Article

Information Geometry of the Exponential Family of Distributions with Progressive Type-II Censoring

Fode Zhang [1][ID]**, Xiaolin Shi** [2][ID] **and Hon Keung Tony Ng** [3,*][ID]

1   Center of Statistical Research, School of Statistics, Southwestern University of Finance and Economics, Chengdu 611130, China; fredzh@swufe.edu.cn
2   School of Electronics Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121, China; linda20016@163.com
3   Department of Statistical Science, Southern Methodist University, Dallas, TX 75275-0332, USA
*   Correspondence: ngh@mail.smu.edu

Abstract: In geometry and topology, a family of probability distributions can be analyzed as the points on a manifold, known as statistical manifold, with intrinsic coordinates corresponding to the parameters of the distribution. Consider the exponential family of distributions with progressive Type-II censoring as the manifold of a statistical model, we use the information geometry methods to investigate the geometric quantities such as the tangent space, the Fisher metric tensors, the affine connection and the α-connection of the manifold. As an application of the geometric quantities, the asymptotic expansions of the posterior density function and the posterior Bayesian predictive density function of the manifold are discussed. The results show that the asymptotic expansions are related to the coefficients of the α-connections and metric tensors, and the predictive density function is the estimated density function in an asymptotic sense. The main results are illustrated by considering the Rayleigh distribution.

Keywords: information geometry; exponential family of distributions; progressively type-II censoring scheme; Bayesian prediction; asymptotic expansions

MSC: 62N05; 62C10; 62E20; 53B05

1. Introduction

From the geometrical viewpoint, a parametric statistical model can be considered a differentiable manifold, and the parameter space can be regarded as a coordinate system of the manifold [1,2]. Let $\mathscr{F} = \{f(x;\theta), \theta \in \Theta\}$ be a parametric statistical model with respect to some σ-finite reference measure μ, where θ is a real k-dimensional parameter vector belonging to some open subset Θ of the k-dimensional real space $\mathbb{R}^k$. For simplicity, a random variable X and its observed value x are uniformly denoted by x in this paper.

When the density function $f(x;\theta)$ is sufficiently smooth in θ and it is differentiable as a function of θ, it is natural to introduce the structure of an k-dimensional manifold in the statistical model $\mathscr{F}$, where θ plays the role of a coordinate system. The geometrical quantities, such as connection, divergence, flatness, curvature and tangent space, play a fundamental role in the statistical inference and asymptotic theory (see, for example, Komaki [3,4] and Harsha and Moosath [5]).

In reliability engineering, a life testing experiment is one of the effective ways to obtain reliability information of a product. To save time and reduce the cost of a life testing experiment, censoring methodologies are often applied so that the experiment is terminated before all the items on the life-testing experiment fail. Some commonly used censoring schemes include the Type-I and Type-II censoring schemes, where the life-testing experiment will be terminated at a prefixed time point and the life-testing experiment will be terminated as soon as the m-th (m is prefixed) failure is observed, respectively. In other

words, the experimental time is prefixed for Type-I censoring scheme and the number of observed failures is prefixed for the Type-II censoring scheme (see, for example, Ng [6]). The Type-I and Type-II censoring schemes have been generalized to a more complicated and flexible censoring scheme such as progressive censoring schemes [7–9] and hybrid censoring schemes [10,11]. For progressive Type-II censoring schemes, the conventional Type-II censoring scheme is extended to situations wherein censoring occurs in multiple stages. A progressive Type-II censored life-testing experiment will be carried out in the following manner. Suppose n items are placed on a life-testing experiment and we assume that these n items have lifetimes following distribution with density function $f(x;\theta)$. It is planned that m failures will be observed and R_r items are randomly removed (i.e., censored) from the experiment at the time of the r-th failure. More specifically, at the time of the first failure (denoted by $X_{1:m:n}$), R_1 randomly selected items from the $n-1$ surviving items are removed from the life testing experiment; then, the experiment continues and at the time of the second failure (denoted by $X_{2:m:n}$), R_2 randomly selected items from the $(n-R_1-2)$ surviving items are removed from the experiment, and so on; finally, at the time of the m-th item failure (denoted by $X_{m:m:n}$), the experiment terminates and all the remaining $R_m = n - m - \sum_{r=1}^{m-1} R_r$ surviving items are censored. Here, $\mathcal{R} = (R_1, R_2, \ldots, R_m)$ is the progressive Type-II censoring scheme for the life testing experiment with $\sum_{r=1}^{m} R_r = n - m$. Note that, when $R_1 = R_2 = \ldots = 0$, $R_m = n - m$, the progressive Type-II censoring scheme reduces to the conventional Type-II censoring scheme.

Since the comprehensive studies of information geometry by Amari [1], information geometry has been productively used in different research fields including statistical learning, machine learning, neural networks, signal processing, information theory and so on (see, for example, Amari et al. [2] and Amari [12].) The information geometry methods are also widely used in statistics and reliability engineering. For example, Zhang et al. [13] discussed the Amari-Chentsov structure on the accelerated life test model with applications to optimal designs with different optimal criteria. The methods of information geometry are also employed to investigate the Bayesian prediction by taking α-divergences as loss functions [14]. In degradation modeling, a robust parameter estimation method was proposed in [15] by minimizing the f-divergence between the true model and suggested models.

In this paper, we investigate the tangent space, affine connection, α-connection, torsion and Riemann-Christoffel curvature of the manifold of the exponential family of distributions with progressive Type-II censoring scheme. These geometric quantities can be applied to different areas in statistics such as Bayesian analysis. Note that one of the challenges of Bayesian analysis is to calculate the integrals for obtaining the posterior distribution, especially when the number of parameters is large. Instead of using numerical methods to approximate those integrals, the geometric quantities developed in this paper can provide an efficient theoretical method to approximate those integrals involved in Bayesian prediction. The main contributions and the organization of this paper are described as follows:

- Asymptotic theory plays an important role in statistical inference, which consider the properties of statistical procedures as the sample size increases. Geometrically, an approximation to a manifold is a local linearization by the tangent space. Thus, the tangent space of the manifold of the exponential family of distributions with progressively Type-II censored data is discussed in Section 2.
- The local linearization accounts only for local properties of a statistical model. It is necessary to investigate the Fisher metric tensors, affine connection, and α-connection of the manifold in order to study the global or large-scale properties of the model. Therefore, these important geometric quantities are studied in Section 3.
- As an application of the geometric quantities, the asymptotic expansions of the posterior density and the posterior Bayesian predictive density of the model are provided in Section 4.
- To illustrate the results presented in this paper, the Rayleigh distribution is considered as an example in Section 5. Moreover, Monte Carlo simulation results and a real data analysis are presented in Section 6 to illustrate the main results.

2. The Statistical Model and Tangent Space

In this paper, we adopt the Einstein summation convention, that is, if an index occurs both as a superscript and as a subscript in a single expression, then the summation over that index is implied. For a density function $f(x;\theta) \in \mathscr{F}$, let $l(x;\theta) = \log f(x;\theta)$, we introduce the following definitions (see [1,2] for more details):

- $g_{ij} \overset{\text{def}}{=} E[\partial_i l(x;\theta)\partial_j l(x;\theta)]$: the Fisher metric tensors, the inverse of g_{ij} is denoted by g^{ij}, where $\partial_i = \partial/\partial\theta^i$;

- $T_{ijk} \overset{\text{def}}{=} E[\partial_i l(x;\theta)\partial_j l(x;\theta)\partial_k l(x;\theta)]$: the skewness tensor;

- $\Gamma_{ijk} \overset{\text{def}}{=} E[\partial_i \partial_j l(x;\theta)\partial_k l(x;\theta)]$: the affine connection;

- $\Gamma_{ijk}^\alpha \overset{\text{def}}{=} \Gamma_{ijk} + \frac{1-\alpha}{2}T_{ijk}$: the α-connection.

