

Article

Bayesian Inference under Small Sample Sizes Using General Noninformative Priors

Jingjing He ¹, Wei Wang ¹, Min Huang ¹, Shaohua Wang ² and Xuefei Guan ^{3,*}

¹ School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China; hejingjing@buaa.edu.cn (J.H.); w.wei@buaa.edu.cn (W.W.); buaahuang@163.com (M.H.);

² China Aviation Power Plant Research Institute, Zhuzhou 412002, China; wangshaohua@buaa.edu.cn

³ Graduate School of China Academy of Engineering Physics, Beijing 100193, China

* Correspondence: xfguan@gscaep.ac.cn

Abstract: This paper proposes a Bayesian inference method for problems with small sample sizes. A general type of noninformative prior is proposed to formulate the Bayesian posterior. It is shown that this type of prior can represent a broad range of priors such as classical noninformative priors and asymptotically locally invariant priors and can be derived as the limiting states of normal-inverse-Gamma conjugate priors, allowing for analytical evaluations of Bayesian posteriors and predictors. The performance of different noninformative priors under small sample sizes is compared using the likelihood combining both fitting and prediction performances. Laplace approximation is used to evaluate the likelihood. A realistic fatigue reliability problem was used to illustrate the method. Following that, an actual aeroengine disk lifing application with two test samples is presented, and the results are compared with the existing method.

Keywords: Bayesian inference; noninformative prior; Jeffreys' prior; invariant



Citation: He, J.; Wang, W.; Huang, M.; Wang, S.; Guan, X. Bayesian Inference under Small Sample Sizes Using General Noninformative Priors. *Mathematics* **2021**, *9*, 2810. <https://doi.org/10.3390/math9212810>

Academic Editor: Francisco-José Vázquez-Polo

Received: 6 September 2021

Accepted: 3 November 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sample sizes are often quite small in many engineering fields due to time, economic, and physical constraints, for example, life testing data of high-reliability mechanical components, large and complex engineering systems, and so on. The influence of sample size on interpreting classical significance test results has been discussed in several studies [1–3]. The predictive models established based on scarce samples may highly depend on the chosen parameter estimation method [4,5]. In those cases, the probabilistic approach is usually preferred over the deterministic approach [6–11]. The Bayes rule provides a consistent and rational mathematical device to incorporate relevant information and prior knowledge for probabilistic inference.

To motivate the discussion, consider an observable random variable X with a conditional probability density function (PDF) (or simply density) of $p(x|\phi)$, where $\phi \in \Phi$ is also a random variable. The inverse problem is to make inferences about ϕ given an observed value x of X . The Bayesian approach to the solution is to use some density $p(\phi)$ over Φ to represent the prior information of ϕ . In this way, the prior knowledge of ϕ can be encoded through the Bayes rule to obtain the posterior PDF of ϕ given x ,

$$p(\phi|x) \propto p(\phi)p(x|\phi). \quad (1)$$

The forward inference, e.g., the PDF of a certain variable or the probability of an event involving ϕ , can be made on the basis of the posterior PDF. Using the method of Markov chain Monte Carlo (MCMC), samples can directly be drawn from the posterior distribution without knowing the normalizing constant $p(x) = \int p(\phi)p(x|\phi)d\phi$ in Equation (1). The Bayesian method has been successfully demonstrated in all important disciplines [12–19].

The choice of $p(\phi)$ can have a great influence on the inference result. The proper choice of priors has been extensively discussed in the probability and statistics communities, and it

can never be overlooked as it is one of the fundamental pieces of Bayesian inference [20–26]. For one thing, the formal rule of constructing a prior regardless of the data and likelihood is sought in the field of physics [9,27]. Jaynes and Bretthorst [28] argued that “problems of inference are ill-posed until we recognize three essential things. (A) The prior probabilities represent our prior *information*, and are to be determined, not by introspection but by *logical analysis* of that information. (B) Since the final conclusions depend necessarily on both the prior information and the data, it follows that, in formulating a problem, one must specify the prior information to be used just as fully as one specifies the data. (C) Our goal is that inferences are to be completely ‘objective’ in the sense that two persons with the same prior information must assign the same prior probabilities.” (p. 373). For another, the choice of a prior may highly depend on data and likelihood in practice. Gelman et al. [29] argued that “a prior can in general only be interpreted in the context of the likelihood with which it will be paired” (p. 12).

To ensure a consistent and objective inference, rules for constructing priors with minimal subjective constraints are sought. Early work on the construction of such priors is based on the “ignorance” over the parameter space using invariance techniques [27,30,31]. The fundamental reasoning is that the priors should carry the same amount of information such that a change of scale and/or a shift of location do not affect the inference results on those parameters. The ignorant prior can systematically be derived using the concept of the transformation group. Using different transformation groups, different priors can be obtained. For simplicity, such priors are loosely referred to as noninformative priors. One of the most notable priors for a scale parameter σ of a distribution is Jeffreys’ prior, i.e., $p(\sigma) \propto 1/\sigma$. Jeffreys’ prior can be obtained using the tool of the transformation group under the condition that a change of scale does not change that state of knowledge. A further extension of the noninformative priors is “reference priors” [32,33]. The theoretical framework has been discussed in many studies, including, but not limited to, [31,34]. Apart from the noninformative approach to derive the priors, there are several less objective approaches to construct the priors. Gelman [35] proposed the weakly informative priors based on the idea of conditional conjugacy for hierarchical Bayesian models. The conditionally conjugate priors provide some computational convenience as a Gibbs sampler can be used to draw samples from a posterior distribution, and some parameters for inverse-Gamma distributions in Bayesian applications are suggested. Simpson et al. [36] proposed a method to build priors. The basic idea is to penalize the complexity induced by deviating from a simpler base model using the Kullback–Leibler divergence as a metric [37].

For critical problems with small sample sizes, the choice of the prior in Bayesian inference can have a great impact on the inference results, rendering an unreliable decision-making. Despite a great deal of research, including the aforementioned, having been conducted, the choice of an optimal prior is still nontrivial to make from the practical point of view, and a systematical method to cope with such cases is rarely seen. Moreover, a quantitative measure to evaluate the performance of a prior is equally important to identify the optimal prior. This study develops a noninformative Bayesian approach to probabilistic inference under small sample sizes. A general $1/\sigma^q$ -type of noninformative prior for the location-scale family of distributions is proposed. This type of prior is further shown to be the limiting state of the commonly used normal-inverse-Gamma conjugate prior, thus allowing for an analytical evaluation of the posterior of the parameter. Given a linear model with a Gaussian error variable, the analytical form of the posterior of the model prediction can also be obtained. The informative or noninformative nature of the prior is centered more on the concept of whether the prior encodes subjective information; therefore, it is less relevant to the likelihood function (and data). The subjective information is the source of inconsistency in the inference result as different practitioners can use different subjective information in the form of assumptions, which in most cases may only be justified by oneself. However, the prior has to be built with a certain state of knowledge. This state of knowledge, or loosely called information, is not related to experience nor assumptions, but is a result of a logical reasoning process [27]. The purpose of the study is not to argue

whether these assumptions are justified or not, nor to question the suitable assumptions made during the inference. This study provides an alternative to existing methods in such cases to avoid introducing subjectiveness into the Bayesian inference. In particular, when the number of samples is small, the influence of the prior can be as important as the likelihood (and data). In such cases, subjective information encoded in the prior can lead to biased results.

The remainder of the paper is organized as follows. First, the Bayesian model with noninformative priors is developed, in particular, a general $1/\sigma^q$ -type of noninformative prior for the location-scale family of distributions is proposed for small sample problems. The noninformative priors are further shown as the limiting states of the normal-inverse-Gamma (NIG) conjugate priors. The closed-form expressions of the Bayesian posterior and predictors using the proposed noninformative priors are obtained under a Gaussian likelihood. Next, a performance measure considering both the fitting performance and predictive performance is proposed using the concept of Bayes factors. Different priors are treated as models in a Bayesian hypothesis testing context for comparisons. Following that, the method is illustrated using two examples.

2. Bayesian Linearized Models with Noninformative Priors

To motivate the discussion, consider a general linear or linearized model:

$$y_i = \mathbf{x}_i\theta + \epsilon_i, \quad (2)$$

where θ is a k -dimensional column vector and ϵ_i are independent and identical distributed random error variables. The distribution of ϵ_i determines the likelihood function or vice versa. Without loss of generality, the distribution belongs to a location-scale family of distributions. Furthermore, it is enough to use a single scale parameter to characterize the distribution of ϵ since any constant nonzero mean, no matter known or unknown, can be grouped into θ . Denote the scale parameter of ϵ as σ^2 . A Bayesian model incorporates both the prior information and the observation through Bayes' rule. Using the matrix form $\mathbf{y} = \mathbf{x}\theta + \epsilon$, the Bayesian posterior of (θ, σ^2) writes:

$$p(\theta, \sigma^2 | \mathbf{y}) \propto p(\theta, \sigma^2) p(\mathbf{y} | \theta, \sigma^2). \quad (3)$$

The common Gaussian error variable, i.e., $\epsilon_i \sim N(0, \sigma^2)$, corresponds to the following likelihood function, for n observations $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$ and n input vector $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ where $\mathbf{x}_i, i = 1, \dots, n$, is a row vector of size k .

$$\begin{aligned} p(\mathbf{y} | \theta, \sigma^2) &= N(\mathbf{x}\theta, \sigma^2 \mathbf{I}) \\ &\propto \sigma^{-n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mathbf{x}_i\theta)^2 \right]. \end{aligned} \quad (4)$$

It should be noted that the error variable does not necessarily follow a Gaussian PDF, and other types of error distributions can be used. For example, the extreme-value PDF for the error corresponds to a Weibull likelihood for the log-transformed model prediction.

