

# Distributions of Risk Functions for the Pareto Model

Fulvio De Santis<sup>1</sup>, Stefania Gubbiotti<sup>1</sup>, Marco Perone Pacifico<sup>2</sup>

<sup>1</sup>Department of Statistical Sciences, Sapienza University of Rome, Rome, Italy

<sup>2</sup>Department of Economics and Finance, Luiss University, Rome, Italy

Email: fulvio.desantis@uniroma1.it, stefania.gubbiotti@uniroma1.it, mperonepacifico@luiss.it

**How to cite this paper:** De Santis, F., Gubbiotti, S. and Perone Pacifico, M. (2024) Distributions of Risk Functions for the Pareto Model. *Open Journal of Statistics*, **14**, 721-736.

<https://doi.org/10.4236/ojs.2024.146032>

**Received:** November 12, 2024

**Accepted:** December 3, 2024

**Published:** December 6, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

In statistical decision theory, the risk function quantifies the average performance of a decision over the sample space. The risk function, which depends on the parameter of the model, is often summarized by the Bayes risk, that is its expected value with respect to a design prior distribution assigned to the parameter. However, since expectation may not be an adequate synthesis of the random risk, we propose to examine the whole distribution of the risk function. Specifically, we consider point and interval estimation for the two parameters of the Pareto model. Using conjugate priors, we derive closed-form expressions for both the expected value and the density functions of the risk of each parameter under suitable losses. Finally, an application to wealth distribution is illustrated.

## Keywords

Bayes Risk, Point Estimation, Interval Estimation, Prior Distributions, Statistical Decision Theory

---

## 1. Introduction

Statistical decision functions—such as point and set estimates or test statistics—are typically compared in terms of loss functions. In frequentist decision theory, the risk function is the expected value of the loss with respect to the sampling distribution of the data, and it depends on the parameter of the model. The risk function is usually summarized by optimality criteria, *i.e.* by applying a suitable real-valued functional (see for instance [1]). We here focus on Bayesian criteria that require a prior probability distribution for the parameter of the model, which induces a distribution on the risk function as well. Among several alternatives, the most popular Bayesian criterion is the Bayes risk, the expected value of the risk

function with respect to the prior. This approach is often referred to as *hybrid frequentist-Bayesian* since it is based on a Bayesian summary of a frequentist risk function. However, relying solely on averaging might be a limitation, since the expected value is not always a good summary of the entire distribution of a random variable. In the context of clinical trials, this point has been raised by several authors for the power function of a test, a quantity that is closely related to the risk. For instance, [2]-[5] argue that examining the whole distribution of the random power guarantees a better insight into the probability of success of an experiment. Specifically, the authors in [3] consider one-sided testing on the location parameter of a normal model and point out that the expected value of the power may be a poor representation of its distribution. Therefore, they derive the expressions of the cumulative distribution function and probability density function of the random power and study their qualitative features. Along these lines, in [5], this approach is adapted to the case of the scale parameter of distributions that belong to exponential families.

In the present article, we borrow the aforementioned ideas, and we propose to study the probability distribution of the risk function induced by a prior in the context of point and interval estimation. Inspection of the shape of the random risk distribution allows one to assess the impact of prior assumptions and sample size on the quality of a candidate estimator. In particular, the proposed approach is developed for both point and set estimation of the parameters of the Pareto model. We show that, for all considered problems, the random risk functions are scale transformations either of the random parameters or of their square. Hence, for any generic design prior, we find explicit expressions for the expected value and the probability density function (pdf) of the risk under standard loss functions. Furthermore, assuming conjugate priors, we show that the resulting risk density functions are still related to the design prior family.

The paper is structured as follows. In Section 2, we introduce notations and formalize the problem for the Pareto model. In Section 3, we derive explicit results for the risk functions of the parameters of the model: specifically, in Sections 3.1 and 3.2, we focus on shape and scale parameters respectively. Section 4 illustrates an application of the proposed methodology for the estimation of the Pareto index for the wealth distribution of the World's Billionaires. Finally, Section 5 contains a discussion.

## 2. Methodology

Let  $\mathbf{X}_n = (X_1, \dots, X_n)$  be a random sample,  $f_n(\mathbf{x}_n; \xi)$  its distribution depending on an unknown parameter  $\xi \in \Xi$ , where  $\Xi$  is the parameter space. Let  $\omega$  be a scalar function of  $\xi$  and let  $d(\mathbf{X}_n)$  be a point or an interval estimator of  $\omega$ , with  $d(\mathbf{x}_n)$  denoting the corresponding estimate. Frequentist decision theory typically employs the *risk function*

$$R(\omega, d) = \mathbb{E}_\omega \left[ \mathbb{L}(\omega, d(\mathbf{X}_n)) \right], \quad (1)$$

to assess the performance of the decision function, where  $\mathbb{L}(\omega, d(\mathbf{X}_n)) \geq 0$  is the *loss function* of  $d$  and  $\mathbb{E}_\omega[\cdot]$  is the expected value with respect to the sampling distribution  $f_n(\cdot|\omega)$ . For any decision function  $d$ ,  $R(\omega, d)$  is a function of  $\omega$ . Within the hybrid frequentist-Bayesian framework, the parameter is thought as a random variable denoted by  $\Omega$ , with prior distribution  $\pi_D(\omega)$ , where the subscript  $D$  is referred to the *design*. In fact, this is often called design prior, to stress its use in pre-experimental evaluation of the decision  $d$ . The expected value of  $R(\Omega, d)$  with respect to  $\pi_D(\omega)$  is the *Bayes risk* of  $d$  defined as  $r_{\pi_D}(d) = \mathbb{E}_{\pi_D}[R(\Omega, d)]$ . If  $\Omega$  is an absolutely continuous random variable and  $\pi_D(\omega)$  its density function,

$$r_{\pi_D}(d) = \int_{\Xi} R(\omega, d) \pi_D(\omega) d\omega. \quad (2)$$

The Bayes risk is typically employed for evaluation of a given  $d$ , comparison of alternative estimators and identification of optimal decision functions. However, a thorough inspection of the features of the random risk function  $R(\Omega, d)$  might be more informative than looking at its expectation only. More precisely, let

$$Y = R(\Omega, d) \quad (3)$$

be the random risk function and  $\mathcal{Y}$  its sample space, and let  $g$  denote the pdf of  $Y$ . The behavior of the random risk can be studied through the density  $g$  and some of its relevant summaries. For instance, the Bayes risk can be retrieved as the expected value of  $Y$ , *i.e.*

$$\mathbb{E}[Y] = \int_{\mathcal{Y}} y \cdot g(y) dy = \int_{\Xi} R(\omega, d) \pi_D(\omega) d\omega. \quad (4)$$

In Section 3, we consider point and interval estimators for the parameters of the Pareto model and we derive the closed-form expression for  $g$  and  $r_{\pi_D}$  under standard loss functions.

