Minicurso 1
Bayesian inference for Poisson processes and applications in reliability

Fabrizio Ruggeri
CNR IMATI, Italia

Organização:
Associação Brasileira de Estatística
Bayesian inference for Poisson processes and applications in reliability

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
web.mi.imati.cnr.it/fabrizio/

Institute for Future Environments
Queensland University of Technology
Brisbane, Australia
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 $M/\text{tt}/c$ queueing systems

• Simple mathematical formulation and relatively straightforward statistical analysis ⇒ very practical, although 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. $P(N(t, t + \Delta t] = 1) = \lambda(t)\Delta t + o(\Delta t)$, as $\Delta t \to 0$
  4. $P(N(t, t + \Delta t] \geq 2) = o(\Delta t)$, as $\Delta t \to 0$

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

$\Rightarrow N(s, t] \sim \text{Po} \int_{s}^{t} \lambda(x) \, dx$
**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 $\lambda$
  - $N(s, t] \sim \text{Po}(\lambda(t - s))$
  - 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)

- $P(N(t, t + \Delta t] \geq 2) = o(\Delta t)$, 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 $\lambda$
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

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

- $n$ 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 gas escapes according to the subnetwork characteristics

• **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
  – $N(t)$ 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 $Y$
    ⇒ $n$ 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 time dependent probabilities $p(t)$, defined on $(0, \infty)$
  – As an example, for an HPP with rate $\lambda$, if events at any time $t$ are kept with probability $p(t)$ ⇒ Poisson process with intensity function $\lambda p(t)$
QUESTION: Which prior for $\lambda$ of HPP?

$$P(T_1, \ldots, T_n) = \lambda^n e^{-\lambda T}$$
HPP: ESTIMATION

- Gamma priors conjugate w.r.t. $\lambda$ in the HPP
- Prior $\text{Ga}(\alpha, \beta)$
- $\Rightarrow f(\lambda| n, T) \propto \lambda^n e^{-\lambda T} \cdot \lambda^{\alpha-1} e^{-\beta \lambda}$
- $\Rightarrow$ posterior $\text{Ga}(\alpha + n, \beta + T)$
- Posterior mean $\hat{\lambda} = \frac{\alpha + n}{\beta + T}$
- Posterior mean combination of
  - Prior mean $\hat{\lambda}_P = \frac{\alpha}{\beta}$
  - MLE $\hat{\lambda}_M = \frac{n}{T}$
HPP: PRIOR AND DATA INFLUENCE

- Posterior mean: $\hat{\lambda} = \frac{\alpha + n}{\beta + T}$

- Prior mean: $\hat{\lambda}_P = \frac{\alpha}{\beta}$ (and variance $\sigma^2 = \frac{\alpha}{\beta^2}$)

- MLE: $\hat{\lambda}_M = \frac{n}{T}$

- $\alpha_1 = k\alpha$ and $\beta_1 = k\beta \Rightarrow \hat{\lambda}_{1P} = \hat{\lambda}_P$ and $\sigma^2_1 = \sigma^2/k$

- Posterior mean: $\hat{\lambda}_{1P} = \frac{k\alpha + n}{k\beta + T}$

- $k \to 0 \Rightarrow$ prior variance $\to \infty \Rightarrow \hat{\lambda} \to \frac{n}{T}$, i.e. MLE (prior does not count)

- $k \to \infty \Rightarrow$ prior variance $\to 0 \Rightarrow \hat{\lambda} \to \hat{\lambda}_P$, i.e. prior mean (data do not count)

- $n \to \infty \Rightarrow \hat{\lambda} \sim \frac{n}{T}$, i.e. MLE (prior does not count)
HPP - PRIOR CHOICE

QUESTION(S):

How to choose $\alpha$ and $\beta$ in $\lambda \sim \text{Ga}(\alpha, \beta)$

- Switch to any interarrival time $X_i$, $i = 1, n$
  - $X_i$’s are i.i.d. (you can prove it!)
  - I have the next slides ready for the exponential case

- $P(X_1 > t) = P(N(t) = 0) = \exp\{-\lambda t\}$

- $X_i \sim \text{Ex}(\lambda)$

What do we know about $\lambda$?
HPP - PRIOR CHOICE

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

⇒ Physical properties of $\lambda$
- $E X = 1/\lambda$
- $Var X = 1/\lambda^2$
- $h(x) = \frac{f(x)}{S(x)} = \frac{\lambda \exp\{-\lambda x\}}{\exp\{-\lambda x\}} = \lambda$ (hazard function)
- $E N(s, t) = \lambda (t - s)$
- $P(N(t) = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$
HPP: PRIOR CHOICE

QUESTION:
Which information is available on $\lambda$ ?
HPP - PRIOR CHOICE

• Exact prior $\pi(\lambda)$ (???)

• Quantiles of $X_i$, i.e. $P(X_i \leq x_q) = q$

• Quantiles of $\lambda$, i.e. $P(\lambda \leq \lambda_q) = q$

• Moments $E\lambda^k$ of $\lambda$, i.e. $\lambda^k \pi(\lambda) d\lambda = a_k \iff (\lambda^k - a_k) \pi(\lambda) d\lambda = 0$

• Generalised moments of $\lambda$, i.e. $h(\lambda) \pi(\lambda) d\lambda = 0$

• Most likely value and upper and lower bounds

• ...

