




The Bayesian Inference of Pareto Models Based on Information Geometry

Fupeng Sun, Yueqi Cao , Shiqiang Zhang  and Huafei Sun * 

School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China; fupeng_sun@bit.edu.cn (F.S.); yueqi_cao@bit.edu.cn (Y.C.); shiqiang@bit.edu.cn (S.Z.)

* Correspondence: huafeisun@bit.edu.cn

Abstract: Bayesian methods have been rapidly developed due to the important role of explicable causality in practical problems. We develop geometric approaches to Bayesian inference of Pareto models, and give an application to the analysis of sea clutter. For Pareto two-parameter model, we show the non-existence of α -parallel prior in general, hence we adopt Jeffreys prior to deal with the Bayesian inference. Considering geodesic distance as the loss function, an estimation in the sense of minimal mean geodesic distance is obtained. Meanwhile, by involving Al-Bayyati's loss function we gain a new class of Bayesian estimations. In the simulation, for sea clutter, we adopt Pareto model to acquire various types of parameter estimations and the posterior prediction results. Simulation results show the advantages of the Bayesian estimations proposed and the posterior prediction.

Keywords: Bayesian inference; Pareto two-parameter model; Jeffreys prior; mean geodesic estimation; Al-Bayyati's loss function

MSC: 62C10; 62F15



Citation: Sun, F.; Cao, Y.; Zhang, S.; Sun, H. The Bayesian Inference of Pareto Models Based on Information Geometry. *Entropy* **2021**, *23*, 45. <https://doi.org/10.3390/e23010045>

Received: 1 December 2020

Accepted: 26 December 2020

Published: 30 December 2020

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



Copyright: © 2020 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

Geometric method plays an important role in Bayesian statistics. At present, there are two main ways to study Bayesian inference through geometric methods. An idea is to regard the prior distribution, the probability distribution of the statistical model and the posterior distribution as the vectors in Hilbert space $L^2(\Theta)$. The research is then carried out through the geometric properties of Hilbert space. M. de Carvalho [1] used the cosine of the angle between vectors to study their relationship with each other, where the cosine of priors represents coherency of opinions of experts, the cosine of prior and probability density represents prior-data agreement and the cosine of prior and posterior represents sensitivity of the posterior to the prior specification. Furthermore, M. de Carvalho used Monte Carlo Markov Chain to give an estimation of the cosine value for further analysis. R. Kulhavy [2] viewed statistical inference as an approximation of the empirical density rather than an estimation of a true density, and built a model by analyzing the trace of orthogonal projection of conditional empirical distributions onto the model manifold. He also used Kerridge inaccuracy as a generalized empirical error. Kerridge inaccuracy is a generalization of Shannon entropy. It is used to measure the difference between observed distribution $Q = (q_1, \dots, q_n)$ and true distribution $P = (p_1, \dots, p_n)$, which is defined by $I(P, Q) = -\sum_i p_i \log(q_i)$. The advantage of this idea is providing a unified treatment to all pieces of Bayesian theorem. However, the finite parameter space measure is required.

Another idea is to give the statistical manifolds Riemannian metrics. J.A. Hartigan [3–5] proposed a reparametrization invariance prior- α -parallel prior and later J. Takeuchi and S. Amari [6,7] clarified an interesting connection between the information geometrical properties of the statistical model and the existence of the α -parallel prior. α -parallel prior, as an uninformed prior, is invariant under the coordinate transformation and can well reflect the intrinsic properties of the model. It is worth noting that the general α -parallel

prior does not always exist, but 0-parallel prior, Jeffreys prior, always exists. After obtaining the corresponding prior, subsequent Bayesian inference and prediction can be carried out. J. Takeuchi and S. Amari studied the asymptotic properties of minimum description length (MDL) estimation and projected Bayesian estimation of general exponential families, and extended it to curve exponential families. The differential geometry of curved exponential families are given by S. Amari and M. Kumon [8,9].

In many confidential fields, historical data is difficult to obtain. Hence one of the advantages of the above idea is giving prior a good theoretical basis. Besides, this idea can illustrate the geometric meaning of the common Bayesian estimations and provide an estimation in the sense of minimal mean geodesic distance. In the application of target detection, maritime radar performance is often seriously interfered and suppressed by sea clutter, especially for the detection of weak targets on the sea. Therefore, whether sea clutter can be effectively suppressed is the key factor to improve the performance of maritime radar systems. Thus the study of sea clutter is of vital importance [10]. With the development of radar hardware technology, the statistical distribution histogram of radar sea clutter appears long “trailing tail”, which is manifested as frequent sea peak phenomenon, and the amplitude distribution of clutter seriously deviates from the Rayleigh distribution proposed before. In order to solve this problem, an improved Rayleigh model-compound Gaussian model is proposed [11]. The compound Gaussian model successfully illustrates the formation mechanism of sea clutter, and is more successful than the single point probability distribution model in terms of data fitting. In the compound Gaussian model, K distribution and Pareto distribution are two typical representatives. When the structural component is Gamma distribution, the compound Gaussian model is K distribution. When the structural component is inverse Gamma distribution, the compound Gaussian model is Pareto distribution. In 2010, M. Farshchian and F.L. Posner [12] analyzed the sea clutter data in X-band, and used Pareto distribution to fit data. They found that the fitting effect of the tail was better than K distribution.

The paper is organized as follows—Section 2 introduces the preliminary including some geometric structure of statistical manifolds and some concepts of Bayesian inference. In Section 3, we introduce the geometric approaches for Bayesian inference by using α -parallel connection, and propose a geometric loss function based on geodesic distance. In Section 4, we prove that Pareto two-parameter model does not have general α -parallel prior. Then we adopt Jeffreys prior to provide the explicit expressions of estimations in the sense of minimal mean geodesic distance. Furthermore, we come up with theorems under Al-Bayyati’s loss function to obtain a new class of Bayesian estimations. The bounds of certain expressions without closed forms are given. Besides, we show the existence of the best parameter in Al-Bayyati’s loss function. In Section 5, we show the advantages of the proposed Bayesian estimations and the posterior prediction distributions.

2. Preliminaries

2.1. α -Parallel Prior

Definition 1 ([7,13]). For a statistical manifold $M = \{p(x;\theta) \mid \int p(x;\theta)dx = 1, p(x;\theta) > 0, \theta \in \Theta \subset \mathbb{R}^d\}$, define an affine connection $\nabla^{(\alpha)}$ on M with the following coefficients

$$\Gamma_{jk}^{(\alpha)i} := \Gamma_{jk}^{(0)i} + \frac{1-\alpha}{2} T_{sjk} g^{is}, \quad \Gamma_{jk}^{(0)i} := \Gamma_{s;jk}^{(0)} g^{is},$$

where $\alpha \in \mathbb{R}$ is an arbitrary real number, $g_{ij} = E[\partial_i l \partial_j l]$, (g^{ij}) is the inverse matrix (g_{ij}) , $\Gamma_{s;jk}^{(0)} = E[\partial_s l \partial_j \partial_k l]$, $T_{sjk} = E[\partial_s l \partial_j l \partial_k l]$, $l := \log p(x;\theta)$ denotes the log likelihood function and $E[\cdot]$ denotes the expectation with respect to the observation x .

Definition 2 ([7]). An affine connection ∇ is called locally equiaffine if around each point x of M , there is a parallel volume element, that is, a nonvanishing d -form w such that $\nabla w = 0$.

An equiaffine connection ∇ on M is a torsion-free affine connection with a parallel volume element w on M .

If w is a volume element on M such that $\nabla w = 0$, then we say that (∇, w) is an affine structure on M .

For a statistical manifold M , we may represent the α -parallel volume element w as

$$w = \pi(\theta)d\theta^1 \wedge \dots \wedge d\theta^d$$

for a certain coordinate $\theta = (\theta^1, \dots, \theta^d) \in \Theta \subset \mathbb{R}^d$, where π is the volume form on the whole manifold. We take $\pi(\theta)$ as a prior distribution on the parameter space Θ .

Definition 3 ([7]). In a statistically equiaffine manifold, for a fixed $\alpha \in \mathbb{R}$, we call the above form of π an α -parallel prior.

When $\alpha = 1$, 1-parallel prior is called maximum likelihood estimation (MLE) prior proposed by J.A. Hartigan [5]. Note that there always exists a $\nabla^{(0)}$ -parallel volume element $w \propto \sqrt{g(\theta)}d\theta^1 \wedge \dots \wedge d\theta^d$, where g is the determinant of the Fisher metric, the invariant volume element in a Riemannian manifold $(M, (g_{ij}))$. This prior distribution $\pi \propto \sqrt{g(\theta)}$ is called the Jeffreys prior.

J. Takeuchi and S. Amari gave a sufficient and necessary condition for the existence of α -parallel prior.

Proposition 1 ([6]). For a statistical manifold M with the α connection $\nabla^{(\alpha)}$, if $\alpha \neq 0$, then there exists an α -parallel prior if and only if

$$\partial_i T_j - \partial_j T_i = 0 \tag{1}$$

where $T_i := T_{ikl}g^{kl}$.

2.2. Bayesian Inference

For the random variate \mathbf{x} subject to the distribution $p(\mathbf{x}; \theta)$, and let $\pi(\theta)$ be the prior distribution of θ . The posterior distribution $\pi(\theta|\mathbf{x})$ is given by the formula

$$\pi(\theta|\mathbf{x}) = \frac{p(\mathbf{x}; \theta)\pi(\theta)}{\int_{\Theta} p(\mathbf{x}; \theta)\pi(\theta) d\theta}.$$

Now, we introduce some notations for later uses. Let $\hat{\theta}_{MD}$ be the maximum posterior estimation. Let $\hat{\theta}_{Me}$ be the posterior median estimation which is the median of the posterior distribution. Let $\hat{\theta}_E$ be the posterior expectation estimation which is the expectation of the posterior distribution.

These three estimations are also known as Bayesian estimations of θ . When $\hat{\theta} = \hat{\theta}_E$, the posterior mean square value reaches the minimum. Hence $\hat{\theta}_E = E[\theta|x]$ is often taken as the Bayesian estimation.

Let the random variable $X \sim p(x; \theta)$. If one does not know the observation data, the marginal distribution $m(x)$ is also known as the prior prediction distribution. If one obtains the observation data $\mathbf{x} = (x_1, \dots, x_n)$, the distribution of unknown observation values could be obtained by the posterior distribution $\pi(\theta|\mathbf{x})$:

1. Predict the future observations of the same population $p(\tilde{x}; \theta)$

$$m(\tilde{x}|\mathbf{x}) = \int_{\Theta} p(\tilde{x}; \theta)\pi(\theta|\mathbf{x}) d\theta.$$

2. Predict the observations of another population $g(z; \theta)$

$$m(z|\mathbf{x}) = \int_{\Theta} g(z; \theta) \pi(\theta|\mathbf{x}) d\theta,$$

where $m(\tilde{x}|\mathbf{x})$ or $m(z|\mathbf{x})$ is called the posterior predictive distribution.

3. The Geometric Approaches for Bayesian Inference

In this section, we introduce the basic methods of Bayesian inference with geometric means. The idea of geometry is embodied in the selection of priors and loss functions.

