Bayesian Robustness

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/

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. $\mathbb{P}(P^C|D^C) = .9$
- Physician's belief on patient having the disease 1%, i.e. $\mathbb{P}(D) = .01$
- Positive test $\Rightarrow \mathbb{P}(D|P)$?

$$\begin{aligned}\mathbb{P}(D|P) &= \frac{\mathbb{P}(D\bigcap P)}{\mathbb{P}(P)} = \frac{\mathbb{P}(P|D)\mathbb{P}(D)}{\mathbb{P}(P|D)\mathbb{P}(D) + \mathbb{P}(P|D^C)\mathbb{P}(D^C)} \\ &= \frac{.95 \cdot .01}{.95 \cdot .01 + .1 \cdot .99} = .0875\end{aligned}$$

- Positive test updates belief on patient having the disease: from 1% to 8.75%
- *Prior* opinion updated into *posterior* one

ILLUSTRATIVE EXAMPLE

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 $\hat{\lambda} = \dfrac{\alpha + n}{\beta + \sum_{i=1}^{n} X_i}$

  $\Rightarrow$ compare with

  - prior mean $\frac{\alpha}{\beta}$

  - MLE $\frac{n}{\sum_{i=1}^{n} X_i}$

- MAP (Maximum a posteriori)

  $\Rightarrow \hat{\lambda} = \dfrac{\alpha + n - 1}{\beta + \sum X_i}$

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

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

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

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

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

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 straightforward computations in complex models

PRIOR AND DATA INFLUENCE

- Posterior mean: $\widehat{\lambda} = \dfrac{\alpha + n}{\beta + \sum X_i}$
- Prior mean: $\widehat{\lambda}_P = \dfrac{\alpha}{\beta}$ (and variance $\sigma^2 = \dfrac{\alpha}{\beta^2}$)
- MLE: $\widehat{\lambda}_M = n / \sum X_i$
- $\alpha_1 = k\alpha$ and $\beta_1 = k\beta \Rightarrow \widehat{\lambda}_{1P} = \widehat{\lambda}_P$ and $\sigma_1^2 = \sigma^2 / k$
- Posterior mean: $\widehat{\lambda} = \dfrac{k\alpha + n}{k\beta + \sum X_i}$
- $k \to 0 \Rightarrow$ prior variance $\to \infty \Rightarrow \widehat{\lambda} \to n / \sum X_i$, i.e. MLE (prior does not count)
- $k \to \infty \Rightarrow$ prior variance $\to 0 \Rightarrow \widehat{\lambda} \to \widehat{\lambda}_P$, i.e. prior mean (data do not count)
- $n \to \infty \Rightarrow \widehat{\lambda} \sim \dfrac{n}{\sum X_i}$, i.e. MLE (prior does not count)

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 $\mathbf{E}\lambda^k$ 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)

- Bets and lotteries

- ...

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 $\mathcal{G}(\alpha, \beta)$ - choice of hyperparameters

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

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

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

- *Empirical Bayes*: choose $(\hat{\alpha}, \hat{\beta}) = \arg\max \int f(X_1, \ldots, X_n|\lambda)\pi(\lambda|\alpha, \beta)d\lambda$

MANY CRITICAL ASPECTS

- Choice of a model $f(X|\lambda)$ for X
- Choice of the prior $\pi(\lambda|\theta)$
  - *physical meaning* of λ
  - functional form of π
  - elicitation of experts' opinions (in a finite time)
  - choice of hyperparameters
- Choice of a loss function (and an estimator)

$\Rightarrow$ How is the statistical analysis affected by such uncertainty and, sometimes, arbitrariness ?

MOTIVATING EXAMPLE (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 Cauchy $\mathcal{C}(0, 1)$ and Gaussian $\mathcal{N}(0, 2.19)$
- Interest in posterior mean $\mu^C(x)$ or $\mu^N(x)$

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

- Decision strongly dependent on the choice of the prior for large x
- Alternative: Posterior median w.r.t. posterior mean

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

A SHORT HISTORY ON BAYESIAN ROBUSTNESS

- Early work by Good in the ’50s
- Kadane and Berger in mid ’80s
- Berger and O’Hagan at Valencia meeting in 1988
- Berger in JSPI (1990) and TEST (1994)
- Workshops in Milano (1992) and Rimini (1995) and their proceedings
- MCMC in mid 90’s
- Rios Insua and Ruggeri (2000)
- Special issue of IJAR (2009)

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

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)*
  - $\Gamma_P = \{\mathcal{G}(\alpha,\beta) : \alpha/\beta = \mu\}$
  - $\Gamma_P = \{\mathcal{G}(\alpha,\beta) : l_1 \le \alpha \le u_1, l_2 \le \beta \le u_2\}$
  - $\Gamma_P = \{\mathcal{G}(\alpha,\beta) : l_1 \le \alpha/\beta \le u_1, l_2 \le \alpha/\beta^2 \le u_2\}$

CLASSES OF PRIORS

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

CLASSES OF PRIORS

- $\Gamma_{GM} = \{P : \int h_i(\theta)dP(\theta) = a_i, i = 1, \ldots, m\}$ *(Generalised 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 with given marginals

- $f(X|\theta_1, \theta_2)$
- $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ known
- $\pi(\theta_1, \theta_2)$ unknown
- Fréchet class of priors $\pi(\theta_1, \theta_2)$ or copulas $C(\Pi_1(\theta_1), \Pi_2(\theta_2))$
- *Fréchet Theorem*: $W(\theta_1, \theta_2) \leq \Pi(\theta_1, \theta_2) \leq M(\theta_1, \theta_2)$
  - $W(\theta_1, \theta_2) = \max\{\Pi_1(\theta_1) + \Pi_2(\theta_2) - 1, 0\}$
  - $M(\theta_1, \theta_2) = \min\{\Pi_1(\theta_1), \Pi_2(\theta_2)\}$

CLASSES OF PRIORS

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

- ... whereas now we introduce others (*Neighbourhood 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, generalised moments contraints class, etc.
  - $\epsilon = \epsilon(\theta)$ (need to normalise!)

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)*

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

COMPARISON OF PROBABILITY MEASURES

$\mathcal{P}$: all probability measures on $(\Theta, \mathcal{F})$, Θ Polish space

$P_0(E) = \frac{\varepsilon}{10}$: ranges of $P(E)$ in neighbourhoods of P_0

1. Variational distance : $|P(A) - P_0(A)| \leq \varepsilon, \forall A \in \mathcal{F}$
   $\Rightarrow P(E) \leq 11\frac{\varepsilon}{10}$

2. ε–contaminations (contaminating measures in $\mathcal{P}$) :
   $-\varepsilon P_0(A) \leq P(A) - P_0(A) \leq \varepsilon P_0(A^C), \forall A \in \mathcal{F}$
   $\Rightarrow (1-\varepsilon)\frac{\varepsilon}{10} \leq P(E) \leq (1-\varepsilon)\frac{\varepsilon}{10} + \varepsilon$

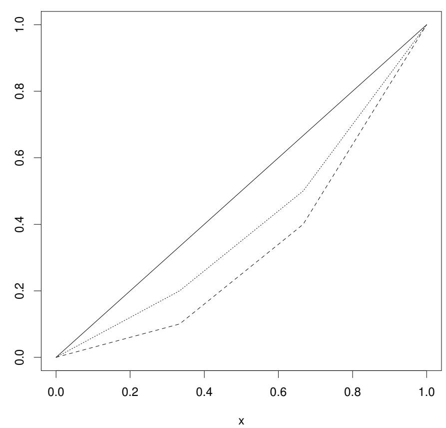
3. $|P(A) - P_0(A)| \leq \varepsilon \min\{P_0(A), P_0(A^C)\}, \forall A \in \mathcal{F}$
   $\Rightarrow (1-\varepsilon)\frac{\varepsilon}{10} \leq P(E) \leq (1+\varepsilon)\frac{\varepsilon}{10}$

4. $|P(A) - P_0(A)| \leq P_0(A)P_0(A^C), \forall A \in \mathcal{F}$
   $\Rightarrow \frac{\varepsilon^2}{100} \leq P(E) \leq (2 - \frac{\varepsilon}{10})\frac{\varepsilon}{10}$

CONCENTRATION FUNCTION CLASS

- n individuals with wealth $x_i, i = 1, \ldots, n \Rightarrow$ ordered $x_{(1)} \leq \ldots \leq x_{(n)}$
- $(k/n, S_k/S_n), k = 0, \ldots, n$, $S_0 = 0$ and $S_k = \sum_{i=1}^{k} x_{(i)}$ (Lorenz curve)
- Comparison of discrete p.m.'s with uniform

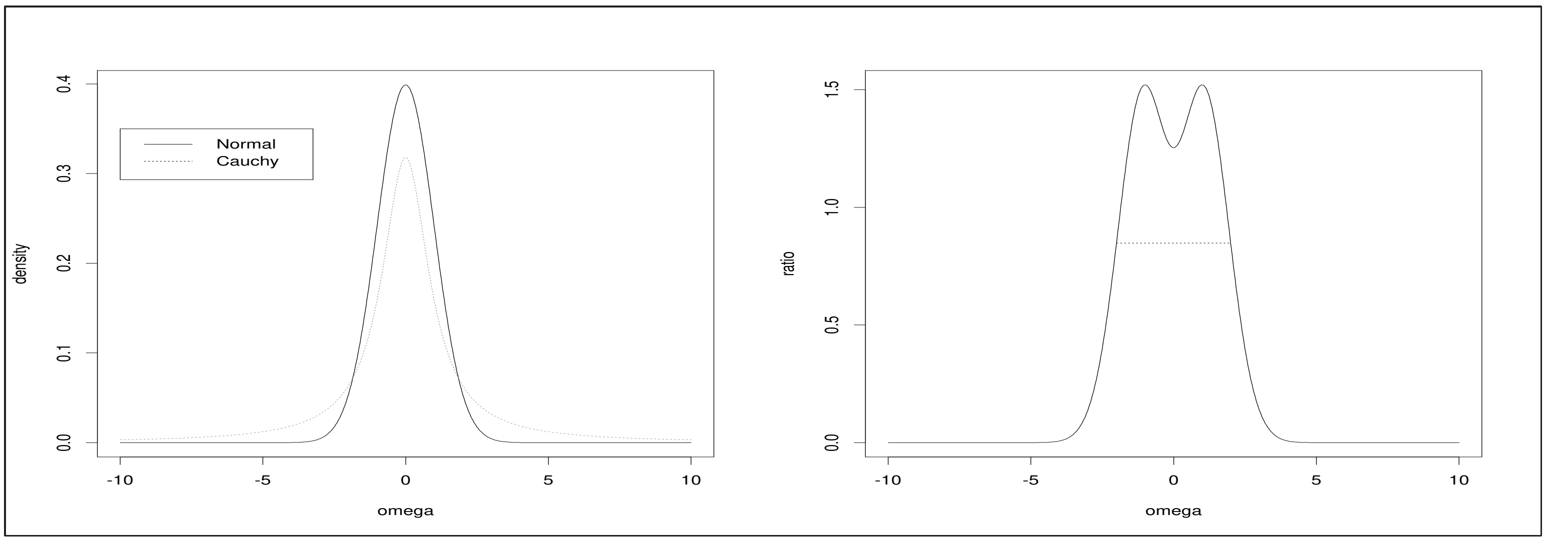
  Example: $(0.2, 0.3, 0.5)$ & $(0.1, 0.3, 0.6)$ vs. $(1/3, 1/3, 1/3)$



