On Probabilistic Parametric Inference

Tomaž Podobnik[1,2] and Tomi Živko[2],

[1]Faculty of Mathematics and Physics, UNI-Ljubljana,

[2]Jožef Stefan Institute, Ljubljana, Slovenia

Second Workshop on Principles and Methods of Statistical Inference with Interval Probability,

München, September 09-15 2009

What is Probabilistic Parametric Inference (PPI)?

$I = \{ \Pr_I(\bullet \mid \theta) : \theta \in V_\Theta \}$; $V_\Theta \subseteq \mathbb{R}^m$ = parameter space

$\Pr_I(\bullet \mid \theta) : \mathfrak{B}^n \to [0,1]$; direct probability distribution

$\Pr_I(B \mid \theta) = \int_B f_I(\mathbf{x} \mid \theta)\, d^n\mathbf{x},\ B \in \mathfrak{B}^n$; continuous distribution

PPI:

$\Pr_I(D \mid \mathbf{x})$; $D \in \mathfrak{B}^m$ $(D \subseteq V_\Theta)$; inverse probability distribution

$\Pr_I(\bullet \mid \mathbf{x}) : \mathfrak{B}^m \to [0,1]$; a special (degenerate) case of interval probabilities, $\underline{P} = \overline{P}$ (Luo and Zhang, 1997)

$\Pr_I(D \mid \mathbf{x}) = \int_D f_I(\theta \mid \mathbf{x})\, d^m\theta,\ D \in \mathfrak{B}^m$; continuous distribution

Who does it?

- Bayesian schools of statistics (as opposed to frequentist schools).
- “In the Bayesian paradigm, it is also possible to make statements regarding the values of the inferred parameters in the absence of data and these statements can be summarized by prior distributions.” (Villegas, 1981)

$f_I(\Theta)$: (non-informative) prior istribution

$$\Rightarrow f_I(\Theta)\, f_I(\mathbf{x}\,|\,\Theta) = f_I(\Theta,\mathbf{x}) = f_I(\mathbf{x})\, f_I(\Theta\,|\,\mathbf{x})$$

$$\Rightarrow f_I(\Theta\,|\,\mathbf{x}) = \frac{f_I(\Theta)\, f_I(\mathbf{x}\,|\,\Theta)}{f_I(\mathbf{x})}$$; (Bayes Theorem)

$$f_I(\mathbf{x}) = \int_{V_\Theta} f_I(\Theta)\, f_I(\mathbf{x}\,|\,\Theta)\, d^m\Theta$$; predictive distribution

Why PPI?

Probabilistic inference is an ideal (Paris, 1994, p. 33): under very general and intuitively appealing assumptions, a system of inference must be identical (isomorphic) to the probability system (The Dutch Book Theorem, Cox's Theorem; recall also M. Goldstein's talk earlier this morning).

But:

- Intuitively appealing (acceptable, natural, in accordance with common sense, perfectly sensible, tempting, meaningful, convincing, justifiable…) **to whom**?
- There are arguments for preferring other kinds of inference, e.g., interval probabilities, Dempster-Shafer belief functions, truth-functional belief (fuzzy logic), …

Why not PPI?

"Each of these theories is based on some system of postulates, and so long as the **postulates** forming one particular system **do not contradict each other** and are sufficient to construct a theory, this is as legitimate as any other." (Neyman, 1937)

1. Every axiomatic system, be it empirical or non-empirical, must be consistent: if, within the system of axioms, a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.

2. "In order to make the theory operational, we must introduce a concept that links mathematics to an external world of measurable phenomena." (Stuart and Ord, 1994)

"The most striking achievement of the physical sciences is prediction." (Pólya, 1954, p. 64)

How to determine (assign,…) $f_I(\theta)$ and preserve consistency of the system ("the quest for the Holy Grail", Fienberg, 2006)?

"A succession of authors have said that the prior probability is a mistake and therefore that the principle of inverse probability, which cannot work without it, is nonsense too." (Jeffreys, 1961, p. 120)

"During the rapid development of practical statistics in the past few decades, the theoretical foundations of the subject have been involved in great obscurity. This obscurity is centred in the so-called inverse methods. …The inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply commited itself)." (Fisher, 1922)

How to make the theory operational?

1. Basic Definitions and Immediate Consequences
2. Objectivity
3. Calibration
4. Interpretations of Probability Distributions
5. Discussion and Conclusions

} mathematics (1–3)

} operationality (4)

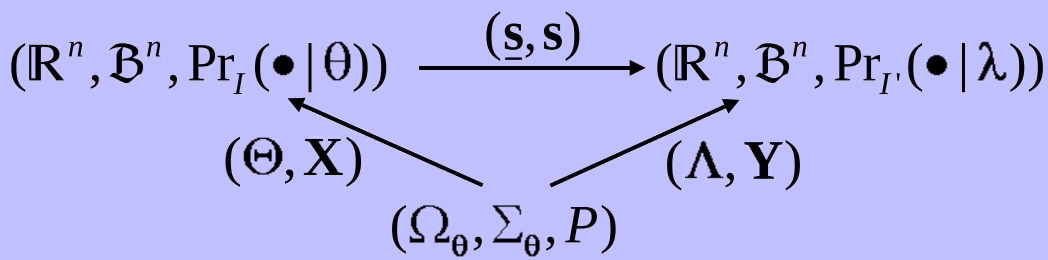
1. Basic Definitions and Immediate Consequences:

Each (direct) probability distribution $\Pr_I(\bullet\,|\,\theta)$ from a parametric family $I=\{\Pr_I(\bullet\,|\,\theta):\theta\in V_{\Theta}\}$ is underlain by a probability space $(\Omega_{\boldsymbol\theta},\Sigma_{\boldsymbol\theta},P)$ and a $\Sigma_{\boldsymbol\theta}$-measurable function $(\Theta,\mathbf{X}):\Omega_{\boldsymbol\theta}\to(\theta,\mathbb{R}^n)$, $A_{\mathbf{X}\le\mathbf{x}}\equiv\{\omega\in\Omega_{\boldsymbol\theta}:\mathbf{X}(\omega)\le\mathbf{x}\}\in\Sigma_{\boldsymbol\theta}\,;\,\mathbf{x}\in\mathbb{R}^n$, called random variable, such that $\Pr_I(\bullet\,|\,\theta)\equiv P\circ(\Theta,\mathbf{X})^{-1}$.

