New classes of priors based on stochastic orders and distortion functions

Fabrizio Ruggeri

Istituto di Matematica Applicata e Tecnologie Informatiche
Consiglio Nazionale delle Ricerche
Via Bassini 15, I-20133, Milano, Italy
fabrizio@mi.imati.cnr.it
web.mi.imati.cnr.it/fabrizio/

Institute for Future Environments, Queensland University of Technology
Brisbane, Australia

Joint work with

• Jose Pablo Arias-Nicolás (Universidad de Extremadura, Spain)
• Alfonso Suárez-Llorens (Universidad de Cadiz, Spain)

to appear in *Bayesian Analysis*
OUTLINE OF THE TALK

- Bayesian robustness
- Stochastic orders
- Distortion function and distortion band of priors
- Relation with concentration functions
- Metrics to measure uncertainty
- Ordering of Bayes actions
- Numerical examples
- Future work
• $X \sim \mathcal{N}(\theta, 1)$

• Expert’s opinion on prior $P$: median at 0, quartiles at $\pm 1$, symmetric and unimodal

• $\Rightarrow$ Possible priors include Cauchy $C(0, 1)$ and Gaussian $\mathcal{N}(0, 2.19)$

• Interest in posterior mean $\mu^C(x)$ or $\mu^N(x)$

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>4.5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu^C(x)$</td>
<td>0</td>
<td>0.52</td>
<td>1.27</td>
<td>4.09</td>
<td>9.80</td>
</tr>
<tr>
<td>$\mu^N(x)$</td>
<td>0</td>
<td>0.69</td>
<td>1.37</td>
<td>3.09</td>
<td>6.87</td>
</tr>
</tbody>
</table>

• Decision strongly dependent on the choice of the prior for large $x$
BAYESIAN ROBUSTNESS

- Practical impossibility of specifying priors exactly matching experts’ knowledge

- Prior elicitation subject to uncertainty and, possibly, some degree of arbitrariness introduced by the analyst, e.g. the functional form of the distribution

- Uncertainty in the choice of priors modelled through a class of distribution (the same might apply for loss functions and statistical models/likelihoods)

- Use of indices to measure the consequences (i.e. perform robustness analysis) of the choice of a class of priors on the quantities of interest (e.g. posterior mean)
BAYESIAN ROBUSTNESS

- Choice of a class $\Gamma$ of priors

- Interest in posterior expectation of $g(\theta)$ (e.g. posterior mean if $g(\theta) = \theta$)

- Computation of a robustness measure, e.g. range $\delta = \bar{\rho} - \underline{\rho}$

  \[
  (\bar{\rho} = \sup_{P \in \Gamma} E_P[g(\theta)] \text{ and } \underline{\rho} = \inf_{P \in \Gamma} E_P[g(\theta)])
  \]

  - $\delta$ “small” $\Rightarrow$ robustness

  - $\delta$ “large”, $\Gamma_1 \subset \Gamma$ and/or new data

  - $\delta$ “large”, $\Gamma$ and same data
BAYESIAN ROBUSTNESS - CLASSES OF PRIORS

- $\Gamma_Q = \{P : \alpha_i \leq P(I_i) \leq \beta_i, i = 1, \ldots, m\}$ (Quantile class)
- $\Gamma_{GM} = \{P : \int h_i(\theta) dP(\theta) = a_i, i = 1, \ldots, m\}$ (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)
- $\Gamma_\varepsilon = \{P : P = (1 - \varepsilon)P_0 + \varepsilon Q, Q \in Q\}$ ($\varepsilon$-contaminations)
- etc.
A SHORT HISTORY OF 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)
STOCHASTIC ORDERS

- **Usual** stochastic order
  - $X$ and $Y$ r.v.’s with d.f.’s $F_X$ and $F_Y$ s.t. $F_X(t) \geq F_Y(t), \quad \forall t \in \mathbb{R}$
  - $\Rightarrow X \leq_{st} Y$, i.e. $X$ is said to be **smaller than** $Y$ **in the usual stochastic order**
  - $X \leq_{st} Y \iff E[g(X)] \leq E[g(Y)]$ 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. $\frac{f_Y(t)}{f_X(t)}$ increases over the union of the supports of $X$ and $Y$ (here $a/0$ is taken to be equal to $\infty$ whenever $a > 0$)
  - $\Rightarrow X \leq_{lr} Y$, 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 $h : [0, 1] \rightarrow [0, 1]$
  - 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 $\pi$ with d.f. $F_\pi(\theta)$ and distortion function $h$

