Statistical methods

Tomaž Podobnik

Oddelek za fiziko, FMF, UNI-LJ

Odsek za eksperimentalno fiziko delcev, IJS

Contents:

VI. Inverse Probability Distributions:

1. Direct probability distributions (a refresher),
2. Inverse probability distributions,
3. (Non-informative) Prior probability distributions,
4. Factorization Theorem and consistency factors,
5. Objectivity,
6. Consistency factors for location-scale families,
7. Consistency factors in the absence of symmetry.

1. Direct probability distributions:

Definition 21 (Parametric family). *The term* parametric family *stands for a collection* $I = \{\Pr_{I,\theta} : \theta \in V_\Theta\}$ *of probability distributions that differ only in the value* θ *of a parameter* Θ.

$$F_X(x) = F_{I,\theta}(x),\ f_X(x) = f_{I,\theta}(x),\ p_X(x) = p_{I,\theta}(x);$$

$$F_{I,\theta}(x) = F_I(x \mid \theta),\ f_{I,\theta}(x) = f_I(x \mid \theta),\ p_{I,\theta}(x) = p_I(x \mid \theta).$$

Example 1. $\Pr_X(B) = \Pr_{I,\theta}(B) = \Pr_I(B \mid \theta) = \int_B f_I(x \mid \theta)\,dx,\ B \in \mathfrak{B}.$

Definitions extend without change to random vectors:

$$F_{\mathbf{X}}(\mathbf{x}) = F_I(\mathbf{x} \mid \boldsymbol{\theta}),\ f_{\mathbf{X}}(\mathbf{x}) = f_I(\mathbf{x} \mid \boldsymbol{\theta}),\ p_{\mathbf{X}}(\mathbf{x}) = p_I(\mathbf{x} \mid \boldsymbol{\theta});$$

$$\Pr_{\mathbf{X}}(B) = \Pr_{I,\boldsymbol{\theta}}(B) = \Pr_I(B \mid \boldsymbol{\theta}) = \int_B f_I(\mathbf{x} \mid \boldsymbol{\theta})\, d^n\mathbf{x};\ B \in \mathfrak{B}^n .$$

Conditional probability distributions:

$$\mathbf{X} = (Y, Z),\ f_{\mathbf{X}}(y,z) = f_I(y,z \mid \boldsymbol{\theta}),\ f_I(z \mid \boldsymbol{\theta}) \equiv \int_{\mathbb{R}} f_I(y,z \mid \boldsymbol{\theta})\, dy > 0$$

$$\Rightarrow f_I(y \mid z, \boldsymbol{\theta}) \equiv \frac{f_I(y,z \mid \boldsymbol{\theta})}{f_I(z \mid \boldsymbol{\theta})} .$$

Example 2 (Reparametrization). Let a probability distribution for a continuous random vector $\mathbf{X}$ belong to a family *I*, $f_{\mathbf{X}}(\mathbf{x}) = f_I(\mathbf{x} \mid \boldsymbol{\theta})$, and let $\mathbf{y} \equiv \mathbf{s}(\mathbf{x})$ and $\boldsymbol{\lambda} \equiv \bar{\mathbf{s}}(\boldsymbol{\theta})$ with $\partial_{\mathbf{x}}\mathbf{s}(\mathbf{x}), \partial_{\boldsymbol{\theta}}\bar{\mathbf{s}}(\boldsymbol{\theta}) \neq 0$. *Then,*

$$f_{I'}(\mathbf{y} \mid \boldsymbol{\lambda}) = f_I\left(\mathbf{s}^{-1}(\mathbf{y}) \mid \bar{\mathbf{s}}^{-1}(\boldsymbol{\lambda})\right)\left|\partial_{\mathbf{y}}\mathbf{s}^{-1}(\mathbf{y})\right| .$$

2. Inverse probability distributions:

Definition 1 (Inverse probability distribution). *Given a probability space* (Ω,Σ,P) *and a function* $(\boldsymbol{\Theta},\mathbf{X}):\Omega\to(\mathbb{R}^m,\mathbf{x}),\ \mathbf{x}\in\mathbb{R}^n,\ \boldsymbol{\Theta}$ is $\Sigma-$measurable, and $\mathbf{X}\sim f_I(\mathbf{x}\,|\,\boldsymbol{\theta})$, *the* inverse probability distribution *is defined as the image measure of* P *by the random variable* $\boldsymbol{\Theta}$, $\Pr_I(\bullet\,|\,\mathbf{x})\equiv P\circ\boldsymbol{\Theta}^{-1}$, *such that*

$$\Pr_I(B\,|\,\mathbf{x})=P\left[\boldsymbol{\Theta}^{-1}(B)\right],\ B\in\mathfrak{B}^m.$$

Distribution and density functions:

$$F_I(\mathbf{x}\,|\,\boldsymbol{\theta})\leftrightarrow F_I(\boldsymbol{\theta}\,|\,\mathbf{x}),\ f_I(\mathbf{x}\,|\,\boldsymbol{\theta})\leftrightarrow f_I(\boldsymbol{\theta}\,|\,\mathbf{x}).$$

Also: probability distributions that are neither purely direct nor purely inverse. Distribution and density functions of such a distribution:

$$F_I(\boldsymbol{\theta},\mathbf{x}_1\,|\,\mathbf{x}_2)\text{ and }f_I(\boldsymbol{\theta},\mathbf{x}_1\,|\,\mathbf{x}_2).$$

From a mathematical perspective, all probability distributions, be they purely direct, purely inverse, or mixtures of the two, are equivalent.

