GENERALIZED MOMENT THEORY AND BAYESIAN ROBUSTNESS FOR HIERARCHICAL MIXTURE MODELS

B. Betrò, *Antonella Bodini*, A. Guglielmi

anto@mi.imati.cnr.it

CNR-IMATI, Milan (Italy)

A robustness problem

$$\{(\Omega, \mathcal{F}, P_\theta),\ \theta \in \Theta \subset \mathbb{R}^n\}\ , \qquad P_\theta \ll \lambda$$

$$g\ : \Theta \longrightarrow \mathbb{R}$$

$$\theta \sim \pi$$

A robustness problem

$$\{(\Omega, \mathcal{F}, P_\theta),\ \theta \in \Theta \subset \mathbb{R}^n\}\ , \qquad P_\theta \ll \lambda$$

$$g\ : \Theta \longrightarrow \mathbb{R}$$

$$\theta \sim \pi$$

$$\pi \in \Gamma$$

$$\sup_{\pi\in\Gamma} \pi(g(\theta)|x_1, \ldots, x_n) - \inf_{\pi\in\Gamma} \pi(g(\theta)|x_1, \ldots, x_n)$$

A robustness problem

$$\{(\Omega, \mathcal{F}, P_\theta),\ \theta \in \Theta \subset \mathbb{R}^n\}\ , \qquad P_\theta \ll \lambda$$

$$g\ : \Theta \longrightarrow \mathbb{R}$$

$$\theta \sim \pi$$

$$\Gamma = \{\pi \in \mathbb{P}(\Theta) : \underbrace{\int_\Theta f_i(\theta)\pi(d\theta)}_{\text{GENERALIZED MOMENT}} \le \alpha_i,\ i = 1, \ldots, m\}$$

B. Betrò, M. Męczarski & F. Ruggeri, 1994

B. Betrò & A. Guglielmi; 1994, 2000

A nonparametric case

$$\{(\Omega, \mathcal{F}, P_\pi),\ \pi \in \mathbb{S} \subset \mathbb{P}(Y)\},\ P_\pi \ll \lambda$$

$$\frac{dP_\pi}{d\lambda} = p(x;\pi) = \int_Y k(x;y)\pi(dy),\ Y \subset \mathbb{R}^n$$

$k \ : \ \mathbb{R}^k \times Y \to \mathbb{R}^+$ transition prob. density

A nonparametric case

$$\{(\Omega, \mathcal{F}, P_\pi),\ \pi \in \mathbb{S} \subset \mathbb{P}(Y)\},\ P_\pi \ll \lambda$$

$$\frac{dP_\pi}{d\lambda} = p(x;\pi) = \int_Y k(x;y)\pi(dy),\ Y \subset \mathbb{R}^n$$

$k\ :\ \mathbb{R}^k \times Y \to \mathbb{R}^+$ transition prob. density

Remark: $\mathbb{S} = \{\delta_y\ ;\ y \in Y\} \Rightarrow$ parametric case.

A nonparametric case

$$\{(\Omega, \mathcal{F}, P_\pi),\ \pi \in \mathbb{S} \subset \mathbb{P}(Y)\},\ P_\pi \ll \lambda$$

$$\frac{dP_\pi}{d\lambda} = p(x;\pi) = \int_Y k(x;y)\pi(dy),\ Y \subset \mathbb{R}^n$$

$k\ :\ \mathbb{R}^k \times Y \to \mathbb{R}^+$ transition prob. density

$\tilde{\pi}\ :\ (\Omega, \mathcal{F}, P) \to (\mathbb{S},\ q), \quad \mathbb{S} \subseteq \mathbb{P}(Y)$ (r.p.m.)

A nonparametric case

$$\{(\Omega, \mathcal{F}, P_\pi),\ \pi \in \mathbb{S} \subset \mathbb{P}(Y)\},\ P_\pi \ll \lambda$$

$$\frac{dP_\pi}{d\lambda} = p(x;\pi) = \int_Y k(x;y)\pi(dy),\ Y \subset \mathbb{R}^n$$

$k\ :\ \mathbb{R}^k \times Y \to \mathbb{R}^+$ transition prob. density

$\tilde{\pi}\ :\ (\Omega, \mathcal{F}, P) \to (\mathbb{S},\ q), \quad \mathbb{S} \subseteq \mathbb{P}(Y)$ (r.p.m.)