- $\Rightarrow$ distorted prior distribution $\pi_h$ with d.f. $F_{\pi_h}(\theta) = h \circ F_\pi(\theta) = h[F_\pi(\theta)]$

- Lemma.
  - $\pi$ prior distribution (absolutely continuous or discrete) with d.f. $F_\pi$
  - $h$ 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
SOME RESULTS

• $X$ r.v. with d.f $F \Rightarrow$ quantile function $F_X^{-1}(p) = \inf\{x : F_X(x) \geq p\}, \forall p \in (0, 1)$

• $X$ and $Y$ r.v.’s with continuous and strictly increasing d.f.’s $F_X$ and $F_Y$
  $\Rightarrow h_{XY}(t) = F_Y(F_X^{-1}(t))$ distortion function mapping the d.f. of $X$ to the d.f. of $Y$

• **Theorem (Lehmann and Rojo, 1992).** Under the above regularity conditions, the likelihood ratio order between $X$ and $Y$ is equivalent to check if the function $h_{XY}(t)$ is convex or, analogously, if $h_{YX}(t)$ is concave

• Therefore, assuming continuous and strictly increasing distributions functions, the existence of a convex or a concave distortion that maps a distribution function to another one is a necessary and sufficient condition for the likelihood ratio order
DISTORTED BAND OF PRIORS

- Uncertainty on prior $\pi$ through concave ($h_1$) and convex ($h_2$) distortion functions

- **Previous Lemma.** Prior $\pi$ and convex (or concave) distortion function $h$ in $[0, 1]$ implies $\pi \leq_{lr} \pi_h$ (or $\pi \geq_{lr} \pi_h$)

- Lemma $\Rightarrow$ distorted distributions $\pi_{h_1}$ and $\pi_{h_2}$ 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 $\pi$, with lower and upper bound given by distorted distributions

- Band defined only through an upper (or lower) bound when considering $h_1$ (or $h_2$) instead of the identity function
DISTORTED BAND OF PRIORS

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

- $\Rightarrow$ distorted band subclass of well known distribution band class, i.e.
  
  $\Gamma_{h_1,h_2,\pi} \subseteq \{\pi' : \pi_{h_1} \leq_{st} \pi' \leq_{st} \pi_{h_2}\}$,  
  $\quad = \{\pi' : F_{\pi_{h_1}}(\theta) \geq F_{\pi'}(\theta) \geq F_{\pi_{h_2}}(\theta), \forall \theta \in \Theta\}$

- Usually d.f. of $\pi_{h_1}$ "upper bound" $F_U$ and d.f of $\pi_{h_2}$ "lower bound" $F_L$

- Interpretation of distortion band in terms of prior probability sets based on Shaked and Shanthikumar (2007):
  
  $\Gamma_{h_1,h_2,\pi} = \{\pi' : \pi_{h_1}(\cdot|A) \leq_{st} \pi'(\cdot|A) \leq_{st} \pi_{h_2}(\cdot|A)\}$, for all measurable $A \subseteq \Theta$

- Note that likelihood ratio order does not apply, in general, when comparing two priors $\pi'_1$ and $\pi'_2$ in $\Gamma_{h_1,h_2,\pi}$, since each of them is just ordered w.r.t. $\pi_{h_1}$ and $\pi_{h_2}$
DISTORTED BAND OF PRIORS

• Given any pair $\pi_1$ and $\pi_2$ in $\Gamma_{h_1,h_2,\pi}$ and any $0 \leq \epsilon \leq 1$
  $\Rightarrow$ consider mixture prior $\pi_\epsilon = (1 - \epsilon)\pi_1 + \epsilon\pi_2$

• It can be proved that $\pi_{h_1} \leq_{lr} \pi_\epsilon$ and $\pi_\epsilon \leq_{lr} \pi_{h_2}$

• $\Rightarrow \pi_\epsilon \in \Gamma_{h_1,h_2,\pi}$

• In particular mixtures of $\pi$ (baseline prior) and any $\pi_1$ in $\Gamma_{h_1,h_2,\pi}$ belong to $\Gamma_{h_1,h_2,\pi}$
CHOICES OF DISTORTION FUNCTIONS