$F_I(\mathbf{x}\,|\,\theta)\equiv\Pr_I(A_{\mathbf{X}\le\mathbf{x}}\,|\,\theta)$; (cumulativ e) distribution function (cdf)

Remark 1. The notion of chance (‘physical randomness’) is avoided in the definition of random variable.

Corrolary. Let $(\mathbf{\Theta},\mathbf{X})$ and $(\mathbf{\Lambda},\mathbf{Y})= (\underline{\mathbf{s}}\ \mathbf{\Theta},\mathbf{s}\ \mathbf{X})$ be continuous random variables, defined on a probability space $(\Omega_{\boldsymbol{\theta}},\Sigma_{\boldsymbol{\theta}},P)$, let $\underline{\mathbf{s}}$ be one-to-one, and let $\left|\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})\right| > 0$.



Then,

$$f_{I'}(\mathbf{y}\,|\,\boldsymbol{\lambda}) = f_I[\mathbf{s}^{-1}(\mathbf{y})\,|\,\underline{\mathbf{s}}^{-1}(\boldsymbol{\lambda})]\left|\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})\right|. \qquad (1)$$

For scalar X and Y we also have

$$F_{I'}(y\,|\,\lambda) = \begin{cases} F_I[s^{-1}(y)\,|\,\underline{\mathbf{s}}^{-1}(\lambda)] & ; s'(x) > 0 \\ 1 - F_I[s^{-1}(y)\,|\,\underline{\mathbf{s}}^{-1}(\lambda)]; & s'(x) < 0 \end{cases}, \qquad (2)$$

Let $\mathbf{X} \equiv (\mathbf{Y}, \mathbf{Z})$ and $f_I(\mathbf{z} \mid \theta) \equiv \int_{\mathbb{R}^{n_1}} f_I(\mathbf{y}, \mathbf{z} \mid \theta)\, d^{n_1}\mathbf{y} > 0$.

$$\Rightarrow f_I(\mathbf{y} \mid \theta, \mathbf{z}) \equiv \frac{f_I(\mathbf{y}, \mathbf{z} \mid \theta)}{f_I(\mathbf{z} \mid \theta)}$$; conditional probability distribution

Consistent definition (Kolmogorov) (Rao, 1993, pp. 25-26, 29-30, and pp. 51-54).

Borel-Kolmogorov paradox (Kolmogorov, 1933, pp. 44-45; Rao, 1993, pp. 65-66).

$\exists\,(\Omega_{\mathbf{x}}, \Sigma_{\mathbf{x}}, P)\,;$

$\exists\,(\Theta, \mathbf{X}) : \Omega_{\mathbf{x}} \to (\mathbb{R}^m, \mathbf{x})\,,\, C_{\Theta \le \theta} \equiv \{\omega \in \Omega_{\mathbf{x}} : \Theta(\omega) \le \theta\} \in \Sigma_{\mathbf{x}}\,;\, \theta \in \mathbb{R}^m\,;$

$\Rightarrow \Pr_I(\bullet \,|\, \mathbf{x}) \equiv P \circ (\Theta, \mathbf{X})^{-1}$; inverse probability distribution

$F_I(\theta \,|\, \mathbf{x}) \equiv \Pr_I(C_{\Theta \le \theta} \,|\, \mathbf{x})$

Let $\Theta \equiv (\Theta_1, \Theta_2)$ and $f_I(\theta_2 \,|\, \mathbf{x}) \equiv \int_{\mathbb{R}^{m_1}} f_I(\theta_1, \theta_2 \,|\, \mathbf{x})\, d^{m_1}\theta_1 > 0\,.$

$\Rightarrow f_I(\theta_1 \,|\, \theta_2, \mathbf{x}) \equiv \dfrac{f_I(\theta_1, \theta_2 \,|\, \mathbf{x})}{f_I(\theta_2 \,|\, \mathbf{x})}$; conditional probability distribution

Random variables $(\Theta, \mathbf{X})$ and $(\Lambda, \mathbf{Y}) = (\underline{\mathbf{s}} \circ \Theta, \mathbf{s} \circ \mathbf{X})$ on a common probability space $(\Omega_{\mathbf{x}}, \Sigma_{\mathbf{x}}, P)$, s one-to-one, $\left|\partial_{\boldsymbol{\lambda}} \underline{\mathbf{s}}^{-1}(\lambda)\right| > 0\,.$

$$\Rightarrow f_{I'}(\lambda \,|\, \mathbf{y}) = f_I[\underline{\mathbf{s}}^{-1}(\lambda) \,|\, \mathbf{s}^{-1}(\mathbf{y})] \left|\partial_{\boldsymbol{\lambda}} \underline{\mathbf{s}}^{-1}(\lambda)\right|. \qquad (3)$$

Proposition 1.