• None of them
HPP: PRIOR CHOICE

QUESTION: How to get information on λ?
HPP - PRIOR CHOICE

- Results from previous experiments (e.g. 75% of car tyres had failed after 5 years of operation ⇒ 5 years is the 75% quantile of $X_i$)

- Split of possible values of $\lambda$ or $X_i$ into equally likely intervals ⇒ median and quartiles

- Two quantiles, with a third given for consistency

- Most likely value and upper and lower bounds

- Expected value of $\lambda$ and confidence on such value (mean and variance)

- Ideal experiment with $\alpha = \text{sample size}$ and $\beta = \text{sum of observations}$
  (from $\lambda \sim G(\alpha, \beta) \Rightarrow \lambda | X \sim G(\alpha + n, \beta + \sum_{i=1}^{n} X_i)$)

- ...
HPP: FORECASTING

- $n$ events observed in the interval $(0, T]$

- Interest in forecasting number of events in $(T, T+s]$: $P(N(T, T+s) = m | n, T)$

QUESTION: How to forecast?
HPP: FORECASTING

- $n$ events observed in the interval $(0, T]$
- Posterior $\text{Ga}(\alpha + n, \beta + T)$
- Interest in forecasting number of events in $(T, T + s]$: $P(N(T, T + s] = m| n, T)$
- For $s > 0$ and integer $m$

$$P(N(T, T + s] = m| n, T) = \int_{0}^{\infty} P(N(T, T + s] = m| \lambda) f(\lambda| n, T) d\lambda$$

$$= \int_{0}^{\infty} (\lambda s)^m e^{-\lambda s} f(\lambda| n, T) d\lambda$$

$$= \int_{0}^{\infty} \frac{m!}{m!} (\lambda s)^m e^{-\lambda s} (\beta + T)^{\alpha+n} \Gamma(\alpha+n) \lambda^{\alpha+n-1} e^{-\lambda(\beta+T)}$$

$$= \frac{s^m (\beta + T)^{\alpha+n} \Gamma(\alpha+n + m)}{m!(\beta + T + s)^{\alpha+n+m} \Gamma(\alpha+n)}$$

$\Rightarrow P(N(T, T + s] = m| n, T)$
HPP: FORECASTING

• $n$ events observed in the interval $(0, T]$

• Posterior $\text{Ga}(\alpha + n, \beta + T)$

• Interest in forecasting expected number of events in $(T, T + s]$

\[
E[N(T, T + s)|n, T)] = \int_0^\infty E[N(T, T + s)|\lambda] f(\lambda|n, T) d\lambda \\
= \int_0^\infty \lambda s f(\lambda|n, T) d\lambda \\
= \frac{\alpha + n}{\beta + T}
\]

• $\Rightarrow E[N(T, T + s)|n, T)] = \frac{\alpha + n}{\beta + T}$
GAS ESCAPES IN A CITY NETWORK

• Setting up an efficient replacement policy in a large metropolitan gas distribution network developed in the last century

• Assessment of failure rate of the pipelines, different for materials (cast iron, steel, polyethylene, etc.) and conditions (diameter, laying depth, etc.)

• Change of pipelines with highest failure rate

• (Traditional) cast iron pipelines with higher failure rate than other materials
  – not subject to corrosion (aging)
  – propensity-to-failure in a unit time period or unit length does not vary significantly with time and space
  – rare events occurring *randomly* and not affecting the next ones
  – \( \Rightarrow \) HPP

• EDA identified diameter, laying depth and location as the most significant factors
FAILURES IN CAST-IRON PIPES

- HPP with parameter $\lambda$ (unit failure rate in time and space)

- $n$ failures in $[0, T] \times S$, $\Rightarrow L(\lambda|n, T, S) = (\lambda sT)^n e^{-\lambda sT}$, with $s = meas(S)$

- Data: $n = 150$ failures in $T = 6$ years on a net $\approx s = 312$ Km long

- $\Rightarrow L(\lambda|n, T, S) = (1872\lambda)^{150} e^{-1872\lambda}$

- MLE $\hat{\lambda} = n/(sT) = 150/1872 = 0.080$

- $\lambda \sim G(\alpha, \beta) \Rightarrow \lambda|n, T, S \sim G(\alpha + n, \beta + sT)$

- Consider 8 classes determined by two levels of relevant covariates: diameter, location and depth
FAILURES IN CAST-IRON PIPE

Overall 0.08003

Street 0.1497

Sidewalk 0.0698

D < 125 mm 0.0755

Depth < 0.9 m 0.0717

Depth > 0.9 m 0.0941

D > 125 mm 0.0656

Depth < 0.9 m 0.0666

Depth > 0.9 m 0.0600

Depth < 0.9 m 0.1773

D < 125 mm 0.1668

Depth > 0.9 m 0.1152

D > 125 mm 0.1371

Depth < 0.9 m 0.1315

Depth > 0.9 m 0.1777
ELICITATION OF EXPERTS’ OPINIONS

- A questionnaire was given to 26 experts from different areas within the company

- Interviewees were unable to say how many failures they expected to see on a kilometer of a given kind of pipe in a year (even upper and lower bounds on them!)