Comparison of two p.m.'s on same $(\Omega, \mathcal{F}, P) \Rightarrow$ concentration function

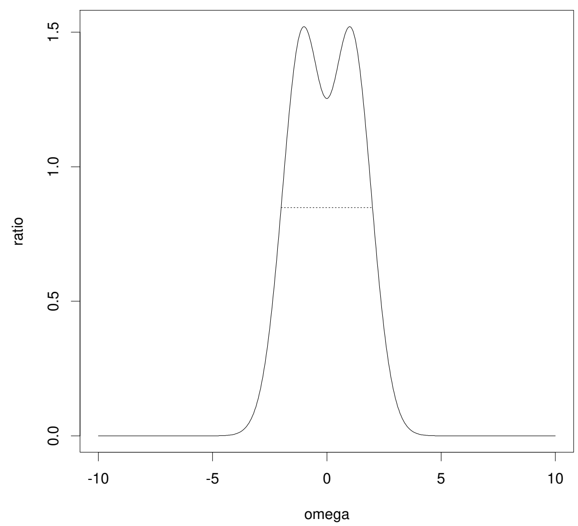
CONCENTRATION FUNCTION CLASS

- P, P_0 probability measures on $(\Omega, \mathcal{F})$
- σ–finite ν dominating $P, P_0 \Rightarrow p(\omega), p_0(\omega)$
- $P \sim \mathcal{N}(0,1), P_0 \sim \mathcal{C}(0,1)$



Densities $\mathcal{N}(0,1)$ and $\mathcal{C}(0,1)$ (left) - likelihood ratio (right)

CONCENTRATION FUNCTION CLASS



- Each horizontal line at $y \Rightarrow$ subset A_y with likelihood ratio $m(\omega) = \dfrac{p(\omega)}{p_0(\omega)} \leq q$
- If $P_0(A_y) = x \Rightarrow A_y$ is the subset of P_0-measure x with smallest P-measure $\varphi(x)$
- The pairs $(x, \varphi(x))$ determine the c.f.

CONCENTRATION FUNCTION CLASS

- (h, N) Lebesgue decomposition of P w.r.t. P_0
- $N = \{\omega \in \Omega : p_0(\omega) = 0\}$
- $m(\omega) = \begin{cases} p(\omega)/p_0(\omega) & \omega \in N^C \\ \infty & \omega \in N \end{cases}$
- $P(A) = P_s(A) + P_a(A), \ \forall A \in \mathcal{F}$
- $P_a(A) = \int_{A \cap N^C} m(\omega) P_0(d\omega), \ P_s(A) = P(A \cap N)$
- $P_a \ll P_0, P_s \perp P_0$

CONCENTRATION FUNCTION CLASS

- $H(y) = P_0(\{\omega \in \Omega : m(\omega) \leq y\})$

- $c_x = \inf\{y \in \Re : H(y) \geq x\}$

- $L_x = \{\omega \in \Omega : m(\omega) \leq c_x\}, L_x^- = \{\omega \in \Omega : m(\omega) < c_x\}$

- $\varphi(x) = \begin{cases} 0 & x = 0 \\ P(L_x^-) + c_x\{x - H(c_{x^-})\} & x \in (0,1) \\ P_a(\Omega) & x = 1 \end{cases}$

CONCENTRATION FUNCTION CLASS

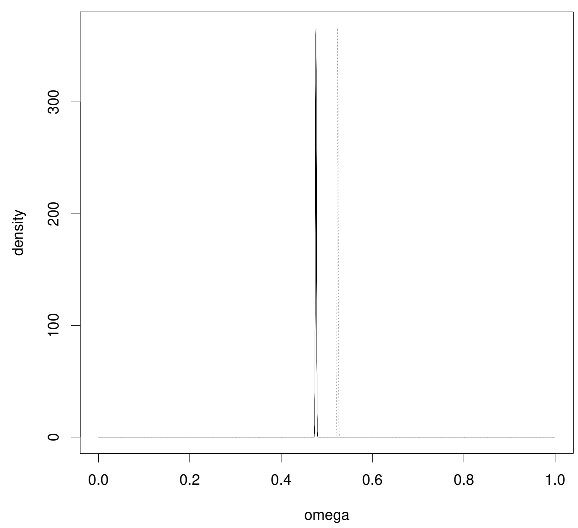
Main properties

- $\varphi(x)$ nondecreasing, continuous and convex, $\varphi(0) = 0$
- $\varphi(x) \equiv 0 \Leftrightarrow P \perp P_0$
- $\varphi(x) = x, \forall x \in [0, 1] \Leftrightarrow P = P_0$
- $P_0(A) = x \Rightarrow \varphi(x) \leq P(A) \leq 1 - \varphi(1 - x)$
- $\varphi(x) = \int_0^{c_x} \{x - H(t)\} \, dt = \int_0^x c_t dt$
- $\lim_{n\to\infty} \varphi_{P_n}(x) = x, \forall x \in [0, 1] \Leftrightarrow \lim_{n\to\infty} \sup_{A\in\mathcal{F}} |P_n(A) - P_0(A)| = 0$

CONCENTRATION FUNCTION CLASS

Two Beta distributions P and P_0 with

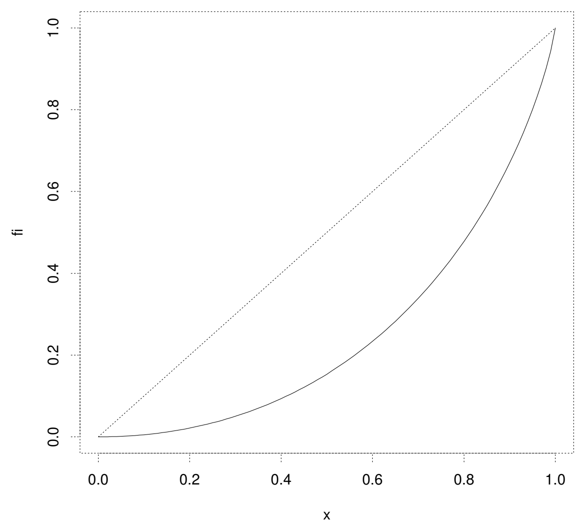
- very close mean, median and mode
- c.f. of P w.r.t. P_0 :$\varphi(x) \approx 0, x \in [0, 1)$
- The two distributions are *very different* since P_0 concentrates mass (i.e. gives very high probability) to a subset of negligible probability under P



CONCENTRATION FUNCTION CLASS

Concentration function of $P \sim \mathcal{G}(2,1)$ w.r.t. $P_0 \sim \mathcal{E}(1)$

- $p_0(\theta) = e^{-\theta}, p(\theta) = \theta e^{-\theta}, \theta \geq 0$
- $m(\theta) = p(\theta)/p_0(\theta) = \theta$
- Find $y : x = P_0\left(\{\theta \in \Theta : m(\theta) \leq y\}\right) = 1 - e^{-y}$
  $\Rightarrow \varphi(x) = P\left(\{\theta \in \Theta : m(\theta) \leq y\}\right) = 1 - (1-x)(1 - \log(1-x))$

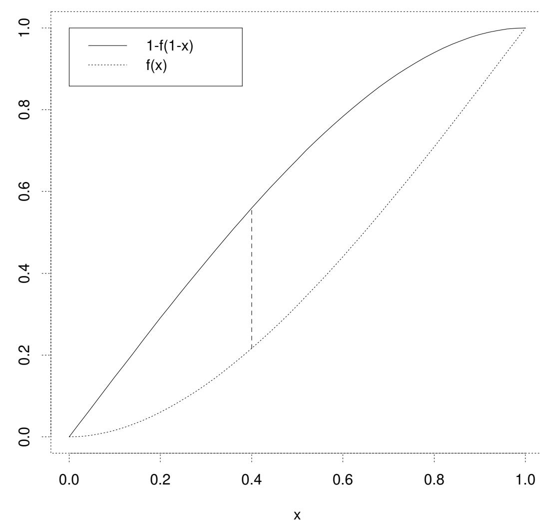


CONCENTRATION FUNCTION CLASS

C.f. of $P \sim \mathcal{G}(2,2)$ w.r.t. $P_0 \sim \mathcal{E}(1)$: $p_0(\theta) = e^{-\theta}, p(\theta) = 4\theta e^{-2\theta}, m(\theta) = 4\theta e^{-\theta}$

- Take $\{y_j\}$ and find $L_{y_j} = \{\theta \in \Theta : m(\theta) \leq y_j\}$
- Compute $x_j = P_0(L_{y_j})$ and $\varphi(x_j) = P(L_{y_j})$

$P_0(A) = x \Rightarrow \varphi(x) \leq P(A) \leq 1 - \varphi(1-x)$
$P_0(A) = .4 \Rightarrow .216 \leq P(A) \leq .559$