The -1-connection and 1-connection are said to be the m-connection and e-connection, denoted by Γ_{ijk}^m and Γ_{ijk}^e, respectively. We also abbreviate some geometric terms by multiplication the metric tensors, i.e., $T_i = T_{ijk}g^{jk}$, $\Gamma_{ij}^l = \Gamma_{ijk}g^{kl}$, $\Gamma_{ij}^{\alpha,l} = \Gamma_{ijk}^\alpha g^{kl}$.

Suppose that $\mathscr{F} = \{f(x,\theta), \theta \in \Theta\}$ is an exponential family of distributions (see, for example, Barndorff-Nielsen [16]) with density function

$$f(x,\theta) = \exp\left\{\sum_{i=1}^{\tau} \alpha^i(\theta)c_i(x) - \psi(\theta)\right\} \tag{1}$$

and reliability function

$$R(x,\theta) = 1 - F(x,\theta) = \exp\left\{\sum_{i=1}^{\tau} \beta^i(\theta)d_i(x) - \phi(\theta)\right\}, \tag{2}$$

where τ is the number of functions of the parameter vector θ, $F(x,\theta)$ is the cumulative distribution function, and $\psi(\theta)$ is the cumulant generating function defined as

$$\exp\{\psi(\theta)\} = \int \exp\left\{\sum_{i=1}^{\tau} \alpha^i(\theta)c_i(x)\right\}\mu(dx),$$

with $\alpha^i(\theta)$ and $\beta^i(\theta)$ are smooth functions of the parameter vector θ, and c_i and d_i are smooth functions of the random variable x. Here are two examples, the exponential and the Rayleigh distributions, of the members in the exponential family of distributions:

- Exponential distribution with density function

$$f(x;\lambda) = \exp\{-\lambda x + \ln \lambda\}, x > 0, \lambda > 0,$$

  and reliability function

$$R(x;\lambda) = \exp\{-\lambda x\}, x > 0, \lambda > 0,$$

  we have $\tau = 1$ and the functions $\alpha^1(\lambda) = \beta^1(\lambda) = -\lambda$ and $c_1(x) = d_1(x) = x$, $\psi(\theta) = -\ln\lambda$ and $\phi(\theta) = 0$. The dimension of the parameter vector θ is $k = 1$.

- Rayleigh distribution with density function

$$f(x;\lambda) = \exp\{-\lambda x^2 + \ln x + \ln 2\lambda\}, x > 0, \lambda > 0, \tag{3}$$

  and reliability function

$$R(x;\lambda) = \exp\{-\lambda x^2\}, x > 0, \lambda > 0,$$

We have $\tau = 2$ and the functions $\alpha^1(\lambda) = \beta^1(\lambda) = -\lambda$, $\alpha^2(\lambda) = 1$, $\beta^2(\lambda) = 0$, $c_1(x) = d_1(x) = x^2$, $c_2(x) = \ln x$, $d_2(x) = 0$, $\psi(\theta) = -\ln 2\lambda$ and $\phi(\theta) = 0$. The dimension of the parameter vector θ is $k = 1$.

Consider the life-testing experiment with progressive Type-II censoring described in Section 1 with n items placed on the life testing experiment and m failures are planned to be observed, let the set of all admissible Type-II PCSs as

$$PC(m, n) = \left\{ \mathcal{R} = (R_1, \dots, R_m) \in \mathbb{N}_0^m \,\middle|\, \sum_{i=1}^{m} R_i = n - m \right\},$$

where $\mathbb{N}_0$ is the set of the non-negative integers. Under a given censoring scheme $\mathcal{R} = (R_1, \dots, R_m) \in PC(m, n)$, the set of progressively Type-II censored order statistics is denoted as $x_{m:n}^{\mathcal{R}} = \{ x_{1:m:n}^{\mathcal{R}}, \dots, x_{m:m:n}^{\mathcal{R}} \}$. The PCS $\mathcal{R} = (R_1, \dots, R_m)$ is prefixed prior to starting the life testing experiment.

Suppose the lifetime distribution of the items in the life testing experiment follows a distribution in the exponential family of distributions with density function in Equation (1), the joint density function of the observed data, $x_{m:n}^{\mathcal{R}}$, can be expressed as [8,10]

$$
\begin{aligned}
L(x_{m:n}^{\mathcal{R}}; \theta) &= c(\mathcal{R}) \prod_{r=1}^{m} f(x_{r:m:n}^{\mathcal{R}}, \theta) \left(1 - F(x_{r:m:n}^{\mathcal{R}}, \theta) \right)^{R_r} \\
&= c(\mathcal{R}) \prod_{r=1}^{m} \exp\left\{ \sum_{i=1}^{\tau} \alpha^i(\theta) c_i(x_{r:m:n}^{\mathcal{R}}) + \sum_{i=1}^{\tau} R_r \beta^i(\theta) d_i\left(x_{r:m:n}^{\mathcal{R}} \right) - \psi(\theta) - R_r \phi(\theta) \right\} \\
&\stackrel{\text{def}}{=} c(\mathcal{R}) \prod_{r=1}^{m} \exp\left\{ \sum_{i=1}^{\tau} \theta^i e_i\left(x_{r:m:n}^{\mathcal{R}} \right) - \varphi(\theta) \right\} \\
&= c(\mathcal{R}) \exp\left\{ \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i e_i\left(x_{r:m:n}^{\mathcal{R}} \right) - m\varphi(\theta) \right\},
\end{aligned}
\tag{4}
$$

where $\theta^i e_i(x_{r:m:n}^{\mathcal{R}}) \stackrel{\text{def}}{=} \alpha^i(\theta) c_i(x_{r:m:n}^{\mathcal{R}}) + R_r \beta^i(\theta) d_i(x_{r:m:n}^{\mathcal{R}})$, $\varphi(\theta) \stackrel{\text{def}}{=} \psi(\theta) + R_r \phi(\theta)$, and

$$c(\mathcal{R}) = n(n - R_1 - 1) \dots (n - R_1 - R_2 - \dots - R_{m-1} - m + 1)$$

is the normalizing constant. By defining a new random variables

$$x_{i;r:m:n}^{\mathcal{R}} = e_i(x_{r:m:n}^{\mathcal{R}}),$$

the joint density function in Equation (4) can be expressed as

$$L(x_{m:n}^{\mathcal{R}}; \theta) = c(\mathcal{R}) \exp\left\{ \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i x_{i;r:m:n}^{\mathcal{R}} - m\varphi(\theta) \right\}.$$

The parameter θ of this form is called the natural parameter of the joint density function of the exponential family of distributions with progressive Type-II censoring.

