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
www.mi.imati.cnr.it/~fabrizio/

BASICS ON BAYESIAN STATISTICS

- X r.v. with density $f(x|\theta)$
- Prior $\pi(\theta)$
- Sample $\underline{X} = (X_1, \ldots, X_n)$
- Bayes theorem $\Rightarrow$ posterior $\pi(\theta|\underline{X})$
- Loss function $L(\theta, a)$, e.g. $(\theta - a)^2$
- Minimize $\mathcal{E}^{\pi(\theta|\underline{X})} L(\theta, a) \Rightarrow$ Bayes estimator, e.g. posterior mean for $(\theta - a)^2$

Is life so easy?

EXERCISE 1

- Car tyres failures
- $X_1, \ldots, X_n$ lifetimes
- How to perform a Bayesian analysis?

EXERCISE 1

Bayesian analysis - what to choose?

- Model $f(x|\theta)$
- Prior $\pi(\theta)$
- Estimator $\widehat{\theta}$

EXERCISE 1 - MODEL SELECTION

Before the analysis - Model chosen according to

- physical laws
- mathematical convenience
- exploratory data analysis
- ...

EXERCISE 1 - MODEL SELECTION

After the analysis - Model chosen according to

- graphical displays (e.g. residuals in regression)
- goodness of fit tests (e.g. χ^2, Kolmogorov-Smirnov) (*not very Bayesian!*)
- Bayes factor to compare
  $\mathcal{M}_1 = \{f_1(x|\theta_1), \pi(\theta_1)\}$ and $\mathcal{M}_2 = \{f_2(x|\theta_2), \pi(\theta_2)\}$

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

- Posterior odds

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

- AIC, BIC, DIC et al.

EXERCISE 1 - MODEL SELECTION

Replacement policy

- New tyre replaced after each failure
  - *Good as new*
  - $X_1, \ldots, X_n$ i.i.d.
  - Renewal process
- Old tyre fixed after each failure
  - *Bad as old*
  - $X_1, \ldots, X_n$ from nonhomogeneous Poisson process

EXERCISE 1 - MODEL SELECTION

Renewal process - model choice

- $X_i \sim \mathcal{E}(\lambda) \Rightarrow f(x|\lambda) = \lambda \,\mathsf{exp}\{-\lambda x\}$
- $X_i \sim \mathcal{G}(\alpha, \beta) \Rightarrow f(x|\alpha, \beta) = \beta^\alpha x^{\alpha-1} \,\mathsf{exp}\{-\beta x\}/\Gamma(\alpha)$
- $X_i \sim \mathcal{LN}(\mu, \sigma^2) \Rightarrow f(x|\mu, \sigma^2) = \{x\sigma\sqrt{2\Pi}\}^{-1} \,\mathsf{exp}\{-(\log x - \mu)^2/(2\sigma^2)\}$
- $X_i \sim \mathcal{GEV}(\mu, \sigma, \lambda) \Rightarrow f(x) = \frac{1}{\sigma}\left[1 + \lambda\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\lambda-1} \mathsf{exp}\left\{-\left[1 + \lambda\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\lambda}\right\}$
- $\ldots$

EXERCISE 1 - MODEL SELECTION

Poisson process - model choice

- $N_t, t \geq 0$ # events by time t

- $N(y, s)$ # events in $(y, s]$

- $\Lambda(t) = \mathcal{E} N_t$ mean value function

- $\Lambda(y, s) = \Lambda(s) - \Lambda(y)$ expected # events in $(y, s]$

EXERCISE 1 - MODEL SELECTION

Poisson process - model choice

$N_t, t \geq 0$, NHPP with intensity function $\lambda(t)$ *iff*

1. $N_0 = 0$
2. independent increments
3. $\mathcal{P}\{\#$ events in $(t, t+h) \geq 2\} = o(h)$
4. $\mathcal{P}\{\#$ events in $(t, t+h) = 1\} = \lambda(t)h + o(h)$

$$\Rightarrow \mathcal{P}\{N(y,s) = k\} = \frac{\Lambda(y,s)^k}{k!} e^{-\Lambda(y,s)}, \forall k \in \mathcal{N}$$

EXERCISE 1 - MODEL SELECTION

Poisson process - model choice

$\lambda(t) \equiv \lambda \; \forall t \Rightarrow$ HPP

- $\lambda(t)$: intensity function of N_t

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

- $\mu(t) := \frac{d\Lambda(t)}{dt}$: Rocof (rate of occurrence of failures)

Property 3. $\Rightarrow \mu(t) = \lambda(t)$ a.e. $\Rightarrow \Lambda(y, s) = \int_y^s \lambda(t) dt$

EXERCISE 1 - MODEL SELECTION

How to choose NHPP?