Conditional inverse probability distributions:

$$\boldsymbol{\Theta} = (\Theta_1, \Theta_2), \exists\, f_I(\theta_1, \theta_2 \mid \mathbf{x}),\ f_I(\theta_2 \mid \mathbf{x}) \equiv \int_{\mathbb{R}} f_I(\theta_1, \theta_2 \mid \mathbf{x})\, d\theta_1 > 0$$

$$\Rightarrow f_I(\theta_1 \mid \theta_2, \mathbf{x}) \equiv \frac{f_I(\theta_1, \theta_2 \mid \mathbf{x})}{f_I(\theta_2 \mid \mathbf{x})}.$$

Example 2 (Reparametrization). Suppose there is $f_I(\boldsymbol{\theta} \mid \mathbf{x})$, and let $\mathbf{y} \equiv \mathbf{s}(\mathbf{x})$ and $\boldsymbol{\lambda} \equiv \bar{\mathbf{s}}(\boldsymbol{\theta})$ with $\partial_{\mathbf{x}} \mathbf{s}(\mathbf{x}), \partial_{\boldsymbol{\theta}} \bar{\mathbf{s}}(\boldsymbol{\theta}) \neq 0$. *Then,*

$$f_{I'}(\boldsymbol{\lambda} \mid \mathbf{y}) = f_I\left(\bar{\mathbf{s}}^{-1}(\boldsymbol{\lambda}) \mid \mathbf{s}^{-1}(\mathbf{y})\right) \left|\partial_{\boldsymbol{\lambda}} \bar{\mathbf{s}}^{-1}(\boldsymbol{\lambda})\right|.$$

3. (Non-informative) Prior probability distributions:

Theorem 1 (Bayes). $f_I(\boldsymbol{\theta}, \mathbf{x}) = f_I(\boldsymbol{\theta} \mid \mathbf{x})\, f_I(\mathbf{x}) = f_I(\mathbf{x} \mid \boldsymbol{\theta})\, f_I(\boldsymbol{\theta})$

$$\Rightarrow f_I(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{f_I(\boldsymbol{\theta})\, f_I(\mathbf{x}|\boldsymbol{\theta})}{f_I(\mathbf{x})}, \quad f_I(\mathbf{x}) = \int_{\mathbb{R}^m} f_I(\boldsymbol{\theta})\, f_I(\mathbf{x}|\boldsymbol{\theta})\, d^m\boldsymbol{\theta}.$$

T. Bayes (1763). Phil. Trans. R. Soc., **53**, 370-418.
P. S. Laplace (1774). Mém. Acad. R. Sci., **6**, 621-656.

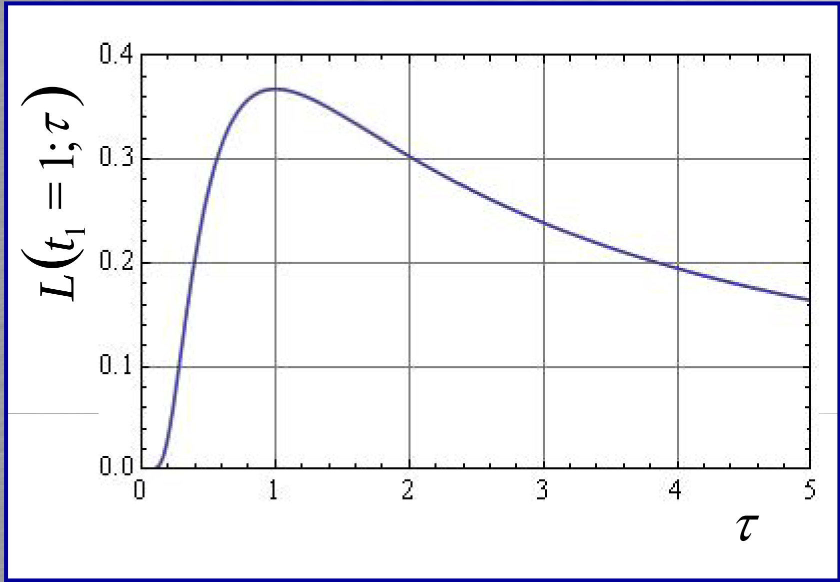
$f_I(\boldsymbol{\theta})$: (non-informative) prior probability distribution.

Uniform $f_I(\boldsymbol{\theta})$: (Bayes, Laplace).

Mathematical difficulties:

$$a)\ f_I(\tau)=\text{const.}\Rightarrow\int_0^\infty f_I(\tau)d\tau=\infty,\int_0^\infty f_I(\tau)f_I(t_1\,|\,\tau)d\tau=\infty,$$

$$b)\ \mu=\ln\tau\Rightarrow f_{I'_{\sigma=1}}(\mu)=e^{\mu}\neq\text{const.}$$



Conceptual (interpretational) difficulties:

a) We do not know anything about $\tau \leftrightarrow$ we know that τ is uniformly distributed,

b) τ is unknown but fixed $\leftrightarrow \tau$ is distributed.

$\Longrightarrow$ " ...inverse probability is a mistake (perhaps the only mistake to which the mathematical world has so deeply committed itself),..."
(R. A. Fisher (1922). Phil. Trans. R. Soc., **A 222**, 309-368)

4. Factorization Theorem and consistency factors:

Theorem 2 (Bayes). $\exists\, f_I(\boldsymbol{\theta}, \mathbf{x}_1 \mid \mathbf{x}_2),\ f_I(\boldsymbol{\theta}, \mathbf{x}_2 \mid \mathbf{x}_1);$

$$f_I(\boldsymbol{\theta} \mid \mathbf{x}_{1,2}) = \int_{\mathbb{R}^n} f_I(\boldsymbol{\theta}, \mathbf{x}_2 \mid \mathbf{x}_1)\, d^n\mathbf{x}_{2,1} > 0\,;$$

$$f_I(\mathbf{x}_{1,2} \mid \mathbf{x}_{2,1}) = \int_{\mathbb{R}^m} f_I(\boldsymbol{\theta}, \mathbf{x}_2 \mid \mathbf{x}_1)\, d^m\boldsymbol{\theta} > 0\,;$$

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

$$\Rightarrow f_I(\boldsymbol{\theta} \mid \mathbf{x}_1, \mathbf{x}_2) = \frac{f_I(\boldsymbol{\theta} \mid \mathbf{x}_1)\, f_I(\mathbf{x}_2 \mid \boldsymbol{\theta})}{f_I(\mathbf{x}_2 \mid \mathbf{x}_1)} = \frac{f_I(\boldsymbol{\theta} \mid \mathbf{x}_2)\, f_I(\mathbf{x}_1 \mid \boldsymbol{\theta})}{f_I(\mathbf{x}_1 \mid \mathbf{x}_2)}\,.$$

