Hellinger Information Matrix in Parametric Estimation and Objective Priors

Arkady Shemyakin

University of St. Thomas

Borovkov Meeting, 2022

Outline

1. Hellinger Information
2. Information Inequalities
3. Objective Priors
4. Hellinger Priors
5. Optimal Design
6. Hellinger Information Matrix
7. Examples
8. References

1. Hellinger Information

A family of probability measures $\{P_\theta, \theta \in \Theta \subset \mathbf{R}\}$ is defined on a measurable space $(\mathcal{X}, \mathcal{B})$ so that all the measures from the family are absolutely continuous with respect to some σ-finite measure λ on $\mathcal{B}$. The square of Hellinger distance between any two parameter values can be defined in terms of densities $p(x,\theta_i) = \dfrac{dP_{\theta_i}}{d\lambda}$ can be defined as

$$h(\theta_1, \theta_2) = \left(\int_{\mathcal{X}} \left(\sqrt{p(x;\theta_1)} - \sqrt{p(x;\theta_2)} \right)^2 \lambda(dx) \right)^{1/2}$$

The following expression for the case of i.i.d. finite samples $x = (x_1, \ldots, x_n)$ and $p^{(n)}(x;\theta) = \dfrac{dP_\theta^{(n)}}{d\lambda^n}$:

$$h^2\left(p^{(n)}(x;\theta_1), p^{(n)}(x;\theta_2)\right) = 2\left[1 - \left(1 - \tfrac{1}{2} h^2\left(p(x;\theta_1), p(x;\theta_2)\right)\right)^n\right]$$

helps to reduce the analysis of finite samples to the case of sample size 1.

If for almost all θ from Θ (w.r.t. measure λ) there exists $\alpha \in (0, 2]$ (***index of regularity***) such that

$$\lim_{\varepsilon \to 0} h(\theta, \theta + \varepsilon) \cdot |\varepsilon|^{-\alpha} = j(\theta),$$

we will define ***Hellinger information at a point*** θ as

$$J(\theta) = j^{2/\alpha}(\theta).$$

Alternatively, the definition can be reformulated as

$$\lim_{\varepsilon \to 0} h^{2/\alpha}(\theta, \theta + \varepsilon) \cdot |\varepsilon|^{-2} = J(\theta).$$

Notice that in regular situations (**Ibragimov and Has'minskii** (1976): $p(x;\theta)$ is twice continuously differentiable w.r.t. θ for almost all $x \in \mathcal{X}$ w.r.t. λ, $E\left(\frac{\partial^2}{\partial\theta^2} p\left(x;\theta\right)\right)^2 < \infty$, Fisher information $I\left(\theta\right) = E\left(\frac{\partial}{\partial\theta}\log p\left(x;\theta\right)\right)^2$ is continuous, strictly positive and finite for almost all θ from Θ) it is true that $\alpha = 2, \quad J\left(\theta\right) = \frac{1}{4} I\left(\theta\right)$.

We will be interested in nonregular situations (including uniform densities $p(x;\theta)$ with support depending on parameter), for which Fisher information is not defined, and Hellinger information may function as a substitute.

Using classification of **Ibragimov and Has'minskii** (1976), $1 < \alpha < 2$ corresponds to singularities of the first and the second type, $\alpha = 1$ to densities with jumps, and $0 < \alpha < 1$ to singularities of the third type.

2. Information Inequalities

We will define quadratic Bayes risk for an independent identically distributed sample $X^{(n)} = \left(X_1,...X_n\right)$ of size n from the model considered above with a prior $\pi\left(\theta\right)$ as

$$R\left(\theta^*\left(X^{(n)}\right)\right) = \int\limits_{\mathbf{X}^{(n)}\times\Theta} \left(\theta^*\left(x^{(n)}\right)-\theta\right)^2 p\left(x^{(n)};\theta\right)\pi\left(\theta\right)\lambda^n(dx)d\theta.$$

Let us consider an integral version of the classical Cramér-Frechet-Rao inequality, which under certain regularity conditions leads to the following asymptotic lower bound for the Bayes risk in terms of Fisher information $I\left(\theta\right) = E\left(\frac{\partial}{\partial\theta}\log p\left(x;\theta\right)\right)^2$

$$R\left(\theta^*\left(X^{(n)}\right)\right) \geq n^{-1}\int\limits_{\Theta} I^{-1}\left(\theta\right)\pi\left(\theta\right)d\theta + O\left(n^{-2}\right)$$

This lower bound, which can be proven to be tight, was obtained by **Borovkov and Sakhanenko** (1980), see also **Bobrovsky et al.** (1987), **Brown and Gajek** (1990).