- Musa-Okumoto
  $\lambda(t;\alpha,\beta) = \alpha/(t+\beta)$ and $\Lambda(t;\alpha,\beta) = \alpha \log(t+\beta)$

- Cox-Lewis
  $\lambda(t;\alpha,\beta) = \alpha \exp\{\beta t\}$ and $\Lambda(t;\alpha,\beta) = (\alpha/\beta)\,[\exp\{\beta t\} - 1]$

- Power law
  $\lambda(t;\alpha,\beta) = \alpha\beta t^{\beta-1}$ and $\Lambda(t;\alpha,\beta) = \alpha t^{\beta}$

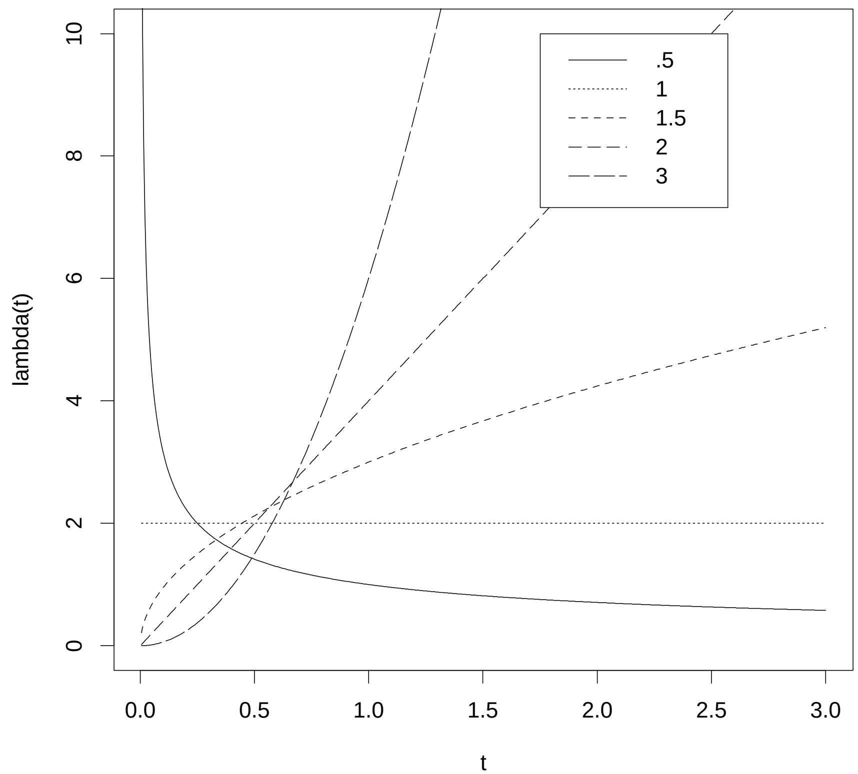
- ...

EXERCISE 1 - MODEL SELECTION

How to choose NHPP?

- $\lim_{t\to\infty} \Lambda(t)$
- $\lim_{t\to 0} \lambda(t)$
- Bounded $\lambda(t)$
- Monotonicity
- Maximum of $\lambda(t)$

EXERCISE 1 - MODEL SELECTION



Some Beta's and M=2

EXERCISE 1 - MODEL SELECTION

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

- $\Rightarrow$ renewal process and HPP

Which prior on λ?

EXERCISE 1 - PRIOR CHOICE

Where to start from?

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

$\Rightarrow$ *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)

EXERCISE 1 - 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 of λ, i.e. $\mathbf{E}\lambda^k$
- 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

EXERCISE 1 - PRIOR CHOICE

How to get information?

- Results from previous experiments (e.g. 75% of car tyres had failed after 5 years of operation $\Rightarrow$ 5 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)
- ...

EXERCISE 1 - PRIOR CHOICE

How to combine information?

Combining opinions of n experts

- Individual analyses and comparison a posteriori

- Opinions as *sample* from the parameter distribution
  $\Rightarrow$ sample mean and sample variance

  - Statements on quantiles $G_q \leftrightarrow \theta$

  - Statements on value of θ

EXERCISE 1 - PRIOR CHOICE

How to use information?

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

  (*empirical Bayes*)
- use information to choose parameters of a random distribution on the space of probability measures

  (*Bayesian nonparametrics*)
- use Jeffreys'/reference/improper priors

  (*objective Bayes*)
- use a class of priors

  (*Bayesian robustness*)

EXERCISE 1 - PRIOR CHOICE

Choice of a prior

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

EXERCISE 1 - 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)

EXERCISE 1 - PRIOR CHOICE

Gamma prior - choice of hyperparameters

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

EXERCISE 1 - PRIOR CHOICE

Gamma prior - choice of hyperparameters

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

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

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

EXERCISE 1 - PARAMETER ESTIMATION

How to estimate λ?