## 2.1. Pareto Model

Suppose that  $X_1, \dots, X_n$  is a random sample from the following Pareto density with parameters  $(\theta, \eta)$

$$f_X(x_i; \theta, \eta) = \frac{\theta \eta^\theta}{x_i^{\theta+1}} I_{[\eta, \infty)}(x_i), \quad \theta > 0, \eta > 0, \quad (5)$$

where  $\theta$  is the shape parameter, often denoted as *Pareto index*, and  $\eta$  is the scale parameter. In the following, when considering the parameters  $\theta$  and  $\eta$  random, they will be denoted as  $\Theta$  and  $H$ .

In Sections 3.1 and 3.2, we derive closed-form expressions for  $r_{\pi_D}$  and  $g$  for point estimators and interval estimators of  $\theta$  and  $\eta$ . Specifically, for point estimation, we consider maximum likelihood estimators and quadratic loss function, whereas for interval estimation, we consider the length of  $1-\gamma$  confidence intervals as loss function. We show that all the corresponding random risk functions are scale transformations either of the random parameters or of their square. We also show that, as a consequence, by adopting conjugate densities as design priors

(generalized Gamma and Pareto respectively for  $\Theta$  and  $H^{-1}$ ), each resulting density  $g$  belongs to the same family of the design prior.

### 2.2. Elicitation of Conjugate Design Prior Distributions

We consider independent conjugate design prior distributions for the two parameters. In particular, we adopt a Gamma density for the Pareto index  $\Theta$  and a Pareto density for the reciprocal of the scale parameter  $H$ . Prior parameters are usually elicited using moments (such as mean and variance) or relevant quantiles of historical data.

*Prior for  $\Theta$ .* For a given pair  $(\nu, \lambda)$  (both positive), the prior for the Pareto index  $\theta$  is the Gamma density

$$\pi_D(\theta) = \frac{\lambda^\nu}{\Gamma(\nu)} \theta^{\nu-1} e^{-\lambda\theta}, \quad \theta > 0. \tag{6}$$

If prior expectation  $m$  and variance  $v$  are elicited for the shape parameter, for instance, on the basis of historical data, the corresponding values for the prior parameters  $\nu$  and  $\lambda$  are

$$\nu = \frac{m^2}{v} \quad \text{and} \quad \lambda = \frac{m}{v}. \tag{7}$$

*Prior for  $H$ .* Adopting a Pareto density for  $H^{-1}$  implies that the prior density of the scale parameter  $H$  is

$$\pi_D(\eta) = \alpha\beta^\alpha \eta^{\alpha-1}, \quad \eta \in \left[0, \frac{1}{\beta}\right]. \tag{8}$$

This density is often referred to as *Inverse Pareto with parameters*  $(\alpha, \beta)$ , but the denomination is not unique in the literature (see, for instance, [6]).

The two parameters  $\alpha$  and  $\beta$  are usually chosen by eliciting: (i) expectation  $m$  and variance  $v$ ; (ii) some quantiles. In case (i), the resulting pair of hyperparameters is

$$\alpha = \sqrt{1 + \frac{m^2}{v}} - 1 \quad \text{and} \quad \beta = \frac{\sqrt{1 + \frac{m^2}{v}} - 1}{m\sqrt{1 + \frac{m^2}{v}}}. \tag{9}$$

In case (ii) two quantiles are elicited (for example  $q_{\gamma_1}$  and  $q_{\gamma_2}$  are the quantiles at level  $\gamma_1$  and  $\gamma_2$  respectively, with  $\gamma_1 \neq \gamma_2$ ), then the prior parameters are

$$\alpha = \frac{\ln\left(\frac{\gamma_1}{\gamma_2}\right)}{\ln\left(\frac{q_{\gamma_1}}{q_{\gamma_2}}\right)} \quad \text{and} \quad \beta = \frac{\frac{1}{q_{\gamma_1}}}{q_{\gamma_1}}. \tag{10}$$

#### Example: Income Data

We now show the elicitation procedure to obtain distributions that will be used

for implementing our proposed methodology in the examples of Sections 3.1 and 3.2. In this regard, we exploit data from a numerical example in Section 5 of [7]. The authors consider a dataset previously analyzed by [8] on annual wage (in multiples of 100 US dollars) relative to a random sample of 30 production-line workers in a large industrial firm. Specifically, income data are analyzed by adopting a Pareto model for the observations and independent conjugate densities for the two parameters. In our examples, we use the same prior assumptions as in [7], that are:

- (i) expectation  $m = 1.9$  and variance  $v = 0.6$  for the shape parameter  $\Theta$ ;
- (ii) median and 5th percentile equal to 100 and 85 respectively, for the scale parameter  $H$ .

Then, according to (7) and (10), we elicit the prior hyperparameters as follows:

- (i) prior hyperparameters for  $\Theta$

$$v = \frac{1.9^2}{0.6} = 6.02 \quad \text{and} \quad \lambda = \frac{1.9}{0.6} = 3.17; \quad (11)$$

- (ii) prior hyperparameters for  $H$

$$\alpha = \frac{\ln\left(\frac{0.50}{0.05}\right)}{\ln\left(\frac{100}{85}\right)} = 14.17 \quad \text{and} \quad \beta = \frac{0.50^{\frac{1}{\alpha}}}{100} = 0.009. \quad (12)$$

These values are employed as hyperparameters in the design priors for  $\Theta$  and  $H$  in Sections 3.1 and 3.2, where we continue this example by deriving the density of the random risk for point and interval estimators.

### 3. Distribution of Risk Functions

In this section, we provide explicit expressions of risk functions, Bayes risks and densities  $g(y)$  for point and interval estimation of the parameters  $(\theta, \eta)$  of the Pareto model. In Section 3.3, we extend the analysis to inference on a Pareto model quantile that is a function of both parameters.

