Bayesian prediction and density estimation for exponential mixtures under type-II censoring [1]

Éric Marchand[a] & Akbar Asgharzadeh[b]
a Département de mathématiques, Université de Sherbrooke, Sherbrooke Qc, CANADA, J1K 2R1
b Department of Statistics, University of Mazandaran, P.O. Box 47146-1407, Balbosar, IRAN
(e-mails: eric.marchand@usherbrooke.ca; a.asgharzadeh@umz.ac.ir)

Abstract

For lifetimes $X_1, \ldots, X_n$ distributed as a scale-mixture of exponentials, and more particularly for a gamma distributed mixing parameter such that $X_i|\beta \sim \text{Exp}(\beta)$ for $i = 1, \ldots, n$ and β is distributed as $\text{Gamma}(\alpha, \theta)$ with known α, we consider the prediction of future order statistics based on having observed the first m order statistics. Focus is placed on estimating the conditional density of the future order statistics given those observed, and in the study of Bayesian predictive densities to meet such an objective. For both Gamma distributed prior densities and the usual non-informative prior density π_0 for θ, we provide explicit representations of Bayesian predictive densities and HPD prediction regions for the vector of future order statistics. For prior π_0, the derived predictive density is multivariate Pareto and it is shown to be the same for all mixing parameter values α, as well as for the degenerate mixing case $X_i|\theta \sim Exp(\theta)$ i.i.d. Predictive distributions for individual order statistics, as well as for sums of future order statistic values, are also considered and shown to bring into play Pareto distributions and linear combinations of multivariate Pareto distributed vectors. Finally, we study the frequentist probability of coverage associated with the choice π_0, showing that it matches the given Bayesian credibility level, and moreover simultaneously so for all mixing parameter α, as well as for the degenerate mixing case $X_i|\theta \sim Exp(\theta)$ i.i.d.

AMS 2020 subject classifications: 62F15, 62N01, 62N05, 62C10, 62C20.

Keywords and phrases: Bayesian predictive density; Coverage probability; Credibility; Highest posterior density; Mixtures of exponentials; Multivariate Pareto; Order statistics; Prediction region; Type-II censoring.

1 Introduction

Consider an experiment with observable lifetimes $X_{(1)} < \cdots < X_{(m)}$, which consist of a fixed number m of lower order statistics, and the problem of predicting future lifetimes $X_{(m+1)} < \cdots < X_{(n)}$ based on $\tilde{X}_1 = (X_{(1)}, \ldots, X_{(m)})^\top$. Our focus is on multivariate prediction, which is central to contemporary statistical theory and practice (e.g., [4]). The described sampling scheme is well-known and referred to as type-II censoring. A common assumption is that the order statistics are drawn from i.i.d. $X_1, \ldots, X_n$, and a common

[1] February 3, 2026.

model is exponential with underlying and unknown rate parameter β. Bayesian prediction regions and predictive densities for future lifetimes have been studied by [10] and [5], among others in a univariate context (i.e., a single future lifetime), and also recently studied by [3] in the multivariate context (i.e., simultaneously for several future lifetimes).

However, it is also natural to consider a common or latent environment for the X_i's. i.e. the rate parameter β as random and unobservable. Specifying β as gamma distributed, such as $\beta \sim lG(\alpha, \theta)$ for known $\alpha > 0$ and unknown $\theta > 0$, is one such appealing choice and leads to the model

$$X_1, \ldots, X_n | \beta \sim \text{Exp}(\beta)\,, \text{ with } \beta \sim \mathcal{G}(\alpha, \theta)\,, \alpha > 0 \text{ (known), } \theta > 0 \text{ (unknown)}\,. \quad (1)$$

Such an assumption introduces a dependence structure for the $X_i' s$, with for instance a Pearson correlation coefficient equal to $\frac{1}{\alpha}$ for $\alpha > 2$ (independently of θ) between X_i and X_j for $i \neq j$. More importantly, the model assumptions result in a multivariate Lomax (or Pareto) distribution, with joint density on $\mathbb{R}_+^n$

$$f(x_1, \ldots, x_n | \theta) = (\alpha)_n \frac{\theta^\alpha}{\left(\theta + \sum_{i=1}^n x_i\right)^{\alpha+n}}\,, \quad (2)$$

with $(\alpha)_n = \frac{\Gamma(\alpha+n)}{\Gamma(\alpha)}$ as Pochhammer's symbol. Such a multivariate model has been studied by several researchers, namely in a context of reliability and inference (e.g., [12]), and others have also studied the joint distribution of the $X_i' s$ for other mixing distributions for β, such as inverse-Gaussian (e.g., [15]). However, the prediction of several future values, in the context of model density (2) or other dependent data models with type-II censoring, has not been considered before.

We consider here prediction of the future values or lifetimes $\tilde{X}_2 = (X_{(m+1)}, \ldots, X_{(n)})^\top$ based on an observed value $\tilde{x}_1$, and are drawn to focus on the estimation of the conditional density $p_\theta(\cdot|\tilde{x}_1)$ of $\tilde{X}_2$ given $\tilde{x}_1$. Here the observed value $\tilde{x}_1$ informs us on both the parameter θ and the form or the conditional density to be estimated. We adopt a Bayesian approach and provide Bayesian predictive densities for $p_\theta(\cdot|\tilde{x}_1)$ associated with Gamma prior densities for θ, as well as the Bayesian predictive density associated with the "non-informative" improper prior density $\pi_0(\theta) = \frac{1}{\theta}\,\mathbb{I}_{(0,\infty)}(\theta)$. The construction of corresponding Bayesian credibility regions is also presented and we derive highest posterior density (HPD) solutions. We also derive predictive densities for: (i) univariate components of $\tilde{X}_2$ and (ii) the sum of the next k residual lifetimes $S_k = \sum_{i=m+1}^k (X_{(i)} - X_{(m)})$, with $k \in \{m+1, \ldots, n\}$, and relate these densities to mixtures of Pareto densities and linear combinations of multivariate Pareto distributions. The strategic importance of inferring about the sum of future residual lifetimes is a recurrent theme in reliability (e.g., [7]; [10]) for some examples of older references). However, previous analyses do not, as far as we can tell, address the mixture model in (1).

In deriving such prediction densities and regions, we consider as in [3] the equivalent representation of the X_i's through the spacings $Z_i = X_{(i)} - X_{(i-1)}, i = 1, \ldots, n$ with $X_{(0)} = 0$, knowing that Bayesian predictive densities $\hat{q}_\pi$ for the conditional density $q_\theta(\cdot|z_{(1)})$ of a future set of spacings $Z_{(2)} = (Z_{m+1}, \ldots, Z_n)^\top$ given $z_{(1)} = (z_1, \ldots, z_m)^\top$, associated with prior density π, can be converted to predictive densities for $\tilde{X}_2$, of functions of $\tilde{X}_2$ such as the order statistic

$$X_{(m+k)} = \sum_{j=1}^{k} Z_{m+j} + x_{(m)}\,; 1 \leq k \leq n-m; \tag{3}$$

and the sum of residual lifetimes

$$S_k = \sum_{i=m+1}^{k} (X_{(i)} - X_{(m)}) = \sum_{j=m+1}^{k} (k-j+1) Z_j \;; m+1 \leq k \leq n. \tag{4}$$

The alternate analysis is guided by the well-known property that the joint distribution of such spacings conditional on β are distributed independently as $Z_j \sim \text{Exp}\big((n-j+1)\beta\big)$ for $i = 1, \ldots, n$.

A notable feature of our findings arises with the predictive density $\hat{q}_{\pi_0}$ for $q_\theta(\cdot|z_{(1)})$ and consists in its simple form (a multivariate Pareto) as well as the property that $\hat{q}_{\pi_0}$ does not depend on the model choice of α in (1). The same necessarily holds for marginal distributions of Z_2 or $\tilde{X}_2$, and for credibility prediction regions $R_{\pi_0}(\cdot; \tilde{x}_{(1)})$. Moreover, the predictive density $\hat{q}_{\pi_0}$ matches the one for the degenerate or baseline case, given by [3], with $X_1, \ldots, X_n | \theta$ independently distributed as $\text{Exp}(\theta)$. Such an appealing robustness finding where a Bayesian procedure remains the same across a class of models is not unique (e.g., [11]; [6], and the references therein), but nevertheless remains surprising, when it does occur as is the case here.

Although the approach put forth is Bayesian, the prediction regions R_{π_0} associated with the specific choice of the prior density π_0, are shown to have exact frequentist validity for all θ, in the sense that the credibility matches the frequentist conditional probability of coverage for all θ. As for $\hat{q}_{\pi_0}$, this matching property remains true for all exponential mixtures in (1).