- MAP (Maximum a posteriori)
  $\Rightarrow \hat{\lambda} = \dfrac{\alpha + n - 1}{\beta + \sum X_i}$

- LPM (Largest posterior mode)
  $\Rightarrow$ here it coincides with MAP (unique posterior mode)

- Minimum expected loss $\mathcal{E}L(\lambda, a)$
  - $L(\lambda, a) = (\lambda - a)^2$
    $\Rightarrow \mathcal{E}\lambda | data = \dfrac{\alpha + n}{\beta + \sum X_i}$ (posterior mean)
  - $L(\lambda, a) = |\lambda - a|$
    $\Rightarrow$ (posterior median)
  - other $L(\lambda, a)$

EXERCISE 1 - CONCLUSIONS

(Bayesian) inference is often the result of many approximations and arbitrary assumptions

- Awareness of it

- Development of *safer* procedures

- $\Rightarrow$ Bayesian robustness is one of them

EXERCISE 1 - CONCLUSIONS

Prior influence

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

EXERCISE 1 - CONCLUSIONS

Influence of prior choice (Berger, 1985)

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

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

- Posterior median w.r.t. posterior mean

CONCERNS ON BAYES

Motivations for Bayesian robustness

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

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

NEED FOR BAYESIAN ROBUSTNESS

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

BAYESIAN ROBUSTNESS

Mathematical tools and *philosophical* approach

- to model uncertainty through classes of priors/models/losses
- to measure uncertainty and its effect
- to avoid arbitrary assumptions
- to favour acceptance of Bayesian approach

BAYESIAN ROBUSTNESS

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

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

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

BAYESIAN ROBUSTNESS

A more formal statement about model and prior sensitivity

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

Bayesian robustness studies these changes

ROBUST BAYESIAN ANALYSIS

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}

EXERCISE 2 - CLASSES OF PRIORS

Specify 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)*

EXERCISE 2 - CLASSES OF PRIORS

Specify reasonable classes of priors

- $\Gamma_P = \{P : p(\theta; \omega), \omega \in \Omega\}$ *(Parametric class)*
- $\Gamma_Q = \{P : \alpha_i \leq P(I_i) \leq \beta_i, i = 1, \ldots, m\}$ *(Quantile class)*
- $\Gamma_{QU} = \{P \in \Gamma_Q, \text{ unimodal}\}$ *(Unimodal quantile class)*
- $\Gamma_{GM} = \{P : \int h_i(\theta) dP(\theta) = 0, i = 1, \ldots, m\}$ *(Generalised moments class)*
- $\Gamma^{DR} = \{P : L(\theta) \leq \alpha p(\theta) \leq U(\theta), \alpha > 0\}$ *(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)*

EXERCISE 2 - CLASSES OF PRIORS

Specify reasonable classes of priors

Neighborhood classes

- $\Gamma_\varepsilon = \{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}\}$ *(ε–contaminations)*
- $\Gamma_\varepsilon^T = \{P : \sup_{A \in \mathcal{F}} |P(A) - P_0(A)| \leq \varepsilon\}$ *(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)*

Classes driven more by mathematical convenience rather than ease of elicitation

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 : $|P(A) - P_0(A)| \le \varepsilon, \forall A \in \mathcal{F}$
   $\Rightarrow P(E) \le 11\frac{\varepsilon}{10}$

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

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

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

CONCENTRATION FUNCTION CLASS

- g monotone nondecreasing, continuous, convex function s.t. $g(0) = 0$ and $g(1) \leq 1$
- $K_g = \{P : P(A) \geq g(P_0(A)) \ \ \forall A \in \mathcal{F}\}$, g-neighbourhood of a nonatomic P_0
- $P \in K_g \Rightarrow g(P_0(A)) \leq P(A) \leq 1 - g(1 - P_0(A))$
- $\{K_g\}$ 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
- The concentration function compares 2 probability measures, extending the Lorenz curve comparing discrete and uniform distributions
- $K_g = \{P : \varphi_P(x) \geq g(x), \forall x \in [0, 1]\}$

CONCENTRATION FUNCTION CLASS

Lorenz curve

- n individuals with wealth $x_i \Rightarrow x_{(1)}, \ldots, x_{(n)}$

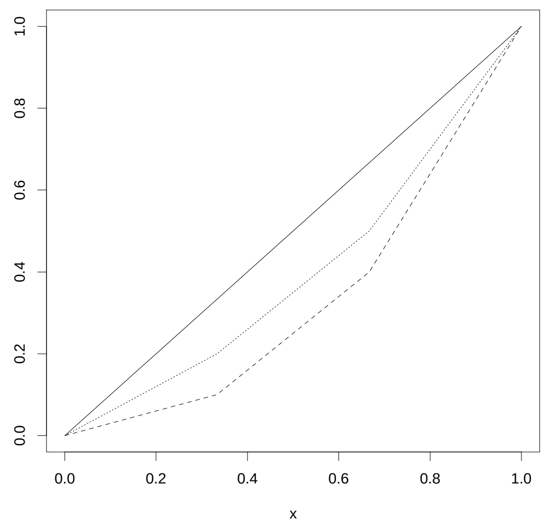
- $(k/n, S_k/S_n), k = 0, \ldots, n,$ $S_0 = 0$ and $S_k = \sum_{i=1}^{k} x_{(i)}$