The tangent vector T_θ of the manifold of the function $L(x_{m:n}^{\mathcal{R}}; \theta)$ is spanned by the vectors $\partial_i = \partial / \partial \theta^i$, and the set $\{\partial_i\}$ is called the natural basis associated with the coordinate system θ. Let

$$l(x_{m:n}^{\mathcal{R}}; \theta) = \log\left[L(x_{m:n}^{\mathcal{R}}; \theta) \right] = \log c(\mathcal{R}) + \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i e_i\left(x_{r:m:n}^{\mathcal{R}} \right) - m\varphi(\theta),$$

and the set

$$T_\theta^{(1)} = \left\{ A(x_{m:n}^{\mathcal{R}}) \,|\, A(x_{m:n}^{\mathcal{R}}) = span\left\{ \partial_i l(x_{m:n}^{\mathcal{R}}; \theta) \right\} \right\}$$

be the linear space of random variables spanned by $\partial_i l(x_{m:n}^{\mathcal{R}}; \theta)$. The space $T_{\theta}^{(1)}$ is called the 1-representation of the tangent space with progressively Type-II censored data. Here, the basis $\partial_i l(x_{m:n}^{\mathcal{R}}; \theta)$ of the 1-representation is given by

$$\partial_i l(x_{m:n}^{\mathcal{R}}; \theta) = \sum_{r=1}^{m} e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_i \varphi(\theta), \tag{5}$$

and the second and the third order derivatives of $l(x_{m:n}^{\mathcal{R}}; \theta)$ are given by

$$\partial_i \partial_j l(x_{m:n}^{\mathcal{R}}; \theta) = -m\partial_i \partial_j \varphi(\theta), \quad \partial_i \partial_j \partial_k l(x_{m:n}^{\mathcal{R}}; \theta) = -m\partial_i \partial_j \partial_k \varphi(\theta). \tag{6}$$

3. The α-Connections of Manifold Model

In this section, we investigate the α-connection of the manifold of the statistical model for the exponential family of distributions with progressively Type-II censored data. From Equation (4), the normalization factor $\varphi(\theta)$ can be defined as

$$\varphi(\theta) = \frac{1}{m} \log \int c(\mathcal{R}) \exp\left\{ \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i e_i\left(x_{r:m:n}^{\mathcal{R}}\right) \right\} \mu\left(dx_{r:m:n}^{\mathcal{R}}\right).$$

Since the function under the integral is assumed to be continuously differentiable, the order of integration and differentiation can be switched, and hence, the first three derivatives of the function $\varphi(\theta)$ with respect to the natural parameter θ^i are given by

$$\partial_i \varphi(\theta) = \frac{1}{m} \sum_{r=1}^{m} E_L\left[e_i\left(x_{r:m:n}^{\mathcal{R}}\right) \right], \tag{7}$$

$$\partial_i \partial_j \varphi(\theta) = \frac{1}{m} E_L\left[\left(\sum_{r=1}^{m} e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_i \varphi(\theta) \right) \left(\sum_{r=1}^{m} e_j\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_j \varphi(\theta) \right) \right]$$

$$= \frac{1}{m} E_L\left[\partial_i l(x_{m:n}^{\mathcal{R}}; \theta) \partial_j l(x_{m:n}^{\mathcal{R}}; \theta) \right], \tag{8}$$

$$\partial_i \partial_j \partial_k \varphi(\theta) = \frac{1}{m} E_L\left[\left(\sum_{r=1}^{m} e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_i \varphi(\theta) \right) \left(\sum_{r=1}^{m} e_j\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_j \varphi(\theta) \right) \right.$$

$$\left. \times \left(\sum_{r=1}^{m} e_k\left(x_{r:m:n}^{\mathcal{R}}\right) - m\partial_k \varphi(\theta) \right) \right]$$

$$= \frac{1}{m} E_L\left[\partial_i l(x_{m:n}^{\mathcal{R}}; \theta) \partial_j l(x_{m:n}^{\mathcal{R}}; \theta) \partial_k l(x_{m:n}^{\mathcal{R}}; \theta) \right], \tag{9}$$

where the expectations $E_L[\cdot]$ are taken with respect to the joint density function in Equation (4). The derivatives in Equations (7)–(9) can be considered as the expected value, the covariance and the third-order central moments of $\sum_{r=1}^{m} e_i\left(x_{r:m:n}^{\mathcal{R}}\right)$, respectively. The derivative in Equation (7) can also be obtained from the condition

$$E_L[\partial_i l(x_{m:n}^{\mathcal{R}}; \theta)] = 0,$$

that is

$$0 = \partial_i \int L(x_{m:n}^{\mathcal{R}}; \theta) \mu\left(dx_{m:n}^{\mathcal{R}}\right) = \int \partial_i L(x_{m:n}^{\mathcal{R}}; \theta) \mu\left(dx_{m:n}^{\mathcal{R}}\right) = E_L[\partial_i l(x_{m:n}^{\mathcal{R}}; \theta)].$$

The derivatives in Equations (8) and (9) can be obtained by calculating, respectively,

$$E_L[\partial_i \partial_j l(x_{m:n}^{\mathcal{R}}; \theta)] \quad \text{and} \quad E_L[\partial_i \partial_j \partial_k l(x_{m:n}^{\mathcal{R}}; \theta)].$$

Equations (8) and (9) show that the (i, j) element of the metric tensors is given by

$$g_{ij}(\theta) = E_L\left[\partial_i l(x_{m:n}^{\mathcal{R}}; \theta) \partial_j l(x_{m:n}^{\mathcal{R}}; \theta) \right] = m\partial_i \partial_j \varphi(\theta),$$

the (i, j, k) element of the skewness tensor is given by

$$T_{ijk}(\theta) = E_L\left[\partial_i l(x_{m:n}^{\mathcal{R}};\theta)\partial_j l(x_{m:n}^{\mathcal{R}};\theta)\partial_k l(x_{m:n}^{\mathcal{R}};\theta)\right] = m\partial_i\partial_j\partial_k\varphi(\theta),$$

and the (i, j, k) element of the affine connection is given by

$$\Gamma_{ijk}(\theta) = E_L\left[\partial_i\partial_j l(x_{m:n}^{\mathcal{R}};\theta)\partial_k l(x_{m:n}^{\mathcal{R}};\theta)\right] = -m\partial_i\partial_j\varphi(\theta)E_L\left[\partial_k l(x_{m:n}^{\mathcal{R}};\theta)\right] = 0.$$

Therefore, based on the joint density function $L(x_{m:n}^{\mathcal{R}};\theta)$, the α-connection of the manifold of an exponential family of distributions is given by

$$\Gamma_{ijk}^{\alpha}(\theta) = \frac{(1-\alpha)m}{2}\partial_i\partial_j\partial_k\varphi(\theta),$$

which means that the natural parameter θ is 1-affine, i.e., $\Gamma_{ijk} = 0$. Based on the information carried by the joint density function in Equation (4), we can obtain the following results.

Theorem 1. *The metric tensors and the α-connection of the exponential family of distributions are given by*

$$\begin{aligned}
g_{ij}(\theta) &= m\partial_i\partial_j\varphi(\theta) \\
\text{and } \Gamma_{ijk}^{\alpha}(\theta) &= \frac{(1-\alpha)m}{2}\partial_i\partial_j\partial_k\varphi(\theta),
\end{aligned}$$

respectively.

From the α-connection, we can obtain the torsion and the Riemann-Christoffel curvature of the manifold. The torsion is represented by the torsion tensor whose components are given by [1,2]

$$S_{ijk}(\theta) = \Gamma_{ijk}(\theta) - \Gamma_{jik}(\theta),$$

which is a tensor anti-symmetric with respect to indices i, j. Note that the coefficients $\Gamma_{ijk}^{\alpha}(\theta)$ of the α-connections are symmetric with respect to the first two indices i and j, then the tensor $S_{ijk}(\theta)$ vanishes for any α-connection. This shows that the manifold of the statistical model of the exponential family of distributions with progressively Type-II censored data is torsion-free.

The Riemann-Christoffel curvature of the manifold can be obtained as [1,2]

$$R_{ijkm} = (\partial_i\Gamma_{jk}^s - \partial_j\Gamma_{ik}^s)g_{sm} + (\Gamma_{irm}\Gamma_{jk}^r - \Gamma_{jrm}\Gamma_{ik}^r),$$

where $\Gamma_{ij}^k = g^{km}\Gamma_{ijm}$. The Riemann-Christoffel curvature based on the α-connection is called the α-Riemann-Christoffel curvature and its tensor is given by

$$R_{ijkm}^{\alpha} = (\partial_i\Gamma_{jk}^{\alpha,s} - \partial_j\Gamma_{ik}^{\alpha,s})g_{sm} + (\Gamma_{irm}^{\alpha}\Gamma_{jk}^{\alpha,r} - \Gamma_{jrm}^{\alpha}\Gamma_{ik}^{\alpha,r}),$$

where $\Gamma_{ij}^{\alpha,k} = g^{km}\Gamma_{ijm}^{\alpha}$. The tangent space of the manifold is said to be α-flat if the α-Riemann-Christoffel curvature $R_{ijkm}^{\alpha} = 0$. We can also obtain the α-covariant derivative and the Laplace operator based on the α-connection and the metric tensors.