2.1. A General Form of Noninformative Priors— $\propto 1/\sigma^q$

A general $\propto 1/\sigma^q$ ($q \geq 0$) form of priors is considered here. The classical Jeffreys' prior and the related asymptotically locally invariant priors are introduced first for the purpose of completeness.

Consider the PDF of a random variable x characterized by a parameter vector ϕ ; Jeffreys' noninformative prior distribution of ϕ is proportional to the square root of the determinant of the Fisher information matrix, e.g.,

$$p(\phi) \propto \sqrt{\det \mathbf{I}(\phi)}, \quad (5)$$

where $\det(\cdot)$ is the determinant operator and $\mathbf{I}(\cdot)$ is the Fisher information matrix. The key feature of it is invariance under the monotone transformation of ϕ . This feature is achieved by using the change of variables theorem. Denote the reparameterized variable or vector as ψ ; it can be shown that:

$$p(\psi) = p(\phi) \left| \det \frac{\partial \phi_i}{\partial \psi_j} \right| = \sqrt{\det \mathbf{I}(\cdot)}, \tag{6}$$

For a Gaussian likelihood with unknown parameters μ and σ^2 ,

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x - \mu)^2}{2\sigma^2}\right]. \tag{7}$$

Jeffreys' prior for the joint parameter (μ, σ^2) is:

$$p(\mu, \sigma^2) \propto \sqrt{\det \mathbf{I}(\mu, \sigma^2)}, \tag{8}$$

where $\mathbf{I}(\mu, \sigma^2)$ is:

$$\mathbf{I} = \mathbb{E} \begin{bmatrix} \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} \\ \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} \\ \frac{\partial \ln p(x|\cdot)}{\partial \mu} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} & \frac{\partial \ln p(x|\cdot)}{\partial (\sigma^2)} \end{bmatrix} \tag{9}$$

and $\mathbb{E}(\cdot)$ is the expectation operator. Using algebraic deduction and integration, $\mathbf{I}(\mu, \sigma^2)$ is simplified to:

$$\mathbf{I} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^4} \end{bmatrix}. \tag{10}$$

As a result, Jeffreys' prior for (μ, σ^2) is:

$$p(\mu, \sigma^2) \propto 1/\sigma^3. \tag{11}$$

It is noted that the distribution of interest is (μ, σ^2) , not (μ, σ) ; therefore, the derivative is taken with respect to σ^2 as a whole instead of σ . In other word, the parameter space is on σ^2 , not σ . For example, the Jeffreys' prior for σ^2 , with a fixed value of μ , is:

$$p(\sigma^2) \propto \frac{1}{\sigma^2}. \tag{12}$$

The Jeffreys' prior for (μ, σ) and σ with a fixed μ are $\propto 1/\sigma^2$ and $\propto 1/\sigma$, respectively. It is shown in Appendix B that the Fisher information matrix is also the Hessian matrix of the Kullback–Leibler (KL) distance of a deviated distribution with respect to the true distribution evaluated at the true parameters. For exponential families of distributions, the KL distance has analytical forms, allowing for the evaluation of the Fisher information matrix without involving integrals in Equation (9).

The asymptotically locally invariant (ALI) prior is another type of prior that satisfies the invariance under certain transformations. An ALI prior can be uniquely determined using the following equation according to [31],

$$\frac{\partial \ln p(\phi)}{\partial \phi} = -\mathbb{E}[f_1 f_2] / \mathbb{E}[f_2], \tag{13}$$

where $f_1 = \partial \ln p(x|\cdot)/\partial \phi$ and $f_2 = \partial^2 \ln p(x|\cdot)/\partial \phi^2$. For a Gaussian distribution with a fixed mean and a random σ , the two terms are $\mathbb{E}[f_1 f_2] = -6/\sigma^3$ and $\mathbb{E}[f_2] = -2/\sigma^2$. Solve:

$$\frac{\partial \ln p(\sigma)}{\partial \sigma} = -\frac{3}{\sigma} \tag{14}$$

to obtain the ALI prior for σ :

$$p(\sigma) \propto \frac{1}{\sigma^3}. \tag{15}$$

Similarly, the ALI prior for σ^2 is:

$$p(\sigma^2) \propto \frac{1}{\sigma^4}. \tag{16}$$

When both μ and σ are random, the joint ALI prior for (μ, σ) is:

$$p(\mu, \sigma) \propto \frac{1}{\sigma^5}. \tag{17}$$

It is noticed that the Jeffreys', ALI, and uniform priors are reproduced from $\propto 1/\sigma^q$ as q takes different integer values. In the following, the derivation of a $1/\sigma^q$ prior from NIG conjugates is shown.

2.2. Derivation of the $1/\sigma^q$ Priors as the Limiting States of NIG Conjugates under Gaussian Likelihood

In Bayesian models, for the given likelihood function, the posterior $p(\phi|x)$ and the prior $p(\phi)$ are called conjugate distributions if both are in the same family of distributions. It can be considered as the prior can be reconditioned by encoding the evidence through the likelihood; therefore, the evidence or data merely change the distribution parameters of the prior and yield another distribution of the same type, but with a different set of parameters.

For a location-scale parameter vector (θ, σ^2) used in linear or linearized Bayesian models with a Gaussian likelihood, the corresponding conjugate prior is the NIG distribution. The closed-form posterior PDFs for the parameter and prediction are given in Appendix A.

Noninformative priors, including the Jeffreys', ALI, and reference priors, for (θ, σ^2) are mostly in the form of $1/\sigma^q$, $q \in 1, 2, \dots$. The uniform prior can be seen as a special case of $1/\sigma^q$ as $q = 0$. It is shown as follows that these noninformative priors can be obtained as certain limiting states of NIG conjugates of (θ, σ^2) . For example, the NIG distribution with parameters $(\alpha, \beta, \Sigma, \mu)$ given by Equation (A3) can reduce to the Jeffreys' prior:

$$p(\theta, \sigma^2) \rightarrow \frac{1}{\sigma^2} \tag{18}$$

as:

$$\begin{cases} \alpha & \rightarrow -k/2 \\ \beta & \rightarrow 0^+ \\ \Sigma^{-1} & \rightarrow \mathbf{0} \\ |\mu| & < \infty \end{cases} \tag{19}$$

Furthermore, a $1/\sigma^q$ -type of prior can all be derived as reduced NIG distributions. By assigning the initial values for α , β , and Σ of the NIG distribution, different q values are obtained. Table 1 presents priors with different q values as reduced NIG distributions and the corresponding α , β , and Σ of the NIG distributions and α^* , β^* , and Σ^* of the resulting NIG posterior distributions. The posterior distribution of σ^2 is an inverse-Gamma distribution $IG(\alpha^*, \beta^*)$, and the PDF is given by:

$$p(\sigma^2|\mathbf{y}) = \frac{\beta^{*\alpha^*}}{\Gamma(\alpha^*)} \left(\frac{1}{\sigma^2}\right)^{\alpha^*+1} \exp\left(-\frac{\beta^*}{\sigma^2}\right). \tag{20}$$

Table 1. $1/\sigma^q$ -type of prior as limiting state cases of the NIG distribution and its NIG posteriors.

Prior (θ, σ^2)	NIG $(\alpha, \beta, \theta, \sigma^2) \rightarrow$ Prior	Posterior \rightarrow NIG $(\alpha^*, \beta^*, \theta, \sigma^2)$
flat	$\alpha = -k/2 - 1$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k - 2)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$
$\frac{1}{\sigma}$	$\alpha = -k/2 - 1/2$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k - 1)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$
$\frac{1}{\sigma^2}$	$\alpha = -k/2$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$
$\frac{1}{\sigma^3}$	$\alpha = -k/2 + 1/2$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k + 1)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$
$\frac{1}{\sigma^4}$	$\alpha = -k/2 + 1$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k + 2)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$
$\frac{1}{\sigma^5}$	$\alpha = -k/2 + 3/2$ $\beta \rightarrow 0$ $\Sigma^{-1} \rightarrow 0$	$\alpha^* = \alpha + n/2 = (n - k + 3)/2$ $\beta^* = \beta + \text{SSE}/2 = \text{SSE}/2$ $\Sigma^* = (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} = (\mathbf{x}^T \mathbf{x})^{-1}$

The posterior distribution of θ is obtained by integrating out σ^2 from the joint posterior distribution of Equation (A6) as,

$$\begin{aligned}
 p(\theta|\mathbf{y}) &= \int p(\theta, \sigma^2|\mathbf{y})d\sigma^2 \\
 &\propto \int \left(\frac{1}{\sigma^2}\right)^{\alpha^*+1} \exp\left\{-\frac{1}{\sigma^2}\left[\beta^* + \frac{1}{2}(\theta - \mu^*)^T \Sigma^{*-1}(\theta - \mu^*)\right]\right\}d\sigma^2 \quad (21) \\
 &\propto \left[1 + \frac{1}{2\beta^*}(\theta - \mu^*)^T \Sigma^{*-1}(\theta - \mu^*)\right]^{-(2\alpha^*+k)/2},
 \end{aligned}$$

which is a $(2\alpha^*)$ degree-of-freedom multivariate t - distribution with a location vector of μ^* and a shape matrix of $\left(\frac{\beta^*}{\alpha^*}\right)\Sigma^*$. In particular, when the NIG prior for (θ, σ^2) is reduced to the Jeffreys’ prior $1/\sigma^2$ by Equation (19), the resulting posteriors of σ^2 and θ are:

$$p(\sigma^2|\mathbf{y}) = \text{IG}\left(\frac{n - k}{2}, \frac{\text{SSE}}{2}\right), \quad (22)$$

and:

$$p(\theta|\mathbf{y}) = \text{MVT}_{n-k}\left((\mathbf{x}^T \mathbf{x})^{-1}(\mathbf{x}^T \mathbf{y}), \frac{\text{SSE}}{n - k}(\mathbf{x}^T \mathbf{x})^{-1}\right), \quad (23)$$

respectively. The term SSE in Equation (22) is the sum of squared errors, given by:

$$\text{SSE} = \sum_i^n (y_i - \mathbf{x}_i \theta)^2. \quad (24)$$

The prediction posterior in this case is:

$$p(\tilde{\mathbf{y}}|\mathbf{y}) = \text{MVT}_{n-k}\left(\tilde{\mathbf{x}}(\mathbf{x}^T \mathbf{x})^{-1}(\mathbf{x}^T \mathbf{y}), \frac{\text{SSE}}{n - k}\left(\mathbf{I} + \tilde{\mathbf{x}}(\mathbf{x}^T \mathbf{x})^{-1}\tilde{\mathbf{x}}\right)\right). \quad (25)$$