• \( h_1(x) = 1 - (1 - x)^\alpha \) and \( h_2(x) = x^\alpha \), \( \forall \alpha > 1 \)
  - \( \alpha = n \in \mathbb{N} \Rightarrow F_{\pi h_1}(\theta) = 1 - (1 - F_\pi(\theta))^n \) 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 \( \pi \)

• \( h_1(x) = \min\{\frac{x}{\alpha}, 1\} \) and \( h_2(x) = \max\{\frac{x-\alpha}{1-\alpha}, 0\}, \quad 0 < \alpha < 1 \)
  - \( \Rightarrow \) truncated distributions \( \pi_{h_1} = st \pi(\cdot | A_1) \) and \( \pi_{h_2} = st \pi(\cdot | A_2) \)
    * \( =_{st} \) means equality in law
    * \( A_1 = (-\infty, F_{\pi}^{-1}(\alpha)] \)
    * \( A_2 = (F_{\pi}^{-1}(\alpha), \infty) \)
CHOICES OF DISTORTION FUNCTIONS

- Skewed distributions
- $\pi$ absolutely continuous, symmetric around 0 prior with density $\pi(\theta)$ and d.f. $F_\pi(\theta)$
- $\Rightarrow$ skew-$\pi$ with parameter $\alpha$ 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

- $h_{\pi \pi, \alpha}(x) = \int_{-\infty}^{F_{\pi}^{-1}(x)} 2\pi(\theta) F_{\pi}(\alpha \theta) d\theta$
  - maps d.f. prior $\pi$ to d.f. of skewed $\pi_{\alpha}$
  - $\Rightarrow h'_{\pi \pi, \alpha}(x) = 2F_{\pi}(\alpha F_{\pi}^{-1}(x))$
  - Both $F_{\pi}$ and $F_{\pi}^{-1}$ increasing and differentiable
    $\Rightarrow h'_{\pi \pi, \alpha}(x)$ increasing $\forall \alpha > 0$ and decreasing $\forall \alpha < 0$
    $\Rightarrow h_{\pi \pi, \alpha}(x)$ convex or concave, respectively
  - Concave and convex functions given, $\forall \beta \geq 0$, by
    * $h_{1}(x) = \int_{-\infty}^{F_{\pi}^{-1}(x)} 2\pi(\theta) F_{\pi}(-\beta \theta) d\theta$
    * $h_{2}(x) = \int_{-\infty}^{F_{\pi}^{-1}(x)} 2\pi(\theta) F_{\pi}(\beta \theta) d\theta$
CHOICES OF DISTORTION FUNCTIONS

- $\pi \sim N(0, 1)$ prior with standard normal d.f. $\Phi_Z$

- 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. $\Phi_Z$
- 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 $\pi_1$ and $\pi_2$ s.t. $\pi_1 \leq_{lr} \pi_2$
  $\Rightarrow$ posteriors s.t. $\pi_{1x} \leq_{lr} \pi_{2x}$

- **Proposition.** $\pi$ prior and $\Gamma_{h_1, h_2, \pi}$ distorted band around $\pi$ 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 $\Gamma_{x}$, 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
CONCENTRATION FUNCTION CLASS

• $n$ individuals with wealth $x_i, i = 1, \ldots, n$ ⇒ 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 $(\Omega, \mathcal{F}, P)$ ⇒ concentration function
CONCENTRATION FUNCTION CLASS

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

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

- Each horizontal line at $y \Rightarrow \text{subset } A_y$ with likelihood ratio $m(\omega) = \frac{p(\omega)}{p_0(\omega)} \leq q$

- 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

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

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

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

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

Main properties

• \( \varphi(x) \) nondecreasing, continuous and convex, \( \varphi(0) = 0 \)

• \( \varphi(x) \equiv 0 \iff P \perp P_0 \)

• \( \varphi(x) = x, \forall x \in [0, 1] \iff P = P_0 \)

• \( P_0(A) = x \Rightarrow \varphi(x) \leq P(A) \leq 1 - \varphi(1 - x) \)

• \( \varphi(x) = \int_0^{c_x} \{x - H(t)\} dt = \int_0^x c_t dt \)

• \( \lim_{n \to \infty} \varphi_{P_n}(x) = x, \forall x \in [0, 1] \iff \lim_{n \to \infty} \sup_{A \in \mathcal{F}} |P_n(A) - P_0(A)| = 0 \)
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 $y : x = P_0 (\{\theta \in \Theta : m(\theta) \leq y\}) = 1 - e^{-y}$