In the above process for obtaining those geometric quantities, we only use the information from the joint density function $L(x_{m:n}^{\mathcal{R}};\theta)$. There is, in fact, another kind of information in the progressively Type-II censored order statistics $x_{r:m:n}^{\mathcal{R}}$ $(r = 1, \ldots, m)$. We can consider the marginal density function of the r-th progressively Type-II censored order statistic,

$x^{\mathcal{R}}_{r:m:n}$ (see, for example, Kamps and Cramer [17], Balakrishnan [18], Balakrishnan and Aggarwala [8], and Balakrishnan and Cramer [10])

$$f_{x^{\mathcal{R}}_{r:m:n}}(x) = c_{r-1} \sum_{s=1}^{r} a_{s,r} f(x)(1 - F(x))^{\gamma_s - 1}, x > 0, r = 1, \ldots, m, \tag{10}$$

where

$$\gamma_s = n - s + 1 + \sum_{r=1}^{s-1} R_r \text{ for } s = 1, \ldots, m,$$

$$c_{r-1} = \prod_{s=1}^{r} \gamma_s \text{ for } r = 1, \ldots, m,$$

$$a_{s,r} = \prod_{k=1,k\neq s}^{r} \frac{1}{\gamma_k - \gamma_s} \text{ for } 1 \leq s \leq r \leq m \text{ with } a_{1,1} = 1.$$

Based on the marginal density in Equation (10), the expectations of the random variables $e_i(x^{\mathcal{R}}_{r:m:n})$ $(r = 1, \ldots, m)$ in Equation (5) can be obtained as

$$
\begin{aligned}
h_{i,r}(\theta) &\stackrel{\text{def}}{=} E_f[e_i(x^{\mathcal{R}}_{r:m:n})] = \int e_i(x) f_{x^{\mathcal{R}}_{r:m:n}}(x)\mu(dx) \\
&= c_{r-1} \sum_{s=1}^{r} a_{s,r} \int e_i(x) \exp\left\{ \sum_{i=1}^{\tau} \theta^i e_{i,\gamma_s}(x) - \varphi_{\gamma_s}(\theta) \right\} \mu(dx),
\end{aligned}
$$

where $\theta^i e_{i,\gamma_s}(x) - \varphi_{\gamma_s}(\theta) = \alpha^i(\theta)c_i(x) + (\gamma_s - 1)\beta^i(\theta)d_i(x) - \psi(\theta) - (\gamma_s - 1)\phi(\theta)$, the expectation $E_f[\cdot]$ is taken with respect to the density function in Equation (10). Suppose that the random variables $e_i(x)$ $(r = 1, \ldots, m)$ are independent, and let

$$h_i(\theta) = \sum_{r=1}^{m} h_{i,r}(\theta) = \sum_{r=1}^{m} \sum_{s=1}^{r} c_{r-1} a_{s,r} \int e_i(x) \exp\left\{ \sum_{i=1}^{\tau} \theta^i e_{i,\gamma_s}(x) - \varphi_{\gamma_s}(\theta) \right\} \mu(dx),$$

we can obtain the following results.

Theorem 2. *The Fisher metric tensors and the α-connection of the exponential family of distribution with progressively Type-II censored data are given by*

$$
\begin{aligned}
\tilde{g}_{ij}(\theta) &= g_{ij}(\theta) = m\partial_i\partial_j\varphi(\theta), \\
\tilde{\Gamma}^{\alpha}_{ijk}(\theta) &= g_{ij}(\theta)(m\partial_k\varphi(\theta) - h_k(\theta)) + \frac{1-\alpha}{2} \prod_{l=i,j,k} (h_l(\theta) - m\partial_l\varphi(\theta)),
\end{aligned}
$$

respectively.

Proof. For the metric tensors, they can be obtained directly from the definition of $g_{ij}(\theta)$. For the affine connection, from $E[\partial_i l(x^{\mathcal{R}}_{m:n}; \theta)] = \sum_{r=1}^{m} E_f[e_i(x^{\mathcal{R}}_{r:m:n})] - m\partial_i\varphi(\theta) = h_i(\theta) - m\partial_i\varphi(\theta)$, Equations (7) and (8), we can obtain

$$
\begin{aligned}
\tilde{\Gamma}_{ijk}(\theta) &= E[\partial_i\partial_j l(x^{\mathcal{R}}_{m:n}; \theta)\partial_k l(x^{\mathcal{R}}_{m:n}; \theta)] \\
&= -m\partial_i\partial_j\varphi(\theta)E[\partial_k l(x^{\mathcal{R}}_{m:n}; \theta)] \\
&= g_{ij}(\theta)(m\partial_k\varphi(\theta) - h_k(\theta)).
\end{aligned}
$$

Then, the third-order tensor $T_{ijk}(\theta)$ can be specified as

$$\tilde{T}_{ijk}(\theta) = E[\partial_i l(x^{\mathcal{R}}_{m:n}; \theta)\partial_j l(x^{\mathcal{R}}_{m:n}; \theta)\partial_k l(x^{\mathcal{R}}_{m:n}; \theta)] = \prod_{l=i,j,k} (h_l(\theta) - m\partial_l\varphi(\theta)).$$

□

4. Applications in Bayesian Predictive Inference and Asymptotic Expansions

In Bayesian inference for the exponential family of distributions, the parameter vector θ is considered as a random variable. Given a prior density function for θ, $\pi(\theta)$, the joint posterior density function of the exponential family of distributions with progressively Type-II censored data can be expressed as

$$f_\pi(\theta|x_{m:n}^{\mathcal{R}}) = \frac{L(x_{m:n}^{\mathcal{R}};\theta)\pi(\theta)}{\int L(x_{m:n}^{\mathcal{R}};\theta)\pi(\theta)d\theta}, \tag{11}$$

and the posterior Bayesian predictive distribution is given by

$$\hat{f}_\pi(x|x_{m:n}^{\mathcal{R}}) = \int f(x;\theta)f_\pi(\theta|x_{m:n}^{\mathcal{R}})d\theta, \tag{12}$$

where x is an unobserved set of observations to be predicted and it is independently distributed according to the same density $f(x;\theta) \in \mathscr{F}$. The predictive density $\hat{f}(x|\hat{\theta})$ is called the plug-in density function or the estimative density function, where $\hat{\theta} = \hat{\theta}(x_{m:n}^{\mathcal{R}})$ is an estimate of θ based on the observed progressively Type-II censored sample $x_{m:n}^{\mathcal{R}}$ (see, for example, Geisser [19]). Consider the Kullback-Leibler divergence as the loss function, the predictive distribution in Equation (12) is the best predictive distribution in the sense that it minimizes the Bayes risk defined as [20]

$$\int \pi(\theta) \int f(x_{m:n}^{R};\theta) \int f(x;\theta) \log\left(\frac{f(x;\theta)}{\hat{f}(x|x_{m:n}^{R})}\right) \mu(dx)\, \mu(dx_{m:n}^{R})\, d\theta.$$

The integral defined in the predictive density in Equation (12) can be difficult to integrate or the form is too complicated to be used in practice. In these situations, asymptotic or large-sample theory (see, for example, Barndorff-Nielsen and Cox [21]) can be considered. In this section, we adopt the metric tensors and the α-connection introduced in Sections 2 and 3 to study the asymptotic expansions of the posterior joint density and the Bayesian predictive density of the exponential family of distributions with progressively Type-II censored data. A similar asymptotic expansion of Bayesian prediction based on a full sample can be found in Zhang et al. [14]. For simplicity, we only consider the information carried by the joint density function in Equation (4), a similar process can be applied for the situation where the information obtained from the joint density function in Equation (4) and the marginal density function in Equation (10) together.