Under analogous conditions:

$$f_I(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{x}_1, \mathbf{x}_2) = \frac{f_I(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{x}_1)\, f_I(\mathbf{x}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{f_I(\mathbf{x}_2 \mid \boldsymbol{\theta}_2, \mathbf{x}_1)} = \frac{f_I(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2, \mathbf{x}_2)\, f_I(\mathbf{x}_1 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{f_I(\mathbf{x}_1 \mid \boldsymbol{\theta}_2, \mathbf{x}_2)}\,.$$

Proof. $f_I(\boldsymbol{\theta}, \mathbf{x}_{1,2} \mid \mathbf{x}_{2,1}) = f_I(\boldsymbol{\theta} \mid \mathbf{x}_{2,1}) f_I(\mathbf{x}_{1,2} \mid \boldsymbol{\theta}) = f_I(\mathbf{x}_{1,2} \mid \mathbf{x}_{2,1}) f_I(\boldsymbol{\theta} \mid \mathbf{x}_1, \mathbf{x}_2)\,.$ □

Theorem 3 (Factorization).

$$f_I(\boldsymbol{\theta}\,|\,\mathbf{x}_1,\mathbf{x}_2)=\frac{f_I(\boldsymbol{\theta}\,|\,\mathbf{x}_1)\,f_I(\mathbf{x}_2\,|\,\boldsymbol{\theta})}{f_I(\mathbf{x}_2\,|\,\mathbf{x}_1)}=\frac{f_I(\boldsymbol{\theta}\,|\,\mathbf{x}_2)\,f_I(\mathbf{x}_1\,|\,\boldsymbol{\theta})}{f_I(\mathbf{x}_1\,|\,\mathbf{x}_2)}$$

$$\Rightarrow\ f_I(\boldsymbol{\theta}\,|\,\mathbf{x}_{1,2})=\frac{\zeta_I(\boldsymbol{\theta})\,f_I(\mathbf{x}_{1,2}\,|\,\boldsymbol{\theta})}{\eta_I(\mathbf{x}_{1,2})};\ \eta_I(\mathbf{x}_{1,2})=\int\zeta_I(\boldsymbol{\theta})\,f_I(\mathbf{x}_{1,2}\,|\,\boldsymbol{\theta})\,d^m\boldsymbol{\theta}\,.$$

Similarly:

$$f_I(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2,\mathbf{x}_1,\mathbf{x}_2)=\frac{f_I(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2,\mathbf{x}_1)\,f_I(\mathbf{x}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{f_I(\mathbf{x}_2\,|\,\boldsymbol{\theta}_2,\mathbf{x}_1)}=\frac{f_I(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2,\mathbf{x}_2)\,f_I(\mathbf{x}_1\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{f_I(\mathbf{x}_1\,|\,\boldsymbol{\theta}_2,\mathbf{x}_2)}$$

$$\Rightarrow\ f_I(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2,\mathbf{x}_{1,2})=\frac{\zeta_{I,\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1)\,f_I(\mathbf{x}_{1,2}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{\eta_{I,\boldsymbol{\theta}_2}(\mathbf{x}_{1,2})};\ \eta_{I,\boldsymbol{\theta}_2}(\mathbf{x}_{1,2})=\int_{\mathbb{R}^{m_1}}\zeta_{I,\boldsymbol{\theta}_2}(\boldsymbol{\theta}_1)\,f_I(\mathbf{x}_{1,2}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\,d^{m_1}\boldsymbol{\theta}_1\,.$$

Proof. $\dfrac{\kappa(\boldsymbol{\theta},\mathbf{x}_1)}{\kappa(\boldsymbol{\theta},\mathbf{x}_2)}=h(\mathbf{x}_1,\mathbf{x}_2);\ \kappa(\boldsymbol{\theta},\mathbf{x}_{1,2})\equiv\dfrac{f_I(\boldsymbol{\theta}\,|\,\mathbf{x}_{1,2})}{f_I(\mathbf{x}_{1,2}\,|\,\boldsymbol{\theta})},\ h(\mathbf{x}_1,\mathbf{x}_2)\equiv\dfrac{f_I(\mathbf{x}_2\,|\,\mathbf{x}_1)}{f_I(\mathbf{x}_1\,|\,\mathbf{x}_2)}$

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

□

Theorem 1 (Bayes) vs. Theorem 3 (Factorization):

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

$$f_I(\boldsymbol{\theta}) \leftrightarrow \zeta_I(\boldsymbol{\theta});$$

Difference: $\zeta_I(\boldsymbol{\theta})$ not a pdf (need not be integrable).

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

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

Transformation of $\zeta_I(\boldsymbol{\theta})$ under reparametrization:

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

5. Objectivity:

Definition 2 (Objectivity). A probabilistic parametric inference is called objective 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: $\mathbf{y} \equiv \boldsymbol{l}(a,\mathbf{x}),\, \boldsymbol{\lambda} \equiv \bar{\boldsymbol{l}}(a,\boldsymbol{\theta}),\, a \in G;$

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

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

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

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

6. Consistency factors for location-scale families:

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

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

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

Invariance of a location-scale family:

$$F_I(x \mid \mu,\sigma) = \Phi\left(\frac{x-\mu}{\sigma}\right),$$

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

$$\zeta_I(\mu,\sigma) = \frac{\chi(a,b)}{a^2}\chi(a,b)\zeta_I\left(\frac{\mu-b}{a},\frac{\sigma}{a}\right); a\in\mathbb{R}^+, b\in\mathbb{R},$$

$$\zeta_{I,\sigma}(\mu) = \chi(b)\zeta_{I,\sigma}(\mu-b); b\in\mathbb{R},$$

$$\zeta_{I,\mu}(\sigma) = \frac{\chi(a)}{a}\zeta_{I,\mu}\left(\frac{\sigma}{a}\right); a\in\mathbb{R}^+.$$

Proposition 1. $f_I(\mu,\sigma\,|\,x_1,x_2),\ f_I(\mu\,|\,\sigma,x_1),\ f_I(\sigma\,|\,\mu,x_1)$ objective

$$\Rightarrow \zeta_I(\mu,\sigma) = \zeta_{I,\mu}(\sigma) = \sigma^{-1},\ \zeta_{I,\sigma}(\mu) = 1.$$

