allows to compare short- vs long-range dependence. The value of ε measures the degree of prior belief about short-range dependence of the original series. Moreover, the diffuse component π_0^* on $(\frac{1}{2},1)$ is taken to be a uniform distribution so to reflect lack of prior information concerning the strength of the long-range dependence behaviour, if present.

As far as the prior for μ and σ^2 are concerned, they depend upon the hyperparameters λ, v, w and μ_c. A discussion of their choice is postponed to Section 4, where sensitivity of posterior estimates is considered as well.

3.1. Estimation of H

Since Bayes' theorem can be applied in our case, the posterior distribution of (H, μ, σ^2), given the vector of observations $\boldsymbol{x} = (x_1, \ldots, x_n)'$, is

$$\pi(\mathrm{d}H, \mathrm{d}\mu, \mathrm{d}\sigma^2|\boldsymbol{x}) \propto \pi_0(\mathrm{d}H)\pi_1(\mu|\sigma^2)\pi_2(\sigma^2)f(H, \mu, \sigma^2; \boldsymbol{x})\,\mathrm{d}\mu\,\mathrm{d}\sigma^2$$

where $\propto$ means that equality holds true up to a normalizing constant. In order to obtain a posterior estimate of the self-similarity parameter H, we determine its posterior distribution by integrating out μ and σ^2 in the joint posterior distribution, so that one has

$$\pi(\mathrm{d}H|\boldsymbol{x}) \propto \left\{\frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}}\delta_{1/2}(\mathrm{d}H) + \frac{(1-\varepsilon)|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H\right\}$$

where

$$\zeta_H := \boldsymbol{t}'\Omega_H^{-1}\boldsymbol{t} + \frac{1}{w}, \qquad \xi_H := \boldsymbol{t}'\Omega_H^{-1}\boldsymbol{x} + \frac{\mu_c}{w}, \qquad M_H := \lambda + \frac{\boldsymbol{x}'\Omega_H^{-1}\boldsymbol{x}}{2} + \frac{\mu_c^2}{2w} - \frac{\xi_H^2}{2\zeta_H}$$

are, for any H in $[\frac{1}{2}, 1)$, computable. On the contrary, the normalizing constant

$$\begin{aligned} k^*(\boldsymbol{x}) &= \varepsilon\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}} + (1-\varepsilon)\int_{(1/2,1)}\frac{|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H \\ &=: \varepsilon k_d(\boldsymbol{x}) + (1-\varepsilon)k_c(\boldsymbol{x}) \end{aligned}$$

has to be approximated. With the posterior distribution $\pi(\mathrm{d}H|x)$ at hand, one can obtain the posterior mean for H

$$E(H|\boldsymbol{x}) = (k*(\boldsymbol{x}))^{-1}\left\{\frac{\varepsilon}{2}\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2}M_{1/2}^{v+n/2}} + (1-\varepsilon)\int_{(1/2,1)}\frac{H|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{v+n/2}}\pi_0^*(H)\,\mathrm{d}H\right\} \tag{10}$$

Since the integral appearing in the right-hand side of (10) cannot be exactly evaluated, we approximated it numerically. Details on the approximation procedures for $k^*(x)$ and $E(H|x)$ are illustrated in Appendix A.1.

3.2. Estimation of loss probability

As far as the problem of evaluating posterior loss probability, as a measure of performance of the system, is concerned, using formula (9) and the dominated convergence theorem, we have that

$$\begin{aligned} &P\left(\sup_{t\geqslant 0}(X(t)-ct) > u|\boldsymbol{x}\right) \\ &\quad\sim \int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} q_u(H,\mu,\sigma^2)\pi(\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2|\boldsymbol{x})\boldsymbol{I}_{(\mu<c)} + P(\mu\geqslant c|\boldsymbol{x}) \end{aligned}$$

holds true for large values of u. Hence, one has

$$\begin{aligned}\int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} & q_u(H,\mu,\sigma^2)1_{(\mu<c)}\pi(\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2|x) \\ & \leqslant \varepsilon \frac{k_d(\boldsymbol{x})}{k^*(\boldsymbol{x})}\int_{\mathbb{R}\times\mathbb{R}^+} \bar{q}_u\left(\tfrac{1}{2},\mu,\sigma^2\right)\pi^*\left(\tfrac{1}{2},\mu,\sigma^2|\boldsymbol{x}\right)\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}\mu\,\mathrm{d}\sigma^2 \\ & \quad + (1-\varepsilon)\frac{k_c(\boldsymbol{x})}{k^*(\boldsymbol{x})}\int_{(1/2,1)\times\mathbb{R}\times\mathbb{R}^+} \bar{q}_u(H,\mu,\sigma^2)\pi^{**}(H,\mu,\sigma^2|\boldsymbol{x})\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2 \end{aligned} \tag{11}$$

where $\bar{q}_u$ is an upper bound for q_u, obtained by substituting H_{2H} with 3.1. Moreover, $\pi^*(\cdot|x)$ and $\pi^{**}(\cdot|x)$ are probability distributions on $\mathbb{R}\times\mathbb{R}^+$ and on $(\frac{1}{2},1)\times\mathbb{R}\times\mathbb{R}^+$, respectively. One easily checks that the determination of an upper bound for the loss probability requires the numerical evaluation of $k_c(x)$ and of the two integrals appearing in (11) above. We implemented an MCMC algorithm whose features are fully described in the appendix.