Theorem 3. *Given a prior distribution $\pi(\theta)$ for θ, the posterior distribution in Equation (11) can be expressed asymptotically as*

$$\begin{aligned}
&f_\pi(\theta|x_{m:n}^{\mathcal{R}}) \\
&= \frac{\det(g_{ij}(\theta))}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}g_{ij}(\hat{\theta})\tilde{\theta}^i\tilde{\theta}^j\right\}\left(1 + \frac{1}{6}T_{ijk}(\hat{\theta})\tilde{\theta}^i\tilde{\theta}^j\tilde{\theta}^k + (\partial_i \log \pi(\hat{\theta}))\tilde{\theta}^i + o^{-1}(n)\right),
\end{aligned}$$

where $\tilde{\theta}^i = \theta^i - \hat{\theta}^i$ and $\hat{\theta}$ is an estimator of parameter set θ.

Proof. Using the Laplace method suggested by Nielsen and Cox [21], the posterior distribution can be expressed asymptotically as

$$\begin{aligned}
f_\pi(\theta|x_{m:n}^{\mathcal{R}}) &= \frac{\det(-\partial_i\partial_j l(x_{m:n}^{\mathcal{R}};\hat{\theta}))}{(2\pi)^{k/2}} \exp\left\{-\frac{1}{2}\left(-\partial_i\partial_j l(x_{m:n}^{\mathcal{R}};\hat{\theta})\right)\tilde{\theta}^i\tilde{\theta}^j\right\} \\
&\quad \times \left(1 + \frac{1}{6}\left(\partial_i\partial_j\partial_k l(x_{m:n}^{\mathcal{R}};\hat{\theta})\right)\tilde{\theta}^i\tilde{\theta}^j\tilde{\theta}^k + (\partial_i \log \pi(\hat{\theta}))\tilde{\theta}^i + o^{-1}(n)\right).
\end{aligned}$$

We have

$$-\partial_i \partial_j l(x_{m:n}^{\mathcal{R}}; \theta) = m \partial_i \partial_j \varphi(\theta) = g_{ij}(\theta), \quad -\partial_i \partial_j \partial_k l(x_{m:n}^{\mathcal{R}}; \theta) = m \partial_i \partial_j \partial_k \varphi(\theta) = T_{ijk}(\theta),$$

which implies that

$$\begin{aligned}
&f_\pi(\theta | x_{m:n}^{\mathcal{R}}) \\
&= \frac{\det(g_{ij}(\hat{\theta}))}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} g_{ij}(\hat{\theta}) \tilde{\theta}^i \tilde{\theta}^j \right\} \left(1 - \frac{1}{6} T_{ijk}(\hat{\theta}) \tilde{\theta}^i \tilde{\theta}^j \tilde{\theta}^k + (\partial_i \log \pi(\hat{\theta})) \tilde{\theta}^i + o^{-1}(n) \right).
\end{aligned}$$

$\square$

Based on the asymptotic expansion presented in Theorem 3, we can obtain the following result.

Theorem 4. *Given a prior distribution $\pi(\theta)$ for θ, the predictive distribution in Equation (12) can be expressed asymptotically as*

$$\begin{aligned}
\hat{f}_\pi\left(x | x_{m:n}^{\mathcal{R}} \right) &= f(x; \hat{\theta}) + \frac{1}{2n} g^{ij}(\hat{\theta}) \left\{ -\partial_i \partial_j \psi(\hat{\theta}) + \Gamma_{ij}^{m,l}(\hat{\theta})(c_l(x) - \partial_l \psi(\hat{\theta})) \right\} \\
&\quad + \frac{1}{n} \left(\partial_i \log \pi(\hat{\theta}) - \Gamma_{ij}^{e,j}(\hat{\theta}) \right) g^{il}(\hat{\theta})(c_l(x) - \partial_l \psi(\hat{\theta})) + o(n^{-1}).
\end{aligned}$$

Proof. The proof is similar to the proof of Theorem 2 in Komaki [3]. The proof can be completed by substituting $\partial_i \partial_j f(x; \hat{\theta})$ and $\partial_i f(x; \hat{\theta})$ with $-\partial_i \partial_j \psi(\hat{\theta})$ and $c_i(x) - \partial_i \psi(\hat{\theta})$, respectively. $\square$

If the prior distribution $\pi(\theta)$ is the Jeffreys prior $\pi_J(\theta) \propto |g_{ij}|^{\frac{1}{2}}$, then from the relationship

$$\partial_i \log \pi(\theta) = \partial_i \log |g_{ij}(\theta)|^{\frac{1}{2}} = \frac{1}{2} \partial_i g_{ij}(\theta) g^{ij}(\theta) = \Gamma_{ij}^j(\theta) = \Gamma_{ij}^{e,j}(\theta) + \frac{1}{2} T_i(\theta),$$

we have

$$\partial_i \log \pi(\theta) - \Gamma_{ij}^{e,j}(\theta) = \frac{1}{2} T_i(\theta).$$

The following results can be immediately obtained.

Corollary 1. *Given the Jeffreys prior $\pi_J(\theta) \propto |g_{ij}(\theta)|^{1/2}$, the posterior distribution in Equation (11) can be asymptotically expanded as*

$$\begin{aligned}
f_{\pi_J}(\theta | x_{m:n}^{\mathcal{R}}) &= \frac{\det(g_{ij}(\hat{\theta}))}{(2\pi)^{k/2}} \exp\left\{ -\frac{1}{2} g_{ij}(\hat{\theta}) \tilde{\theta}^i \tilde{\theta}^j \right\} \left(1 + \frac{1}{6} T_{ijk}(\hat{\theta}) \tilde{\theta}^i \tilde{\theta}^j \tilde{\theta}^k \right. \\
&\quad \left. + \left(\Gamma_{ij}^{e,j}(\hat{\theta}) + \frac{1}{2} T_i(\hat{\theta}) \right) \tilde{\theta}^i + o^{-1}(n) \right),
\end{aligned}$$

Corollary 2. *Given the Jeffreys prior $\pi_J(\theta) \propto |g_{ij}(\theta)|^{1/2}$, the prediction (12) can be asymptotically expanded as*

$$\begin{aligned}
\hat{f}_{\pi_J}\left(x | x_{m:n}^{\mathcal{R}} \right) &= f(x; \hat{\theta}) - \frac{1}{2n} g^{ij}(\hat{\theta}) \left\{ \partial_i \partial_j \psi(\hat{\theta}) + \Gamma_{ij}^{m,l}(\hat{\theta})(c_l(x) - \partial_l \psi(\hat{\theta})) \right\} \\
&\quad + \frac{1}{2n} T_i(\hat{\theta}) g^{il}(\hat{\theta})(c_l(x) - \partial_l \psi(\hat{\theta})) + o(n^{-1}).
\end{aligned}$$

These results show that the predictive density function, when the sample size n approaches infinity, is the estimative density function in the asymptotic sense.

