ELSEVIER

Statistics & Probability Letters 59 (2002) 159–167

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

Fisher information matrix for the Feller–Pareto distribution

Vytaras Brazauskas *

Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201 USA

Received December 2001

Abstract

In this paper, the exact form of Fisher information matrix for the Feller–Pareto (FP) distribution is determined. The FP family is a very general unimodal distribution which includes a variety of distributions as special cases. For example:

- A hierarchy of Pareto models: Pareto (I), Pareto (II), Pareto (III), and Pareto (IV) (see Arnold (Pareto Distributions, International Cooperative Publishing House, Fairland, MD, 1983)); and
- Transformed beta family which in turn includes such general families as Burr, Generalized Pareto, and Inverse Burr (see Klugman et al. (Loss Models: From Data to Decisions, Wiley, New York, 1998)).

Application of these distributions covers a wide spectrum of areas ranging from actuarial science, economics, finance to biosciences, telecommunications, and extreme value theory.

Keywords: Digamma function; Fisher information; Pareto models; Transformed beta family; Trigamma function

1. Introduction

In this paper, the exact form of Fisher information matrix for the Feller–Pareto (FP) distribution is determined. It is well known that Fisher information matrix serves as a valuable tool for derivation of covariance matrix in the asymptotic distribution of maximum likelihood estimators (MLE). Further, under suitable regularity conditions, the determinant (divided by the sample size) of the asymptotic covariance matrix of MLE reaches an optimal lower bound for the volume of the "spread ellipsoid" of joint estimators (see Serfling (1980, Section 4.1)). In the univariate case, this optimality property of MLE is widely used in the "robustness versus efficiency" studies as a quantitative benchmark

* Tel.: +1-414-229-5656; fax: +1-414-229-4907.
E-mail address: vytaras@uwm.edu (V. Brazauskas).

0167-7152/02/$ - see front matter
PII: S0167-7152(02)00143-8

for efficiency considerations. See, for example, Brazauskas and Serfling (2000a, b), Hampel et al. (1986), Huber (1981), Kimber (1983a, b), and Lehmann (1983, Chapter 5).

The FP family is a very general unimodal distribution which includes a variety of distributions as special cases. As shown in Section 2, it includes:

- A hierarchy of Pareto models which is constituted of Pareto (I), Pareto (II), Pareto (III), and Pareto (IV) distributions; and
- Transformed beta family which in turn includes such general families as Burr, Generalized Pareto, and Inverse Burr.

A key feature of these distributions is a relatively high probability in the upper tail. However, it is also interesting to note that there are some distributions that exhibit distinctly non-Paretian behavior in the upper tail. For instance, Loglogistic, Inverse Pareto, and Inverse Paralogistic—each is a special case of Inverse Burr—have relatively "light" tails (see Section 2.2 below).

Application of such models covers a wide spectrum of areas ranging from actuarial science, economics, finance to medicine and telecommunications, for distributions of variables such as sizes of insurance claims, incomes in a population of people, stock price fluctuations, duration of responses to medical treatment, and length of telephone calls. (See Arnold, 1983; Johnson et al., 1994; Klein and Moeschberger, 1997.) Moreover, some of these distributions are relevant within much broader classes of models. For example, a Generalized Pareto distribution arises in semiparametric modeling of upper observations in samples from distributions which are regularly varying or in the domain of attraction of extreme value distributions (Embrechts et al., 1997).

The paper is organized as follows. In Section 2, we describe two different representations of the FP family and specify some general distributions which the FP family includes as special cases. In Section 3, we provide elements of the Fisher information matrix for FP, Pareto (IV), Inverse Burr, and Generalized Pareto distributions. Intermediate technical results (integrals) are presented in the Appendix.

2. Feller–Pareto and related distributions

The FP family traces its roots back to Feller (1971), but in the form we consider here it was first defined and investigated by Arnold and Laguna (1977) (see also Arnold (1983, Section 3.2)).

Let random variable X_1 have a Beta distribution with parameters $\gamma_1 > 0$ and $\gamma_2 > 0$, then $X_2 = X_1^{-1} - 1$ has (according to Feller (1971, p. 50)) a Pareto distribution. Next, if for $-\infty < \mu < +\infty$, $\sigma > 0$, and $\gamma > 0$ we define $X = \mu + \sigma X_2^{\gamma}$, then X has a Feller–Pareto distribution with the density function

$$f(x) = \frac{\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1)\Gamma(\gamma_2)} \frac{\left(\frac{x-\mu}{\sigma}\right)^{(\gamma_2/\gamma)-1}}{\gamma\sigma\left[1+\left(\frac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{\gamma_1+\gamma_2}}, \quad x > \mu, \tag{1}$$

where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,dt$ is the gamma function. We denote this distribution by FP$(\mu, \sigma, \gamma, \gamma_1, \gamma_2)$.