#### 3.1. Inference on the Pareto Index $\theta$ ( $\eta$ Known)

##### 3.1.1. Point Estimation

It is easy to check that (see [9]) the maximum likelihood (ML) estimator of  $\theta$  is

$$\hat{\theta} = \frac{n}{\sum \ln \frac{x_i}{\eta}}$$

that is distributed as an Inverse Gamma with shape  $n$  and scale  $n\theta$  [10]. Noting

that  $\mathbb{E}[\hat{\theta}] = \frac{n\theta}{n-1}$  and that  $\mathbb{V}[\hat{\theta}] = \frac{n^2\theta^2}{(n-1)^2(n-2)}$ , using the quadratic loss

function we obtain

$$R(\theta, \hat{\theta}) = \frac{n^2 + n - 2}{(n-1)^2(n-2)}\theta^2 = \frac{n+2}{(n-1)(n-2)}\theta^2 = k\theta^2.$$

Let us now consider the random risk of  $\hat{\theta}$ , that is  $Y = R(\Theta, \hat{\theta}) = k\Theta^2$ . For any design prior density  $\pi_D(\theta)$  on  $\Theta$ , we have that

$$r_{\pi_D}(\hat{\theta}) = k\mathbb{E}_{\pi_D}[\Theta^2] \tag{13}$$

$$g(y) = \frac{1}{2}\sqrt{\frac{1}{ky}}\pi_D\left(\sqrt{\frac{y}{k}}\right), \quad y \geq 0. \tag{14}$$

Assuming for  $\Theta$  a conjugate Gamma prior with shape parameter  $\nu$  and rate parameter  $\lambda$ , from Equations (13) and (14) it follows that

$$r_{\pi_D}(\hat{\theta}) = \frac{k\nu(\nu+1)}{\lambda^2} \tag{15}$$

$$g(y) = \frac{\lambda^\nu}{2k^2\Gamma(\nu)}y^{\frac{\nu}{2}-1}e^{-\lambda\sqrt{\frac{y}{k}}}, \quad y \geq 0 \tag{16}$$

that is a Generalized Gamma density [10] with parameters  $(c_1 = \nu, c_2 = \frac{1}{2}, c_3 = \frac{\lambda^2}{k})$  in the parameterization of [11]. This result is a consequence of the closure of the Generalized Gamma distribution under both scale and power transformations.

### 3.1.2. Interval Estimation

Since  $\hat{\theta}$  follows an Inverse Gamma distribution, it is easy to check that (see [9])  $n\theta/\hat{\theta}$  is a pivotal quantity for  $\theta$  with distribution  $\text{Gamma}(n, 1)$ . Therefore, the  $1-\gamma$  equal-tails confidence interval is

$$C_\theta = \left[ \frac{q_{\frac{\gamma}{2}}}{n}\hat{\theta}, \frac{q_{1-\frac{\gamma}{2}}}{n}\hat{\theta} \right],$$

where  $q_\epsilon$  denotes the  $\epsilon$ -level quantile of the  $\text{Gamma}(n, 1)$ . Using the length of the interval as loss function, we have

$$\mathbb{L}(\theta, C_\theta) = \frac{q_{1-\frac{\gamma}{2}} - q_{\frac{\gamma}{2}}}{n}\hat{\theta}$$

and the resulting risk function is

$$R(\theta, C_\theta) = \frac{q_{1-\frac{\gamma}{2}} - q_{\frac{\gamma}{2}}}{n}\mathbb{E}_\theta[\hat{\theta}] = h\theta,$$

where  $h = \left( q_{1-\frac{\gamma}{2}} - q_{\frac{\gamma}{2}} \right) / n$ . Note that, for any  $\theta$ ,  $R(\theta, C_\theta)$  tends to 0 as  $n$  diverges.

In this case, the random risk is  $Y = R(\Theta, C_\theta) = h\Theta$  and it is straightforward to check that

$$r_{\pi_D}(C_\theta) = h\mathbb{E}_{\pi_D}(\Theta),$$

$$g(y) = \frac{1}{h}\pi_D\left(\frac{y}{h}\right), \quad y \geq 0.$$

Using again for  $\Theta$  a Gamma prior density of parameters  $(\nu, \lambda)$ , we have that

$$r_{\pi_D}(C_\theta) = \frac{\nu h}{\lambda}$$

$$g(y) = \frac{\left(\frac{\lambda}{h}\right)^\nu}{\Gamma(\nu)} y^{\nu-1} e^{-\frac{\lambda}{h}y}, \quad y \geq 0. \tag{17}$$

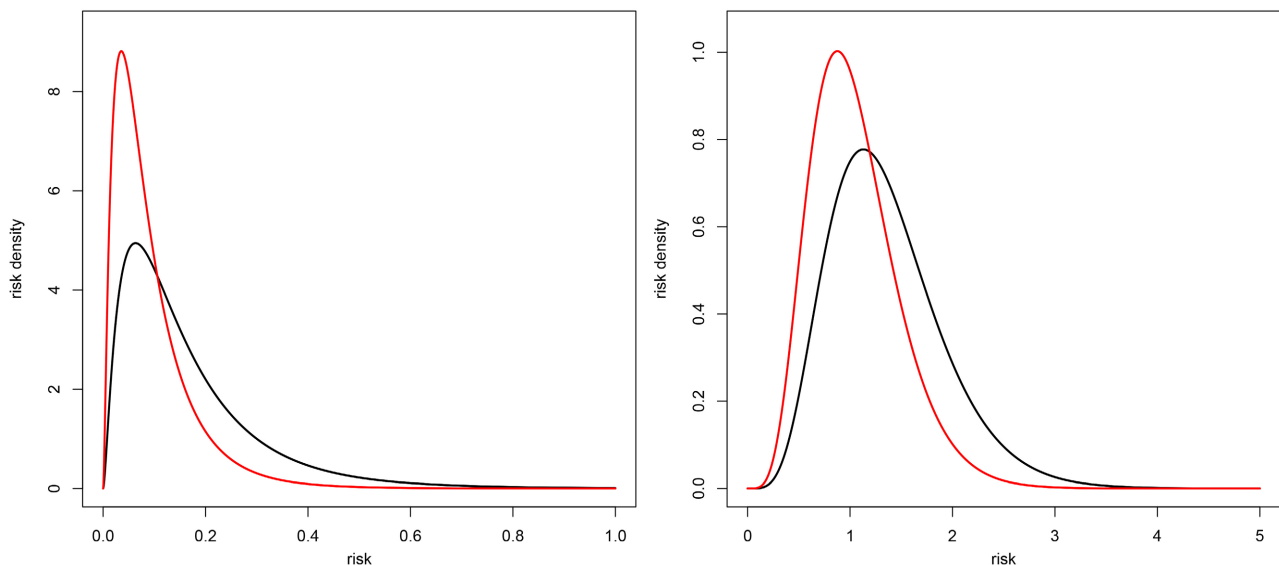
As a consequence of the above-mentioned closure under scale transformation,  $g$  is still a Gamma density with parameters  $\left(\nu, \frac{\lambda}{h}\right)$ .

**3.1.3. Example: Income Data (Cont.)**

Let us consider the income data example in Section 2.2 again. Assume now that the scale parameter  $\eta$  is known while the Pareto index  $\theta$  needs to be estimated. Values of prior hyperparameters are given by (11).