5. Illustration Example

The illustration of the geometric quantities for exponential distribution has been provided in the literature (see, for example, [12]). In this section, we use the Rayleigh distribution, a member of the exponential family of distributions, presented in Section 2 as an example to illustrate our results. Suppose that $x_{m:n}^{\mathcal{R}}$ is the progressively Type-II censored order statistics form items with lifetimes follow the Rayleigh distribution with density function in Equation (3), then the joint density function of $x_{m:n}^{\mathcal{R}}$ can be expressed as

$$
\begin{aligned}
L(x_{m:n}^{\mathcal{R}}; \theta) &= c(\mathcal{R}) \prod_{r=1}^{m} f(x_{r:m:n}^{\mathcal{R}}, \theta) \left[1 - F(x_{r:m:n}^{\mathcal{R}}, \theta) \right]^{R_r} \\
&= c(\mathcal{R}) 2^m \exp\left\{ \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i e_i \left(x_{r:m:n}^{\mathcal{R}} \right) - m\varphi(\theta) \right\},
\end{aligned}
$$

where $\ell = 2$. $\theta^1 = \lambda$, $\theta^2 = 1$, $e_1\left(x_{r:m:n}^{\mathcal{R}}\right) = -(1 + R_r)\left(x_{r:m:n}^{\mathcal{R}}\right)^2$, $e_2\left(x_{r:m:n}^{\mathcal{R}}\right) = \ln\left(x_{r:m:n}^{\mathcal{R}}\right)$ and $\varphi(\theta) = \psi(\theta) = -\ln(\lambda)$. Let

$$
\begin{aligned}
l(x_{m:n}^{\mathcal{R}}; \theta) &= \log L(x_{m:n}^{\mathcal{R}}; \theta) \\
&= \log(c(\mathcal{R})2^m) + \sum_{r=1}^{m} \sum_{i=1}^{\tau} \theta^i e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - m\varphi(\theta).
\end{aligned}
$$

Then, the first three derivatives of the function $l(x_{m:n}^{\mathcal{R}}; \theta)$ can be obtained as

$$
\begin{aligned}
\partial_1 l(x_{m:n}^{\mathcal{R}}; \theta) &= -\sum_{r=1}^{m} (1 + R_r)\left(x_{r:m:n}^{\mathcal{R}}\right)^2 + \frac{m}{\lambda}, \\
\partial_1 \partial_1 l(x_{m:n}^{\mathcal{R}}; \theta) &= -\frac{m}{\lambda^2}, \\
\partial_1 \partial_1 \partial_1 l(x_{m:n}^{\mathcal{R}}; \theta) &= \frac{2m}{\lambda^3}.
\end{aligned}
$$

The maximum likelihood estimator (MLE) of the parameter λ can be derived as

$$
\hat{\lambda} = \frac{m}{\sum_{r=1}^{m} (1 + R_r)\left(x_{r:m:n}^{\mathcal{R}}\right)^2}.
$$

We first consider the information carried by the joint density in Equation (4). The metric tensors have one element, that is,

$$
g_{11}(\theta) = m\partial_1 \partial_1 \varphi(\theta) = \frac{m}{\lambda^2}. \tag{13}
$$

The skewness tensor can be written as

$$
T_{111}(\theta) = m\partial_1 \partial_1 \partial_1 \varphi(\theta) = -\frac{2m}{\lambda^3}. \tag{14}
$$

The affine connection and the α-connection can be obtained as

$$
\begin{aligned}
\Gamma_{111}(\theta) &= -m\partial_1 \partial_1 \varphi(\theta) E_L\left[\partial_1 l(x_{m:n}^{\mathcal{R}}; \theta)\right] = 0, \\
\Gamma_{111}^{\alpha}(\theta) &= \frac{(1-\alpha)m}{2} \partial_1 \partial_1 \partial_1 \varphi(\theta) = \frac{(\alpha-1)m}{\lambda^3},
\end{aligned}
$$

respectively. The coefficients of the *m*-connection and *e*-connection are

$$
\begin{aligned}
G_{11}^{m,1}(\theta) &= \Gamma_{111}^{m}(\theta)g^{11}(\theta) = -2/\lambda, \\
G_{11}^{e,1}(\theta) &= \Gamma_{111}^{e}(\theta)g^{11}(\theta) = 0,
\end{aligned}
$$

respectively. For Bayesian inference, we consider the Jeffreys prior for the parameter λ, i.e.,

$$\pi_J(\theta) \propto \sqrt{m}/\lambda,$$

then the posterior distribution of λ is

$$f_{\pi_J}(\theta|x_{m:n}^{\mathcal{R}}) = \frac{\exp\left\{\sum_{r=1}^{m}\sum_{i=1}^{\tau}\theta^i e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - (m-1)\varphi(\theta)\right\}}{\int \exp\left\{\sum_{r=1}^{m}\sum_{i=1}^{\tau}\theta^i e_i(x_{r:m:n}^{\mathcal{R}}) - (m-1)\varphi(\theta)\right\}d\theta},$$

which can be written as

$$
\begin{aligned}
&f_{\pi_J}(\theta|x_{m:n}^{\mathcal{R}})\\
&= \frac{\det(g_{ij}(\hat{\theta}))}{(2\pi)^{k/2}}\exp\left\{-\frac{1}{2}g_{ij}(\hat{\theta})\tilde{\theta}^i\tilde{\theta}^j\right\}\left(1 - \frac{1}{6}T_{ijk}(\hat{\theta})\tilde{\theta}^i\tilde{\theta}^j\tilde{\theta}^k\right.\\
&\quad\left. + \left(\Gamma_{ij}^{e,j}(\hat{\theta}) + \frac{1}{2}T_i(\hat{\theta})\right)\tilde{\theta}^i + o^{-1}(n)\right)\\
&= \frac{m}{\hat{\lambda}^2(2\pi)^{1/2}}\exp\left\{-\frac{m}{2\hat{\lambda}^2}(\lambda-\hat{\lambda})^2\right\}\left(1 + \frac{m}{3\hat{\lambda}^3}(\lambda-\hat{\lambda})^3 - \frac{1}{\hat{\lambda}}(\lambda-\hat{\lambda}) + o(n^{-1})\right).
\end{aligned}
$$

Here, the predictive distribution is

$$\hat{f}_{\pi_J}\left(x|x_{m:n}^{\mathcal{R}}\right) = c(\mathcal{R})2^m\int\exp\left\{\sum_{r=1}^{m}\sum_{i=1}^{\tau}\theta^i e_i\left(x_{r:m:n}^{\mathcal{R}}\right) - \theta^1 x^2 + \ln(x) - (m+1)\varphi(\theta)\right\}d\theta,$$

which can be expanded asymptotically as

$$
\begin{aligned}
\hat{f}_{\pi_J}\left(x|x_{m:n}^{\mathcal{R}}\right) &= f(x;\hat{\theta}) - \frac{1}{2n}g^{ij}(\hat{\theta})\left\{\partial_i\partial_j\psi(\hat{\theta}) + \Gamma_{ij}^{m,l}(\hat{\theta})(c_l(x) - \partial_l\psi(\hat{\theta}))\right\}\\
&\quad + \frac{1}{2n}T_i(\hat{\theta})g^{il}(\hat{\theta})(c_l(x) - \partial_l\psi(\hat{\theta})) + o(n^{-1})\\
&= f(x,\hat{\lambda}) - \frac{\hat{\lambda}^2}{2nm}\left(-\frac{1}{2\hat{\lambda}^2} - \frac{2}{\hat{\lambda}}\left(\frac{1}{2\hat{\lambda}} - x^2\right)\right)\\
&\quad + \frac{1}{2n}\left(-\frac{2m}{\hat{\lambda}^3}\right)\frac{\hat{\lambda}^2}{m}\frac{\hat{\lambda}^2}{m}\left(\frac{1}{2\hat{\lambda}} - x^2\right) + o(n^{-1})\\
&= f(x,\hat{\lambda}) + \frac{1}{4mn} + o(n^{-1}).
\end{aligned}
$$