3.1. The Geometric Prior

The idea of geometric methods is to extend the uniform distribution naturally and construct geometric priors suitable for multidimensional and measure infinite-dimensional parameter space according to the idea that probability measure is proportional to volume element. The studied probability distribution family can be regarded as a statistical manifold with Riemannian metrics.

Fisher information matrix is the most widely used Riemannian metric on statistical manifolds, and the prior generated by its corresponding volume element is Jeffreys prior. α -connection is a natural extension of the Levi-Civita connection corresponding to Fisher information matrix. Its corresponding volume element is α -parallel volume element, and the generated prior is called α -parallel prior. In particular, the 0-parallel prior is the Jeffreys prior.

α -parallel prior reflects the intrinsic property of the model and does not depend on the selection of parameters. Although Jeffreys prior must exist, general α -parallel prior does not necessarily exist. (1) gives the necessary and sufficient conditions for the existence of general α -parallel prior.

Therefore, when one deals with Bayesian inference by geometric methods, the first step is to select the appropriate geometric priors, that is, to verify the existence of α -parallel prior in a specific statistical manifold.

With Riemannian metric, we can acquire geometric information of the statistical manifold, such as connection, curvature, geodesic and geodesic distance [14,15]. Through geometric priors, the joint posterior density of the parameters can be obtained, and then the corresponding Bayesian estimation and Bayesian posterior prediction are carried out [16].

3.2. The Geometric Loss Functions

In this subsection, we show the geometric meaning of the common Bayesian estimations and propose a new geometric approach of choosing loss functions.

Proposition 2. Suppose that $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$. Let $\pi(\theta|x)$ be the joint posterior distribution and $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d)$ be the estimation of θ . For the loss function $l_1(\theta, \hat{\theta}) = \sum_{i=1}^d |\hat{\theta}_i - \theta_i|$, the corresponding estimation is $\hat{\theta}_{Me} = (\hat{\theta}_{1Me}, \dots, \hat{\theta}_{dMe})$. For the loss function $l_2(\theta, \hat{\theta}) = \sum_{i=1}^d (\hat{\theta}_i - \theta_i)^2$, the corresponding estimation is $\hat{\theta}_E = (\hat{\theta}_{1E}, \dots, \hat{\theta}_{dE})$.

Proof. Denote $d\tilde{\theta}_i = d\theta_1 \cdots d\theta_{i-1} d\theta_{i+1} \cdots d\theta_d$. Let $\pi(\theta_i|x)$ be the marginal posterior density and $\Pr(\theta_i \leq t|x)$ be its cumulative distribution. Assume that $\Pr(\theta_i \leq a|x) = 0$ and $\Pr(\theta_i \leq b|x) = 1$, where $a, b \in \mathbb{R}$ and they may be infinite.

For the loss function $l_1(\theta, \hat{\theta})$, we define

$$\begin{aligned}
 R_1(\theta, \hat{\theta}) &= \int_{\Theta} l_1(\theta, \hat{\theta}) \pi(\theta|x) d\theta \\
 &= \left(\int_a^{\hat{\theta}_i} - \int_{\hat{\theta}_i}^b \right) \int_{\hat{\theta}_i} (\hat{\theta}_i - \theta_i) \pi(\theta|x) d\theta_i d\theta + \sum_{j \neq i} \int_{\Theta} |\hat{\theta}_j - \theta_j| \pi(\theta|x) d\theta \\
 &= \hat{\theta}_i \left(\int_a^{\hat{\theta}_i} - \int_{\hat{\theta}_i}^b \right) \pi(\theta_i|x) d\theta_i - \left(\int_a^{\hat{\theta}_i} - \int_{\hat{\theta}_i}^b \right) \theta_i \pi(\theta_i|x) d\theta_i + \sum_{j \neq i} \int_{\Theta} |\hat{\theta}_j - \theta_j| \pi(\theta|x) d\theta.
 \end{aligned}$$

Then we get

$$\frac{\partial R_1}{\partial \hat{\theta}_i} = \int_a^{\hat{\theta}_i} \pi(\theta_i|x) d\theta_i - \int_{\hat{\theta}_i}^b \pi(\theta_i|x) d\theta_i.$$

Let $\frac{\partial R_1}{\partial \hat{\theta}_i} = 0$, we have $\hat{\theta}_i = \hat{\theta}_{iMe}$.

For the loss function $l_2(\theta, \hat{\theta})$, we define

$$R_2(\theta, \hat{\theta}) = \int_{\Theta} l_2(\theta, \hat{\theta}) \pi(\theta|x) d\theta = \sum_{i=1}^d \int_{\Theta} (\hat{\theta}_i - \theta_i)^2 \pi(\theta|x) d\theta.$$

Then we obtain

$$\frac{\partial R_2}{\partial \hat{\theta}_i} = 2 \int_a^b (\hat{\theta}_i - \theta_i) \pi(\theta_i|x) d\theta_i$$

Hence, $\frac{\partial R_2}{\partial \hat{\theta}_i} = 0$ implies that $\hat{\theta}_i = \hat{\theta}_{iE}$. \square

If the loss function is the distance induced by $\| \cdot \|_1$, then by Proposition 2 we see that the corresponding risk function represents the average distance between the estimated value and the true value. Besides, the posterior median estimation of parameters minimizes risk function, which means that this estimation has the minimum mean distance from the posterior density.

If the loss function is the distance induced by $\| \cdot \|_2$, then the corresponding risk function represents the average value of the square of the distance between the estimated value and the true value. The obtained estimation is the posterior expectation of parameters, which has the minimum mean square error from the posterior density.

These two kinds of loss functions above are distances or increasing functions of distances in \mathbb{R}^n . However, in the parameter space endowed with corresponding Riemannian metric, the distance between two points is geodesic distance instead of Euclidean distance.

Hence, in order to make the estimation more accurate, we take the geodesic distance or its increasing function as a loss function, the corresponding risk function represents the geodesic distance between the estimated value and the true value. Before that, we need the following definition.

Definition 4. (Mean Geodesic Estimation) Assume that the statistical manifold $M = \{p(x; \theta)\}$ is equipped with Fisher Riemannian metric (g_{ij}) . Let $\pi(\theta|x)$ be the joint posterior distribution and $d(\theta, \hat{\theta})$ be the geodesic distance between θ and $\hat{\theta}$, where $\hat{\theta}$ is the estimation of θ . Let $F: \mathbb{R} \rightarrow \mathbb{R}$ be an increasing function. Denote $D(\theta, \hat{\theta}) = F \circ d(\theta, \hat{\theta})$. The risk function with the loss function $D(\theta, \hat{\theta})$ is

$$R(\theta, \hat{\theta}) = \int_{\Theta} D(\theta, \hat{\theta}) \pi(\theta|x) d\theta. \tag{2}$$

The estimation minimizing $R(\theta, \hat{\theta})$ is called mean geodesic estimation (MGE) and denoted by $\hat{\theta}_{MGE}$.

The geometric priors, the corresponding geodesic distance and the corresponding Bayesian inference depend on the choice of the Riemannian metric. Hence choosing a proper Riemannian metric is of great importance to Bayesian inference.

4. Bayesian Inference on Pareto Model

4.1. The Geometric Structure of Pareto Two-Parameter Model

The probability density function of Pareto two-parameter distribution satisfies

$$p(x; \alpha, \beta) = \frac{\beta \alpha^\beta}{x^{\beta+1}} I_{[x \geq \alpha]}, \quad \alpha > 0, \beta > 0, \tag{3}$$

where α is called the scale parameter and β is called the shape parameter.

Its logarithmic likelihood function is

$$l(x; \alpha, \beta) = \log p(x; \alpha, \beta) = \log \beta + \beta \log \alpha - (\beta + 1) \log x \tag{4}$$

Noting that the Pareto distribution family does not meet the common regularity condition, hence the Fisher-Rao metric on the Pareto distribution family is not equal to the negative Hessian matrix.

Furthermore, from References [17,18] we can get the geometric structure of Pareto model. On Pareto two-parameter distribution family, the tensor form of Fisher-Rao metric satisfies

$$g = \frac{\beta^2}{\alpha^2} d\alpha \otimes d\alpha + \frac{1}{\beta^2} d\beta \otimes d\beta, \tag{5}$$

which is isometric to the upper half of the Poincaré plane. Hence, Pareto two-parameter model (P, g) is a Riemannian manifold endowed with Riemannian metric g . The volume form, the connection form, the curvature form, the Christoffel symbols and the geodesic distance formula on (P, g) are given as follows

$$dv = \theta^1 \wedge \theta^2 = \frac{1}{\alpha} d\alpha \wedge d\alpha \tag{6}$$

$$w_2^1 = \frac{\beta}{\alpha} d\alpha \tag{7}$$

$$\Omega_2^1 = dw_2^1 = -\frac{1}{\alpha} d\alpha \wedge d\alpha = -Kdv \tag{8}$$

$$\begin{cases} \nabla_{\partial_1} \partial_1 = -\frac{1}{\alpha} \partial_1 - \frac{\beta^3}{\alpha^2} \partial_2 \\ \nabla_{\partial_2} \partial_1 = \nabla_{\partial_1} \partial_2 = \frac{1}{\beta} \partial_1 \\ \nabla_{\partial_2} \partial_2 = -\frac{1}{\beta} \partial_2 \end{cases} \tag{9}$$

$$d((\alpha_0, \beta_0), (\alpha_1, \beta_1)) = \operatorname{arcosh} \left(1 + \frac{\beta_0 \beta_1 (\log \alpha_0 - \log \alpha_1)^2}{2} + \frac{(\beta_0 - \beta_1)^2}{2\beta_0 \beta_1} \right). \tag{10}$$

4.2. The Existence of α -Parallel Prior on Pareto Two-Parameter Model

Theorem 1. When $\alpha \neq 0$, Pareto two-parameter model does not have any α -parallel prior.

Proof. Denote $\partial_1 = \frac{\partial}{\partial \alpha}$, $\partial_2 = \frac{\partial}{\partial \beta}$. Then by calculation, we can obtain

$$\begin{cases} T_{111} = E[\partial_1 l \partial_1 l \partial_1 l] = \frac{\beta^3}{\alpha^3} \\ T_{222} = E[\partial_2 l \partial_2 l \partial_2 l] = -\frac{2}{\beta^3} \\ T_{112} = T_{121} = T_{211} = E[\partial_1 l \partial_1 l \partial_2 l] = 0 \\ T_{122} = T_{221} = T_{212} = E[\partial_1 l \partial_2 l \partial_2 l] = \frac{1}{\alpha \beta} \end{cases} \tag{11}$$

Hence, we get

$$\begin{cases} T_1 = T_{1ik} g^{ik} = \frac{2\beta}{\alpha} \\ T_2 = T_{2ik} g^{ik} = -\frac{2}{\beta} \end{cases}$$

and

$$\partial_i T_j - \partial_j T_i = \pm \frac{2}{\alpha} d\beta \wedge d\alpha.$$

It is obvious that $\alpha \neq 0$ means $\partial_i T_j - \partial_j T_i \neq 0$. Therefore, according to Proposition 1, we find that Pareto two-parameter model does not have any α -parallel prior when $\alpha \neq 0$. \square

From Theorem 1, we see that Pareto two-parameter model only has the 0-parallel prior (Jeffreys prior). Its Jeffreys prior $\pi(\alpha, \beta)$ satisfies $\pi(\alpha, \beta) \propto \sqrt{\frac{\alpha^2}{\beta^2} \frac{1}{\beta^2}} = \frac{1}{\alpha}, \alpha > 0, \beta > 0$, which is a generalized prior density whose integral is infinite. We assume that $\pi(\alpha, \beta) = \frac{1}{\alpha}, \alpha > 0, \beta > 0$.