**Figure 1** shows: risk density (16) relative to the ML estimator (left panel); risk density (17) relative to the equal-tails 95%-confidence interval (right panel). In both cases, as an example, we consider sample sizes of  $n = 30$  and  $n = 50$ . Note that the density of the risk function shrinks towards 0 as the sample size increases, as a consequence of the consistency of the estimators. Moreover, for larger and larger values of  $n$ , the skewness of the distribution reduces, which results in closer and closer values of the main summaries of  $g$ . Expectation, median and mode of  $g$  are reported in **Table 1** for  $n = 30$  and  $n = 50$  for point estimation.

Since  $g$  is remarkably right-skewed, the mean is larger than the median and the mode; discrepancies between the three summaries are smaller when the sample size increases.



**Figure 1.** Example (income data). Density  $g(y)$  of the risk function when estimating Pareto index for  $n = 30$  (black curve) and  $n = 50$  (red curve). Left panel: point estimation, density (16); right panel: interval estimation, density (17).

**Table 1.** Example (income data). Summaries of the density  $g$  of the random risk function for point estimation of  $\theta$ .

|        | $n = 30$ | $n = 50$ |
|--------|----------|----------|
| mean   | 0.166    | 0.093    |
| median | 0.127    | 0.071    |
| mode   | 0.063    | 0.036    |

### 3.2. Inference on the Scale Parameter $\eta$ ( $\theta$ Known)

#### 3.2.1. Point Estimation

The maximum likelihood estimator for the parameter  $\eta$  is the sample minimum  $\hat{\eta} = X_{(1)}$  that is a Pareto random variable with parameters  $(n\theta, \eta)$ . Therefore, under the quadratic loss function,

$$R(\eta, \hat{\eta}) = \frac{2}{(n\theta - 1)(n\theta - 2)} \eta^2 = k\eta^2.$$

Then, the random risk for  $\hat{\eta}$  is  $Y = R(H, \hat{\eta}) = kH^2$  and, similarly to the previous case,

$$r_{\pi_D}(\hat{\eta}) = k \mathbb{E}_{\pi_D} [H^2]$$

$$g(y) = \frac{1}{2} \sqrt{\frac{1}{ky}} \pi_D \left( \sqrt{\frac{y}{k}} \right), \quad y \geq 0,$$

where  $\pi_D(\eta)$  is the design prior density for  $H$  and  $k = 2/[(n\theta - 1)(n\theta - 2)]$ .

When we assume a conjugate prior (8) for  $H$ , i.e. an inverse Pareto density with parameters  $(\alpha, \beta)$ , we obtain

$$r_{\pi_D}(\hat{\eta}) = \frac{\alpha k}{\beta^2 (\alpha + 2)} \tag{18}$$

$$g(y) = \frac{\alpha}{2} \left( \frac{\beta^2}{k} \right)^{\frac{\alpha}{2}} y^{\frac{\alpha}{2}-1}, \quad y \in \left[ 0, \frac{k}{\beta^2} \right], \tag{19}$$

that is an inverse Pareto density with parameters  $(\alpha/2, \beta^2/k)$ . Note that closure with respect to scale and power transformations also holds for the distribution of the reciprocal of a Pareto random variable.

#### 3.2.2. Interval Estimation

A pivotal quantity for determining an interval estimator for  $\eta$  is  $\hat{\eta}/\eta$  that is a Pareto random variable with parameters  $(n\theta, 1)$ . Recalling that the  $1 - \gamma$  quantile of this distribution is  $\gamma^{-\frac{1}{n\theta}}$ , the  $1 - \gamma$  confidence interval for  $\eta$  is

$$C_\eta = \left[ \gamma^{\frac{1}{n\theta}} \hat{\eta}, \hat{\eta} \right].$$

The length of the interval is  $\left( 1 - \gamma^{\frac{1}{n\theta}} \right) \hat{\eta}$  and the risk function is



$$R(\eta, C_\eta) = \left(1 - \gamma^{\frac{1}{n\theta}}\right) \mathbb{E}_\eta[\hat{\eta}] = \frac{n\theta \left(1 - \gamma^{\frac{1}{n\theta}}\right)}{n\theta - 1} \eta = h\eta$$

with  $h$  decreasing to 0 with  $n$ . The random risk of  $C_\eta$  is  $Y = R(H, C_\eta) = hH$ . Assuming again the conjugate prior density (8) for  $\eta$  we obtain

$$r_{\pi_D}(C_\eta) = \frac{\alpha h}{\beta(\alpha + 1)} \tag{20}$$

and the resulting  $g$  is again a density of the form (8) with parameters  $(\alpha, \beta/h)$

$$g(y) = \alpha \left(\frac{\beta}{h}\right)^\alpha y^{\alpha-1}, \quad y \in \left[0, \frac{h}{\beta}\right]. \tag{21}$$

### 3.2.3. Remark

We notice that the forms  $kH^2$  and  $hH$  of the risk functions are not characteristic of the estimation method considered in the previous section. As an example, the moment estimator for  $\eta$  is  $\hat{\eta}_m = \theta^{-1}(\theta - 1)\bar{X}_n$ , with random risk function  $Y = R(H, \hat{\eta}_m) = k_m H^2$ , where  $k_m = [n\theta(\theta - 2)]^{-1}$ .

As regards interval estimation, recalling that the asymptotic distribution of  $\hat{\eta}_m$  is normal with mean  $\eta$  and variance  $k_m \eta^2$ , the asymptotic  $1 - \gamma$  confidence interval is  $\hat{\eta}_m \pm z_{\frac{1-\gamma}{2}} \hat{\eta}_m \sqrt{k_m}$ , with length  $2z_{1-\gamma/2} \hat{\eta}_m \sqrt{k_m}$ , where  $z_\epsilon$  is the  $\epsilon$  quantile of a standard normal. Then the random risk is given by  $h_m H$ , where  $h_m = 2z_{\frac{1-\gamma}{2}} [n\theta(\theta - 2)]^{-\frac{1}{2}}$  that decreases to 0 with  $n$ .