- Uniformly distributed wealth $\Rightarrow$ straight line

CONCENTRATION FUNCTION CLASS

Lorenz curve

Example: $(0.2, 0.3, 0.5)$ vs. $(0.1, 0.3, 0.6)$



OBSERVABLE QUANTITIES

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

CLASSES OF MODELS

Finite classes (Shyamalkumar, 2000)

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

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

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

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

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

CLASSES OF MODELS

Finite classes (Shyamalkumar, 2000)

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

CLASSES OF MODELS

Parametric models

Box-Tiao, 1962

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

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

CLASSES OF MODELS

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)

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

Critical aspects: parameter and probabilistic interpretation

CLASSES OF MODELS

- Class of NHPPs N_t, $t \geq 0$

- Intensity function $\lambda(t)$

- Mean value function $M(t) = \mathcal{E}N_t = \int_0^t \lambda(u)du$

- $[M(t)]' = \dfrac{\alpha M(t) + \beta t}{\gamma + \delta t}$

CLASSES OF MODELS

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

CLASSES OF 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 $\Delta_1 > 0$
  - $|\alpha| \approx 0 \Rightarrow L(\Delta_1) \approx \sigma^2 \Delta_1^2/2$ (i.e. squared loss)
- $\Delta_2 = (a/\theta - 1)$ (Basu and Ebrahimi, 1991)

CLASSES OF LOSSES

Example for $L(a, \theta) = \exp\{\alpha(a/\theta - 1)\} - \alpha(a/\theta - 1) - 1, \alpha \neq 0$

Estimate the mean failure time (in hours) of a freeze seal gate valve when 20 valves are tested until 5-th failure (Martz and Waller, Basu and Ibrahimi, F.R.)

- $f(x|\theta) = (1/\theta) \exp\{-x/\theta\}$
- $0.5 \leq \alpha \leq 2.5$
- $\pi_1(\theta) = 1/\theta \Rightarrow 21808.6 \leq \mathcal{E}(\theta|data) \leq 25585.8$
- $\pi_2(\theta)\ \mathcal{IG}(a, b), a = 8.5, b = 286000 \Rightarrow 28253.1 \leq \mathcal{E}(\theta|data) \leq 30234.3$

CLASSES OF LOSSES

- $\mathcal{L}_U = \{L : L(\theta, a) = L(|\theta - a|), L(\cdot)$ 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)
  - $\{C_1, \ldots, C_n\}$ 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
- $\underline{a} = \inf_{L_\lambda \in \Psi} a_{L_\lambda}$, $\overline{a} = \sup_{L_\lambda \in \Psi} a_{L_\lambda} \Rightarrow \underline{a} \leq a_L \leq \overline{a}$, $\forall L \in \Omega$
  - $L_\lambda(\theta, a) = |\theta - a|^\lambda$, $\lambda \geq 1$
  - $L_\lambda(\theta, a) = e^{\lambda(a-\theta)} - \lambda(a - \theta) - 1$, $, \lambda_1 \leq \lambda \leq \lambda_2$
  - $L_\lambda(\theta, a) = \chi_{[a-\lambda, a+\lambda]^C}(\theta)$, $\lambda > 0$

CLASSES OF LOSSES

Mixtures of convex loss functions - examples

- $L_\lambda(\theta, a) = |\theta - a|^\lambda,\ \lambda \geq 1$
  - $\Pi \in \Gamma = \{$All symmetric probability measures w.r.t.$\mu\}$
  - $\Rightarrow a_L = \mu, \forall L \in \Omega, \forall \Pi \in \Gamma$
- $L_\lambda(\theta, a) = \chi_{[a-\lambda, a+\lambda]^C}(\theta),\ \lambda > 0$
  - $\Rightarrow \mathcal{E}L_\lambda = 1 - \Pi([a-\lambda, a+\lambda])$
  - $\Rightarrow a_{L_\lambda}$ midpoint of interval of size 2λ with the highest probability
  - $\Pi \sim \mathcal{B}eta(3,2) \Rightarrow \underline{a} = 1/2,\ , \overline{a} = 2/3$

CLASSES OF LOSSES

Bands of convex loss functions

- $\Lambda(\theta, a) = \Lambda(\theta - a) : \Lambda'(t) = \lambda(t)$
- $\lambda(t) < 0$ for $t < 0, \lambda(0) = 0, \lambda(t) > 0$ for $t > 0$
- $\lambda'(t) > 0$
- L, U losses: $L'(t) = l(t)$ and $U'(t) = u(t)$
- $\Omega = \{\Lambda : l(t) \leq \lambda(t) \leq u(t),\ \forall t\}$
- Π probability measure: $\Pi(A) > 0$ for any interval A
- $L_1, L_2 : L_1'(t) \leq L_2'(t) \Rightarrow$ Bayes actions: $a_{L_1} \leq a_{L_2}$
- $\underline{a} = \inf_{\Lambda \in \Omega} a_\Lambda,\ \overline{a} = \sup_{\Lambda \in \Omega} a_\Lambda \Rightarrow \underline{a} = a_L,\ \overline{a} = a_U$