$\Rightarrow \varphi(x) = P (\{\theta \in \Theta : m(\theta) \leq y\}) = 1 - (1 - x)(1 - \log(1 - x))$
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

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

- \( P_0 \) uniform over \([0, 1]\)

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

- \( \Rightarrow \varphi(x) \geq x^2 = g(x), \quad \forall x \in [0, 1] \)
CONCENTRATION FUNCTION AND DISTORTION BANDS

- Nondecreasing, continuous and convex distortion function $h(x)$

- R.v. $X$ with d.f. $F(x)$ and density $f(x)$

- Distorted r.v. $X_h$ with d.f. $F_h(x) = h[F(x)]$ and density $f_h(x) = h'[F(x)]f(x)$

- Likelihood ratio $m(x) = \frac{f_h(x)}{f(x)} = h'[F(x)]$ increasing since $h'' > 0$ because of convexity of $h$

- Consider likelihood subsets $L_z = (-\infty, z]$ with probability $F(z)$ and $F_h(z)$ under the two probability measures

- Take $x_z = F(z)$ and assume $F$ invertible so that $z = F^{-1}(x_z)$

- $\varphi_h(x_z) = F_h(z) = h[F(z)] = h[F(F^{-1}(x_z))] = h(x_z)$
CONCENTRATION FUNCTION AND DISTORTION BANDS

- Dropping the dependence on \( z \) in \( \varphi_h(x_z) = h(x_z) \)
  \[ \Rightarrow \varphi_h(x) = h(x) \text{ c.f. of p.m. } \Pi \text{ (for r.v. } X_h) \text{ w.r.t. p.m. } \Pi_0 \text{ (for r.v. } X) \]

- Given a distorted measure \( \Rightarrow \) its distortion function interpreted as c.f. of the distorted measure w.r.t. baseline one

- Given a nondecreasing, continuous and convex function \( h(x) \), there exists an infinite number of p.m.’s (including the corresponding distorted measure) whose c.f.’s w.r.t. the baseline measure are given by \( h(x) \)

- \( \Psi_{\pi_0,h} \) c.f. class of priors given by all the p.m.’s whose c.f.’s are above a nondecreasing, continuous and convex function \( h(x) \)

- \( \Gamma_{\pi_0,h} = \{\pi' : \pi_0 \leq_{lr} \pi' \leq_{lr} \pi_h\} \) distorted band

- **Theorem.** The distorted band class \( \Gamma_{\pi_0,h} \) is properly included in the concentration function class \( \Psi_{\pi_0,h} \)
CONCENTRATION FUNCTION AND DISTORTION BANDS

- **Theorem.** The distorted band class $\Gamma_{\pi_0, h}$ is properly included in the concentration function class $\Psi_{\pi_0, h}$

- **Example** showing the inclusion is proper:
  - Uniform distribution on $[0, 1]$ as a baseline prior $\pi_0$
  - Distortion function $h(x) = x^2$
  - $\Rightarrow$ corresponding distorted distribution with density $\pi_h(x) = 2x$, whose likelihood ratio w.r.t. the uniform density is increasing $(m(x) = \pi_h(x)) = 2x$
  - Distribution $\pi^*$ with density $\pi^*(x) = 2(1 - x)$ has same c.f. $\varphi^*(x) = x^2$ as distorted distribution (w.r.t. $\pi_0$) but ratio of its density w.r.t. the uniform one is decreasing
  - $\Rightarrow \pi^* \not\in \Gamma_{\pi_0, h}$
CONCENTRATION FUNCTION AND DISTORTION BANDS

- As a consequence of the Representation Theorem 3 in Fortini and Ruggeri (1995), all priors in the distorted band class can be represented as mixture of extremal distributions in $\Psi_{\pi_0, h}$.

- As a consequence of the previous theorem and Theorem 4 in Fortini and Ruggeri (1995), it is possible to provide an upper bound on the supremum of the expectation of an integrable function $g(x)$ w.r.t. the class of priors $\Gamma_{\pi_0, h}$, since

$$\sup_{\pi \in \Gamma_{\pi_0, h}} E^\pi(g(X)) \leq \sup_{\pi \in \Psi_{\pi_0, h}} E^\pi(g(X))$$

- The supremum of the expectation of $g(x)$ over the class $\Psi_{\pi_0, h}$ is obtained for a distribution with c.f. $h(x)$ w.r.t. $\pi_0$, as proved in Fortini and Ruggeri (1995).