Example 3 (Normal family).

$$f_I(\mu \mid \sigma, x_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x_1-\mu)^2}{2\sigma^2}\right\},$$

$$f_I(\sigma \mid \mu, x_1) = \sqrt{\frac{2}{\pi}} \frac{|x_1-\mu|}{\sigma^2} \exp\left\{-\frac{(x_1-\mu)^2}{2\sigma^2}\right\},$$

$$f_I(\mu, \sigma \mid x_1, x_2) = \frac{|x_1-x_2|}{\pi\sigma^3} \exp\left\{-\frac{(x_1-\mu)^2+(x_2-\mu)^2}{2\sigma^2}\right\},$$

$$f_I(\sigma \mid x_1, x_2) = \int_{-\infty}^{\infty} f_I(\mu, \sigma \mid x_1, x_2)\, d\mu = \frac{|x_1-x_2|}{\pi\sigma^2} \exp\left\{-\frac{(x_1-x_2)^2}{4\sigma^2}\right\},$$

$$f_I(\mu \mid x_1, x_2) = \int_0^{\infty} f_I(\mu, \sigma \mid x_1, x_2)\, d\sigma = \frac{1}{\pi} \frac{|x_1-x_2|}{x_1^2+x_2^2+2\mu^2-2\mu(x_1+x_2)}.$$

Example 3 (cont'd).

$$f_I\left(\mu \mid x_1,\ldots,x_n\right)=\int_0^\infty f_I\left(\mu,\sigma \mid x_1,\ldots,x_n\right)d\sigma$$

$$=\sqrt{\frac{1}{\pi}}\frac{(\frac{n}{2}-1)!}{(\frac{n}{2}-\frac{3}{2})!}\frac{s_n'^{\,n-1}}{[s_n'^2+(\bar{x}_n-\mu)^2]^{\frac{n}{2}}},\; s_n'^2\equiv\frac{1}{n}\sum_{i=1}^n(x_i-\bar{x}_n)^2 .$$

(Student's distribution with $n-1$ degrees of freedom).

Example 4 (Exponential family).

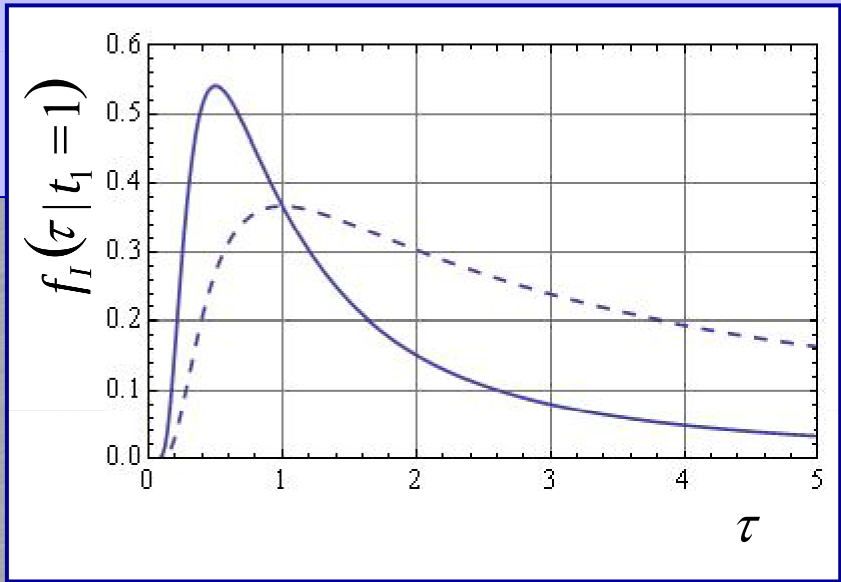
$x_1 \equiv \ln t_1\,,\ \mu \equiv \ln \tau$

$\Rightarrow\ f_{I'_{\sigma=1}}(x_1\,|\,\mu) = e^{x_1-\mu}\exp\{-e^{x_1-\mu}\};\quad I' = \{f_{I'_{\sigma=1}}(x\,|\,\mu) = e^{x-\mu}\exp(-e^{x-\mu});\ \mu\in\mathbb{R}\}$

$\Rightarrow\ \zeta_{I'_{\sigma=1}}(\mu) = 1$

$\Rightarrow\ \zeta_I(\tau) = \tau^{-1}$

$\Rightarrow\ f_I(\tau\,|\,t_1) = \dfrac{t_1}{\tau^2}\exp\left\{-\dfrac{t_1}{\tau}\right\}.$



7. Consistency factors in the absence of symmetry:

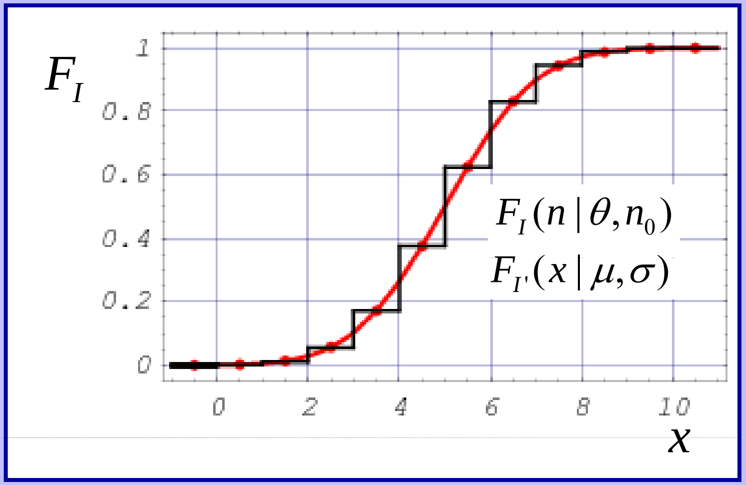
Example 5 (Binomial vs. Normal distribution).