- 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

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>“equally likely as”</td>
<td>A → 1</td>
<td>B</td>
</tr>
<tr>
<td>“a little more likely than”</td>
<td>A → 3</td>
<td></td>
</tr>
<tr>
<td>“much more likely than”</td>
<td>A → 5</td>
<td></td>
</tr>
<tr>
<td>“clearly more likely than”</td>
<td>A → 7</td>
<td></td>
</tr>
<tr>
<td>“definitely more likely than”</td>
<td>A → 9</td>
<td></td>
</tr>
</tbody>
</table>

• Pairwise comparison for alternatives $A_1, \ldots, A_n$

  • $\Rightarrow$ square matrix of size $n$
  
  • $\Rightarrow$ (normalized) eigenvector associated with the largest eigenvalue

  • $\Rightarrow (P(A_1), \ldots, P(A_n))$

• **Question**: if a gas escape occurs, where do you think it will occur if you have to choose between subnetwork A and subnetwork B?
**ANALYTIC HIERARCHY PROCESS**

*An expert’s opinion on propensity to failure of cast-iron pipes*

<table>
<thead>
<tr>
<th>Class</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>1/6</td>
<td>1</td>
<td>1/6</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1/3</td>
<td>1</td>
<td>1/4</td>
<td>2</td>
<td>1/6</td>
<td>1/2</td>
<td>1/5</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1/3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/6</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1/3</td>
<td>1/2</td>
<td>1</td>
<td>1</td>
<td>1/5</td>
<td>1</td>
<td>1/5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
<td>1/6</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>5</td>
<td>1/4</td>
<td>6</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>1/3</td>
<td>1</td>
<td>1/2</td>
<td>1</td>
<td>1/5</td>
<td>1</td>
<td>1/4</td>
<td>1</td>
</tr>
</tbody>
</table>
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)
⇒ find the “most risky” class

- Failures in the network occur at rate $\lambda$ and allocated to class $A_i$ with probability $P(A_i)$ ⇒ failures in class $A_i$ occur at rate $\lambda_i = \lambda P(A_i)$ (*Coloring Theorem*)
- $P(A_i)$ given by AHP for any expert

- Choice of $\lambda$ ⇒ 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 expected number of gas escapes for given period and length of network ⇒ statements on $\lambda$, unit failure rate for entire network, and get a gamma prior on it
  - What we did
    * Estimate $\lambda$ by MLE $\hat{\lambda}$ with a unique HPP for the network
    * Use experts' opinions through AHP to get a prior on $\lambda_i = \hat{\lambda} P(A_i)$
MODELS FOR CAST-IRON PIPES

• Choice of priors
  – Gamma vs. Lognormal [informal sensitivity]
  – For each expert, eigenvector from AHP multiplied by $\hat{\lambda} \Rightarrow \text{sample}$ about $(\lambda_1, \ldots, \lambda_8)$
  – Mean and variance of priors on $\lambda_i$’s estimated from the sample of size 26 (number of experts) using the method of moments

• Posterior mean of failure rate $\lambda_i$ for each class

• Classes ranked according to posterior means (largest $\Rightarrow$ most keen to gas escapes)

• Sensitivity
  – Classes of Gamma priors on $\lambda$ with mean and/or variance in intervals
  – Quantile class on $\lambda$

• Non-dominated (Bayes) actions under classes of priors/losses
ESTIMATES’ COMPARISON

• Location: W (under walkway) or T (under traffic)
• Diameter: S (small, < 125 mm) or L (large, ≥ 125 mm)
• Depth: N (not deep, < 0.9 m) or D (deep, ≥ 0.9 m)

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
• LN and G ⇒ similar answers
NON-DOMINATED ACTIONS

Range of prior opinions (14 experts)
NON-DOMINATED ACTIONS

Range of non-dominated actions for quantile class (given by empirical quantiles of experts' opinions): 3 quantiles (left) and 7 (right)
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
    $\lambda(t) = at + \beta$ (linear ROCOF model)
    $\lambda(t) = at^2 + \beta t + \gamma$ (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)
    $\lambda(t) = \alpha \beta(at)^{\beta - 1} \exp\{\alpha t^\beta\}$ (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 - PLP)
    ∗ Different behavior for different $\beta$s
NONHOMOGENEOUS POISSON PROCESS

• Different mathematical properties
  – Periodicity (Lewis)
    \* \( \lambda(t) = \alpha \exp\{\rho \cos(\omega t + \phi)\} \)
    \* 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
      \[ \alpha \log (1 + \beta t) \]
    \* \( \lambda(t) = \frac{\alpha \log (1 + \beta t)}{1 + \beta t} \)
    \* 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
    ⇒ 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
    ⇒ ratio-logarithmic \( \lambda(t) \) (Pievatolo et al., 2003)
  – New product ⇒ life cycle described by bathtub intensity (e.g. PLP with change points)
  – Finite number of bugs to be detected during software testing
    ⇒ \( m(t) \) finite over an infinite horizon
  – Unlimited number of death in a population
    ⇒ \( m(t) \) infinite over an infinite horizon (as a good approximation)
NONHOMOGENEOUS POISSON PROCESS