Alternatively, the FP distribution can also be represented via two independent gamma variables. More specifically, if Y_1 and Y_2 are independent gamma random variables with unit scale parameter and γ_1 and γ_2 respective shape parameters, then $X = \mu + \sigma(Y_2/Y_1)^\gamma$ has FP$(\mu, \sigma, \gamma, \gamma_1, \gamma_2)$ distribution.

2.1. A hierarchy of Pareto models

As discussed in Arnold (1983, pp. 44–45), a hierarchy of Pareto models is established by starting with the classical Pareto distribution, Pareto (I), and subsequently introducing additional parameters which relate to *location*, *scale*, *shape*, and *inequality*. Such an approach leads to the Pareto (IV) family with the density function

$$g_1(x) = \frac{\alpha}{\gamma\sigma} \frac{\left(\dfrac{x-\mu}{\sigma}\right)^{1/\gamma-1}}{\left[1+\left(\dfrac{x-\mu}{\sigma}\right)^{1/\gamma}\right]^{\alpha+1}}, \quad x > \mu,$$

where $-\infty < \mu < +\infty$ is the location parameter, $\sigma > 0$ is the scale parameter, $\gamma > 0$ is the inequality parameter, and $\alpha > 0$ is the shape parameter which characterizes the tail of the distribution. This is a very general family of distributions which itself includes Pareto (I), Pareto (II), Pareto (III), and the Burr distributions. Nevertheless, Pareto (IV) and, consequently, the other related distributions can be identified as special cases of the FP family by appropriately choosing parameters in (1):

Pareto (I) $(\sigma, \alpha) = \text{FP}\ (\sigma, \sigma, 1, \alpha, 1)$,

Pareto (II) $(\mu, \sigma, \alpha) = \text{FP}\ (\mu, \sigma, 1, \alpha, 1)$,

Pareto (III) $(\mu, \sigma, \gamma) = \text{FP}\ (\mu, \sigma, \gamma, 1, 1)$,

Pareto (IV) $(\mu, \sigma, \gamma, \alpha) = \text{FP}\ (\mu, \sigma, \gamma, \alpha, 1)$.

2.2. Transformed beta family

Another special case of the FP distribution is the transformed beta family. It can be found in Klugman et al. (1998, p. 573), that the density function of the transformed beta distribution is given by

$$g_2(x) = \frac{\Gamma(\alpha+\tau)}{\Gamma(\alpha)\Gamma(\tau)} \frac{\gamma(x/\theta)^{\gamma\tau}}{x[1+(x/\theta)^\gamma]^{\alpha+\tau}}, \quad x > 0.$$

Hence, it follows from (1) and the definition of $g_2(\cdot)$ that

Transformed Beta $(\theta, \gamma, \alpha, \tau) = \text{FP}\ (0, \theta, 1/\gamma, \alpha, \tau)$.

This family itself includes a variety of distributions thus offering great flexibility in modeling. Moreover, as it is seen from a below provided summary, those more specialized families are still quite general and include other important distributions themselves. Specifically,

- Burr (θ,γ,α) = Transformed Beta $(\theta,\gamma,\alpha,1)$:

  Loglogistic (θ,γ) = Burr $(\theta,\gamma,1)$,

  Paralogistic (θ,α) = Burr (θ,α,α),

  Pareto (II) $(0,\theta,\alpha)$ = Burr $(\theta,1,\alpha)$,

- Generalized Pareto (θ,α,τ) = Transformed Beta $(\theta,1,\alpha,\tau)$:

  Pareto (II) $(0,\theta,\alpha)$ = Generalized Pareto $(\theta,\alpha,1)$,

  Inverse Pareto (θ,τ) = Generalized Pareto $(\theta,1,\tau)$,

- Inverse Burr (θ,γ,τ) = Transformed Beta $(\theta,\gamma,1,\tau)$:

  Loglogistic (θ,γ) = Inverse Burr $(\theta,\gamma,1)$,

  Inverse Pareto (θ,τ) = Inverse Burr $(\theta,1,\tau)$,

  Inverse Paralogistic (θ,γ) = Inverse Burr $(\theta,\gamma,1)$.

For further illustrations and discussion see Klugman et al. (1998, Section 2.7).

In order to make the FP distribution a regular family (in terms of maximum likelihood estimation), we assume that parameter μ is known and, without loss of generality, equal to 0. Thus, further treatment is based on the density function

$$f_0(x)=\frac{\Gamma(\gamma_1+\gamma_2)}{\Gamma(\gamma_1)\Gamma(\gamma_2)}\,\frac{(x/\sigma)^{(\gamma_2/\gamma)-1}}{\gamma\sigma[1+(x/\sigma)^{1/\gamma}]^{\gamma_1+\gamma_2}},\quad x>0. \tag{2}$$

As the above-presented list of special cases suggests, the assumption of μ known is not too restrictive for modeling purposes. Also, in typical applications the lower limit of variables of interest is known. For instance, in insurance and reinsurance context, the lower limit of *severity* of claims is pre-defined by a contract and can be represented as a deductible or a retention level (see Daykin et al., 1994).