$$p_I(n \mid \theta, n_0) = \binom{n_0}{n} \theta^n (1-\theta)^{n_0 - n} \; ; \quad \begin{matrix} n_0 \in \mathbb{N} \\ n \in \mathbb{N}_0 ; n \le n_0 \end{matrix}$$

$$n_0\theta, n_0(1-\theta) \gg 1 : \quad F_I(n \mid \theta, n_0) = \sum_{i=0}^{n} p(i \mid \theta, n_0) \simeq \int_{-\infty}^{n+0.5} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(x'-\mu)^2}{2\sigma^2} \right\} dx'$$

$$\mu = n_0\theta \,, \quad \sigma = \sqrt{n_0\theta(1-\theta)} \approx \sqrt{n(n_0 - n)/n_0}$$

$$\Rightarrow f_I(\theta \mid \sigma, n) \approx \frac{n_0}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{(n-\mu)^2}{2\sigma^2} \right\} .$$



VII. Interpretation of Inverse Probability Distributions:

1. Credible intervals and regions,
2. Degrees of belief and calibration,
3. Fiducial argument,
4. Theorem of Stein, Chang and Villegas,
5. Calibration of marginal distributions.

1. Credible intervals and regions :

Definition 3.a (Credible interval): $\exists\, f_I(\theta \mid x);\ \theta, x \in \mathbb{R}$

$\Rightarrow$ Credible interval $(\theta_a, \theta_b] \subseteq V_\Theta$ with (inverse-) probability content δ :

$$\int_{\theta_a}^{\theta_b} f_I(\theta \mid x)\, d\theta = \delta\,.$$

Definition 3.b (Credible region): $\exists\, f_I(\boldsymbol{\theta} \mid \mathbf{x});\ \boldsymbol{\theta} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n$

$\Rightarrow$ Credible region $U(\mathbf{x}) \subseteq V_{\boldsymbol{\Theta}} \subseteq \mathbb{R}^m$ with probability content δ :

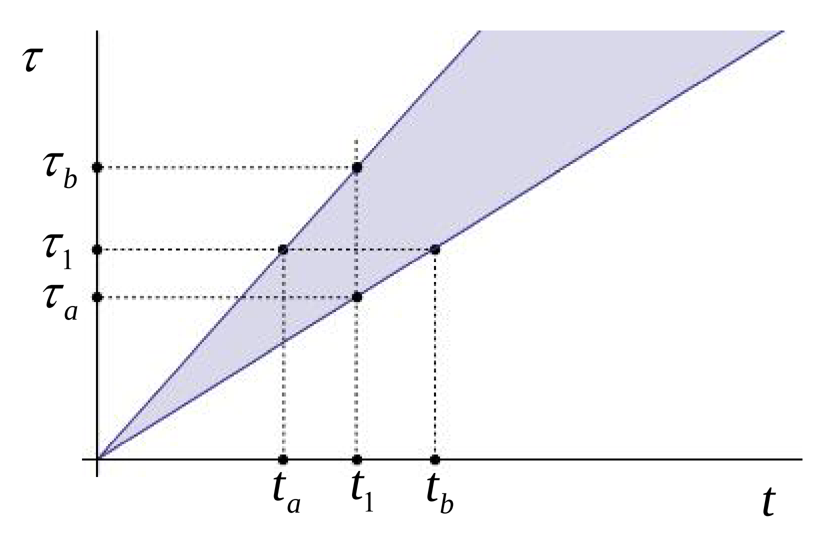
$$\int_U f_I(\boldsymbol{\theta} \mid \mathbf{x})\, d^m\boldsymbol{\theta} = \delta\,.$$

2. Degrees of belief and calibration :

The inferred parameter, although unknown, is usually fixed. What is distributed is *our degree of belief* that, given observed $\mathbf{x}$, the *true value* of the inferred parameter is in the credible region $U(\mathbf{x})$.

Definition 4 (Calibration): *The inverse probability density* $f_I(\boldsymbol{\theta} \mid \mathbf{x})$ *is called* calibrated *if there is at least one algorithm to construct credible regions whose (inverse-) probability content is equal to the long-term relative frequency, called* coverage, *of the regions that cover the true value(s) of the parameter, be the distribution of the true values in the* ensemble *what it may.*

3. Fiducial argument.



1) $\alpha \in [0, 1-\gamma]$

2) τ_1

3) $t_a : F_I(t_a \mid \tau_1) = \alpha$

4) $t_b : F_I(t_b \mid \tau_1) = \alpha + \gamma$

$\Rightarrow \Pr_I(t_a < t \le t_b \mid \tau_1)$

$= F_I(t_b \mid \tau_1) - F_I(t_a \mid \tau_1)$

$= \gamma$

5) $\tau \in (0, \infty)$

6) t_1; τ_1 true value

7) $\tau_1 \in (\tau_a, \tau_b) \Leftrightarrow t_1 \in (t_a, t_b)$

Remark 4. $$\left.\begin{aligned} F_I(t_1 \mid \tau_b) &= \alpha \\ F_I(t_1 \mid \tau_a) &= \alpha + \gamma \end{aligned}\right\} \Rightarrow \gamma = F_I(t_1 \mid \tau_a) - F_I(t_1 \mid \tau_b).$$

$$\left.\begin{aligned} F_I(\tau_a \mid t_1) &= \alpha \\ F_I(\tau_b \mid t_1) &= \alpha + \delta \end{aligned}\right\} \Rightarrow \delta = F_I(\tau_b \mid t_1) - F_I(\tau_a \mid t_1) = \int_{\tau_a}^{\tau_b} f_I(\tau \mid t_1)\, d\tau\ .$$

Proposition 2 (Fiducial argument). $\delta = \gamma$ for all admissible α, γ

$\Rightarrow f_I(\tau \mid t_1) = -\partial_\tau F_I(t_1 \mid \tau).$

Proof. For infinitesimal λ, $\delta = f_I(\tau_a \mid t_1)\Delta\tau$ and $\gamma = -\Delta\tau\, \partial_\tau F_I(t_1 \mid \tau)\big|_{\tau=\tau_a}$.

□

Remark 1. General argument. .

Example 6. $\zeta_{I,\mu}(\sigma) = \sigma^{-1}$, $\zeta_{I,\sigma}(\mu) = 1$ satisfy the fiducial condition.