This bound can be extended to the nonregular case, when Fisher information may not exist. One of such extensions is ***Hellinger information inequality*** providing an asymptotic lower bound

$$\inf_{\theta^*\left(X^{(n)}\right)} R\left(\theta^*\left(X^{(n)}\right)\right) \geq C(\alpha)\, n^{-2/\alpha} \int_{\Theta} J^{-1}(\theta)\, \pi(\theta)\, d\theta + o\left(n^{-2/\alpha}\right),$$

obtained in **Shemyakin** (1992) under the assumptions of Hellinger information $J(\theta)$ being strictly positive, almost surely continuous and bounded on any compact subset of Θ, and

$$\int_{\Theta} J^{-1}(\theta)\pi(\theta)\, d\theta < \infty \qquad ,$$

where Θ is an open subset of real numbers and the constant $C(\alpha)$ is related to technical details of the proof and is not necessarily tight.

Why was nobody in the Bayesian world interested? One possible reason: double integration in the definition of Bayes' risk is not a popular approach. Classical statistics does not integrate over parameter, Bayesian statistics does not integrate over data.

3. Objective Priors

Bayesian statistical analysis (both estimation and hypothesis testing) is based on the use of continuous version of the Bayes formula

$$\pi(\theta|x) = \frac{p(x;\theta)\pi(\theta)}{\int p(x;t)\pi(t)\,dt} \propto p(x;\theta)\pi(\theta),$$

where

$x -$ data (fixed)

$\theta -$ parameter (random)

$p(x;\theta) -$ likelihood

$\pi(\theta) -$ prior

$\pi(\theta|x) -$ posterior

How to choose priors?

A. **Subjective Bayes** (elicitation of priors using background information)

B. **Empirical Bayes** (double-dipping: using data to obtain priors)

C. **Objective Bayes** (in absence of definitive information, determining priors based on certain mathematical criteria: opting for universal, non-intrusive, non-informative options)

Motivation for objective Bayes: multiparameter problems with sparse prior information on some parameters (components of the vector parameter). Objective priors are not always evident: say, for $X \sim Bin(n,\theta),\ \pi(\theta) = Beta(1/2, 1/2) \propto \dfrac{1}{\sqrt{\theta(1-\theta)}}$ - horseshoe (bathtub) shape, not uniform (as could be expected).