However, the algorithm we have resorted to might be computationally cumbersome in some cases, because of the presence of the term $\Psi(A\sigma^{-1}u^{1-H})$ in (9). Considerable simplifications are obtained by virtue of the following well-known inequality for the Mills' ratio of the Gaussian distribution

$$\Psi(A\sigma^{-1}u^{1-H}) < \frac{\sigma}{\sqrt{2\pi}A\sigma^{1-H}}\exp\left\{-\frac{1}{2\sigma^2}A^2u^{2-2H}\right\} \tag{12}$$

see, e.g. Reference [20, p. 49]. Using (12), we obtain

$$q_u(H,\mu,\sigma^2) < \frac{3.1(c-\mu)^{1-2H}(1-H)^{(5/2)-2H-(1/H)}u^{(H-1)(2H-1)/H}}{2^{1/(2H)}H^{3/2-2H}\sigma^{1/H-2}}\exp\left\{-\frac{A^2u^{2-2H}}{2\sigma^2}\right\}$$

and integration w.r.t. σ^2 gives, for an appropriate function $\bar{q}_u^*(H,\mu;\boldsymbol{x})$.

$$\begin{aligned}\int_{[1/2,1)\times\mathbb{R}\times\mathbb{R}^+} & q_u(H,\mu,\sigma^2)\pi(\mathrm{d}H,\mathrm{d}\mu,\mathrm{d}\sigma^2|\boldsymbol{x}) \\ & < \int_{[1/2,1)\times\mathbb{R}} \bar{q}_u^*(H,\mu;\boldsymbol{x})\pi(\mathrm{d}H|\boldsymbol{x})\pi(\mu|H,\boldsymbol{x})\,\mathrm{d}\mu\end{aligned} \tag{13}$$

where

$$\pi(\mu|H,\boldsymbol{x}) = \frac{\Gamma((n+2v+1)/2)}{\sqrt{\pi}\,\Gamma((n+2v)/2)}\sqrt{\frac{\zeta_H}{2M_H}}\left\{1+\frac{\zeta_H}{2M_H}\left(\mu-\frac{\xi_H}{\zeta_H}\right)^2\right\}^{-(n+2v+1)/2}$$

In order to provide an approximation of the integral in the right-hand side of (13), we implement both a Monte Carlo i.i.d. sampling and a Metropolis–Hastings scheme as described in appendix A.2.

3.3. A model comparison problem: short- vs long-range dependence

An important problem is to study the essential features of traffic, and in particular the presence/absence of self-similarity. Apropos of this we mention that there is a great debate in the literature in order to assess the characteristics of traffic in telecommunication systems, see the key paper by Willinger *et al.* [11] and the references in Reference [13]. See also Reference [15] for further important remarks and bibliographic references. The classical approach [11,12] consists of constructing an asymptotic confidence interval for the Hurst parameter H, and to check whether it contains 0.5, which implies short-range dependence, or not, this latter case implying long-range dependence. Two points have to be stressed. First of all, in References [11,12] the (frequentist) analysis involves Ethernet traffic packets; no analysis is made for ATM traffic, neither frequentist nor Bayesian. In our knowledge this is the first paper where a Bayesian analysis for ATM traffic data is carried out. As already mentioned in Section 2, the value of H does have a great influence on the performance of the system: the greater H, the worse the performance in terms of loss probability.

The problem of identifying the traffic data characteristics can be formally written down as an hypothesis problem

$$H_0 : H = \tfrac{1}{2}, \quad \text{vs} \quad H_1 : H > \tfrac{1}{2}$$

On the basis of results in previous section, a Bayesian test can be easily performed. In principle, a Bayesian test is based on the probability ratio

$$\frac{P(H = \frac{1}{2} \,|\, \boldsymbol{x})}{P(H \in (\frac{1}{2}, 1) | \boldsymbol{x})} \tag{14}$$

It is apparent from the exposition in Section 3 that the probabilities in (14) cannot be computed analytically. Using the same notation as in Section 3, the numerator of (14) is approximated by

$$\frac{\varepsilon |\Omega_{1/2}|^{-1/2}}{\hat{k}^*(\boldsymbol{x}) \zeta_{1/2}^{1/2} M_{1/2}^{v+n/2}}$$

Hence ratio (14) is approximated by

$$\frac{\varepsilon |\Omega_{1/2}|^{-1/2}}{\hat{k}^*(\boldsymbol{x}) \zeta_{1/2}^{1/2} M_{1/2}^{v+n/2} - \varepsilon |\Omega_{1/2}|^{-1/2}}$$

4. APPLICATION TO REAL DATA

We consider data from the experimental European ATM network [21], a project jointly developed by the leading telecommunications company in the European Union. Engineering aspects of the measurement problem are thoroughly described in Reference [22].