4.3. Bayesian Estimations of Pareto Model

Before we proceed, we state necessary results from Reference [17].

The joint probability density of a simple random sample on Pareto model is

$$p(x; \alpha, \beta) = \beta^n \alpha^{n\beta} \left(\prod_{i=1}^n x_i \right)^{-\beta-1} I_{[\min_{i=1}^n x_i \geq \alpha]}. \tag{12}$$

The posterior distribution of Pareto model under Jeffreys prior is obtained by Bayesian formula

$$\pi(\alpha, \beta | x) = \frac{n(q_2(x) - n \log q_1(x))^n}{\tau(n)} \beta^n \alpha^{n\beta-1} \exp(-q_2(x)\beta) I_{[0 \leq \alpha \leq q_1(x)]}, \tag{13}$$

where $q_1(x) = \min_{i=1}^n x_i$, $q_2(x) = \sum_{i=1}^n \log x_i$. Furthermore, by calculation we can see that the maximum likelihood estimation and the maximum posterior estimation of α, β are given as

$$\begin{cases} \hat{\alpha}_{MLE} = \hat{\alpha}_{MD} = q_1(x) \\ \hat{\beta}_{MLE} = \hat{\beta}_{MD} = \frac{n}{\sum_{i=1}^n \log x_i - n \log(\min_{i=1}^n x_i)} = \frac{n}{q_2(x) - n \log q_1(x)}. \end{cases} \tag{14}$$

The marginal posterior density of α determined by the joint posterior density $\pi(\alpha, \beta | x)$ is

$$\pi(\alpha | x) = \frac{n^2 (q_2(x) - n \log q_1(x))^n}{\alpha (q_2(x) - n \log \alpha)^{n+1}} I_{[0 \leq \alpha \leq q_1(x)]} \tag{15}$$

and its cumulative distribution function is

$$\Pr(\alpha \leq t|x) = \left(1 + \hat{\beta}_{MLE} \log \frac{\hat{\alpha}_{MLE}}{t}\right)^{-n}, \quad 0 \leq t \leq q_1(x). \tag{16}$$

The marginal posterior density of β determined by the joint posterior density $\pi(\alpha, \beta|x)$ is

$$\pi(\beta|x) = \frac{\beta^{n-1}(q_2(x) - n \log q_1(x))^n}{\tau(n)} \exp(-(q_2(x) - n \log q_1(x))\beta). \tag{17}$$

Under posterior distribution (13), when β is known, the conditional posterior density of α is

$$\pi(\alpha|x, \beta) = \frac{n\beta\alpha^{n\beta-1}}{q_1^{n\beta}(x)} I_{[0 \leq \alpha \leq q_1(x)]} \tag{18}$$

and its cumulative distribution function is

$$\Pr(\alpha \leq t|x, \beta) = \left(\frac{t}{q_1(x)}\right)^{n\beta}, \quad 0 \leq t \leq q_1(x). \tag{19}$$

When α is known, the conditional posterior density of β is

$$\pi(\beta|x, \alpha) = \frac{\beta^n(q_2(x) - n \log \alpha)^{n+1}}{\tau(n+1)} \exp(-(q_2(x) - n \log \alpha)\beta). \tag{20}$$

4.3.1. Mean Geodesic Estimation

The geodesic distance between (α, β) and $(\hat{\alpha}, \hat{\beta})$ is expressed as

$$d((\alpha, \beta), (\hat{\alpha}, \hat{\beta})) = \operatorname{arcosh}\left(1 + \frac{\beta\hat{\beta}(\log \alpha - \log \hat{\alpha})^2}{2} + \frac{(\beta - \hat{\beta})^2}{2\beta\hat{\beta}}\right). \tag{21}$$

Hence, the distance function is a monotone function of

$$1 + \frac{\beta\hat{\beta}(\log \alpha - \log \hat{\alpha})^2}{2} + \frac{(\beta - \hat{\beta})^2}{2\beta\hat{\beta}}. \tag{22}$$

We denote (22) as

$$D((\alpha, \beta), (\hat{\alpha}, \hat{\beta})) = 1 + \frac{\beta\hat{\beta}(\log \alpha - \log \hat{\alpha})^2}{2} + \frac{(\beta - \hat{\beta})^2}{2\beta\hat{\beta}}. \tag{23}$$

By the discussion in the Section 3, when α and β are unknown, $D((\alpha, \beta), (\hat{\alpha}, \hat{\beta}))$ can be taken as the loss function, and the estimations $\hat{\alpha}_{MGE}$ and $\hat{\beta}_{MGE}$ in the sense of minimum mean geodesic distance can be obtained. When $\beta = \beta_0$ or $\alpha = \alpha_0$, $D_{\beta_0}((\alpha, \beta_0), (\hat{\alpha}, \beta_0))$ or $D_{\alpha_0}((\alpha_0, \beta), (\alpha_0, \hat{\beta}))$ can be taken as the loss function respectively, and we can get the mean geodesic estimation $\hat{\alpha}_{MGE}(x|\beta)$ or $\hat{\beta}_{MGE}(x|\alpha)$.

Theorem 2. When α and β are unknown, we have

$$\hat{\alpha}_{MGE} = \frac{\hat{\alpha}_{MLE}}{\exp\left(\frac{1}{n\hat{\beta}_{MLE}}\right)}, \quad \hat{\beta}_{MGE} = \sqrt{\frac{2n^2(n-1)}{2n^3+n+1}} \hat{\beta}_{MLE}.$$

Proof. Denote $f(\alpha) = E_{\beta}[\pi(\alpha, \beta|x)]$, then have

$$f(\alpha) = \frac{n^2(n+1)(q_2(x) - n \log q_1(x))^n}{\alpha(q(x) - n \log \alpha)^{n+2}}, \quad 0 \leq \alpha \leq q_1(x).$$

Let $F(t) = \int_0^t f(\alpha) d\alpha$, then we have

$$F(t) = \frac{n(q_2(x) - n \log q_1(x))^n}{(q(x) - n \log t)^{n+1}}, \quad 0 \leq t \leq q_1(x).$$

Denote $g(t) = \frac{F(t)}{t}$, then

$$G(s) = \int_0^s g(t) dt = \frac{(q_2(x) - n \log q_1(x))^n}{n(q(x) - n \log s)^n}, \quad 0 \leq s \leq q_1(x).$$

Denote $h(t) = \frac{G(t)}{t}$, then have

$$H(v) = \int_0^v h(t) dt = \frac{(q_2(x) - n \log q_1(x))^n}{n^2(n-1)(q(x) - n \log v)^{n-1}}, \quad 0 \leq v \leq q_1(x).$$

The risk function is

$$R((\alpha, \beta), (\hat{\alpha}, \hat{\beta})) = \int_0^{\hat{\alpha}_{MLE}} \int_0^{+\infty} D((\alpha, \beta), (\hat{\alpha}, \hat{\beta})) \pi(\alpha, \beta|x) d\beta d\alpha.$$

Firstly, we show that $\hat{\alpha}_{MGE} = \frac{\hat{\alpha}_{MLE}}{\exp\left(\frac{1}{n\hat{\beta}_{MLE}}\right)}$. Since

$$\frac{\partial R}{\partial \hat{\alpha}} = \frac{\hat{\beta}}{\hat{\alpha}} \int_0^{\hat{\alpha}_{MLE}} \int_0^{+\infty} (\log \alpha - \log \hat{\alpha}) \beta \pi(\alpha, \beta|x) d\beta d\alpha,$$

when $\frac{\partial R}{\partial \hat{\alpha}} = 0$, we have

$$\log \hat{\alpha}_{MGE} = \frac{\int_0^{\hat{\alpha}_{MLE}} \log \alpha \cdot E_{\beta}[\pi(\alpha, \beta|x)] d\alpha}{\int_0^{\hat{\alpha}_{MLE}} E_{\beta}[\pi(\alpha, \beta|x)] d\alpha}.$$

Combining

$$\int_0^{\hat{\alpha}_{MLE}} E_{\beta}[\pi(\alpha, \beta|x)] d\alpha = F(\hat{\alpha}_{MLE}) = \hat{\beta}_{MLE}$$

with

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} \log \alpha \cdot E_{\beta}[\pi(\alpha, \beta|x)] d\alpha &= \int_0^{\hat{\alpha}_{MLE}} \log t dF(t) \\ &= \hat{\beta}_{MLE} \log \hat{\alpha}_{MLE} - G(\hat{\alpha}_{MLE}) \\ &= \hat{\beta}_{MLE} \left(\log \hat{\alpha}_{MLE} - \frac{1}{n\hat{\beta}_{MLE}} \right) \end{aligned}$$

we get

$$\hat{\alpha}_{MGE} = \frac{\hat{\alpha}_{MLE}}{\exp\left(\frac{1}{n\hat{\beta}_{MLE}}\right)}.$$

Secondly, we will show that $\hat{\beta}_{MGE} = \sqrt{\frac{2n^2(n-1)}{2n^3+n+1}} \hat{\beta}_{MLE}$. In fact, from

$$\frac{\partial R}{\partial \hat{\beta}} = \int_0^{\hat{\alpha}_{MLE}} \int_0^{+\infty} \left(\frac{\beta(\log \alpha - \log \hat{\alpha})^2}{2} - \frac{\beta}{\hat{\beta}^2} + \frac{1}{\beta} \right) \pi(\alpha, \beta|x) d\beta d\alpha$$

we see that when $\frac{\partial R}{\partial \hat{\beta}} = 0$, we have

$$\frac{\hat{\beta}_{MLE}}{\hat{\beta}_{MGE}^2} = \frac{F(\hat{\alpha}_{MLE})}{\hat{\beta}_{MGE}^2} = \int_0^{\hat{\alpha}_{MLE}} \int_0^{+\infty} \frac{1}{\beta} \pi(\alpha, \beta|x) d\beta d\alpha + \frac{1}{2} \int_0^{\hat{\alpha}_{MLE}} (\log \alpha - \log \hat{\alpha})^2 f(\alpha) d\alpha.$$

Noting that

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} \int_0^{+\infty} \frac{1}{\beta} \pi(\alpha, \beta|x) d\beta d\alpha &= \int_0^{+\infty} \frac{1}{\beta} \pi(\beta|x) d\beta \\ &= \frac{(q_2(x) - n \log q_1(x))^n}{\tau(n)} \int_0^{+\infty} \beta^{n-2} \exp(-(q_2(x) - n \log q_1(x))\beta) d\beta \\ &= \frac{q_2(x) - n \log q_1(x)}{n-1} \\ &= \frac{n}{(n-1)\hat{\beta}_{MLE}} \end{aligned}$$

and

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} (\log \alpha - \log \hat{\alpha})^2 f(\alpha) d\alpha &= \int_0^{\hat{\alpha}_{MLE}} (\log t - \log \hat{\alpha})^2 dF(t) \\ &= \hat{\beta}_{MLE} (\log \hat{\alpha}_{MLE} - \log \hat{\alpha})^2 - 2 \int_0^{\hat{\alpha}_{MLE}} (\log t - \log \hat{\alpha}) g(t) dt \\ &= \hat{\beta}_{MLE} \frac{1}{(n\hat{\beta}_{MLE})^2} - 2G(\hat{\alpha}_{MLE}) \frac{1}{n\hat{\beta}_{MLE}} + 2H(\hat{\alpha}_{MLE}) \\ &= -\frac{1}{n^2\hat{\beta}_{MLE}} + \frac{2}{n(n-1)\hat{\beta}_{MLE}} \\ &= \frac{n+1}{n^2(n-1)\hat{\beta}_{MLE}} \end{aligned}$$

we have

$$\frac{\hat{\beta}_{MLE}}{\hat{\beta}_{MGE}^2} = \frac{n}{(n-1)\hat{\beta}_{MLE}} + \frac{n+1}{2n^2(n-1)\hat{\beta}_{MLE}} = \frac{2n^3+n+1}{2n^2(n-1)\hat{\beta}_{MLE}}.$$