- A lower bound on the infimum is obtained similarly.

- The finding can be useful, especially when the difference between upper and lower bounds is small, when performing a sensitivity analysis about a posterior expected value (e.g. of the function $g(x)$) aimed to measure the influence of the choice of a prior in a class.
METRICS TO MEASURE UNCERTAINTY

- Interest in probability metrics to evaluate how a prior belief differs from its distorted version and how the corresponding posterior distributions differ

- Mathematical tractability (but not only!) ⇒ interest in Kolmogorov and Kantorovich metrics

- R.v.’s $X$ and $Y$ with d.f.’s $F_X$ and $F_Y$

- Kolmogorov metric $K(X, Y)$
  
  $K(X, Y) = \sup_{x \in \mathbb{R}} |F_X(x) - F_Y(x)|$

  ⇒ largest absolute difference between $F_X$ and $F_Y$

- Kantorovich (or Wasserstein) metric $KW(X, Y)$

  $KW(X, Y) = \int_{-\infty}^{\infty} |F_X(x) - F_Y(x)| dx$

- $|F_X(x) - F_Y(x)| \to 0$ as $x \to \pm \infty$ ⇒ Kolmogorov metric completely insensitive to differences in the tails of the distributions, unlike Kantorovich’s
KOLMOGOROV METRIC

Lemma.

- $\pi$ absolutely continuous prior

- $h$ differentiable (concave or convex) distortion function

- Kolmogorov distance between $\pi$ and $\pi_h$ given by

$$K(\pi, \pi_h) = \sup_{x \in \mathbb{R}} |F_\pi(x) - F_{\pi_h}(x)|,$$

$$= \begin{cases} p_0 - h(p_0) & \text{if } h \text{ is convex}, \\ h(p_0) - p_0 & \text{if } h \text{ is concave}. \end{cases}$$

- $p_0$ satisfies $h'(p_0) = 1$

- maximum achieved at $\theta_0 = F_\pi^{-1}(p_0)$
KOLMOGOROV METRIC

• Distance in previous Lemma depends only on the distortion function
  – $p_0 - h(p_0)$ and $p_0$ satisfying $h'(p_0) = 1$

• $\Rightarrow$ Kolmogorov distance useful to measure (and tune) uncertainty in distorted band

• Example.
  – $h_1(x) = 1 - (1 - x)^\alpha$ and $h_2(x) = x^\alpha$, $\forall \alpha > 1$
  – $\Rightarrow K(\pi, \pi_{h_1}) = K(\pi, \pi_{h_2}) = \frac{\alpha - 1}{\sqrt[\alpha - 1]{\alpha^\alpha}}$

  – $K(\pi, \pi_{h_1})$ increasing function of $\alpha$

  – $\alpha = 1.2$

  – $\Rightarrow K(\pi, \pi_{h_1}) = K(\pi, \pi_{h_2}) = 0.067$, i.e. the d.f.'s differs at most for 0.067
KANTOROVICH METRIC

• $X$ and $Y$ with d.f. $F_X$ and $F_Y$ s.t. $X \leq_{st} Y$

$$KW(X, Y) = \int_{-\infty}^{\infty} |F_X(x) - F_Y(x)| \, dx$$

$$= \int_{-\infty}^{\infty} (F_X(x) - F_Y(x)) \, dx$$

$$= E(Y) - E(X)$$

when expectations exist

• $X \leq_{st} Y$ and $E(X) = E(Y)$
  
  $\Rightarrow KW(X, Y) = 0$

  $\Rightarrow X$ and $Y$ equal in distributions, i.e., $X =_{st} Y$
KANTOROVICH METRIC

- $KW(\pi_{h_1}, \pi_{h_2}) = E^{\pi_{h_2}}(\theta) - E^{\pi_{h_1}}(\theta)$
- $KW(\pi, \pi_{h_1}) = E^{\pi}(\theta) - E^{\pi_{h_1}}(\theta)$
- $KW(\pi, \pi_{h_2}) = E^{\pi_{h_2}}(\theta) - E^{\pi}(\theta)$
- Same for posterior distributions
- $KW(\pi_{h_1}, \pi_{h_2}) = KW(\pi, \pi_{h_2}) + KW(\pi, \pi_{h_1})$
- $\Rightarrow$ possible identify which bound, $h_1$ or $h_2$, is contributing the most to uncertainty
KANTOROVICH METRIC

