Journal of Statistical Modelling: Theory and Applications
Vol. 3, No. 2, 2022, pp. 145-155
Yazd University Press 2022

JSMTA

Research Paper

Conditions for interior based constrained prior distributions to ensure probability density

Amir Hossein Ghatari[1], Elham Tabrizi*[2]
[1]Department of Statistics, Amirkabir University of Technology, Tehran, Iran.
[2]Department of Mathematics, Kharazmi University, Tehran, Iran.

Received: August 05, 2023/ Revised: September 10, 2023/ Accepted: November 16, 2023

Abstract: In Bayesian inference, the acquisition of prior distributions plays a fundamental role. While authorized priors need not conform to traditional probability densities and may be improper priors, obtaining proper prior densities remains a challenge in the Bayesian literature. This article explores a set of conditions that enable the establishment of specific assumptions, ensuring that maximum entropy priors and restricted reference priors become proper and transform into probability density priors. By examining these conditions, this study contributes to the advancement of proper prior acquisition in Bayesian analysis.

Keywords: Constrained prior; Jensen inequality; Maximum entropy prior; Restricted reference priors.
Mathematics Subject Classification (2010): 62F15.

1 Introduction

In Bayesian literature, inference is made by mixing prior information about model parameters and available data which is called posterior probability. The posterior probability is a conditional probability distribution obtained by applying the distributional form of Bayes theorem. Indeed, given prior distribution $\pi(\theta)$, posterior distribution obtained as:

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{p(x)},$$

where $f(x \mid \theta)$ is likelihood function and $p(x) = \int f(x \mid \theta)\pi(\theta)d\theta$. Accordingly, the posterior distribution changes by changing $\pi(\theta)$. Gelman (2006) proposed two key

*Corresponding author: elham.tabrizi@khu.ac.ir

issues about setting up a prior distribution as:
• what information is going into the prior distribution.
• the properties of the resulting posterior distribution.

Priors can be classified in several ways such as: 1- the information about the model parameters that it divides them into informative, weakly informative and uninformative priors. For example, Jeffreys prior, proposed by Jeffreys (1946) is an uninformative prior. 2- the proper or improper priors. Indeed, improper priors are not actual probability distribution but make posterior probability distribution. 3- conjugate or non-conjugate priors. For example, beta distribution can be considered as conjugate priors for data with Bernoulli, Binomial and Geometry distributions. Let us note that a prior distribution can be in several class of the mentioned classes. We can define another class of priors with an interior prior π_0 in their structure. Maximum entropy prior is one of these kind of priors. We call this class as interior based priors. The main focus of this paper is around interior based priors and their conditions to be proper. For more information about prior distributions and their classification, see Kass (2005), Gelman (2006), Bernardo and Smith (2009) and their references.

Prior distribution makes a wide-range concept. Hence, the study of their mathematical properties is an important subject in the concept of Bayesian statistics. Case and Keats (1982) examined the relationship between defectives in the sample and defectives in the rest of the lot for each of five prior distributions in the concept of Bayesian acceptance sampling. Chen et al. (2020) studied several theoretical properties of Jeffreys prior for binomial regression models and showed that Jeffreys prior is symmetric and unimodal for a class of binomial regression models. Rojas et al. (2009) presented an assessment of prior knowledge and a sensitivity analysis of the prior in groundwater modeling, emphasizing the importance of selecting proper prior probabilities. Tang et al. (2016) described tools for the evaluation of parameter sensitivity to the prior distribution to provide guidelines for defining meaningful priors using Kullback-Leibler Divergence and prior information elasticity. Gelman et al. (2017) precisely subjected the challenge in choosing prior distributions in Bayesian analysis, mentioning various common types of priors and their conceptual tensions. Their paper offers a resolution by considering the choice of priors within the broader context of Bayesian analysis, encompassing inference, prediction, and model evaluation. Banner et al. (2020) provided the underutilization of Bayesian data analysis in ecology due to insufficient attention to prior specification. Their work showed the importance of choosing priors. Kosmidis and Firth (2021) studied the properties of Jeffreys prior when it was used in the concept of generalized linear models.

Our findings in the paper contribute to the broader field of Bayesian inference, specifically focusing on the challenges related to prior distributions and the critical role they play in Bayesian analyses. Additionally, we wrote this paper as the readers even with little statistical knowledge can better comprehend our research.