### 3.2.4. Example: Income Data (Cont.)

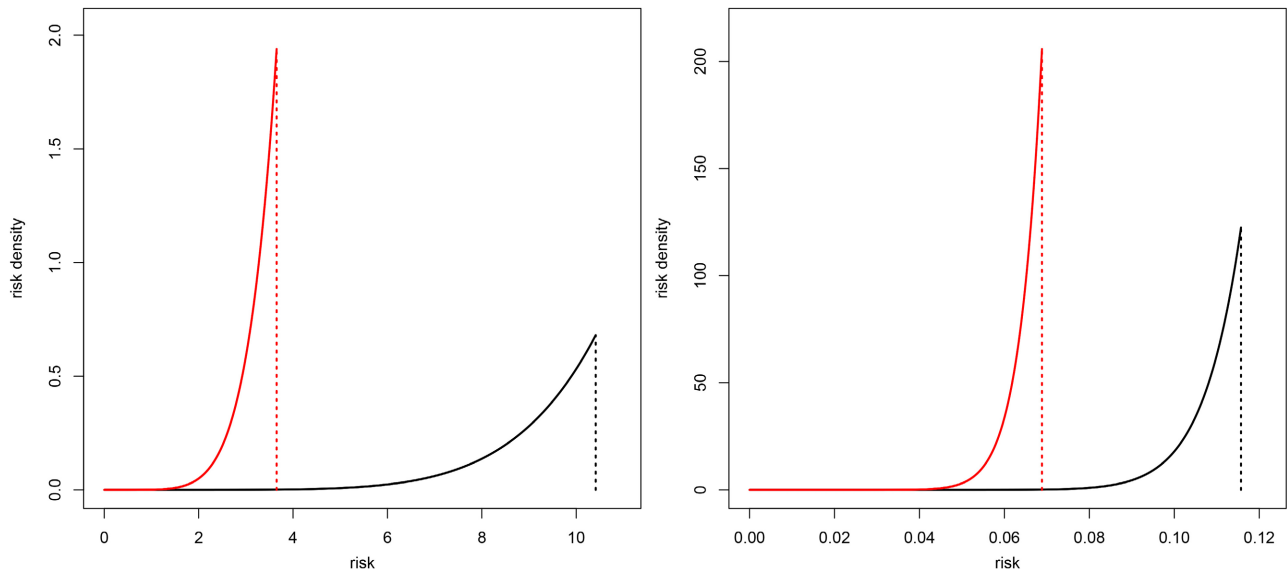
Let us consider the income data example in Section 2.2 again. Assume now that the scale parameter  $\eta$  needs to be estimated while the Pareto index is known to be equal to  $\theta = 1.584$  (that corresponds to the mode of the Gamma prior density elicited in 0). Values of prior hyperparameters for  $H$  are given by (12).

**Table 2.** Example (income data). Summaries of the density  $g$  of the random risk function for point estimation of  $\eta$ .

|        | $n = 30$ | $n = 50$ |
|--------|----------|----------|
| mean   | 9.124    | 3.201    |
| median | 9.442    | 3.312    |
| mode   | 10.413   | 3.652    |

**Figure 2** shows: risk density (19) relative to the ML estimator (left panel); risk density (21) relative to the equal-tails 95%-confidence interval (right panel). In both cases, as an example, sample sizes of  $n = 30$  and  $n = 50$  have been considered. Again, a larger value of the sample size induces higher densities for lower values of the risk function of  $\eta$ . Note however that in this case,  $g$  is always an increasing function (which is remarkably left-skewed) but, again, the values of the

main summaries of  $g$  become closer and closer as  $n$  increases. Numerical values of expectation, median and mode of  $g$  are given in **Table 2** for  $n=30$  and  $n=50$  for point estimation.



**Figure 2.** Example (income data). Density  $g(y)$  of the risk function when estimating the scale parameter for  $n=30$  (black curve) and  $n=50$  (red curve). Left panel: point estimation, density (19); right panel: interval estimation, density (21).

### 3.3. Inference on Quantiles

In many relevant applications that involve the Pareto model, the quantity of inferential interest is often represented by a specific quantile of the distribution, for instance, Value at Risk [12]. In this section, we extend the approach of the previous sections to this problem. Specifically, for a Pareto distribution with parameters  $(\theta, \eta)$ , the  $\gamma$ -quantile is  $q_\gamma = \eta(1-\gamma)^{\frac{1}{\theta}}$ , which is a function of both parameters. The maximum likelihood estimator for this quantity is

$$\hat{q}_\gamma = \hat{\eta}(1-\gamma)^{\frac{1}{\hat{\theta}}}. \tag{22}$$

For the sake of analytical tractability, let us consider the following loss function

$$\mathbb{L}(q_\gamma, \hat{q}_\gamma) = (\ln \hat{q}_\gamma - \ln q_\gamma)^2,$$

*i.e.* the quadratic loss on the logarithmic scale. The corresponding risk function is

$$R(q_\gamma, \hat{q}_\gamma) = \mathbb{E}_{\theta, \eta} [\mathbb{L}(q_\gamma, \hat{q}_\gamma)] = \frac{2 + \ln(1-\gamma) + n \ln(1-\gamma)^2}{n^2 \theta^2} = \frac{k}{\theta^2}$$

with  $k = n^{-2} [2 + \ln(1-\gamma) + n \ln(1-\gamma)^2]$ .

It follows that  $Y = k\Theta^{-2}$  and its density is

$$g(y) = \frac{\sqrt{k}}{2} y^{-\frac{3}{2}} \pi_D \left( \sqrt{\frac{k}{y}} \right), \quad y \geq 0.$$

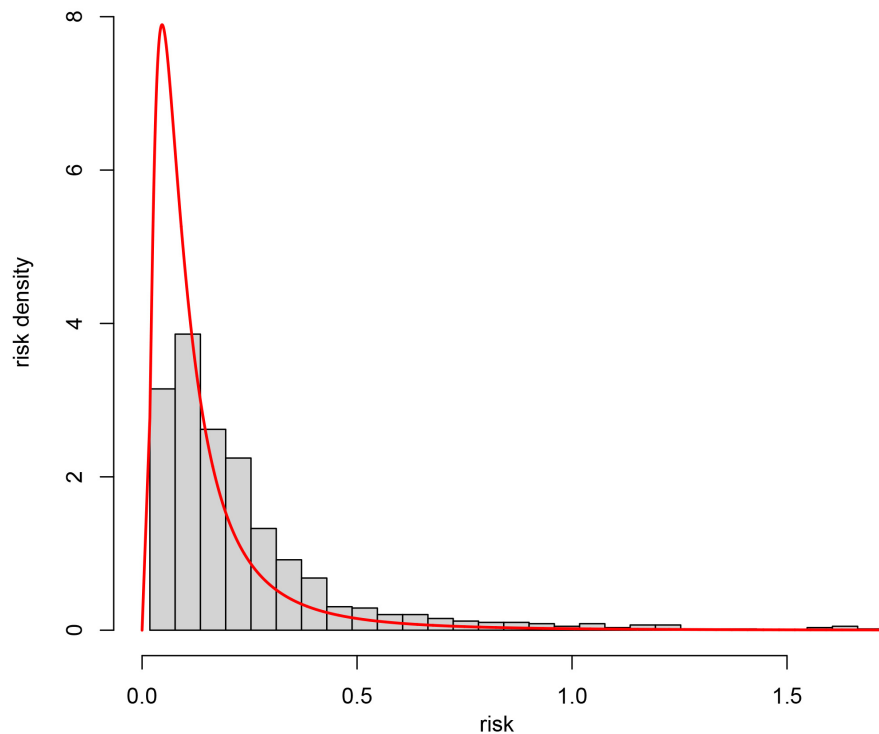
When a Gamma prior with parameters  $(\nu, \lambda)$  is adopted, the expression of  $g$  becomes

$$g(y) = \frac{(\lambda^2 k)^{\frac{\nu}{2}}}{2\Gamma(\nu)} y^{-\frac{\nu}{2}-1} e^{-\frac{\lambda\sqrt{k}}{\sqrt{y}}}, \quad y \geq 0, \quad (23)$$

that is the density function of a Generalized Inverse Gamma of parameters  $(c_1 = \nu, c_2 = 1/2, c_3 = \lambda^2 k)$  in the parametrization of [11].