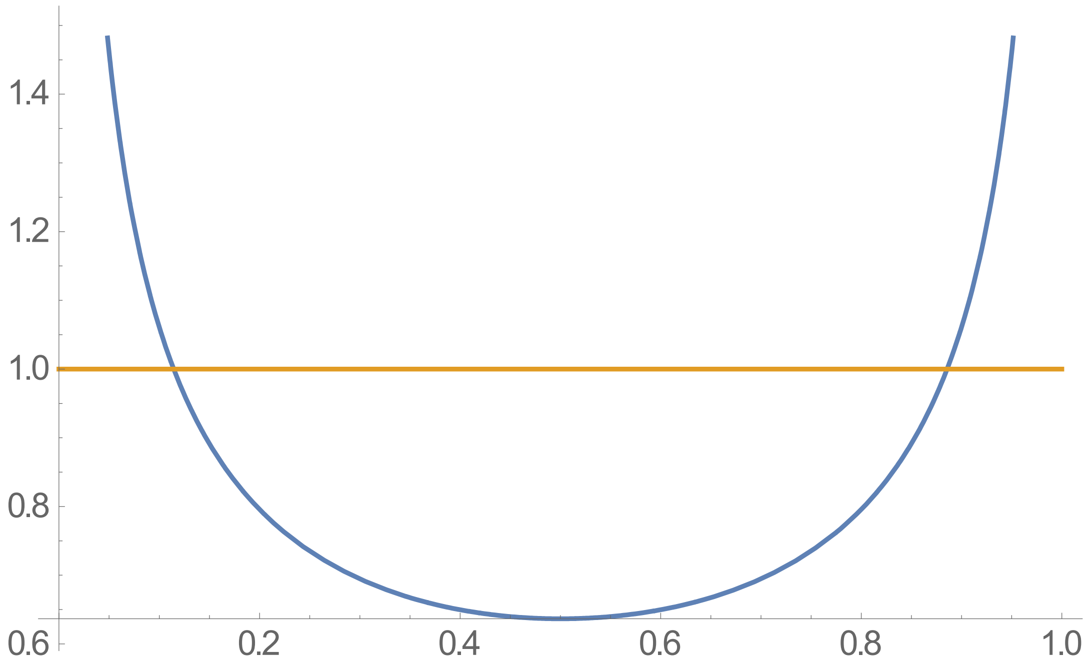

Figure 1. Beta density: uniform $Beta(1,1)$ and horseshoe (bathtub) $Beta(1/2,1/2)$.

Further: let us suggest some information criteria which can help to determine such non-informative “objective” priors.

Three most popular approaches:

A. The Jeffreys Rule: **Jeffreys** (1946).

B. Probability Matching Priors: **Ghosal, Ghosh and Samanta** (1995), **Ghosal** (1999).

C. Reference Priors: **Berger, Bernardo** (1992), **Ghosal and Samanta** (1997).

A. The Jeffreys Rule: $\pi(\theta) \propto \sqrt{I(\theta)}$

B. Probability Matching Priors (PMP)

Posterior probabilities of certain regions (credible sets) coincide (exactly or approximately) with their coverage probabilities by confidence sets:

$$P\left(\theta \in A_\alpha(X) \middle| \theta\right) = \alpha = P\left(\theta \in A_\alpha(x) \middle| x\right).$$

C. Reference Priors

Reference prior is both

- permissible (does not lead to an improper posterior);
- maximizes global **Kullback-Leibler divergence** between the prior and the posterior.

For many regular parameter families, probability matching and reference priors both satisfy the Jeffreys rule. However, it is not necessary in case of multi-parametric families and the loss of regularity. Let us focus on the nonregular case, when Fisher information may not be defined. Main progress in this direction was achieved in **Berger, Bernardo and Sun** (2009).

4. Hellinger Priors: One Parameter

Define ***Hellinger prior*** for the parametric set Θ as (**Shemyakin**, 2011, 2012, 2014)

$$\pi_H(\theta) \propto \sqrt{J(\theta)} = j^{1/\alpha}(\theta).$$

Hellinger priors will often coincide with well-known priors obtained by different approaches. However, there are some distinctions. A special role might be played by Hellinger priors in case when Fisher information is not defined.

Example 1: Uniform $X \sim Unif\left(\theta^{-1}, \theta\right), \theta \in (1, \infty)$.

$$\alpha = 1, \qquad \pi_H(\theta) \propto j(\theta) = \frac{\theta^2 + 1}{\theta\left(\theta^2 - 1\right)}$$

The same prior can be constructed as the probability matching prior (**Ghosal**, 1999) or the reference prior (**Berger, Bernardo, and Sun**, 2009).

Example 2: Uniform $X \sim Unif\left(\theta, \theta^2\right), \theta \in (1, \infty)$

$$\alpha = 1, \quad \pi_H(\theta) \propto j(\theta) = \frac{2\theta + 1}{\theta(\theta - 1)}.$$

This prior is different from the reference prior obtained by **Berger et al.** (2009):

$$\pi_{R1}(\theta) \propto \frac{2\theta - 1}{\theta(\theta - 1)} \exp\left\{\psi\left(\frac{2\theta}{2\theta - 1}\right)\right\},$$

where $\psi(z)$ is the digamma function defined as $\psi(z) = \frac{d}{dz}\log\Gamma(z), z > 0$. It is also different from the prior obtained by **Tri Minh Le** (2013) by a similar approach, maximizing Hellinger distance between prior and posterior:

$$\pi_{R2}(\theta) \propto \frac{(2\theta + 1)^2 (2\theta - 1)^5}{\theta^3 (\theta - 1)\left[20\sqrt{\pi}(2\theta - 1)^2 - 2(2\theta + 1)^2\right]}.$$

However, all three priors exhibit a very similar behavior (to within a multiplicative constant):