This paper employs Jensen's inequality in the context of statistical literature to introduce two general inequalities for integrating prior distributions. These inequalities are then utilized to establish the probability density conditions for maximum entropy and restricted reference priors (continuous versions of them). These priors are particularly useful when dealing with scenarios where there is limited information available about the model parameters. Obtaining a probability density as a prior under such circumstances is advantageous.

Section 2 of the paper provides a comprehensive overview of fundamental concepts, including the definition of entropy. In Section 3, a generalized form of inequality encompassing improper integrals is proven, leading to the derivation of the probability density conditions for maximum entropy and restricted reference priors.

2 Basic concepts

In this section, we recall basic definitions which are required to obtain the main results of the paper. Maximum entropy prior, restricted reference prior and Jensen inequality are the concept that will be reviewed in continue.

2.1 Maximum entropy prior

In most of the cases, there is a low information about the unknown parameter. On the other hand, it is better to use non-informative priors since by using them the Bayesian inference has at least equal accuracy versus classical statistical inference. For example, may be we know mean and variance of prior distribution. In this case, we are looking for the most non-informative density between all of prior densities with the given constrains. When we have a continuous parameter space, a useful approach is using entropy concept.

Definition 2.1. *When $\theta \in \Theta$ and Θ is continuous, the entropy of prior density (π) is defined as follows*

$$En(\pi) = -\int_{\Theta} \pi(\theta) \log(\frac{\pi(\theta)}{\pi_0(\theta)}) d\theta, \tag{1}$$

where $\pi_0(\theta)$ is an invariant non-informative prior for θ.

If our knowledge about unknown variable parameter θ be functions as $g_k(\theta)$, $k = 1, .., m$ with these constrains:

$$E(g_k(\theta)) = \int_{\Theta} \pi(\theta) g_k(\theta) d\theta = \mu_k \ , \ k = 1, ..., m.$$

Maximum Entropy prior is density that maximize (1) and it can calculate as follows

$$\bar{\pi}(\theta) = \frac{\pi_0(\theta) \exp(\sum_{k=1}^{m} \lambda_k g_k(\theta))}{\int_{\Theta} \pi_0(\theta) \exp(\sum_{k=1}^{m} \lambda_k g_k(\theta))}, \tag{2}$$

where λ_k is constant that calculated based on knowledges $g_k(\theta)$. For more information about maximum entropy priors see Berger and Berger (1985), Cover (1999), and Bernardo and Smith (2009). As we can see $\bar{\pi}(\theta)$ in (2) is depended to $\pi_0(\theta)$ and we call it as interior prior for $\bar{\pi}(\theta)$.

2.2 Restricted reference prior

The reference prior framework can be applied to allow constraints much like the kind imposed on the maximum entropy priors. Hence, we want the prior that maximizes the mutual information between the prior and posterior while satisfying the constraints

$$E_p\left[g_k(\theta)\right] = \beta_k, \quad k = 1, \dots, m,$$

where m is the number of constraints (similar to maximum entropy). The common analytical form of restricted reference prior is

$$\pi_{RR}(\theta) = \pi_0(\theta) \exp\left(\sum_{k=1}^{m} \lambda_k g_k(\theta)\right), \tag{3}$$

where $\pi_0(\theta)$ is the prior for the unconstrained form of problem (simultaneously interior prior for $\pi_{RR}(\theta)$). This prior is known as a tilted distribution. For more information about reference prior and its various types see Bernardo (2005).

2.3 Jensen inequality

Jensen inequality relates the value of a convex function of an integral to the integral of the convex function. The inequality proposed and proved by Jensen (1906). If g is a real-valued function that is integrable on its domain, and if φ is a convex function on the real line, then Jensen inequality is defined as

$$\phi\left(\int\limits_{D_x} g(x)dx\right) \leq \int\limits_{D_x} \phi\left(g(x)\right) dx.$$

The same result can be equivalently stated in a probability theory setting Perlman (1974), let X be an integrable real-valued random variable and ϕ a convex function. Then:

$$\phi\left(E(X)\right) \leq E\left(\phi\left(X\right)\right). \tag{4}$$

Note that the sign of the inequality is changed when ϕ be a concave function. Also Jensen inequality has different types and special cases. For more information about Jensen inequality, see Rudin (1987).