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

- Two parameterizations:
  - $\lambda(t; \alpha, \beta) = \frac{\beta}{\alpha} \cdot \frac{t^{-1}}{\alpha}$ and $m(t; \alpha, \beta) = \frac{t^{-\beta}}{\alpha}$, $\alpha, \beta, t > 0$
  - $\lambda(t; M, \beta) = M \beta t^{\beta-1}$ 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

\[
\lambda(t | M=3, \beta)
\]

\begin{align*}
\beta &= 0.5 \\
\beta &= 1.0 \\
\beta &= 1.5 \\
\beta &= 2.0 \\
\beta &= 3.0
\end{align*}
PLP: FREQUENTIST ANALYSIS

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

$$I(\alpha, \beta \mid T) = (\beta/\alpha)^n \prod_{i=1}^{n} \left(\frac{T_i}{\alpha}\right)^{\beta-1} e^{-\left(y/\alpha\right)^\beta}$$

- **Failure truncation** $\Rightarrow y = T_n$
  
  MLE: $\hat{\beta} = \frac{n^{-1}}{\sum_{i=1}^{n-1} \log(\frac{T_n}{T_i})}$ and $\hat{\alpha} = \frac{T_n}{n^{1/\beta}}$

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

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

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

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

Unbiased estimators, $\hat{\lambda}(t)$, approx. C.I., hypothesis testing, goodness-of-fit, etc.
PLP: BAYESIAN ANALYSIS

Failure truncation \equiv\ Time truncation

\[ l(\alpha, \beta \mid \mathcal{I}) = (\beta / \alpha)^n \prod_{i=1}^{n} \left( T_i / \alpha \right)^{\beta - 1} e^{-\left( y / \alpha \right)^{\beta}} \]

- \( \pi(\alpha, \beta) \propto (\alpha \beta y)^{\gamma - 1} \) \quad \alpha > 0, \beta > 0, \gamma = 0, 1 \quad \Rightarrow \beta \mid \mathcal{T} \sim \beta \chi^2_{2(n - \gamma)} / (2n)
  - Posterior exists, except for \( \gamma = 0 \) and \( n = 1 \)
  - \( \hat{\beta} = n^{-1} \sum_{i=1}^{n} \log \left( T_i / T \right) \)
  - Posterior mean \( \bar{\beta} = (n - \gamma) / \sum_{i=1}^{n} \log \left( T_i / T \right) \)
  - Credible intervals easily obtained with standard statistical software

- \( \pi(\alpha \mid \beta) \propto \alpha^{-\frac{a \beta}{a}} \alpha^{-\frac{a \beta}{s}} e^{-\left( s / \alpha \right)^{\beta}} \) \quad a, b, s > 0 and \( \beta \sim \text{U}(\beta_1, \beta_2) \)

\[ \Rightarrow \pi(\beta \mid \mathcal{I}) \propto \beta^n \prod_{i=1}^{n} \left( \frac{T_i}{s} \right)^{\beta - 1} \left( \frac{T_n}{s} \right)^{\beta - \frac{a}{s}} \text{I}_{[\beta_1, \beta_2]}(\beta) \]

- In all case \( \alpha \mid \mathcal{I} \) by simulation (but \( \alpha \mid \beta, \mathcal{I} \) inverse of a Weibull)
BAYESIAN ANALYSIS

Other parametrization

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

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

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

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

\[
M\mid T_1, \ldots, T_n\beta \sim \text{Ga}(\alpha + n, \delta^\beta + T_\beta^n)
\]

\[
\beta\mid T_1, \ldots, T_nM \propto \beta^{\mu+n-1} \exp\{\beta(\log T - \nu) - MT^\beta - M\delta^\beta\}
\]

• Sample from posterior applying Metropolis step within Gibbs sampler

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

Failure of a water pump in a car

- Water pump ⇒ non-repairable system
- Car ⇒ 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 ⇒ reliability growth but not “as new”

Different models used in disjoint time intervals
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 (⇒ back to previous reliability)

Repairs could worsen the reliability
RELIABILITY MEASURES

• System reliability (for a PLP)
  – Data on the same system (observed up to $y$):
    $$R((y, s]) = P (N(y, s) = 0 | M, \beta) = e^{-M^\beta s^{\beta} - y^{\beta}}$$
  – Data on equivalent system:
    $$R(s) = P (N(s) = 0 | M, \beta) = e^{-Ms^{\beta}}$$

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

• Intensity function at $y$:
  Reliability growth models without further improvements $\Rightarrow$ constant intensity $\lambda(y)$
GAS ESCAPES IN STEEL PIPES

• Replacements not affecting network reliability ⇒ repairable system

• Steel pipes subject to corrosion (aging) ⇒ NHPP and relevance of installation date

• Network split into subnetworks based upon year of installation, as if pipes were installed on July, 1st each year

• Independent PLP’s $N_i(t)$ for each subnetwork, with $\lambda_i(t) = M_i\beta_i t^{\beta_i-1}$

• Superposition Theorem: Sum of independent NHPPs $N_i(t)$ with intensity functions $\lambda_i(t)$ is still a NHPP $N(t)$ with intensity function $\lambda(t) = \sum \lambda_i(t)$