CONCENTRATION FUNCTION CLASS

- g monotone nondecreasing, continuous, convex: $g(0) = 0$ and $g(1) \leq 1$
- $K_g = \{P : P(A) \geq g(P_0(A)) \;\; \forall A \in \mathcal{F}\}$, g-neighborhood of non-atomic P_0
  - $g(P_0(A)) = P_0(A)P_0(A^C)$
  - $g(P_0(A)) = \min\{P_0(A), P_0(A^C)\}$
- $P \in K_g \Rightarrow g(P_0(A)) \leq P(A) \leq 1 - g(1 - P_0(A))$
- $\{K_g\}$ generates a topology over $\mathcal{P}$
- $\exists$ at least one P : g is the concentration function $\varphi_P(x)$ of P w.r.t. P_0
- $K_g = \{P : \varphi_P(x) \geq g(x), \forall x \in [0, 1]\}$
- $P \in K_g$ mixture of extremal p.m.'s in $E_g = \{P : \varphi_P(x) = g(x), \forall x \in [0, 1]\}$
- $\Rightarrow \sup_{P \in K_g} E[k(\theta)] = \sup_{P \in E_g} E[k(\theta)]$

CONCENTRATION FUNCTION CLASS

Neighbourhood of the uniform distribution

- $X \sim$*Bin*$(2, \theta)$
  $\Rightarrow f(x|\theta) = \binom{2}{x} \theta^x (1-\theta)^{2-x}, \theta \in [0,1], x = 0, 1, 2$

- P_0 uniform over $[0, 1]$

- Choose a class of priors P s.t.
  $|P_0(A) - P(A)| \leq P_0(A) P_0(A^C), \forall A \in \mathcal{F}$

- $\Rightarrow \varphi(x) \geq x^2 = g(x), \forall x \in [0, 1]$

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 generalised moments constrained class)

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

POISSON PROCESS

- Intensity function: $\lambda(t) = \lim_{\Delta t \to 0} \dfrac{P(N(t, t+\Delta t] \geq 1)}{\Delta t}$
  - HPP (homogeneous Poisson process): constant $\lambda(t) = \lambda$, $\forall t$
  - NHPP (nonhomogeneous Poisson process): o.w.
- HPP with rate λ
  - $N(s, t] \sim \mathcal{P}\lambda(t-s)$
  - 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

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 $0 \le t \le t_0$ (uniform distribution for HPP)

OBSERVABLE QUANTITIES

- Actual prior elicitation better performed if done on observable quantities
- Failures in repairable systems modelled by NHPP
- PLP (Power Law process) $\Rightarrow \lambda(t) = M\beta t^{\beta-1}$ and $\Lambda(t) = Mt^{\beta}$
- Expert asked about lower and upper bounds on time of first failure T_1
  i.e. $l_i \leq \mathcal{P}(T_1 > s_i) = \mathcal{P}(N(s_i) = 0) \leq u_i, i = 1, n$
  - $\mathcal{P}(T_1 > s_i | M, \beta) = \exp\{-M s_i^{\beta}\}$
  - $\mathcal{P}(T_1 > s_i) = \int \mathcal{P}(T_1 > s_i | M, \beta)\pi(M, \beta) dM d\beta$
- Suppose M known
- Generalised moments constrained class on β given by
  $l_i \leq \int_0^{\infty} \exp\{-M s_i^{\beta}\}\pi(\beta) d\beta \leq u_i,\ i = 1, n$

NEAR IGNORANCE

- Improper priors
- Uniform distribution on *large* interval (for unbounded Θ)
- Neighbourhood 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 $\theta, \sigma > 0, \beta \in (-1, 1]$

Skew-normal class of distributions

$$\Lambda_{SN} = \left\{ f(y|\alpha,\xi,\tau) = \frac{2}{\tau}\,\phi\left(\frac{y-\xi}{\tau}\right)\;\Phi\left(\alpha\frac{\theta-\xi}{\tau}\right) \right\}$$

for any α and ξ, and $\tau > 0$

CLASSES OF NHPPs

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

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 MODELS

Neighbourhood 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 neighbourhood)

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 normalise $f(x|\theta)$

CLASSES OF LOSSES

Interest in behaviour 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 minimising the posterior expected loss, but very different posterior expected loss $\Rightarrow$ robustness calibration

CLASSES OF LOSSES

Mixtures of convex loss functions

- $L_\lambda \in \Psi$, family of convex loss functions, $\lambda \in \Lambda$
- $G \in \mathcal{P}$, class of all probability measures on $(\Lambda, \mathcal{A})$
- $\Omega = \{L : L(\theta, a) = \int_\Lambda L_\lambda(\theta, a) dG(\lambda)\}$
- a_L Bayes action for loss L, under probability measure π
- $\underline{a} = \inf_{L_\lambda \in \Psi} a_{L_\lambda}, \ \overline{a} = \sup_{L_\lambda \in \Psi} a_{L_\lambda} \Rightarrow \underline{a} \leq a_L \leq \overline{a}, \ \forall L \in \Omega$
  - $L_\lambda(\theta, a) = |\theta - a|^\lambda, \ \lambda \geq 1$
  - $L_\lambda(\theta, a) = e^{\lambda(a-\theta)} - \lambda(a - \theta) - 1, \ , \lambda_1 \leq \lambda \leq \lambda_2$
  - $L_\lambda(\theta, a) = \chi_{[a-\lambda, a+\lambda]^C}(\theta), \ \lambda > 0$

LOSS ROBUSTNESS

Preference among losses

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

L_1 preferred to L_2 (Makov, 1994) if

- $\sup_x \inf_a \rho_{L_1}(\pi, x, a) < \sup_x \inf_a \rho_{L_2}(\pi, x, a)$
  (posterior minimax)
- $\mathcal{E}_X \rho_{L_1}(\pi, x, a_\pi^{L_1}) < \mathcal{E}_X \rho_{L_2}(\pi, x, a_\pi^{L_2})$
  (preposterior)
- $\sup_x \left| \frac{\partial}{\partial x} \rho_{L_1}(\pi, x, a_\pi^{L_1}) \right| < \sup_x \left| \frac{\partial}{\partial x} \rho_{L_2}(\pi, x, a_\pi^{L_2}) \right|$
  (influence approach)

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

SENSITIVITY MEASURES

Relative sensitivity: $\sup_\pi R_\pi$

with $R_\pi = \dfrac{(\rho_\pi - \rho_0)^2}{V^\pi}$, $\rho_0 = E_{\Pi_0^*}[h(\theta)]$, $\rho_\pi = E_{\Pi^*}[h(\theta)]$ and $V^\pi = Var_{\Pi^*}[h(\theta)]$

- Scale invariant (*calibration*)
- Decision theoretic interpretation (relative posterior expected loss increase when considering π_0 instead of *true* π)
  - Square loss function ($\Rightarrow$ posterior mean as Bayes action)
  - π_0 chosen prior $\Rightarrow$ ρ_0 Bayes action
  - $\pi \neq \pi_0$, $\pi \in \Gamma$, *true* prior $\Rightarrow$ ρ_π *true* Bayes action
  - Use of ρ_0 instead of ρ_π $\Rightarrow$ expected loss $(\rho_\pi - \rho_0)^2 + V^\pi$ instead of V^π
  - $R_\pi = \dfrac{(\rho_\pi - \rho_0)^2}{V^\pi}$ relative increase in expected loss
  - $\sup_\pi R_\pi$ maximum expected loss over all possible *true* priors in Γ
- Asymptotic behaviour *as expected*

SENSITIVITY MEASURES

Local sensitivity

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

Measures

- Derivatives of extrema in $\{K_\varepsilon\}, \varepsilon \geq 0$, neighbourhood 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 minimised by a_π^L

- $\sup_{L \in \mathcal{L}} \rho_L(\pi, x, a) - \inf_{L \in \mathcal{L}} \rho_L(\pi, x, a)$

- $\sup_{L \in \mathcal{L}} a_\pi^L - \inf_{L \in \mathcal{L}} a_\pi^L$

- $\sup_x \left| \frac{\partial}{\partial x} \rho_L(\pi, x, a_\pi^L) \right| - \inf_x \left| \frac{\partial}{\partial x} \rho_L(\pi, x, a_\pi^L) \right|$

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

- Linearisation 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 $c(q) = 0$
- Discretisation of $\Theta \Rightarrow$ Linear programming
- Linear Semi-infinite Programming (for Generalised 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 (see next)
- Acquisition of new data

APPLICATIONS

- Number of accidents X_k for a company with n_k workers at time period k
- $X_k|\theta, n_k \sim \mathcal{P}(n_k\theta)$
- $\Gamma = \{\pi : \pi(0, .38] = .25, \pi(.38, .58] = .25, \pi(.58, .98] = .25,\ \pi(.98, \infty) = .25\}$
- Year 1988: $\underline{E}[X_k|D_k]/n_k = 0.05$ and $\bar{E}[X_k|D_k]/n_k = 0.58$
- Fréchet derivative of $E[X_k|D_k]/n_k \Rightarrow$ sum of contributions from each interval
- Split interval with largest contribution (here first)
- Year 1988: $\underline{E}[X_k|D_k]/n_k = 0.15$ and $\bar{E}[X_k|D_k]/n_k = 0.24$

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

HIERARCHICAL MODEL

- $Y_i|\lambda_i \sim \mathcal{P}(\lambda_i), i = 1, n$
- $\lambda_i|\alpha, \beta \sim \mathcal{G}(\alpha, \beta)$
- $\pi(\alpha, \beta)$
- "Pure" Bayesian approach $\Rightarrow$ prior on (α, β)
- Improper priors
- Empirical Bayes
  - $\lambda_i|\alpha, \beta, \underline{Y} \sim \mathcal{G}(\alpha + y_i, \beta + 1)$, $\lambda_i \perp \lambda_j|\underline{Y}$
  - $f(\underline{Y}|\alpha, \beta) = \int f(\underline{Y}|\underline{\lambda})\pi(\underline{\lambda}|\alpha, \beta)d\underline{\lambda}$ maximized by $(\widehat{\alpha}, \widehat{\beta})$

    $\Rightarrow \lambda_i|\widehat{\alpha}, \widehat{\beta}, \underline{Y} \sim \mathcal{G}(\widehat{\alpha} + y_i, \widehat{\beta} + 1), \forall i$

BAYESIAN NONPARAMETRICS

- Dirichlet process
  - $P \sim \mathcal{DP}(\eta)$ if $\forall (A_1, \ldots, A_m)$
    $\Rightarrow (P(A_1), \ldots, P(A_m)) \sim \mathcal{D}(\eta(A_1), \ldots, \eta(A_m))$
  - $Z_1, \ldots, Z_n$ sample of size n from P
    $\Rightarrow P|Z_1, \ldots, Z_n \sim \mathcal{DP}(\eta + \sum_1^n \delta_{Z_i})$
- Embed parametric model $P_0(x)$ in a Dirichlet process with parameter $\eta(x) = \alpha P_0(x)$ since $EP(A) = P_0(A)$