3 Problem and results

In the realm of Bayesian statistics, the selection of prior distributions poses significant challenges. As discussed in Section 1, priors exhibit various properties, which include the classification into proper and improper priors. While improper priors do not conform to probability density functions, the resulting posterior distributions can still be expressed as probability densities under any circumstances. Within this section, we present a set of inequalities, including an interval, that applies to a broad range of prior distributions encompassing two specific types of constrained priors known as interior based priors. These priors consist of the restricted reference priors and the maximum entropy priors. By making certain assumptions, we establish that both types of priors become proper if their corresponding interior prior is proper. Next Lemma provides a generalized form of Jensen inequality which is used to obtain inequalities (including an interval) for prior distributions.

Lemma 3.1. *suppose that f,g : $D_x \to \mathbb{R}^+$ and concave function $\phi(x) = \log(x)$ exist.*
i. *If $\int_{D_x} f(x)dx = k$ and $1 \le k \le \infty$ then:*

$$\int_{D_x} \frac{f(x)}{k} \log(g(x))dx \le \log\left(\int_{D_x} f(x)g(x)dx\right).$$

ii. *If $\int_{D_x} f(x)dx = k$ and $0 < k < 1$ then for h=f/k:*

$$\int_{D_x} h(x)\log(g(x))dx \le \log\left(\int_{D_x} h(x)g(x)dx\right).$$

Proof. i. If $k = \infty$ based on intermediate value theorem for integral there exists ξ_x so that

$$\log(\int_{D_x} f(x)g(x)dx) = \log(g(\xi_x)\int_{D_x} f(x)dx) = k\infty = \infty.$$

So we have

$$\int_{D_x} \frac{f(x)}{k} \log(g(x))dx \le \infty = \log(\int_{D_x} f(x)g(x)dx).$$

If $1 \le k < \infty$, we have

$$\int_{D_x} f(x)\log(g(x))dx = k\int_{D_x} (f(x)/k)\log(g(x))dx = kE(\log(g(X))).$$

Since $\frac{f(x)}{k}$ is a density function, we have

$$\int_{D_x} \frac{f(x)}{k} \log(g(x))dx = E(\log(g(x))),$$

according to (4) for concave functions, we have

$$\begin{aligned} E(\log(g(x))) &\le \log(E(g(X))) \\ &= \log(\int_{D_x} (f(x)/k)g(x)dx) \le \log(\int_{D_x} f(x)g(x)dx). \end{aligned}$$

ii. if $\int_{D_x} f(x)dx = k$ then $\int_{D_x} h(x)dx = 1$ so we have

$$\int_{D_x} h(x)\log(g(x))dx \le E(\log(g(X))) \le \log(E(g(X))) = \log(\int_{D_x} h(x)g(x)dx).$$

□

Similar to the Lemma 3.1, if we put the concave function $\phi(x) = \exp(x)$, we reach to
i. If $\int\limits_{D_x} f(x)dx = k$ and $1 \leq k \leq \infty$ then

$$\int\limits_{D_x} \frac{f(x)}{k} \exp(g(x))dx \geq \exp\left(\int\limits_{D_x} f(x)g(x)dx\right).$$

ii. If $\int\limits_{D_x} f(x)dx = k$ and $0 < k < 1$ then for $h = f/k$

$$\int\limits_{D_x} h(x) \exp(g(x))dx \geq \exp\left(\int\limits_{D_x} h(x)g(x)dx\right).$$

Using Lemma 3.1, the following theorem provides preliminaries to reach the conditions of obtaining making probability density by maximum entropy and restricted reference priors.

Theorem 3.2. *Suppose that $\theta \in \Theta$ and Θ is continuous and $\pi(\theta)$ be a interior based prior for θ. If $\frac{\pi(\theta)}{\pi_0(\theta)} \geq 1$, then*

$$\int\limits_{\Theta} \pi(\theta)d\theta \geq \frac{1}{M\sqrt{\int\limits_{\Theta} \exp\left|\log(\pi_0(\theta))\right| d\theta}},$$

where $M = \sup\limits_{\theta\in\Theta} \pi(\theta)$, and $\pi_0(\theta)$ is same as an interior prior for $\pi(\theta)$.