$f_I(\mathbf{x}_1,\mathbf{x}_2\,|\,\theta) = f_I(\mathbf{x}_1\,|\,\theta)\, f_I(\mathbf{x}_2\,|\,\theta)$ ($\mathbf{X}_1$ and $\mathbf{X}_2$ i.i.d.);

$\exists\; f_I(\theta\,|\,\mathbf{x}_1,\mathbf{x}_2)$, $f_I(\theta\,|\,\mathbf{x}_1)$, and $f_I(\theta\,|\,\mathbf{x}_2)$;

$\exists\; f_I(\theta,\mathbf{x}_1\,|\,\mathbf{x}_2), \exists\; f_I(\theta,\mathbf{x}_2\,|\,\mathbf{x}_1)$;

$$f_I(\theta\,|\,\mathbf{x}_2) = \int_{\mathbb{R}^{n_1}} f_I(\theta,\mathbf{x}_1\,|\,\mathbf{x}_2)\, d^{n_1}\mathbf{x}_1 > 0,\; f_I(\mathbf{x}_1\,|\,\mathbf{x}_2) \equiv \int_{\mathbb{R}^m} f_I(\theta,\mathbf{x}_1\,|\,\mathbf{x}_2)\, d^m\theta > 0;$$

$$f_I(\theta\,|\,\mathbf{x}_1) = \int_{\mathbb{R}^{n_1}} f_I(\theta,\mathbf{x}_2\,|\,\mathbf{x}_1)\, d^{n_2}\mathbf{x}_2 > 0,\; f_I(\mathbf{x}_2\,|\,\mathbf{x}_1) \equiv \int_{\mathbb{R}^m} f_I(\theta,\mathbf{x}_2\,|\,\mathbf{x}_1)\, d^m\theta > 0;$$

$$\Rightarrow f_I(\theta\,|\,\mathbf{x}_1,\mathbf{x}_2) = \frac{f_I(\theta\,|\,\mathbf{x}_2)\, f_I(\mathbf{x}_1\,|\,\theta)}{f_I(\mathbf{x}_1\,|\,\mathbf{x}_2)} = \frac{f_I(\theta\,|\,\mathbf{x}_1)\, f_I(\mathbf{x}_2\,|\,\theta)}{f_I(\mathbf{x}_2\,|\,\mathbf{x}_1)}. \qquad (4)$$

Remark 2. $f_I(\theta, \mathbf{x}_1 \mid \mathbf{x}_2)$ and $f_I(\theta, \mathbf{x}_2 \mid \mathbf{x}_1)$ neither purely direct nor purely inverse.

Remark 3. Existence of $f_I(\theta)$ (of $f_I(\theta, \mathbf{x})$) has **not** been assumed.

Proposition 1 (cont'd).

$$f_I(\theta_1 \mid \theta_2, \mathbf{x}_1, \mathbf{x}_2) = \frac{f_I(\theta_1 \mid \theta_2, \mathbf{x}_2)\, f_I(\mathbf{x}_1 \mid \theta_1, \theta_2)}{f_I(\mathbf{x}_1 \mid \mathbf{x}_2)} = \frac{f_I(\theta_1 \mid \theta_2, \mathbf{x}_1)\, f_I(\mathbf{x}_2 \mid \theta_1, \theta_2)}{f_I(\mathbf{x}_2 \mid \mathbf{x}_1)}. \quad (5)$$

Proposition 2 (Factorization).

a) $f_I(\theta|\mathbf{x}_1) = \dfrac{\zeta_I(\theta)\, f_I(\mathbf{x}_1|\theta)}{\eta_I(\mathbf{x}_1)}$; $\eta_I(\mathbf{x}_1) = \int_{\mathbb{R}^m} \zeta_I(\theta)\, f_I(\mathbf{x}_1|\theta)\, d^m\theta$,

b) $f_I(\theta_1|\theta_2,\mathbf{x}_1) = \dfrac{\zeta_{I,\boldsymbol{\theta}_2}(\theta_1)\, f_I(\mathbf{x}_1|\theta_1,\theta_2)}{\eta_{I,\boldsymbol{\theta}_2}(\mathbf{x}_1)}$; $\eta_{I,\boldsymbol{\theta}_2}(\mathbf{x}_{1,}) = \int_{\mathbb{R}^{m_1}} \zeta_{I,\boldsymbol{\theta}_2}(\theta_1)\, f_I(\mathbf{x}_1|\theta_1,\theta_2)\, d^{m_1}\theta_1$.

Proof. $\dfrac{f_I(\theta|\mathbf{x}_2)\, f_I(\mathbf{x}_1|\theta)}{f_I(\mathbf{x}_1|\mathbf{x}_2)} = \dfrac{f_I(\theta|\mathbf{x}_1)\, f_I(\mathbf{x}_2|\theta)}{f_I(\mathbf{x}_2|\mathbf{x}_1)}$ (Eq. (4))

$$\underbrace{\frac{f_I(\theta|\mathbf{x}_2)}{f_I(\mathbf{x}_2|\theta)}}_{\kappa(\theta,\mathbf{x}_2)} \frac{f_I(\mathbf{x}_1|\theta)}{f_I(\theta|\mathbf{x}_1)} = \underbrace{\frac{f_I(\mathbf{x}_1|\mathbf{x}_2)}{f_I(\mathbf{x}_2|\mathbf{x}_1)}}_{h(\mathbf{x}_1,\mathbf{x}_2)} \Rightarrow \frac{\kappa(\theta,\mathbf{x}_2)}{\kappa(\theta,\mathbf{x}_1)} = h(\mathbf{x}_1,\mathbf{x}_2) \Rightarrow h(\mathbf{x}_1,\mathbf{x}_2) = \frac{\eta_I(\mathbf{x}_1)}{\eta_I(\mathbf{x}_2)}$$

$$\Rightarrow \kappa(\theta,\mathbf{x}_2)\eta_I(\mathbf{x}_2) = \kappa(\theta,\mathbf{x}_1)\eta_I(\mathbf{x}_1) \Rightarrow \kappa(\theta,\mathbf{x}_1)\eta_I(\mathbf{x}_1) = \zeta_I(\theta)$$

□

Bayes Theorem vs. Proposition 2 (Factorization):

$$f_I(\theta|\mathbf{x}) = \frac{f_I(\theta)\, f_I(\mathbf{x}|\theta)}{f_I(\mathbf{x})} \leftrightarrow f_I(\theta|\mathbf{x}) = \frac{\zeta_I(\theta)\, f_I(\mathbf{x}|\theta)}{\eta_I(\mathbf{x})};$$

$$f_I(\theta) \leftrightarrow \zeta_I(\theta);$$

Difference: consistency factor $\zeta_I(\theta)$ not a pdf (needs not be integrable).