Data streams are produced by superimposing the traffic generated by three different kinds of applications: videoconference, teleteaching and transportation of routing information between IP network routers. All these applications use IP packets over ATM, so that the overflow probability must be estimated on the basis of the available data. The measurement process produces the number of cells arriving in a time slot (1/80 000 s) at an ATM multiplexer, composed by an ATM link and a buffer to store cells not immediately transmitted by the link.

The data we have used come from measurements taken for three different kind of applications: videoconference, teleteaching, and transportation of routing information. In order to study the effects of simultaneous transmissions from several different sources, as described in Section 2 (see (1)), the stream corresponding to every application has been split into substreams (1 s length), and only the data from the first quarter in each substream have been considered. Data from different quarters can be considered approximately independent, each of them coming from a different 'virtual' source. They can be superimposed as if they were coming from sources simultaneously connected to the same ATM multiplexer. We consider three different traffic scenarios: light, medium and heavy traffic. In the first case (light traffic) we have superimposed 10 'virtual sources' for each kind of applications. We have obtained a medium traffic situation by considering 30 teleteaching sources and 20 sources each for both videoconference and transportation of routing information. Finally, the heavy traffic scenario is obtained by superimposing 30 teleteaching sources and 34 sources each from both videoconference and transportation of routing information.

Data from the three scenarios are depicted in Figures 1(a)–(c). Arrivals follow a typical pattern in telecommunications, already observed in non-ATM cases; see References [11,12]. In fact, the patterns are far from being generated by a process with independent increments (i.e. with $H=0.5$). On the opposite, Figures 1(a)–(c) shows the presence of self-similarity in the arrival processes, i.e. long-range dependence in the corresponding increments ($H > 0.5$). As already mentioned in Section 2, FBM is a natural tool for modelling purposes.

Formulas in Section 3 require dealing with matrices whose dimension is given by the data stream length, in our case more than 20 000. Computational burdens have lead us to reduce the size by grouping the data considering the number of cells arriving in 300 consecutive time slots. Therefore, matrices from grouped data have been inverted by FORTRAN routines and their determinants have been computed, as well.

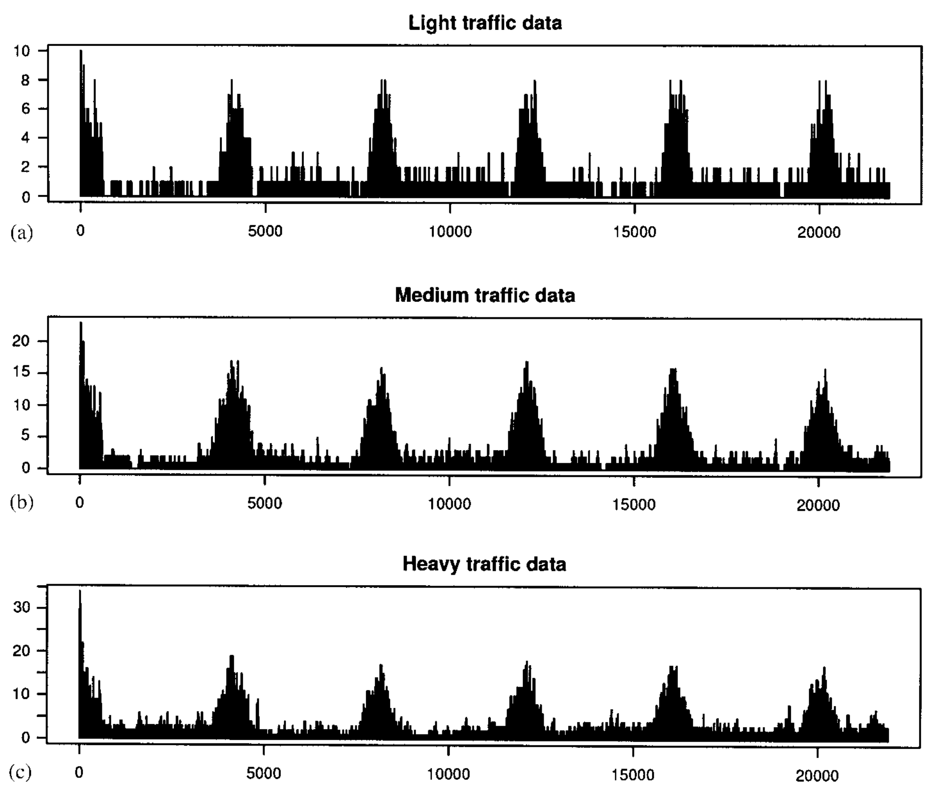


Figure 1. Number of cell arrivals under (a) light, (b) medium and (c) heavy traffic.