The manuscript is organized as follows. Sections 2.1 and 2.2 contain preliminary results and definitions, namely for multivariate Pareto type II distributions and their linear combinations, as well as for mixtures of exponential distributions. Section 2.3 contains sample properties related to model (1). Section 3 is about Bayes predictive densities for the non-informative prior density π_0 (Section 3.1) and for Gamma prior densities (Section 3.2). Corresponding Bayesian point predictors and credibility regions, including HPD solutions, are presented in Section 3.3 and Section 3.4, respectively. Finally, analytical developments leading to matching frequentist coverage probability and credibility are obtained in Section 4.

2 Preliminary results

2.1 Definitions and properties

We begin with a definition for multivariate Pareto type-II distributions. Properties which we will make use of appear a little below in Lemma 2.1.

Definition 2.1. *A random vector $Z = (Z_1, \ldots, Z_N)^\top$ has a multivariate Pareto type II distribution, denoted $\mathcal{P}_2(c, h_1, \ldots, h_N)$, whenever Z admits the representation*

$$Y =^d \left(\frac{E_1}{G}, \ldots, \frac{E_N}{G}\right)^\top,$$

with $E_1, \ldots, E_N, G$ independently distributed as $E_j \sim Exp(h_j)$ and $G \sim \mathcal{G}(c, 1)$. Equivalently, Z is also defined as a random vector with joint density

$$g_{c,h_1,\ldots,h_N}(z) = \frac{(c)_N \prod_{i=1}^N h_i}{(1 + \sum_{i=1}^N h_i z_i)^{c+N}} \, \mathbb{I}_{\mathbb{R}_+^N}(z), \tag{5}$$

The above defines a density with shape parameter c and scale parameters $\frac{1}{h_1}, \ldots, \frac{1}{h_N}$.

Remark 2.1. *The model density in (2) for $(X_1, \ldots, X_n)^\top$ is that of a $\mathcal{P}_2(\alpha, \frac{1}{\theta}, \ldots, \frac{1}{\theta})$ distribution.*

We will denote a gamma distribution as $\mathcal{G}(a, b), a, b > 0$, with density $f(x) = \frac{b^a \, x^{a-1} e^{-bx}}{\Gamma(a)} \, \mathbb{I}_{(0,\infty)}(x)$. Exponential distributions $\text{Exp}(b)$ arise as $\mathcal{G}(1, b)$. We will refer to Beta type-II distributions (also called Beta prime or inverted Beta in the literature) as $W \sim \mathcal{B}2(a, b, \sigma)$ with p.d.f. $f_W(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\sigma^b \, w^{a-1}}{(\sigma+w)^{a+b}}$, for $w \in (0, \infty)$, $a, b > 0$ being shape parameters, and $\sigma > 0$ being a scale parameter. Univariate type II Pareto distributions $\mathcal{P}_2(b, \frac{1}{\sigma})$ arise for $a = 1$. We will refer to Kummer type-II distributions, denoted $\mathcal{K}_2(\gamma_1, \gamma_2, \gamma_3, \sigma)$ with $\gamma_1, \sigma > 0$, and either $\gamma_2 \in \mathbb{R}, \gamma_3 > 0$ pr $\gamma_2 > 0, \gamma_3 = 0$, as those with densities

$$\frac{\sigma^{\gamma_2} \, u^{\gamma_1 - 1} \, e^{-\frac{\gamma_3 u}{\sigma}}}{\Gamma(\gamma_1) \, \psi(\gamma_1, 1 - \gamma_2, \gamma_3) \, (\sigma + u)^{\gamma_1 + \gamma_2}} \, \mathbb{I}_{(0,\infty)}(u),$$

where ψ is the confluent hypergeometric function of type II given by

$$\psi(\gamma_1, \gamma_2, \gamma_3) = \frac{1}{\Gamma(\gamma_1)} \int_0^\infty \frac{t^{\gamma_1 - 1}}{(1+t)^{\gamma_1 - \gamma_2 + 1}} \, e^{-\gamma_3 t} \, dt.$$

We point out that the Kummer type-II family contains both Beta type-II and gamma distributions, with $\mathcal{K}_2(\gamma_1, \gamma_2, 0, \sigma)$ reducing to $\mathcal{B}_2(\gamma_1, \gamma_2, \sigma)$, and $\mathcal{K}_2(\gamma_1, -\gamma_1, \gamma_3, \sigma)$ reducing to $\mathcal{G}(\gamma_1, \frac{\gamma_3}{\sigma})$.

Next, we collect properties of multivariate Pareto type II distributions which will appear below as model and predictive densities. Notable features include the preservation of the multivariate Pareto nature for marginal and conditional distributions. The properties follow from the above definitions and are readily verifiable. They have appeared in various forms in the literature (e.g., [2]; [8]). Section 2.2 treats linear combinations of multivariate Pareto distributed vectors and their study is also of independent interest.

Lemma 2.1. *(Multivariate Pareto properties) Consider* $Y = (Y_1, \ldots, Y_N)^\top \sim \mathcal{P}_2(c, h_1, \ldots, h_N)$. *Then, we have:* **(i)** *then* $Y_i \sim \mathcal{P}_2(c, h_i),\ i = 1, \ldots, N$; **(ii)** $Y_j \mid Y_i = y_i \sim \mathcal{P}_2\big(c+1, \frac{h_j}{1+h_i y_i}\big)$ *for* $i \neq j$; **(iii)** $(h_1 Y_1, \ldots, h_N Y_N)^\top \sim \mathcal{P}_2(c, 1, \ldots, 1)$; **(iv)** Y *has survival function* $P\big(\cap_{i=1}^n (Y_i > y_i)\big) = \big(1 + \sum_{i=1}^N h_i y_i\big)^{-c},\ y_i > 0$; **(v)** Y *has expectation* $\mathbb{E}(Y) = \frac{1}{c-1}\left(\frac{1}{h_1}, \ldots, \frac{1}{h_N}\right)^\top$; **(vi)** *for* $Y_{(1)} = (Y_1, \ldots, Y_m)^\top$ *and* $Y_{(2)} = (Y_{m+1}, \ldots, Y_N)^\top$, *we have*

$$Y_{(1)} \sim \mathcal{P}_2(c, h_1, \ldots, h_m), \text{ and } Y_{(2)} \mid Y_{(1)} = y_{(1)} \sim \mathcal{P}_2\big(c+m, \frac{h_{m+1}}{\Delta}, \ldots, \frac{h_N}{\Delta}\big), \tag{6}$$

with $\Delta = 1 + \sum_{i=1}^m h_i y_i$; *and* **(vii)** $U = \sum_{i=1}^N h_i Y_i$ *is distributed as* $\mathcal{B}_2(N, c, 1)$.

We next present properties for mixtures of exponential distributions that relate to model (1) expressed in terms of spacings Z_i.

Lemma 2.2. *Let* $Z_j|\beta \sim Exp(a_j\beta), j = 1, \ldots, N$, *be conditionally independent, with known* a_j*'s,* $\beta|\theta \sim g_\theta$, g_θ *being a density supported on* $S \subseteq \mathbb{R}_+$ *with respect to a* $\sigma-$*finite measure* μ, *and* θ *being an unknown parameter. Then,*

- **(a)** *Based on* $Z_1, \ldots, Z_m$ *with* $1 \le m \le N$, $T = \sum_{j=1}^m a_j Z_j$ *is a sufficient statistic for* θ;
- **(b)** *For* $\beta|\theta \sim \mathcal{G}a(\alpha, \theta)$, $Z = (Z_1, \ldots, Z_N)^\top$ *is distributed as* $\mathcal{P}_2(\alpha, \frac{a_1}{\theta}, \ldots, \frac{a_N}{\theta})$;
- **(c)** *For* $\beta|\theta \sim \mathcal{G}a(\alpha, \theta)$, *the statistic* T *is distributed as* $\mathcal{B}2(m, \alpha, \theta)$.