### Example: Income Data (Cont.)

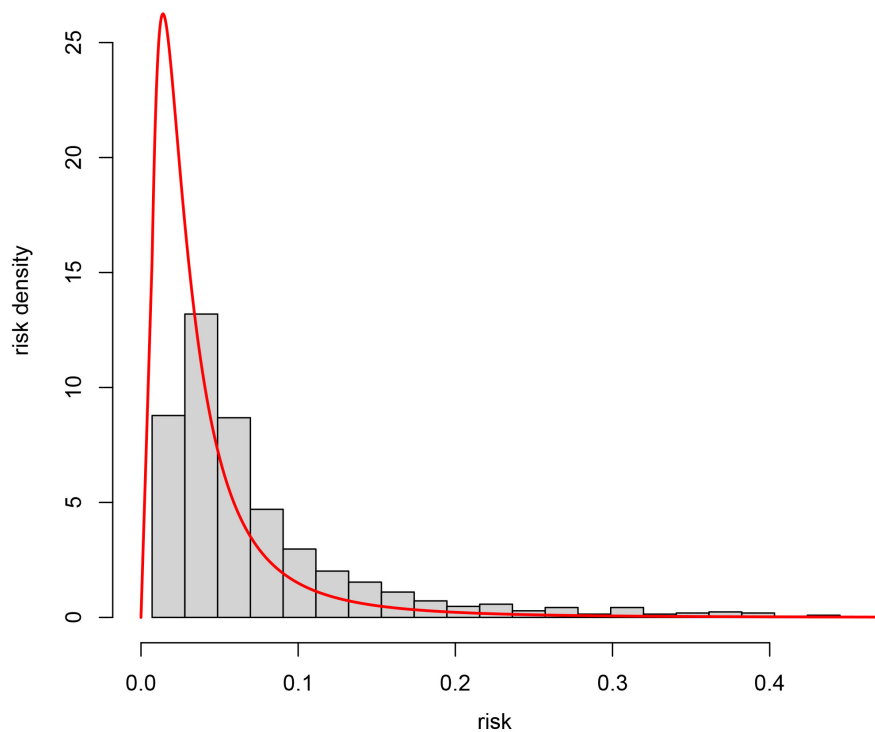
Risk density can be a useful tool when comparing different estimators. Using the same prior information as in the income data example of Section 2.2, *i.e.* (11) and (12), we consider the maximum likelihood estimator (22) of the 95th percentile of a Pareto population and the nonparametric estimator obtained as the 95th sample percentile. Since we do not have an analytic expression for the distribution of the risk relative to the nonparametric estimator, we have simulated the distribution of the risk, whose histogram is reported in **Figure 3** and **Figure 4**. The red curves in the same figure are the plot of the density (23) of the risk relative to the maximum likelihood estimator. The advantage, in terms of risk, of the maximum likelihood estimator seems to be quite evident from **Figure 3** and **Figure 4**. **Table 3** reports the mean and median of the risk associated with the two estimators for moderate ( $n = 30$ ) and large ( $n = 100$ ) sample sizes.



**Figure 3.** Example (income data). Histogram of the simulated risk relative to the nonparametric estimator of the 95th quantile and risk density (red curve) relative to the maximum likelihood estimator for samples of size  $n = 30$ .

**Table 3.** Example (income data). Summaries of the density  $g$  of the random risk for the ML estimator and the nonparametric estimator of the quantile.

| ML estimator            |          |           |
|-------------------------|----------|-----------|
|                         | $n = 30$ | $n = 100$ |
| mean                    | 0.148    | 0.045     |
| median                  | 0.092    | 0.028     |
| Nonparametric estimator |          |           |
|                         | $n = 30$ | $n = 100$ |
| mean                    | 0.278    | 0.095     |
| median                  | 0.171    | 0.057     |



**Figure 4.** Example (income data). Histogram of the simulated risk relative to the nonparametric estimator of the 95th quantile and risk density (red curve) relative to the maximum likelihood estimator for samples of size  $n = 100$ .

#### 4. Application: Prediction of Pareto Index for the World’s Billionaires

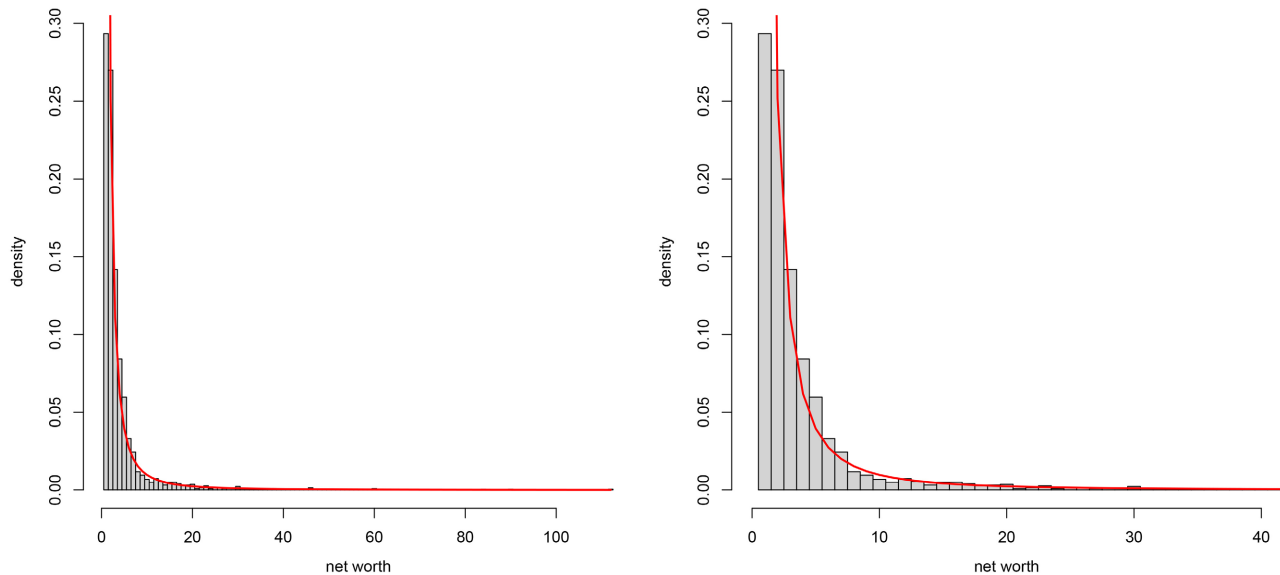
As argued by [13]: “The Pareto distribution is commonly used to represent situations where a small portion of the population controls a disproportionately large share of resources, such as income or wealth distribution.”