$$\Gamma = \{q \in \mathbb{P}(\mathbb{S}) : \int_{\mathbb{S}} f_i(\pi) q(d\pi) \le \alpha_i,\ i = 1, \ldots, m\}$$

the hierarchical model

$$X_1, \ldots, X_r \mid Y_1, \ldots, Y_r \text{ independent}$$

$$X_i \mid Y_i \sim \mathcal{L}(X_i | Y_i)$$

$$Y_1, \ldots, Y_r \mid \tilde{\pi} \quad \text{i.i.d.,} \qquad Y_1 | \tilde{\pi} = \pi \sim \pi$$

$$\tilde{\pi} \sim q$$

the hierarchical model

$$X_1, \ldots, X_r \mid Y_1, \ldots, Y_r \text{ independent}$$

$$X_i \mid Y_i \sim \mathcal{L}(X_i | Y_i)$$

$$Y_1, \ldots, Y_r \mid \tilde{\pi} \quad \text{i.i.d.,} \qquad Y_1 | \tilde{\pi} = \pi \sim \pi$$

$$\tilde{\pi} \sim q$$

$\tilde{\pi}$ = Dirichlet process $\Rightarrow$
mixture of Dirichlet process models (Lo, 1984)

the hierarchical model

$$X_1, \ldots, X_r \mid Y_1, \ldots, Y_r \text{ independent}$$

$$X_i \mid Y_i \sim \mathcal{L}(X_i|Y_i)$$

$$Y_1, \ldots, Y_r \mid \tilde{\pi} \quad \text{i.i.d.,} \qquad Y_1|\tilde{\pi} = \pi \sim \pi$$

$$\tilde{\pi} \sim q$$

$\tilde{\pi}$ = Dirichlet process $\Rightarrow$
mixture of Dirichlet process models (Lo, 1984)

- $x \mapsto k(x, y)$ defines the density of $X_i|Y_i = y$

the hierarchical model

$$X_1, \ldots, X_r \mid Y_1, \ldots, Y_r \text{ independent}$$

$$X_i \mid Y_i \sim \mathcal{L}(X_i | Y_i)$$

$$Y_1, \ldots, Y_r \mid \tilde{\pi} \quad \text{i.i.d.,} \qquad Y_1 | \tilde{\pi} = \pi \sim \pi$$

$$\tilde{\pi} \sim q$$

$\tilde{\pi}$ = Dirichlet process $\Rightarrow$
mixture of Dirichlet process models (Lo, 1984)

- $x \mapsto k(x, y)$ defines the density of $X_i | Y_i = y$
- $q \in \Gamma$

A simple example, I

$$X_i \mid Y_i \sim bern(Y_i)$$

$$k(x;y) = y^x(1-y)^{1-x},\ (x,y) \in \{0,1\} \times [0,1],$$

$$Y_i | \tilde{\pi} = \pi \sim \pi \in \mathbb{S} = \mathbb{P}([0,1])$$

$\tilde{\pi} \sim q \in \mathbb{P}(\mathbb{S})$ such that

$P(X_1 = 1) = \int_{\mathbb{S}} \left(\int_0^1 y\pi(dy) \right) q(d\pi) \leq \alpha_1$, or

$q(\mathbb{S}_2) = \int_{\mathbb{S}} I_{\mathbb{S}_2}(\pi) q(d\pi) \leq \alpha_2$, $\mathbb{S}_2 \subset \mathbb{S}$

$$\Rightarrow \Gamma$$

the objective function...

- $g(\pi) = I_{\mathbb{S}_1}(\pi), \mathbb{S}_1 \in \mathcal{B}(\mathbb{S}) \;\Rightarrow$

$$\sup_{q \in \Gamma} \; q(\tilde{\pi} \in \mathbb{S}_1 \mid \underline{x})$$

the objective function...

- $g(\pi) = I_{\mathbb{S}_1}(\pi), \mathbb{S}_1 \in \mathcal{B}(\mathbb{S}) \;\Rightarrow$

$$\sup_{q \in \Gamma} \; q(\tilde{\pi} \in \mathbb{S}_1 \mid \underline{x})$$

- $g(\pi) = \pi(A), \quad A \in \mathcal{B}(Y) \;\Rightarrow$

$$\sup_{q \in \Gamma} \; E_q(\tilde{\pi}(A) \mid \underline{x})$$

...and the constraints

- $Y \subset \mathbb{R}$, $\quad f_i(\pi) = \int_Y y^i \pi(dy) \Rightarrow$

$$\int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{S}} \Big(\int_Y y^i \pi(dy) \Big) q(d\pi)$$