Proof. With the mixture representation, we have the joint density

$$f_Z(z_1, \ldots, z_m) = \int_S \beta^m \,(\prod_{j=1}^m a_j)\, exp\{-\beta \sum_{j=1}^m a_j z_j\}\, g_\theta(\beta)\, d\mu(\beta)\,,$$

and the sufficiency in part **(a)** follows from the factorization theorem. For part **(b)**, the result follows by a straightforward calculation or, by exploiting the hierarchical model to infer that

$$\begin{aligned}(Z_1, \ldots, Z_N)^\top &=^d \big(\frac{\theta\beta Z_1}{\theta\beta}, \ldots, \frac{\theta\beta Z_N}{\theta\beta}\big)^\top \\ &=^d \big(\frac{E_1}{G}, \ldots, \frac{E_N}{G}\big)^\top,\end{aligned}$$

with $G =^d \theta\beta|\theta \sim \mathcal{G}(\alpha, 1)$, $E_j =^d \theta\beta Z_j|\theta \sim Exp(\frac{a_j}{\theta})$ for $j = 1, \ldots, N$ independent, in which case the result follows with Definition 2.1. Finally, part **(c)** is a consequence of part **(b)** and Lemma 2.1, **(vii)**. □

2.2 Linear combinations of multivariate Pareto type II distributions

We are interested here in linear combinations $a^\top V$ with $a = (a_1, \ldots, a_N)^\top$ and $V = (V_1, \ldots, V_N)^\top$ distributed as multivariate Pareto type II $\mathcal{P}_2(c, h_1, \ldots, h_N)$, and notably in applications to linear combinations such as those represented in (3) and (4) which will arise in the below analysis. We record properties for general a's such that $a_i > 0$ and $a_i h_i$ are all distinct, for $i = 1, \dots, N$. The distributions below relate to sums of independent exponential distributed random variables, referred to in the literature as hypoexponential, Generalized Erlang or Generalized Gamma, among others (e.g., [13]; [9]), and can derived through the mixture representation in Definition 2.1. Such distributions are expressible in terms of Lagrange basis polynomials which we briefly describe.

For a given set $\mathcal{Y} = \{y_1, \dots, y_k\}$ of $k > 1$ distinct real numbers and $\mathcal{Y}_{(j)} = \mathcal{Y} - \{j\}$, the Lagrange interpolating polynomial is defined as $L(y) = \sum_{j=1}^{k} y_j \, l_{\mathcal{Y},j}(y)$, with the basis polynomials $l_{\mathcal{Y},j}(y)$'s given by

$$l_{\mathcal{Y},j}(y) = \prod_{i \in \mathcal{Y}_{(j)}} \left(\frac{y - y_i}{y_j - y_i}\right). \tag{7}$$

The above well-known construction is such that: (i) $l_{\mathcal{Y},j}(y_j) = 1$ for $y_j \in \mathcal{Y}$, (ii) $l_{\mathcal{Y},j}(y_i) = 0$ for $y_i \in \mathcal{Y}_{(j)}$, (iii) $L(y_j) = y_j$ for $y_j \in \mathcal{Y}$, and (iv) $\sum_{j=1}^{k} l_{\mathcal{Y},j}(y) = 1$ for $y \in \mathbb{R}$.

For $\Lambda = \{\lambda_1, \dots, \lambda_N\}$ a set of $N \geq 2$ distinct real numbers and $Y_j \sim Exp(\lambda_j)$, the sum $S = \sum_{j=1}^{N} Y_j$ follows a hypoexponential distribution, and has density

$$f(s) = \sum_{j=1}^{N} l_{\Lambda,j}(0) \, \lambda_j \, e^{-\lambda_j s} \, \mathbb{I}_{(0,\infty)}(s). \tag{8}$$

This is a mixture of $Exp(\lambda_j)$ densities with weights $l_{\Lambda,j}(0)$ obtained from (7), adding to 1, but with alternating signs. We now derive the density of a multivariate Pareto type II linear combination.

Theorem 2.1. *Let $N > 1$, $V \sim \mathcal{P}_2(c, h_1, \dots, h_N)$, $a = (a_1, \dots, a_N)^\top$, and $\mathcal{Y}^* = \{a_1 h_1, \dots, a_N h_N\}$, such that $a_j > 0$ and the $a_j h_j$'s are distinct. Then, then $S = a^\top V$ has density function*

$$f(s) = \sum_{j=1}^{N} l_{\mathcal{Y}^*,j}(0) \, g_{c, a_j h_j}(s), \tag{9}$$

where $g_{c,a_j h_j}$ is the density of a $\mathcal{P}_2(c, a_j h_j)$ distribution, and the $l_{\mathcal{Y}^*,j}(0)$'s are the Lagrange basis polynomials evaluated at 0.

Proof. With Definition 2.1, the distribution of $V = (V_1, \dots, V_N)^\top$ admits the representation

$$V_j | X_0 \text{ independent } Exp(h_j X_0) \Longrightarrow a_j V_j | X_0 \text{ independent } Exp(a_j h_j X_0), \tag{10}$$

so that the conditional distribution of S given X_0 is hypoexponential with density function

$$f(s|x_0) = \sum_{j=1}^{N} l_{\mathcal{Y}^*,j}(0) \; a_j h_j x_0 \, e^{-a_j h_j x_0 \, s} \, \mathbb{I}_{(0,\infty)}(s), \tag{11}$$

where we have used the fact that the basis polynomials associated with the sets $\{a_1 h_1 x_0, \dots, a_N h_N x_0\}$ are identical for all $x_0 > 0$, and therefore equal to those for $x_0 = 1$. Finally the result follows by integrating out x_0 with respect to a $X_0 \sim \mathbb{G}(c, 1)$ density. $\square$

2.3 Model and sampling properties

We re-express the observed order statistics $X_{(1)} < \cdots < X_{(m)}$ and the future order statistics $X_{(m+1)} < \cdots < X_{(n)}$ of model (1) in terms of the spacings

$$Z_i = X_{(i)} - X_{(i-1)},\ i = 1, \ldots, n\,, \tag{12}$$

and the vectors of spacings

$$Z_{(1)} = (Z_1, \ldots, Z_m)^\top, \qquad Z_{(2)} = (Z_{m+1}, \ldots, Z_n)^\top\,, \tag{13}$$

with $X_{(0)} = 0$ and $1 \leq m < n$. The transformation of the original data to the spacings simplifies and facilitates, with no loss of information, the sampling properties and analysis, by virtue of the well-known property (e.g., [14], for an early reference).

$$Z_i|\beta \sim \text{Exp}\big((n-i+1))\,\beta\big) \text{ independent,}$$

with respect to model (1).

Lemma 2.3. *Based on model (1) expressed in terms of the spacings (12, 13), we have*

- **(a)** $(Z_1, Z_2, \ldots, Z_n)^\top|\theta \sim \mathcal{P}_2(\alpha, \frac{n}{\theta}, \frac{n-1}{\theta}, \ldots, \frac{1}{\theta})$;
- **(b)** *Based on* $Z_{(1)}$*, the statistic* $T = \sum_{i=1}^m (n-i+1) Z_i$ *is a sufficient statistic for* θ*, and distributed as* $T \sim \mathcal{B}_2(m, \alpha, \theta)$*;*
- **(c)** $Z_{(2)} \mid Z_{(1)} = z_{(1)}, \theta \sim \mathcal{P}_2\big(m+\alpha, \frac{n-m}{\theta+t}, \frac{n-m-1}{\theta+t}, \ldots, \frac{1}{\theta+t}\big)$*, where* $t = \sum_{i=1}^m (n-i+1) z_i$ *is the observed value of* T*.*

Proof. Parts **(a)** and **(b)** follow, respectively, from parts **(b)** and **(a)** of Lemma 2.2. Part **(c)** follows from part **(vi)** of Lemma 2.1 given the joint distribution in **(a)**. □

In view of the previous result when referring to model (1), we will hereafter also be referring to the sufficient statistic $T \sim \mathcal{B}_2(m, \alpha, \theta)$ having observed $Z_{(1)}$, and the model density $q_\theta(\cdot|z_{(1)})$ for $Z_{(2)}$ which is that of a $\mathcal{P}_2\big(m+\alpha, \frac{n-m}{\theta+t}, \frac{n-m-1}{\theta+t}, \ldots, \frac{1}{\theta+t}\big)$ distribution.

3 Bayesian predictive densities and regions

Our findings below relate to Bayesian predictive densities for the conditional distribution of $Z_{(2)}$ given $Z_{(1)} = z_{(1)}$, based on prior (Lebesgue) density π and corresponding posterior density $\pi(\cdot|z_{(1)})$ for θ. Such predictive densities are of the form

$$\hat{q}_\pi(z_{(2)}|z_{(1)}) \;=\; \int_\Theta q_\theta(z_{(2)}|z_{(1)})\, \pi(\theta|z_{(1)})\, d\theta\,, z_{(2)} \in \mathbb{R}_+^{n-m}, \tag{14}$$

and can also serve to generate point predictors or prediction regions for $Z_{(2)}$, as well as for subvectors and univariate marginal components Z_i. We begin with the non-informative prior density $\pi_0(\theta) \;=\; \frac{1}{\theta}\, \mathbb{I}_{(0,\infty)}(\theta)$, which yields a simple form for the predictive density (Section 3.1). Furthermore, we reveal a surprising robustness property with respect to the model choice of α (Section 4). We provide Bayesian predictive densities for gamma priors in Section 3.2, and credible regions are discussed in Section 3.3.

