International Journal of Mathematical Analysis and Modelling

(Formerly Journal of the Nigerian Society for Mathematical Biology)

Volume 5, Issue 4 (Nov.), 2022

ISSN (Print): 2682 - 5694
ISSN (Online): 2682 - 5708

International Journal of Mathematical Analysis and Modelling
Volume 5, Issue 4, November 2022, Pages 89 – 94

Construction of prior probability distribution using a mollified functional

S.J. Yaga*†, N.P. Dibal*, I.E. Gongsin*, H.R. Bakari*, S.C. Nwaosu‡, and Zara Wudiri§

Abstract

In this paper, we proposed a new class of prior distribution, the Mollified Functional Data Informative Prior using the convolution of standard scaled mollified functional and sampling distribution of the data. We also show that the MFDIP possesses the invariance property to reparameterization and apply the MFDIP to derive the prior for exponential probability distribution.

Keywords: mollifier; mfdip; invariance property; prior probability distribution; parametric functions

1 Introduction

Bayes introduced the notion of inverse probability in doctrine of chances and Laplace did further extensive work. The basic idea in both of their work is that, when there is no prior specified in advance, the uniform prior is use as the prior distribution, but uniform prior has limitations, one of them, is that it does not possess the property of invariance to reparameterization. Kass and Wasserman [1] discussed formal rules of constructing prior distribution, and reference priors [2].

Non-informative prior is formulated in [3], [4]. Structural rules for determining prior distribution as a standard of reference is considered in [5], these rules, use the case when the parameter space is compact. Several modifications were done to this scheme and later, justified the method by looking at the invariance property of the prior under reparameterization. A frequentist approach for constructing prior was discussed in [6] as second order probability matching priors, probability matching equation for parametric function [7], is based on a differential equation that a prior must satisfy if the posterior density is one-sided and the frequentist equalize up to $O(n^{-1})$. Analytic results for Maximal Data Informative Prior (MDIP) densities are described [8], [9], [10], [11], formally adopting calculus of variation. The Summary of past and recent results relating MDIP densities and extensions are found in [12], [13] and an application of MDIP on two-parameter Gamma distribution [14]. [15] Estimate quadratic function of multivariate normal mean and identified the drawbacks of some alternative Bayesian Non-informative priors. The aim of this paper

*Department of Mathematical Sciences, University of Maiduguri, Maiduguri, Nigeria
† Correspondence Author. E-mail: Samailajyaga@unimaid.edu.ng
‡ Department of Statistics, University of Agriculture, Makurdi, Nigeria
§ College of Medical Sciences, Department of Community Medicine, University of Maiduguri, Maiduguri, Nigeria

is to propose a new class prior called mollified functional prior for one-dimensional parametric functions.

2 Mollifier

A mollifier is a family of functions ϕ_d, or non-negative function satisfying the following conditions
$\phi_d(\theta) \geq 0 \; for \; all \; \theta \in \mathbb{R}^n$, for all $d > 0$
$\phi \in C_c^\infty(\mathbb{R}^n)$
$\|\phi\|_1 = 1$ or $\int \phi_d(\theta) d^n \theta = 1$.

A standard mollifier, is given by the function

$$\phi(\theta) = K_v \psi(1 - \|\theta\|^2) = \begin{cases} 0, & for \; \|\theta\| \geq 1 \\ K_\sigma \exp\left(-\frac{1}{1 - \|\theta\|^2}\right), & for \; \|\theta\| < 1 \end{cases} \tag{1}$$

K_v is define such that the integral is normalized. The function can be generalized to multi-dimensional case but we will stick to one-dimension. The function $\phi(\theta)$ is smooth, when we use substitution and differentiate the function. We will scale the mollifier using the function

$$\phi_d(\theta) = \frac{1}{d^v} \phi\left(\frac{\theta}{d}\right). \tag{2}$$

The function in (2) is such a case where the mass is preserved $\|\phi\|_1 = 1$, and concentrate more around the origin like Dirac delta function, where the density is zero everywhere except at point zero, then is an approximate identity. We normalized the mollifier over the region $\mathbb{R}^v$ to get

$$\int \phi_d(\theta) d\theta = \int \frac{1}{d^v} \phi\left(\frac{\theta}{d}\right) d\theta = \int \phi(y) dy = 1. \tag{3}$$

3 Convolution of function by mollifiers

Given a smooth function f, the convolution of f by the mollifier $\phi_d(\theta)$ is

$$f_d(\theta) = (f * \phi_d)(\theta) = \int f(\theta - u)\phi_d(u) du = \int f(u)\phi_d(\theta - u) du, \tag{4}$$