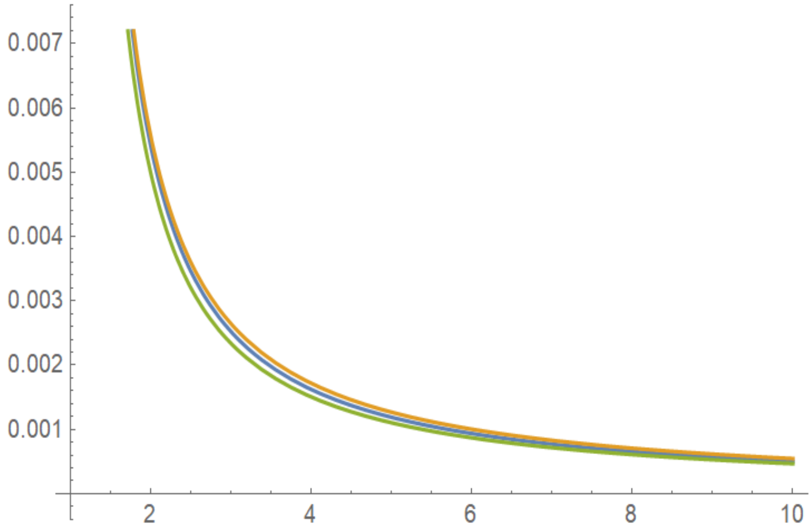

Figure 2. Hellinger prior (blue), reference priors (green and yellow).

Problem: Hellinger prior is not obtained as the solution of some optimization problem (like reference priors). Also, how does it work in case of vector parameter?

5. Optimal Experimental Design

Polynomial model of experimental design (**Smith**, 1994) may be presented as

$$y_i = \sum_{k=0}^{q} \theta_k x_i^k + \varepsilon_i,\ i = 1, \ldots, n,$$

where x_i are scalars, $\theta \in \mathbf{R}^{q+1}$ is the unknown parameter of interest and errors ε_i are non-negative i.i.d variables with density $p_0(y;\alpha) \sim \alpha c(\alpha) y^{\alpha-1},\ y \to 0,\ \alpha \geq 1$ (e.g., Weibull or Gamma). The space of balanced designs is defined as

$$\Xi = \left\{ \xi = (w_i, x_i) : \sum_{i=1}^{n} w_i x_i = 0, x_i \in [-A, A] \right\},$$

and there exist several definitions of optimal design. **Lin, Martin and Yang** (2018, 2019) suggest using criterion

$$\xi_{opt} = \arg\max_{\xi} J_{\xi}(\theta),\ \ J_{\xi}(\theta) = \inf_{u} J_{\xi}(\theta; u),$$

where $J(\theta;u)$ is ***Hellinger information*** in the direction $u \in \mathbf{R}^{q+1}$, defined as

$$\lim_{\varepsilon \to 0} \frac{h(\theta, \theta + \varepsilon u)}{|\varepsilon|^{\alpha}} = J(\theta;u)$$

(generalizing the definition on page 4). In case of nonregular polynomial design $\alpha \in [1,2)$, Lin et al. prove that if for some $\delta > 0$ it holds that

$$\int_{\delta}^{\infty} \left(\frac{d}{dy} \log p_0(y;\alpha) \right)^2 p_0(y;\alpha) < \infty,$$

Hellinger information for design ξ can be represented as

$$J_{\xi}(\theta;u) = c(\alpha)\left[1 + \alpha r(\alpha)\right] \sum_{i=1}^{n} \left\{ w_i \left| \sum_{k=0}^{q} x_i^k u_{k+1} \right|^{\alpha} \right\},$$

where

$$r(\alpha) = \int_{0}^{\infty} \left[(w+1)^{(\alpha-1)/2} - (w)^{(\alpha-1)/2} \right]^2 dw.$$