In relationship (3) we have taken $u = 1000$, which is a rather high buffer size, whereas the choice $c = 300$ follows from the link speed. Moreover, in the prior distribution of μ we set $w = 0.01$, in order to reduce the probability of sampling negative values for μ.

Our main goals are the detection of long-range dependence and the estimation of the overflow probability $Q(u)$. The former goal has been achieved by both estimating the parameter H and computing the posterior probability ratio for the problem $H = \frac{1}{2}$ vs $H > \frac{1}{2}$. The main finding is the detection of long-range dependence even when considering priors heavily concentrated around $H = \frac{1}{2}$. The values of ratio (14), for different values of prior hyperparameters, are shown in Tables I and II as well as the Bayes estimates of $Q(u)$. Table III shows the same quantities when the r.h.s. of inequality (12) is considered. It is worth mentioning that plots of estimated $Q(u)$'s vs number of iterations show a quick convergence.

It appears that even when the prior probability of $H = 0.5$ is very high ($\varepsilon = 0.999$), ratio (14) takes small values. By the way, the smaller ε, the smaller (14). Hence, our first conclusion is that the data show the presence of strongly correlated increments in the input processes corresponding to the three scenarios considered. Bayes' estimates of H (Table IV) are generally far from 0.5.

The value $\varepsilon = 1$ has been considered in order to study the effect of neglecting long-range dependence. Such an effect is particularly relevant in case of 'heavy' traffic (Tables I and II), which is the most important for applications. As expected, a value $\varepsilon = 1$ could produce a severe underestimation of the overflow probability. The use of the bound in (13) is less expensive from a computational point of view, but produces less accurate results. Compare Tables I and II with Table III. As far as the sensitivity of the Bayes' estimates of the overflow probabilities (with

Table I. Upper bound and probability ratio with $v = 500$, $\lambda = 10$, $\mu = 1000$.

Traffic	ε	Upper	Ratio
Heavy	1	0.0	$+\infty$
Heavy	0.999	0.267×10^{-9}	0.159×10^{-19}
Medium	1	0.0	$+\infty$
Medium	0.999	0.0	0.222×10^{-21}
Light	1	0.0	$+\infty$
Light	0.999	0.0	0.222×10^{-21}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table II. Upper bound and probability ratio with $v = 300$, $\lambda = 10$, $\mu = 1000$.

Traffic	ε	Upper	Ratio
Heavy	1	0.49×10^{-16}	$+\infty$
Heavy	0.999	0.907×10^{-7}	0.687×10^{-12}
Medium	1	0.0	$+\infty$
Medium	0.999	0.996×10^{-16}	0.198×10^{-12}
Light	1	0.0	$+\infty$
Light	0.999	0.0	0.581×10^{-19}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table III. Upper bound and probability ratio using Mills' ratio, with $\lambda = 10$.

Traffic	ε	v	Upper	Ratio
Heavy	0.999	500	0.556×10^{-7}	0.16×10^{-20}
Heavy	0.999	300	0.498×10^{-4}	0.66×10^{-12}
Heavy	1	300	0.0	$+\infty$
Medium	0.999	300	0.232×10^{-13}	0.193×10^{-12}
Light	0.999	300	0.423×10^{-41}	0.57×10^{-19}

The column labelled 'Upper' features posterior estimates of the upper bound of the overflow probability. The column labelled 'Ratio' provides estimates of the posterior probability ratio for the Bayesian test in Section 4.

Table IV. Posterior estimates of H.

	$v = 300,\ \lambda = 10$			$v = 3,\ \lambda = 10$		
	HT	MT	LT	HT	MT	LT
Monte Carlo	0.665	0.666	0.726	0.867	0.865	0.894
MCMC	0.694	0.688	0.756	0.821	0.818	0.832
MC error	0.00164	0.00165	0.00159	0.0015	0.0015	0.0017

The abbreviation HT stands for 'High Traffic', MT for 'Medium Traffic' and LT for 'Light Traffic'.

respect to the choice of the hyperparameters) is concerned, we may note that it is moderate although not negligible.