BAYESIAN NONPARAMETRICS

Uncertainty in the parameter $\eta \Rightarrow \eta \in \Gamma \Rightarrow$ changes in

- Dirichlet process
  - P and Q chosen by two Dirichlet processes with different η
  - $d_{DP}(P,Q) = \sup_{A \in \mathcal{A}} d(P(A), Q(A))$
  - $d(X,Y) = \left\{ \int (\sqrt{p} - \sqrt{q})^2 \, d\mu \right\}^{1/2}$ Hellinger distance
- Probability of subsets of p.m.'s on $(\mathcal{X}, \mathcal{A})$
  - $\Theta = \{P \in \mathcal{M} : P(A) \in B\}$, $A \in \mathcal{A}$, $B \in \mathcal{B}([0,1])$ (e.g. $\Theta = \{F : F(1/2) \leq 1/2\}$)
  - $P \sim \mathcal{DP}(\eta) \Rightarrow P(A) \sim \mathcal{B}(\eta(A), \eta(A^C)) \Rightarrow$ compute $\mathcal{P}(\Theta) = \mathcal{P}(P(A) \in B)$

BAYESIAN NONPARAMETRICS

Uncertainty in the parameter $\eta \Rightarrow \eta \in \Gamma \Rightarrow$ changes in

- Probabilities of set probabilities and random functionals
  - $P(A) \sim \mathcal{B}(\eta(A), \eta(A^C))$
  - $(P(A_1), \ldots, P(A_n)) \sim \mathcal{D}(\eta(A_1), \ldots, \eta(A_n))$
  - $\displaystyle\int_{\Re} Z dP$
- Bayes estimators of random distributions and functionals
  - Bayes estimator of the mean: $\dfrac{\int_{\Re} x\eta(x)dx}{\int_{\Re} \eta(x)dx}$
  - Distribution function $F^*(x) = \dfrac{\alpha\eta(x) + \sum_1^n \delta_{Z_i}(x)}{\alpha + n}$

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: $\{y_{ij}, i = 1 \ldots k_j\}_{j=1}^{n}$ 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 $\beta \Rightarrow \widetilde{M}S = \dfrac{\beta}{1 + n\beta}[\alpha S + \sum_{j=1}^{n} \sum_{i=1}^{k_j} \mathbb{I}_S(y_{ij})]$

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 $\beta \Rightarrow \widetilde{R}S = \left(1 + \frac{\beta}{1+n\beta}\right)^{-(\alpha S + \sum_{j=1}^{n} \xi_j S)}$

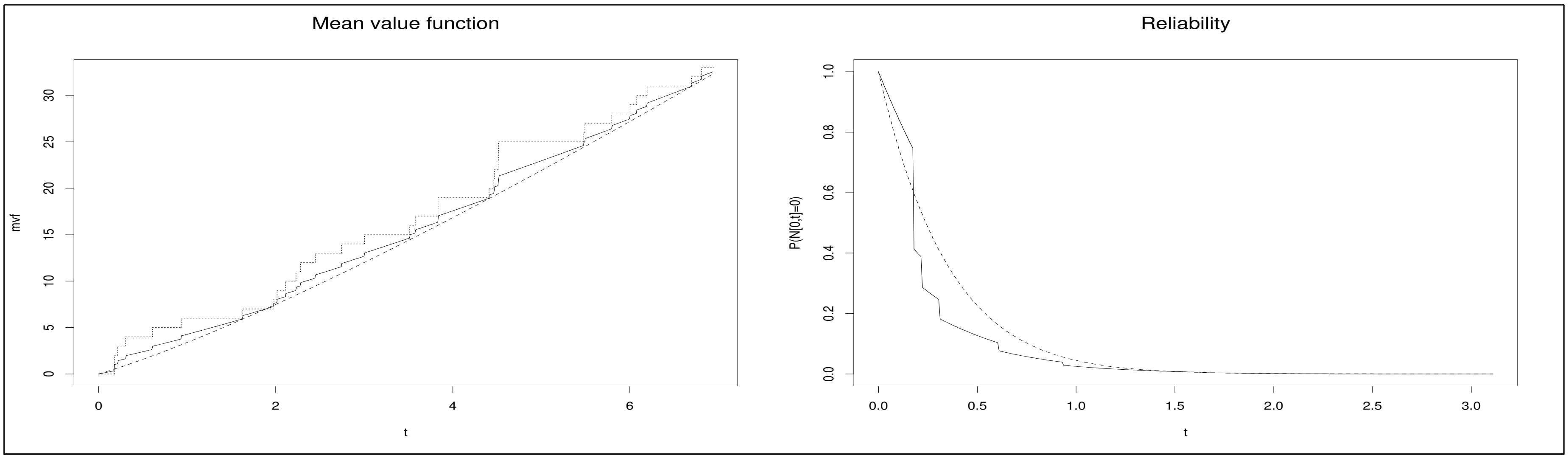
STEEL PIPES

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: $P(\underline{k} \mid H_P) = \int_{\mathbb{R}_+^3} e^{-\Lambda_\theta(T)} \prod_{j=1}^{n} \frac{[\Lambda_\theta I_j]^{k_j}}{k_j!} \pi(\theta) \, d\theta$

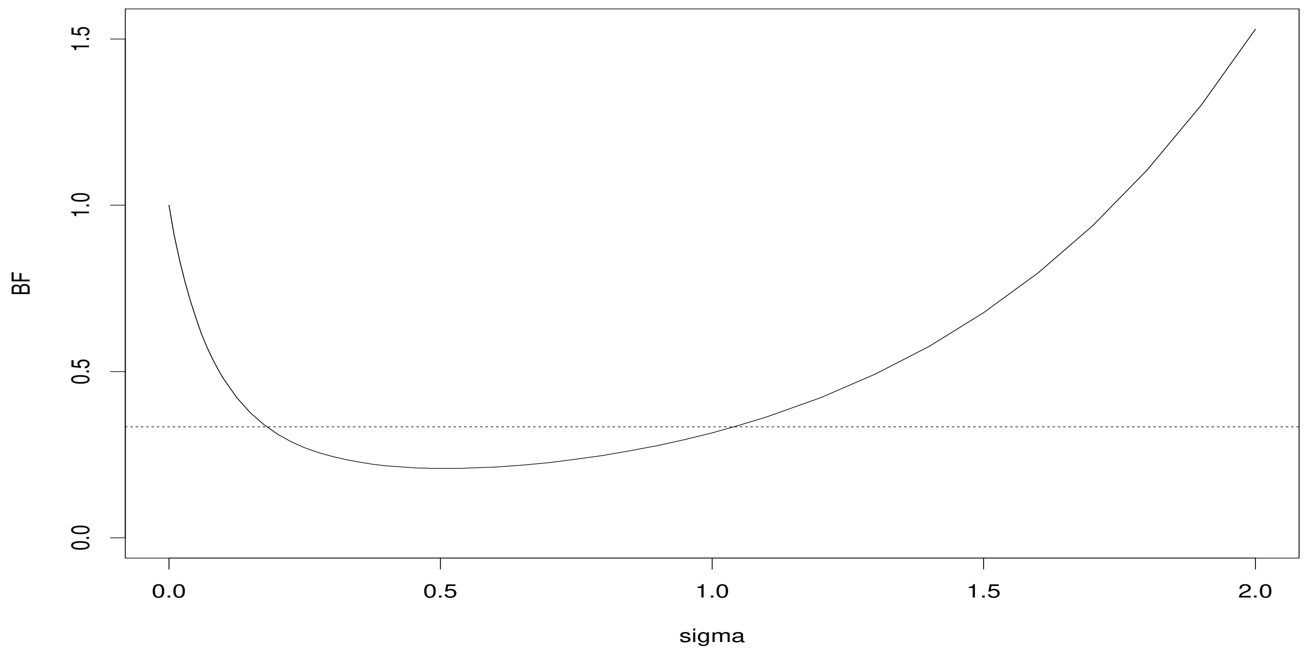
Nonparametric: $\underline{k} \mid M, \theta \sim f(\underline{k} \mid M_\theta)$, $M \mid \theta \sim \mathcal{GG}(\Lambda_\theta / \sigma, \sigma)$ and $\theta \sim \pi$:

$$P(\underline{k} \mid H_N) = \int_{\mathbb{R}_+^3} \prod_{j=1}^{n} \left[\frac{\prod_{i=0}^{k_j - 1} (\Lambda_\theta I_j + i\sigma)}{k_j! \exp\left[\left(\frac{\Lambda_\theta I_j}{\sigma} + k_j \right) \ln(1 + \sigma) \right]} \right] \pi(\theta) \, d\theta$$

Bayes Factor: $BF_{PN} = \frac{P(\underline{k} \mid H_P)}{P(\underline{k} \mid H_N)} = \frac{\int_{\mathbb{R}_+^3} e^{-\Lambda_\theta(T)} \prod_{j=1}^{n} (\Lambda_\theta I_j)^{k_j} \pi(\theta) \, d\theta}{\int_{\mathbb{R}_+^3} \prod_{j=1}^{n} \left[(1 + \sigma)^{-(\Lambda_\theta I_j / \sigma + k_j)} \prod_{i=0}^{k_j - 1} (\Lambda_\theta I_j + i\sigma) \right] \pi(\theta) \, d\theta}$

PARAMETRIC VS. NONPARAMETRIC

Bayes factor BF_{PN} as a function of σ



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, minimised by a_π

- $\rho_C = \mathsf{inf}_{a \in \mathcal{A}} \mathsf{sup}_{\pi \in \Gamma} \rho(\pi, a)$
  (Posterior Γ-minimax expected loss)

  Optimal action by interchanging inf and sup for convex losses

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

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

APPLICATIONS

- Very few applications of *these* robust Bayesian procedures
- Typically, either
  - informal analysis (a finite family of priors) or
  - choice of robust procedures (e.g. hierarchical models), robust distributions (e.g. Student) and robust estimators (e.g. median)
- Need for sensitivity checks is nowadays widely accepted within the Bayesian community
- Classes and tools often driven more by maths rather than by practice
- Lack of adequate software