• Characteristics of pipes installed vs. PLP parameters
  - Equal pipes ⇒ same $M$ and $\beta$ for each $N_i(t)$
  - Completely different pipes ⇒ different, independent $M_i$ and $\beta_i$ for each $N_i(t)$
  - Similar pipes ⇒ $M_i$ and $\beta_i$ for each $N_i(t)$ coming from a common distribution (exchangeability)
GAS ESCAPES IN STEEL PIPES

- Experts asked about interval of first gas escape $X_1$
  - 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 $(M, \beta)$
    \[ P(X_1 \in [T_0, T_1]) = \exp{-lMT^\beta_0} - \exp{-lMT^\beta_1} = 0.95 \]
  - Check for consistency, e.g. $A \subset B \Rightarrow P(A) > P(B)$

- Pooling of experts’ opinions
  - $\Rightarrow$ sample from priors
  - $\Rightarrow$ hyperparameters in priors, matching moments
GAS ESCAPES IN STEEL PIPES

- **Known** length \( l_s \) of network installed in year \( s = 1, \ldots, r \)
- **Known** installation date \( \delta_k \) of \( k \)-th failed pipe

- Likelihood \( L(M, \beta; t, \delta) = \prod_{k=1}^{n} \beta_{\delta_k}^{l_{\delta_k}}(t_{k} - \delta_{k})^{\beta_{\delta_k} - 1} e^{-r_{s=1} l_{s} M_{s}[(T_1 - s)^{\beta_{s}} - (s \lor T_0 - s)^{\beta_{s}}]} \)
- \( M_s \sim E(\theta_M) \perp \beta_s \sim E(\theta_\beta), s = 1, \ldots, r \), but exchangeable among themselves
- \( \theta_M \sim E(\tau_M) \) and \( \theta_\beta \sim E(\tau_\beta) \)
- Posterior \( \pi(M, \beta|t, \delta) \) obtained integrating out \( \theta_M \) and \( \theta_\beta \)

- \( \pi(M, \beta|t, \delta) \propto \prod_{s=1}^{r} (l_{s} M_{s}^{\beta_{s}}) |l_{s}| \prod_{k=1}^{r} (t_{k} - \delta_{k})^{\beta_{\delta_k} - 1} e^{-r_{s=1} l_{s} M_{s}[(T_1 - s)^{\beta_{s}} - (s \lor T_0 - s)^{\beta_{s}}]} \)
MODELS FOR STEEL PIPES

Exchangeable $M$ and $\beta$; known installation dates

95% credible intervals for reliability measures:

- System reliability over 5 years: $P\{N(1998, 2002) = 0\} \Rightarrow [0.0000964, 0.01]$  
- Mean value function (solid) vs. cumulative # failures (points)
A VERY SIMPLE NHPP MODEL

MLE (dashed) vs. Bayes (solid) for $\lambda_\theta(t) = a \ln(1 + bt) + c$
NONPARAMETRIC APPROACH

# events in \([T_0, T_1] \sim P(\Lambda[T_0, T_1])\), 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: \(\Lambda[T_0, T_1] \sim G(\cdot, \cdot) \Rightarrow \Lambda\) d.f. of the random measure \(M\)

Notation: \(\mu B := \mu(B)\)

**Definition 1** Let \(\alpha\) be a finite, \(\sigma\)-additive measure on \((S, S)\). The random measure \(\mu\) follows a **Standard Gamma** distribution with shape \(\alpha\) (denoted by \(\mu \sim GG(\alpha, 1)\)) if, for any family \(\{S_j, j = 1, \ldots, k\}\) of disjoint, measurable subsets of \(S\), the random variables \(\mu S_j\) are independent and such that \(\mu S_j \sim G(\alpha S_j, 1)\), for \(j = 1, \ldots, k\).

**Definition 2** Let \(\beta\) be an \(\alpha\)-integrable function and \(\mu \sim GG(\alpha, 1)\). The random measure \(M = \beta \mu\), s.t. \(\beta \mu(A) = \int_A \beta(x) \mu(dx), \forall A \in S\), follows a **Generalised Gamma** distribution, with shape \(\alpha\) and scale \(\beta\) (denoted by \(M \sim GG(\alpha, \beta)\)).
NONPARAMETRIC APPROACH

Consequences:

- \( \mu \sim P_{\alpha,1}, \) \( P_{\alpha,1} \) unique p.m. on \((\Omega, M)\), space of finite measures on \((S, S)\), with these finite dimensional distributions

- \( M \sim P_{\alpha,\beta}, \) weighted random measure, with \( P_{\alpha,\beta} \) p.m. induced by \( P_{\alpha,1} \)

- \( EM = \beta \alpha, \) i.e. \( \int_{\Omega} M(A)P_{\alpha,\beta}(dM) = \int_{A} \beta(x)\alpha(dx), \forall A \in S \)

**Theorem 1** Let \( \xi = (\xi_1, \ldots, \xi_n) \) be \( n \) Poisson processes with intensity measure \( M \). If \( M \sim GG(\alpha, \beta) \) a priori, then \( M \sim GG(\alpha + \sum_{i=1}^{n} \xi_i, \beta/(1 + n\beta)) \) a posteriori.
NONPARAMETRIC APPROACH