6. Hellinger Information Matrix

Extending definitions of section 1 to the multi-parameter case $\Theta \subset \mathbf{R}^m, m = 1, 2, \ldots$ we will introduce ***Hellinger distance matrix*** with elements

$$H_{ij}(\theta, \mathbf{U}) = \int_X \left(\sqrt{p(x;\theta)} - \sqrt{p(x;\theta + \mathbf{u}_i}\right)\left(\sqrt{p(x;\theta)} - \sqrt{p(x;\theta + \mathbf{u}_j}\right) d\lambda$$

where $\mathbf{U}$ is an $m \times m$ matrix with columns $\mathbf{u}_i$. Define also vectors $\alpha = (\alpha_1, \ldots \alpha_m)^T$ (index of regularity with components $0 < \alpha_i < 2$) and $\delta = (\delta_1, \ldots \delta_m)^T$ with components $\delta_i = \varepsilon^{2/\alpha_i}$, $\mathbf{\Delta} = Diag(\delta)$ such that for all $i = 1, \ldots, m$ there exist finite non-degenerate limits

$$0 < \lim_{\varepsilon \to 0} |\varepsilon|^{-2} H_{ii}^{2/\alpha_i}(\theta, \mathbf{\Delta}) < \infty,$$

Then ***Hellinger information matrix*** $J(\theta)$ will be defined by its components

$$J_{ij}(\theta) = \lim_{\varepsilon \to 0} |\varepsilon|^{-2} H_{ij}^{(1/\alpha_i + 1/\alpha_j)}(\theta, \mathbf{\Delta}),$$

If $J > 0,$ ***vector Hellinger prior*** can be defined as

$$\pi_H(\theta) \propto \sqrt{\det J(\theta)}.$$

In case of all components $\alpha_i \equiv 2$, Hellinger information reduces to Fisher information matrix and "vector Hellinger" approach leads to the Jeffreys prior (Shemyakin, 2014).

Lin et al. notice that in their experimental design context index of regularity is the same for all components of vector parameter. For more complex nonregularities one should define Hellinger information matrix. It was done in 2014 paper without a rigorous study of its properties: is it true that $J(\theta)$ is positive definite?

Theorem 1:

Let $\boldsymbol{\alpha} = \left(2,...,2,\alpha_m\right), \alpha_m \in [1,2)$, $J_{ij}\left(\theta\right) = I_{ij}\left(\theta\right), i,j = 1,...,m-1; J_{im} = 0; J_{mm} \neq 0.$

$$J\left(\theta\right) = \begin{bmatrix} \bullet & & \bullet & 0 \\ & I\left(\theta\right) & & \dots \\ \bullet & & \bullet & 0 \\ 0 & \dots & 0 & J_{mm}\left(\theta\right) \end{bmatrix}$$

Then, if $\mathbf{E}\, I^{-1}\left(\theta\right) < \infty, \mathbf{E}\, J_{mm}^{-1}\left(\theta\right) < \infty,$

$$(1)\ J\left(\theta\right) > 0$$

$$(2)\ \inf_{\theta^*} R\left(\theta^*\right) \succ C_1\left(\boldsymbol{\alpha}\right) Diag\left(n^{-2/\boldsymbol{\alpha}}\right) \mathbf{E}\, J^{-1}\left(\theta\right)$$

where matrix ordering $A\left(n\right) \succ B\left(n\right)$ is understood as "asymptotic nonnegative definite property"

$$\inf_{|v|=1} \sum_{i,j=1}^{m} \left(A(n) - B(n)\right)_{ij} v_i v_j \geq -\delta_n, \quad \lim_{n\to\infty} \delta_n \left\|A(n) - B(n)\right\|^{-1} = 0,\ \delta_n > 0.$$

Theorem 2:

Let $\alpha_1 = \ldots = \alpha_m = 1$, $J_{ij}(\theta) = 0, i \neq j;\ J_{ii} \neq 0.$

$$J(\theta) = \begin{bmatrix} J_{11}(\theta) & 0 & \ldots & 0 \\ 0 & J_{22}(\theta) & 0 & \ldots \\ \ldots & 0 & \ldots & 0 \\ 0 & \ldots & 0 & J_{mm}(\theta) \end{bmatrix}$$

Then, if $\mathbf{E}\, J_{ii}^{-1}(\theta) < \infty$,

$$(1)\ J(\theta) > 0$$

$$(2)\ \inf_{\theta^*} R(\theta^*) \succ C_2 n^{-2} \mathbf{E}\, J^{-1}(\theta)$$

In both theorem statements, the matrix of Bayes risk is defined as

$$R(\theta^*) = \int_{\mathbf{X}^{(n)} \times \Theta} \left(\theta^*\left(x^{(n)}\right) - \theta\right)\left(\theta^*\left(x^{(n)}\right) - \theta\right)^T p\left(x^{(n)};\theta\right)\pi(\theta)\lambda^n(dx)d\theta.$$

7. Examples

Example 4: Truncated Weibull distribution (see Theorem 1):

$$p(x;\beta,\varphi,\tau)=\beta\varphi^{\beta}x^{\beta-1}\exp\left\{-\varphi^{\beta}\left(x^{\beta}-\tau^{\beta}\right)\right\},\ x\in[\tau,\infty),\ \beta>2,$$

with two parameters of interest: threshold $\tau>0$ and pseudo-scale $\varphi>0$.