As far as posterior estimates of H are concerned, our results are summarized in Table IV, where all three different frameworks of heavy traffic (HT), medium traffic (MT) and light traffic (LT) are considered. Two different procedures have been employed. The first one relies upon the classical Monte Carlo procedure as illustrated in Section 3.1. The second one resorts to a Metropolis–Hastings sampling scheme for drawing from the posterior $\pi(\mathrm{d}H|\boldsymbol{x})$ and the resulting sample is used for estimating H. The estimates have been obtained after 12 000 runs and with a burn-in of 10 000 iterations. Diagnostic tests performed with the BOA package (see Reference [23]) have provided strong evidence of convergence of the estimation procedure. The reason for considering an MCMC scheme in this setting is two-fold. On one hand, it is desirable to sketch some comparison, both in terms of computational time and in terms of numerical outcomes, with the classical Monte Carlo procedure. From a computational point of view, the Monte Carlo method is much faster, since it requires on average 15 min with a HP machine (processor PA8000, 180 MHz) to be completed, whereas the MCMC sampler has been running for 95 min. From a numerical point of view, the estimates are not significantly different. On the other hand, having performed an MCMC algorithm one can use the MCMC output in order to get a kernel density estimate of the posterior distribution of H. This can give some insight on the dispersion of H around its posterior mean. In Figure 2, we provide graphs both of the histogram of the MCMC sample and the kernel density estimate of the posterior distribution of H.

The results obtained are fairly insensitive to different choices of the parameters of the prior for μ. The situation is different as far as the prior of σ^2 is concerned. From Table IV, it is argued that different values of v could have a rather strong influence on the estimation of H. However, as it appears from Tables I and II, such a negative effect is mitigated when one considers the

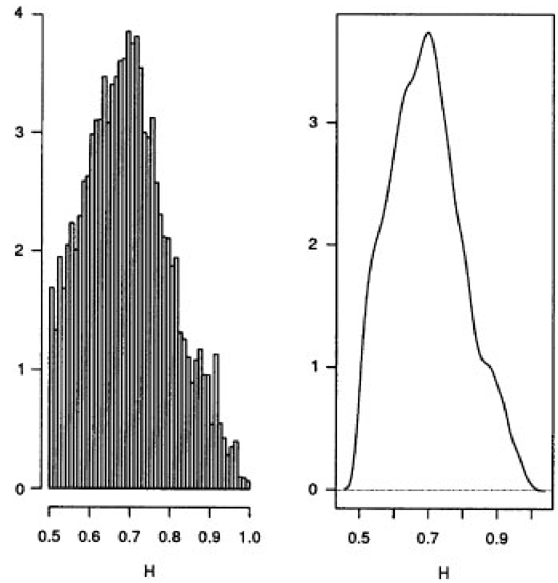


Figure 2. Histogram (on the left) and kernel density estimate of H (on the right) obtained with $v=10$, $\lambda=300$ and $\varepsilon=0.999$. The posterior estimate is, in this case, $E(H|\mathbf{x}) = \hat{H} \approx 0.694$. The estimates are obtained basing on the 'heavy traffic' data.

Table V. Sensitivity of posterior estimates of H (with heavy traffic data).

	$\mathrm{Var}(\sigma^2)=0.001$		$\mathrm{Var}(\sigma^2)=1$	
$E(\sigma^2)$	E(H|x)	Acceptance ratio (%)	E(H|x)	Acceptance ratio (%)
0.5	0.6831	50.14	0.8251	58.11
1	0.6828	50.06	0.8239	57.9
10	0.6827	50.04	0.6931	58.29
100	0.6828	50.01	0.6829	57.43

overflow probability (which is the *real* goal of our analysis). In fact, large variations of v produce only moderate variations of the Bayes' estimates of the overflow probability. Here we provide some tables with estimates of H corresponding to different values of $E(\sigma^2)$ and $\mathrm{Var}(\sigma^2)$, i.e. to different choices of λ and v. See Table V. One can notice that the posterior estimates of H are more sensitive to changes in $E(\sigma^2)$ when $\mathrm{Var}(\sigma^2)$ than in the case in which $\mathrm{Var}(\sigma^2)$ is very low.

4.1. Discussion on the choice of the hyperparameters

The Bayes estimates of the overflow probability exhibit some sensitivity w.r.t. the hyperparameters w, λ, v. For this reason, it is of interest to discuss their choice. In telecommunications, prior information is frequently available in the form of *trial samples*, i.e. small samples of traffic measurements.

Assume that the available prior data consist of $k \geqslant 1$ independent samples. Each sample is obtained by observing the system under consideration for a (short) period of time. The whole observation period is split into m time intervals of length $\Delta > 0$. Let $Y_{i,j}$ be the amount of traffic entering the system in the ith time interval of the jth sample ($i = 1, \ldots, m$, $j = 1, \ldots, k$). From our previous assumptions, conditionally on μ, σ, H, the random vectors ($Y_{i,j}; i = 1, \ldots, m$) have independent multinormal distributions, with $E[Y_{i,j}|\mu,\sigma,H] = \Delta\mu$, $\mathrm{Var}[Y_{i,j}|\mu,\sigma,H] = \Delta^2\sigma^2$, $\mathrm{Cov}[Y_{i,j},\ Y_{i+l,j}\ |\mu,\sigma,H] = \Delta^2\rho(l)$.