Proposition 3 (Lindley). $f_I(\theta \mid x) = \pm\partial_\theta F_I(x \mid \tau)$ and

$$f_I(\theta \mid x) = \frac{\zeta_I(\theta)}{\eta_I(x)} f_I(x \mid \theta) \Rightarrow \exists\, \mu(\theta), y(x) : f_{I'}(y \mid \mu) = \phi(y - \mu).$$

Proof. D.Lindley (1958), J. Roy. Statist Soc. Ser. **B 20**, pp. 102-107. □

4. Theorem of Stein, Chang and Villegas:

Theorem 4 (Stein, Chang, Villegas). $f_I(\mathbf{x} \mid \boldsymbol{\theta})$ invariant under a topological (e.g., Lie) group G, $\zeta_I(\boldsymbol{\theta})$ coincides with the element of the right - invaiant Haar measure on $G \Rightarrow f_I(\boldsymbol{\theta} \mid \mathbf{x})$ calibrated on equivariant credible regions, $U[\boldsymbol{I}(a, \mathbf{x})] = \bar{\boldsymbol{I}}[U(\mathbf{x})]$.

Proof. T. Chang, C. Villegas (1986), Canad. J. Statist. **14**, 289 - 296.

□

Example 7. $\zeta_I(\mu, \sigma) = \sigma^{-1}$ is the element of the right - invariant Haar measure on $G = \mathbb{R} \times \mathbb{R}^+$.

Example 7 (Cont'd). Equivariant rectangles $U = [\mu_a, \mu_b] \times [\sigma_a, \sigma_b]$,

$\Pr_I(U_1 \mid \mathbf{x}) = \alpha$, $\Pr_I(U_2 \mid \mathbf{x}) = \beta$, $\Pr_I(U_3 \mid \mathbf{x}) = \varepsilon$, $\Pr_I(U \mid \mathbf{x}) = \gamma$,

$U_1 \equiv (-\infty, \mu_a] \times \mathbb{R}^+$, $U_2 \equiv [\mu_a, \mu_b] \times \mathbb{R}^+$, $U_3 \equiv [\mu_a, \mu_b] \times (0, \sigma_a]$,

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

$\mathbf{x} \equiv (x_1, \ldots, x_n)$, $n \ge 2$.

5. Calibration and marginal distributions:

Example 7 (Cont'd). Equivariant rectangles $U = [\mu_a, \mu_b] \times [\sigma_a, \sigma_b]$,

$$\Pr_I(U_1 \mid \mathbf{x}) = \alpha,\ \Pr_I(U_2 \mid \mathbf{x}) = \beta,\ \Pr_I(U_3 \mid \mathbf{x}) = \varepsilon,\ \Pr_I(U \mid \mathbf{x}) = \gamma,$$

$$U_1 \equiv (-\infty, \mu_a] \times \mathbb{R}^+,\ U_2 \equiv [\mu_a, \mu_b] \times \mathbb{R}^+,\ U_3 \equiv [\mu_a, \mu_b] \times (0, \sigma_a],$$

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

$\beta = \gamma \Rightarrow U = \mathbb{R} \times [\sigma_a, \sigma_b] \Rightarrow f_I(\sigma \mid \mathbf{x})$ calibrated,

$\beta = 1 \Rightarrow U = [\mu_a, \mu_b] \times \mathbb{R}^+ \Rightarrow f_I(\mu \mid \mathbf{x})$ calibrated.

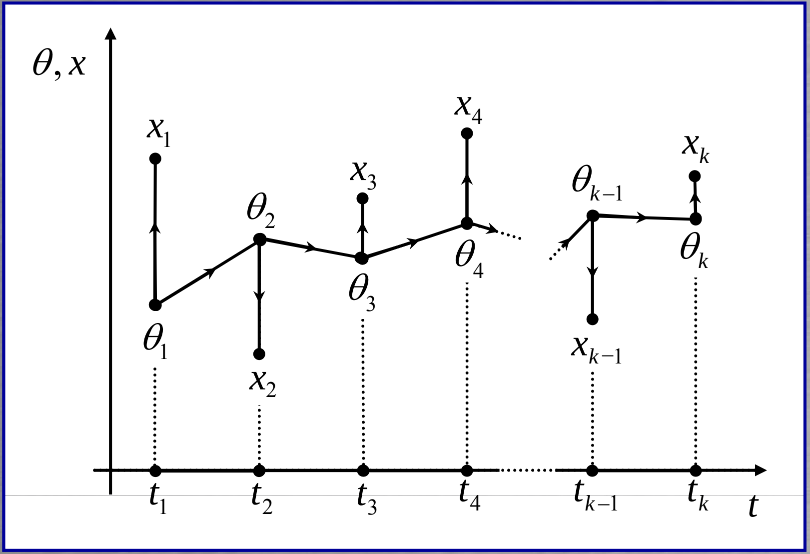
Remark 2. Not conserved under general one-to-one reparametrization.

3.a Kalmanov filter – osnovne predpostavke

$$\mathbf{\Theta}_k \equiv \{\theta_1,\dots,\theta_k\}; \quad \theta_j = \theta(t_j) \in \mathbb{R}^n$$
$$\mathbf{X}_k \equiv \{x_1,\dots,x_k\}; \quad x_j = x(t_j) \in \mathbb{R}^n$$

$$f(\theta_{j+1} | \mathbf{\Theta}_j, \mathbf{X}_j) = f(\theta_{j+1} | \theta_j) \sim N(A_j\theta_j, Q_j)$$

$f(\theta_{j+1} | \theta_j)$: frekvenčna interpretacija



$$f(x_j | \mathbf{\Theta}_j, \mathbf{X}_{j-1}) = f(x_j | \theta_j) \sim N(\theta_j, V_j)$$