A. Using notation $\theta=(\varphi,\tau)$, obtain $\alpha_1=2,\alpha_2=1,\delta_1=\varepsilon,\delta_2=\varepsilon^2$, and

$$H(\theta,\Delta)=\begin{bmatrix}\delta_1\varphi^{-2} & o(\delta_1\delta_2)\\ o(\delta_1\delta_2) & \delta_2\varphi^{\beta}\tau^{\beta-1}\end{bmatrix},\ I_H(\theta)=\begin{bmatrix}\varphi^{-2} & 0\\ 0 & \varphi^{2\beta}\tau^{2\beta-2}\end{bmatrix}$$

Therefore, $\pi_H(\tau,\varphi)\propto\tau^{\beta-1}\varphi^{\beta-1}$, which is also the reference prior for the vector parameter $\theta=(\tau,\varphi)$, see **Ghosal** (1997).

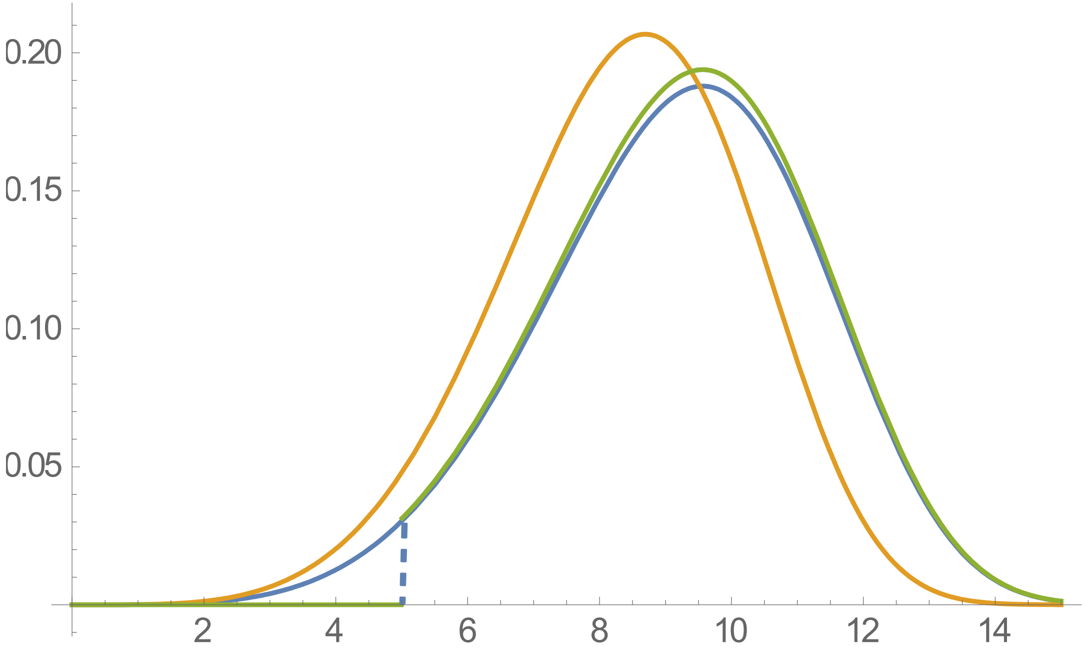

Figure 3. Graphs of $p(x;5,0.1,0)$, $p(x;5,0.1+0.01,0)$, $p(x;5,0.1,0+5)$

Example 5: Truncated Beta in polar coordinates on a disc (see Theorem 1):

$$p\left(r,\varphi;\beta_1,\beta_2,\rho\right)\propto\frac{r}{\pi\rho^2}\left(\frac{\varphi}{2\pi}\right)^{\beta_1-1}\left(1-\frac{\varphi}{2\pi}\right)^{\beta_2-1},0\le\varphi\le 2\pi,r\le\rho,$$

with three parameters of interest: Parameters of Beta $\beta_1,\beta_2>1$ and radius of the disc $\rho>0$.