Proof. i. Based on the assumptions of the theorem, $\pi_0(\theta)$ and $\pi(\theta)$ are positive. According to (2.1), the entropy of $\pi(\theta)$ is defined as (1) and we can write

$$0 \leq |En(\pi)| = \left|-\int\limits_{\Theta} \pi(\theta)\log(\frac{\pi(\theta)}{\pi_0(\theta)})d\theta\right| = \left|\int\limits_{\Theta} \pi(\theta)\log(\frac{\pi(\theta)}{\pi_0(\theta)})d\theta\right|.$$

Hence, we have

$$\begin{aligned} 0 \leq |En(\pi)| &= \int\limits_{\Theta} \pi(\theta)ln(\frac{\pi(\theta)}{\pi_0(\theta)})d\theta \\ &= \int\limits_{\Theta} \pi(\theta)\log(\pi(\theta))d\theta - \int\limits_{\Theta} \pi(\theta)\log(\pi_0(\theta))d\theta. \end{aligned}$$

On the other hand, according to Lemma 3.1, we have:

$$\int\limits_{\Theta} \pi(\theta)\log(\pi_0(\theta))d\theta \leq \int\limits_{\Theta} \pi(\theta)\log(\pi(\theta))d\theta$$

$$\leq \log \int_{\Theta} k\pi(\theta)\pi(\theta)d\theta \leq \log kM \int_{\Theta} \pi(\theta)d\theta \leq \log(Mk^2), \quad (5)$$

where $k = \int_{\Theta} \pi(\theta)d\theta$. We can obtain from (5) and using the properties of absolute and logarithmic functions

$$\exp\{\int_{\Theta} \pi(\theta) \left|\log(\pi_0(\theta))\right| d\theta\} \geq \frac{1}{Mk^2}.$$

Again according to Lemma 3.1 for convex function $f(x) = e^x$ we have

$$\int_{\Theta} \pi(\theta) \exp\left(\left|\log(\pi_0(\theta))\right|\right) d\theta \geq \frac{1}{Mk^2}.$$

Hence, using the property of M, we can conclude that

$$M \int_{\Theta} \exp\left|\log(\pi_0(\theta))\right| d\theta \geq \frac{1}{Mk^2}.$$

Therefore, we have

$$k = \int_{\Theta} \pi(\theta)d\theta \geq \frac{1}{M\int_{\Theta} \exp\left|\log(\pi_0(\theta))\right| d\theta}.$$

□

Before presentation of Corollary for Theorem 3.2, we need to consider a set like Π_1 with following definition

$$\Pi_1 = \{\pi(\theta) \mid \pi(\theta) \geq 1 \ , \theta \in \Theta\}.$$

It contains all possible priors which their range is greater than 1. In continue, we obtain an interval for all $\pi(\theta)$ satisfying the terms of Theorem 3.2. Then, we show that the restricted reference priors satisfy the terms of Theorem 3.2. In addition, we prove that maximum entropy priors hold the both terms of Theorem 3.2 under different assumptions.

Corollary 3.3. *if* $\dfrac{\pi(\theta)}{\pi_0(\theta)} \geq 1$ *and* $\pi_0(\theta) \in \Pi_1$*, then*

$$\frac{1}{M^2 \int_{\Theta} \pi_0(\theta)d\theta} \leq \int_{\Theta} \pi(\theta)d\theta \leq \frac{1}{I_1} \log \int_{\Theta} \pi_0(\theta)d\theta,$$

where M *is defined in Theorem 3.2 and* $I_1 = \inf_{\theta\in\Theta} \log \pi_0(\theta)$

Proof. Form Theorem 3.2, we have

$$\int_{\Theta} \pi(\theta)d\theta \geq \frac{1}{M^2 \int_{\Theta} \exp\left(|\log(\pi_0(\theta))|\right) d\theta},$$

and because $\pi_0(\theta) \in \Pi_1$ it will be obtained that $|\log(\pi_0(\theta))| = \log(\pi_0(\theta))$. Hence, we can write

$$\int_{\Theta} \pi(\theta)d\theta \geq \frac{1}{M^2 \int_{\Theta} \exp\left(|\log(\pi_0(\theta))|\right) d\theta} = \frac{1}{M^2 \int_{\Theta} \pi_0(\theta)d\theta}. \tag{6}$$

On the other hand, since $\pi_0(\theta) \geq 1$ then $\pi(\theta) \geq 1$. Therefore, we can immediately conclude that

$$\pi(\theta)\log(\pi_0(\theta)) \leq \log(\pi_0(\theta)).$$

Hence, we have

$$I_1 \int_{\Theta} \pi(\theta)d\theta \leq \int_{\Theta} \pi(\theta)\log(\pi_0(\theta))d\theta \leq \log \int_{\Theta} \pi_0(\theta)d\theta, \tag{7}$$

where we know $I_1 \geq 0$. Finally, according to (6) and (7), we reach to

$$\frac{1}{M^2 \int_{\Theta} \pi_0(\theta)d\theta} \leq \int_{\Theta} \pi(\theta)d\theta \leq \frac{1}{I_1} \log \int_{\Theta} \pi_0(\theta)d\theta.$$