$f(x_j | \theta_j)$: frekvenčna interpretacija

$$f(x_j | \theta_j) \sim N(H_j\theta_j, V_j)$$

H_j: obrnljiva

$$f(\theta_k | \mathbf{X}_k) = ?$$

$$\exists\ f(\theta_1 \mid x_1) \overset{?}{\sim} N(r_1, R_1); \quad r_1 = r_1(x_1) \in \mathbb{R}^n$$

$$\exists\ f(\boldsymbol{\Theta}_j, x_j \mid \mathbf{X}_{j-1}); \quad j = 2,\dots,k$$

$$\Rightarrow f(\boldsymbol{\Theta}_k \mid \mathbf{X}_k) = f(\theta_1 \mid x_1) \prod_{j=2}^{k} \frac{f(x_j \mid \theta_j) f(\theta_j \mid \theta_{j-1})}{f(x_j \mid \mathbf{X}_{j-1})}$$

$$\Rightarrow f(\theta_k \mid \mathbf{X}_k) = \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} f(\boldsymbol{\Theta}_k \mid \mathbf{X}_k) d^n\theta_1 \dots d^n\theta_{k-1} \sim N(r_k, R_k)$$

$$r_j = A_{j-1} r_{j-1} + R_j V_{j-1}^{-1}(x_j - A_{j-1} r_{j-1}); \; r_j = r_j(\mathbf{X}_j) \in \mathbb{R}^n$$

$$R_j^{-1} = V_j^{-1} + (A_{j-1} R_{j-1} A_{j-1}^T + Q_{j-1})^{-1}$$

T. N. Thiele (1880). *Om Anvendelse af mindste Kvadraters Methode...*, § 2.
P. Swerling (1959). J. Astronaut. Sci. **6**, str. 46-52.
R. L. Stratonovich (1959). Radiofizika **2:6**, str. 892-901.
-- (1960). Radio Eng. Electr. Phys. **5:11**, str. 1-19.
R. E. Kalman (1960). Trans. ASME J. Basic Eng. 82, str. 34-45.

3.b Inicializacija Kalmanovega filtra

$$\nexists\ f(\theta_1) \Rightarrow f(\theta_1 \mid x_1) = \frac{\varsigma(\theta_1)\, f(x_1 \mid \theta_1)}{\eta(x_1)}$$

Invarianca $f(x_1 \mid \theta_1)$ glede na translacije in inverzijo

+

objektivnost

$\Downarrow$

Rešitev funkcijskih enačb za $\varsigma(\theta_1)$ **:**

$$\varsigma(\theta_1) = 1 \Rightarrow f(\theta_1 \mid x_1) = f(x_1 \mid \theta_1)$$

$$\Rightarrow f(\theta_1 \mid x_1) \sim N(r_1, R_1); \quad r_1 = x_1\ ,\ R_1 = V_1$$

3.c Interpretacija $f(\theta_k|\mathbf{X}_k)$:

$f(\theta_1|x_1) \neq$ porazdelitev θ_1

$\Rightarrow f(\theta_k|\mathbf{X}_k) \neq$ porazdelitev θ_k

Merljive napovedi:

$f(\theta_k|\mathbf{X}_k) = f(r_k|\theta_k)$;

$f(r_k|\theta_k) \sim N(\theta_k, R_k)$; invariantnost na translacije;

$f(r_k|\theta_k)$: frekvenčna interpretacija.

$C(r_k)$ = ekvivariantno območje zaupanja:

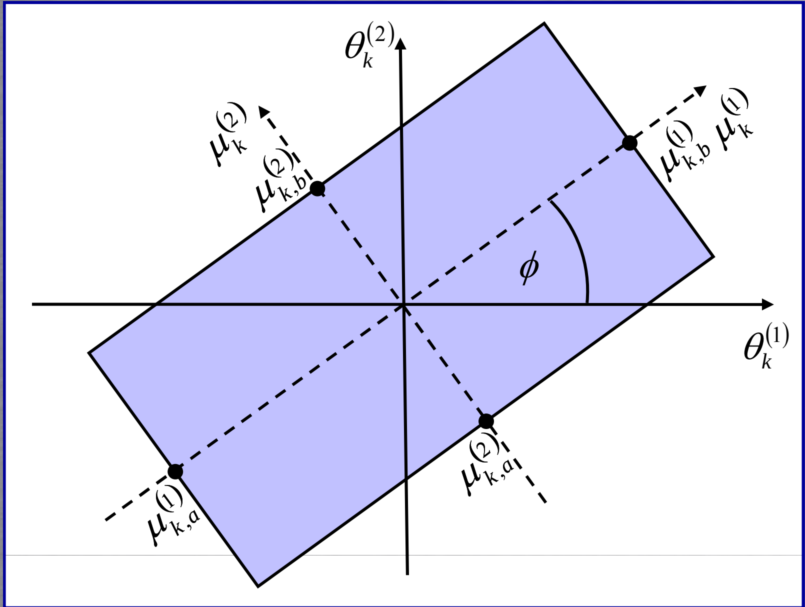
$C(r_k + a) = C(r_k) + a:\ a \in \mathbb{R}^n$.

$\int_C f(\theta_k|\mathbf{X}_k)\, d^n\theta_k = \gamma$ = delež $C(r_k)$, ki vsebujejo resnične θ_k, ne glede na porazdelitev θ_k v ansamblu.

C. Stein (1965). Proc. Int. Research Seminar UC Berkeley, str. 217-240.

T. Chang, C. Villegas (1986). Canad. J. Statist. **14**, str. 289-296.

Ekvivariantni (n-dim.) pravokotniki



$$\theta_k = \left(\theta_k^{(1)},\ldots,\theta_k^{(n)}\right),$$

$$f\left(\theta_k^{(i)} | r_k\right) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f\left(\theta_k | r_k\right) d\theta_k^{(1)} \ldots d\theta_k^{(i-1)} d\theta_k^{(i+1)} \ldots d\theta_k^{(n)},$$

$$A \equiv \left(\theta_{k,a}^{(i)}, \theta_{k,b}^{(i)}\right): \int_{-\infty}^{\theta_{k,a}^{(i)}} f\left(\theta_k^{(i)} | r_k\right) d\theta_k^{(i)} = \alpha,$$

$$\int_{\theta_{k,a}^{(i)}}^{\theta_{k,b}^{(i)}} f\left(\theta_k^{(i)} | r_k\right) d\theta_k^{(i)} = \gamma$$

$\Rightarrow \gamma = P\left(A | r_k\right) \neq$ delež $A(r_k)$, ki vsebujejo resnične $\theta_k^{(i)}$.