Data: \( \{ y_{ij}, i = 1 \ldots k_j \}_j \) from \( \xi = (\xi_1, \ldots, \xi_n) \)

Bayesian estimator of \( M \): measure \( \tilde{M} \) s.t., \( \forall S \in S \),

\[
\tilde{M}_S = \frac{\beta(x)}{s'} \alpha(dx) + \sum_{j=1}^n \frac{k_j}{1 + n \beta(y_{ij})} I_S(y_{ij})
\]

Constant \( \beta \Rightarrow \tilde{M}_S = \frac{\beta}{1 + n \beta} [\alpha + \sum_{j=1}^n \sum_{i=1}^k I_S(y_{ij})] \)

Bayesian estimator of reliability \( R \), \( RS = P(\xi S = 0), S \in S \):

\[
\tilde{R} S = \exp \left( - \sum_{j=1}^n \frac{\beta(x)}{s'} \ln \left( 1 + \frac{\alpha(x)}{1 + n \beta(x)} \right) \right) - \sum_{j=1}^n \frac{k_j}{1 + n \beta(y_{ij})} I_S(y_{ij})
\]

Constant \( \beta \Rightarrow \tilde{R} S = 1 + \frac{\beta}{1 + n \beta} \right) -(\alpha + \sum_{j=1}^n \xi_S) \)
STEEL PIPES

Parametric NHPP: \( \Lambda_\theta(t) = \int_0^t [\bar{a} \log(1 + \bar{b}t)]dt + \hat{c}t \)

Nonparametric model: \( M \sim P_{\alpha, \beta} : \alpha(ds) := \Lambda_\theta(s)/\sigma ds, \beta(s) := \sigma \)

\( \Rightarrow E\{M\} = \Lambda_\theta S \) and \( \text{Var}\{M\} = \sigma \Lambda_\theta S \)

\( \Rightarrow M \) “centered” at parametric estimator \( \Lambda_\theta S \) and closeness given by \( \sigma \)

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: $k = (k_1, \ldots, k_n)$, with $k_j = \# \text{obs. in } I_j \Rightarrow f(k \mid \Lambda) = e^{-\Lambda(T)} \sum_{j=1}^{n} \frac{(\Lambda I_j)^{k_j}}{k_j!}$

Parametric: $P(k \mid H_P) = e^{-\Lambda(T)} \prod_{j=1}^{n} \frac{[\Lambda_{i,j}]^{k_j}}{k_j!} \pi(\theta) d\theta$

Nonparametric: $k \mid M, \theta \sim f(k \mid M_{\theta})$, $M \mid \theta \sim \text{CG}(\Lambda_{\theta}/\sigma, \sigma)$ and $\theta \sim \pi$:

$$P(k \mid H_N) = \prod_{i=0}^{n} \sum_{j=1}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}{\prod_{i=0}^{n} \sum_{j=1}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}{\prod_{i=0}^{n} \sum_{j=1}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}} \pi(\theta) d\theta$$

Bayes Factor: $BF_{PN} = \frac{P(k \mid H_P)}{P(k \mid H_N)} = \frac{\sum_{i=0}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}{\prod_{i=0}^{n} \sum_{j=1}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}} \pi(\theta) d\theta}{\sum_{i=0}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}{\prod_{i=0}^{n} \sum_{j=1}^{\Delta_i I_j} \frac{k_j! \exp^{\sigma} + k_j \ln(1 + \sigma)}} \pi(\theta) d\theta}$
PARAMETRIC VS. NONPARAMETRIC

Bayes factor $BF_{P \mid N}$ as a function of $\sigma$
SUBWAY TRAINS FAILURES

*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

\[ \lambda(t; \theta_1, \theta_2) = \mu(g(t); \theta_1) s(t; \theta_2) \]

\[ \cdot \mu(k; \theta_1) = \beta_0 \frac{\log(1 + \beta_1 k)}{(1 + \beta_1 k)} \]

\[ \mu(0; \theta_1) = 0, \text{ maximum at } (e - 1)/b_1 \text{ and } \lim_{k \to \infty} \mu(k; \theta_1) = 0 \]

\[ \cdot \text{m.v.f. } \Lambda(k) = \beta_0 \log^2(1 + \beta_1 k)/(2\beta_1) \]

suitable for actual cumulative number of failures

\[ \cdot s(t; \theta_2) = \exp\{\rho \cos(\omega t + \phi)\}\text{(periodic component)} \]

\[ \cdot \text{From EDA we could take } k = g(t) = at + bt^2 \text{ and substitute above} \]

\[ \cdot \text{We actually took kilometers } k|t \sim N(g(t), \sigma^2) \]
MODEL FOR DOORS FAILURES

- $j$-th train monitored in $[0, T_j]$
- Failures at times $(t_1, \ldots, t_{n_j}) = t_j$ and kilometers $(k_1, \ldots, k_{n_j}) = k_j$
- 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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MLE</th>
<th>C.I.</th>
<th>Parameter</th>
<th>MLE</th>
<th>C.I.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a \times 10^{-2}$</td>
<td>1.209</td>
<td>[1.171, 1.247]</td>
<td>$b \times 10^2$</td>
<td>2.025</td>
<td>[1.862, 2.188]</td>
</tr>
<tr>
<td>$\sigma^2 \times 10^{-7}$</td>
<td>5.809</td>
<td>[4.214, 6.345]</td>
<td>$\rho \times 10$</td>
<td>3.234</td>
<td>[0.000, 6.779]</td>
</tr>
<tr>
<td>$\beta_0 \times 10^2$</td>
<td>7.358</td>
<td>[5.640, 9.076]</td>
<td>$\beta_1 \times 10^5$</td>
<td>2.239</td>
<td>[1.938, 2.540]</td>
</tr>
</tbody>
</table>
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.