As a result, we obtain $\hat{\beta}_{MGE} = \sqrt{\frac{2n^2(n-1)}{2n^3+n+1}} \hat{\beta}_{MLE}$. \square

Theorem 3. When β is known, we have $\hat{\alpha}_{MGE}(x|\beta) = \frac{\hat{\alpha}_{MLE}}{\exp(\frac{1}{n\beta})}$. And when α is known, we have

$$\hat{\beta}_{MGE}(x|\alpha) = \frac{\sqrt{n(n+1)}}{q_2(x) - n \log \alpha}.$$

Proof. When β is known, the risk function is

$$R((\alpha, \beta), (\hat{\alpha}, \beta)) = \int_0^{\hat{\alpha}_{MLE}} D_\beta((\alpha, \beta), (\hat{\alpha}, \beta)) \pi(\alpha|x, \beta) d\alpha.$$

Since

$$\frac{dR}{d\hat{\alpha}} = \frac{\beta^2}{\hat{\alpha}} \int_0^{\hat{\alpha}_{MLE}} (\log \alpha - \log \hat{\alpha}) \pi(\alpha|x, \beta) d\alpha$$

when $\frac{dR}{d\hat{\alpha}} = 0$, we have

$$\begin{aligned} \log \hat{\alpha}_{MGE}(x|\beta) &= \int_0^{\hat{\alpha}_{MLE}} \log \alpha \cdot \pi(\alpha|x, \beta) d\alpha \\ &= \int_0^{\hat{\alpha}_{MLE}} \log t d\Pr(\alpha \leq t|x, \beta) \\ &= \log \hat{\alpha}_{MLE} - \int_0^{\hat{\alpha}_{MLE}} \frac{1}{t} \Pr(\alpha \leq t|x, \beta) dt \\ &= \log \hat{\alpha}_{MLE} - \frac{1}{n\beta}. \end{aligned}$$

Hence

$$\hat{\alpha}_{MGE}(x|\beta) = \frac{\hat{\alpha}_{MLE}}{\exp\left(\frac{1}{n\beta}\right)}.$$

When α is known, the risk function is

$$R((\alpha, \beta), (\alpha, \hat{\beta})) = \int_0^{+\infty} D_{\alpha}((\alpha, \beta), (\alpha, \hat{\beta})) \pi(\beta|x, \alpha) d\beta.$$

Since

$$\frac{dR}{d\hat{\beta}} = \int_0^{+\infty} \left(-\frac{\beta}{\hat{\beta}^2} + \frac{1}{\hat{\beta}}\right) \pi(\beta|x, \alpha) d\beta,$$

when $\frac{dR}{d\hat{\beta}} = 0$, we have

$$\hat{\beta}_{MGE}(x|\alpha) = \frac{\int_0^{+\infty} \beta \pi(\beta|x, \alpha) d\beta}{\int_0^{+\infty} \frac{1}{\beta} \pi(\beta|x, \alpha) d\beta}.$$

Noting that

$$\begin{aligned} \int_0^{+\infty} \beta \pi(\beta|x, \alpha) d\beta &= \frac{(q_2(x) - n \log \alpha)^{n+1}}{\tau(n+1)} \int_0^{+\infty} \beta^{n+1} \exp(-(q_2(x) - n \log \alpha)\beta) d\beta \\ &= \frac{n(n+1)}{(q_2(x) - n \log \alpha)^2} \frac{(q_2(x) - n \log \alpha)^{n+1}}{\tau(n+1)} \int_0^{+\infty} \beta^{n-1} \exp(-(q_2(x) - n \log \alpha)\beta) d\beta \\ &= \frac{n(n+1)}{(q_2(x) - n \log \alpha)^2} \int_0^{+\infty} \frac{1}{\beta} \pi(\beta|x, \alpha) d\beta \end{aligned}$$

hence we have $\hat{\beta}_{MGE}(x|\alpha) = \frac{\sqrt{n(n+1)}}{q_2(x) - n \log \alpha}$. \square

4.3.2. Bayesian Estimations under Al-Bayyati’s Loss Function

The Al-Bayyati’s loss function was stated by Reference [19] as

$$l(\hat{\theta}, \theta) = \theta^c (\hat{\theta} - \theta)^2, \tag{24}$$

where c is a real number. Next, we use Al-Bayyati’s loss function to derive the Bayesian estimation of Pareto model.

Proposition 3. Assume that $\theta \in \Theta \subset R$. Under Al-Bayyati’s loss function, the Bayesian estimation of parameter θ is given by

$$\hat{\theta}_{B_c} = \frac{\int_{\Theta} \theta^{c+1} \pi(\theta|x) d\theta}{\int_{\Theta} \theta^c \pi(\theta|x) d\theta}. \tag{25}$$

Proof. Since the risk function

$$R(\hat{\theta}, \theta) = \int_{\Theta} \theta^c (\hat{\theta} - \theta)^2 \pi(\theta|x) d\theta,$$

we have

$$\frac{dR(\hat{\theta}, \theta)}{d\hat{\theta}} = \hat{\theta} \int_{\Theta} \theta^c \pi(\theta|x) d\theta - \int_{\Theta} \theta^{c+1} \pi(\theta|x) d\theta.$$

Then we have

$$\hat{\theta}_{B_c} = \frac{\int_{\Theta} \theta^{c+1} \pi(\theta|x) d\theta}{\int_{\Theta} \theta^c \pi(\theta|x) d\theta},$$

by letting $\frac{dR(\hat{\theta}, \theta)}{d\hat{\theta}} = 0$. \square

Using Al-Bayyati’s loss function, $\hat{\alpha}_{B_c}$ lacks the simple display expression. Thus we give the upper and lower bound estimations.

Theorem 4. Using Al-Bayyati’s loss function and assuming $c \geq 0$, we find that when β is unknown, then $\hat{\alpha}_{B_c}$ satisfies

$$\frac{(n\hat{\beta}_{MLE} + c)((n - 1)\hat{\beta}_{MLE} - (c + 1))}{(n - 1)n\hat{\beta}_{MLE}^2} \hat{\alpha}_{MLE} \leq \hat{\alpha}_{B_c} \leq \frac{(n - 1)n\hat{\beta}_{MLE}^2}{(n\hat{\beta}_{MLE} + c + 1)((n - 1)\hat{\beta}_{MLE} - c)} \hat{\alpha}_{MLE}.$$

And when α is unknown, then

$$\hat{\beta}_{B_c} = \left(1 + \frac{c}{n}\right) \hat{\beta}_{MLE}.$$

Furthermore, there exists c_0 such that $\hat{\beta}_{B_{c_0}} = \beta_0$, where β_0 is the real value of shape parameter β .

Proof. When β is unknown, we have

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} \alpha^{c+1} \pi(\alpha|x) d\alpha &= \int_0^{\hat{\alpha}_{MLE}} t^{c+1} d\Pr(\alpha \leq t|x) \\ &= \left(t^{c+1} \Pr(\alpha \leq t|x)\right) \Big|_0^{\hat{\alpha}_{MLE}} - (c + 1) \int_0^{\hat{\alpha}_{MLE}} t^c \Pr(\alpha \leq t|x) dt \\ &= \hat{\alpha}_{MLE}^{c+1} - (c + 1) \int_0^{\hat{\alpha}_{MLE}} t^c \Pr(\alpha \leq t|x) dt. \end{aligned}$$

Noting that

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} t^c d\Pr(\alpha \leq t|x) &= \int_0^{\hat{\alpha}_{MLE}} t^c \left(1 + \hat{\beta}_{MLE} \log \frac{\hat{\alpha}_{MLE}}{t}\right)^{-n} dt \\ &= \hat{\alpha}_{MLE}^{c+1} \int_0^{+\infty} \exp(-(c + 1)u) (1 + u\hat{\beta}_{MLE})^{-n} du \\ &\leq \hat{\alpha}_{MLE}^{c+1} \int_0^{+\infty} (1 + u\hat{\beta}_{MLE})^{-n} du \\ &= \frac{\hat{\alpha}_{MLE}^{c+1}}{(n - 1)\hat{\beta}_{MLE}} \end{aligned}$$

and that

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} t^c dPr(\alpha \leq t|x) &= \hat{\alpha}_{MLE}^{c+1} \int_0^{+\infty} \exp(-(c+1)u) (1 + u\hat{\beta}_{MLE})^{-n} du \\ &\geq \hat{\alpha}_{MLE}^{c+1} \int_0^{+\infty} \exp(-(c+1)u) (\exp(u\hat{\beta}_{MLE}))^{-n} du \\ &= \hat{\alpha}_{MLE}^{c+1} \int_0^{+\infty} \exp(-(n\hat{\beta}_{MLE} + c + 1)u) du \\ &= \frac{\hat{\alpha}_{MLE}^{c+1}}{n\hat{\beta}_{MLE} + c + 1} \end{aligned}$$

we have

$$\begin{aligned} \int_0^{\hat{\alpha}_{MLE}} \alpha^{c+1} \pi(\alpha|x) d\alpha &\geq \hat{\alpha}_{MLE}^{c+1} - (c+1) \frac{\hat{\alpha}_{MLE}^{c+1}}{(n-1)\hat{\beta}_{MLE}} = \frac{(n-1)\hat{\beta}_{MLE} - (c+1)}{(n-1)\hat{\beta}_{MLE}} \hat{\alpha}_{MLE}^{c+1} \\ \int_0^{\hat{\alpha}_{MLE}} \alpha^{c+1} \pi(\alpha|x) d\alpha &\leq \hat{\alpha}_{MLE}^{c+1} - \frac{(c+1)\hat{\alpha}_{MLE}^{c+1}}{n\hat{\beta}_{MLE} + c + 1} = \frac{n\hat{\beta}_{MLE}}{n\hat{\beta}_{MLE} + c + 1} \hat{\alpha}_{MLE}^{c+1}. \end{aligned}$$

Similarly, we can get

$$\frac{(n-1)\hat{\beta}_{MLE} - c}{(n-1)\hat{\beta}_{MLE}} \hat{\alpha}_{MLE}^c \leq \int_0^{\hat{\alpha}_{MLE}} \alpha^c \pi(\alpha|x) d\alpha \leq \frac{n\hat{\beta}_{MLE}}{n\hat{\beta}_{MLE} + c} \hat{\alpha}_{MLE}^c.$$