$\zeta_I(\theta)$ unique only up to a multiplier $\chi(\mathbf{x})$:

$$\frac{\chi(\mathbf{x})\zeta_I(\theta)\, f_I(\mathbf{x}|\theta)}{\int_{\mathbb{R}^m} \chi(\mathbf{x})\zeta_I(\theta)\, f_I(\mathbf{x}|\theta)\, d^m\theta} = \frac{\zeta_I(\theta)\, f_I(\mathbf{x}|\theta)}{\int_{\mathbb{R}^m} \zeta_I(\theta)\, f_I(\mathbf{x}|\theta)\, d^m\theta}.$$

Transformation of $\zeta_I(\theta)$ under reparametrization:

$$\zeta_{I'}(\lambda) = \zeta_I[\underline{\mathbf{s}}^{-1}(\lambda)]\left|\partial_\lambda \underline{\mathbf{s}}^{-1}(\lambda)\right|, \lambda \equiv \underline{\mathbf{s}}(\theta), \left|\partial_\lambda \underline{\mathbf{s}}^{-1}(\lambda)\right| \neq 0. \quad (6)$$

Example 1(Location - scale family).

$$\left.\begin{aligned} F_I(x\,|\,\mu,\sigma) &= \Phi\left(\frac{x-\mu}{\sigma}\right) \\ f_I(x\,|\,\mu,\sigma) &= \partial_x F_I(x\,|\,\mu,\sigma) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \end{aligned}\right\} x,\mu \in \mathbb{R},\ \ \sigma \in \mathbb{R}^+$$

$$f_I(\mu,\sigma\,|\,x_1,x_2) = \frac{\zeta_I(\mu,\sigma)}{\eta_I(x_1,x_2)}\, f_I(x_1,x_2\,|\,\mu,\sigma)\,;\ X_1, X_2 \text{ i.i.d.},$$

$$f_I(\mu\,|\,\sigma,x_1) = \frac{\zeta_{I,\sigma}(\mu)}{\eta_{I,\sigma}(x_1)}\, f_I(x_1\,|\,\mu,\sigma)\,,$$

$$f_I(\sigma\,|\,\mu,x_1) = \frac{\zeta_{I,\mu}(\sigma)}{\eta_{I,\mu}(x_1)}\, f_I(x_1\,|\,\mu,\sigma)\,.$$

2. Objectivity:

Definition (Objectivity). A probabilistic parametric inference is called objective, or internally consistent (McCullagh, 1992), if a particular likelihood function always leads to the same posterior density function.

Motivation: at the beginning of the inference only the parametric family is known, and inferences based on identical information should be the same.

Invariance: $\mathbf{y} \equiv \boldsymbol{l}(a, \mathbf{x}),\ \lambda \equiv \bar{\boldsymbol{l}}(a, \theta),\ a \in G;$

$$F_{I'}(\mathbf{y} \mid \lambda) = F_I(\mathbf{y} \mid \lambda);$$

$$f_I(\mathbf{y} \mid \lambda) = f_I(\boldsymbol{l}^{-1}[a, \mathbf{y}] \mid \bar{\boldsymbol{l}}^{-1}[a, \lambda]) \left| \partial_{\mathbf{y}} \boldsymbol{l}^{-1}(a, \mathbf{y}) \right|.$$

$\Longrightarrow$ Relative invariance of $\zeta_I(\theta)$:

$$\zeta_I(\theta) = \chi(a)\, \zeta_I[\bar{\boldsymbol{l}}^{-1}(a, \theta)] \left| \partial_{\boldsymbol{\theta}} \bar{\boldsymbol{l}}^{-1}(a, \theta) \right|. \qquad (7)$$

Example 2 (Invariance of a location-scale family).

$$\left.\begin{aligned} l[(a,b),x] &\equiv ax+b \\ \bar{l}[(a,b),(\mu,\sigma)] &\equiv (a\mu+b, a\sigma) \end{aligned}\right\} \quad (a,b)\in \mathbb{R}^{+}\times\mathbb{R} = G$$

$$\Rightarrow \zeta_I(\mu,\sigma) = \chi(a,b)\,\zeta_I\left(\frac{\mu-b}{a},\frac{\sigma}{a}\right)\frac{1}{a^2}$$

Proposition 3.

$I = \{ f_I(x\,|\,\theta) : x\in\mathbb{R} \wedge \theta \in V_\Theta \subseteq \mathbb{R}\}$, invariant under a scalar Lie group G

$\Rightarrow \exists\, s, \underline{s} :\ f_{I'}(y\,|\,\mu) = \phi(y-\mu)$, $y \equiv s(x)$ and $\mu \equiv \underline{s}(\theta)$.

Example 3. Every scale parameter σ of a location - scale family I is reducible to a location parameter ν of a location - scale family $\tilde{I}^{\pm}$.

$$I = \left\{ f_I(x \mid \mu, \sigma) = \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^{+} \right\}$$

$$\tilde{I}^{\pm} = \left\{ f_{\tilde{I}^{\pm}}(z \mid \nu, \tau) = \frac{1}{\tau}\tilde{\phi}\left(\frac{z-\nu}{\tau}\right) : (\nu, \tau) \in \mathbb{R} \times \mathbb{R}^{+} \right\}$$

$$\tilde{\phi}(u) \propto \exp(u)\phi[\pm\exp(u)],\ \nu \equiv \ln\sigma,\ \text{and}\ z \equiv \ln[\pm(x-\mu)];\ x-\mu \gtrless 0$$

Proposition 4.

a) $\exists$ objective $f_I(\mu,\sigma \mid x_1, x_2)$, $f_I(\mu \mid \sigma, x_1) \Rightarrow \boxed{\zeta_{I,\sigma}(\mu) = 1}$;

b) $\exists$ objective $f_I(\sigma \mid \mu, x_1)$, $f_{\tilde{I}^\pm}(\nu,\tau \mid z_1, z_2)$, $f_{\tilde{I}^\pm}(\nu \mid \tau, z_1) \Rightarrow \boxed{\zeta_{I,\mu}(\sigma) = \frac{1}{\sigma}}$;

c) $\exists$ objective $f_I(\mu,\sigma \mid x_1, x_2)$, $f_I(\sigma \mid \mu, x_1)$, $f_{\tilde{I}^\pm}(\nu,\tau \mid z_1, z_2)$, $f_{\tilde{I}^\pm}(\nu \mid \tau, z_1) \Rightarrow \boxed{\zeta_I(\mu,\sigma) = \frac{1}{\sigma}}$.