As a prior for H, it is reasonable to assume a mixture of a Dirac $\delta_{1/2}$ and a uniform distribution. As seen in the example, the posterior of H is robust w.r.t. the weight ε. As far as the choice of μ_c, w, λ, v is concerned, the basic idea consists in matching them with 'empirical quantities' evaluated on the basis of $Y_{i,j}$s. First of all, let $z > 0$, and

$$\tau(z) := E[z^H] = \varepsilon z^{1/2} + 2(1-\varepsilon)\frac{z - z^{1/2}}{\log z}$$

It is immediate to see that the equalities

$$E[\mu] = \mu_c, \quad E[\sigma^2] = \frac{\lambda}{v-1}, \quad E[\sigma^3] = \frac{\lambda^{3/2}}{\Gamma(v)}\Gamma\left(v - \frac{3}{2}\right)$$

$$\mathrm{Var}(\mu) = E[\mathrm{Var}(\mu|\sigma)] + \mathrm{Var}(E[\mu|\sigma]) = E[w\sigma^2] = w\frac{\lambda}{v-1}$$

$$E[\rho(l)] = \tfrac{1}{2}(\tau((l+1)^2) - 2\tau(l^2) + \tau((l-1)^2))$$

hold true, provided that $v > \frac{3}{2}$. Consider now the hth sample moments

$$\bar{Y}_{h,j} = m^{-1}\sum_{i=1}^{m} Y_{i,j}^h, \quad \bar{Y}_{h..} = k^{-1}\sum_{j=1}^{k} \bar{Y}_{h,j}, \quad h = 1, 2, 3$$

Their expected values are equal to

$$a_1 = E[\bar{Y}_{1,j}] = \Delta\mu_c$$

$$a_2 = E[\bar{Y}_{2,j}] = E|Y_{i,j}^2] = \Delta^2\left(\frac{\lambda}{v-1}(1+w) + \mu_c^2\right)$$

$$a_3 = E[\bar{Y}_{3,j}] = E|Y_{i,j}^3] = \Delta^2\left(3\frac{\lambda}{v-1}(1+w)\mu_c + \mu_c^3\right)$$

$$\begin{aligned} a_4 &= \mathrm{Var}(Y_{1,j}) = E[\mathrm{Var}(\bar{Y}_{1,j}|\mu,\sigma,H)] + \mathrm{Var}(\mu) \\ &= \Delta^2\left((m^{-1} + w)E|\sigma^2] + \frac{2}{m^2}\sum_{l=1}^{m-1}(m-l)E[\sigma^2\rho(l)]\right) \\ &= \Delta^2\frac{\lambda}{v-1}\left\{m^{-1} + w + \frac{1}{m^2}\sum_{l=1}^{m-1}(m-l)(\tau((l+1)^2) - 2\tau(l^2) + \tau((l-1)^2))\right\} \end{aligned}$$

respectively. A possible (and simple, as well) criterion to set the prior hyperparameters consists in choosing μ_c, w, λ, v in such a way that the relationships

$$a_1 = \bar{Y}_{1..}, \quad a_2 = \bar{Y}_{2..}, \quad a_3 = \bar{Y}_{3..}, \quad a_4 = \frac{1}{k-1}\sum_{j=1}^{k}(\bar{Y}_{1\cdot j} - \bar{Y}_{1..})^2$$

hold true.

5. CONCLUSIONS

In this paper, we have proposed a new, Bayesian approach to estimate the overflow probability in ATM networks, when FBM is used to model the traffic. We have analysed short- and long-range dependence using Telecom Italia data and we have discussed important issues in Bayesian analysis, like the choice of the hyperparameters and the sensitivity of the inference with respect to changes in their values. The choice of the prior distributions has been motivated by their flexibility and relative ease in their use. More general classes could have been used, but at the cost of making the computational algorithm even more cumbersome. The choice of the FBM is justified by its (relatively) tractable mathematical structure, although the request of Gaussianity of traffic data could be attenuated. For this purpose, other self-similar processes could be considered, but their use would be a very challenging task.

APPENDIX A

A description of the main computational issues associated with the estimation procedure set forth in Section 3 will be now provided.