CLASSES OF LOSSES

Bands of convex loss functions

- $l(t) = \begin{cases} 3t & t < 0 \\ t & t \geq 0 \end{cases}$

- $u(t) = \begin{cases} t & t < 0 \\ 3t & t \geq 0 \end{cases}$

- $\Omega = \{\Lambda : 1/2(\theta - a)^2 \leq \Lambda(\theta, a) \leq 3/2(\theta - a)^2\}$

- $\Lambda(\theta, a) = (\theta - a)^2 \in \Omega$

- $\Pi \sim \mathcal{N}(0, 1) \Rightarrow \underline{a} = -.3989, \; \overline{a} = .3989$

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

NON-DOMINATED ACTIONS

Foundations (Giron and Rios, 1980)

- Associate $a \rightarrow L(\theta, a)$, $\theta \in \boldsymbol{\Theta}$
- $\mathcal{D} = \{h \mid \exists a \in \mathcal{A}, h(\theta) = L(a, \theta), \forall \theta \in \theta\}$
- Preferences $\preceq$ are established over these functions

NON-DOMINATED ACTIONS

Foundations (Giron and Rios, 1980)

$(\mathcal{D}, \preceq)$ satisfies the following conditions

- $(\mathcal{D}, \preceq)$ is a quasi order (reflexive and transitive)
- If $L(a,\theta) < L(b,\theta), \forall\theta \in \Theta$, then $b \prec a$
- For $a, b, c \in \mathcal{A}$, $\lambda \in (0,1)$, then $L(a,\theta) \leq L(b,\theta)$ if and only if $\lambda L(a,\theta) + (1-\lambda)L(c,\theta) \preceq \lambda L(b,\theta) + (1-\lambda)L(c,\theta)$
- For $f_n, g, h \in \mathcal{D}$, if $f_n \to f$ and $f_n \preceq g$, $h \preceq f_n$, $\forall n$, then $f \preceq g$, $h \preceq f$

$\Rightarrow \exists \Gamma = \{\pi : \pi(\theta), \theta \in \Theta\}$ s.t.

$a \preceq b \iff \int L(a,\theta)\pi(\theta)d\theta \geq \int L(b,\theta)\pi(\theta)d\theta, \forall \pi(\cdot) \in \Gamma$

NON-DOMINATED ACTIONS

Foundations (Giron and Rios, 1980)

- Provide a qualitative framework for sensitivity analysis in Statistical Decision Theory

- non-dominated actions as basic computational objective in sensitivity analysis, when interested in decision theoretic problems

NON-DOMINATED ACTIONS

$\rho_L(\pi, x, a) = \mathcal{E}^{\pi(\cdot|x)} L(\theta, a)$

- $a, b \in \mathcal{A}$ actions

- $b \preceq a \iff \rho_L(\pi, x, a) \leq \rho_L(\pi, x, b), \forall L \in \mathcal{L}, \forall \pi \in \Gamma$

  (Action b at most as preferred as a)

- Strict inequality for some L and/or $\pi \Rightarrow b \prec a$

  (a dominates b)

NON-DOMINATED ACTIONS

Properties of the non-dominated set $\mathcal{ND}$

- Non-empty action set $\mathcal{A} \Rightarrow$ non-empty $\mathcal{ND}$

- Compact $\mathcal{A}$ and $\mathcal{L}$ generated by a finite number of loss functions, continuous in a, uniformly w.r.t. $\theta \Rightarrow$ non-empty $\mathcal{ND}$

- Unique Bayes action a_π^L for any $L \in \mathcal{L}$ and $\pi \in \Gamma \Rightarrow \mathcal{B} \subseteq \mathcal{ND}$, $\mathcal{B}$ set of Bayes actions

SENSITIVITY MEASURES

Global sensitivity

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

Measures

- Range: $\delta = \overline{\rho} - \underline{\rho}$, with $\overline{\rho} = \sup_{P \in \Gamma} E_{P^*}[h(\theta)]$ and $\underline{\rho} = \inf_{P \in \Gamma} E_{P^*}[h(\theta)]$
  Simple interpretation
- Relative sensitivity $\sup_\pi R_\pi$, with $R_\pi = \frac{(\rho_\pi - \rho_0)^2}{V^\pi}$, $\rho_0 = E_{\Pi_0^*}[h(\theta)]$, $\rho_\pi = E_{\Pi^*}[h(\theta)]$ and $V^\pi = Var_{\Pi^*}[h(\theta)]$
  Scale invariant, decision theoretic interpretation, asymptotic behaviour

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 $K_0 = \{P_0\}$

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

- Gatêaux differential

SENSITIVITY MEASURES

Measures

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

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

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

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

- $\Gamma_\varepsilon = \{P : P = (1-\varepsilon)P_0 + \varepsilon Q, Q \in \mathcal{Q}_A\}$