□

It can be said that Corollary 3.3 shows that when $\pi_0(\theta) \in \Pi_1$, any unknown prior $\pi(\theta)$ is bounded just if it holds the assumptions. Note that if the lower bound of interval in Corollary 3.3 be infinity then without any calculation we can conclude that all of the interior based priors obtained by $\pi_0(\theta)$ are improper priors. For example, maximum entropy ($\bar{\pi}(\theta)$) and reference (restricted or not) priors are improper if they satisfy the conditions of Theorem 3.2 or Corollary 3.3.

Consider restricted reference prior (3). It is clear that

$$\frac{\pi_{RR}(\theta)}{\pi_0(\theta)} = \exp\left(\sum_{k=1}^{m} \lambda_k g_k(\theta)\right) \geq 1. \tag{8}$$

Hence, we can conclude that $\pi_{RR}(\theta)$ holds the condition of the first term in Theorem 3.2 in any case. Also, if $\pi_0(\theta) \in \Pi_1$, then we have the results of Corollary 3.3 for $\pi_{RR}(\theta)$. In continue, we check the required conditions for $\bar{\pi}(\theta)$ to satisfy the assumptions of Theorem 3.2. According to (2), we have

$$\bar{\pi}(\theta) = \frac{\pi_0(\theta)\exp(\sum_{k=1}^{m} \lambda_k g_k(\theta))}{\int_{\Theta} \pi_0(\theta)\exp(\sum_{k=1}^{m} \lambda_k g_k(\theta))}. \tag{9}$$

The denominator of $\bar{\pi}(\theta)$ is a constant value and we call it C. Hence, if $C \leq 1$ then we have:

$$\frac{\bar{\pi}(\theta)}{\pi_0(\theta)} = \frac{1}{C}(\theta) \exp\left(\sum_{k=1}^{m} \lambda_k g_k(\theta)\right) \geq 1.$$

Therefore, the terms of Theorem 3.2 holds. If $C \geq 1$, there is no certain answer but, it depends on the numerator of $\bar{\pi}(\theta)$. Note that we have the results of Corollary 3.3 for $\bar{\pi}(\theta)$ when $C \leq 1$.

In the concept of prior distributions, sometimes the obtained prior is improper and actually cannot considered as probability density. The following theorem propose situation about interior based priors which says that they are proper if we choose a proper interior prior ($\pi_0(\theta)$ be proper).

Theorem 3.4. *Suppose that $\pi_0(\theta)$ is proper,*
i. if $\frac{\pi(\theta)}{\pi_0(\theta)} \geq 1$ and $\pi_0(\theta) \in \Pi_1$, then $\pi(\theta)$ is proper and we can reach to a prior probability density on Θ.
ii. If $\frac{\pi(\theta)}{\pi_0(\theta)} \leq 1$, then $\pi(\theta)$ is proper and we can reach to a prior probability density.

Proof. According to the assumptions, $\int_\Theta \pi_0(\theta)d\theta = K < \infty$. Also, $\pi_0(\theta) \geq 1$. Hence, $I_1 = \inf_{\theta\in\Theta} \log \pi_0(\theta) < \infty$.
i. From Corollary 3.3 we have

$$\int_\Theta \pi(\theta)d\theta \leq \frac{1}{I_1} \log \int_\Theta \pi_0(\theta)d\theta = \frac{\log K}{I_1} = K^* < \infty,$$

therefore,

$$\exists \ C_1 \quad \ni \qquad \int_\Theta \pi(\theta)d\theta = C_1 < \infty \Longrightarrow \int_\Theta \frac{\pi(\theta)}{C_1}d\theta = 1.$$

Hence, $\pi^*(\theta) = \frac{\pi(\theta)}{C_1}$ is a probability density on Θ.
ii. Based on the assumptions and Theorem 3.2, $\log \frac{\pi_0(\theta)}{\pi(\theta)} \geq 0$. Hence, $I = \inf_{\theta\in\Theta} \log \frac{\pi_0(\theta)}{\pi(\theta)} < \infty$. Consequently, we have $\int_\Theta \pi(\theta)d\theta \leq \frac{K}{I} < \infty$ therefore,

$$\exists \ C_2 \quad \ni \quad \int_\Theta \pi(\theta)d\theta = C_2 < \infty \Longrightarrow \int_\Theta \frac{\pi(\theta)}{C_1}d\theta = 1.$$