Furthermore, by Proposition 3, we have

$$\hat{\alpha}_{B_c} = \frac{1 - (c+1) \int_0^{+\infty} \exp(-(c+1)u) (1 + u\hat{\beta}_{MLE})^{-n} du}{1 - c \int_0^{+\infty} \exp(-cu) (1 + u\hat{\beta}_{MLE})^{-n} du} \hat{\alpha}_{MLE}$$

and

$$\begin{aligned} \hat{\alpha}_{B_c} &\geq \frac{(n-1)\hat{\beta}_{MLE} - (c+1)}{(n-1)\hat{\beta}_{MLE}} \hat{\alpha}_{MLE}^{c+1} \frac{n\hat{\beta}_{MLE} + c}{n\hat{\beta}_{MLE}} \hat{\alpha}_{MLE}^{-c} = \frac{(n\hat{\beta}_{MLE} + c)((n-1)\hat{\beta}_{MLE} - (c+1))}{(n-1)n\hat{\beta}_{MLE}^2} \hat{\alpha}_{MLE} \\ \hat{\alpha}_{B_c} &\leq \frac{n\hat{\beta}_{MLE}}{n\hat{\beta}_{MLE} + c + 1} \hat{\alpha}_{MLE}^{c+1} \frac{(n-1)\hat{\beta}_{MLE}}{(n-1)\hat{\beta}_{MLE} - c} \hat{\alpha}_{MLE}^{-c} = \frac{(n-1)n\hat{\beta}_{MLE}^2}{(n\hat{\beta}_{MLE} + c + 1)((n-1)\hat{\beta}_{MLE} - c)} \hat{\alpha}_{MLE}. \end{aligned}$$

When α is unknown, we have

$$\begin{aligned} \int_0^{+\infty} \beta^{c+1} \pi(\beta|x) d\beta &= \frac{(q_2(x) - n \log q_1(x))^n}{\tau(n)} \int_0^{+\infty} \beta^{n+c} \exp(-(q_2(x) - n \log q_1(x))\beta) d\beta \\ &= \frac{n+c}{q_2(x) - n \log q_1(x)} \int_0^{+\infty} \beta^c \pi(\beta|x) d\beta. \end{aligned}$$

Thus, by Proposition 3, we have

$$\hat{\beta}_{B_c} = \left(1 + \frac{c}{n}\right) \hat{\beta}_{MLE}.$$

And when $c_0 = n\left(\frac{\beta_0}{\hat{\beta}_{MLE}} - 1\right)$, we have $\hat{\beta}_{B_{c_0}} = \beta_0$. \square

Remark 1. When β is unknown, we have

$$\hat{\alpha}_{B_c} = \frac{1 - (c+1) \int_0^{+\infty} \exp(-(c+1)u) (1 + u\hat{\beta}_{MLE})^{-n} du}{1 - c \int_0^{+\infty} \exp(-cu) (1 + u\hat{\beta}_{MLE})^{-n} du} \hat{\alpha}_{MLE}. \tag{26}$$

In particular, when $c = 0$, we get

$$\hat{\alpha}_{B_c} = \hat{\alpha}_E = \left(1 - \int_0^{+\infty} \exp(-u)(1 + u\hat{\beta}_{MLE})^{-n} du\right) \hat{\alpha}_{MLE}. \tag{27}$$

From (27), we know that $\hat{\alpha}_E < \hat{\alpha}_{MLE}$. By analyzing (26), when $c > 0$ gradually increases, $\hat{\alpha}_{B_c}$ will increase firstly and then decrease, and finally $\hat{\alpha}_{B_c}$ will converge to $\hat{\alpha}_{MLE}$. Let α_0 be the real value of scale parameter α , then when $\hat{\alpha}_E \leq \alpha_0 \leq \hat{\alpha}_{MLE}$, there exists $c_0 \neq 0$ such that $\hat{\alpha}_{B_{c_0}} = \alpha_0$. When $\alpha_0 > \hat{\alpha}_E$, $\hat{\alpha}_E$ is the closest estimation among all $\hat{\alpha}_{B_c}$. When $\alpha_0 > \hat{\alpha}_{MLE}$, there exists $c_0 \neq 0$ such that $\hat{\alpha}_{c_0}$ is the closest estimation among all $\hat{\alpha}_{B_c}$.

Theorem 5. Using Al-Bayyati’s loss function, we find that when β is known, $\hat{\alpha}_{B_c}(x|\beta) = \frac{n\beta+c}{n\beta+c+1}\hat{\alpha}_{MLE}$. When α is known, $\hat{\beta}_{B_c}(x|\alpha) = \frac{n+1+c}{q_2(x)-n\log \alpha}$. In both cases, there exist c_1 and c_2 such that $\hat{\alpha}_{B_{c_1}}(x|\beta) = \alpha_0$ and $\hat{\beta}_{B_{c_2}}(x|\alpha) = \beta_0$, where α_0 and β_0 are the real value of scale and shape parameter respectively.

Proof. When β is known, we have

$$\int_0^{\hat{\alpha}_{MLE}} \alpha^{c+1} \pi(\alpha|x, \beta) d\alpha = \int_0^{\hat{\alpha}_{MLE}} t^{c+1} d\Pr(\alpha \leq t|x, \beta) = \frac{n\beta}{n\beta+c+1} \hat{\alpha}_{MLE}^{c+1}$$

and

$$\int_0^{\hat{\alpha}_{MLE}} \alpha^c \pi(\alpha|x, \beta) d\alpha = \frac{n\beta}{n\beta+c} \hat{\alpha}_{MLE}^c.$$

Then by Proposition 3, we have

$$\hat{\alpha}_{B_c}(x|\beta) = \frac{\int_0^{\hat{\alpha}_{MLE}} \alpha^{c+1} \pi(\alpha|x, \beta) d\alpha}{\int_0^{\hat{\alpha}_{MLE}} \alpha^c \pi(\alpha|x, \beta) d\alpha} = \frac{n\beta+c}{n\beta+c+1} \hat{\alpha}_{MLE}.$$

When α is known, we get

$$\begin{aligned} \int_0^{+\infty} \beta^{c+1} \pi(\beta|x, \alpha) d\beta &= \frac{(q_2(x) - n \log \alpha)^{n+1}}{\tau(n+1)} \int_0^{+\infty} \beta^{n+c+1} \exp(-(q_2(x) - n \log \alpha)\beta) d\beta \\ &= \frac{n+c+1}{q_2(x) - n \log q_1(x)} \int_0^{+\infty} \beta^c \pi(\beta|x, \alpha) d\beta. \end{aligned}$$

Thus by Proposition 3, we get

$$\hat{\beta}_{B_c}(x|\alpha) = \frac{n+1+c}{q_2(x) - n \log \alpha}.$$

When $\beta = \beta_0$, we have

$$\hat{\alpha}_{B_c}(x|\beta_0) = \frac{n\beta_0+c}{n\beta_0+c+1} \hat{\alpha}_{MLE}$$

Noting that if $\hat{\alpha}_{MLE} \neq \alpha_0$, then $\hat{\alpha}_{B_{c_0}}(x|\beta_0) = \alpha_0$, where

$$c_0 = \frac{\alpha_0}{\hat{\alpha}_{MLE} - \alpha_0} - n\beta_0.$$

Hence either $\hat{\alpha}_{MLE}$ or $\hat{\alpha}_{B_{c_0}}(x|\beta_0)$ is the true value of α .

When $\alpha = \alpha_0$, we have

$$\hat{\beta}_{B_c}(x|\alpha_0) = \frac{n + 1 + c}{q_2(x) - n \log \alpha_0}.$$

Hence we can take

$$c_0 = (q_2(x) - n \log \alpha_0)\alpha_0 - (n + 1),$$

such that $\hat{\beta}_{B_{c_0}}(x|\alpha_0)$ is the true value of β . \square

4.4. Bayesian Posterior Prediction

Let $\tilde{X} \sim \pi(\tilde{x}; \alpha, \beta)$ be the value that needs to be observed from Pareto distribution. In the sense of posterior distribution (13), if the sample X is given, we can make relevant posterior prediction of \tilde{X} . The discussion will be divided into the following three cases.

1. When neither α nor β are unknown, then we have

$$\begin{aligned} m(\tilde{x}|x) &= \int_0^{+\infty} \int_0^{+\infty} \pi(\tilde{x}; \alpha, \beta)\pi(\alpha, \beta|x)dad\beta \\ &= - \int_0^{+\infty} \int_0^{+\infty} \frac{n(q_2(x) - n \log q_1(x))^n}{\tau(n)} \beta^n \alpha^{n\beta-1} \exp(-q_2(x)\beta) I_{[0 \leq \alpha \leq q_1(x)]} \cdot \frac{\beta \alpha^\beta}{\tilde{x}^{\beta+1}} I_{[\tilde{x} \geq \alpha]} dad\beta \\ &= - \int_0^{+\infty} \frac{n(q_2(x) - n \log q_1(x))^n}{\tau(n)} \exp(-q_2(x)\beta) \frac{\beta^{n+1}}{\tilde{x}^{\beta+1}} d\beta \int_0^{\min(\tilde{x}, q_1(x))} \alpha^{(n+1)\beta-1} d\alpha \\ &= - \int_0^{+\infty} \frac{n(q_2(x) - n \log q_1(x))^n}{\tau(n)} \exp(-q_2(x)\beta) \frac{\beta^{n+1}}{\tilde{x}^{\beta+1}} \cdot \frac{(\min(\tilde{x}, q_1(x)))^{(n+1)\beta}}{(n+1)\beta} d\beta \\ &= - \int_0^{+\infty} \frac{n(q_2(x) - n \log q_1(x))^n}{\tau(n)} \beta^n \frac{\exp(-(-(n+1) \log \min(\tilde{x}, q_1(x)) + \log \tilde{x} + q_2(x))\beta)}{(n+1)\tilde{x}} d\beta \\ &= \frac{n^2(q_2(x) - n \log q_1(x))^n}{(n+1)\tilde{x}(-(n+1) \log(\min(\tilde{x}, q_1(x))) + \log \tilde{x} + q_2(x))^{n+1}} I_{[\tilde{x} > 0]} \\ &= \frac{n^2(q_2(x) - n \log q_1(x))^n}{(n+1)\tilde{x}} N(x, \tilde{x}), \end{aligned}$$

where

$$N(x, \tilde{x}) = \begin{cases} (q_2(x) - n \log \tilde{x})^{-(n+1)}, & 0 < \tilde{x} < \hat{\alpha}_{MLE}(x) \\ ((n+1) \log q_1(x) + \log \tilde{x} + q_2(x))^{-(n+1)}, & \tilde{x} \geq \hat{\alpha}_{MLE}(x). \end{cases}$$

2. When α is known and β is unknown, we have

$$m(\tilde{x}|x, \alpha) = \int_0^{+\infty} \pi(\tilde{x}; \alpha, \beta)\pi(\beta|x, \alpha)d\beta = \frac{(n+1)(q_2(x) - n \log \alpha)^{n+1}}{\tilde{x}(-(n+1) \log \alpha + \log \tilde{x} + q_2(x))^{n+2}} I_{[\tilde{x} > \alpha]}.$$

3. When β is known and α is unknown, we have

$$\begin{aligned} m(\tilde{x}|x, \beta) &= \int_0^{+\infty} \pi(\tilde{x}; \alpha, \beta)\pi(\alpha|x, \beta)d\alpha \\ &= \frac{n}{n+1} \beta \hat{\alpha}_{MLE}^{-n\beta}(x) \cdot \tilde{x}^{-\beta-1} (\min(\tilde{x}, \hat{\alpha}_{MLE}(x)))^{(n+1)\beta} I_{[\tilde{x} > 0]} \\ &= \frac{n}{n+1} \beta \cdot \begin{cases} \hat{\alpha}_{MLE}^{-n\beta}(x) \cdot \tilde{x}^{n\beta-1}, & 0 < \tilde{x} < \hat{\alpha}_{MLE}(x) \\ \hat{\alpha}_{MLE}^\beta(x) \cdot \tilde{x}^{-\beta-1}, & \tilde{x} \geq \hat{\alpha}_{MLE}(x). \end{cases} \end{aligned}$$

For the above posterior prediction distribution, given the prediction credibility k , we can make Bayesian prediction inference in practical applications. The specific process is as follows. From $\int_{\tilde{x}_L}^{\tilde{x}_U} m(\tilde{x}|x)d\tilde{x} = k$, multiple sets of \tilde{x}_L, \tilde{x}_U can be got. By choosing

appropriately the upper and lower bounds for \tilde{X} such that $\tilde{x}_U - \tilde{x}_L$ is smaller, then we can obtain higher prediction accuracy.