- $\Gamma_Q = \{P : P(I_i) = p_i, i = 1, \ldots, m\}$

Probability measures as mixture of extremal ones

COMPUTATIONAL TECHNIQUES

- Linearisation technique
  - $\overline{\rho} = \inf\{q | c(q) = 0\}$ where
  - $c(q) = \sup_{P \in \Gamma} \int_{\Theta} c(\theta, q) P(d\theta) = 0$
  - $c(\theta, q) = l(\theta)\ (h(\theta) - q)$
  - 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)

QUEST FOR ROBUSTNESS

Range δ "large" and possible refinement of Γ

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

QUEST FOR ROBUSTNESS

Inherently robust procedures

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

LACK OF ROBUSTNESS

Range δ "large" and no further possible refinement of Γ

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

- Choice of an estimate of $E_{P^*}[h(\theta)]$ according to an optimality criterion, e.g.
  - Γ–minimax posterior expected loss
  - Γ–minimax posterior regret

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

GAMMA-MINIMAX

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

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

  Optimal action by interchanging inf and sup for convex losses

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

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

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

APPLICATIONS

- 8 different configurations of pipelines (diameter, depth, location)
- Gas escapes modelled by $\mathcal{P}(\lambda_i)$, $i = 1, 8$
- Gamma priors on λ_i
- Pipelines ranked according to posterior mean of λ_i's
- Classes of gamma priors with parameters in intervals
- $\Rightarrow$ Sensitivity of ranking w.r.t. priors

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 $\bar{E}[X_k|D_k]/n_k = 0.24$

APPLICATIONS

Wavelets in nonlinear regression

- $y_i = f(x_i) + \varepsilon_i,\ \varepsilon_i$ i.i.d. $\mathcal{N}(0, \sigma^2),\ i = 1, N (= 2^n)$
- y_i: noisy measurements
- $x_i = i/N$
- f: unknown signal
- ε_i: noise
- wavelet transform $W \Rightarrow d_i = \theta_i + \eta_i,\ i = 1, N$
  $[\underline{y} \rightarrow \underline{d} = W\underline{y}, \underline{f} \rightarrow \underline{\theta} = W\underline{f}, \underline{\varepsilon} \rightarrow \underline{\eta} = W\underline{\varepsilon}]$

APPLICATIONS

Wavelets in nonlinear regression

- Model for $d_i = \theta_i + \eta_i$

- $d_i|\theta \sim f(d_i|\theta) = f(d_i - \theta)$, symmetric and unimodal
  e.g. $d_i|\theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$

- Loss $L(\theta, a) = (\theta - a)^2 \Rightarrow E^{\theta|d_i}\theta$ optimal

- Signal smoothed by thresholding or shrinkage

Are Bayesian estimators shrinkers?

APPLICATIONS

Wavelets in nonlinear regression

How to choose prior to have shrinkage, i.e. $\Delta = |E^{\theta|d}\theta/d| < 1$*?*

- $\Gamma_S = \{$all symmetric$\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta > 1$ ($= \infty$ for normal model)

- $\Gamma_{Sp} = \{$all symmetric + mass p at 0$\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta < \infty$ but > 1 for "small" p

- $\Gamma_{SU} = \{$all symmetric, unimodal$\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta \leq 1$

- $\Gamma_S = \{$all symmetric, unimodal+ mass p at 0$\}$
  $\Rightarrow \sup_{\pi \in \Gamma_S} \Delta < 1$

PREFERENCES AMONG PRIORS

- Expert able not only to specify a class Γ of priors but also preferences among them or its subsets (e.g. elicitation of a quantile class, allowing even for discrete distributions, but *absolutely continuous unimodal priors* preferred to *step functions* and even more to *discrete distributions*)

- Sensitivity analysis over Γ could lead to lack of robustness but robustness might be achieved in the subset of Γ more likely according to the expert

- Instead of reporting lack of robustness in the larger class and choosing a convenient prior in it (providing both Bayes estimator under it and range over Γ), analyst could report a robust Bayesian estimator along with the subset not considered in the computation of the range

- How to make this formal in a probabilistic framework?

PREFERENCES AMONG PRIORS

- Given $X \sim f(x|\theta) \Rightarrow$ interest in posterior mean of θ
- In $\Gamma_P = \{P : p(\theta;\omega), \omega \in \Omega\}$ preferences can be described by a function $\pi(\omega)$
- $\pi(\omega)$ can be treated as a prior $\Rightarrow$ *formally* a hierarchical model $X \sim f(x|\theta), \theta \sim p(\theta;\omega)$ and $\pi(\omega)$
- Posterior mean of θ unique under hierarchical model but the original problem, $\omega \in \Omega$, leads to a set of values for the posterior mean
- Compute the range on a subset of Ω such that its *probability* under π is high but the range is as small as possible
- How to make the procedure formally acceptable in a probabilistic framework and how to extend it to a nonparametric class?