A.1. Estimation of H

In order to provide posterior estimates of the Hurst parameter H, a simple Monte Carlo procedure is adopted. Such a choice is suggested by the expression appearing on the right-hand side of (10). A sample of N i.i.d. observations $H_1, \ldots, H_N$ from $\pi_0{}^*(H)$ can be generated. Such a sample is used to approximate the normalizing constant $k^*(\boldsymbol{x})$ by means of the empirical mean

$$\hat{k}^*(\boldsymbol{x}) = \frac{\varepsilon|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2} M_{1/2}^{v+n/2}} + \frac{1-\varepsilon}{N}\sum_{i=1}^{N}\frac{|\Omega_{H_i}|^{-1/2}}{\zeta_{H_i}^{1/2} M_{H_i}^{v+n/2}}$$

so that $E(H|x)$ can be approximated by a ratio of empirical means, namely

$$E(H|\boldsymbol{x}) \approx (\hat{k}^*(\boldsymbol{x}))^{-1}\left\{\frac{\varepsilon}{2}\frac{|\Omega_{1/2}|^{-1/2}}{\zeta_{1/2}^{1/2} M_{1/2}^{v+n/2}} + \frac{(1-\varepsilon)}{N}\sum_{i=1}^{N}\frac{H_i|\Omega_{H_i}|^{-1/2}}{\zeta_{H_i}^{1/2} M_{H_i}^{v+n/2}}\right\}$$

A.2. Estimation of loss probability

An algorithm providing the desired approximations of the two integrals appearing in the right-hand side of (11) works as follows. If, for any H in $[\frac{1}{2}, 1)$,

$$\pi(\sigma^2|H,\boldsymbol{x}) = \frac{M_H^{v+n/2}}{\Gamma(v+n/2)}\left(\frac{1}{\sigma^2}\right)^{v+n/2+1}\exp\left\{-\frac{M_H}{\sigma^2}\right\}$$

and

$$\pi(\mu|H,\sigma^2,\boldsymbol{x}) = \frac{\zeta_H^{1/2}}{\sigma\sqrt{\pi}}\exp\left\{-\frac{\zeta_H}{2\sigma^2}\left(\mu-\frac{\xi_H}{2\zeta_H}\right)^2\right\}$$

one has $\pi^*(\frac{1}{2},\mu,\sigma^2|\boldsymbol{x}) = \pi(\mu|\frac{1}{2},\sigma^2,\boldsymbol{x})\pi(\sigma^2|\frac{1}{2},\boldsymbol{x})$. It follows that the first integral in the right-hand side of (11), corresponding to the case in which $H=\frac{1}{2}$, can be easily handled by generating an i.i.d. sample (σ_i^2,μ_i), $i=1,\ldots,N$, with σ_i^2 and μ_i drawn from $\pi(\sigma^2|\frac{1}{2},\boldsymbol{x})$ and from $\pi(\mu|\frac{1}{2},\sigma_i^2,\boldsymbol{x})$, respectively. Hence

$$\int_{\mathbb{R}\times\mathbb{R}^+}\bar{q}_u\left(\tfrac{1}{2},\mu,\sigma^2\right)\pi^*\left(\tfrac{1}{2},\mu,\sigma^2|\boldsymbol{x}\right)\boldsymbol{I}_{(\mu<c)}\,\mathrm{d}\mu\,\mathrm{d}\sigma^2 \approx \frac{1}{N}\sum_{i=1}^{N}\bar{q}_u\left(\tfrac{1}{2},\mu_i,\sigma_i^2\right)\boldsymbol{I}_{(\mu_i<c)}$$

As far as the second integral in the right-hand side of (11) is concerned, note that $\pi**(dH,\mu,\sigma^2|\boldsymbol{x}) = \pi(H|\boldsymbol{x})\pi(\sigma^2|H,\boldsymbol{x})\pi(\mu|H,\sigma^2,\boldsymbol{x})\boldsymbol{I}_{1/2<H<1}$, where

$$\pi(H|\boldsymbol{x}) \propto \frac{|\Omega_H|^{-1/2}}{\zeta_H^{1/2}M_H^{\nu+n/2}}\pi_0*(H)$$

A Metropolis–Hastings algorithm applies in this case, since it is not possible to sample directly from $\pi(H|x)$. The proposal we employ is

$$p(H_{i+1,}\sigma_{i+1}^2,\mu_{i+1}|H_i,\sigma_i^2,\mu_i) = \eta(H_{i+1})\pi(\sigma_{i+1}^2|H_{i+1},\boldsymbol{x})\pi(\mu_{i-1}|\sigma_{i+1}^2,H_{i+1},\boldsymbol{x})$$

$\eta(H) \propto (1-2H)^{\gamma-1}(1-H)^{\delta-1}$ being a probability density function on $(\frac{1}{2},1)$. Therefore, the adopted scheme corresponds to an independence sampler with acceptance ratio given by

$$\alpha((H_i,\mu_i,\sigma_i),(H_{i+1},\mu_{i+1},\sigma_{i+1})) = \min\left\{1,\frac{\eta(H_{i+1})\pi(H_i|x)}{\eta(H_i)\pi(H_{i+1}|\boldsymbol{x})}\right\}$$

The same MCMC output is used to estimate $k_c(\boldsymbol{x})$ and, then, $k^*(\boldsymbol{x})$.