A. Using notation $\theta=\left(\beta_1,\beta_2,\rho\right)$, obtain $\alpha_1=\alpha_2=2,\alpha_3=1,\delta_1=\delta_2=\varepsilon,\delta_3=\varepsilon^2$, and

$$I_H(\theta)=\begin{bmatrix}\psi\left(\beta_1\right)-\psi\left(\beta_1+\beta_2\right) & -\psi\left(\beta_1+\beta_2\right) & 0\\ -\psi\left(\beta_1+\beta_2\right) & \psi\left(\beta_2\right)-\psi\left(\beta_1+\beta_2\right) & 0\\ 0 & 0 & 4\rho^{-2}\end{bmatrix}$$

Therefore, $\pi_H\left(\beta_1,\beta_2,\rho\right)\propto\rho^{-1}\sqrt{\psi\left(\beta_1\right)\psi\left(\beta_2\right)-\left[\psi\left(\beta_1\right)+\psi\left(\beta_2\right)\right]\psi\left(\beta_1+\beta_2\right)}$.

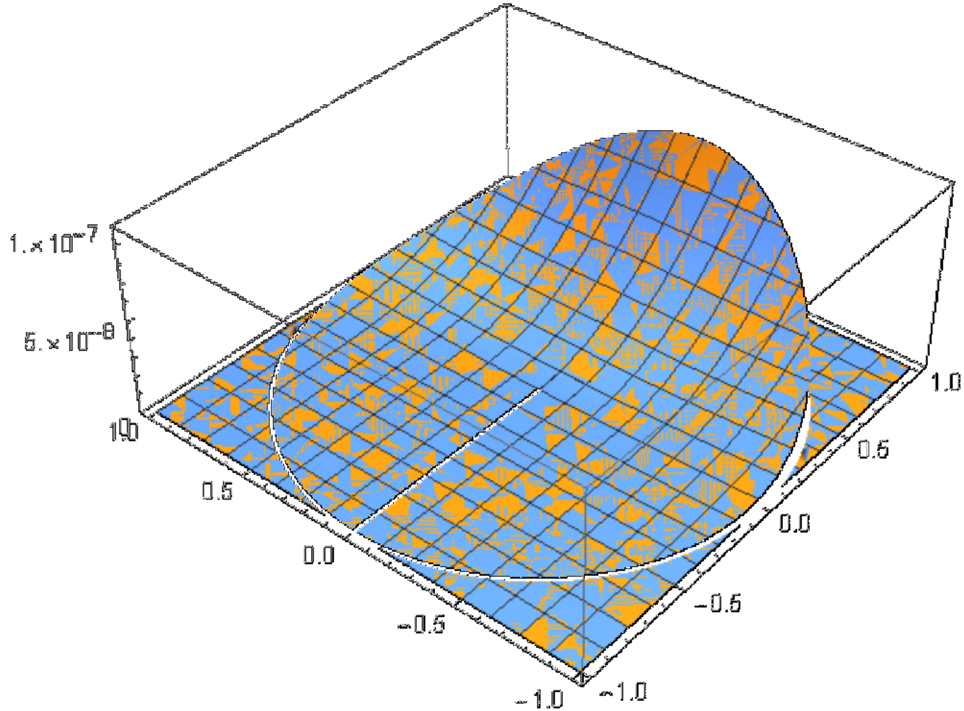

Figure 4. Plots of $p(r,\varphi;12,12,1)$, $p(r,\varphi;12,12,1+0.1)$.