...and the constraints

- $Y \subset \mathbb{R}$, $\quad f_i(\pi) = \int_Y y^i \pi(dy) \Rightarrow$

$$\int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{S}} \Big(\int_Y y^i \pi(dy) \Big) q(d\pi)$$

- $f_i(\pi) = P_\pi(C_i),\ C_i \in \mathcal{B}(\mathbb{R}^k) \quad \Rightarrow$
$\int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{S}} P_\pi(C_i) q(d\pi) = P(X_1 \in C_i)$

...and the constraints

- $Y \subset \mathbb{R}, \quad f_i(\pi) = \int_Y y^i \pi(dy) \Rightarrow$

$$\int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{S}} \Big(\int_Y y^i \pi(dy) \Big) q(d\pi)$$

- $f_i(\pi) = P_\pi(C_i), \; C_i \in \mathcal{B}(\mathbb{R}^k) \quad \Rightarrow$
  $\int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{S}} P_\pi(C_i) q(d\pi) = P(X_1 \in C_i)$

- $X \; : \; \Omega \to \mathbb{R}, \quad f_i(\pi) = \int_{\mathbb{R}} x^i P_\pi(dx)$
  $\Rightarrow \int_{\mathbb{S}} f_i(\pi) q(d\pi) = \int_{\mathbb{R}} x^i m_X(x) \lambda(dx)$

The problem

$l(\pi)$ = likelihood

$$\sup_{q \in \Gamma} \frac{\int_{\mathbb{S}} g(\pi) l(\pi) q(d\pi)}{\int_{\mathbb{S}} l(\pi) q(d\pi)}$$

The problem

$l(\pi)$ = likelihood

$$\sup_{q \in \Gamma} \frac{\int_{\mathbb{S}} g(\pi) l(\pi) q(d\pi)}{\int_{\mathbb{S}} l(\pi) q(d\pi)}$$

is NOT LINEAR in q

hint

Betrò & Guglielmi (1994):

The map

$$\psi \ : \ \mathbb{P}(\mathbb{S}) \to \mathcal{M}(\mathbb{S}) = \{\underline{\textit{finite}} \textit{ measures on } \mathbb{S}\}$$

$$\psi(q)(A) =: \mu(A) = \frac{q(A)}{\int_{\mathbb{S}} l(\pi) q(d\pi)} \quad (A \in \mathcal{B}(\mathbb{S}))$$

is injective and

$\int_{\mathbb{S}} l(\pi)\mu(d\pi) = 1 \Leftrightarrow \mu \in \psi(\mathbb{P}(\mathbb{S}))$, *provided*
$0 < \int_{\mathbb{S}} l(\pi) q(d\pi) < +\infty$.

toward a solution

$$\sup_{q \in \Gamma} \frac{\int_{\mathbb{S}} g(\pi) l(\pi) q(d\pi)}{\int_{\mathbb{S}} l(\pi) q(d\pi)}$$

$$\Updownarrow$$

$$\sup_{\mu \in \mathcal{M}_1} \int_{\mathbb{S}} g(\pi) l(\pi) \mu(d\pi)$$

$$\mathcal{M}_1 = \{\mu \in \mathcal{M}(\mathbb{S}) :$$

$$\int_{\mathbb{S}} \left(f_i(\pi) - \alpha_i\right) \mu(d\pi) \leq 0,\ i = 1, \ldots, m;\ \int_{\mathbb{S}} l(\pi) \mu(d\pi) = 1\}$$

toward a solution

$$\sup_{q\in\Gamma} \frac{\int_{\mathbb{S}} g(\pi)l(\pi)q(d\pi)}{\int_{\mathbb{S}} l(\pi)q(d\pi)}$$

$$\Updownarrow$$

$$\sup_{\mu\in\mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi)$$

$$\mathcal{M}_1 = \{\mu \in \mathcal{M}(\mathbb{S}) :$$

$$\int_{\mathbb{S}} (f_i(\pi) - \alpha_i)\mu(d\pi) \leq 0,\ i = 1, \ldots, m;\ \int_{\mathbb{S}} l(\pi)\mu(d\pi) = 1\}$$

GENERALIZED MOMENT PROBLEM

KEMPERMAN 1971, 1983, 1987.

The main result

$\mathbb{S}$ separable

Theorem.

Under the hypotheses 1-7, if $\mathcal{M}_1 \neq \emptyset$ then

$$\sup_{\mu\in\mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi) =$$
$$= \inf\left\{\beta_0 \in \mathbb{R} : \exists\, \beta \in \mathbb{R}^m_+ \text{ such that } \beta_0 l(\pi) + \sum_{i=1}^m \beta_i(f_i(\pi) - \alpha_i) \geq g(\pi)l(\pi) \ \forall\, \pi \in \mathbb{S}\right\}.$$

The supremum is attained and finite.