Remark 4. $\zeta_{I,\sigma}(\mu)$, $\zeta_{I,\mu}(\sigma)$, and $\zeta_I(\mu,\sigma)$ **not** integrable.

Different axiomatizations:

- Qualitative probabilities (Villegas, 1967, 1977, 1981)
- Non-unitary probabilities (Hartigan, 1983, pp. 14 and 18)

} $P(\Omega_x) \neq 1$

- Finite additivity (de Finetti, 1974, pp. 119 and 228-252)

"Each of them opens its own particular can of worms" (Schervish, 1996, p. 20), e.g., non-conglomerability, $f_I(x) \neq \int f_I(\theta)\, f_I(x \mid \theta)\, d\theta$, … (O'Hagan, 1994, pp. 100-102)

The special axiomatizations are unnecessary for the consistency factors.

Non-integrable consistency factors (priors) supposedly lead to inconsistencies, such as, for example, the marginalization paradox (Stone and Dawid, 1972) and the strong inconsistency (Stone, 1976).

The inconsistencies stem from applying to the rules outside PPI, and from wrong conclusions.

Example 4 (Strong inconsistency).

$$I = \{ f_I(x\,|\,\mu) \sim N(\mu,\sigma) : \mu \in \mathbb{R}\},\ f_I(\mu) = \exp\{4\mu\},\ B \equiv \{(\mu,x) \in \mathbb{R}\times\mathbb{R} : \mu - x > 2\};$$

$$\Rightarrow \begin{Bmatrix} \Pr_I(B\,|\,\mu) = [1-\mathrm{erf}\{\sqrt{2}\}]/2\,;\ \forall\mu \overset{?}{\Rightarrow} \Pr_I(B\,|\,\mu) = \Pr_I(B) \\ \Pr_I(B\,|\,x) = [1+\mathrm{erf}\{\sqrt{2}\}]/2\,;\ \forall x \overset{?}{\Rightarrow} \Pr_I(B\,|\,x) = \Pr_I(B) \end{Bmatrix} \Rightarrow \underbrace{\Pr_I(B) < \Pr_I(B)}_{\rightarrow\mid\leftarrow}.$$

But… $\int_B f_I(\mu,x)\,d\mu\,dx = \int_{-\infty}^{\infty} d\mu\ f_I(\mu) \int_{-\infty}^{\mu-2} dx\ f_I(x\,|\,\mu) = \infty \Rightarrow \nexists\ \Pr_I(B)$

$$\Rightarrow \Pr_I(B\,|\,\mu), \Pr_I(B\,|\,x) \neq \Pr_I(B).$$

3. Calibration:

Definition (Calibration). $f_I(\theta \mid \mathbf{x}), \theta \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$, is calibrated iff

a) for every $\gamma \in [0,1]$ there is at least one algorithm to construct a credible region $D(\mathbf{x}) \in \mathfrak{B}^m$ whose inverse probability distribution $\Pr_I(D \mid \mathbf{x}) = \gamma$ is equal to the direct probability distribution (unconditional or conditional) of a set $B(\theta) \in \mathfrak{B}^n$,

b) $\theta \in D \Leftrightarrow \mathbf{x} \in B$ holds.

Similarly, we define calibrated $f_I(\theta_1 \mid \theta_2, \mathbf{x}), \theta_1 \in \mathbb{R}^{m_1}, \theta_2 \in \mathbb{R}^{m_2}, \mathbf{x} \in \mathbb{R}^n$,

calibrated $f_I(\theta_1 \mid \mathbf{x}) = \int_{\mathbb{R}^{m_2}} f_I(\theta_1, \theta_2 \mid \mathbf{x})\, d^{m_2}\theta_2$, and calibrated

$f_I(\mathbf{x}_{n+1} \mid \mathbf{x}_1, \dots, \mathbf{x}_n) = \int_{\mathbb{R}^m} f_I(\theta \mid \mathbf{x}_1, \dots, \mathbf{x}_n)\, f_I(\mathbf{x}_{n+1} \mid \theta)\, d^m\theta, \mathbf{x}_i \in \mathbb{R}^n$.

$\zeta_{I,\sigma}(\mu), \zeta_{I,\mu}(\sigma)$, and $\zeta_I(\mu,\sigma)$ coincide with the elements of the right Haar measures for the translation group $\mathbb{R}$, for the scale group $\mathbb{R}^+$, and for the group $\mathbb{R}^+ \times \mathbb{R}$ whose group operation is $(a_1,b_1)\circ(a_2,b_2)=(a_1a_2, a_1b_2+b_1) \Rightarrow$ calibrated $f_I(\mu\,|\,\sigma, x_1,\ldots,x_k)$, $f_I(\sigma\,|\,\mu, x_1,\ldots,x_k)$, and $f_I(\mu,\sigma\,|\,x_1,\ldots,x_j)$; $k \geq 1$, $j \geq 2$ (Chang and Villegas, 1986; $D[\boldsymbol{I}(a,\mathbf{x})] = \bar{\boldsymbol{I}}[a, D(\mathbf{x})]$).