• Example.
  - $\pi \sim U(0, 1)$
  - $h_1(x) = 1 - (1 - x)^\alpha$ and $h_2(x) = x^\alpha$, $\forall \alpha > 1$
  - $\Rightarrow KW(\pi, \pi_{h_1}) = KW(\pi, \pi_{h_2}) = \frac{1}{2} - \frac{1}{\alpha + 1}$
  - $KW(\pi, \pi_{h_1})$ increasing function of $\alpha$
  - $\Rightarrow$ distance between expectations not greater than 0.5
  - If $h_1(x) = 1 - (1 - x)^2$ and $h_2(x) = x^3$
  - $\Rightarrow KW(\pi, \pi_{h_1}) = \frac{1}{6} < \frac{1}{4} = KW(\pi, \pi_{h_2})$
  - $\Rightarrow h_2$ contributes more than $h_1$ to uncertainty
CHOICE OF DISTORTION FUNCTIONS

- Parameter $\theta$ meaningful in a financial context
  - class allows for both risk aversion and proneness through use of convex and concave distortion functions, respectively, giving more (or less) weight to larger events
  - proper choice of parameters to model more uncertainty towards aversion or proneness

- $h_1(x) = 1 - (1 - x)^\alpha$ and $h_2(x) = x^\alpha, \forall \alpha > 1$
  - satisfactorily represents uncertainty in the tails of the prior

- Concave and convex functions given by
  
  \[
  h_1(x) = \int_{-\infty}^{F^{-1}_\pi(x)} 2\pi(\theta)F_\pi(-\beta\theta)d\theta \quad \text{and} \quad h_2(x) = \int_{-\infty}^{F^{-1}_\pi(x)} 2\pi(\theta)F_\pi(\beta\theta)d\theta, \forall \beta \geq 0
  \]
  - useful to elicit prior knowledge with "normal-like" shape but with lack of symmetry
(ROBUST) DECISION ANALYSIS

- $X$ r.v. with distribution $P_{\theta}$ and density $p_{\theta}(x)$
- $\pi$ prior, with density $\pi(\theta)$ over the set of states $\theta \in \Theta$, in class $\Gamma$
- $l(\theta)$ likelihood function for observed $x$
- $m_{\pi}(x) = \int l(\theta)\pi(\theta)d\theta$ marginal density
- $\pi_{x}$ posterior with density $\pi_{x}(\theta)$ in class $\Gamma_{x}$
- $\mathcal{A}$ set of alternatives (actions) $a$
- $L(\theta, a)$ loss function in class $\mathcal{L}$
(ROBUST) DECISION ANALYSIS

• $\rho(\pi, L, a)$ posterior expected loss of $a$, i.e.

$$
\rho(\pi, L, a) = \frac{\int L(a, \theta)l(\theta)\pi(\theta)d\theta}{m_\pi(x)} = E_{\pi_x}[L(a, \theta)]
$$

• $\forall (L, \pi) \in \mathcal{L} \times \Gamma$, $a^*_{{(L,\pi)}}$ Bayes action corresponding to $(L, \pi)$ given by

$$
\rho(\pi, L, a^*_{{(L,\pi)}}) = \min_{a \in \mathcal{A}} \rho(\pi, L, a)
$$

• $B(L, \pi)$ set of all Bayes actions associated with pair $(L, \pi)$

• $a^*_{{(L,\pi)}}$ and $\bar{a}^*_{{(L,\pi)}}$: infimum and supremum of all Bayes actions
ORDERING BAYES ACTIONS