$f_d(\theta)$, is a weighted smoothed average over all the values of, f in the ball $B_d(\theta)$ of radius d and center θ. We illustrate this using the function

$$\psi_d(\theta) = a\big(B_1(0)\big)^{-1} \chi_{B_1(0)}(\theta), \tag{5}$$

ψ_d , satisfies similar properties as ϕ_d except it is not smooth. We define

$$\psi_d(\theta) = \frac{1}{d^v} \psi\left(\frac{\theta}{d}\right). \tag{6}$$

The convolution

$$(f * \psi_d)(\theta) = \int f(u)\psi_d(\theta - u)du = a\big(\mathrm{B}_1(0)\big)^{-1} \int f(u)du, \tag{7}$$

is the average value of f over the ball $\mathrm{B}_d(\theta)$ of radius d and center θ, since $\phi_d \geq 0$ and has integral one over $\mathrm{B}_d(\theta)$. We can consider $(f * \psi_d)(\theta)$ to be weighted average of f over $\mathrm{B}_d(\theta)$.

4 New class of mollified functional priors

In constructing this new class of prior density, we consider a prior that will have the property of invariance and reflect the frequentist density function. First, we will use the standard mollifier

$$\phi(\theta) = K_v \psi(1 - \|\theta\|^2) = K_\sigma \exp\left(-\frac{1}{1 - \|\theta\|^2}\right), for\ \ \|\theta\| < 1, \tag{8}$$

and we scale (8) with (2) to get

$$\phi_d(\theta) = \frac{1}{d^v} k_\sigma \psi\left(1 - \left\|\left(\frac{\theta}{d}\right)^2\right\|\right), for\ \ \|\theta\| < 1. \tag{9}$$

Our prior is a function of the Fisher information since is sensitive to parameter changes.

We define $d = I(\theta)$ and equation (9) becomes

$$\phi_d(\theta) = \frac{1}{|I(\theta)|^v} k_\sigma \psi\left(1 - \left\|\left\{\frac{\theta}{\mathrm{I}(\theta)}\right\}^2\right\|\right), for\ \ \|\theta\| < 1, v = 1,2,3, \dots, n \tag{10}$$

where v , is the dimension for the parameter of the distribution and for one-dimensional parameter $v = 1$ and when two $v = 2$ etc. hence, the function

$$\psi\left(1 - \left\|\left(\frac{\theta}{|\mathrm{I}(\theta)|}\right)^2\right\|\right) = \exp\left\{-\frac{\mathrm{I}(\theta)}{[\mathrm{I}(\theta)]^2 - \|\theta\|^2}\right\}. \tag{11}$$

Substituting in (10) we have

$$\phi_d(\theta) = \frac{1}{|I(\theta)|^v} k_\sigma \exp\left\{-\frac{\mathrm{I}(\theta)}{[\mathrm{I}(\theta)]^2 - \|\theta\|^2}\right\}. \tag{12}$$

Theorem 1
A density function ϕ_d, (mollifier) with zero mass everywhere except at point zero, and given a smooth function $f,$ without changing the properties of the function, the convolution is the average of the mollifier and the density function.

Proof

Using equation (7) and $\psi_d(\theta) = a\big(B_1(0)\big)^{-1}\chi_{B_1(0)}(\theta)$, where $\chi_{B_1(0)}(\theta)$ is the characteristic function, the convolution

$$(f * \psi_d)(\theta) = \int f(u)\psi_d(\theta - u)du = a\big(\mathrm{B}_1(0)\big)^{-1}\int f(u)du. \tag{13}$$

Using the ball with radius 1 at the center 0, we get the relation $a\big(\mathrm{B}_1(0)\big)^{-1}\int f(u)du =1$, over a region $\mathbb{R}$. The normalization is because of area of the ball $\psi_d(\theta)$.

Theorem 2

Suppose we have a data model that follows an exponential density function $\psi(x|\theta)$. Under this condition, the prior distribution Mollified Functional Data Informative (MFDIP) Prior denoted by $\phi_d(\theta)$ for the parameter θ, is given by

$$\phi_d = \frac{\theta^2}{n}k_\sigma \exp\left(-\frac{n}{n^2-\theta^4}\right), \qquad \theta^4 < n^2. \tag{14}$$

Proof

For the proof, we will make use of equation (12) Firstly, we have to evaluate the Fishers Information $\mathrm{I}(\theta)$ for the exponential density, which is obtained after some algebra

$$\frac{d^2}{d\theta^2}\ln L[\psi(x|\theta)] = -\frac{n}{\theta^2}. \tag{15}$$

The Fisher information after substitution becomes

$$\mathrm{I}(\theta) = -E\left(\frac{d^2}{d\theta^2}\ln L[\psi(x|\theta)]\right) = \frac{n}{\theta^2}. \tag{16}$$