The elements $\zeta_{I,\sigma}(\mu), \zeta_{I,\mu}(\sigma)$, and $\zeta_I(\mu,\sigma)$ of the right Haar measures also ensure calibrated predictive distributions (on HPD credible sets)

$$f_I(x_{k+1}\,|\,\sigma, x_1,\ldots,x_k) = \int_{\mathbb{R}} f_I(\mu\,|\,\sigma, x_1,\ldots,x_k)\, f_I(x_{k+1}\,|\,\mu,\sigma)\, d\mu,$$

$$f_I(x_{k+1}\,|\,\mu, x_1,\ldots,x_k) = \int_{\mathbb{R}^+} f_I(\sigma\,|\,\mu, x_1,\ldots,x_k)\, f_I(x_{k+1}\,|\,\mu,\sigma)\, d\sigma,$$ and

$$f_I(x_{j+1}\,|\,x_1,\ldots,x_j) = \int_{\mathbb{R}\times\mathbb{R}^+} f_I(\mu,\sigma\,|\,x_1,\ldots,x_j)\, f_I(x_{j+1}\,|\,\mu,\sigma)\, d\mu\, d\sigma$$

(Eaton and Sudderth, 2004).

Example 5. Calibration of $f_I(\mu,\sigma\,|\,\mathbf{x}), \mathbf{x}=(x_1,\dots,x_n), n\ge 2.$

$$f_I(\mathbf{x}\,|\,\mu,\sigma)=\frac{1}{\sigma^n}\prod_{i=1}^{n}\phi\left(\frac{x_i-\mu}{\sigma}\right);$$

$D_1=(-\infty,\mu_1)\times\mathbb{R}^+,\ D_2=(\mu_1,\mu_2)\times\mathbb{R}^+,\ D_3=(\mu_1,\mu_2)\times(0,\sigma_1)\,,\ D=(\mu_1,\mu_2)\times(\sigma_1,\sigma_2),;$

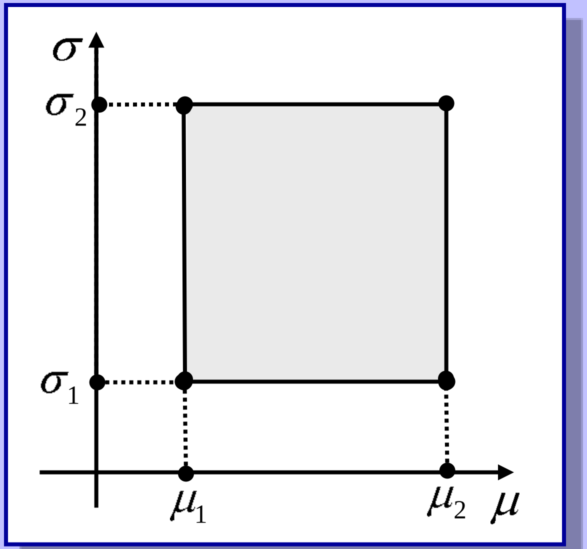
(1) $\Pr_I(D_1\,|\,\mathbf{x})=\alpha$, **(2)** $\Pr_I(D_2\,|\,\mathbf{x})=\beta$, **(3)** $\Pr_I(D_3\,|\,\mathbf{x})=\varepsilon$, **(4)** $\Pr_I(D\,|\,\mathbf{x})=\gamma$;

$0\le\alpha,\beta,\varepsilon,\gamma\le 1\,,\,1-\alpha\ge\beta\ge 1-\varepsilon\ge\gamma$.

- Rectangles D with sides parallel to the axes μ and σ.
- The rectangles include bands $\mathbb{R}\times(\sigma_1,\sigma_2)$ and $(\mu_1,\mu_2)\times\mathbb{R}^+$ (set $\beta=\gamma$ and $\beta=1$, respectively).

  $\Rightarrow\ f_I(\mu\,|\,\mathbf{x})=\int_{\mathbb{R}^+} f_I(\mu,\sigma\,|\,\mathbf{x})\,d\sigma$ and

  $f_I(\sigma\,|\,\mathbf{x})=\int_{\mathbb{R}} f_I(\mu,\sigma\,|\,\mathbf{x})\,d\mu$ also calibrated.
- This property is not generally conserved under (1-1) reparametrizations $(\mu,\sigma)\to(\nu,\tau)$.



Example 6. Behrens - Fisher problem (Behrens, 1929; Fisher, 1935).

$$f_I(\mathbf{x},\mathbf{y}\mid\mu_x,\sigma_x,\mu_y,\sigma_y)=\prod_{i=1}^{n} f_{I_x}(x_i\mid\mu_x,\sigma_x)\prod_{j=1}^{m} f_{I_y}(y_j\mid\mu_y,\sigma_y)\,,$$

$\mathbf{x}=(x_1,\ldots,x_n)$, $\mathbf{y}=(y_1,\ldots,y_m)$; $n,m\geq 2$,

$$I_x=\{\,f_I(x\mid\mu_x,\sigma_x)\sim N(\mu_x,\sigma_x)\;:(\mu_x,\sigma_x)\in\mathbb{R}\times\mathbb{R}^+\}\,,$$

$$I_y=\{\,f_I(y\mid\mu_y,\sigma_y)\sim N(\mu_y,\sigma_y):(\mu_y,\sigma_y)\in\mathbb{R}\times\mathbb{R}^+\}\,.$$

$$\zeta_I(\mu_x,\sigma_x,\mu_y,\sigma_y)=(\sigma_x\,\sigma_y)^{-1}$$

$\Rightarrow\ f_I(\mu_x,\sigma_x,\mu_y,\sigma_y\mid\mathbf{x},\mathbf{y})$ calibrated on 4 - dim. rectangles

$(\mu_{x,1},\mu_{x,2})\times(\sigma_{x,1},\sigma_{x,2})\times(\mu_{y,1},\mu_{y,2})\times(\sigma_{y,1},\sigma_{y,2})$