$\exists\, O_k : O_k^T = O_k^{-1}$, $O_k R_k O_k^T = D_k$ = diagonalna;

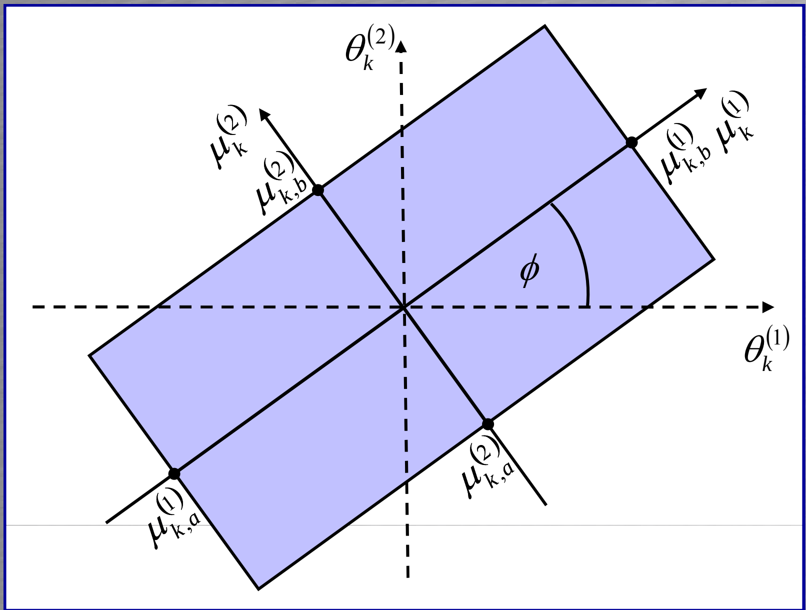
$\mu_k \equiv O_k \theta_k$, $s_k \equiv O_k r_k$;

$\mu_k = \left(\mu_k^{(1)},\ldots,\mu_k^{(n)}\right)$, $B \equiv \left(\mu_{k,a}^{(i)}, \mu_{k,b}^{(i)}\right)$

$\Rightarrow \gamma = P\left(B | s_k\right) =$ delež $B(s_k)$, ki vsebujejo resnične $\mu_k^{(i)}$.

$$\exists\, O_k : O_k^T = O_k^{-1}\ ,\ \ O_k R_k O_k^T = D_k\ ; \mu_k \equiv O_k \theta_k\ ,\ \ s_k \equiv O_k r_k\ ;$$
$$f\left(s_k \mid \mu_k\right) = N\left(\mu_k, D_k\right),\ \ f\left(\mu_k \mid s_k\right) = N\left(s_k, D_k\right).$$

Ekvivariantni (n-dim.) pravokotniki



$$\int_{-\infty}^{\mu_{k,a}^{(1)}} d\mu_k^{(1)} \int_{-\infty}^{\infty} d\mu_k^{(2)} \cdots \int_{-\infty}^{\infty} d\mu_k^{(n)}\, f\left(\mu_k \mid s_k\right) = \alpha_1\ ,$$

$$\int_{\mu_{k,a}^{(1)}}^{\mu_{k,b}^{(1)}} d\mu_k^{(1)} \int_{-\infty}^{\infty} d\mu_k^{(2)} \cdots \int_{-\infty}^{\infty} d\mu_k^{(n)}\, f\left(\mu_k \mid s_k\right) = \beta_1\ ,$$

$$\vdots$$

$$\int_{\mu_{k,a}^{(1)}}^{\mu_{k,b}^{(1)}} d\mu_k^{(1)} \int_{\mu_{k,a}^{(2)}}^{\mu_{k,b}^{(2)}} d\mu_k^{(2)} \cdots \int_{\mu_{k,a}^{(n)}}^{\mu_{k,b}^{(n)}} d\mu_k^{(n)}\, f\left(\mu_k \mid s_k\right) = \beta_n = \gamma\ ;$$

$$\theta_k = \left(\theta_k^{(1)}, \ldots, \theta_k^{(n)}\right),\ \ \mu_k = \left(\mu_k^{(1)}, \ldots, \mu_k^{(n)}\right);$$

$$\alpha_j, \beta_j \in [0,1],\ 1-\alpha_1 \geq \beta_1 \geq 1-\alpha_2 \geq \beta_2 \geq \ldots \geq \gamma\ .$$

$\alpha_{j-1} = 0,\ \beta_{j-1} = 1,\ \beta_j = \gamma \Rightarrow f\left(\mu_k^{(j)} \mid s_k\right) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f\left(\mu_k \mid s_k\right) d\mu_k^{(1)} \ldots d\mu_k^{(j-1)} d\mu_k^{(j+1)} \ldots d\mu_k^{(n)}$ kalibrirane

$f\left(\theta_k^{(j)} \mid s_k\right) = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f\left(\theta_k \mid s_k\right) d\theta_k^{(1)} \ldots d\theta_k^{(j-1)} d\theta_k^{(j+1)} \ldots d\theta_k^{(n)}$ **niso** kalibrirane

3.b Inicializacija Kalmanovega filtra

$$\nexists\ f(\theta_1) \Rightarrow f(\theta_1 \mid x_1) = \frac{\varsigma(\theta_1)\, f(x_1 \mid \theta_1)}{\eta(x_1)}$$

Invarianca $f(x_1 \mid \theta_1)$**:**

$$l_1(a, x_1) = x_1 + a,\ \bar{l}_1(a, \theta_1) = \theta_1 + a;\ a \in \mathbb{R}$$
$$l_2(x_1) = -x_1, \qquad \bar{l}_2(\theta_1) = -\theta_1$$

Rešitev funkcijskih enačb za $\varsigma(\theta_1)$ **:**

$$\varsigma(\theta_1) = 1 \Rightarrow f(\theta_1 \mid x_1) = f(x_1 \mid \theta_1)$$
$$\Rightarrow f(\theta_1 \mid x_1) \sim N(r_1, R_1); \quad r_1 = x_1,\ R_1 = V_1$$

Text box
With shadow