- $\hat{\Lambda}(t)$ estimated from data $T_1, T_2, \ldots$
- Suppose $T_1, T_2, \ldots$ from NHPP with m.v.f. $\hat{\Lambda}(t)$
- $Y_1 = \hat{\Lambda}(T_1), Y_2 = \hat{\Lambda}(T_2), \ldots$ data from HPP with rate 1
- Interarrival times $X_i = Y_i - Y_{i-1}$ i.i.d. $E(1)$
- $U_i = \exp\{-X_i\}$ i.i.d. $U[0, 1]$
- Should $\hat{\Lambda}(t)$ 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

• ⇒ 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 \( E(N(2, 2 + i)|N(0, 2)) = \int \Lambda((2, 2 + i)|\theta) \pi(\theta|N(0, 2)) d\theta \), 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 $g(t)$ realization of a Gamma process

$$
\begin{align*}
g(t) & \sim G(at, b) \\
\theta & \sim \pi(\theta) \\
[t \mid g, \theta] & = \text{NHP } P\{\mu(g(t); \theta_1) s(t; \theta_2)\} \\
[k \mid t, g] & = \delta_{g(t)}(\cdot) \\
\end{align*}
$$

- $g$ 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 $g(t_{i-1})$

- $g(t)$ updated with an acceptance/rejection step
An example of $g$ during the MCMC run
INTENSITY AND MEAN VALUE FUNCTION ESTIMATION

![Graphs of Intensity and Mean Value Function for Train 19 and Train 20](image)
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**

<table>
<thead>
<tr>
<th>Code</th>
<th>Subsystem</th>
<th>No. of parts</th>
<th>Total failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>opening commands (electrical)</td>
<td>14</td>
<td>530</td>
</tr>
<tr>
<td>2</td>
<td>cables and clamps</td>
<td>4</td>
<td>33</td>
</tr>
<tr>
<td>3</td>
<td>mechanical parts</td>
<td>67</td>
<td>1182</td>
</tr>
<tr>
<td>4</td>
<td>electrical protections</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>power supply circuit</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>6</td>
<td>pneumatic gear</td>
<td>31</td>
<td>295</td>
</tr>
<tr>
<td>7</td>
<td>electro-valves</td>
<td>8</td>
<td>39</td>
</tr>
</tbody>
</table>

*Classification of failure modes and total failures per mode for all trains in nine years*
• Failure modes 4 and 5 very rare ⇒ not enough information for fitting a stochastic process model

• Failure modes 6 (QUESTION: Comments on it?) and 7 show change-points
• Failure modes 1, 2 and 3 display a more regular pattern
• Mode 2 failures are only 0.11 per train and per year
  • ⇒ concentrate on failure modes 1 and 3
• Different average daily distance

• More recent trains are used less daily
BIVARIATE INTENSITY FUNCTION

For each train \( i \)

\[
\lambda_i(t, s) = \mu \exp^{-\gamma (s - a_i - c_i(t - t_{0i}))^2 w(t - t_{0i})} \cdot \exp\{A \cos(\omega(t - d))\} \lambda_0(t - t_{0i})
\]

- \( t_{0i} \) 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. \( w(z) = 1 + \frac{1}{1 + z} \), 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
• Periodogram of monthly time series of failure modes 1 and 3 (after detrending)

• No clear frequency for failure mode 1 ⇒ 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)

- \( \Lambda_0(u) = \frac{1 - e^{-bt}}{b} \) (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 \( \Lambda(t; \hat{\theta}) = \sum_{i=1}^{40} \int_{t_0}^{t} \lambda_i(u; \hat{\theta}) \, du, \quad t = 1, \ldots, 3287 \)
  - \( \hat{\theta} \) estimate of \( \theta \) from MCMC run
  - \( \lambda_i(t) = \mu \frac{n}{\gamma w(t-t_{0i})} \Phi(\gamma w(t-t_{0i})) \cdot 2 \gamma w(t-t_{0i}) \cdot \exp\{A \cos(\omega (t-d))\} \lambda_0(t-t_{0i}) \)
  (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 $\theta$

Predictive distribution

$$P(N_{T_0 + u} - N_{T_0} \mid D_{T_0}) = \frac{\cdot e^{-\{\Lambda(T_0 + u;\theta) - \Lambda(T_0;\theta)\} \cdot \{\Lambda(T_0 + u;\theta) - \Lambda(T_0;\theta)\}^x \cdot \pi(\theta \mid D_{T_0}) \cdot d\theta}}{x!}$$