APPLICATION: GAS ESCAPE

- Interest in replacement policy for pipelines more prone to gas escapes in a metropolitan distribution network at low-pressure (20 mbar over atmospheric pressure)

- Identification of the most prone material (traditional cast iron [CI]) and the most influential technical and environmental features (diameter of pipes, laying location, depth)

- Cast iron pipes not subject to corrosion $\Rightarrow$ homogeneous Poisson process (HPP)

- Two levels for each feature $\Rightarrow$ eight subnetworks modelled by independent HPP's with parameters λ_i

- Experts' opinions on propensity to gas escapes through pairwise comparison of eight subnetworks and use of AHP (Analytic Hierarchy Process)

- Bayesian (and ML) estimation of λ_i's and ranking of subnetworks according to their value

- Sensitivity analysis

FAILURES IN CAST-IRON PIPES

- CI pipes cover more than a quarter of the whole network, with about 6000 different pipe sections with homogeneous characteristics, ranging in length from 3 to 250 meters for a total of 312 kilometers

- Cast-iron not aging $\Rightarrow$ HPP in space and time 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 $s = meas(\mathcal{S})$

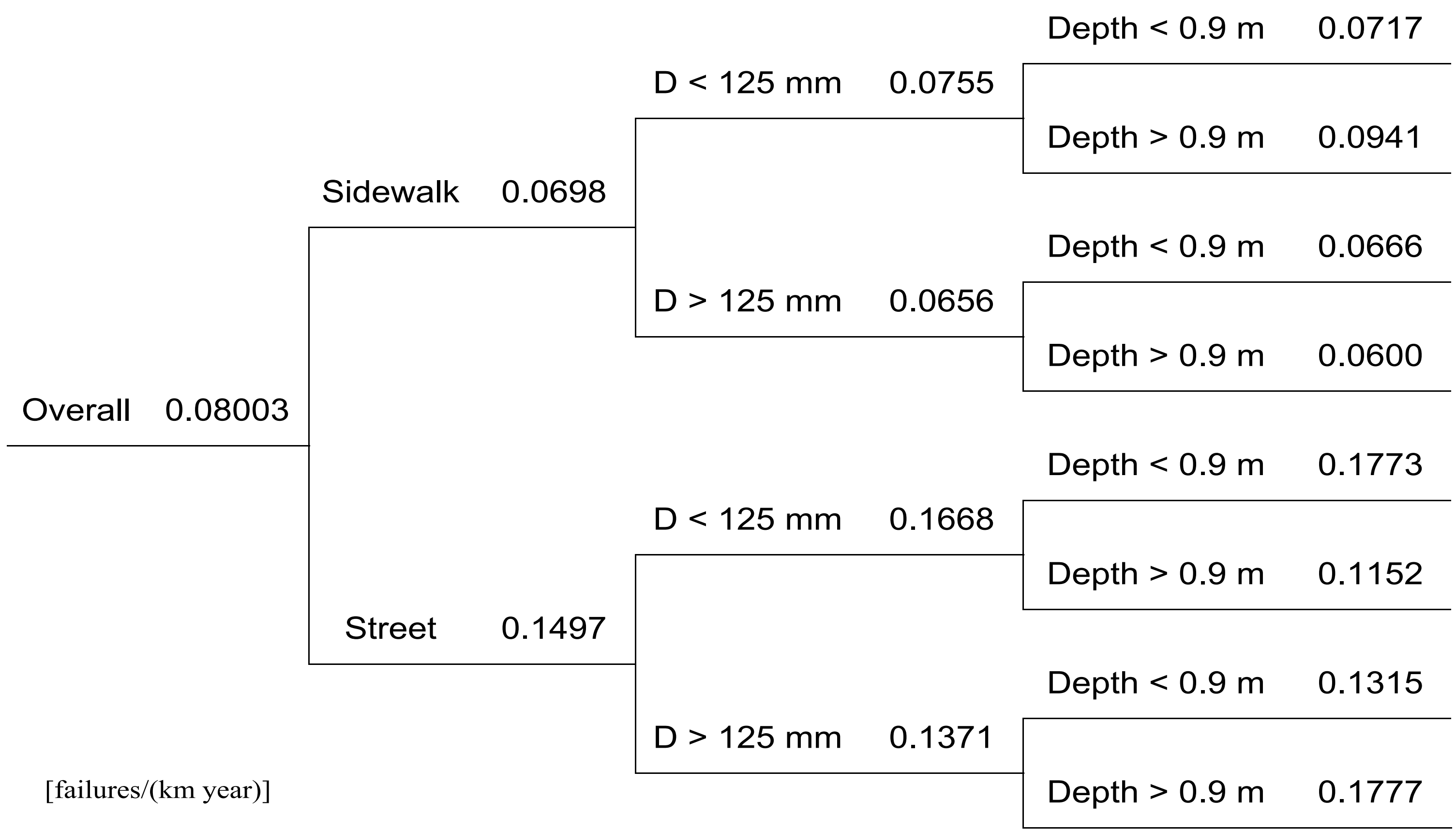
- Data: $n = 150$ failures in $T = 6$ years on a net $\approx s = 312$ Km long

  $\Rightarrow l(\lambda|n,T,\mathcal{S}) = (1872\lambda)^{150} e^{-1872\lambda}$ (if considering all failures together)

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

- Consider 8 classes determined by two levels of the relevant covariates: diameter, location and depth

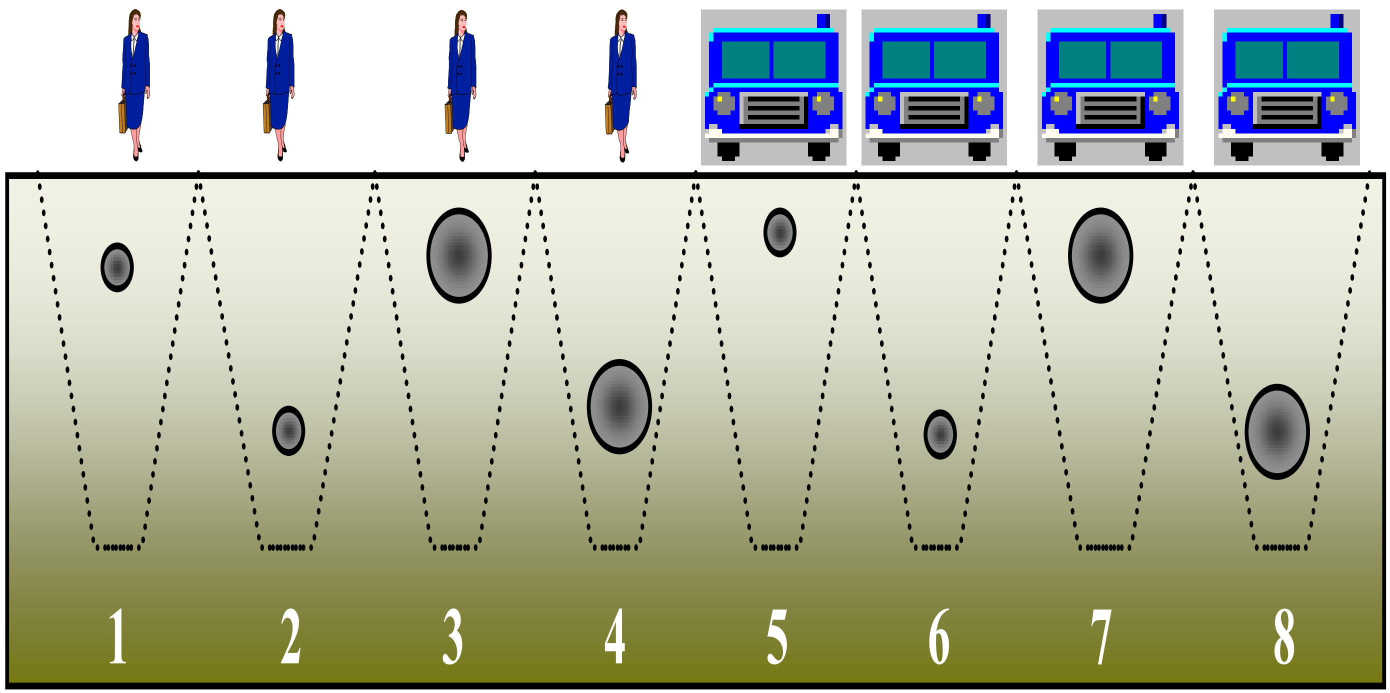
FAILURES IN CAST-IRON PIPE



ELICITATION OF EXPERTS' OPINIONS

- Importance of experts' judgements (and Bayesian approach ...)
  - relatively few data from the company
  - 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
- 26 company experts (from different areas) filled an ad hoc questionnaire based on pairwise comparisons of propensity to gas escapes in the 8 subnetworks
  - interviewees unable to say how many failures they expected to see on a kilometer of a given kind of pipe in a year or how much a factor influenced the failure
  - interviewees able to compare the performance against failure of different pipeline classes through a linguistic judgement $\Rightarrow$ transformed into numerical judgements via AHP and reported in a 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
- $\Rightarrow$ (normalized) eigenvector associated with the largest eigenvalue
- $\Rightarrow (P(A_1), \ldots, P(A_n))$
- **Question**: if a gas escape occurs, where do 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