In the present application, we consider the World’s Billionaires List which is published yearly by Forbes Magazine at <https://www.forbes.com/billionaires/>.

Historical data reporting wealth (net worth) of billionaires are available at

[https://stats.areppim.com/stats/links\\_billionairexlists.htm](https://stats.areppim.com/stats/links_billionairexlists.htm).

We specifically refer to data relative to 2018, which is the last complete list available. The two plots in **Figure 5** show that the Pareto density fits quite accurately the data.



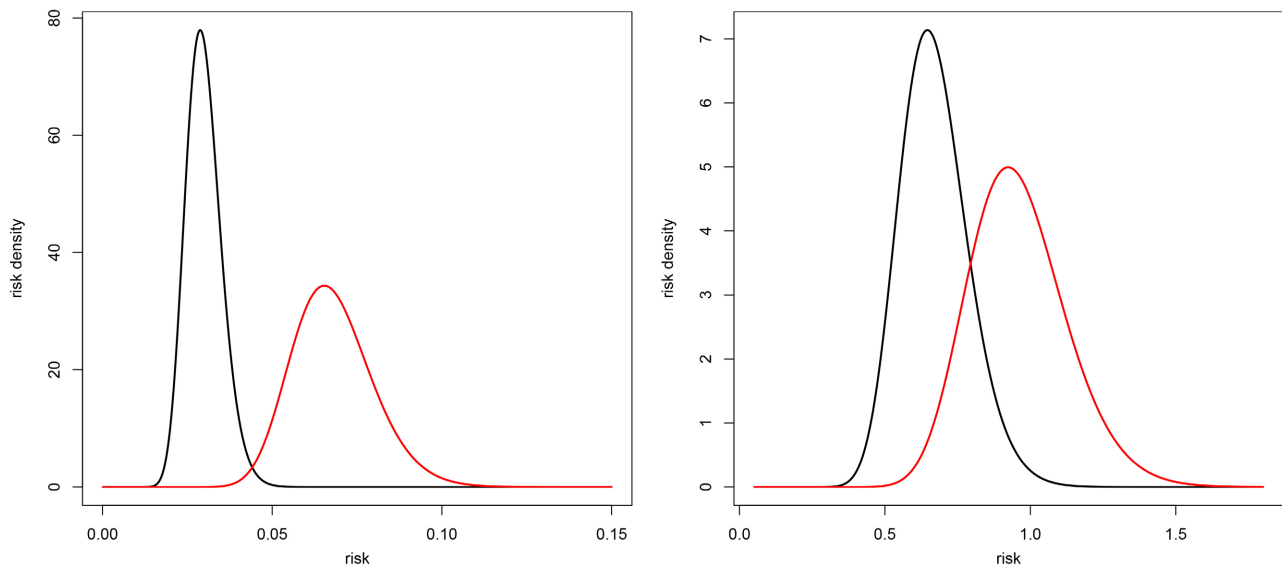
**Figure 5.** Application (World's Billionaires). Fit of the Pareto distribution with the 2018 billionaires' net worths (left panel: all distribution; right panel: most crowded net worth interval 0 - 40 billion).

Our goal is to obtain the density  $g(y)$  of the random risk for point and interval estimators of next year's Pareto index. To elicit the design prior  $\pi_D(\theta)$ , we use the above mentioned 2018 list as historical data. As in Section 3.1 we adopt a Gamma design prior density of parameters  $(\nu, \lambda)$ . These values are fixed using Equation (7) where  $m$  and  $\nu$  are based on historical data. Specifically, we set  $m = 1.024$  (i.e. the ML estimate of  $\theta$  based on 2018 historical data) and  $\nu = 0.008$  (i.e. estimated variance of the ML estimates of  $\theta$  in the period 2001-2018). The resulting values for the prior hyperparameters are  $\nu = 130.724$  and  $\lambda = 127.712$ . Using (16) and (17), we obtain **Figure 6** which shows the densities of the random risk for point and interval estimation of  $\theta$ . We consider two different sample sizes ( $n = 20$  and  $n = 40$ ) to highlight the advantage that larger samples produce in terms of loss: as  $n$  increases, the distribution of the risk tends to be more concentrated on values close to 0. In this case,  $g$  is substantially symmetric and the values of Bayes risk, median and mode for both point and interval estimation for  $n = 20$  and  $n = 40$  are almost coincident, as reported in **Table 4**.

For additional insight, in **Figure 7**, we consider a much larger prior variance (sixteen times as much), with other prior choices being equal. The increase in prior variance amplifies the skewness of the distributions and, eventually, reduces the gain in terms of loss produced by a larger sample size. As expected, in this case, the values of the mean are larger than those of the median and mode for both point and interval estimators (see **Table 5**).

**Table 4.** Application (World’s Billionaires). Summaries of the density  $g$  of the random risk for point and interval estimator of  $\theta$ .

| ML estimator        |          |          |
|---------------------|----------|----------|
|                     | $n = 20$ | $n = 40$ |
| mean                | 0.068    | 0.030    |
| median              | 0.067    | 0.030    |
| mode                | 0.065    | 0.029    |
| Confidence interval |          |          |
|                     | $n = 20$ | $n = 40$ |
| mean                | 0.893    | 0.633    |
| median              | 0.891    | 0.631    |
| mode                | 0.886    | 0.628    |