$\Rightarrow\ f_{I'}(\delta,\sigma_x,\mu_y,\sigma_y\mid\mathbf{x},\mathbf{y})=f_I(\delta+\mu_y,\sigma_x,\mu_y,\sigma_y\mid\mathbf{x},\mathbf{y})$, $\boxed{\delta\equiv\mu_x-\mu_y}$ calibrated, but **not** on

$(\delta_1,\delta_2)\times(\sigma_{x,1},\sigma_{x,2})\times(\mu_{y,1},\mu_{y,2})\times(\sigma_{y,1},\sigma_{y,2})$

$\Rightarrow\ f_{I'}(\delta\mid\mathbf{x},\mathbf{y})=\int_{\mathbb{R}^+}d\sigma_x\int_{\mathbb{R}^+}d\sigma_y\int_{\mathbb{R}}d\mu_y\ f_{I'}(\delta,\sigma_x,\mu_y,\sigma_y\mid\mathbf{x},\mathbf{y})$ **not** calibrated.

4. Interpretation:

"Jede axiomatische (abstrakte) Theorie läßt bekanntlich unbegrentzt viele konkrete Interpretationen zu." (Kolmogorov, 1933, p. 1.)

Direct probability distribution. A necessary condition for a sequence $\{\mathbf{x}_i\}_{i=1}^n, \mathbf{x}_i \in \mathbb{R}^n$, to be $\Pr_I(\bullet \mid \theta)$ - random is

$$\lim_{n\to\infty}\left|\frac{n(\mathbf{x}_i \in B)}{n} - \Pr_I(B \mid \theta)\right| < \varepsilon \text{ for every } \varepsilon > 0 \text{ and } B \in \mathfrak{B}^n.$$

$\Rightarrow$ Probability distribution $=$ (frequency) distribution of $\{\mathbf{x}_i\}_{i=1}^\infty$.

$$\{\mathbf{x}_i\}_{i=1}^\infty, \mathbf{x}_i \in \mathbb{R}^n, B \in \mathfrak{B}^n, \Pr_I(B \mid \theta) = \gamma, \{b_i\}_{i=1}^\infty, b_i = \begin{cases} 1\,; \mathbf{x}_i \in B \\ 0\ \mathbf{x}_i \notin B \end{cases};$$

$\Rightarrow \{\mathbf{x}_i\}_{i=1}^\infty$ is $\Pr_I(\bullet \mid \theta)$ - random iff $\{b_i\}_{i=1}^\infty$ is Martin-Löf random (or typical) with respect to the Bernoulli measure Ber_γ for every $B \in \mathfrak{B}^n$ (Martin-Löf, 1966).

Inverse probability distribution. The inferred (true) value θ of the parameter is unknown, but (often) fixed. What is distributed is the rational degree of belief that, given a realization $\mathbf{x}$ of $\mathbf{X}$, the true value of θ is in $D(\mathbf{x}_i)$.

Still, calibrated inverse distributions are operational because they provide verifiable predictions:

$\{\mathbf{x}_i\}_{i=1}^{\infty}$ sampled from $\Pr_I(\bullet \mid \theta_i)$ - random sequences,

$f_I(\theta \mid \mathbf{x})$ calibrated,

$D(\mathbf{x}_i)$ calibrated credible regions with $\Pr_I(D \mid \mathbf{x}_i) = \gamma$

$$\Rightarrow \lim_{n\to\infty} \left| \frac{n[\theta_i \in D(\mathbf{x}_i)]}{n} - \gamma \right| < \varepsilon \,.$$

“A typical setting occurs when estimating physical quantities like the speed of light, c. An answer in this particular setting is that the limited accuracy of the measurement instruments implies that the true value of c will never be known, and thus it is justified to consider c as being uniformly distributed on $[c_0-\varepsilon, c_0+\varepsilon]$, if ε is the maximal precision of the measuring instruments and c_0 the obtained value.” (Robert, 2001, p. 10)

“Whether or not the difference (between the two interpretations of probability) is important philosophically, we do not feel it should make any difference operationally.” (Berger, 1985, pp. 76-77)

The difference between the two interpretations **does** make the difference also operationally.

For example, the inverse and the predictive distributions must be calibrated in order to be operational.

Caveat against the unreserved use of marginalization; important because the marginalization is the principal practical advantage of the Bayesian approach over other methods of inference (Gelman et al., 2004, pp. 73-74 ; Robert, 2001, p. 168 ; Jaynes, 2003, p. xxviii ; Cox and Hinkley, 2000, pp. 53 and 364-365).

5. Conclusion and discussion:

- PPI without invoking prior probability distributions.
- The theory is operational, i.e., it provides for verifiable predictions.
- Not assumed that the solution (the objective inverse distribution) always exists.
- The solutions found for location-scale families (symmetry under Lie groups,…).
- Other approaches:

· **The Laplace Principle of insufficient reason** (Laplace, 1886, p. XVII), **The Principle of maximum entropy** (Catlin, 1989, p. 82) : contradict Eq. (6).

· **Jeffreys' priors** (Jeffreys, 1946), **form invariant priors** ($\chi(a) = 1$ in Eq. (7); Harney, 2003) : violate the product rule

$$f_I(\mu, \sigma \mid x_1, \ldots, x_n) = f_I(\sigma \mid \mu, x_1, \ldots, x_n)\, f_I(\mu \mid x_1, \ldots, x_n), n \geq 2 .$$

· **Fiducial inference** (Fisher, 1930), **empirical Bayes** (Robbins, 1964) : not generally compatible with Proposition 2 (Factorization).

· **Reference priors** (Bernardo, 1979) : inconsistent (multiple solutions) .