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

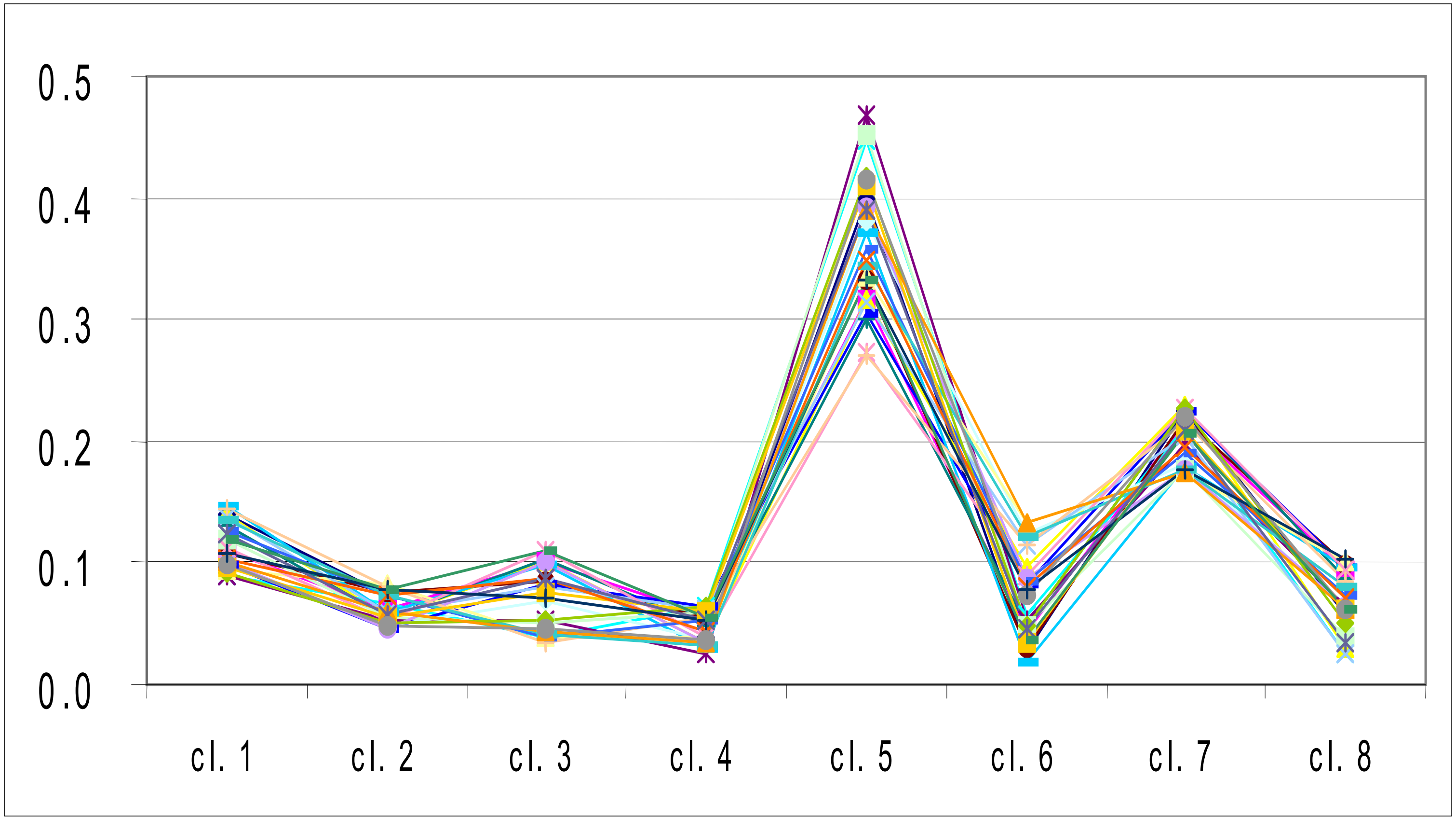
- A not consistent $\Rightarrow$ take eigenvector of $A\mathbf{w} = \lambda_{max}\mathbf{w}$, with λ_{max} largest eigenvalue (always $\lambda_{max} \geq n$ for positive reciprocal matrices and $\lambda_{max} = n$ for consistent ones)

- $\dfrac{\lambda_{max} - n}{n - 1}$ 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 $\Rightarrow$ 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 λ and allocated to class A_i with probability $P(A_i)$ $\Rightarrow$ 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 expected number of gas escapes for given period and length of network $\Rightarrow$ statements on λ, unit failure rate for entire network, 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 $\lambda_i = \widehat{\lambda} P(A_i)$

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

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 $\widehat{\underline{\beta}}$
    $\Rightarrow \lambda_i|\widehat{\underline{\beta}}, \underline{d} \sim \mathcal{G}(\alpha e^{\underline{X}_i^T \widehat{\underline{\beta}}} + y_i, \alpha + t_i)$, $\forall i$
- “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

CRITICAL REVIEW OF PAST RESULTS

- Qualitative judgements manipulated via AHP instead of assessments on observable quantities, like
  - expected number of gas escapes in a 1 Km long pipe in one year
  - median of the distribution of the time of the first gas escape in a 1 Km long pipe

  $\Rightarrow$ their use as *sample* on λ or, better, conditions determining classes of priors

- Mixed use of MLE and prior assessment

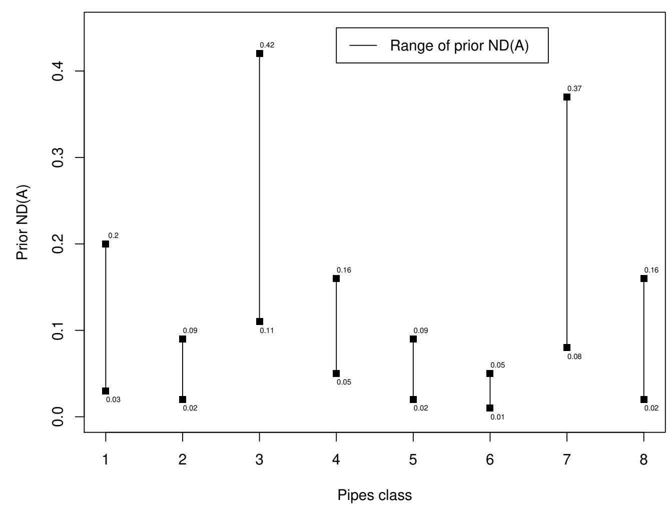
- Choice of a functional form (convenient from a mathematical viewpoint but not corresponding to what the experts think)

- Ranking based on posterior means, justified by the choice of squared loss function and not by the company's preferences

- Use (as much as possible) just actual beliefs and preferences $\Rightarrow$ classes of priors and losses $\Rightarrow$ ranking based on *adequate* actions

BAYES ACTIONS

- Range of prior opinions on λ_i's, $i = 1, 8$, by 14 experts
- Class 3 (TSN) looks worse than others but its lower bound is below upper bounds of classes 1 (WSN), 4 (TSD), 7 (TLN) and 8 (WLD)
- Large variability



BAYES ACTIONS

Quartiles determined by experts' opinions

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
min	0.03	0.02	0.11	0.05	0.02	0.01	0.08	0.02
max	0.2	0.09	0.42	0.16	0.09	0.05	0.37	0.16
$q_{0.250}$	0.090	0.040	0.290	0.060	0.040	0.020	0.120	0.040
$q_{0.500}$	0.105	0.050	0.320	0.090	0.060	0.030	0.185	0.045
$q_{0.750}$	0.120	0.080	0.350	0.130	0.080	0.040	0.230	0.060

BAYES ACTIONS

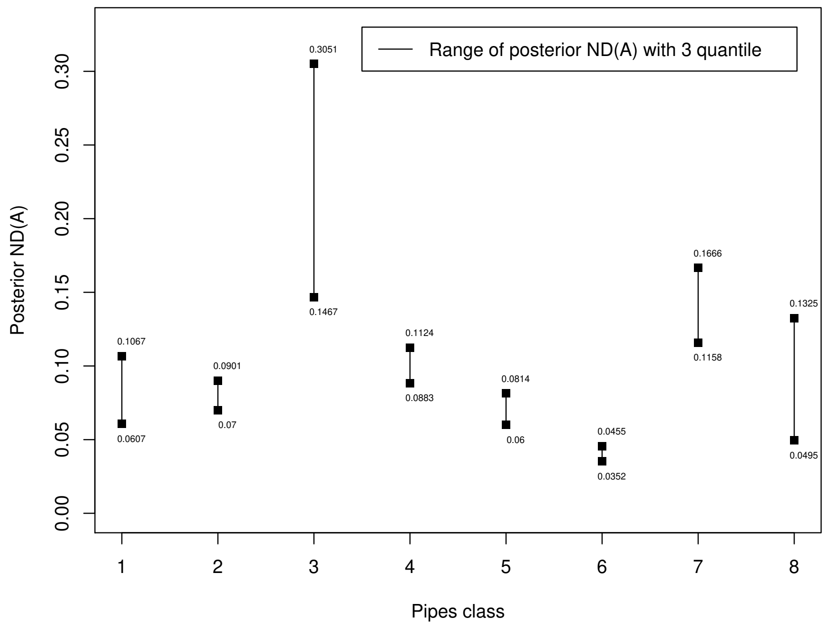
- Quartiles determine intervals I_i, $i = 1, 4$ with probability 0.25 each
- In general, quantiles assign probabilities p_i to intervals I_i, $i = 1, n$
- Class of priors $\Gamma = \{\pi : \int_{I_i} \pi(d\lambda) = p_i, i = 1, n\}$
- Interest in posterior mean $E_\pi \lambda = \dfrac{\int \lambda l(\lambda)\pi(\lambda)d\lambda}{\int l(\lambda)\pi(\lambda)d\lambda}$
- $\sup\limits_{\pi \in \Gamma} E_\pi \lambda = \sup\limits_{\lambda_i \in \bar{I}_i, i=1,n} \dfrac{\sum_{i=1}^n \lambda_i l(\lambda_i) p_i}{\sum_{i=1}^n l(\lambda_i) p_i}$
- Ranges compared for classes with 3 or 7 quantiles

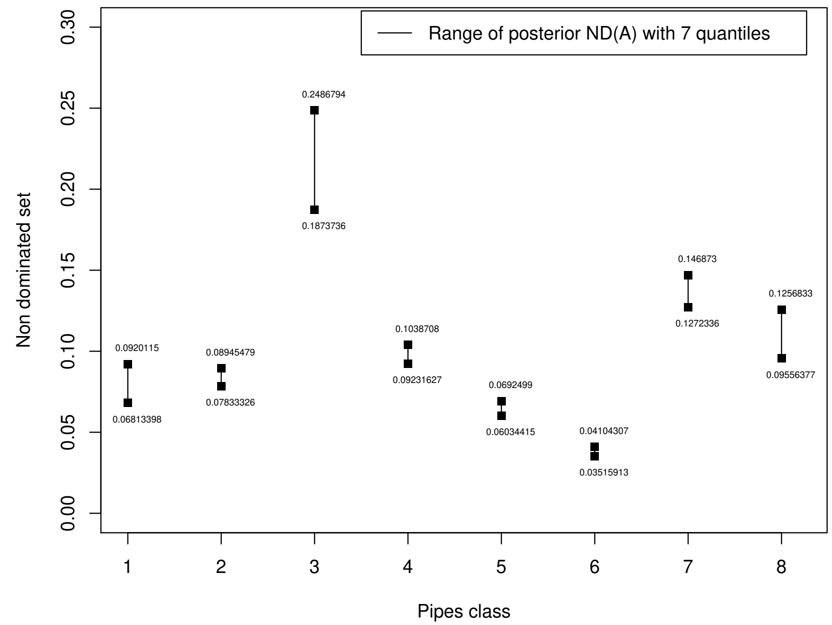
BAYES ACTIONS

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8
min	0.03	0.02	0.11	0.05	0.02	0.01	0.08	0.02
max	0.2	0.09	0.42	0.16	0.09	0.05	0.37	0.16
$q_{0.125}$	0.060	0.030	0.210	0.050	0.020	0.020	0.110	0.030
$q_{0.250}$	0.090	0.040	0.290	0.060	0.040	0.020	0.120	0.040
$q_{0.375}$	0.090	0.040	0.310	0.070	0.050	0.030	0.150	0.040
$q_{0.500}$	0.105	0.050	0.320	0.090	0.060	0.030	0.185	0.045
$q_{0.625}$	0.110	0.060	0.330	0.110	0.060	0.030	0.220	0.050
$q_{0.750}$	0.120	0.080	0.350	0.130	0.080	0.040	0.230	0.060
$q_{0.875}$	0.130	0.100	0.390	0.130	0.090	0.040	0.260	0.150