PARTIAL AND INCOMPATIBLE INFORMATION

Combination of opinions of conflicting experts in different fields (e.g. e-democracy) $\Rightarrow$ partial and incompatible information

Three experts provide information on different pairs

Marginal	(0,0)	(0,1)	(1,0)	(1,1)
$f_1(x_1,x_2)$	0.47	0.13	0.13	0.27
$f_2(x_2,x_3)$	0.47	0.13	0.13	0.27
$f_3(x_1,x_3)$	0.30	0.30	0.30	0.10

X_1, X_2 and X_3: Bernoulli $\mathcal{B}e(0.4)$

No joint density $f(x_1,x_2,x_3)$ with those marginals

How to combine partial and incompatible priors, possibly in an automatic way?

PARTIAL AND INCOMPATIBLE INFORMATION

Given the random quantities X_1, ..., X_n, how to combine them?

1. Choose a rule: chain's rule!

2. Agree on an order for the chain's rule, e.g.
$f(X_1, \ldots, X_n) = f(X_n | X_{n-1}, \ldots, X_1) \cdot f(X_{n-1} | X_{n-2} \ldots X_1) \cdot \ldots \cdot f(X_1)$

3. For each component $f(X_k | X_{k-1}, \ldots, X_1)$ look for all the contributions of the stakeholders on it

4. Combine the contributions into $\widetilde{f}(X_k | X_{k-1}, \ldots, X_1)$

5. Get the joint density $\widetilde{f}(X_1, \ldots, X_n)$ via chain's rule, combining all $\widetilde{f}(X_k | X_{k-1}, \ldots, X_1)$

PARTIAL AND INCOMPATIBLE INFORMATION

Marginal	(0,0)	(0,1)	(1,0)	(1,1)
$f_1(x_1, x_2)$	0.47	0.13	0.13	0.27
$f_2(x_2, x_3)$	0.47	0.13	0.13	0.27
$f_3(x_1, x_3)$	0.30	0.30	0.30	0.10

Chain's rule $f(x_1, x_2, x_3) = f(x_3|x_2, x_1)f(x_2|x_1)f(x_1)$

Assumptions like $f_2(x_3, x_2) = f_2(x_3|x_2)f_2(x_2) = f_2(x_3|x_2, x_1)f_2(x_2)$

Contributions to each components

$$\begin{aligned} f(x_1) &= \alpha f_1(x_1) + (1-\alpha)f_3(x_1) \\ f(x_2|x_1) &= \beta f_1(x_2|x_1) + (1-\beta)f_2(x_2) \\ f(x_3|x_2, x_1) &= \gamma f_2(x_3|x_2) + (1-\gamma)f_3(x_3|x_1) \end{aligned}$$

$\Rightarrow X_1$ still Bernoulli $\mathcal{B}e(0.4)$ (and α disappears)

Conditional	(0,0)	(0,1)	(1,0)	(1,1)
$f_1(x_2|x_1)$	47/60	13/60	13/40	27/40
$f_2(x_3|x_2)$	47/60	13/60	13/40	27/40
$f_3(x_3|x_1)$	0.50	0.50	0.75	0.25

PARTIAL AND INCOMPATIBLE INFORMATION

Conditional densities of f on their support

	(0,0)	(0,1)
$f(x_2\|x_1)$	$(36+11\beta)/60$	$(24-11\beta)/60$
	(1,0)	(1,1)
$f(x_2\|x_1)$	$(24-11\beta)/40$	$(16+11\beta)/40$

	(0,0,0)	(0,0,1)
$f(x_3\|x_2,x_1)$	$(30+17\gamma)/60$	$(30-17\gamma)/60$
	(0,1,0)	(0,1,1)
$f(x_3\|x_2,x_1)$	$(20-7\gamma)/40$	$(20+7\gamma)/40$
	(1,0,0)	(1,0,1)
$f(x_3\|x_2,x_1)$	$(45+2\gamma)/60$	$(15-2\gamma)/60$
	(1,1,0)	(1,1,1)
$f(x_3\|x_2,x_1)$	$(30-17\gamma)/40$	$(10+17\gamma)/40$

PARTIAL AND INCOMPATIBLE INFORMATION

Joint density $f(x_1, x_2, x_3)$

(x_1, x_2, x_3)	$f(x_1, x_2, x_3)$
(0,0,0)	$(30 + 17\gamma)(36 + 11\beta)/6000$
(0,0,1)	$(30 - 17\gamma)(36 + 11\beta)/6000$
(0,1,0)	$(20 - 7\gamma)(24 - 11\beta)/4000$
(0,1,1)	$(20 + 7\gamma)(24 - 11\beta)/4000$
(1,0,0)	$(45 + 2\gamma)(24 - 11\beta)/6000$
(1,0,1)	$(15 - 2\gamma)(24 - 11\beta)/6000$
(1,1,0)	$(30 - 17\gamma)(16 + 11\beta)/4000$
(1,1,1)	$(10 + 17\gamma)(16 + 11\beta)/4000$