The advantage of treating the $1/\sigma^q$ -type of noninformative prior as reduced NIG conjugates is that the Bayesian posteriors of the model prediction and parameters all have analytical forms, allowing for efficient evaluations without resorting to the MCMC techniques, as done in regular Bayesian analysis. It is worth noting that the analytical results are obtained under the condition (or assumption) that the likelihood is a Gaussian. This condition implies that the conjugate prior is an NIG. The values of the NIG parameters are obtained by equating the noninformative prior to the NIG prior. In other words, equating the two distributions for the purpose of mathematical convenience is the premise of resolving those parameter values.

3. Assessment of Noninformative Priors

To evaluate the performance of priors, the Bayesian model assessment method using efficient asymptotic approximations is proposed. The idea is to recast the assessment as a model comparison problem. The participating models here are the posteriors obtained with different priors. The comparison can then be made using Bayes' factors in a Bayesian hypothesis testing context.

3.1. Fitting Performance

The Bayes factor, on the basis of observed data \mathbf{y} , evaluating the plausibility of two different models, M_1 and M_2 , can be expressed as:

$$B_{12} = \frac{P(\mathbf{y}|M_1)}{P(\mathbf{y}|M_2)} = \frac{\int p(\mathbf{y}|\theta, \sigma^2, M_1)p(\theta, \sigma^2|M_1)d\theta d\sigma^2}{\int p(\mathbf{y}|\theta, \sigma^2, M_2)p(\theta, \sigma^2|M_2)d\theta d\sigma^2} \tag{26}$$

The comparison of two models can be made based on the ratio of the posterior probabilities of the models:

$$\frac{P(M_1|\mathbf{y})}{P(M_2|\mathbf{y})} = \frac{P(\mathbf{y}|M_1)P(M_1)}{P(\mathbf{y}|M_2)P(M_2)} = B_{12} \cdot \frac{P(M_1)}{P(M_2)}. \tag{27}$$

The ratio, when the prior probabilities of the models are equal, is reduced to the Bayes factor of Equation (26). The assessment of the Bayes factors involves two integrals over the parameter space of (θ, σ^2) . The integrands for models M_1 and M_2 are $p(\mathbf{y}|\theta, \sigma^2, M_1) \cdot p(\theta, \sigma^2|M_1)$ and $p(\mathbf{y}|\theta, \sigma^2, M_2) \cdot p(\theta, \sigma^2|M_2)$, respectively.

For general multidimensional integration, asymptotic approximation or simulation-based estimation are two commonly used methods. The Laplace approximation method evaluates the integral as a multivariate normal distribution. Consider the above integrand term in Equation (26), i.e., $p(\mathbf{y}|\theta, \sigma^2, M_1) \cdot p(\theta, \sigma^2|M_1)$. Drop the model symbol for simplicity of the derivation, and denote (θ, σ^2) as ϕ . The natural logarithm of the integrand, $p(\mathbf{y}|\phi)p(\phi)$, i.e., $p(\mathbf{y}, \phi)$, can be expressed using Taylor expansion around its mode ϕ' as:

$$\begin{aligned} \ln p(\mathbf{y}, \phi) &= \ln p(\mathbf{y}, \phi') + (\phi - \phi')^T \nabla \ln p(\mathbf{y}, \phi') \\ &+ \frac{1}{2}(\phi - \phi')^T \left[\nabla^2 \ln p(\mathbf{y}, \phi') \right] (\phi - \phi') \\ &+ O[(\phi - \phi')^3], \end{aligned} \tag{28}$$

where $\nabla \ln p(\mathbf{y}, \phi')$ is the gradient of $\ln p(\mathbf{y}, \phi)$ evaluated at ϕ' , $\nabla^2 \ln p(\mathbf{y}, \phi')$ is the Hessian matrix of $\ln p(\mathbf{y}, \phi)$ evaluated at ϕ' , and $O[\cdot]$ are higher-order terms. When the higher-order terms are negligible,

$$\begin{aligned} \ln p(\mathbf{y}, \phi) &\approx \ln p(\mathbf{y}, \phi') + \underbrace{(\phi - \phi')^T \nabla \ln p(\mathbf{y}, \phi')}_{(o)} \\ &+ \frac{1}{2}(\phi - \phi')^T \left[\nabla^2 \ln p(\mathbf{y}, \phi') \right] (\phi - \phi'). \end{aligned} \tag{29}$$

The term (\circ) is zero at the mode of the distribution where the gradient is zero; therefore, expanding $\ln p(\mathbf{y}, \phi)$ around ϕ' eliminates the term (\circ) and yields:

$$\ln p(\mathbf{y}, \phi) \approx \ln p(\mathbf{y}, \phi') + \frac{1}{2}(\phi - \phi')^T \left[\nabla^2 \ln p(\mathbf{y}, \phi') \right] (\phi - \phi'). \tag{30}$$

Exponentiate the above equation to obtain:

$$p(\mathbf{y}, \phi) \approx p(\mathbf{y}, \phi') \exp \left\{ -\frac{1}{2}(\phi - \phi')^T \left[-\nabla^2 \ln p(\mathbf{y}, \phi') \right] (\phi - \phi') \right\}. \tag{31}$$

Realizing the first term of the above equation is a constant and the last term is the variable part of a multivariate normal distribution with a mean vector of ϕ' and a covariance matrix $\Sigma' = [-\nabla^2 \ln p(\mathbf{y}, \phi')]^{-1}$, the integration of $p(\mathbf{y}, \phi)$ writes,

$$\begin{aligned} P(\mathbf{y}) &= p(\mathbf{y}, \phi') \cdot \int_{\Phi} \exp \left[-\frac{1}{2}(\phi - \phi')^T [\Sigma']^{-1} (\phi - \phi') \right] d\phi \\ &= p(\mathbf{y}, \phi') \sqrt{(2\pi)^{k+1} |\Sigma'|}. \end{aligned} \tag{32}$$

Notice that θ is a k -dimensional vector, so $\phi = (\theta, \sigma^2)$ is a $k + 1$ -dimensional vector. Using the result of Equation (32), the integral associated with the j th model in Equation (26) is:

$$P(\mathbf{y}|M_j) \approx p(\mathbf{y}, (\theta, \sigma^2)' | M_j) \sqrt{(2\pi)^{k+1} |\Lambda'_j|}. \tag{33}$$

where $(\theta, \sigma^2)'$ is the mode of the distribution and Λ'_j is the covariance matrix, i.e., the inverse of the negative Hessian matrix of $\ln p(\mathbf{y}, \theta, \sigma^2 | M_j)$ evaluated at $(\theta, \sigma^2)'$. Proper numerical derivative methods based on finite difference schemes can achieve reliable results for the gradient and the Hessian in the following equation.

$$\begin{cases} (\theta, \sigma^2)' = \arg \max_{(\theta, \sigma^2)} \left[\nabla \ln p(\mathbf{y}, \theta, \sigma^2 | M_j) \right] \\ \Lambda' = \left[-\nabla^2 \ln p(\mathbf{y}, \theta, \sigma^2 | M_j) \Big|_{(\theta, \sigma^2)'} \right]^{-1}, \end{cases} \tag{34}$$

Experience has shown that for a distribution with a single mode and approximately symmetric, the Laplace approximation method can achieve accurate results for engineering applications [38,39].

It should be noted that the $1/\sigma^q$ -type of noninformative prior is not a proper prior, i.e., the normalizing constant for $1/\sigma^q$ is not bounded for $\sigma \in (0, +\infty)$; therefore, it is necessary to limit the support of $1/\sigma^q$ to a proper range and to have a finite normalizing constant. Assume the effective range of σ is (σ_-, σ_+) for $q \in (0, 1, \dots)$; the normalizing constant is:

$$Z(\sigma) = \begin{cases} \ln(\sigma_{+1}) - \ln(\sigma_-) & q = 1 \\ \frac{1}{1-q} (\sigma_+^{1-q} - \sigma_-^{1-q}) & q \neq 1 \end{cases}. \tag{35}$$

Substitute $p(\theta, \sigma^2)$ with the $1/(Z(\sigma)\sigma^q)$ in Equation (3). The likelihood of Equation (33) compares the fitting performance of different $1/\sigma^q$ -type of priors.

3.2. Predictive Performance

The predictive performance for a model is measured by the likelihood of the data that are not used for model parameter estimation. The likelihood of the future measurement data $\tilde{\mathbf{y}}$ with the posterior PDF of $p(\theta, \sigma^2 | \mathbf{y})$ writes:

$$\begin{aligned}
 P(\tilde{\mathbf{y}}) &= \int p(\tilde{\mathbf{y}}|\theta, \sigma^2)p(\theta, \sigma^2|\mathbf{y})d\theta d\sigma^2 \\
 &= \frac{1}{P(\mathbf{y})} \int p(\tilde{\mathbf{y}}|\theta, \sigma^2)p(\mathbf{y}|\theta, \sigma^2)p(\theta, \sigma^2)d\theta d\sigma^2,
 \end{aligned}
 \tag{36}$$

where $P(\mathbf{y})$ is evaluated using Equation (33).