Example 6: Rectangular uniform distribution (see Theorem 2) –

$$Unif\left(0,\theta_1\right)\times\left(0,\theta_2\right);\ \theta_1\in\left(0,\infty\right),\theta_2\in\left(0,\infty\right).$$

A. Using notation $\theta=\left(\theta_1,\theta_2\right)$, obtain $\alpha_1=\alpha_2=1, \delta_1=\delta_2=\varepsilon^2$, and

$$H\left(\theta,\mathbf{\Delta}\right)=\begin{bmatrix}\dfrac{\delta}{\theta_1} & \dfrac{\delta^2}{\theta_1\theta_2}\\ \dfrac{\delta^2}{\theta_1\theta_2} & \dfrac{\delta}{\theta_2}\end{bmatrix}, J(\theta)=\begin{bmatrix}\dfrac{1}{\theta_1^2} & 0\\ 0 & \dfrac{1}{\theta_2^2}\end{bmatrix}>0$$

Therefore, $\pi_H\left(\mu,\tau\right)\propto\dfrac{1}{\theta_1\theta_2}$ (also, the reference prior).

Example 7: Shifted uniform distribution -

$$Unif\left(\theta_1, \theta_1 + \theta_2\right), \theta_1 \in \left(-\infty, \infty\right), \theta_2 \in \left(0, \infty\right).$$

A. Using notation $\theta = \left(\theta_1, \theta_2\right)$, obtain $\alpha_1 = \alpha_2 = 1, \delta = \delta_1 = \delta_2 = \varepsilon^2$, and

$$H\left(\theta, \Delta\right) = \begin{bmatrix} \frac{2\delta}{\theta_2} & \frac{\delta}{\theta_2} \\ \frac{\delta}{\theta_2} & \frac{\delta}{\theta_2} \end{bmatrix}, J(\theta) = \begin{bmatrix} \frac{4}{\theta_2^2} & \frac{1}{\theta_2^2} \\ \frac{1}{\theta_2^2} & \frac{1}{\theta_2^2} \end{bmatrix} > 0$$

Therefore, $\pi_H\left(\mu, \tau\right) \propto \frac{1}{\theta_2^2}$ (also, the reference prior).

9. References

[1] J. O. Berger, and J. M. Bernardo, “On the development of reference priors” (with discussion), In: *Bayesian Statistics*, Oxford University Press (1992), **4**, 35-60.

[2] J. O. Berger, J. M. Bernardo, and D. Sun, “The formal definition of reference priors”, *Ann. Statist.* (2009), **37**, 1, 905-938.

[3] B. Clarke and A. Barron, “Jeffreys’ prior is asymptotically least favorable under entropy risk”, *J. Statist. Plann. Inference* (1994), **41**, 37-60.

[4] S. Ghosal, J. Ghosh, and T. Samanta, “On convergence of posterior distributions”, *Ann. Statist.* (1995), 23, 6, pp. 2145-2152. [5] S. Ghosal and T. Samanta, “Expansion of Bayes risk for entropy loss and reference prior in nonregular cases”, *Statist. Decisions* (1997), **15**, 129-140, MR1475184.

[6] S. Ghosal, “Probability matching priors for non-regular cases”, *Biometrika* (1999), **86**, 956-964.

[7]] I.A. Ibragimov, R.Z. Has'minskii, (1976) "Statistical estimation: asymptotic theory", Springer Verlag, New York (in English 1981).

[7] H. Jeffreys (1946), "An invariant form for the prior probability in estimation problems". *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences.* **186** (1007): *453–461.*

[8] T. M. Le, “Formal Reference Priors”, O-Bayes-2013, Duke University.

[9] Y. Lin, R. Martin, and M. Yang, “On Optimal Designs for Nonregular Models”, *Ann. Statist.* (2019), 47, 6, 3335-3359.

[10] A. Shemyakin, “Rao-Cramer type multidimensional integral inequalities for parametric families with singularities”, *Sib. Math. J.* (1991), 32, 4, *Translated from Russian*, Plenum, 1992, 706-715.

[11] A. Shemyakin, “On Information Inequalities in the Parametric Estimation”, *Theory Probab. Appl*. (1992), **37**, 1, pp.89-91.

[12] A. Shemyakin, “A New Approach to Construction of Objective Priors: Hellinger Information”, *Applied Econometrics* (In Russian), (2012), **28**, 2, 124-137.

[13] A. Shemyakin, “Hellinger Distance and Non-informative Priors”, *Bayesian Analysis*, (2014), **9**, 4, 923-938.

[14] D. Sun, “A note on non-informative priors for Weibull distribution”, *J. of Statist. Plann. Inf.,* (1997), **61**, 2, 319-338.

[15] D. Sun and J. O. Berger, “Reference priors with partial information”, *Biometrika* (1998), **85**, 55-71.