3.1 Non-informative prior density

Theorem 3.1. *For model (1), the Bayesian predictive density* $\hat{q}_{\pi_0}(z_{(2)}|z_{(1)})$; $z_{(2)} \in \mathbb{R}_+^{n-m}$; *for the conditional distribution* $Z_{(2)}|Z_{(1)} = z_{(1)}$*, and associated with prior density* $\pi_0(\theta) = \frac{1}{\theta}\,\mathbb{I}_{(0,\infty)}(\theta)$*, is that of a* $\mathbb{P}2(m, \frac{n-m}{t}, \ldots, \frac{1}{t})$ *density with* $t = \sum_{i=1}^{m}(n-i+1)z_i$ *.*

Proof. It follows from Lemma 2.3 that $T = \sum_{i=1}^{m}(n-i+1)Z_i$ is a sufficient statistic for θ, and distributed as $\mathcal{B}2(m, \alpha, \theta)$. With the prior density π_0, we obtain as a posterior density

$$\pi(\theta|z_{(1)}) \propto \frac{\theta^{\alpha}\, t^{m-1}}{(\theta+t)^{\alpha+m}}\, \frac{1}{\theta}\, \mathbb{I}_{(0,\infty)}(\theta)\,, \tag{15}$$

or equivalently that $\theta|t \sim \mathcal{B}2(\alpha, m, t)$. From this and part **(c)** of Lemma 2.3, we obtain the Bayes predictive density

$$\begin{aligned}
\hat{q}_{\pi_0}(z_{(2)}|z_{(1)}) &= \int_0^{\infty} q_{\theta}(z_{(2)}|z_{(1)})\, \pi(\theta|z_{(1)})\, d\theta \\
&\propto \int_0^{\infty} \frac{1}{(\theta+t)^{n-m}}\, \frac{1}{\left(1+\sum_{i=m+1}^{n} \frac{(n-i+1)z_i}{\theta+t}\right)^{n+\alpha}}\, \frac{\theta^{\alpha-1}}{(\theta+t)^{\alpha+m}}\, d\theta \\
&\propto \int_0^{\infty} \frac{\theta^{\alpha-1}}{\left(\theta+t+\sum_{i=m+1}^{n}(n-i+1)z_i\right)^{n+\alpha}}\, d\theta \\
&\propto \frac{1}{\left(t+\sum_{i=m+1}^{n}(n-i+1)z_i\right)^{n}}\,,
\end{aligned}$$

where we have used the identity $\int_0^{\infty} \frac{x^{a-1}}{(c+x)^{a+b}}\, dx = \frac{1}{c^b}\, \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ for $a, b, c > 0$. Finally, the proof is complete as the above density is indeed a Pareto density as stated.

Remark 3.1. *We point out that a Bayesian predictive density for the conditional distribution* $Z_{(2)}|Z_{(1)} = z_{(1)}$ *based on a prior density* π *for* θ *can be converted to a predictive density for the conditional distribution of* $\tilde{X}_2$ *given an observed* $\tilde{x}_1$ *(equivalently* $z_{(1)}$*) with the inverse transformation*

$$X_{(m+j)} = \sum_{i=1}^{j} Z_{m+i} + x_{(m)}\,,\; j = 1, \ldots, n-m\,,$$

and Jacobian equal to 1*. As an illustration, for the case of the prior density* π_0*, we have*

$$\hat{q}_{\pi_0}(z_{(2)}|z_{(1)}) = \frac{(n-1)!}{(m-1)!}\, \frac{(n-m)!\, t^m}{\left(t + \sum_{j=1}^{n-m}(n-m-j+1)z_{m+j}\right)^{n}}\,;\, z_{(2)} \in \mathbb{R}_+^{n-m}\,;$$

and therefore the predictive density

$$\hat{q}^{*}_{\pi_0}(\tilde{x}_2|z_{(1)}) = \frac{(n-1)!}{(m-1)!}\, \frac{(n-m)!\, t^m}{\left(t + \sum_{j=1}^{n-m} x_{(m+j)} - (n-m)\, x_{(m)}\right)^{n}}\,,$$

for $\tilde{X}_2 = (X_{(m+1)}, \ldots, X_{(n)})^{\top}$*.*

A characteristic feature of the above predictive density $\hat{q}_{\pi_0}$ is that it does not depend on the model specification, i.e., the specification of α in model (1). This is in contrast to the posterior distribution, which is $\mathcal{B}2(\alpha, m, t)$ and depends on α. Moreover, the predictive density $\hat{q}_{\pi_0}$ matches that of the non-mixture $\hat{q}_{\pi_0}$ version with $X_1, \ldots, X_n$ i.i.d. $Exp(\theta)$ (see [3]). Additional insight on this robustness property will be provided with Remark 4.1 and is an attribute of the prior density π_0 - other prior density choices which are studied below in this section do not lead to such a property.

The same robustness property will naturally carry-over to the joint predictive distribution of the future order statistics $\tilde{X}_2$, and functions of $\tilde{X}_2$ including the marginal predictive distributions of the components Z_i (or $X_{(i)}$) of $Z_{(2)}$ (or $\tilde{X}_2$), as well as the sum of residual lifetimes S_k. For the future order statistics $X_{(m+k)}$ and residual sums S_k, we obtain their predictive densities by exploiting their representation as a linear combination of a multivariate Pareto distributed vector.

Corollary 3.1. *Consider model (1), prior density π_0, and $t = \sum_{i=1}^m (n-i+1) z_i$.*

(a) *the Bayes predictive density of $Z_{m+j} | Z_{(1)} = z_{(1)}$ is that of a $\mathbb{P}2(m, \frac{n-m-j+1}{t})$ distribution;*

(b) *the Bayes predictive density of $X_{(m+1)}$ is given by $\hat{q}'(x_{(m+1)} | x_{(m)})) = g_{m, \frac{n-m}{t}}(x - x_m)$;*

(c) *the Bayesian predictive density of $X_{(m+k)} | Z_{(1)} = z_{(1)}$, $k \in \{2, \ldots, n-m\}$, is equal to*

$$\hat{q}_{\pi_0}(x | z_{(1)}) = \sum_{j=1}^{k} \Big\{ \prod_{i \neq j} \frac{n-m-i+1}{j-i} \Big\} \, g_{m, \frac{n-m-j+1}{t}}(x - x_{(m)}) \,, \tag{16}$$

i.e., a mixture of translated $\mathcal{P}_2(m, n-m-j+1)$ densities with weights $\prod_{i \neq j} \frac{n-m-i+1}{j-i}$;

(d) *the Bayesian predictive density of $S_n | Z_{(1)} = z_{(1)}$ is that of a $\mathcal{B}_2(n-m, m, t)$ distribution;*

(e) *the Bayesian predictive density of $S_k | Z_{(1)} = z_{(1)}$, $k \in \{m+2, \ldots, n-1\}$, is given by*

$$\hat{q}_{S_k, \pi_0}(s | z_{(1)}) = \sum_{j=1}^{k-m} \Big\{ \prod_{i \neq j} \frac{c_i}{c_i - c_j} \Big\} \, g_{m, c_j/t}(s) \,, \tag{17}$$

where $c_j = (k-j-m+1)\,(n-j-m+1)$.

Proof. *Part* **(a)** *follows from Theorem 3.1 and part* **(i)** *of Lemma 2.1, while part* **(b)** *follows from the translation $z_{m+1} \to x_{(m+1)} = z_{m+1} + x_{(m)}$.*

For part **(c)**, *it follows from Theorem 3.1 and part* **(v)** *of Lemma 2.1 that*

$$V = (V_1, \ldots, V_{k-m})^\top = (Z_{m+1}, \ldots, Z_k)^\top \sim \mathcal{P}_2\big(m, \frac{n-m}{t}, \ldots, \frac{n-m-k+1}{t}\big) \,, \tag{18}$$

under density $\hat{q}_{\pi_0}(\cdot|z_{(1)})$. We now make use of Theorem 2.1 with $c = m$, $h_i = \frac{n-m-i+1}{t}$, $a = (1,\ldots,1)^\top$, $\mathcal{Y}^* = \{\frac{n-m}{t},\ldots,\frac{n-m-k+1}{t}\}$, and $S = X_{(m+k)} - x_{(m)} = a^\top V$ given representation (3). Applying (9), the result follows with

$$l_{\mathcal{Y}^*,j}(0) = \prod_{i\in\mathcal{Y}^*,j} \frac{n-m-i+1}{(n-m-i+1)-(n-m-j+1)} = \prod_{i\neq j} \frac{n-m-i+1}{j-i} .$$

Expression (16) then follows by the change of location $S \to X_{(m+k)} = S + x_{(m)}$.