Let us now move on to the problem of determining the upper bound in (13), which follows from inequality (12) on Mills' ratio for the Gaussian distribution. We still consider separately the cases $H=\frac{1}{2}$ and $H\in(\frac{1}{2},1)$. When $H=\frac{1}{2}$, a simple Monte Carlo integration can be done, by sampling i.i.d. μ_i's from $\pi(\mu|\frac{1}{2},\boldsymbol{x})$. If $H\in(\frac{1}{2},1)$, an independence sampler is employed again. In fact, we set

$$p(H_{i+1,}\mu_{i+1}|H_i,\mu_i) := \eta(H_{i+1})\pi(\mu_{i+1}|H_{i+1},\boldsymbol{x})$$

as the proposal distribution, with $\eta(H)$ as above. Acceptance ratio is

$$\alpha((H_i,\mu_i),(H_{i+1},\mu_{i+1})) = \min\left\{1,\frac{\eta(H_{i+1})\pi(H_i|\boldsymbol{x})}{\eta(H_i)\pi(H_{i+1}|\boldsymbol{x})}\right\}$$

ACKNOWLEDGEMENTS

Thanks are due to two anonymous referees, who carefully read this paper, and made very interesting and stimulating comments. The authors are also grateful to Giovanni Petris, Antonio Pievatolo and Mike Wiper for their valuable comments. The research was performed when the second author was affiliated with CNR-IMATI.

REFERENCES

1. Giroux N, Ganti S. *Quality of Service in ATM Networks*. Prentice-Hall: London, 1999.
2. De Pricker M. *Asynchronous Transfer Mode*: *Solution for Broadband ISDN*. Ellis Horwood: New York, 1993.
3. IETF RFC 2225. Classic IP and ARP over ATM. *Internet Engineering Task Force*, 1998.
4. IETF RFC 3035. MLPS using LPD and ATM VC switching. *Internet Engineering Task Force*, 2001.
5. Roberts R, Mocci U, Virtamo J (eds). *Broadband Network Teletraffic*. Lectures Notes in Computer Science, vol. 1155. Springer: Berlin, 1996.
6. ITU-T Recommendation I.356. ATM layer cell transfer performance. *International Telecommunications Union—Telecommunication*, 1996.
7. Bruneel H, Kim BG. *Discrete-Time Models for Communication Systems including ATM*. Kluwer: Boston, 1993.
8. Beneš VE. *General Stochastic Processes in the Theory of Queues*. Addison-Wesley: Reading, MA, 1963.
9. Norros I. A storage model with self-similar input. *Queueing Systems Theory and Applications* 1994; **16**:387–396.
10. Norros I. On the use of fractional Brownian motion in the theory of connectionless networks. *IEEE Journal on Selected Areas in Communications* 1995; **13**:953–962.
11. Willinger W, Taqqu MS, Leland WE, Wilson DV. Self-similarity in high-speed packet traffic: analysis and modelling of Ethernet traffic measurements. *Statistical Science* 1995; **10**:67–85.
12. Willinger W, Taqqu MS, Sherman R, Wilson DV. Self-similarity through high-variability: statistical analysis of Ethernet *LAN* traffic at the source level. In *Proceedings of ACM SIGCOMM 95*. Cambridge, ACM: 1995; 100–113.
13. Willinger W, Taqqu MS, Erramilli A. A bibliographic guide to self-similar traffic and performance modeling for modern high-speed networks. In *Stochastic Network*, Kelly FP, Zachery S, Ziedins I (eds). Oxford University Press: Oxford, 1997.
14. Beran J. *Statistics for Long-Memory Processes*. Chapman & Hall: London, 1994.
15. Resnick SI. Heavy tail modeling, teletraffic data. *The Annals of Statistics* 1997; **25**:1805–1896.
16. Tsybakov B, Georganas ND. On self-similar traffic in ATM queues: definitions, overflow probability bound and cell delay distribution. *IEEE Transactions on Networking* 1997; **5**:397–409.
17. Hüsler J, Piterbarg V. Extremes of a certain class of Gaussian processes. *Stochastic Processes and their Applications* 1999; **83**: 257–271.
18. Shao QM. Bounds and estimators of a basic constant in extreme value theory of Gaussian processes. *Statistica Sinica* 1996; **6**: 245–257.
19. Hüsler J. Extremes of a Gaussian process and the constant H_α. *Extremes* 1999; **1**:59–70.
20. Chow YS, Teicher H. *Probability Theory*. Springer: New York, 1995.
21. Parker, M. The European ATM Pilot. In *Proceedings of XV International Switching Symposium ISS'95*, Berlin, 1995.
22. Gnetti AS. Characterizations and measurements of teletraffic generated by IP protocol on ATM network. *M.S. Thesis*, Università di Roma 'La Sapienza' Facoltà di Ingegneria (in Italian).
23. Smith B. *Bayesian Output Analysis program: version 1.0.0*. Department of Biostatistics, University of Iowa, 2001, Available at http://www.public-health.uiowa/edu/boa.