The comprehensive performance integrating both fitting performance and predictive performance in terms of likelihood in the joint space of $(\mathbf{y} \times \tilde{\mathbf{y}})$ is:

$$P(\tilde{\mathbf{y}})P(\mathbf{y}) = \int p(\tilde{\mathbf{y}}|\theta, \sigma^2)p(\mathbf{y}|\theta, \sigma^2)p(\theta, \sigma^2)d\theta d\sigma^2.
 \tag{37}$$

Realize that Equations (32), (36), and (37) can all be evaluated using the method of Laplace approximation, and the performance of different priors can quantitatively be compared. It is noted that the likelihood is the product of the fitting component and the prediction component, and the influence of the prior is incorporated in the fitting component, i.e., $P(\mathbf{y})$.

4. Application Examples

To illustrate the proposed Bayesian inference with noninformative priors, two examples are presented. The results were compared with the classical least squares approach. The fitting and comprehensive performance of the $1/\sigma^q$ -type of priors were evaluated using the method of Laplace approximation.

4.1. Low-Cycle Fatigue Reliability Assessment with Sparse Samples

The above numerical example reveals the advantage of using a noninformative prior over the flat prior in probabilistic inference under small sample sizes. To examine its usefulness in realistic problems and further compare the performance of different noninformative priors, a fatigue life prediction problem is presented.

The low-cycle fatigue testing data reported in ASTM E739-10 [40] were used and shown in Table 2. The cyclic strain amplitudes ($\Delta\varepsilon/2$) were (~ 0.016 , ~ 0.0068 , ~ 0.0016 , ~ 0.0005), and each level had two or three data points. The following log-linear model was adopted,

$$\ln N = a_0 + a_1 \ln(\Delta\varepsilon/2),
 \tag{38}$$

where a_0 and a_1 are model parameters that need to be identified. It should be noted that other variants of the above equation can be used to include material plasticity, the mean stress effect, etc. The discrepancy between the experimental data and the model prediction can be modeled using an uncertain variable with zero mean and a standard deviation of σ_e . To compare with the regular least squares method, a Gaussian likelihood was assumed, and the Bayesian posterior is:

$$\begin{aligned}
 p(a_0, a_1, \sigma_e | N) &\propto \frac{1}{\sigma^q} \cdot \left(\frac{1}{\sigma^2}\right)^{n/2} \\
 &\times \exp\left[-\frac{\sum_{i=1}^n (\ln N_i - a_0 - a_1 \ln(\Delta\varepsilon_i/2))^2}{2\sigma^2}\right].
 \end{aligned}
 \tag{39}$$

Table 2. Low-cycle fatigue testing data. Source: [40].

<i>i</i>	1	2	3	4	5	6	7	8	9
$\Delta\varepsilon/2$	0.01636	0.01609	0.00675	0.00682	0.00179	0.00160	0.00165	0.00053	0.00054
<i>N</i>	168	200	1000	1180	4730	8035	5254	28,617	32,650

The exponent q takes the values of $0, 1, 2, \dots$ to generate a flat prior and different noninformative priors, and $(\Delta\varepsilon_i, N_i), i = 1, \dots, n$ is the i th experimental data points of the total n data points used for parameter estimation.

Notice that when $q = 0$, the above equation reduces to a regular least squares format. One data point ($i = 6$) was arbitrarily chosen for prediction performance evaluation, and the rest eight data points were used to estimate the model parameters using Equation (39).

To compare the performance of $1/\sigma^q$ -type priors, $q = 0, \dots, 5$ were used to represent the regular least squares, Jeffreys', and ALI priors. The likelihood of the fitting performance and prediction performance was evaluated using the method of Laplace approximation, and results are shown in Figure 1. The total likelihood of the fitting and prediction results indicates that the Jeffreys' prior $1/\sigma^2$ outperformed the others. The result associated with $q = 0$ is the Bayesian version of the regular least squares estimation.

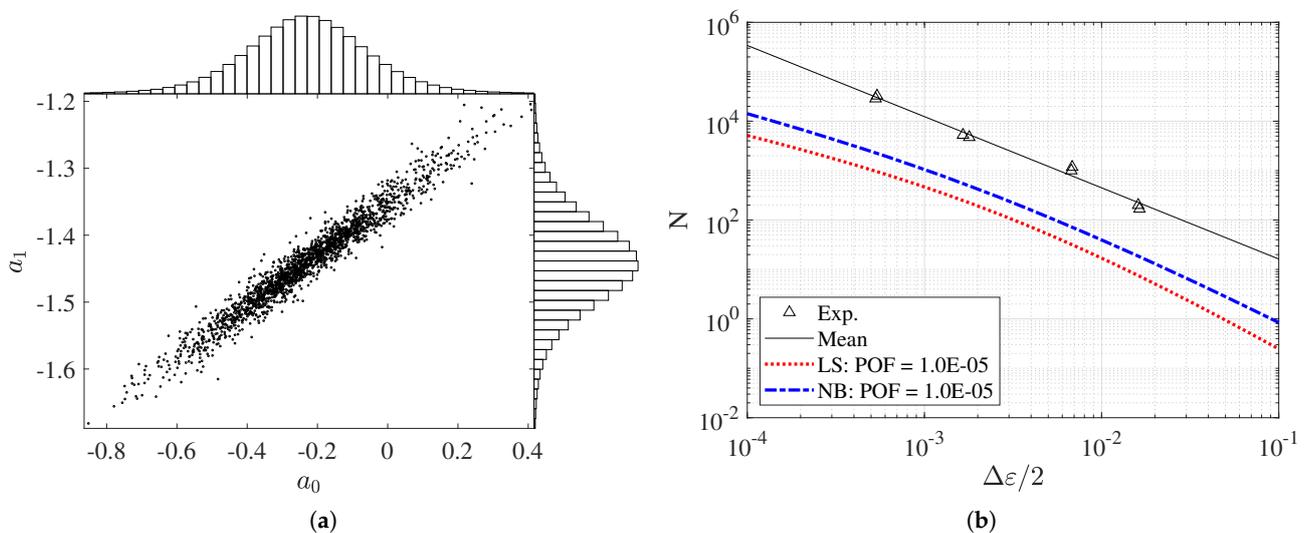


Figure 1. (a) MCMC samples drawn from the noninformative Bayesian posterior with Jeffreys' prior ($q = 2$) and (b) the fatigue life results at $POF = 10^{-5}$ obtained with the regular least squares (LS) estimator and noninformative Bayesian (NB).

To further demonstrate the difference between the regular least squares and noninformative Bayesian approach under this sample size, the following reliability problem was considered. The fatigue life given a prescribed probability of failure (POF) was calculated for the cyclic strain range ($\Delta\varepsilon/2$) between 10^{-4} and 10^{-1} . Given a cyclic strain range and a POF, the fatigue life N_{POF} can be expressed as:

$$\Pr(N > N_{POF}) = 1 - POF. \tag{40}$$

For the regular least squares estimator, it can be obtained using the t-statistics given by Equation (A9) where $\gamma/2$ should be replaced with the prescribed POF because only the lower one-sided bound is needed. For the Bayesian with Jeffreys' prior, the one-side bound can trivially be obtained using the POF-quantile of the prediction results evaluated using MCMC samples of the posterior POF. Figure 2 presents the comparisons of the fatigue life results where the obvious difference was observed. The results of the noninformative Bayesian were larger than those of regular least squares, indicating a longer fatigue life for the same risk level, i.e., 10^{-5} POF.

4.2. Aeroengine Turbine Disk Lifting and Failure Rate Assessment

The prescribed safe life is a critical parameter for airworthiness and service health management of turbine disks in aeroengines. To determine the prescribed safe life for turbine disks, life testing with a number of disks is performed using disk samples in a rotating pit. The number of tested piece is usually less than five due to the cost and time

constraints; sometimes, only one or two test pieces are available. The determination of the prescribed safe life must be carefully made to mitigate the risk. A general and standard procedure described in [41] is centered on the idea that two source of uncertainty, namely the sampling uncertainty and the life distribution uncertainty, should be incorporated. To account for the sampling uncertainty, the 95% single-sided lower bound of the sample mean is used as the mean of the life distribution. To account for the uncertainty of the life distribution itself, no more than one in seven-hundred fifty service disks is allowed to develop an engineering crack. The total safety margin is a 1/750 cumulative failure probability to a 95% confidence.