Expected value

$$E(N_{T_0 + u} - N_{T_0} \mid D_{T_0}) = \cdot \{\Lambda(T_0 + u;\theta) - \Lambda(T_0;\theta)\} \cdot \pi(\theta \mid D_{T_0}) \cdot d\theta$$
FORECAST OF FUTURE FAILURES OF MODE 1

<table>
<thead>
<tr>
<th>end of recording period</th>
<th>forecasting horizon (years)</th>
<th>95% credibility interval</th>
<th>true value</th>
<th>posterior mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1992</td>
<td>1</td>
<td>(86, 143)</td>
<td>83</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(79, 140)</td>
<td>72</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(71, 138)</td>
<td>62</td>
<td>105</td>
</tr>
<tr>
<td>1993</td>
<td>1</td>
<td>(69, 124)</td>
<td>72</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(59, 121)</td>
<td>62</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(50, 119)</td>
<td>42</td>
<td>85</td>
</tr>
<tr>
<td>1994</td>
<td>1</td>
<td>(50, 100)</td>
<td>62</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(41, 95)</td>
<td>42</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(32, 91)</td>
<td>35</td>
<td>59</td>
</tr>
<tr>
<td>1995</td>
<td>1</td>
<td>(38, 81)</td>
<td>42</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(30, 74)</td>
<td>35</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>(24, 68)</td>
<td>23</td>
<td>44</td>
</tr>
<tr>
<td>1996</td>
<td>1</td>
<td>(27, 60)</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>(20, 52)</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>(19, 46)</td>
<td>23</td>
<td>39</td>
</tr>
</tbody>
</table>
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

$$
\Pr(N_H(T_0) > x_U \mid D_t) = 1 - \sum_{x=0}^{x_U} \frac{\Lambda_H(T_0; \theta)}{x!} \pi(\theta \mid D_t) \, d\theta
$$

<table>
<thead>
<tr>
<th>f. mode 1</th>
<th>$x_U$</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>f. mode 3</th>
<th>$x_U$</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>12</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.00</td>
</tr>
</tbody>
</table>
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.

- 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)

- Here interest in bugs detection during software testing
SOFTWARE RELIABILITY

Failures at $T_1, T_2, \ldots, T_n$

Inter-failure times $T_i - T_{i-1} \sim E(\lambda_i)$, independent, $i = 1, \ldots, n$

- $\lambda_i = \varphi(N - i + 1)$, $\varphi \in \mathbb{R}^+$, $N \in \mathbb{N}$,  
  (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 $N \sim P(\nu)$ and $\varphi \sim G(\alpha, \beta)$
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 $P$ with rows $P_i = (P_{i1}, \ldots, P_{ik})$, $i = 1, \ldots, k$
HIDDEN MARKOV MODEL

- Interarrival times of $m$-th failure $X_m|Y_m = i \sim E(\lambda(i)), i = 1, \ldots, k, m = 1, \ldots, n$

- $X_m$'s independent given $Y \Rightarrow f(X_1, \ldots, X_n|Y) = \prod_{m=1}^{n} f(X_m|Y)$

- $P_i \sim Dir(\alpha_i^1, \ldots, \alpha_i^k), \forall i \in E$, i.e. $\pi(P_i) \propto \prod_{j=1}^{k} P_{ij}^{\alpha_i^j - 1}$

- Independent $\lambda(i) \sim G(a(i), b(i)), \forall i \in E$

- Interest in posterior distribution of $\Theta = (\lambda^{(k)}, P, Y^{(n)})$
  - $\lambda^{(k)} = (\lambda(1), \ldots, \lambda(k))$
  - $Y^{(n)} = (Y_1, \ldots, Y_n)$
HIDDEN MARKOV MODEL: CRITICAL ISSUES

• Ranking of reliability states through ordered $\lambda$’s
  
  – Order preserving prior leads to unjustified (by data) equality of adjacent $\lambda$’s
  
  – Label switching when considering independent $\lambda$’s

• Number of states $K$
  
  – Reversible jump MCMC
  
  – Bayes factor out of MCMC (Chib’s method)
SOFTWARE RELIABILITY - MUSA DATA

• Musa System 1 data: 136 software failure times

• Hidden Markov model with 2 unknown states
SOFTWARE RELIABILITY - MUSA DATA


<table>
<thead>
<tr>
<th>Lambda[1]</th>
<th>0.0004</th>
<th>0.0008</th>
<th>0.0012</th>
<th>0.0016</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lambda[2]</td>
<td>0.005</td>
<td>0.010</td>
<td>0.015</td>
<td>0.020</td>
<td>0.0</td>
</tr>
</tbody>
</table>
SOFTWARE RELIABILITY - MUSA DATA

Posterior Distribution of $P_{1,1}$

Posterior Distribution of $P_{1,2}$

Posterior Distribution of $P_{2,1}$

Posterior Distribution of $P_{2,2}$
SOFTWARE RELIABILITY - MUSA DATA

Time Series Plot of Failure Times

Time Series Plot of Posterior Probabilities of $Y(t)=1$

Longer failure times $\Rightarrow$ higher Bayes estimator of probability of "good" state