5. Simulation

In real life, the proposed algorithm for target detection of maritime radar needs to be tested and verified on sea clutter data. In order to determine the sea clutter better, it is often necessary to estimate the parameters of sea clutter model.

Therefore, in this section, we will use the conclusion of Section 4 to estimate the parameters of Pareto model of sea clutter and show the simulation results.

5.1. The Influence of Parameters on Sea Clutter

In this subsection, we show the effect of scale parameter α and shape parameter β on sea clutter.

Figures 1 and 2 show the probability density curve of Pareto distribution with respect to two parameters. It can be seen from the figures that when the scale parameter α is larger, the density curve is even. The proportion of small clutter amplitude increases, and the decline of the whole curve is gentle. As the shape parameter β becomes larger, the proportion of small clutter amplitude increases significantly and becomes more concentrated, and the tail descends faster. On the whole, for Pareto model, the energy is concentrated on the small clutter. The trailing phenomenon is apparent. The essential reason is that when the radar is grazing incident, the overall backscatter echo is relatively weak.

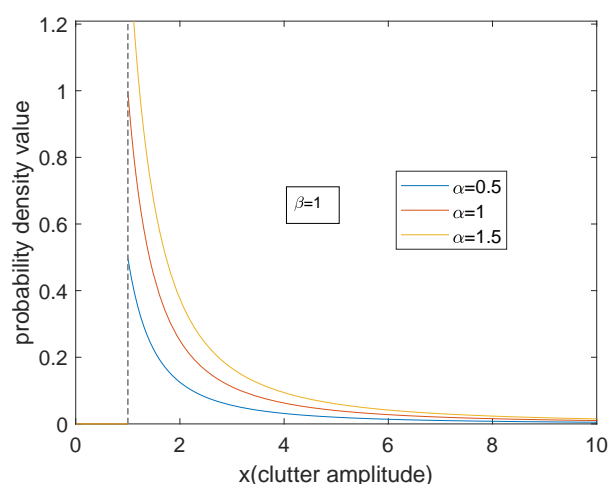


Figure 1. The influence of the change of α probability distributions.

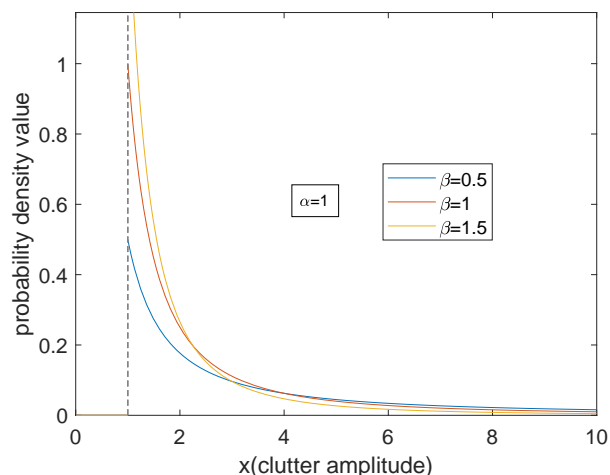


Figure 2. The influence of the change of β on probability distributions.

5.2. Various Types of Bayesian Estimation on Sea Clutter Models

In this subsection, we show the aforementioned Bayesian estimations of sea clutter.

In order to generate random samples of Pareto distribution with parameters α_0 and β_0 , we use the inverse distribution function and take the inverse transformation method to extract Pareto samples: $X = \frac{\alpha_0 U^{-1}}{\beta_0}$, where U is the uniformly distributed random variable on $[0, 1]$. We carry out numerical simulations where $\alpha_0 = 0.5, 1, 1.5$ and $\beta_0 = 0.5, 1, 1.5$, respectively. Using the inverse transformation method, we generate 1000 random samples subject to Pareto distribution. To show the geometry of Pareto model of sea clutter, we take $(\alpha_0, \beta_0) = (0.5, 1 \pm 0.5)$ as the center and draw the unit geodesic circumference with dotted line. We draw 64 uniformly distributed geodesics with directions $\theta_0 = \frac{k\pi}{32}, k = 0, 1, \dots, 63$ with solid line. See Figures 3 and 4.

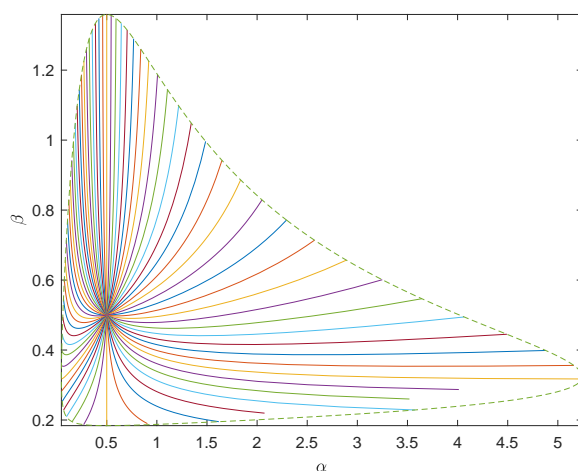


Figure 3. Sixty-four uniformly distributed geodesics centered on (0.5, 0.5).

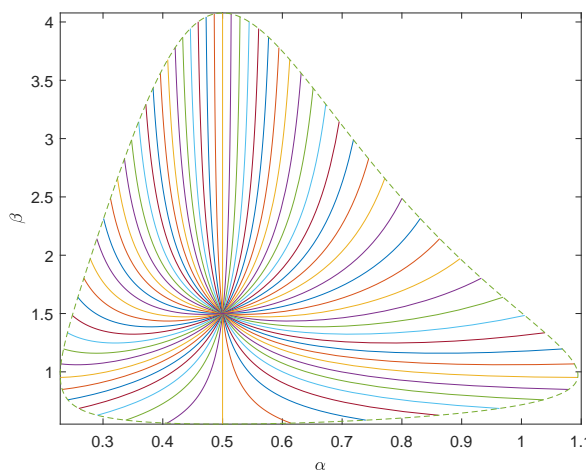


Figure 4. Sixty-four uniformly distributed geodesics centered on (0.5, 1.5).

To describe the proximity between the estimated values of each group and the predetermined parameter value (α_0, β_0) , we need to calculate the geodesic distance $d\{(\alpha_0, \beta_0), (\hat{\alpha}(x), \hat{\beta}(x))\}$. If the distance between the estimated value and the predetermined parameter value is close, we believe that the estimation is accurate. By (22), $d((\alpha_0, \beta_0), (\hat{\alpha}(x), \hat{\beta}(x))) \propto 1 + \frac{\beta_0 \hat{\beta}(x) (\log \alpha_0 - \log \hat{\alpha}(x))^2}{2} + \frac{(\beta_0 - \hat{\beta}(x))^2}{2\beta_0 \hat{\beta}(x)}$. Hence the smaller $|\log \alpha_0 - \log \hat{\alpha}|$ and $|\beta_0 - \hat{\beta}|$ are, the more accurate the estimation is.

Next we will make a comparative analysis of various types of Bayesian estimations.

5.2.1. Mean Geodesic Estimation and the Common Bayesian Estimations

Case 1. Both scale parameter α and shape parameter β are unknown.

From Table 1 we know that $|\hat{\alpha}_E - \hat{\alpha}_{MGE}|$ and $|\hat{\beta}_E - \hat{\beta}_{MGE}|$ are less than 10^{-4} , hence $\hat{\alpha}_E$ and $\hat{\alpha}_{MGE}$ are almost equal. Since $\hat{\alpha}_E$ does not have explicit expression, $\hat{\alpha}_{MGE}$ can take the place of $\hat{\alpha}_E$ and it also has more precise geometric explanation.

Table 1. Mean geodesic estimations and the common Bayesian estimations (α, β are not known).

(α, β)	$(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$	$(\hat{\alpha}_{Me}, \hat{\beta}_{Me})$	$(\hat{\alpha}_E, \hat{\beta}_E)$	$(\hat{\alpha}_{MGE}, \hat{\beta}_{MGE})$
(0.5,0.5)	(0.5024,0.5171)	(0.5017,0.5169)	(0.5014,0.5171)	(0.5014,0.5171)
(0.5,1.0)	(0.5001,1.0292)	(0.4998,1.0288)	(0.4997,1.0292)	(0.4997,1.0292)
(0.5,1.5)	(0.5003,1.5566)	(0.5001,1.5561)	(0.5000,1.5566)	(0.5000,1.5566)
(1.0,0.5)	(1.0053,0.4743)	(1.0038,0.4742)	(1.0032,0.4743)	(1.0032,0.4743)
(1.0,1.0)	(1.0009,1.0209)	(1.0002,1.0205)	(0.9999,1.0209)	(0.9999,1.0209)
(1.0,1.5)	(1.0001,1.4639)	(0.9996,1.4634)	(0.9994,1.4639)	(0.9994,1.4639)
(1.5,0.5)	(1.5033,0.4883)	(1.5012,0.4881)	(1.5002,0.4883)	(1.5002,0.4883)
(1.5,1.0)	(1.5003,1.0226)	(1.4993,1.0223)	(1.4988,1.0226)	(1.4988,1.0226)
(1.5,1.5)	(1.5010,1.4978)	(1.5003,1.4973)	(1.5003,1.4978)	(1.5000,1.4978)

In most simulation tests, $(\hat{\alpha}_{MGE}, \hat{\beta}_{MGE})$ is more accurate than $(\hat{\alpha}_{MLE}, \hat{\beta}_{MLE})$ and $(\hat{\alpha}_{ME}, \hat{\beta}_{ME})$. Hence, in general the estimation MGE that we proposed is better than the common Bayesian estimations.

Case 2. Either shape parameter β or scale parameter α is known.

When one parameter is known, the statistical manifold is degenerated. Hence taking the Euclidean distance or the geodesic distance does not make much difference. This can be seen in Tables 2 and 3.

Table 2. Mean geodesic estimations and the common Bayesian estimations (β is known).