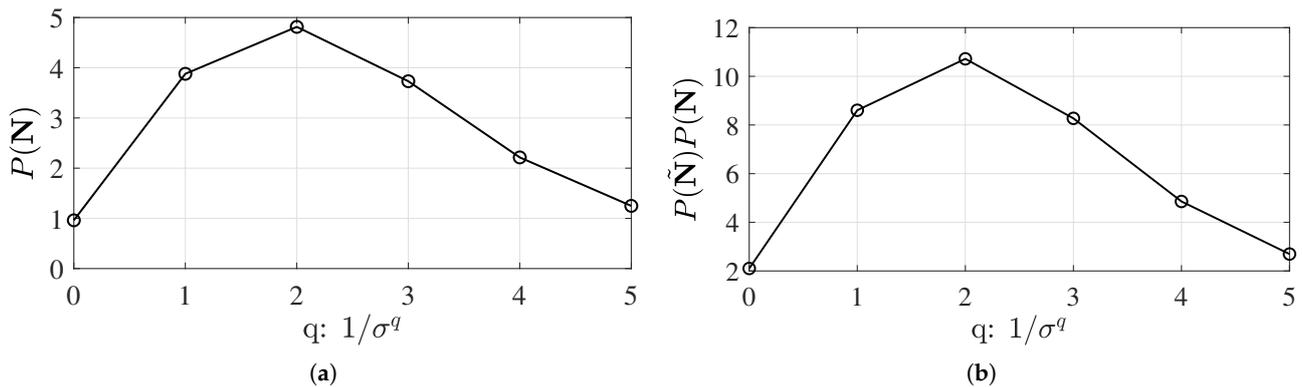


Figure 2. (a) The fitting performance $P(N)$ evaluated with different $1/\sigma^q$ priors and (b) the prediction performance $P(\tilde{N})P(N)$ evaluated with different $1/\sigma^q$ priors. One data point \tilde{N} is considered for the prediction.

The prescribed safe life is assumed to be a log-normal random variable, and the log-life is a normally distributed variable. Denote the mean and standard deviation of the normal distribution of the log-life as $\ln N_\mu$ and σ . The 95% single-sided lower bound of the sample mean can be expressed as:

$$\ln N_{\mu}^{95\%} = \ln N_{\mu} - \lambda_{\alpha} \frac{\sigma}{\sqrt{n}}, \tag{41}$$

where N_{μ} is the geometric mean of the life, $N_{\mu,95\%}$ is the single-sided lower bound of the geometric mean, $\lambda_{\alpha} = \Phi^{-1}(1 - \alpha)$, and $\Phi^{-1}(\cdot)$ is the inverse cumulative density function (CDF) of the standard normal variable when the standard deviation σ is known. For the 95% single-sided lower bound, $z_{\alpha} = \Phi^{-1}(0.95) = 1.645$. When the standard deviation is unknown, the estimate of the standard deviation based on the samples should be used. In such cases, the value of λ_{α} is obtained using the Student distribution with a degree of freedom (DOF) of $n - 1$, i.e., $\lambda_{\alpha} = t^{-1}(1 - \alpha, n - 1)$. The function $t^{-1}(\cdot, n - 1)$ is the inverse CDF of the Student distribution with a DOF of $n - 1$. The 1/750 failure probability log-life corresponds to the -3σ shift from the mean of the log-life distribution. When using $\ln N_{\mu}^{95\%}$ as the mean of the log-life distribution, the 1/750 failure probability life $N_{\mu-3\sigma}^{95\%}$ writes:

$$\ln N_{\mu-3\sigma}^{95\%} = \ln N_{\mu}^{95\%} - 3\sigma. \tag{42}$$

Combine Equations (41) and (42) to have:

$$\ln N_{\mu} - \ln N_{\mu-3\sigma}^{95\%} = \lambda_{\alpha} \frac{\sigma}{\sqrt{n}} + 3\sigma, \tag{43}$$

The safety factor that ensures a 1/750 failure to a 95% confidence can be expressed by transforming Equation (43) into the linear scale as:

$$S = \frac{N_{\mu}}{N_{\mu-3\sigma}^{95\%}} = \exp \left[\left(\frac{\lambda_{\alpha}}{\sqrt{n}} + 3 \right) \sigma \right] \tag{44}$$

It is seen that the prescribed safe life depends on the number of tested samples n and the standard deviation of the distribution. The deterministic prescribed safe life A_r is finally obtained as:

$$A_r = N_\mu / S. \tag{45}$$

The cumulative burst probability when the disk is used y cycles can be expressed as:

$$F_b(y|x') = \int_0^y \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-\frac{(t-x')^2}{2\sigma_b^2}\right] dt, \tag{46}$$

where x' is the log-transformed geometric mean of the samples and y is the log-transformed number of cycles, i.e., $y = \ln N$. The failure rate in terms of the failures per cycle can be expressed as:

$$h(y|x') = \frac{\partial F_b(y|x') / \partial y}{1 - F_b(y|x')} = \frac{f_b(y|x')}{1 - F_b(y|x')} \tag{47}$$

where $f_b(y|x')$ is the burst PDF:

$$f_b(y|x') = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left[-\frac{(y-x')^2}{2\sigma_b^2}\right]. \tag{48}$$

The sampling uncertainty in the log-transformed geometric mean x' can be quantified using the sampling PDF. The sample mean x' distributes according to:

$$f_s(x') = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_s} \exp\left[-\frac{(x' - \ln N_\mu)^2}{2\sigma_s^2}\right] & \sigma \text{ known} \\ \ln N_\mu + \frac{s}{\sqrt{n}} \cdot t(n-1) & \sigma \text{ unknown} \end{cases}, \tag{49}$$

where s is the standard deviation of the sample, $t(n-1)$ is the standard Student distribution with a DOF of $n-1$, and $\sigma_s = \frac{\sigma}{\sqrt{n}}$.

The final failure rate can be obtained by marginalizing out the sampling error variable x' as:

$$h(y) = \int h(y|x') f_s(x') dx'. \tag{50}$$

With the above discussion, turbine disk life testing results on two disks were obtained as shown in Table 3.

Table 3. Spin rig testing results on two disks.

Disk No.	No. of Cycles
1	27,000
2	21,700

It is seen that the key information is the standard deviation of the population. For one thing, whether it is considered as a known quantity requires a justification, and its actual value relies on empirical evidence and historical data. For another, when it needs to be estimated from the samples, the proper method should be used due to its limited sample size.

4.2.1. Existing Method

Previous disk life testing results showed that the scatter factor, defined as the ratio of the maximum life (N_{\max}) over the minimum life (N_{\min}), is less than six. Furthermore, N_{\max} and N_{\min} are considered as the $+3\sigma$ and -3σ points of the normal distribution of

the log-life. With the above empirical evidence and assumption The empirical evidence of $N_{\max}/N_{\min} = 6$, combined with the assumption that $\ln N_{\max} - \ln N_{\min} = 6\sigma$, yields:

$$\sigma = \frac{\ln N_{\max} - \ln N_{\min}}{6} = \frac{1}{6} \ln\left(\frac{N_{\max}}{N_{\min}}\right) = \frac{\ln 6}{6}. \tag{51}$$

The safety factor that ensures a 1/750 failure to a 95% confidence can be expressed, by substituting Equation (51) into Equation (43) as,

$$S = \frac{N_{\mu}}{N_{\mu-3\sigma}^{95\%}} = 6^{\frac{1}{2} + \frac{1.645}{6\sqrt{n}}} \tag{52}$$

Given that assumption that the standard deviation of the log-life distribution is known, the first equation of the right-hand side in Equation (49) is used. In addition, the term $1 - F_b(y|x') \approx 1$ for rare events. In such a case, the hazard rate function Equation (50) reduces to the convolution of two normal PDFs, and the resulting hazard rate function is:

$$h(y) = \int h(y|x')f_s(x')dx' \approx \text{Norm}(y, \ln N_{\mu}, \sigma_b^2 + \sigma_s^2). \tag{53}$$

Note that $y = \ln N$; the hazard rate in terms of failure per flight hour is found by using the chain rule as,

$$h(H) = \frac{1}{H} \cdot \text{Norm}(\ln(\beta H), \ln N_{\mu}, \sigma_b^2 + \sigma_s^2), \tag{54}$$

where β is the exchange rate (cycles/hour) converting the cycle to equivalent engine flight hours. Using the data in Table 3, the log-transformed geometric mean $\ln N_{\mu} = 24,205$, the standard deviation of the log-life distribution is $\sigma_b = (\ln 6)/6$, and the standard deviation of the sampling distribution is $\sigma_s = (\ln 6)/(6\sqrt{2})$. Substitute these parameters into Equation (45) to obtain the prescribed safe life as $A_r = 6981$ cycles.

4.2.2. The Proposed Noninformative Bayesian Method

Without relying on the empirical evidence and the assumption, the noninformative Bayesian estimator for (μ, σ) of the log-life distribution can be obtained by writing the posterior of the likelihood of samples with a noninformative prior. For a demonstration, the prior of $\propto 1/\sigma^2$ was used, and the posterior writes,

$$p(\mu, \sigma|D) \propto \frac{1}{\sigma^2} \cdot \left(\frac{1}{\sigma^2}\right)^{n/2} \cdot \exp\left[-\frac{\sum_{i=1}^n (\ln N_i - \mu)^2}{2\sigma^2}\right], \tag{55}$$

where D denotes the event of observed life samples N_1, \dots, N_n . To estimate μ and σ , 1×10^7 samples were drawn from Equation (55) using the MCMC method. The resulting histogram of the samples is shown in Figure 3. The 95% single-sided confidence bound $\mu^{95\%}$ can be estimated using the 0.05-quantile of the MCMC samples as 9.86565 or 19,259 in the linear scale. Centering on $\mu^{95\%} = 9.86565$, the value corresponding to a 1/750 cumulative failure probability was found by subtracting $\mu^{95\%}$ by 3σ . The term σ was estimated using the MCMC sample mean, and in this case, it was found as 0.1893. Consequently, after applying the two safety margins, the prescribed safe life (in log scale) using the noninformative Bayesian method was found as $\mu_{-3\sigma}^{95\%} = 9.2976$ or $A_r = 10,911$ cycles.

The failure rates calculated using the existing method and the proposed noninformative Bayesian method are compared in Figure 4. The failure rate per flight hour was based on an exchange rate $\beta = 2.5$.