In the following, we consider the information obtained from the marginal density function in Equation (10) and the joint density function in Equation (4) together. Notice that

$$
\begin{aligned}
f_{x_{r:m:n}^{\mathcal{R}}}(x) &= c_{r-1}\sum_{s=1}^{r}a_{s,r}f(x)(1-F(x))^{\gamma_s-1}\\
&= c_{r-1}\sum_{s=1}^{r}a_{s,r}2\lambda x\exp\left\{-\lambda\gamma_s\left(x_{r:m:n}^{\mathcal{R}}\right)^2\right\}, r = 1,\ldots,m.
\end{aligned}
$$

The n-order moment about the origin of the r-th progressively Type-II censored order statistic is given by

$$E_f\left[\left(x_{r:m:n}^{\mathcal{R}}\right)^n\right] = \int x^n f_{x_{r:m:n}^{\mathcal{R}}}(x)\mu(dx) = c_{r-1}\sum_{s=1}^{r}a_{s,r}\lambda\frac{\Gamma\left(\frac{n}{2}+1\right)}{(\lambda\gamma_s)^{\frac{n}{2}+1}},$$

which implies

$$E_f\left[\partial_1 l\left(x_{m:n}^{\mathcal{R}};\theta\right)\right] = -\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)c_{r-1}\frac{a_{s,r}}{\lambda\gamma_s^2} + \frac{m}{\lambda}.$$

Thus, the affine connection is specified as

$$\tilde{\Gamma}_{111}(\theta) = E_f\left[\partial_1\partial_1 l(x_{m:n}^{\mathcal{R}};\theta)\partial_1 l(x_{m:n}^{\mathcal{R}};\theta)\right] = \frac{m}{\lambda^3}\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)c_{r-1}\frac{a_{s,r}}{\gamma_s^2} - \frac{m^2}{\lambda^3}.$$

The metric tensor $\tilde{g}_{11}(\theta)$ and the skewness tensor $\tilde{T}_{111}(\theta)$ are the same as the expressions in Equations (13) and (14). The α-connection is reduces to

$$\tilde{\Gamma}_{111}^{\alpha}(\theta) = \frac{m}{\lambda^3}\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)c_{r-1}\frac{a_{s,r}}{\gamma_s^2} - \frac{m^2}{\lambda^3} - \frac{(1-\alpha)m}{\lambda^3}.$$

The coefficients of the m-connection and the e-connection are

$$\tilde{\Gamma}_{11}^{m,1}(\theta) = \frac{1}{\lambda}\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)c_{r-1}\frac{a_{s,r}}{\gamma_s^2} - \frac{2+m}{\lambda},$$

$$\text{and } \tilde{\Gamma}_{11}^{e,1}(\theta) = \frac{1}{\lambda}\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)c_{r-1}\frac{a_{s,r}}{\gamma_s^2} - \frac{m}{\lambda},$$

respectively. Therefore, based on the Jeffreys prior $\pi_J(\theta) \propto \sqrt{m}/\lambda$, the Bayesian predictive density function of the Rayleigh distribution with progressively Type-II censored data can be asymptotically expanded as

$$
\begin{aligned}
\hat{f}_{\pi_J}\left(x|x_{m:n}^{\mathcal{R}}\right) &= f(x;\hat{\theta}) - \frac{1}{2n}\tilde{g}^{ij}(\hat{\theta})\left\{\partial_i\partial_j\psi(\hat{\theta}) + \tilde{\Gamma}_{ij}^{m,l}(\hat{\theta})(c_l(x) - \partial_l\psi(\hat{\theta}))\right\} \\
&\quad + \frac{1}{2n}\tilde{T}_i(\hat{\theta})g^{il}(\hat{\theta})(c_l(x) - \partial_l\psi(\hat{\theta})) + o(n^{-1}) \\
&= f(x,\hat{\lambda}) + \frac{1}{4mn}\left(1 + \left(2\hat{\lambda}x^2 - 1\right)\left(\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)\frac{c_{r-1}a_{s,r}}{\gamma_s^{-2}} - m\right)\right) \\
&\quad + o(n^{-1}).
\end{aligned}
$$

This shows that the predictive density function, with the increase of the sample size n and the observed sample size m, is the estimative density function in the asymptotic sense. The term

$$\frac{2\hat{\lambda}x^2 - 1}{4mn}\left(\sum_{r=1}^{m}\sum_{s=1}^{r}(1+R_r)\frac{c_{r-1}a_{s,r}}{\gamma_s^{-2}} - m\right)$$

can be considered the correction term due to the information carried by the density function in Equation (10).

6. Monte Carlo Simulation Study and Real Data Analysis

　　In this section, we present a Monte Carlo simulation study of the Bayesian prediction based on progressively Type-II censored data described in Section 4. We also present a real data analysis based on the progressive Type-II censored data discussed in the literature. In the Monte Carlo simulation study, we consider different sample sizes $(n, m) = (10, 30)$, $(10, 35)$, $(15, 40)$ and $(20, 40)$ and three different censoring schemes:

$\mathcal{R}_1$: $R_1 = R_2 = \cdots = R_{m-1} = 0, R_m = n - m$;
$\mathcal{R}_2$: $R_1 = n - m, R_2 = R_3 = \cdots = R_m = 0$;
$\mathcal{R}_3$: $R_1 = \cdots = R_{m-1} = 1, R_m = n - 2m + 1$.

The progressively Type-II censored data, $x_{m:n}^{\mathcal{R}}$, are generated based on the Rayleigh distribution in Equation (3) with parameter $\lambda = 2$ for different sample sizes and censoring schemes. For the proposed Bayesian prediction (BP), we consider two different priors: (i) the Jeffreys prior $\pi_J(\theta) \propto \sqrt{m}/\lambda$; and (ii) the uniform prior π_I on interval $(0,3)$. For comparative purposes, we also consider the plug-in prediction (PP) approach in which the estimative density function $\hat{f}(x,\hat{\lambda})$ is also considered. For the plug-in approach, the parameter is estimated by using the maximum likelihood method based on the simulated progressive Type-II censored sample $x_{m:n}^{\mathcal{R}}$. The estimated biases and mean square errors (MSEs) of different prediction approaches for predicting the probability density at $x = 2.5$ based on 10,000 simulations are presented in Table 1.

From Table 1, we observe that the performances of all prediction methods improve in terms of MSEs as the sample sizes m and n increase. In other words, the number of items being removed during the progressively Type-II censored experiment affects the performance of prediction. Moreover, we observe that the Bayesian prediction method with the Jeffreys prior can provide smaller biases and smaller MSEs compared to the plug-in prediction method in some cases.

Table 1. Simulated biases and mean square errors (MSEs) of different prediction methods based on Rayleigh distribution with $\lambda = 2$.

(m,n)	Schemes	PP $\hat{f}(x;\hat{\lambda})$		BP $\hat{f}_{\pi_J}\left(x\|x_{m:n}^{\mathcal{R}}\right)$		BP $\hat{f}_{\pi_I}\left(x\|x_{m:n}^{\mathcal{R}}\right)$	
		Bias	**MSE**	**Bias**	**MSE**	**Bias**	**MSE**
	$\mathcal{R}_1$	−0.045	0.032	−0.026	0.027	−0.038	0.035
$(10,30)$	$\mathcal{R}_2$	0.035	0.031	0.024	0.023	−0.021	0.034
	$\mathcal{R}_3$	−0.034	0.032	−0.025	0.026	0.013	0.037
	$\mathcal{R}_1$	−0.030	0.029	−0.031	0.023	−0.016	0.033
$(10,35)$	$\mathcal{R}_2$	0.025	0.027	−0.020	0.022	0.028	0.031
	$\mathcal{R}_3$	−0.022	0.028	0.024	0.024	−0.027	0.032
	$\mathcal{R}_1$	−0.023	0.019	0.021	0.022	0.022	0.025
$(15,40)$	$\mathcal{R}_2$	−0.017	0.020	−0.013	0.018	−0.013	0.027
	$\mathcal{R}_3$	0.024	0.021	0.021	0.019	−0.023	0.023
	$\mathcal{R}_1$	−0.023	0.019	0.021	0.010	−0.026	0.014
$(20,40)$	$\mathcal{R}_2$	0.018	0.014	0.015	0.008	0.024	0.015
	$\mathcal{R}_3$	−0.013	0.017	−0.010	0.009	−0.017	0.012

To illustrate the practical applications of the approximate methods based on geometric quantities proposed in this paper, we analyze a real data set which contains the times to breakdown of an insulating fluid at 34 kV originally presented in Nelson [22] (Table 6.1). A progressively Type-II censored sample of size $m = 8$ generated from the $n = 19$ observations by Balakrishnan et al. [9] is analyzed here. The progressively censored sample and the progressive censoring scheme are presented in Table 2.