BAYES ACTIONS

Range of Bayes actions for quantile class: 3 quantiles (left) and 7 (right)



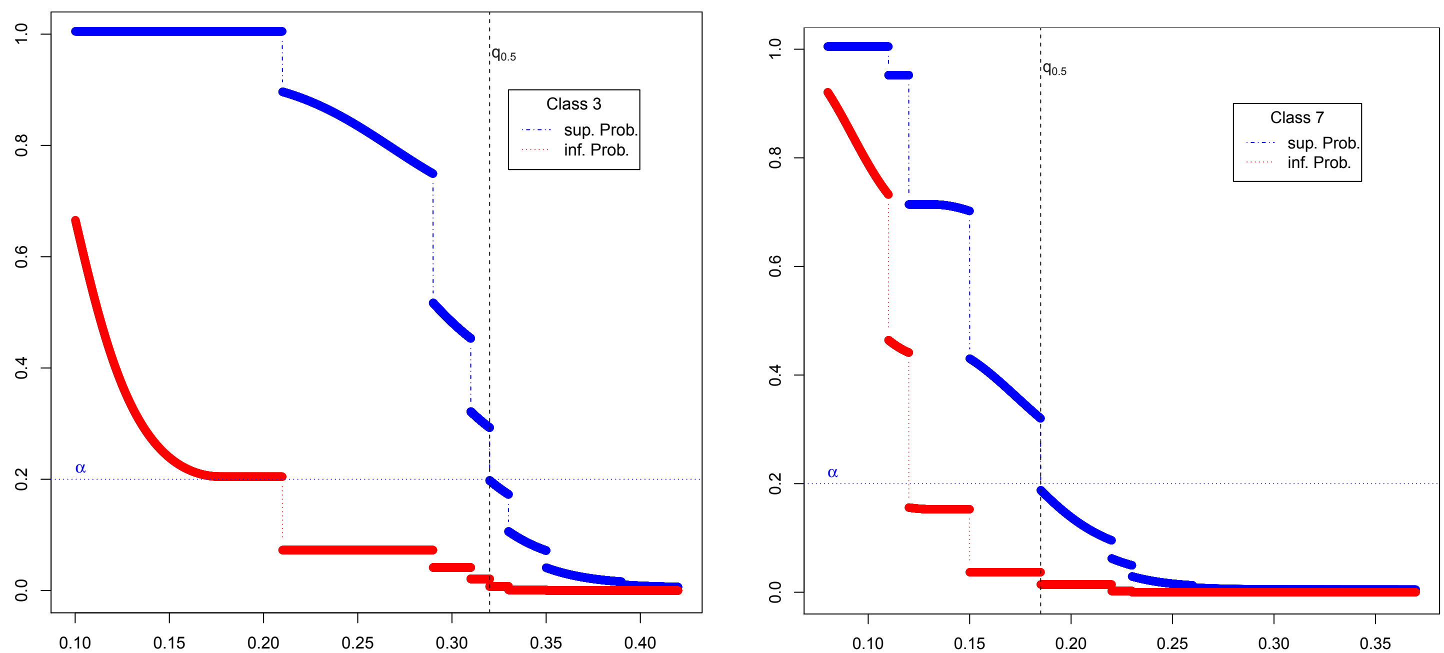


THRESHOLD EXCEEDANCE

- Interest not only in the class to be replaced first but also **if** any has to be replaced
- $\Rightarrow$ set a critical threshold $\tilde{\lambda}$ on the failure rate
- Given prior π and data $\underline{d} \Rightarrow$ interest in posterior $P_\pi(\lambda > \tilde{\lambda}|\underline{d})$
- Do not replace pipes if $P_\pi(\lambda > \tilde{\lambda}|\underline{d}) \leq \alpha$, with α related to the *acceptable risk*
- Under a class of priors Γ
  - do not replace if $\sup_{\pi \in \Gamma} P_\pi(\lambda > \tilde{\lambda}|\underline{d}) \leq \alpha$
  - replace if $\inf_{\pi \in \Gamma} P_\pi(\lambda > \tilde{\lambda}|\underline{d}) > \alpha$
  - further investigation o.w.
- Sup and Inf obtained for discrete distributions with one point for each interval (actually its closure) determined by the prior quantiles

THRESHOLD EXCEEDANCE

- Upper and lower bound on threshold exceedance probability as a function of $\tilde{\lambda}$ for class 3 (TSN - left) and class 7 (TLN - right), compared with prior median $q_{0.5}$ assessed by experts
- $\alpha = 0.2$
- For class 3: replace if $\tilde{\lambda} < 0.21$ and do not replace if $\tilde{\lambda} > 0.32$



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*, c*learly 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.

STOCHASTIC ORDERS

- *Usual* stochastic order
  - X and Y r.v.'s with d.f.'s F_X and F_Y s.t. $F_X(t) \geq F_Y(t), \quad \forall t \in \mathbb{R}$
  - $\Rightarrow X \leq_{\text{st}} Y$, i.e. X is said to be *smaller than* Y *in the usual stochastic order*
  - $X \leq_{\text{st}} Y \Leftrightarrow E[g(X)] \leq E[g(Y)]$ holds for all increasing functions g for which the expectations exist

- Likelihood ratio order
  - X and Y be (discrete) absolutely continuous r.v.'s with d.f.'s F_X and F_Y and (discrete) densities f_X and f_Y s.t. $\dfrac{f_Y(t)}{f_X(t)}$ increases over the union of the supports of X and Y (here $a/0$ is taken to be equal to ∞ whenever $a > 0$)
  - $\Rightarrow X \leq_{\text{lr}} Y$, i.e. X is said to be *smaller than* Y *in the likelihood ratio order*

- $X \leq_{lr} Y \Rightarrow X \leq_{st} Y$

DISTORTION FUNCTIONS

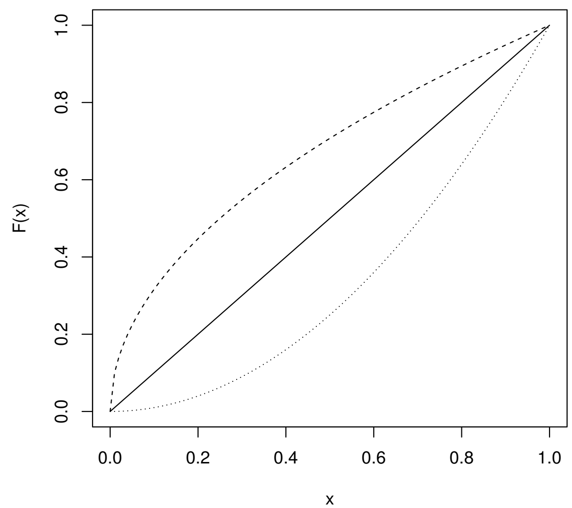
- X r.v. with d.f. F_X
- h distortion function
  - non-decreasing continuous function $h : [0,1] \to [0,1]$
  - s.t. $h(0) = 0$ and $h(1) = 1$
- Given h, cumulative probability modified by

  $F_h(x) = h \circ F(x) = h[F(x)]$
- $\Rightarrow X_h$ distorted r.v. with d.f. $F_h(x)$
- Distortion functions used to build classes of priors, with stochastic order properties

SOME RESULTS

- Prior distribution π with d.f. $F_\pi(\theta)$ and distortion function h
- $\Rightarrow$ distorted prior distribution π_h with d.f. $F_{\pi_h}(\theta) = h \circ F_\pi(\theta) = h\left[F_\pi\left(\theta\right)\right]$
- **Lemma.**
  - π prior distribution (absolutely continuous or discrete) with d.f. F_π
  - h convex distortion function in $[0, 1] \Rightarrow \pi \leq_{lr} \pi_h$
  - h concave distortion function in $[0, 1] \Rightarrow \pi \geq_{lr} \pi_h$
- Important result for the construction of classes of priors through stochastic ordering

CONCAVE AND CONVEX DISTORTION FUNCTIONS



- Solid: $F_\pi(\theta) = \theta$
- Dashed: $F_{\pi_{h_1}}(\theta) = \sqrt{\theta}$ (concave distortion)
  $\Rightarrow$ decreasing l.r. $= 1/(2\sqrt{\theta}) \Rightarrow \pi \geq_{lr} \pi_{h_1}$
- Dotted: $F_{\pi_{h_2}}(\theta) = \theta^2$ (convex distortion)
  $\Rightarrow$ increasing l.r. $= 2\theta \Rightarrow \pi \leq_{lr} \pi_{h_2}$

DISTORTED BAND OF PRIORS

- Uncertainty on prior π through concave (h_1) and convex (h_2) distortion functions
- **Previous Lemma.** Prior π and convex (or concave) distortion function h in $[0, 1]$
  $\Rightarrow \pi \leq_{lr} \pi_h$ (or $\pi \geq_{lr} \pi_h$)
- Lemma $\Rightarrow$ distorted distributions π_{h_1} and π_{h_2} s.t. $\pi_{h_1} \leq_{lr} \pi \leq_{lr} \pi_{h_2}$
- **Definition.** Distorted band $\Gamma_{h_1,h_2,\pi}$ s.t. $\Gamma_{h_1,h_2,\pi} = \{\pi' : \pi_{h_1} \leq_{lr} \pi' \leq_{lr} \pi_{h_2}\}$
- Lemma $\Rightarrow \pi \in \Gamma_{h_1,h_2,\pi}$
- $\Rightarrow$ distorted band as a particular "neighborhood" band of π, with lower and upper bound given by distorted distributions
- Band defined only through an upper (or lower) bound when considering h_1 (or h_2) the identity function