For part **(d)**, *we have $\frac{1}{t} S_n = \sum_{j=m+1}^{n} \frac{(n-j+1)}{t} Z_j \sim \mathcal{B}_2(n-m, m, t)$ as a consequence of part (**vii**) of Lemma 2.1. The result follows.*

For part **(e)**, *observe that (i.e., (4)) that $S_k = \sum_{j=m+1}^{k}(k-j+1)Z_j = \sum_{j=1}^{k-m}(k-j-m+1)V_j$, with $V = (V_1,\ldots,V_{k-m})^\top$ as in (18). We can therefore apply Theorem 2.1 with $S = a^\top V$, $N = k-m$, $a = (a_1,\ldots,a_{k-m})^\top$ with $a_j = k-j-m+1$, and $\mathcal{Y}^* = \{a_1h_1,\ldots,a_Nh_N\}$, and the result follows.* □

3.2 Gamma priors

In this section, we provide Bayesian predictive densities for the conditional density of $Z_{(2)}$, given an observed value $z_{(1)}$, associated with a gamma prior density. As shown with the next result, the posterior and predictive densities are conveniently expressible in terms of Kummer type-II distributions and the confluent hypergeometric function of type II.

Theorem 3.2. *For $Z = (Z_{(1)}, Z_{(2)})^\top$ distributed as in model (1), based on observation $z_{(1)}$ with $t = \sum_{i=1}^{m}(n-i+1)\,z_i$ and a $\mathcal{G}(a,b)$ prior density $\pi_{a,b}$ for θ,*

(a) the posterior distribution of $\theta|z_{(1)}$ is $\mathcal{K}_2(\gamma_1 = a+\alpha, \gamma_2 = m-a, \gamma_3 = bt, \sigma = t)$;

(b) The Bayesian predictive density of $Z_{(2)}|Z_{(1)} = z_{(1)}$ is given by

$$\hat{q}_{\pi_{a,b}}(z_{(2)}|z_{(1)}) = \frac{(n-m)!\,\Gamma(n+\alpha)}{\Gamma(m+\alpha)} \times \frac{\psi\big(\alpha+a\,,\, a-n+1\,,\, b\,(t+w_1)\big)}{\psi\big(\alpha+a\,,\, a-m+1\,,\, b\,t\big)} \frac{t^{m-a}}{(w_1+t)^{n-a}}\,, \tag{19}$$

with $w_1 = \sum_{i=m+1}^{n}(n-i+1)\,z_i$.

Proof.

(a) Since T is a sufficient statistic for θ and distributed as $\mathcal{B}_2(m,\alpha,\theta)$ (Lemma 2.3), we have for the posterior density of θ:

$$\pi_{a,b}(\theta|z_{(1)}) \propto f_T(t|\theta)\,\pi_{a,b}(\theta) \propto \frac{\theta^{\alpha+a-1}}{(\theta+t)^{\alpha+m}}\, e^{-b\theta}\,\mathbb{I}_{(0,\infty)}(\theta)\,,$$

which is indeed the stated Kummer type-II density with the given parameters.

(b) Calculations, making use of Lemma 2.3, the definition of ψ, and the above posterior density, yield

$$\begin{aligned}
\hat{q}_{\pi_{a,b}}(z_{(2)}|t) &= \int_0^\infty q_\theta(z_{(2)}|z_{(1)})\,\pi_{a,b}(\theta|z_{(1)})\,d\theta \\
&= \frac{\Gamma(n+\alpha)\,(n-m)!\,t^{m-a}}{\Gamma(m+\alpha)\,\Gamma(\alpha+a)\,\psi(\alpha+a,a-m+1,bt)} \int_0^\infty \frac{\theta^{\alpha+a-1}\,e^{-b\theta}}{(t+w_1+\theta)^{n+\alpha}}\,d\theta \\
&= \frac{\Gamma(n+\alpha)\,(n-m)!\,t^{m-a}}{\Gamma(m+\alpha)\,\psi(\alpha+a,a-m+1,bt)}\,\frac{\Gamma(\alpha+a)\,\psi(\alpha+a,a-n+1,b(t+w_1))}{(t+w_1)^{n-a}},
\end{aligned}$$

for $z_{(2)} \in \mathbb{R}_+^{n-m}$, yielding the result. □

Remark 3.2. *Using the same steps as above with $a = b = 0$ for the gamma density, i.e., $\pi_0(\theta) = \frac{1}{\theta}\,\mathbb{I}_{(0,\infty)}(\theta)$, the $\mathcal{P}_2(m, \frac{n-m}{t}, \ldots, \frac{1}{t})$ predictive density in Theorem 3.1 may be derived directly from (19) with simplifications that follow from the identity $\psi(\gamma_1,\gamma_2,0) = \frac{\Gamma(1-\gamma_2)}{\Gamma(\gamma_1+1-\gamma_2)}$.*

3.3 Point predictors

In a Bayesian framework with prior density π, it is reasonable to propose the expectations $\mathbb{E}_\pi(Z_{m+k}|z_{(1)})$ and $\mathbb{E}_\pi(X_{(m+k)}|z_{(1)})$ as point predictors of the future spacing Z_{m+k} and order statistic $X_{(m+k)}$, respectively. We elaborate on such expressions for: **(I)** prior density π_0, and **(II)** gamma prior densities $\pi_{a,b}$.

(I) For prior density π_0, in accordance with Theorem 3.1, such conditional expectations will be independent of α in model (1). From Theorem 3.1 and part **(v)** of Lemma 2.1, we obtain for $m > 1$ the simple expression

$$\mathbb{E}_{\pi_0}(Z_{m+k}|z_{(1)}) = \frac{t}{(m-1)\,(n-m-k+1)}\,,\ k = 1,\ldots,n-m\,. \tag{20}$$

An alternative derivation, which does not require explicitly the form of the predictive density $\hat{q}_{\pi_0}$ is as follows. We have

$$\begin{aligned}
\mathbb{E}_{\pi_0}(Z_{m+k}|z_{(1)}) &= \mathbb{E}_{\pi_0}^{\theta|z_{(1)}}\{\mathbb{E}(Z_{m+k}|z_{(1),\theta})\} \\
&= \mathbb{E}_{\pi_0}^{\theta|z_{(1)}}\{\frac{\theta+t}{(m+\alpha-1)(n-m-k+1)}\} \\
&= \frac{\mathbb{E}_{\pi_0}(\theta|z_{(1)})\,+\,t}{(m+\alpha-1)(n-m-k+1)} \\
&= \frac{\frac{t\alpha}{m-1}\,+\,t}{(m+\alpha-1)(n-m-k+1)} \\
&= \frac{t}{(m-1)\,(n-m-k+1)}\,;
\end{aligned} \tag{21}$$

where we have used part (c) of Lemma 2.3 and parts **(i), (v)** of Lemma 2.1 to infer that $\mathbb{E}(Z_{m+k}|z_{(1),\theta})\} = \frac{\theta+t}{(m+\alpha-1)\,(n-m-k+1)}$, as well as the posterior expectation

$\mathbb{E}_{\pi_0}(\theta|z_{(1)}) = \frac{t\alpha}{m-1}$ for $m > 1$ which follows from the the posterior distribution $\theta|z_{(1)} \sim \mathbb{B}2(\alpha, m, t)$ (see (15)) and corresponding Beta type-II expectation.

For predicting $X_{(m+k)}$, since $X_{(m+k)} = X_{(m)} + \sum_{j=1}^{k} Z_{m+j}$, we obtain

$$\begin{aligned}\mathbb{E}_{\pi_0}(X_{(m+k)}|z_{(1)}) &= x_{(m)} + \sum_{j=1}^{k} \mathbb{E}_{\pi_0}(Z_{m+j}|z_{(1)}) \\ &= x_{(m)} + \frac{t}{m-1} \sum_{j=1}^{k} \frac{1}{n-m-j+1},\end{aligned}$$

making use of (20).