Substituting (16) in (12) the MFDIP for the parameter θ is given by

$$\phi_d = \frac{\theta^2}{n}\left[k_\sigma exp\left\{-\frac{n}{n^2-\theta^4}\right\}\right]. \tag{17}$$

5 Invariance property of MFDIP

Proposition

Consider the Mollified Functional Data Informative Prior (MFDIP) ϕ_d with parameter θ. Let the function $\eta = h(\theta)$ be a smooth reparameterization, the prior density ϕ_d is invariant to change of scale defined as

$$\phi_d(\eta) = \frac{1}{|I(\eta)|^v}\left|\frac{\partial\theta(\eta)}{\partial\eta}\right| k_\sigma \exp\left\{-\frac{\mathrm{I}(\eta)}{[\mathrm{I}(\eta)]^2 - (h^{-1}(\eta))^2}\right\}. \tag{18}$$

Proof

Let $\eta = h(\theta)$ be the smooth reparameterization. We determine the inverse of function for θ and the Jacobian of the transformation, we obtain $\theta = h^{-1}(\eta)$, and the Jacobian is
$\mathcal{J}_{(\theta)} = \left|\frac{\partial\theta(\eta)}{\partial\eta}\right|$.

Since we are able to obtain the Jacobian, the new density function, is derived using the relation

$$\phi_d(\eta) = \left|\frac{\partial\theta(\eta)}{\partial\eta}\right| \phi_d\big((h^{-1}(\eta)\big),$$

to get the density

$$\phi_d(\eta) = \frac{1}{|I(\eta)|^v}\left|\frac{\partial\theta(\eta)}{\partial\eta}\right| k_\sigma \exp\left\{-\frac{\mathrm{I}(\eta)}{[\mathrm{I}(\eta)]^2 - (h^{-1}(\eta))^2}\right\}. \tag{19}$$

6 Summary and discussion

Mollifiers are functions that behave like the Dirac delta function, with zero mass everywhere except at the origin. In this study, we constructed a new class of prior for a v-dimensional parameter density function for the frequentist data model. We constructed the prior using the convolution of the scaled mollifier by preserving normalization to one of the area of the mollifier with the sampling distribution of the data. The new class of prior has the invariance property to smooth reparameterization. We obtain the prior using MFDIP for exponential density. The prior is a function of the Fisher Information.

References

[1] Kass, R and L. Wasserman (1996). The selection of prior distribution by formal rules. Journal of the American Statistical Association 431, 1343-1370
[2] Bernardo, J.M. (1979). Reference posterior distribution for Bayesian inference. Journal of Royal statistical Society Ser B, 41, 113-147
[3] Box, G.E.P, Tiao, G. (1973). Bayesian inference in statistical analysis, John Wiley and Sons.
[4] Bernardo, J.M., Smith, A.F.M (1994). Bayesian theory. New York; John Wiley and Sons, Inc.
[5] Jeffrey, H. (1946). An invariant form of prior probability in estimation problems. Proc. Royal Statistical Society. Ser. A, 186, 453-461.
[6] Mukerjee, R. andGhosh, M. (1997). Second Order probability matching priors. Biometrika, 84, 4:970-975.
[7] Datta, G.S, and Ghosh, J.K. (1994). Probability matching equations for a parametric function. Purdue University, Tech Report, 94-23C.
[8] Good, I.J. (1956). Some terminology and notation in Information theory. Proc of the Institute of Electrical Engineers, C, 103, 200-204.
[9] Lindley, D.V. (1956). On a measure of information provided by an experiment. Annals of Mathematical Statistics, 27, 986-1005.
[10] Shannon, C.E. (1948). A Mathematical theory of communication. Bell System Technical

Journal, 27, 623-659.

[11] Zellner, A. (1977). Maximal data informative prior distribution in new developments in the application of Bayesian methods. Eds. A. Aykac and C. Brumal, Amsterdam North-Holland, 201-215

[12] Zellner, A. (1984). Maximal data informative prior distributions, basic issues in econometrics, The University of Chicago Press, USA.

[13] Tibshirani, R. (1989). Non-informative priors for one parameter of many. Biometrika 3, 604-608

[14] Moala, F. A, Ramos, P.L, Achcar, J.A. (2013). Bayesian inference for two-parameter gamma distribution assuming different non-informative priors. Revista Colombiana de Estadistica, 36, 321-338.

[15] Berger, J.O, Philippe, A., Robert, C.P. (1998). Estimation of quadratic functions: non-informative priors for non-centrality parameters. Statistica Sinica, 8, 358-375.