$\beta = \beta_0$	$\hat{\alpha}_{Me}(x \beta_0)$	$\hat{\alpha}_E(x \beta_0)$	$\hat{\alpha}_E(x \beta_0)$	$\hat{\alpha}_{MGE}(x \beta_0)$
(0.5,0.5)	0.5024	0.5017	0.5014	0.5014
(0.5,1.0)	0.5001	0.4998	0.4996	0.4996
(0.5,1.5)	0.5003	0.5000	0.4999	0.4999
(1.0,0.5)	1.0053	1.0039	1.0033	1.0033
(1.0,0.5)	1.0009	1.0002	0.9999	0.9999
(1.0,1.5)	1.0001	0.9996	0.9994	0.9994
(1.5,0.5)	1.5033	1.5012	1.5003	1.5003
(1.5,1.0)	1.5003	1.4992	1.4988	1.4988
(1.5,1.5)	1.5010	1.5003	1.5000	1.5000

Table 3. Mean geodesic estimations and the common Bayesian estimations (α is known).

$\alpha = \alpha_0$	$\hat{\beta}_{MLE}(x \alpha_0)$	$\hat{\beta}_{Me}(x \alpha_0)$	$\hat{\beta}_E(x \alpha_0)$	$\hat{\beta}_{MGE}(x \alpha_0)$
(0.5,0.5)	0.5158	0.5162	0.5163	0.5161
(0.5,1.0)	1.0289	1.0295	1.0299	1.0294
(0.5,1.5)	1.5553	1.5563	1.5568	1.5560
(1.0,0.5)	0.4732	0.4735	0.4736	0.4734
(1.0,0.5)	1.0199	1.0206	1.0210	1.0205
(1.0,1.5)	1.4638	1.4647	1.4652	1.4645
(1.5,0.5)	0.4877	0.4881	0.4882	0.4880
(1.5,1.0)	1.0224	1.0231	1.0234	1.0229
(1.5,1.5)	1.4921	1.4931	1.4936	1.4928

Comparing Tables 2 and 3, when β is known, the accuracy of the estimations is 10^{-3} and when α is known, the accuracy of all kinds of estimations is 10^{-2} . Therefore, the

accuracy of various types of estimations will improve if β is known. This indicates that scale parameter α is more easily obtained from samples in the sea clutter model and has strong robustness. Shape parameter β is more sensitive than scale parameter α .

5.2.2. The Estimations under Al-Bayyati’s Loss Function

In this subsection, we give the simulation of various types of Bayesian estimations when $(\alpha_0, \beta_0) = (1.5, 1.5)$.

Case 1. Both scale parameter α and shape parameter β are unknown.

When α and β are unknown, the variation trend of Bayesian estimation of the two parameters under Al-Bayyati’s loss function with parameter c is shown in Figures 5 and 6, respectively. When β is unknown, by (26)

$$\hat{\alpha}_{B_c} = \frac{1 - (c + 1) \int_0^{+\infty} \exp(- (c + 1)u) (1 + u\hat{\beta}_{MLE})^{-n} du}{1 - c \int_0^{+\infty} \exp(-cu) (1 + u\hat{\beta}_{MLE})^{-n} du} \hat{\alpha}_{MLE}.$$

Figure 5 shows the case when $\hat{\alpha}_E \geq \alpha_0$. Hence from the discussion in Remark 1, $\hat{\alpha}_E$ is the closest estimate among all $\hat{\alpha}_{B_c}$ with the increasing of positive c .

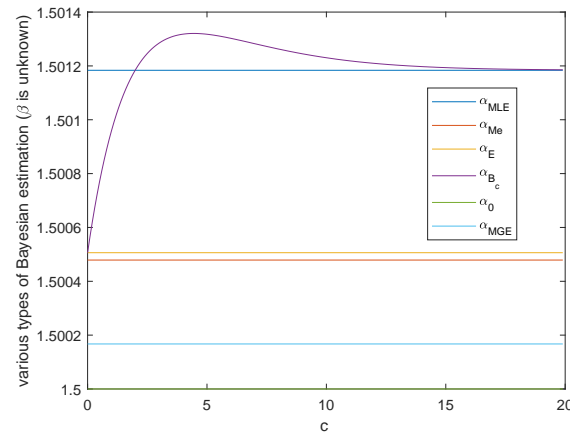


Figure 5. The variation of various Bayesian estimations of α with Al-Bayyati’s loss function parameter c .

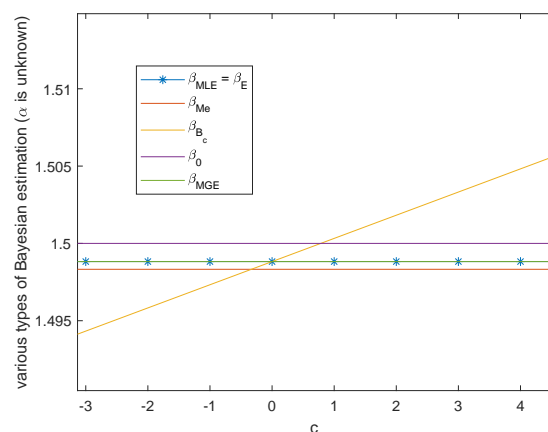


Figure 6. The variation of various Bayesian estimations of β with Al-Bayyati’s loss function parameter c .

When α is unknown, by Theorem 4

$$\hat{\beta}_{B_c} = \left(1 + \frac{c}{n}\right) \hat{\beta}_{MLE}.$$

When $c = 0$, $\hat{\beta}_{B_c} = \hat{\beta}_{MLE} = \hat{\beta}_E$ and when $c_0 = n\left(\frac{\beta_0}{\hat{\beta}_{MLE}} - 1\right)$, $\hat{\beta}_{B_{c_0}} = \beta_0$. As shown in Figure 6, there always exists infinitely many c such that $\hat{\beta}_c$ is closer than the common

Bayesian estimations. And $c_0 = n\left(\frac{\beta_0}{\hat{\beta}_{MLE}} - 1\right)$ is nothing but the real value of parameter β . These two figures also show that MGE are better than the common Bayesian estimations when both of the parameters are unknown. Therefore, when α and β are unknown, to obtain closer estimations, we are able to change c_1 and c_2 to make $|\log \alpha_0 - \log \hat{\alpha}_{c_1}|$ and $|\beta_0 - \hat{\beta}_{c_2}|$ smaller and even obtain the minimum value. Through the previous discussions, the choice of the best c_1 depends on the inequality among real parameter α_0 , $\hat{\alpha}_{MLE}$ and $\hat{\alpha}_E$. The best c_2 is $c_2 = n\left(\frac{\beta_0}{\hat{\beta}_{MLE}} - 1\right)$.

Case 2. Either shape parameter β or scale parameter α is known.

When β or α is known, the variation trend of various Bayesian estimation of parameter α or parameter β in Al-Bayyati’s loss function with parameter c is shown in Figure 7 or Figure 8, respectively. When $\beta = \beta_0$, by Theorem 5, we get

$$\hat{\alpha}_{B_c}(x|\beta) = \frac{n\beta + c}{n\beta + c + 1} \hat{\alpha}_{MLE}.$$

If $\hat{\alpha}_{MLE} \neq \alpha_0$, we have $\hat{\alpha}_{B_c}(x|\beta) = \alpha_0$ when $c = \frac{\alpha_0}{\hat{\alpha}_{MLE} - \alpha_0} - n\beta$. Hence either $\hat{\alpha}_{MLE}$ is the true value of α , or we can take $c_0 = \frac{\alpha_0}{\hat{\alpha}_{MLE} - \alpha_0} - n\beta$ such that $\hat{\alpha}_{B_{c_0}}(x|\beta)$, which is the true value of α . This is shown in Figure 7.

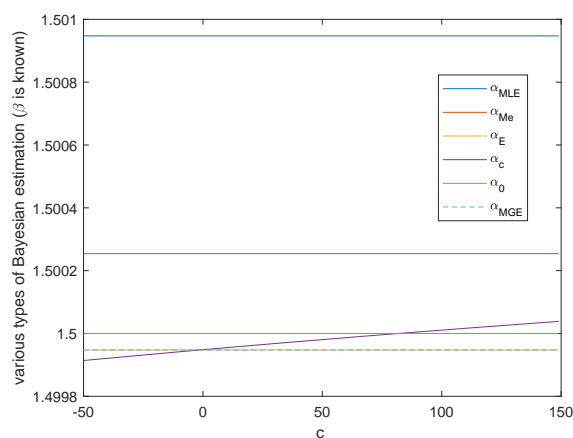


Figure 7. The variation of various Bayesian estimations of α with loss function parameter c when $\beta = \beta_0$.

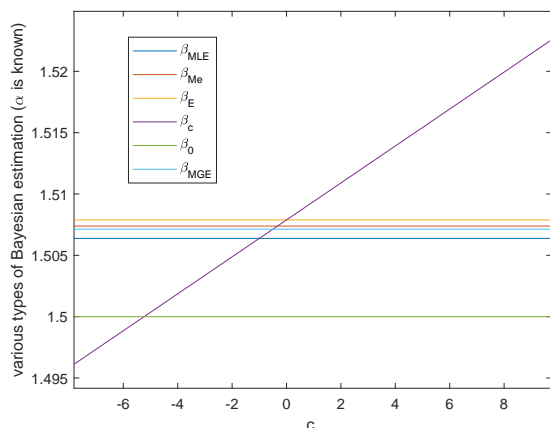


Figure 8. The variation of various Bayesian estimations of β with loss function parameter c when $\alpha = \alpha_0$.

When $\alpha = \alpha_0$, by Theorem 5, we get

$$\hat{\beta}_{B_c}(x|\alpha) = \frac{n + 1 + c}{q_2(x) - n \log \alpha}.$$

If $c = (q_2(x) - n \log \alpha)\alpha_0 - (n + 1)$, then $\hat{\beta}_{B_c}(x|\alpha) = \beta_0$. Hence we can take $c_0 = (q_2(x) - n \log \alpha)\alpha_0 - (n + 1)$ such that $\hat{\beta}_{B_{c_0}}(x|\alpha)$ is the true value of β . This is shown in Figure 8.

In above two cases, there are infinitely many c such that $\hat{\alpha}_c(x|\beta)$ or $\hat{\alpha}_c(x|\alpha)$ is closer than the common Bayesian estimations.

5.3. Simulation of Posterior Predictive Distribution

In order to observe the simulation effect of the posterior prediction distribution, according to the samples generated in Section 5.2, we drew the posterior prediction distribution of sea clutter and the real Pareto distribution $\pi(x|\alpha_0, \beta_0)$ of sea clutter where $(\alpha_0, \beta_0) = (1.5, 1.5)$ for comparative analysis. See Figures 9–11.

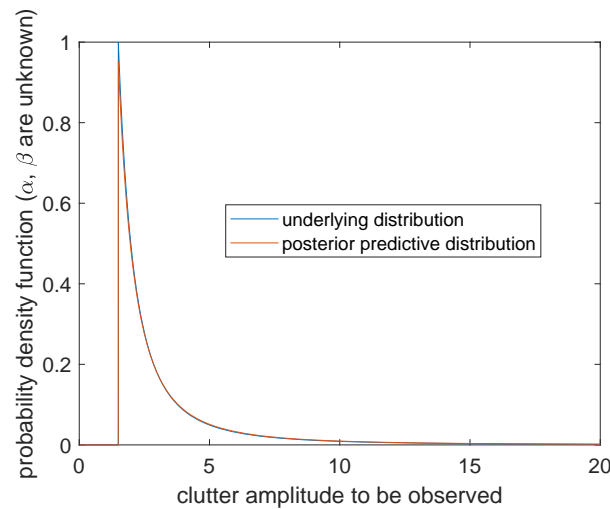


Figure 9. Posterior predictive distribution and underlying distribution (α and β are unknown).