(II) For Gamma prior densities $\pi_{a,b}$, the point predictor $\mathbb{E}_{\pi_{a,b}}(Z_{m+k}|z_{(1)})$ can be evaluated from (21) and requires the posterior expectation $\mathbb{E}_{\pi_{a,b}}(\theta|z_{(1)})$. Since $\theta|z_{(1)}$ is Kummer type-II distributed and since a $\mathcal{K}_2(\gamma_1, \gamma_2, \gamma_3, \sigma))$ distribution for $\gamma_3 > 0$ has expectation $\sigma\,\gamma_1 \frac{\psi(\gamma_1+1, 2-\gamma_2, \gamma_3)}{\psi(\gamma_1+1, 1-\gamma_2, \gamma_3)}$, it follows from part **(a)** of Theorem 3.2 that

$$\mathbb{E}_{\pi_{a,b}}(\theta|z_{(1)}) = \sigma\,(a+\alpha)\,\frac{\psi(a+\alpha+1, a+2-m, b)}{\psi(a+\alpha, a+1-m, b)}. \tag{22}$$

We therefore obtain from (21) the Bayesian point predictor

$$\begin{aligned}\mathbb{E}_{\pi_{a,b}}(Z_{m+k}|z_{(1)}) &= \frac{1}{(m+\alpha-1)\,(n-m-k+1)} \\ &\quad \times \left(t + \sigma\,(a+\alpha)\,\frac{\psi(a+\alpha+1, a+2-m, b)}{\psi(a+\alpha, a+1-m, b)}\right),\end{aligned}$$

for $k \in \{1, \ldots, n-m\}$. Finally, as above for **(I)**, we obtain for predicting the order statistic $X_{(m+k)}$:

$$\begin{aligned}\mathbb{E}_{\pi_{a,b}}(X_{(m+k)}|z_{(1)}) &= x_{(m)} + \sum_{j=1}^{k} \mathbb{E}_{\pi_{a,b}}(Z_{m+j}|z_{(1)}) \\ &= x_{(m)} + \sum_{j=1}^{k} \frac{1}{n-m-j+1} \Big(\frac{t}{m+\alpha-1} \\ &\quad + \frac{\sigma\,(a+\alpha)}{m+\alpha-1}\,\frac{\psi(a+\alpha+1, a+2-m, b)}{\psi(a+\alpha, a+1-m, b)}\Big).\end{aligned}$$

3.4 Credibility regions for $Z_{(2)}$

Under model (1), given an observed value $z_{(1)}$ of $Z_{(1)}$, a Bayesian prediction region $R(z_{(1)})$ for $Z_{(2)}$ associated with a $\mathcal{G}(a, b)$ prior density of credibility $1-\lambda$ is such that

$$\int_{R(z_{(1)})} \hat{q}_{\pi_{a,b}}(z_{(2)}|(z_{(1)})\,dz_{(2)} = 1-\lambda. \tag{23}$$

We provide for: **(I)** prior density π_0, and **(II)** gamma prior densities $\pi_{a,b}$ with shape parameter $a \leq n$, the prediction region for $z_{(2)}$ with the smallest volume among all regions with credibility $1-\lambda$, derived as the highest posterior density region of the form

$$R_{HPD}(z_{(1)}) = \{z_{(2)} \in \mathbb{R}_+^{n-m} : \hat{q}_{\pi_{a,b}}(z_{(2)}|(z_{(1)}) \geq k\}\,. \tag{24}$$

For **(I)**, the result is directly obtained and matches the HPD procedure obtained by [3] applicable to Exponential data, i.e., the non-mixture model with β degenerate at θ in model (1). For **(II)**, we will make use of the following intermediate result.

Lemma 3.1. *Let $Z \in \mathbb{R}_+^N$ has joint density $f\big(\sum_{i=1}^N a_i z_i\big)$ with positive a_i's. Then $W_1 = \sum_{i=1}^N a_i Z_i$ has joint density*

$$h_{W_1}(w_1) = \frac{w_1^{N-1}\, f(w_1)}{(N-1)!\,\prod_{i=1}^N a_i}\,. \tag{25}$$

Proof. Let $W_k = \sum_{i=1}^{N-k+1} a_i Z_i$, for $k = 2,\ldots,N$, and transform Z to $W = FZ = (W_1,\ldots,W_N)^\top$ with the ith row of $F(N\times N)$ equal to $f_i = (a_1,\ldots,a_{N-i+1})$ for $i = 1,\ldots,N$. Since $|F| = \prod_{i=1}^N a_i$, the joint density of W is given by

$$h_{W_1}(w_1) = \frac{f(w_1)}{\prod_{i=1}^N a_i}\, \mathbb{I}_{T(w_1)}(w_2,\ldots,w_N)\, \mathbb{I}_{(0,\infty)}(w_1)\,,$$

with $T(w_1) = \big\{(w_2,\ldots,w_N) \in \mathbb{R}_+^{N-1} : w_{N-i+1} < w_{N-i} \text{ for } i = 1,\ldots,N-1\big\}$. Finally, integrating out $w_2,\ldots,w_N$, we obtain

$$h_{W_1}(w_1) = \Big(\int_{T(w_1)} dw_2\ldots dw_N\Big)\frac{f(w_1)}{\prod_{i=1}^N a_i}\,,$$

and then (25) since the volume of $T(w_1)$ is equal to $\frac{w_1^{N-1}}{(N-1)!}$. □

Example 3.1. *Here is an illustrative example for a multivariate Pareto type-II distribution which provides an alternative justification of part* **(vii)** *in Lemma 2.1. Let $Z \sim \mathcal{P}_2(c,h_1,\ldots,h_N)$ and $W_1 = \sum_{i=1}^N h_i Z_i$. From Lemma 3.1 and (5), we obtain the density*

$$\begin{aligned} h_{W_1}(w_1) &= \frac{w_1^{N-1}\,(c)_N\,(\prod_{i=1}^N h_i)}{(N-1)!\,(\prod_{i=1}^N h_i)\,(1+w_1)^{c+N}} \\ &= \frac{\Gamma(c+N)}{\Gamma(c)\,\Gamma(N)}\,\frac{w_1^{N-1}}{(1+w_1)^{c+N}}\,, \end{aligned}$$

which is a $\mathbb{B}_2(N,c,1)$ density.

We now proceed with the HPD regions.

(I) Prior density π_0.

Theorem 3.3. *Under model (1), given an observed value $z_{(1)}$, the HPD Bayesian credibility prediction region for $Z_{(2)}$ associated with a prior density π_0 is given by*

$$R_{HPD}(z_{(1)}) \,=\, \left\{z_{(2)} \in \mathbb{R}_+^{n-m} : w_1 \,=\, \sum_{i=m+1}^{n} (n-i+1) z_i \leq t\, k_0 \right\}, \tag{26}$$

with $t = \sum_{i=1}^{m} (n-i+1) z_i$ and k_0 the quantile of order $1-\lambda$ of a $\mathcal{B}2(n-m, m, 1)$ distribution.

Proof. From Theorem 3.1, the Bayesian predictive density $\hat{q}_{\pi_0}(z_{(2)}|z_{(1)})$ is that of a $\mathcal{P}_2(m, \frac{n-m}{t}, \ldots, \frac{1}{t})$ distribution, depends on $z_{(2)}$ through w_1 only and is decreasing in $w_1 > 0$. Therefore, the HPD region is of the form (26). From part **(d)** of Corollary 3.1 (or Example 3.1) we obtain that $\frac{W_1}{t} \sim \mathcal{B}2(n-m, m, 1)$, and the result follows. □

(II) Gamma prior densities $\pi_{a,b}$. We have the following

Theorem 3.4. *Under model (1), given an observed value $z_{(1)}$, the HPD Bayesian credibility prediction region for $Z_{(2)}$ associated with a $\mathcal{G}(a, b)$ density with $a \leq n$ and credibility $1-\lambda$, is given by*

$$R_{HPD}(z_{(1)}) \,=\, \left\{z_{(2)} \in \mathbb{R}_+^{n-m} : w_1 \,=\, \sum_{i=m+1}^{n} (n-i+1) z_i \leq c_0 \right\}, \tag{27}$$

with $c_0 \in (0, \infty)$ the unique solution of

$$\begin{aligned} &\int_0^{c_0} \frac{w_1^{n-m-1}\, \psi(\alpha+a, a-n+1, b(t+w_1))}{(w_1+t)^{n-a}}\, dw_1 \\ &= \frac{(1-\lambda)\,(n-m-1)!}{t^{m-a}}\, \frac{\Gamma(m+\alpha)\, \psi(\alpha+a, a-m+1, bt)}{\Gamma(n+\alpha)}\,. \end{aligned}$$