- **Subjective Bayes** (deFinetti, 1974; O'Hagan, 1994) :

"For the subjectivist, $f(\theta)$, $f(x\,|\,\theta)$, and $f(\theta\,|\,x)$ are all subjective probability distributions." (O'Hagan, p. 13)

"In contrast (to classical, or frequentist, statistician, there is a unique Bayesian solution to any problem." (O'Hagan, p. 20)

"The Bayesian method requires the posterior distribution to be derived via Bayes' theorem." (O'Hagan, pp. 13 and 20)

The (subjective) posteriror distributions **lack operationality** (unless the prior distributions coincide with the (true frequency) distributions of the inferred parameters).

Probabilistic inference is an ideal (Paris, 1994, p. 33), but ideals cannot always be reached.

In order to set a theory with general applicability, we would then have to seek beyond the probabilistic framework (interval probabilities, belief functions, truth-functions,…).

References:

Behrens, W. V. (1929). Ein Beitrag zur Fehlerberechnung bei Wenige Beobachtungen, Landwirtch. Jb. **68**, pp. 807-937.

Berger, J. O. (1980). *Statistical Decision Theory and Bayesian Analysis,* Springer Series in Statistics. Springer-Verlag.

Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian inference. J. Roy. Statist. Soc. Ser. B **41**, pp. 113-147.

Bernardo, J. M., Smith, A. F. M. (1994). *Bayesian Theory.* John Wiley & Sons.

Catlin, D. E. (1989). *Estimation, Control, and the Discrete Kalman Filter,* Applied mathematical Sciences. Springer-Verlag.

Chang, T., Villegas, C. (1986). On a Theorem of Stein Relating Bayesian and Classical Inferences in Group Models. Canad. J. Statist. **14**, pp. 289-296.

Cox, D. R., Hinkley, D. V. (2000). *Theoretical Statistics.* Chapman & Hall/CRC.

de Finetti, B. (1974). *Theory of Probability – Vol. 1.* John Wiley & Sons.

Eaton, M. L., Sudderth, W. D. (2004). Properties of Right Haar Predictive Inference. Sankhyã **66**, pp. 487-512.

Fienberg, S. E. (2006). Does is Make Sense to be an “Objective Bayesian”? Bayes. Anal. **1**, pp. 429-432.

Fisher, R. A. (1922). On the Mathematical Foundations of Theoretical Statistics. Philos. Trans. R. Soc. Lond. Ser. A, **222**, pp. 309-368.

Fisher, R. A. (1935). The Fiducial Argument in Statistical Inference. Ann. Eugen. **6**, pp. 391-398.

Fisher, R. A. (1930). Inverse Probability. Proc. Cambridge Philos. Society 26, pp. 528-535.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B. (2004). *Bayesian Data Analysis,* 2nd Edition. Chapman & Hall/CRC.

Harney, H. L. (2003). *Bayesian Inference – Parameter Estimation and Decisions,* Advanced Texts in Physics. Springer.

Hartigan, J. A. (1983). *Bayes Theory,* Springer Series in Statistics. Springer-Verlag.

Jaynes, E. T. (2003). *Probability Theory - The Logic of Science.* Cambridge University Press.

Jeffreys, H. (1946) An Invariant Form for the Prior Probability Estimation Problems. Proc. R. Soc. Lond. Ser. A **186**, pp. 453-461.

Jeffreys, H. (1961). *Theory of Probability,* 3rd Edition. Oxford: Clarendon Press.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrehnung,* Vol. 2 of Ergebnisse der Matematik und ihrer Granzgebeite. Berlin: Springer.

Laplace, P. S. (1886). *Théorie Analitique des Probabilités,* Vol. 8. of Œuvres Complètes. Paris: Guthier-Villars.

Luo, X., Yhang, C. (1997). An Axiomatic Framework for Interval Probability. In Santos, E., editor, Proceedings of the Eighth Midwest Artificial Intelligence and Cognitive Science Society, pp. 47-52. AAAI Press, CON-97-01.

Martin-Löf, P. (1966). The Definition of Random Sequences. Inform. And Control **9**, pp. 602-619.

McCullagh, P. (1992). Conditional Inference and Cauchy Models. Biometrika **79**, pp. 247-259.

Neyman, J. (1937). Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. Philos. Trans. R. Soc. Lond. Ser. A, **236**, pp. 333-380.

O'Hagan, A. (1994). *Bayesian Inference,* Vol. 2B of Kendall's Advanced Theory of Statistcs. London: Arnold.

Paris, J. B. (1994). *The Uncertain Reasoner's Companion – A Mathematical Perspective,* Vol. 39 of Cambridge Tracts in Theoretical Computer Science. Cambridge: Cambridge University Press.

Pólya, G. (1954). *Patterns of Plausible Inference,* Vol. 2 of Mathematics and Plausible reasoning. Princeton: Princeton University Press.

Rao, M. M. (1993). *Conditional Measures and Applications.* Marcel Dekker, Inc..

Robbins, H. (1964). The Empirical Bayes Approach to Statistical Decision Problems. Ann. Math. Statist. **35**, pp. 1-20.

Robert, C. P. (2001). *The Bayesian Choice,* Springer Texts in Statistics. Springer.

Schervish, M. J. (1995). *Theory of Statistics,* Springer Series in Statistics. Springer-Verlag.

Stone, M., Dawid., P. (1972). Un-Bayesian Implications of Improper Bayes Inference in Routine Statistical Problems. Biometrika **59**, pp. 369-357.

Stone, M. (1976). Strong Inconsistency from Uniform Priors. J. Amer. Statist. Assoc. **71**, pp. 114-116.

Stuart, A., Ord, J. K. (2000). *Distribution Theory,* Vol. 1. of Kendall's Advanced Theory of Statistics. London: Arnold.

Villegas, C. (1967). On *Qualitative* Probability. Amer. Math. Monthly **74**, pp. 661-669.

Villegas, C. (1972). Inner Statistical Inference. J. Amer. Statist. Assoc. **72**, pp. 453-458.

Villegas, C. (1981). Inner Statistical Inference II. Ann. Statist. **9**, pp. 768-776.