Based on the results shown in Figure 4, the usable flight hours obtained using the noninformative Bayesian method were about 1000 h more than those obtained using the existing method given the same risk constraint. For example, the flight hours corresponding to a failure rate of $10^{-8}/\text{h}$ were 2734 and 1734, respectively, for the two methods. When the failure rate was required to be $10^{-6}/\text{h}$, the usable flight hours were 3817 and 2720,

respectively. It should be noted that the proposed noninformative Bayesian method does not rely on empirical evidence and assumptions other than the log-normal distribution for disk life.

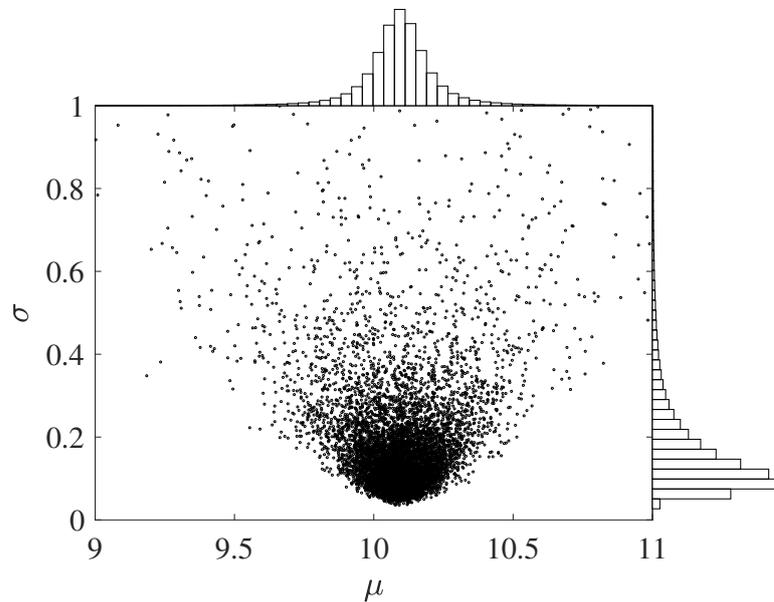


Figure 3. Samples drawn from the posterior PDF and histograms.

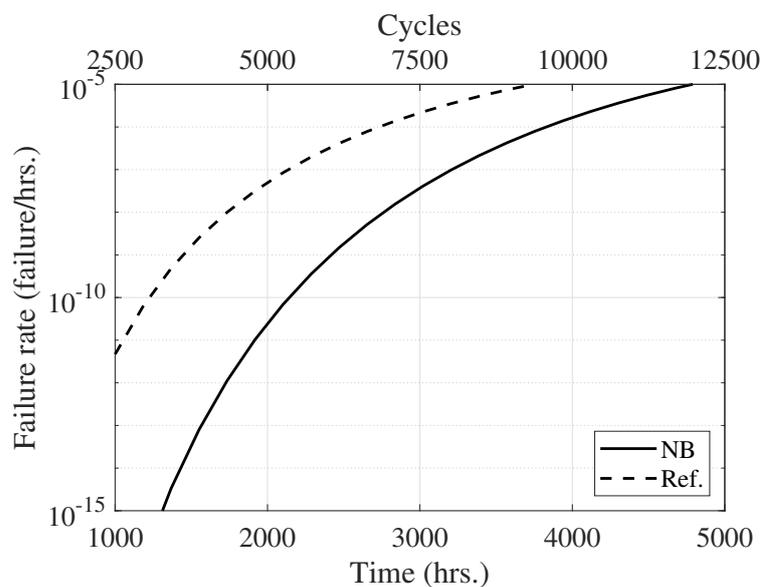


Figure 4. Comparisons of the failure rate results of the two methods. The solid line and the dashed line are results from the noninformative Bayesian (NB) and existing method (Ref.), respectively.

5. Conclusions

The study developed a noninformative Bayesian inference method for small sample problems. In particular, a $1/\sigma^q$ -type of prior was proposed to formulate the Bayesian posterior. It was shown that this type of prior can represent the classical Jeffreys’ prior (Fisher information), flat prior, and asymptotic locally invariant prior. More importantly, this type of prior was derived as the limiting state of normal-inverse-Gamma (NIG) conjugate priors, allowing for fast and efficient evaluations of the Bayesian posteriors and predictors in closed-form expressions under a Gaussian likelihood. To illustrate the overall method and compare the performance of noninformative priors, a realistic fatigue reliability problem was discussed in detail. The combined fitting and prediction performance in terms of

the likelihood was used to evaluate different priors. It was observed that Jeffreys' prior, $1/\sigma^2$, yielded the maximum likelihood. The developed method was further applied to an actual aeroengine disk life problem with two test samples, and the results on the prescribed safe life and risk in terms of failure/hour were comparable with those obtained using the existing method incorporating the empirical evidence and assumptions. It should be emphasized that the study did not promote abandoning justified experience and/or information in prior construction. It was argued in this study that in the absence of those data or experience, the noninformative prior should be used to avoid introducing unjustified information into the prior. The following conclusions were drawn based on the current results:

- The $1/\sigma^q$ -type of prior can be used equivalently as the NIG conjugate priors at their limiting states. The great features of conjugate priors, such as having analytical posterior and prediction PDFs under a Gaussian likelihood, can be retained when using noninformative priors for Bayesian linear regression analysis;
- For the $1/\sigma^q$ -type of noninformative prior, the classical Jeffreys' prior $1/\sigma^2$ yielded an optimal fitting and prediction performance in terms of the likelihood or Bayes factors. When $q = 0$, the Bayesian estimator reduces to the regular least squares estimator. The results of the two case studies showed the advantage of using the noninformative Bayesian estimator with the prior of $\propto 1/\sigma^2$ over the regular least squares estimator and other empirical methods under small sample sizes.

Author Contributions: Conceptualization, X.G.; Data curation, S.W.; Formal analysis, W.W. and M.H.; Methodology, J.H., W.W. and X.G.; Resources, M.H. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by National Natural Science Foundation of China, Nos. 11872088, 12088101, NSAF U1930403, CAEP CX20200035, and 2020-JCJQ-JJ-465. The support is gratefully acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The study was supported by National Natural Science Foundation of China, Nos. 11872088, 12088101, NSAF U1930403, CAEP CX20200035, and 2020-JCJQ-JJ-465. The support is gratefully acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Conjugate Prior, Posterior, and Prediction

Assume a general linear model with a parameter vector θ and a Gaussian likelihood with a scale parameter σ^2 , i.e., the modeling error $\epsilon_i = (y_i - \mathbf{x}_i\theta) \sim \text{Norm}(0, \sigma^2)$.

Appendix A.1. Conjugate Prior

It is known that when the likelihood function belongs to the exponential family, a conjugate prior exists and belongs also to the exponential family. To obtain the joint conjugate using $p(\theta, \sigma^2) = p(\theta|\sigma^2)p(\sigma^2)$, it is necessary to look at the conjugate prior of σ^2 first. For a Gaussian likelihood with an unknown variance σ^2 and zero mean, the conjugate prior for σ^2 is an inverse Gamma distribution, i.e., $p(\sigma^2) \sim \text{IG}(\alpha, \beta)$, and is expressed as:

$$p(\sigma^2) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \cdot \exp\left(-\frac{\beta}{\sigma^2}\right), \quad (\text{A1})$$

where $\alpha > 0$ and $\beta > 0$ are shape and scale parameters, respectively. The posterior distribution of σ^2 , given n data points, has the following new shape and scale parameters:

$$\begin{aligned}
 (\alpha', \beta') &= \left(\alpha + \frac{n}{2}, \beta + \frac{\sum_i^n (y_i - \mathbf{x}_i \theta)^2}{2} \right) \\
 &= \left(\alpha + \frac{n}{2}, \beta + \frac{\text{SSE}}{2} \right),
 \end{aligned}
 \tag{A2}$$

where $\text{SSE} = \sum_i^n (y_i - \mathbf{x}_i \theta)^2$ is the sum of squared errors. Consider the joint conjugate of (θ, σ^2) , a widely adopted conjugate prior for Bayesian linear regression is the normal-inverse-Gamma (NIG) prior for (θ, σ^2) , and it can be written as Equation (A3).

$$\begin{aligned}
 p(\theta, \sigma^2) &= p(\theta | \sigma^2) p(\sigma^2) = \text{NIG}(\mu_\theta, \Sigma_\theta, \alpha, \beta) \\
 &= Z \cdot \exp \left\{ -\frac{1}{\sigma^2} \left[\beta + \frac{1}{2} (\theta - \mu_\theta)^T \Sigma_\theta^{-1} (\theta - \mu_\theta) \right] \right\} \\
 &\propto Z' \cdot \exp \left\{ -\frac{1}{\sigma^2} \left[\beta + \frac{1}{2} (\theta - \mu_\theta)^T \Sigma_\theta^{-1} (\theta - \mu_\theta) \right] \right\},
 \end{aligned}
 \tag{A3}$$

where:

$$Z = \frac{\beta^\alpha}{(2\pi)^{k/2} \Gamma(\alpha) \sqrt{|\Sigma_\theta|}} \left(\frac{1}{\sigma^2} \right)^{\alpha+k/2+1},
 \tag{A4}$$

and:

$$Z' = \left(\frac{1}{\sigma^2} \right)^{\alpha+k/2+1}.
 \tag{A5}$$

Using the inverse gamma distribution for σ^2 , the initial distribution parameter (α, β) still needs to be determined. Unfortunately, there is no formal rule to do that. Gelman [35] discussed the inverse-Gamma distribution $\text{IG}(\epsilon, \epsilon)$ where ϵ was set to a low value such as 1, 0.01, or 0.001. A difficulty of this prior is that ϵ must be set to a proper value. Inferences become very sensitive to ϵ for cases where low values of σ^2 are possible. Due to this reason, the $\text{IG}(\epsilon, \epsilon)$ family of priors for σ^2 was not recommended by the author. It was further argued that the resulting inference based on it for cases where σ^2 is estimated near zero was the case for which classical and Bayesian inferences differed the most [35]. Browne and Draper [42] also mentioned this disadvantage and suggested using a uniform one $\text{U}(0, 1/\epsilon)$ to reduce this defect. However, choosing an appropriate value for ϵ for $\text{U}(0, 1/\epsilon)$ still requires attention and understanding of the data. Indeed, the effect of the prior and its parameters is highly subjected to the data and problem, and its sensitivity to Bayesian inferences based on sparse data or extensive data can be dramatically different as the data dominants the posterior due to *the law of large numbers*.