Hence, $\pi^{**}(\theta) = \frac{\pi(\theta)}{C_1}$ is a probability density on Θ. □

Corollary 3.5. *Suppose that $\pi_0(\theta)$ is proper. If $\pi_0(\theta) \in \Pi_1$, then restricted reference prior provides a proper prior.*

Proof. From the inequality (8), we have $\frac{\pi_{RR}(\theta)}{\pi_0(\theta)} \geq 1$. Also, $\pi_0(\theta) \in \Pi_1$, then according to Theorem 3.4,

$$\exists \ C_3 \quad \ni \quad \int_\Theta \pi_{RR}(\theta)d\theta = C_3 < \infty \Longrightarrow \int_\Theta \frac{\pi_{RR}(\theta)}{C_3}d\theta = 1.$$

□

Similarly, for maximum entropy prior, if $C \leq 1$ in (9) then we can obtain C_4 such that

$$\int_{\Theta} \frac{\bar{\pi}(\theta)}{C_4} d\theta = 1.$$

Indeed, if we choose a proper interior prior, then we can build a certain density probability prior based on restricted reference prior and maximum entropy prior as interior based priors.

4 Discussion and conclusions

In Bayesian literature, posterior distributions are inherently represented as probability densities. However, the priors employed in such analyses can take on either proper or improper forms. By utilizing a prior with a probability density function for the parameters within a statistical model, we gain access to valuable insights into the parameter distribution prior to observing data and computing posterior densities. The main contribution of this paper lies in establishing the conditions which are necessary for obtaining proper priors when the parameter space is continuous and particularly in scenarios where limited information is available regarding the model parameters (such as when maximum entropy and restricted reference priors are employed). For future research endeavors, it would be worthwhile to explore the properties of priors utilized for linear model coefficients, especially within the context of non-Gaussian distributed models.

Acknowledgment

We acknowledged the editor-in-chief, the associate editor and the reviewers for their thorough observations and useful comments that improved the quality of this paper.

References

Banner, K.M., Irvine, K.M. and Rodhouse, T.J. (2020). The use of bayesian priors in ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, **11**(8):882–889.

Berger, J.O. and Berger, J.O. (1985). *Bayesian Analysis.* New York: Springer.

Bernardo, J.M. (2005). Reference analysis. *Handbook of Statistics*, **25**:17–90.

Bernardo, J.M. and Smith, A.F. (2009). *Bayesian Theory.* Vol. 405, John Wiley & Sons.

Case, K.E. and Keats, J.B. (1982). On the selection of a prior distribution in Bayesian acceptance sampling. *Journal of Quality Technology*, **14**(1):10–18.

Chen, M.-H., Ibrahim, J.G. and Kim, S. (2008). Properties and implementation of jeffreys's prior in binomial regression models. *Journal of the American Statistical Association*, **103**(484):1659–1664.

Cover, T.M. (1999). *Elements of Information Theory.* New Jersey: John Wiley & Sons.

Gelman, A. (2006). Prior distribution. In *Encyclopedia of Environmetrics*, Vol. 3, 1634–1637.

Gelman, A., Simpson, D. and Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, **19**(10):555.

Jeffreys, H. (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.

Jensen, J.L.W.V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, **30**(1):175–193.

Kass, R. (2005). Prior distribution. *Encyclopedia of Biostatistics*, **6**.

Kosmidis, I. and Firth, D. (2021). Jeffreys-prior penalty, finiteness and shrinkage in binomial-response generalized linear models. *Biometrika*, **108**(1):71–82.

Perlman, M.D. (1974). Jensen's inequality for a convex vector-valued function on an infinite-dimensional space. *Journal of Multivariate Analysis*, **4**(1):52–65.

Rojas, R., Feyen, L. and Dassargues, A. (2009). Sensitivity analysis of prior model probabilities and the value of prior knowledge in the assessment of conceptual model uncertainty in groundwater modelling. *Hydrological Processes: An International Journal*, **23**(8):1131–1146.

Rudin, W. (1987). *Real and Complex Analysis.* McGraw-HiII Book Company.

Tang, Y., Marshall, L., Sharma, A. and Smith, T. (2016). Tools for investigating the prior distribution in bayesian hydrology. *Journal of Hydrology*, **538**:551–562.