Table 2. Progressively Type-II censored sample of the times to breakdown data on insulating fluid tested at 34 KV with $n = 19$ and $m = 8$ obtained from Balakrishnan et al. [9]

r	1	2	3	4	5	6	7	8
$x_{r:n}$	0.19	0.78	0.96	1.31	2.78	4.85	6.50	7.35
R_r	0	0	3	0	3	0	0	5

Suppose that we assume the lifetimes of the insulating fluid tested at 34 kV follow a Rayleigh distribution and we are interested in predicting the probability density, based on the progressive Type-II censored data presented in Table 2, the predicted density curves obtained from the plug-in prediction approach and the proposed Bayesian prediction approach with two different priors are presented in Figure 1. From Figure 1, we observe that the three prediction methods provide similar predicted density curves in this case.

For instance, if we are interested in predicting density at $x = 2.8$, based on the progressive Type-II censored data presented in Table 2, the predicted values of plug-in prediction density $\hat{f}(x, \hat{\lambda})$ is 0.230, and the Bayesian prediction densities $\hat{f}_{\pi_J}(x|x_{m:n}^{\mathcal{R}})$ with Jeffreys prior π_J and $\hat{f}_{\pi_I}(x|x_{m:n}^{\mathcal{R}})$ uniform prior π_I are 0.229 and 0.232, respectively.
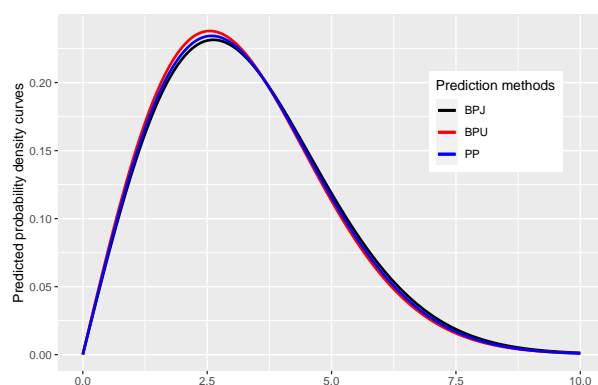

Figure 1. The predicted density curves of Rayleigh distribution obtained from the Bayesian prediction approach with Jeffreys prior (BPJ), with uniform prior to interval $(0, 3)$ (BPU), and the plug-in prediction (PP) approach based on the data presented in Table 2.

7. Conclusions

In this paper, we discussed the tangent space, affine connection, α-connection, torsion and Riemann-Christoffel curvature of statistical manifold induced by the exponential family of distributions. As applications of these geometric quantities, the asymptotic expansions of the Bayesian posterior distribution and prediction function with progressively Type-II censored data were discussed. The results showed that the asymptotic expansions are related to the geometric quantities. We also illustrated the main results by studying the Rayleigh distribution. Note that more theoretical results and applications of information geometry in reliability in addition to the main results of this paper can be found in the Ph.D. thesis [23].

References

1. Amari, S. *Differential-Geometrical Methods in Statistics*; Lecture Notes in Statistics; Springer: New York, NY, USA, 1985; Volume 28.
2. Amari, S.; Barndorff-Nielsen, O.E.; Kass, R.E.; Lauritzen, S.L.; Rao, C.R. *Differential Geometry in Statistical Inference. IMS Lecture Notes: Monograph Series 10*; Institute of Mathematical Statistics; IMS: Hayward, CA, USA, 1987.
3. Komaki, F. On asymptotic properties of predictive distributions. *Biometrika* **1996**, *83*, 299–313. [CrossRef]
4. Komaki, F. Asymptotic Properties of Bayesian Predictive Densities When the Distributions of Data and Target Variables are Different. *Bayesian Anal.* **2015**, *10*, 31–51. [CrossRef]
5. Harsha, K.V.; Moosath, K.S.S. Dually flat geometries of the deformed exponential family. *Phys. A* **2015**, *433*, 136–147.

6. Ng, H.K.T. Censoring Methodology. In *International Encyclopedia of Statistical Science*; Lovric, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
7. Algarni, A.; Almarashi, A.M.; Okasha, H.; Ng, H.K.T. E-Bayesian Estimation of Chen distribution Based on Type-I Censoring Scheme. *Entropy* **2020**, *22*, 636. [CrossRef] [PubMed]
8. Balakrishnan, N.; Aggarwala, R. *Progressive Censoring: Theory, Methods, and Applications*; Birkhäuser: Boston, MA, USA, 2000.
9. Viveros, R.; Balakrishnan, N. Interval estimation of parameters of life from progressively censored data. *Technometrics* **1994**, *36*, 84–91. [CrossRef]
10. Balakrishnan, N.; Cramer, E. *The Art of Progressive Censoring. Applications to Reliability and Quality*; Birkhäuser: New York, NY, USA, 2014.
11. Balakrishnan, N.; Kundu, D. Hybrid censoring: Models, inferential results and applications. *Comput. Stat. Data Anal.* **2013**, *57*, 166–209. [CrossRef]
12. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
13. Zhang, F.D.; Ng, H.K.T.; Shi, Y.M.; Wang, R.B. Amari-Chentsov structure on the statistical manifold of models for accelerated life tests. *TEST* **2019**, *28*, 77–105. [CrossRef]
14. Zhang, F.D.; Shi, Y.M.; Ng, H.K.T.; Wang, R.B. Information Geometry of Generalized Bayesian Prediction Using α-divergences as Loss Functions. *IEEE Trans. Inf. Theory* **2018**, *64*, 1812–1824. [CrossRef]
15. Zhang, F.D.; Ng, H.K.T.; Shi, Y.M. Mis-specifcation analysis of Wiener degradation models by using f-divergence with outliers. *Reliab. Eng. Syst. Saf.* **2020**, *195*, 106751 [CrossRef]
16. Barndorff-Nielsen, O.E. *Information and Exponential Families in Statistical Theory*; Wiley Series in Probability and Mathematical Statistics; John Wiley and Sons: Hoboken, NJ, USA, 1978.
17. Kamps, U.; Cramer, E. On distributions of generalized order statistics. *Statistics* **2001**, *35*, 269–280. [CrossRef]
18. Balakrishnan, N. Progressive censoring methodology: An appraisal (with discussions). *Test* **2007**, *16*, 211–296. [CrossRef]
19. Geisser, S. *Predictive Inference: An Introduction*; Chapman and Hall: New York, NY, USA, 1993.
20. Hartigan, J.A. The maximum likelihood prior. *Ann. Stat.* **1998**, *26*, 2083–2103. [CrossRef]
21. Barndorff-Nielsen, O.E.; Cox, D.R. *Inference and Asymptotics*; Chapman and Hall: London, UK, 1994.
22. Nelson, W. *Applied Life Data Analysis*; Wiley: New York, NY, USA, 1982.
23. Zhang, F.D. On the Information Geometry and Tsallis Statistics in the Reliability Analysis and Its Applications. Ph.D. Thesis, Northwestern Polytechnical University, Xi'an, China, 2017.