Appendix A.2. Posterior Distribution

The posterior distribution of (θ, σ^2) is expressed as:

$$p(\theta, \sigma^2 | \mathbf{y}) = \frac{p(\theta, \sigma^2) p(\mathbf{y} | \theta, \sigma^2)}{p(\mathbf{y})}.
 \tag{A6}$$

The term $p(\mathbf{y}) = \int p(\theta, \sigma^2) p(\mathbf{y} | \theta, \sigma^2) d\theta d\sigma^2$ is the marginal distribution of the data. It can be shown that after some algebraic operations, the posterior can be written as:

$$\begin{aligned}
 p(\theta, \sigma^2 | \mathbf{y}) &\propto \left(\frac{1}{\sigma^2} \right)^{\alpha^*+k/2+1} \\
 &\times \exp \left\{ -\frac{1}{\sigma^2} \left[\beta^* + \frac{1}{2} (\theta - \mu^*)^T \Sigma^{*-1} (\theta - \mu^*) \right] \right\},
 \end{aligned}
 \tag{A7}$$

where:

$$\begin{cases} \mu^* &= (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} (\Sigma^{-1} \mu + \mathbf{x}^T \mathbf{y}) \\ \Sigma^* &= (\Sigma^{-1} + \mathbf{x}^T \mathbf{x})^{-1} \\ \alpha^* &= \alpha + n/2 \\ \beta^* &= \beta + \frac{1}{2} (\mu^T \Sigma^{-1} \mu + \mathbf{y}^T \mathbf{y} - \mu^{*T} \Sigma^{*-1} \mu^*) \end{cases}$$

The initial NIG conjugate parameter $(\mu, \Sigma, \alpha, \beta)$ is updated to a new set of parameters $(\mu^*, \Sigma^*, \alpha^*, \beta^*)$.

Appendix A.3. Prediction

The Bayesian prediction of the response given the data \mathbf{y} can be obtained using the above distributions with a new set of input variable $\tilde{\mathbf{x}}$. Denote the corresponding prediction as $\tilde{\mathbf{y}}$. The prediction posterior distribution of $\tilde{\mathbf{y}}$ can be expressed as:

$$\begin{aligned} p(\tilde{\mathbf{y}}|\mathbf{y}) &= \int p(\tilde{\mathbf{y}}|\theta, \sigma^2) p(\theta, \sigma^2|\mathbf{y}) d\theta d\sigma^2 \\ &= \int \text{MVN}(\tilde{\mathbf{x}}\theta, \sigma^2 \mathbf{I}) \times \text{NIG}(\mu^*, \Sigma^*, \alpha^*, \beta^*) d\theta d\sigma^2 \\ &= \text{MVT}_{2\alpha^*} \left(\tilde{\mathbf{x}}\mu^*, \frac{\beta^*}{\alpha^*} (\mathbf{I} + \tilde{\mathbf{x}}\Sigma^* \tilde{\mathbf{x}}^T) \right). \end{aligned} \tag{A8}$$

when a single point is evaluated, Equation (A8) is reduced to a one-dimensional Student distribution, and the confidence interval is readily evaluated using the inverse Student distribution. For multipoint simultaneous evaluations, the interval contours can be approximated using a few methods as suggested in [43].

To compare, the simple linear regression results of the prediction mean and bounds are given. The prediction mean is $\tilde{\mathbf{x}}\hat{\theta}$, and the prediction bounds of level $(1 - \gamma)\%$ are:

$$\tilde{\mathbf{y}}_{1-\gamma} = \tilde{\mathbf{x}}\hat{\theta} \pm \text{tinv}(\gamma/2, n - k) \cdot \alpha_{\mathbf{y}}, \tag{A9}$$

where $\hat{\theta} = (\mathbf{x}^T \mathbf{x})^{-1} (\mathbf{x}^T \mathbf{y})$ is the maximum likelihood estimator of θ . It is known that $\hat{\theta}$ is also identical to the least squares estimator and the Bayesian estimator, i.e., Equation (23). The term $\text{tinv}(\gamma/2, n - k)$ represents the $\gamma/2$ -percentile value of the standard Student t-distribution with $n - k$ degrees of freedom, and $\alpha_{\mathbf{y}}$ is the standard error of the prediction given by:

$$\sigma_{\tilde{\mathbf{y}}}^2 = \frac{\mathbf{e}^T \mathbf{e}}{n - k} \cdot [\mathbf{I} + \tilde{\mathbf{x}}(\mathbf{x}^T \mathbf{x})^{-1} \tilde{\mathbf{x}}^T], \tag{A10}$$

where the term $\mathbf{e} = \mathbf{y} - \mathbf{x}\hat{\theta}$ is the discrepancy vector between the model and observation, also called the residual or error. As multiple points are estimated, the F-statistic can be used instead of the t-statistic to yield the Working–Hotelling confidence intervals [44],

$$\tilde{\mathbf{y}}_{1-\gamma} = \tilde{\mathbf{x}}\hat{\theta} \pm \sqrt{2\text{Finv}(\gamma, k, n - k)} \cdot \alpha_{\mathbf{y}}, \tag{A11}$$

where $\text{Finv}(\gamma, k, n - k)$ represents the γ -percentile value of the F-distribution with $(k, n - k)$ degrees of freedom.

Appendix B. Fisher Information Matrix and Kullback–Leibler Divergence

The equivalence of the Fisher information matrix and the Kullback–Leibler divergence can be shown as follows. The matrix form of the Fisher information matrix of a continuous PDF with a parameter vector θ writes,

$$\mathbf{I}(\theta) = \mathbb{E}_{\theta} \left[\nabla \ln p_{\theta}(x) \cdot \nabla \ln p_{\theta}(x)^T \right]. \tag{A12}$$

Under the condition that the integration and derivation may be exchanged and the log-function is twice differentiable, the following Equation (A13) can be attained.

$$\begin{aligned} \nabla^2 \ln p_\theta(x) &= \frac{\nabla^2 p_\theta(x)}{p_\theta(x)} - \frac{\nabla p_\theta(x) \nabla \cdot p_\theta(x)^T}{[p_\theta(x)]^2} \\ &= \frac{\nabla^2 p_\theta(x)}{p_\theta} - \nabla \ln p_\theta(x) \cdot \nabla \ln p_\theta(x)^T. \end{aligned} \tag{A13}$$

It is noted that:

$$\mathbb{E}_\theta \left[\frac{\nabla^2 p_\theta(x)}{p_\theta(x)} \right] = \int \nabla^2 p_\theta(x) dx = \nabla^2 1 = 0. \tag{A14}$$

The following result, that is the negative expected Hessian matrix of the log-function is equal to the Fisher information matrix, can be obtained.

$$\mathbf{I}(\theta) = \mathbb{E}_\theta \left[\nabla \ln p_\theta(x) \cdot \nabla \ln p_\theta(x)^T \right] = -\mathbb{E}_\theta \left[\nabla^2 \ln p_\theta(x) \right], \tag{A15}$$

where $\nabla^2 \ln p_\theta(x)$ is recognized as the Hessian matrix of $\ln p_\theta(x)$, denoted as $\mathbf{H}_{\ln p_\theta(x)}$. The Fisher information matrix can loosely be thought of as the curvature matrix of the log-function graph.

The Kullback–Leibler divergence, also called the relative entropy, between two distributions can be written as:

$$\mathbf{D}(\theta || \theta') = \int p_\theta(x) \ln \left[\frac{p_\theta(x)}{p_{\theta'}(x)} \right] dx = \mathbb{E}_\theta [\ln p_\theta(x) - \ln p_{\theta'}(x)]. \tag{A16}$$

Expand $\ln p_{\theta'}(x)$ around θ to the second order to have,

$$\begin{aligned} \ln p_{\theta'}(x) &\approx \ln p_\theta(x) + (\theta - \theta') \nabla \ln p_\theta(x) \\ &\quad + \frac{1}{2} (\theta - \theta')^T \nabla^2 \ln p_\theta(x) (\theta - \theta'). \end{aligned} \tag{A17}$$

Substitute Equation (A17) into Equation (A16):

$$\mathbf{D}(\theta || \theta') \approx -\mathbb{E}_\theta [\Delta\theta \cdot \nabla \ln p_\theta(x)] - \frac{1}{2} \mathbb{E}_\theta \left[\Delta\theta \cdot \nabla^2 \ln p_\theta(x) \cdot \Delta\theta^T \right], \tag{A18}$$

where $\Delta\theta = (\theta - \theta')$. The first term of the right-hand side of Equation (A18) is zero; and the second term of the right-hand side is related to Equation (A15). The KL divergence can finally be expressed as,

$$\mathbf{D}(\theta || \theta') \approx \frac{1}{2} \Delta\theta^T \cdot \mathbf{I}(\theta) \cdot \Delta\theta. \tag{A19}$$