- $X \leq_{lr} Y \Rightarrow X \leq_{st} Y$
- $X \leq_{st} Y \Leftrightarrow E[g(X)] \leq E[g(Y)]$ for all increasing $g$ s.t. expectations exist
- $\Rightarrow E^{\pi_{h_1}}(g(\theta)) \leq E^{\pi'}(g(\theta)) \leq E^{\pi_{h_2}}(g(\theta)), \forall \pi' \in \Gamma_{h_1, h_2, \pi}$
- Same for posteriors
- $L(a, \theta) : \mathbb{R}^2 \rightarrow \mathbb{R}$ submodular function if for all $(a_1, \theta_1), (a_2, \theta_2) \in \mathbb{R}^2$
  \[ L(a_1, \theta_1) \lor L(a_2, \theta_2) \leq L(a_1, \theta_1) + L(a_2, \theta_2) \]
  \[ (a_1, \theta_1) \lor (a_2, \theta_2) = (\max\{a_1, a_2\}, \max\{\theta_1, \theta_2\}) \]
  \[ (a_1, \theta_1) \land (a_2, \theta_2) = (\min\{a_1, a_2\}, \min\{\theta_1, \theta_2\}) \]
- $L$ submodular $\Rightarrow -L$ supermodular
- $L$ twice differentiable $\Rightarrow$ submodular if $\frac{\partial^2 L(a, \theta)}{\partial \theta \partial a} \leq 0, \forall a, \theta$
- For $a_1$ and $a_2$ s.t. $a_1 \leq a_2 \Rightarrow L$ submodular $\Leftrightarrow L(a_2, \theta) - L(a_1, \theta)$ decreasing in $\theta$
CLASS OF CONVEX SUBMODULAR LOSS FUNCTIONS

- From now on, consider $L_{sm}$, class of convex submodular loss functions

- Widely used (classes of) loss functions included in $L_{sm}$
  - $\mathcal{L} = \{L_p(a, \theta) = |a - \theta|^p, p \geq 1\}$
    (absolute error loss for $p = 1$ and squared error loss for $p = 2$)
  - Quantile losses: $\mathcal{L} = \{L_q(a, \theta) = |a - \theta| + a(2q - 1), q \in [0, 1]\}$
    (absolute error loss for $q = 1/2$)
  - LINEX losses: $\mathcal{L} = \{L_k(a, \theta) = \exp(k(a - \theta)) - k(a - \theta) - 1, (k \neq 0)\}$
  - Linear losses $\mathcal{L} = \{L_{\alpha,\beta}(a, \theta) = \begin{cases} \alpha(a - \theta) & a \geq \theta \\ \beta(\theta - a) & a < \theta \end{cases}, \alpha, \beta > 0\}$
  - $\mathcal{L} = \{L_{\lambda(.)}(a, \theta) = \int_{0}^{\theta-a} \lambda(t)dt\}$,
    with $\lambda(t)$ positive (negative, null) if and only if $t > 0 (t < 0, t = 0)$ and $\lambda'(t) > 0$
  - $\mathcal{L} = \{L_{\phi(.)}(a, \theta) = \phi(\theta - a)\}$ for differentiable convex function $\phi$
ORDERING BAYES ACTIONS

Theorem.

- \( \pi \) prior distribution
- \( h_1 \) concave and \( h_2 \) convex distortion functions
- \( \Gamma_{h_1,h_2,\pi} \) corresponding distorted band
- \( B_{(L,\pi)} = [a^*_{(L,\pi)}, \overline{a}^*_{(L,\pi)}] \) set of all Bayes actions associated with pair \((L, \pi)\)
  - \( a^*_{(L,\pi)} \) and \( \overline{a}^*_{(L,\pi)} \): infimum and supremum of all Bayes actions
- \( \Rightarrow a^*_{(L,\pi_{h_1})} \leq a^*_{(L,\pi')} \leq a^*_{(L,\pi_{h_2})} \) and \( \overline{a}^*_{(L,\pi_{h_1})} \leq \overline{a}^*_{(L,\pi')} \leq \overline{a}^*_{(L,\pi_{h_2})} \)
  \( \forall L \in \mathcal{L}_{sm} \) and \( \forall \pi' \in \Gamma_{h_1,h_2,\pi} \), provided the set of Bayes actions is not empty
SAMPLE FROM POSTERIOR DISTORTED DISTRIBUTIONS

- Difficult in general to compute the exact distributions of the posterior distorted distributions
- $\pi$ with probability density function
- Differentiable distortion functions
- Posterior sample generated using acceptance-rejection method
EXAMPLE