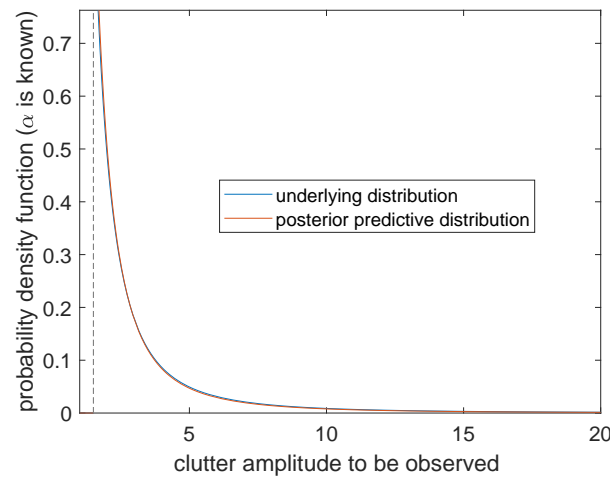


Figure 10. Posterior predictive distribution and underlying distribution ($\alpha = \alpha_0$).

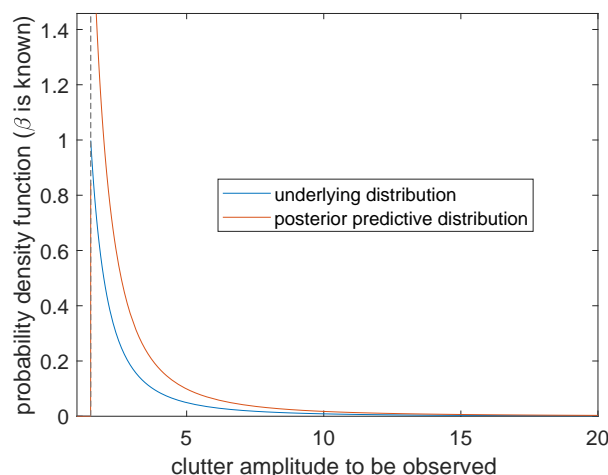


Figure 11. Posterior predictive distribution and underlying distribution ($\beta = \beta_0$).

Case 1. Scale parameter α and shape parameter β are unknown
 The posterior prediction distribution of sea clutter is

$$m(\tilde{x}|x) = \frac{n^2(q_2(x) - n \log q_1(x))^n}{(n + 1)\tilde{x}} N(x, \tilde{x}), \tag{28}$$

where

$$N(x, \tilde{x}) = \begin{cases} (q_2(x) - n \log \tilde{x})^{-(n+1)}, & 0 < \tilde{x} < \hat{\alpha}_{MLE}(x) \\ (- (n + 1) \log q_1(x) + \log \tilde{x} + q_2(x))^{-(n+1)}, & \tilde{x} \geq \hat{\alpha}_{MLE}(x). \end{cases}$$

The image is shown in Figure 9. The blue curve represents the probability distribution of sea clutter $\pi(x|\alpha_0, \beta_0)$, which gives positive values at the right side of the boundary point $x = \alpha_0$. The orange curve represents the posterior prediction distribution of sea clutter $m(\tilde{x}|x)$, which changes continuously when $\tilde{x} > 0$, but forms a cusp at $\tilde{x} = \hat{\alpha}_{MLE}$. Compared with the two curves, the curve of the predicted distribution of sea clutter is connected by a continuous curve and shifts slightly to the left. It is worth noting that although $m(\tilde{x}|x)$ tends to infinity as $\tilde{x} \rightarrow 0^+$, it is not reflected in the image and can be ignored in the actual calculation of the probability.

Case 2. α is known and β is unknown

The posterior prediction distribution of sea clutter is

$$m(\tilde{x}|x, \alpha) = \frac{(n + 1)(q_2(x) - n \log \alpha)^{n+1}}{\tilde{x}((n + 1) \log \alpha + \log \tilde{x} + q_2(x))^{n+2}} I_{[\tilde{x} > \alpha]}. \tag{29}$$

It can be seen from Figure 10 that the probability distribution $\pi(x|\alpha_0, \beta_0)$ (the blue curve) and the posterior prediction distribution $m(\tilde{x}|x, \beta)$ (the orange curve) can only obtain positive values at the right side of the boundary point $x = \alpha_0$. There is a very high degree of overlap, which means that when α is known, the prediction is going to be very accurate. We come to the conclusion that more effective information can be obtained for parameter α than β .

Case 3. β is known and α is unknown

The posterior prediction distribution of sea clutter is

$$m(\tilde{x}|x, \beta) = \frac{n}{n + 1} \beta \cdot \begin{cases} \hat{\alpha}_{MLE}^{-n\beta}(x) \cdot \tilde{x}^{n\beta-1}, & 0 < \tilde{x} < \hat{\alpha}_{MLE}(x) \\ \hat{\alpha}_{MLE}^{\beta}(x) \cdot \tilde{x}^{-\beta-1}, & \tilde{x} \geq \hat{\alpha}_{MLE}(x). \end{cases} \tag{30}$$

The probability distribution $\pi(x|\alpha_0, \beta_0)$ (the blue curve) obtains a positive value at the right side of boundary point $x = \alpha_0$. The posterior prediction distribution $m(\tilde{x}|x, \alpha)$

(the orange curve) changes continuously at $\tilde{x} > 0$ and forms a cusp at $\tilde{x} = \hat{\alpha}_{MLE}$. By comparing these two curves, the posterior prediction distribution shows a significant right shift, and the simulation effect is not ideal near $\tilde{x} = \hat{\alpha}_{MLE}$. However, with the continuous increasing of \tilde{x} , the two curves gradually coincide and the prediction accuracy becomes higher. Therefore, when β is known and α is unknown, the larger the clutter amplitude is to be observed, the higher the prediction accuracy will be.

To sum up, for the sea clutter model, the Bayesian posterior prediction results under the above three conditions are ideal, and the prediction model can well reflect the characteristics of sea clutter towing.

6. Conclusions and Future Work

In this paper, we presented systematic methods for Bayesian inference from geometric viewpoints and applied it to Pareto model. We carried out simulations on sea clutter to show the effectiveness.

For Pareto model, there does not exist general α -parallel prior. Using the Jeffreys prior and by using geodesic distance and Al-Bayyati's loss function, we obtain two new classes of Bayesian estimations. We call the estimation in the sense of mean geodesic distance MGE and it is proved that MGE has following advantages: it has the explicit expression, and is more accurate than the common Bayesian estimations which has shown in our simulation. We also prove that the estimations under the Al-Bayyati's loss function are more accurate than the common Bayesian estimations. Actually, there are infinitely many c such that the new estimations are better. These results are important for the estimation of parameters when studying sea clutter model. Finally we show that the Bayesian posterior prediction results can well reflect the characteristics of sea clutter towing in any case.

In the future, more in-depth researches are worth discussing. From statistical viewpoints, we can apply Bayesian inference for the Pareto model to non-linear regression models [20]. From geometrical viewpoints, we expect to generalize our framework and combine more tools from information geometry. We want to carry out more experiments and applications in different fields.

Author Contributions: F.S. participated in raising questions and completing the calculation process; H.S. participated in raising questions and confirming the results; Y.C. and S.Z. participated in part of the calculation process and confirming the results. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Plan of China, No. 2020YFC2006201.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is synthetically generated and described in the paper.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. de Carvalho, M.; Page, G.L.; Barney, B.J. On the geometry of Bayesian inference. *Bayesian Anal.* **2017**, *14*, 1013–1036.
2. Kulhavy, R. Recursive nonlinear estimation: A geometric approach. *Automatica* **1990**, *26*, 545–555.
3. Hartigan, J.A. Invariant prior distributions. *Ann. Math. Stat.* **1964**, *35*, 836–845.
4. Hartigan, J.A. The asymptotically unbiased density. *Ann. Math. Stat.* **1965**, *36*, 1137–1152.
5. Hartigan, J.A. The maximum likelihood prior. *Ann. Math. Stat.* **1998**, *26*, 2083–2103.
6. Takeuchi, J.; Amari, S. α -parallel prior and its properties. *IEEE Trans. Inf. Theory* **2005**, *51*, 1011–1023.
7. Tanaka, F. Curvature form on statistical model manifolds and its application to Bayesian analysis. *J. Stat. Appl. Probab.* **2012**, *1*, 35–43.
8. Amari, S. Differential geometry of curved exponential families—Curvature and information loss. *Ann. Stat.* **1982**, *10*, 357–385.

9. Amari, S.; Kumon, M. Differential geometry of edgeworth expansions in curved exponential family. *Ann. Inst. Stat. Math.* **1983**, *35*, 1–24.
10. Ward, K.D.; Watts, S. Use of sea clutter models in radar design and development. *IET Radar Sonar Navig.* **2010**, *4*, 146–157.
11. Ollila, E.; Tyler, D.E.; Koivunen, V.; Poor, H.V. Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution. *IEEE Signal Process. Lett.* **2012**, *19*, 876–879.
12. Farshchian, M.; Posner, F.L. The Pareto distribution for low grazing angle and high resolution X-band sea clutter. In Proceedings of the 2010 IEEE Radar Conference, Washington, DC, USA, 10–14 May 2010.
13. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; AMS: Oxford, UK, 2000.
14. Cao, L.; Sun, H.; Wang, X. The geometric structures of the Weibull distribution manifold and the generalized exponential distribution manifold. *Tamkang J. Math.* **2008**, *39*, 45–51.
15. Zhang, Z.; Sun, H.; Zhong, F. Information geometry of the power inverse Gaussian distribution. *Appl. Sci.* **2007**, *9*, 194–203.
16. Al-Kutubi, H.S.; Ibrahim, N.A. Bayes estimator for exponential distribution with extension of Jeffery prior information. *Malays. J. Math. Sci.* **2009**, *3*, 282–282.
17. Li, M.; Sun, H.; Peng, L. Fisher-Rao geometry and Jeffreys prior for Pareto distribution. *Commun. Stat. Theory Methods* **2020**, 1–16, doi:10.1080/03610926.2020.1771593.
18. Peng, L.; Sun, H.; Jiu, L. The geometric structure of the Pareto distribution. *Bol. Asoc. Mat. Venez.* **2007**, *14*, 5–13.
19. Dibal, N.P.; Mobolaji, A.T.; Musa, Y.A. Bayes' estimators of an exponentially distributed random variables using Al-Bayyati's loss function. *Int. J. Entific Res. Publ. (IJSRP)* **2019**, *9*, 674–84.
20. Contreras-Reyes, J.E.; Quintero, F.O.L.; Wiff, R. Bayesian Modeling of Individual Growth Variability Using Back-calculation: Application to Pink Cusk-eel (*Genypterus Blacodes*) off Chile. *Ecol. Model.* **2018**, *385*, 145–153.