Furthermore, differentiate Equation (A16) with respect to θ' to obtain:

$$\begin{aligned} \nabla_{\theta'} \mathbf{D}(\theta || \theta') &= \nabla_{\theta'} \mathbb{E}_\theta [\ln p_\theta(x) - \ln p_{\theta'}(x)] \\ &= -\mathbb{E}_\theta \left[\frac{\nabla p_{\theta'}(x)}{p_{\theta'}(x)} \right]. \end{aligned} \tag{A20}$$

Continue to differentiate Equation (A20) with respect to θ' to obtain:

$$\nabla_{\theta'}^2 \mathbf{D}(\theta || \theta') = -\mathbb{E}_\theta \left[\nabla p_{\theta'}(x) \frac{\nabla p_{\theta'}(x)}{p_{\theta'}^2} - \frac{\nabla^2 p_{\theta'}(x)}{p_{\theta'}(x)} \right]. \tag{A21}$$

Notice that the expectation of the second term in Equation (A21), when $\theta' = \theta$, reads,

$$\mathbb{E}_\theta \left[\frac{\nabla^2 p_{\theta'}(x)}{p_{\theta'}(x)} \right] = \nabla^2 \int p_\theta(x) = \nabla^2 1 = 0, \tag{A22}$$

and the first term is the Fisher information matrix. As a result, the Fisher information matrix is the Hessian matrix of the Kullback–Leibler distance evaluated at the true parameter θ .

$$\nabla_{\theta'}^2 \mathbf{D}(\theta || \theta')|_{\theta'=\theta} = \mathbf{I}(\theta). \quad (\text{A23})$$

References

- Royall, R.M. The effect of sample size on the meaning of significance tests. *Am. Stat.* **1986**, *40*, 313–315.
- Raudys, S.J.; Jain, A.K. Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 252–264. [[CrossRef](#)]
- Nelson, C.R.; Kim, M.J. Predictable stock returns: The role of small sample bias. *J. Financ.* **1993**, *48*, 641–661. [[CrossRef](#)]
- Fan, X.; Thompson, B.; Wang, L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Struct. Equ. Model. A Multidiscip. J.* **1999**, *6*, 56–83. [[CrossRef](#)]
- McNeish, D. On using Bayesian methods to address small sample problems. *Struct. Equ. Model. A Multidiscip. J.* **2016**, *23*, 750–773. [[CrossRef](#)]
- Winkler, R.L. Probabilistic prediction: Some experimental results. *J. Am. Stat. Assoc.* **1971**, *66*, 675–685. [[CrossRef](#)]
- Melchers, R.E.; Beck, A.T. *Structural Reliability Analysis and Prediction*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
- He, J.; Chen, J.; Guan, X. Lifetime distribution selection for complete and censored multi-level testing data and its influence on probability of failure estimates. *Struct. Multidiscip. Optim.* **2020**, *62*, 1–17. [[CrossRef](#)]
- Du, Y.M.; Ma, Y.H.; Wei, Y.F.; Guan, X.; Sun, C. Maximum entropy approach to reliability. *Phys. Rev. E* **2020**, *101*, 012106. [[CrossRef](#)]
- Zhou, D.; He, J.; Du, Y.M.; Sun, C.; Guan, X. Probabilistic information fusion with point, moment and interval data in reliability assessment. *Reliab. Eng. Syst. Saf.* **2021**, *213*, 107790. [[CrossRef](#)]
- Gao, C.; Fang, Z.; Lin, J.; Guan, X.; He, J. Model averaging and probability of detection estimation under hierarchical uncertainties for Lamb wave detection. *Mech. Syst. Signal Process.* **2022**, *165*, 108302. [[CrossRef](#)]
- Gregory, P. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support*; Cambridge University Press: Cambridge, UK, 2005.
- Guan, X.; Jha, R.; Liu, Y. Model selection, updating, and averaging for probabilistic fatigue damage prognosis. *Struct. Saf.* **2011**, *33*, 242–249. [[CrossRef](#)]
- Hamdia, K.M.; Zhuang, X.; He, P.; Rabczuk, T. Fracture toughness of polymeric particle nanocomposites: Evaluation of models performance using Bayesian method. *Compos. Sci. Technol.* **2016**, *126*, 122–129. [[CrossRef](#)]
- Yang, J.; He, J.; Guan, X.; Wang, D.; Chen, H.; Zhang, W.; Liu, Y. A probabilistic crack size quantification method using in-situ Lamb wave test and Bayesian updating. *Mech. Syst. Signal Process.* **2016**, *78*, 118–133. [[CrossRef](#)]
- Wang, D.; He, J.; Guan, X.; Yang, J.; Zhang, W. A model assessment method for predicting structural fatigue life using Lamb waves. *Ultrasonics* **2018**, *84*, 319–328. [[CrossRef](#)]
- Guan, X.; He, J. Life time extension of turbine rotating components under risk constraints: A state-of-the-art review and case study. *Int. J. Fatigue* **2019**, *129*, 104799. [[CrossRef](#)]
- He, J.; Huo, H.; Guan, X.; Yang, J. A Lamb wave quantification model for inclined cracks with experimental validation. *Chin. J. Aeronaut.* **2021**, *34*, 601–611. [[CrossRef](#)]
- Huo, H.; He, J.; Guan, X. A Bayesian fusion method for composite damage identification using Lamb wave. *Struct. Health Monit.* **2020**, 1475921720945000. [[CrossRef](#)]
- Berger, J.O. Robust Bayesian analysis: Sensitivity to the prior. *J. Stat. Plan. Inference* **1990**, *25*, 303–328. [[CrossRef](#)]
- Berger, J.O.; Moreno, E.; Pericchi, L.R.; Bayarri, M.J.; Bernardo, J.M.; Cano, J.A.; De la Horra, J.; Martín, J.; Ríos-Insúa, D.; Betrò, B.; et al. An overview of robust Bayesian analysis. *Test* **1994**, *3*, 5–124. [[CrossRef](#)]
- Kass, R.E.; Wasserman, L. The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* **1996**, *91*, 1343–1370. [[CrossRef](#)]
- Bishop, P.; Bloomfield, R.; Littlewood, B.; Povyakalo, A.; Wright, D. Toward a formalism for conservative claims about the dependability of software-based systems. *IEEE Trans. Softw. Eng.* **2010**, *37*, 708–717. [[CrossRef](#)]
- Gelman, A.; Hennig, C. Beyond subjective and objective in statistics. *J. R. Stat. Soc. Ser. A* **2017**, *180*, 967–1033. [[CrossRef](#)]
- Zhang, J.; Shields, M.D. The effect of prior probabilities on quantification and propagation of imprecise probabilities resulting from small datasets. *Comput. Methods Appl. Mech. Eng.* **2018**, *334*, 483–506. [[CrossRef](#)]
- Schervish, M.J.; DeGroot, M.H. *Probability and Statistics*; Pearson Education: London, UK, 2014.
- Jaynes, E.T. Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [[CrossRef](#)]
- Jaynes, E.; Bretthorst, G. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
- Gelman, A.; Simpson, D.; Betancourt, M. The prior can often only be understood in the context of the likelihood. *Entropy* **2017**, *19*, 555. [[CrossRef](#)]
- Jeffreys, H. An Invariant Form for the Prior Probability in Estimation Problems. *Proc. R. Soc. Lond. Ser. A* **1946**, *186*, 453–461.
- Hartigan, J. Invariant prior distributions. *Ann. Math. Stat.* **1964**, *35*, 836–845. [[CrossRef](#)]
- Bernardo, J.M. Reference posterior distributions for Bayesian inference. *J. R. Stat. Soc. Ser. B* **1979**, *41*, 113–147. [[CrossRef](#)]

33. Berger, J.O.; Bernardo, J.M. Ordered group reference priors with application to the multinomial problem. *Biometrika* **1992**, *79*, 25–37. [[CrossRef](#)]
34. Jeffreys, H. *The Theory of Probability*; Oxford University Press: Oxford, UK, 1998.
35. Gelman, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **2006**, *1*, 515–534. [[CrossRef](#)]
36. Simpson, D.; Rue, H.; Riebler, A.; Martins, T.G.; Sørbye, S.H. Penalising model component complexity: A principled, practical approach to constructing priors. *Stat. Sci.* **2017**, *32*, 1–28. [[CrossRef](#)]
37. Kullback, S.; Leibler, R. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [[CrossRef](#)]
38. Guan, X.; He, J.; Jha, R.; Liu, Y. An efficient analytical Bayesian method for reliability and system response updating based on Laplace and inverse first-order reliability computations. *Reliab. Eng. Syst. Saf.* **2012**, *97*, 1–13. [[CrossRef](#)]
39. He, J.; Guan, X.; Jha, R. Improve the accuracy of asymptotic approximation in reliability problems involving multimodal distributions. *IEEE Trans. Reliab.* **2016**, *65*, 1724–1736. [[CrossRef](#)]
40. American Society for Testing and Materials. *ASTM E739-10(2015)-Standard Practice for Statistical Analysis of Linear or Linearized Stress-Life (S-N) and Strain-Life (ϵ -N) Fatigue Data*; ASTM International: West Conshohocken, PA, USA, 2015.
41. Boyd-Lee, A.; Harrison, G.; Henderson, M. Evaluation of standard life assessment procedures and life extension methodologies for fracture-critical components. *Int. J. Fatigue* **2001**, *23*, 11–19. [[CrossRef](#)]
42. Browne, W.J.; Draper, D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Anal.* **2006**, *1*, 473–514. [[CrossRef](#)]
43. Kotz, S.; Nadarajah, S. *Multivariate T-Distributions and Their Applications*; Cambridge University Press: Cambridge, UK, 2004.
44. Graybill, F.A.; Bowden, D.C. Linear segment confidence bands for simple linear models. *J. Am. Stat. Assoc.* **1967**, *62*, 403–408. [[CrossRef](#)]