- $X_1, X_2, \ldots, X_n$ i.i.d. r.v.'s $N(\theta, \sigma^2)$, with unknown mean $\theta$ and known variance $\sigma^2$
- $\pi: N(\mu, \tau^2)$ prior distribution on $\theta$
- Squared loss function $L_2(\theta, a) = (\theta - a)^2$
- Distorted class $\Gamma_{h_1, h_2, \pi}$ defined by skewed distributions
  
  $$
  h_1(x) = \int_{-\infty}^{F_{\pi}^{-1}(x)} 2\pi(\theta) F_{\pi}(-\beta \theta) d\theta 
  \text{ and } h_2(x) = \int_{-\infty}^{F_{\pi}^{-1}(x)} 2\pi(\theta) F_{\pi}(\beta \theta) d\theta, \forall \beta \geq 0
  $$

- Gaussian posterior $\pi_x$
  - mean $$(\sigma^2 \mu + n \bar{x} \tau^2) / (\sigma^2 + n \tau^2)$$
  - variance $$(\sigma^2 \tau^2) / (\sigma^2 + n \tau^2)$$
- Here: $\mu = 0$, $\tau^2 = 1$, $\sigma^2 = 1$ and $\beta = 1.2$ distortion parameter
- $\beta$ provides degree of distortion using Kolmogorov and Kantorovich metrics
  
  $$
  K(\pi_{h_2}, \pi_{h_1}) = 0.5577 \text{ and } KW(\pi_{h_2}, \pi_{h_1}) = 1.2259
  $$
## EXAMPLE

Range of Bayes actions (posterior means), $n = 1, \beta = 1.2$

<table>
<thead>
<tr>
<th>$x$</th>
<th>$E^{\pi_{x}}(\theta)$</th>
<th>$E^{\pi_{h_{2},x}}(\theta)$</th>
<th>$E^{\pi_{h_{1},x}}(\theta)$</th>
<th>$K(\pi_{h_{2},x}, \pi_{h_{1},x})$</th>
<th>$KW(\pi_{h_{2},x}, \pi_{h_{1},x})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.541</td>
<td>0.492</td>
<td>0.460</td>
<td>0.450</td>
<td>0.492</td>
</tr>
<tr>
<td>2</td>
<td>0.913</td>
<td>0.812</td>
<td>0.752</td>
<td>0.720</td>
<td>0.813</td>
</tr>
<tr>
<td>1</td>
<td>0.541</td>
<td>0.492</td>
<td>0.460</td>
<td>0.450</td>
<td>0.492</td>
</tr>
<tr>
<td>0</td>
<td>0.450</td>
<td>0.460</td>
<td>0.752</td>
<td>0.720</td>
<td>0.813</td>
</tr>
<tr>
<td>-1</td>
<td>0.460</td>
<td>0.450</td>
<td>0.720</td>
<td>0.752</td>
<td>0.813</td>
</tr>
<tr>
<td>-2</td>
<td>0.492</td>
<td>0.460</td>
<td>0.752</td>
<td>0.720</td>
<td>0.813</td>
</tr>
<tr>
<td>-3</td>
<td>0.541</td>
<td>0.492</td>
<td>0.460</td>
<td>0.450</td>
<td>0.492</td>
</tr>
</tbody>
</table>

- $KW(\pi_{h_{2},x}, \pi_{h_{1},x}) = E^{\pi_{h_{2},x}}(\theta) - E^{\pi_{x}}(\theta)$

- $\Rightarrow$ range of Bayes actions coincides with Kantorovich distance
EXAMPLE: POSTERIOR DISTRIBUTION

\[ \bar{x} = -1, \ n = 1 \]

\[ \bar{x} = 0, \ n = 1 \]

\[ \bar{x} = 1, \ n = 1 \]

\[ \bar{x} = -1, \ n = 10 \]

\[ \bar{x} = 0, \ n = 10 \]

\[ \bar{x} = 1, \ n = 10 \]
EXAMPLE: POSTERIOR MEAN

Uncertainty decreases as $n$ increases and increases as $\beta$ increases (red: upper bound ($h_2$), black: lower bound ($h_1$), blue: baseline ($\pi$))

$n = 1, \beta = 0.5$

$n = 1, \beta = 1.2$

$n = 1, \beta = 2.5$

$n = 10, \beta = 0.5$

$n = 10, \beta = 1.2$

$n = 10, \beta = 2.5$
FUTURE WORK

- Application to real case study, with proper choice of distortion functions

- $n$-dimensional model parameter, stemming from works by
  - Shaked and Shanthikumar (2007) on the definitions of multivariate likelihood ratio order and multivariate stochastic order and
  - Di Bernardino and Rulliere (2013) on the extension of the notion of distortion to the multivariate case