Proof. It is immediately seen from (19) that the predictive $\hat{q}_{\pi_{a,b}}(\cdot|z_{(1)})$ depends on $z_{(2)}$ only through w_1. Furthermore, the density is decreasing in w_1 as both the terms $\psi\big(\alpha+a, a-n+1, b(t+w_1)\big)$ and $\frac{1}{(w_1+t)^{n-a}}$ are decreasing in w_1 given that $a \leq n$. It then follows from (23) and (24) that the HPD region is given by (27) with

$$\mathbb{P}(W_1 \leq c_0) \,=\, 1-\lambda\,, \tag{28}$$

under density $\hat{q}_{\pi_{a,b}}(\cdot|z_{(1)})$ for $Z_{(2)}$. Now, an application of Lemma 3.1 yields the density

$$\begin{aligned} h_{W_1}(w_1) \,&=\, \frac{w_1^{n-m-1}\, \hat{q}_{\pi_{a,b}}(w_1|z_{(1)})}{(n-m-1)!\,(n-m)!} \\ &= \frac{\Gamma(n+\alpha)\, t^{m-a}}{(n-m-1)!\, \Gamma(m+\alpha)}\, \frac{\psi(\alpha+a, a-n+1, b(t+w_1))}{\psi(\alpha+a, a-m+1, bt)}\, \frac{w_1^{n-m-1}}{(w_1+t)^{n-a}}\,. \end{aligned}$$

Finally, the result follows with a re-arrangement of terms from the above and (28)
. □

4 Frequentist coverage probability and credibility

The Bayesian predictive densities of Theorems 3.1 and 3.2 for estimating the conditional density

$$Z_{(2)}|z_{(1)} =^d Z_{(2)}|t \sim \mathcal{P}_2\big(m+\alpha, \frac{n-m}{\theta+t}, \ldots, \frac{1}{\theta+t}\big)\,, \tag{29}$$

with sufficient statistic $t = \sum_{i=1}^{n}(n-i+1)z_i$, lead to Bayesian prediction regions for $Z_{(2)}$, or subvectors of $Z_{(2)}$, with prescribed credibility $1-\lambda$, as seen in Section 3.4. We focus here on the non-informative prior density choice $\pi_0(\theta) = \frac{1}{\theta}\,\mathbb{I}_{(0,\infty)}(\theta)$, the corresponding predictive density

$$\hat{q}_{\pi_0}(\cdot|t) \sim \mathcal{P}_2\big(m, \frac{n-m}{t}, \ldots, \frac{1}{t}\big)\,, \tag{30}$$

and show that prediction regions $R_{\pi_0}(Z_{(1)})$ with credibility $1-\lambda$ have matching frequentist coverage probability for all $\theta > 0$, i.e.,

$$\mathbb{P}_{\pi_0}\big(Z_{(2)} \in R_{\pi_0}(z_{(1)}) \,\big|\, z_{(1)}\big) \;=\; \mathbb{P}\big(Z_{(2)} \in R_{\pi_0}(z_{(1)}) \,\big|\, z_{(1)}, \theta\big) \;=\; 1-\lambda, \text{ for all } z_{(1)}, \theta.$$

Moreover, the above matching is shown to occur simultaneously for all mixing parameter $\alpha > 0$ in model (1). In the above equality, the left-hand side relates to the predictive distribution of $Z_{(2)}$ given $Z_{(1)} = z_{(1)}$ under π_0 (i.e., (30)), while the right-hand side relates to the frequentist conditional distribution of $Z_{(2)}$ given $Z_{(1)} = z_{(1)}$ (Lemma 2.3) which depends on θ. The equality implies the weaker property of unconditional matching coverage (i.e., prior to observing $Z_{(1)}$), which can expressed as $\mathbb{P}\big(Z_{(2)} \in R_{\pi_0}(Z_{(1)}) \,\big|\, \theta\big) = 1-\lambda$. The developments below extend those of [3] for exponentially distributed data, and also differ since the conditional distribution being estimated in (29) does depend on the observed data $z_{(1)}$. To conclude this section, we provide further insight on why the predictive density $\hat{q}_{\pi_0}$ in (30) does not depend on the model setting of $\alpha > 0$.
To pursue, consider the more general setting with densities

$$T \sim p_\sigma(t) \;=\; \frac{1}{\sigma}\,p_1(\frac{t}{\sigma})\,, \text{ and } Y|T=t \sim q_{\sigma+t}(y) \;=\; \frac{1}{(\sigma+t)^d}\,q_1(\frac{y_1}{\sigma+t}, \ldots, \frac{y_d}{\sigma+t})\,, \tag{31}$$

supported on $\mathbb{R}_+$ and $\mathbb{R}_+^d$, respectively. In our setting, p_σ will be the density of a $\mathcal{B}(m,\alpha,\sigma)$ distribution, $q_{\sigma+t}$ the Pareto density in (29), with $d = n-m$ and $\sigma = \theta$. In general, we have the following.

Lemma 4.1. *For model (31) and prior density $\pi_0(\sigma) = \frac{1}{\sigma}\,\mathbb{I}_{(0,\infty)}(\sigma)$, the Bayesian predictive density for the conditional density of $Y|T=t$ is given by*

$$\hat{q}_{\pi_0}(y|t) \;=\; \frac{1}{t^d}\,h(\frac{y}{t})\,, \tag{32}$$

where h is the frequentist density of $R = (R_1, \ldots, R_d)^\top = \frac{Y}{T}|\sigma$, which is free of σ, and given by

$$h(r) \;=\; \int_0^\infty (\frac{u}{1+u})^d\,q_1\big(\frac{ur_1}{u+1}, \ldots, \frac{ur_d}{u+1}\big)\,p_1(u)\,du\,. \tag{33}$$

Stated otherwise, the posterior predictive under π_0 and frequentist distributions match, i.e.,

$$(R_1, \ldots, R_d)|t \;=^d\; (R_1, \ldots, R_d)|\sigma \;\text{ for all } t, \sigma\,. \tag{34}$$

Proof. Identity (34) follows from the other parts of the lemma. Then, with prior density π_0, the posterior density of σ becomes $\pi_0(\sigma|t) = \frac{t}{\sigma^2} p_1(\frac{t}{\sigma}) \mathbb{I}_{(0,\infty)}(\sigma)$, in which case it follows that

$$\begin{aligned}\hat{q}_{\pi_0}(y|t) &= \int_0^\infty \frac{1}{(\sigma+t)^d} q_1\Big(\frac{y_1}{\sigma+t},\ldots,\frac{y_d}{\sigma+t}\Big) \frac{t}{\sigma^2} p_1\Big(\frac{t}{\sigma}\Big) d\sigma \\ &= \frac{1}{t^d}\int_0^\infty \Big(\frac{u}{1+u}\Big)^d q_1\Big(\frac{uy_1}{t(u+1)},\ldots,\frac{uy_d}{t(u+1)}\Big) p_1(u)\, du\,,\end{aligned}$$

which is (32). Finally, a direct evaluation of the frequentist density of the ratio $\frac{Y}{T}$ under model (31) yields the density in (33). □

Theorem 4.1. *Consider model (31), an observed value t of T, and a Bayesian predictive region $R^*_{\pi_0}(t)$ for Y with credibility $1-\lambda$ associated with prior density $\pi_0(\sigma) = \frac{1}{\sigma}\mathbb{I}_{(0,\infty)}(\sigma)$. Then $R^*_{\pi_0}(t)$ has exact frequentist conditional coverage probability, i.e.*

$$\mathbb{P}\big(Y \in R^*_{\pi_0}(T) \,\big|\, T=t,\sigma\big) = 1-\lambda, \text{ for all } t,\sigma>0\,.$$

Proof. Since $R^*_{\pi_0}(t)$ has credibility $1-\lambda$, we have

$$\mathbb{P}_{\pi_0}\big(Y \in R^*_{\pi_0}(t)|T=t\big) = \mathbb{P}_{\pi_0}\Big(\frac{Y}{T} \in S^*_{\pi_0}(t)|T=t\Big) = 1-\lambda\,,$$

where $S^*_{\pi_0}(t) = \{y \in R^d_+ : tY \in R^*_{\pi_0}(t)\}$. Since the Bayesian predictive distribution of $\frac{Y}{t}$ is free of t (see expression 32), it follows that $S^*_{\pi_0}(t) \equiv S^*_{\pi_0}$ is free of t and such that

$$\int_{S^*_{\pi_0}} h(r)\, dr = 1-\lambda.$$

On the other hand, the frequentist conditional coverage probability of the region $R_{\pi_0}(T)$ is given by

$$\mathbb{P}\big(Y \in R^*_{\pi_0}(T)|T=t,\sigma\big) = \mathbb{P}\Big(\frac{Y}{T} \in S^*_{\pi_0}(T)|T=t,\sigma\Big) = \int_{S^*_{\pi_0}} h(r)\, dr = 1-\lambda\,,$$

since $R = \frac{Y}{T}|T=t,\sigma$ has density h free of t and σ by virtue of Lemma 4.1. □

An application of the above in the context of our problem leads to the following main finding.