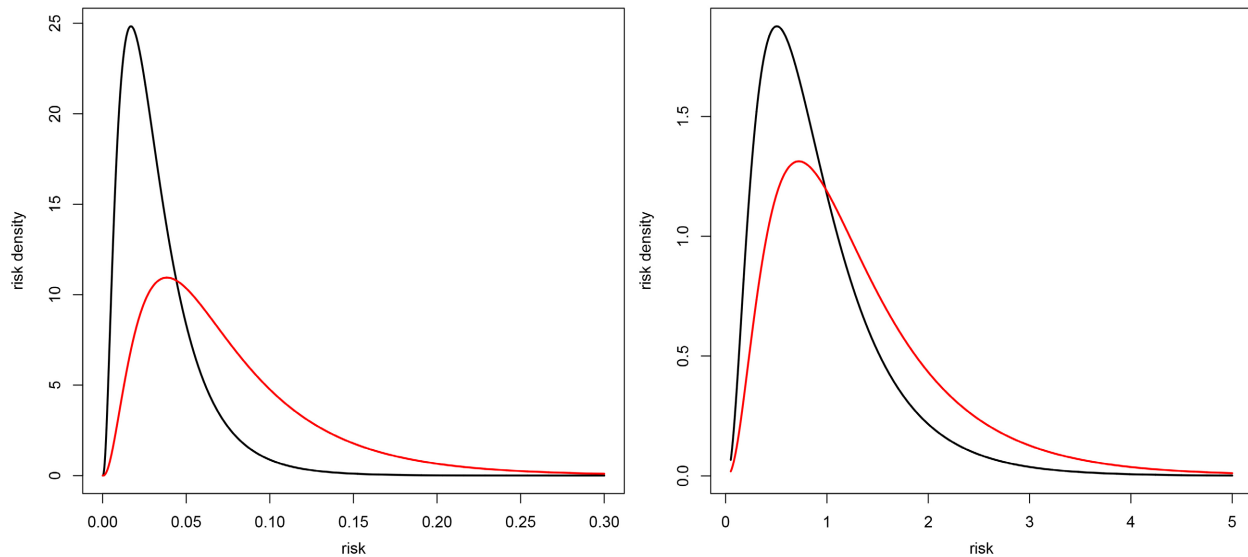


**Figure 6.** Application (World’s Billionaires). Density  $g(y)$  of the random risk for maximum likelihood estimator (left panel) and for 95% confidence interval (right panel) for the Pareto index  $\theta$ . Curves obtained for  $n = 20$  (red) and for  $n = 40$  (black), given prior variance 0.008.

**Table 5.** Application (World’s Billionaires). Summaries of the density  $g$  of the random risk for point and interval estimator of  $\theta$  when considering a much larger prior variance (sixteen times as much).

| ML estimator |          |          |
|--------------|----------|----------|
|              | $n = 20$ | $n = 40$ |
| mean         | 0.076    | 0.033    |
| median       | 0.062    | 0.027    |
| mode         | 0.038    | 0.017    |

|        | Confidence interval |          |
|--------|---------------------|----------|
|        | $n = 20$            | $n = 40$ |
| mean   | 0.893               | 0.633    |
| median | 0.857               | 0.607    |
| mode   | 0.784               | 0.556    |



**Figure 7.** Application (World's Billionaires). Density of the random risk for maximum likelihood estimator (left panel) and for 95% confidence interval (right panel) for the Pareto index  $\theta$ . Curves obtained for  $n = 20$  (red) and for  $n = 40$  (black), given a prior variance  $16 \times 0.008$ .

## 5. Closing Remarks

In this paper, we study the distribution of the risk function for point and interval estimation for the Pareto model, when interest is on the shape parameter  $\theta$  (Pareto index) or on the scale parameter  $\eta$  or on a quantile that is a function of both  $\theta$  and  $\eta$ . Using conjugate priors, we obtain closed-form expressions for both the expected value and the density functions of the risk of each parameter under suitable losses. Interestingly, due to the analytical expressions of the risk function, in all the cases considered, the densities of the risk always belong to the same family as the corresponding design prior. This is a consequence of the closure of both Generalized Gamma and Inverse Pareto families with respect to scale and power transformations. Inspection of the shape of the density functions allows one to evaluate the impact of prior assumptions and sample size on the risk and to select an appropriate summary of the risk distribution. All these ideas are illustrated through a numerical example related to income data [7] and an application based on Forbes World's Billionaires list. Future developments of this work may be devoted to sample size determination in the spirit of [14].

## Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

## References

- [1] Lehman, E.L. and Casella, G. (1998) Theory of Point Estimation. 2nd Edition, Springer-Verlag.
- [2] Spiegelhalter, D.J. and Freedman, L.S. (1986) A Predictive Approach to Selecting the Size of a Clinical Trial, Based on Subjective Clinical Opinion. *Statistics in Medicine*, **5**, 1-13. <https://doi.org/10.1002/sim.4780050103>
- [3] Rufibach, K., Burger, H.U. and Abt, M. (2016) Bayesian Predictive Power: Choice of Prior and Some Recommendations for Its Use as Probability of Success in Drug Development. *Pharmaceutical Statistics*, **15**, 438-446. <https://doi.org/10.1002/pst.1764>
- [4] Liu, C.C. and Yu, R.X. (2023) Epistemic Uncertainty in Bayesian Predictive Probabilities. *Journal of Biopharmaceutical Statistics*, **34**, 394-412. <https://doi.org/10.1080/10543406.2023.2204943>
- [5] De Santis, F. and Gubbiotti, S. (2024) On the Distribution of the Power Function for the Scale Parameter of Exponential Families. *Statistics in Medicine*, **43**, 1973-1992. <https://doi.org/10.1002/sim.10043>
- [6] Kleiber, C. and Kotz, S. (2003) Statistical Size Distributions in Economics and Actuarial Sciences. Wiley. <https://doi.org/10.1002/0471457175>
- [7] Arnold, B.C. and Press, S.J. (1989) Bayesian Estimation and Prediction for Pareto Data. *Journal of the American Statistical Association*, **84**, 1079-1084. <https://doi.org/10.1080/01621459.1989.10478875>
- [8] Dyer, D. (1981) Structural Probability Bounds for the Strong Pareto Law. *Canadian Journal of Statistics*, **9**, 71-77. <https://doi.org/10.2307/3315297>
- [9] Malik, H.J. (1970) Estimation of the Parameters of Thepareto Distribution. *Metrika*, **15**, 126-132. <https://doi.org/10.1007/bf02613565>
- [10] Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) Continuous Univariate Distributions, 2nd Edition. Wiley.
- [11] Mead, M.E. (2013) Generalized Inverse Gamma Distribution and Its Application in Reliability. *Communications in Statistics—Theory and Methods*, **44**, 1426-1435. <https://doi.org/10.1080/03610926.2013.768667>
- [12] McNeil, A., Frey, R., Embrechts, P. (2005) Quantitative Risk Management: Concepts Techniques and Tools. Princeton University Press.
- [13] Pinsky, E., Zhang, W. and Wang, Z. (2024) Pareto Distribution of the Forbes Billionaires. *Computational Economics*. <https://doi.org/10.1007/s10614-024-10730-1>
- [14] De Santis, F., Gubbiotti, S. and Mariani, F. (2024) On the Bayes Risk Induced by Alternative Design Priors for Sample Size Choice. *Optimization in Green Sustainability and Ecological Transition*, Ischia, 4-7 September 2023, 189-197. [https://doi.org/10.1007/978-3-031-47686-0\\_17](https://doi.org/10.1007/978-3-031-47686-0_17)