The main result

$\mathbb{S}$ separable

Theorem.

Under the hypotheses 1-7, if $\mathcal{M}_1 \neq \emptyset$ *then*

$$\sup_{\mu \in \mathcal{M}_1} \int_{\mathbb{S}} g(\pi) l(\pi) \mu(d\pi) =$$
$$= \inf \Big\{ \beta_0 \in \mathbb{R} : \exists\, \beta \in \mathbb{R}^m_+ \text{ such that } \beta_0 l(\pi) + \sum_{i=1}^m \beta_i (f_i(\pi) - \alpha_i) \geq g(\pi) l(\pi) \;\; \forall\, \pi \in \mathbb{S} \Big\}.$$

The supremum is attained and finite.

LINEAR SEMI-INFINITE PROGRAMMING.

some history

Theorem 5

Kemperman 1983

$\Downarrow$

Theorem

$\Bigg\}$ generalized moment theory

$\Updownarrow$ robustness framework

Our main result

some history

Theorem 5		
Kemperman 1983	}	generalized
$\Downarrow$		moment
Theorem		theory

$\Updownarrow$ robustness framework

Our main result

http://www.mi.imati.cnr.it/iami/abstracts/02-06.html

hypotheses

1 f_j l.s.c., $j = 1, \ldots, m$; l continuous;

2 g is u.s.c.;

3 $\forall\, j \; \exists\, C_j < 0 \;:\; f_j(\pi) \geq C_j \;\forall\, \pi \in \mathbb{S}$;

4 l bounded;

5 $\exists\, C \;: g(\pi)l(\pi) \leq C \;\forall\, \pi \in \mathbb{S}$;

6 $\exists\, \tilde{\beta}_0 \in \mathbb{R}_+$, $\tilde{\beta} \in \mathbb{R}_+^{m+1}$: $\forall\, \pi \in \mathbb{S}$
$\tilde{\beta}_0 l(\pi) + \sum_{i=1}^m \tilde{\beta}_i (f_i(\pi) - \alpha_i) - \tilde{\beta}_{m+1} g(\pi) l(\pi) > 1$

7 $\forall\, \varepsilon > 0 \;\exists\, K_\varepsilon$, a compact s. of $\mathbb{S}$, and B, $\beta_i^\varepsilon \geq 0$, $i = 0, \ldots, m+1$, such that $\sum_{i=0}^{m+1} \beta_i^\varepsilon \leq B$ and $\forall\, \pi \in K_\varepsilon^c$ we have:

$$\beta_0^\varepsilon l(\pi) + \sum_{i=1}^m \beta_i^\varepsilon (f_i(\pi) - \alpha_i) - \beta_{m+1}^\varepsilon g(\pi) l(\pi) > \frac{1}{\varepsilon}$$

An optimal solution

$$\int_{\mathbb{S}} g(\pi)l(\pi)\mu^*(d\pi) = \sup_{\mu\in\mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi)$$

An optimal solution

$$\int_{\mathbb{S}} g(\pi)l(\pi)\mu^*(d\pi) = \sup_{\mu \in \mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi)$$

$$\Rightarrow \ \mu^* = \sum_{i=1}^{k} \mu_i^* \delta_{\pi_i^*}, \quad \mu_i^* > 0,\ \pi_i^* \in \mathbb{S} = \mathbb{P}(Y),\ k \leq m+1$$

(Rogosinsky, 1958)

An optimal solution

$$\int_{\mathbb{S}} g(\pi)l(\pi)\mu^*(d\pi) = \sup_{\mu\in\mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi)$$

$$= \inf\Big\{\beta_0 \in \mathbb{R} : \exists\, \beta \in \mathbb{R}^m_+ \text{ such that}$$

$$\beta_0 l(\pi) + \sum_{i=1}^m \beta_i (f_i(\pi) - \alpha_i) \geq g(\pi)l(\pi) \;\; \forall\, \pi \in \mathbb{S}\Big\} =: \beta_0^*$$