PARTIAL AND INCOMPATIBLE INFORMATION

Bivariate marginals

	(0,0)
$f(x_1, x_2)$	$0.36 + 0.11\beta$
$f(x_1, x_3)$	$0.3 + (0.06 + 0.05\beta)\gamma$
$f(x_2, x_3)$	$0.36 - (0.0275 - 0.0275\gamma)\beta + 0.11\gamma$
	(0,1)
$f(x_1, x_2)$	$0.24 - 0.11\beta$
$f(x_1, x_3)$	$0.3 - (0.06 + 0.05\beta)\gamma$
$f(x_2, x_3)$	$0.24 + (0.0275 - 0.0275\gamma)\beta - 0.11\gamma$
	(1,0)
$f(x_1, x_2)$	$0.24 - 0.11\beta$
$f(x_1, x_3)$	$0.3 - (0.06 + 0.05\beta)\gamma$
$f(x_2, x_3)$	$0.24 + (0.0275 - 0.0275\gamma)\beta - 0.11\gamma$
	(1,1)
$f(x_1, x_2)$	$0.16 + 0.11\beta$
$f(x_1, x_3)$	$0.1 + (0.06 + 0.05\beta)\gamma$
$f(x_2, x_3)$	$0.16 - (0.0275 - 0.0275\gamma)\beta + 0.11\gamma$

PARTIAL AND INCOMPATIBLE INFORMATION

Univariate Bernoulli marginals are kept

$\beta = 1 \Rightarrow f(x_1, x_2) = f_1(x_1, x_2)$

$\gamma = 0 \Rightarrow f(x_1, x_3) = f_3(x_1, x_3)$

$\beta = 1, \gamma = 0 \Rightarrow$

	(0,0)	(0,1)	(1,0)	(1,1)
$f_2(x_2, x_3)$	.47	.13	.13	.27
$f(x_2, x_3)$	.3325	.2675	.2675	.1325

$\beta = 0.5, \gamma = 0.5 \Rightarrow$

	(0,0)	(0,1)	(1,0)	(1,1)
$f_1(x_1, x_2)$	.47	.13	.13	.27
$f(x_1, x_2)$	.415	.185	.185	.215
$f_2(x_2, x_3)$	.47	.13	.13	.27
$f(x_2, x_3)$	.408125	.191875	.191875	.208125
$f_3(x_1, x_3)$	.30	.30	.30	.10
$f(x_1, x_3)$	.3425	.2575	.2575	.1425

Now the marginals are compatible!

BAYESIAN NONPARAMETRICS

- Risk analysis $\Rightarrow$ extreme value theory $\Rightarrow$ Generalized Extreme Value (GEV) distribution
- Cdf $F(x) = \exp\left\{-\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\lambda}\right\}$
- Density $f(x) = \frac{1}{\sigma}\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\lambda-1} \exp\left\{-\left[1+\lambda\left(\frac{x-\mu}{\sigma}\right)\right]_+^{-1/\lambda}\right\}$
- q-th quantile: $q = \exp\left\{-\left[1+\lambda\left(\frac{G_q-\mu}{\sigma}\right)\right]_+^{-1/\lambda}\right\}$
- Expert gives 3 quantiles in the tails (e.g. .80, .95, .99) on the observable quantity $X \Rightarrow$ parameters μ, σ and λ determined
- Expert presented with plots of density functions until satisfied with the shape
- Quantile specification in the tail $\Rightarrow$ good approximation in the tail but bad elsewhere

BAYESIAN NONPARAMETRICS

- Generalised moments constrained class, given by $q_i = \int_0^{Q_i} \{\int f(x|\mu,\sigma,\lambda)\pi(\mu,\sigma,\lambda)d\mu d\sigma d\lambda\}dx, i = 1,2,3$

- As an alternative $\Rightarrow$ 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 the parametric model in a Dirichlet process with parameter $\eta(x) = \alpha F(x; \widehat{\mu}, \tilde{\sigma}, \widehat{\lambda})$

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. $\Theta = \{F : F(1/2) \leq 1/2\}$)
  - $G \sim \mathcal{DP}(\eta) \Rightarrow G(A) \sim \mathcal{B}(\eta(A), \eta(A^C)) \Rightarrow$ compute $\mathcal{P}(\Theta) = \mathcal{P}(G(A) \in \ B)$

BAYESIAN NONPARAMETRICS

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

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

IMPRECISE PROBABILITIES

- Similar tools but ...
- ... different philosophy

IMPORTANT PROBLEMS

- Software
- Efficient and parsimonious MCMC simulations for Bayesian robustness (current methods are for a unique prior)
- Classes more problem driven
- Applications

EXERCISE 3

- Flip of a coin
- $P(\text{tail}) = P(X = 1) = \theta$
- Sample $X_1, \ldots, X_n$
- Perform a robust Bayesian analysis