CHOICES OF DISTORTION FUNCTIONS

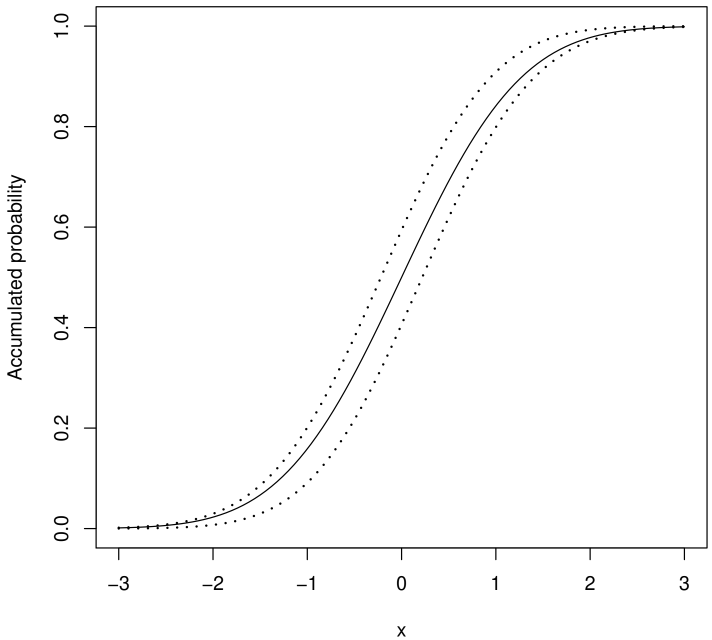
- $h_1(x) = 1 - (1-x)^\alpha$ and $h_2(x) = x^\alpha, \quad \forall \alpha > 1$
  - $\alpha = n \in \mathbb{N} \Rightarrow F_{\pi_{h_1}}(\theta) = 1 - (1 - F_\pi(\theta))^n$ and $F_{\pi_{h_2}}(\theta) = (F_\pi(\theta))^n$
  - $\Rightarrow$ d.f.'s of min and max of i.i.d. random sample of size n from baseline prior π

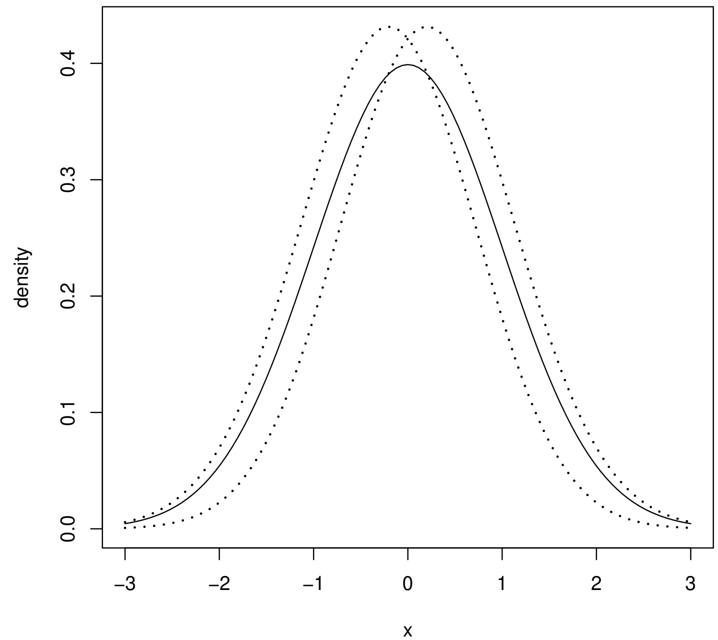
- $h_1(x) = \min\{\frac{x}{\alpha}, 1\}$ and $h_2(x) = \max\{\frac{x-\alpha}{1-\alpha}, 0\}, \quad 0 < \alpha < 1$
  - $\Rightarrow$ truncated distributions $\pi_{h_1} =_{\mathcal{L}} \pi(\cdot|A_1)$ and $\pi_{h_2} =_{\mathcal{L}} \pi(\cdot|A_2)$
    - $=_{\mathcal{L}}$ means equality in law
    - $A_1 = (-\infty, F_\pi^{-1}(\alpha)]$
    - $A_2 = (F_\pi^{-1}(\alpha), \infty)$
  - (π_{h_1} (π_{h_2}) concentrated up to (after) α–quantile of π)

CHOICES OF DISTORTION FUNCTIONS

- Skewed distributions
- π absolutely continuous, symmetric around 0 prior with density $\pi(\theta)$ and d.f. $F_\pi(\theta)$
- $\Rightarrow$ skew-π with parameter α with density $\pi_\alpha(\theta) = 2\pi(\theta)F_\pi(\alpha\theta)$
- Distribution: right skewed if $\alpha > 0$ and left skewed if $\alpha < 0$
- Easy to show $\pi \leq_{lr} \pi_\alpha$ for all $\alpha > 0$ and $\pi_\alpha \leq_{lr} \pi$ for all $\alpha < 0$

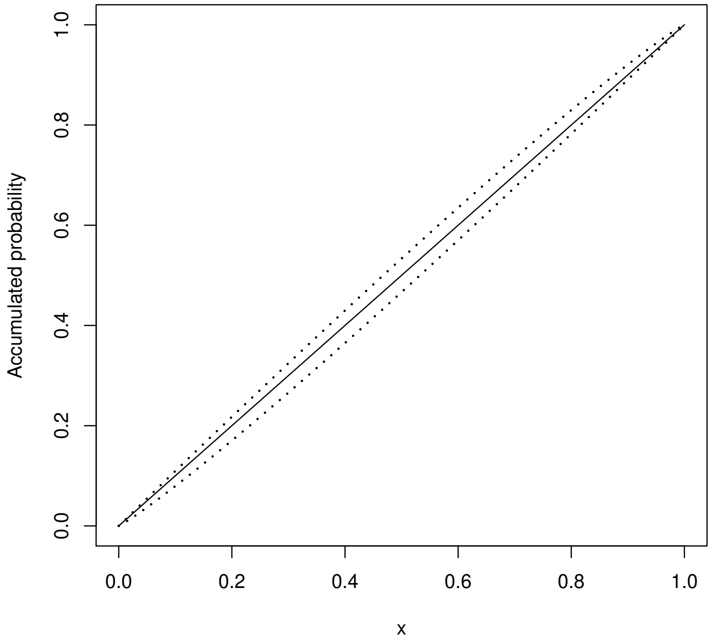
CHOICES OF DISTORTION FUNCTIONS

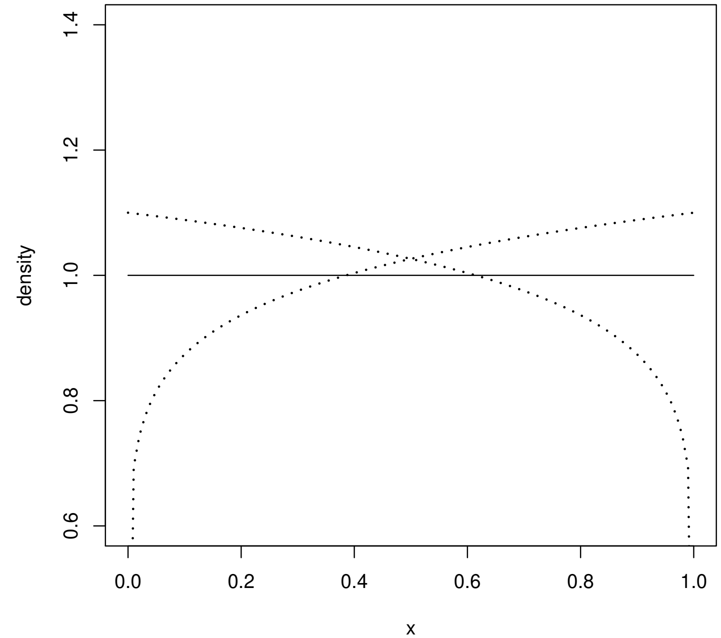




- $\pi \sim N(0,1)$ prior with standard normal d.f. Φ_Z
- Distorted d.f.'s $F_{\pi_{h_1}}(\theta) = 1 - (1 - \Phi_Z(\theta))^{1.3}$ and $F_{\pi_{h_2}}(\theta) = (\Phi_Z(\theta))^{1.3}$

CHOICES OF DISTORTION FUNCTIONS





- $\pi \sim U(0,1)$ prior with d.f. Φ_Z
- Distorted d.f.'s $F_{\pi_{h_1}}(\theta) = 1 - (1 - \Phi_Z(\theta))^{1.1}$ and $F_{\pi_{h_2}}(\theta) = (\Phi_Z(\theta))^{1.1}$

POSTERIOR BAND

- Spizzichino (2001): given two priors π_1 and π_2 s.t. $\pi_1 \leq_{lr} \pi_2$

  $\Rightarrow$ posteriors s.t. $\pi_{1x} \leq_{lr} \pi_{2x}$

- **Proposition.** π prior and $\Gamma_{h_1,h_2,\pi}$ distorted band around π based on h_1 and h_2

  $\Rightarrow \pi_{h_1,x} \leq_{lr} \pi'_x \leq_{lr} \pi_{h_2,x} \; \forall \pi' \in \Gamma_{h_1,h_2,\pi}$

- Posterior of lower and upper bound distributions of the distribution band $\Rightarrow$ lower and upper bounds in the $\leq_{lr}$ order sense for Γ_x, family of posterior distributions

- $\Rightarrow \Gamma_x$ still distortion band of a posterior for some concave and convex functions

- *Closure* property very uncommon among classes of priors

  $\Rightarrow$ dealing with priors or posteriors is the same

OTHER WORKS

- Band of distorted priors used in Fault Tree Analysis
- Band of multivariate distorted priors
- Classes of Dirichlet processes
- Robustness in Adversarial Risk Analysis
- Robust Bayesian Analysis for Generalized Extreme Value Models