Corollary 4.1. *For model (1) and an observed $z_{(1)}$, Bayesian predictive regions $R_{\pi_0}(Z_{(1)})$ for $Z_{(2)}$ of credibility $1-\lambda$, based on the predictive density $\hat{q}_{\pi_0}(\cdot|z_{(1)}) \sim \mathcal{P}_2(,\frac{n-m}{t},\ldots,\frac{1}{t})$ with $t=\sum_{i=1}^m (n-i+1)z_i$, have exact frequentist conditional coverage probability $1-\lambda$, i.e.,*

$$\mathbb{P}\big(R_{\pi_0}(Z_{(1)}) \ni Z_{(2)} \,|\, Z_{(1)}=z_{(1)},\theta\big) = 1-\lambda \text{ for all } z_{(1)} \in \mathbb{R}^m_+, \theta>0\,.$$

Furthermore, the above matching result holds simultaneously for all $\alpha>0$ in model (1), (i.e., for all Gamma mixing $\beta|\theta \sim \mathcal{G}(\alpha,\theta)$), as well as for the degenerate version with $X_1,\ldots,X_n$ i.i.d.. Exp(θ).

Proof. For the degenerate case, the result is due to [3]. Since t is a sufficient statistic for θ, we can express $R_{\pi_0}(Z_{(1)})$ as $R^*_{\pi_0}(T)$ and consider the equivalent property

$$\mathbb{P}\big(R^*_{\pi_0}(T) \ni Z_{(2)} \,|\, T=t, \theta\big) \;=\; 1-\lambda \;\text{ for all } t, \theta > 0\,.$$

We can then apply Theorem 4.1 with T as given, $Y = Z_{(2)}$, the credibility region $R^*_{\pi_0}(T)$, and model (31) with a $\mathbb{B}(m, \alpha, \sigma)$ density p_σ for T and a $\mathbb{P}_2(m+\alpha, \frac{n-m}{\sigma+t}, \ldots, \frac{1}{\sigma+t})$ density $q_{\sigma+t}$ for the conditional distribution $Y|T=t$. The result then follows from Theorem 4.1. □

Remark 4.1. *Here is an alternate derivation of the property to the effect that the predictive density $\hat{q}_{\pi_0}$ given in Theorem 3.1 is free of the model choice of α. We connect it to a more general result, namely Lemma 4.1's representation, and make use of Definition 2.1. To do so, we consider the frequentist distribution of $R = \frac{Y}{T}|\beta$ with $Y = Z_{(2)}$ and $T = \sum_{i=1}^m (n-i+1) Z_i$.*

As a first step, conditioning on β in (1), it is the case that:

(i) $Z_j|\beta, j=1,\ldots,m$, are independently distributed Exp($(n-j+1)\beta$), which implies that $T|\beta \sim \mathcal{G}(m,\beta)$ and $G = \beta T|\beta \sim \mathcal{G}(m,1)$;

(ii) For $j = m+1,\ldots,n$, $Z_j|T,\beta$ are independently distributed Exp($(n-j+1)\beta$), which implies that $E_j =^d \beta Z_j|T,\beta =^d \beta Z_j|\beta$ are independently distributed Exp($(n-j+1)$).

From the above, we infer that

$$R|\beta =^d \frac{Y}{T}|\beta =^d \frac{\beta Y}{\beta T}|\beta =^d \big(\frac{E_{m+1}}{G}, \ldots, \frac{E_n}{G}\big)^\top \sim \mathcal{P}_2(n-m,\ldots,1)\,,$$

independently of β, by virtue of Definition 2.1. By invoking Lemma 4.1, we conclude that the Bayesian predictive density $\hat{q}_{\pi_0}$ is indeed a $\mathcal{P}_2(\frac{n-m}{t},\ldots,\frac{1}{t})$ density, simultaneously for all choices of α in model (1). Moreover, the matching also holds for the baseline exponential model with $X_1,\ldots,X_n$ independently distributed as Exp(θ).

Concluding remarks

In this article, for dependent data distributed as an exponential mixture and for a type-II right-censoring scheme, we have studied and derived Bayesian predictive densities for a vector of future order statistics, and related functions such as individual order statistics and totals of future order statistics. From such Bayesian predictive densities and gamma priors densities on the mixing parameter distribution, we also provide Bayesian HPD prediction regions and point predictors for the set of future order statistics. Notably, we establish that the Bayes prediction procedures associated with a non-informative prior density π_0 do not vary across the class of Gamma mixtures of exponential distributions, and are therefore robust with respect to misspecification within this class as well as for the non-mixing basis exponential distribution. Furthermore, resulting prediction regions with credibility $1-\lambda$ are shown to have matching frequentist coverage $1-\lambda$. The pivotal predictive density $\hat{q}_{\pi_0}$ is a multivariate Pareto density, and our analysis brings into play a novel study of the distribution of linear combinations of such multivariate Pareto densities.

A main motivation for this study has been to explore a dependent data structure as an alternative to i.i.d. data. A promising future avenue would be the exploration of other dependent data structures, possibly more general. It would be of interest as well to extend the analysis to other types of sampling schemes such as type-I and progressive censoring. Finally, a topic that remains unexplored is the frequentist risk performance, in a decision-theoretic framework with Kullback-Leibler divergence loss, of the density $\hat{q}_{\pi_0}$, namely in terms of admissibility and minimaxity, especially for large dimensions.

Acknowledgements

The authors are thankful to the Université de Sherbrooke for support via the visiting scholar fund, as well as to Leïla Chalifour for a diligent reading. Author Marchand is grateful for research support from NSERC of Canada.

References

[1] Arnold, B.C., Balakrishnan, N., & Nagaraja, H.N. (2008). *A first course in order statistics*. SIAM: Classics in Applied Mathematics, Philadelphia, USA.

[2] Arnold, B.C. (2014). Univariate and multivariate Pareto models. *Journal of Statistical Distributions and Applications*, **1**, article 11.

[3] Asgharzadeh, A., Marchand, É. & Saadati A. N. (2025). Bayesian prediction regions and density estimation with type-2 censored data, *IEEE Transactions on Reliability*, 74(2), 2827-2836, https://doi.org/10.1109/TR.2024.3438189

[4] Clarke, B. & Yao, Y. (2025). A cheat sheet for Bayesian prediction. *Statistical Science*, **40**, 3–24.

[5] Dunsmore, I.R. (1974). The Bayesian prediction problem in life testing models. *Technometrics*, **16**, 455–4600.

[6] Fourdrinier, D., Marchand, É. & Strawderman, W.E. (2019). On efficient prediction and predictive density estimation for spherically symmetric models. *Journal of Multivariate Analysis*, **173**, 18–25.

[7] Hewett, J.E. (1968). A note on prediction intervals based on partial observations in certain life test experiments. *Technometrics*, **10**, 850–853.

[8] Kotz, S., Balakrishnan, N. & Johnson, N.L. (2000). *Multivariate continuous distributions: volume 1*, Wiley.

[9] Kotz, S., Balakrishnan, N. & Johnson, N.L. (1995). *Continuous univariate distributions: volume 1*, Second edition. Wiley.

[10] Lawless, J.F. (1971). A prediction problem concerning samples from the exponential distributions, with application in life testing. *Technometrics*, **13**, 725–730.

[11] Maruyama, Y. (2003). A robust generalized Bayes estimator improving on the James-Stein estimator for spherically symmetric distributions. *Statistics & Decisions*, **21**, 69–77.

[12] Nayak, T. K. (1987). Multivariate Lomax distribution: properties and usefulness in reliability theory. *Journal of Applied Probability*, **24**, 170–177.

[13] Ross, S. (2023). *Introduction to Probability Models*. 13th edition. Academic Press.

[14] Sukhatme, S. (1937). Tests of significance for samples of the χ^2 population with two degrees of freedom. *Annals of Eugenics*, **8**, 52–56.

[15] Whitmore, G.A. & Lee, M.L.T. (1991). A multivariate survival distribution generated by an inverse Gaussian mixture of exponentials. *Technometrics*, **33**, 39–50.