An optimal solution

$$\int_{\mathbb{S}} g(\pi)l(\pi)\mu^*(d\pi) = \sup_{\mu\in\mathcal{M}_1} \int_{\mathbb{S}} g(\pi)l(\pi)\mu(d\pi)$$

$$= \inf\Big\{\beta_0 \in \mathbb{R} : \exists\, \beta \in \mathbb{R}^m_+ \text{ such that}$$

$$\beta_0 l(\pi) + \sum_{i=1}^m \beta_i(f_i(\pi) - \alpha_i) \geq g(\pi)l(\pi) \ \ \forall\, \pi \in \mathbb{S}\ \Big\} =: \beta_0^*$$

the optimal measure $\mu^* = \sum_{i=1}^k \mu_i^* \delta_{\pi_i^*}$ satisfies

$$\begin{cases} \beta_0^* l(\pi_j^*) + \sum_{i=1}^m \beta_i^*(f_i(\pi_j^*) - \alpha_i) = g(\pi_j^*)l(\pi_j^*), \ \ \forall\, j \\ \sum_{j=1}^k (f_i(\pi_j^*) - \alpha_i)\mu_j^* = 0 \ \ \forall\, i \ : \ \beta_i^* > 0 \end{cases}$$

and viceversa

A simple example, II

$$X_i \mid Y_i \;\sim\; bern(Y_i)$$

$$Y_i \mid \tilde{\pi} = \pi \sim \pi \in \mathbb{S} = \mathbb{P}([0,1])$$

$$\pi_1 \;:\; \int_0^1 y\pi_1(dy) = 1/4; \qquad \pi_2 \;:\; \int_0^1 y\pi_2(dy) = 3/8$$

$$\mathbb{S}_1 = \{(1-\gamma)\pi_1 + \gamma\pi_2; \gamma \in [0,1]\}$$

$$\mathbb{S}_2 = \{(1-\gamma)\pi_1 + \gamma\pi_2; \gamma \in [1/4, 3/4]\}$$

$$\Gamma = \{q : \; P(X_1 = 1) = \int_{\mathbb{S}} \int_0^1 y\pi(dy) q(d\pi) = 0.5;$$

$$q(\mathbb{S}_2) = \int_{\mathbb{S}} I_{\mathbb{S}_2}(\pi) q(d\pi) \geq 1/8\}$$

$$f_1(\pi) = \int_0^1 y\pi(dy) = -f_2(\pi); \; f_3(\pi) = -I_{\mathbb{S}_2}(\pi)$$

cont.

$$\sup_\Gamma \; q(\tilde\pi \in \mathbb{S}_1 | x_1 = 1, x_2 = 1) \leftrightarrow g(\pi) = I_{\mathbb{S}_1}(\pi)$$

μ_i^*	0.6546	0.3967	2.1226
π_i^*	δ_1	$0.25\pi_1 + 0.75\pi_2$	π_2

$\Leftrightarrow 0.3454$

cont.

$$\sup_{\Gamma}\ q(\tilde{\pi} \in \mathbb{S}_1 | x_1 = 1, x_2 = 1) \leftrightarrow g(\pi) = I_{\mathbb{S}_1}(\pi)$$

μ_i^*	0.6546	0.3967	2.1226
π_i^*	δ_1	$0.25\pi_1 + 0.75\pi_2$	π_2

$\Leftrightarrow 0.3454$

$\inf_{\Gamma}$: $-I_{\mathbb{S}_1}(\pi)$ is not u.s.c. !!

μ_i^*	0.6546	0.0793	2.1226
π_i^*	δ_1	$0.25\pi_1 + 0.75\pi_2$	π_2

$\Leftrightarrow 0.0208$

References

- Betrò, B. and Guglielmi, A. (1994). An algorithm for robust Bayesian analysis under generalized moment conditions. Quaderno IAMI, 94.6, CNR-IAMI, Milan.
- Betrò, B. and Guglielmi, A. (2000). Methods for global prior robustness under generalized moment conditions. *Robust Bayesian analysis, Lecture Notes in Statistics*, Vol. 152, (eds. D. Rios Insua, F. Ruggeri), 273–294, Springer, New York.
- Betrò, B., Męczarski, M. and Ruggeri, F. (1994). Robust Bayesian analysis under generalized moments conditions. *Journal of Statistical Planning and Inference*, **41**, 257-266.
- Kemperman, J.H.B. (1983). On the role of duality in the theory of moments, *Lecture Notes in Economics and Mathematical Systems*, Vol. 215, 63-92, Springer-Verlag, Berlin.
- Lo, A. Y. (1984). On a class of bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* **12**, 351-357.
- Rogosinsky, W.W. (1958). Moments of non-negative mass, *Proceedings of the Royal Society of London*, Ser. A, **245**, 1-27.