3. Information matrix for Feller–Pareto

Suppose X is a random variable with the probability density function $f_\Theta(\cdot)$ where $\Theta=(\theta_1,\ldots,\theta_k)$. Then the information matrix $\mathbf{I}(\Theta)$ is the $k\times k$ symmetric matrix with elements

$$I_{ij}(\Theta)=\mathbf{E}_\Theta\left[\frac{\partial\log f_\Theta(X)}{\partial\theta_i}\,\frac{\partial\log f_\Theta(X)}{\partial\theta_j}\right].$$

If the density $f_\Theta(\cdot)$ has second derivatives $\partial^2 f_\Theta(x)/\partial\theta_i\,\partial\theta_j$ for all i and j, then there is an alternative expression for $I_{ij}(\Theta)$, namely,

$$I_{ij}(\Theta)=-\mathbf{E}_\Theta\left[\frac{\partial^2\log f_\Theta(X)}{\partial\theta_i\,\partial\theta_j}\right]. \tag{3}$$

For the FP $(0,\sigma,\gamma,\gamma_1,\gamma_2)$ distribution all second derivatives exist, therefore formula (3) is appropriate and, most importantly, in our case it significantly simplifies computations. Thus, we have

$\Theta = (\theta_1, \theta_2, \theta_3, \theta_4) = (\sigma, \gamma, \gamma_1, \gamma_2)$, the log-density

$$\log f_0(x) = (\gamma_2/\gamma - 1)\log(x/\sigma) - (\gamma_1 + \gamma_2)\log[1 + (x/\sigma)^{1/\gamma}] - \log(\gamma) - \log(\sigma) + \log\Gamma(\gamma_1 + \gamma_2) - \log\Gamma(\gamma_1) - \log\Gamma(\gamma_2)$$

and the required second partial derivatives

$$\frac{\partial^2 \log f_0(x)}{\partial\sigma^2} = -\frac{\gamma_1}{\gamma\sigma^2} + \frac{\gamma_1 + \gamma_2}{\gamma\sigma^2}\left[\frac{1 - 1/\gamma}{1 + (x/\sigma)^{1/\gamma}} + \frac{1/\gamma}{(1 + (x/\sigma)^{1/\gamma})^2}\right],$$

$$\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma} = -\frac{\gamma_1}{\gamma^2\sigma} + \frac{\gamma_1 + \gamma_2}{\gamma^2\sigma}\left[\frac{1}{1 + (x/\sigma)^{1/\gamma}} - \frac{(x/\sigma)^{1/\gamma}\log(x/\sigma)^{1/\gamma}}{(1 + (x/\sigma)^{1/\gamma})^2}\right],$$

$$\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma_1} = \frac{1}{\gamma\sigma} - \frac{1}{\gamma\sigma}\frac{1}{1 + (x/\sigma)^{1/\gamma}},$$

$$\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma_2} = -\frac{1}{\gamma\sigma}\frac{1}{1 + (x/\sigma)^{1/\gamma}},$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma^2} = \frac{1}{\gamma^2} - \frac{2\gamma_1}{\gamma^2}\log(x/\sigma)^{1/\gamma} + \frac{2(\gamma_1 + \gamma_2)}{\gamma^2}\frac{\log(x/\sigma)^{1/\gamma}}{1 + (x/\sigma)^{1/\gamma}} - \frac{\gamma_1 + \gamma_2}{\gamma^2}\frac{(x/\sigma)^{1/\gamma}\log^2(x/\sigma)^{1/\gamma}}{(1 + (x/\sigma)^{1/\gamma})^2},$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma\,\partial\gamma_1} = \frac{1}{\gamma}\log(x/\sigma)^{1/\gamma} - \frac{1}{\gamma}\frac{\log(x/\sigma)^{1/\gamma}}{1 + (x/\sigma)^{1/\gamma}},$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma\,\partial\gamma_2} = -\frac{1}{\gamma}\frac{\log(x/\sigma)^{1/\gamma}}{1 + (x/\sigma)^{1/\gamma}},$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma_1^2} = \psi'(\gamma_1 + \gamma_2) - \psi'(\gamma_1),$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma_1\,\partial\gamma_2} = \psi'(\gamma_1 + \gamma_2),$$

$$\frac{\partial^2 \log f_0(x)}{\partial\gamma_2^2} = \psi'(\gamma_1 + \gamma_2) - \psi'(\gamma_2),$$

where $\psi(a) = \Gamma'(a)/\Gamma(a)$ and $\psi'(a) = \mathrm{d}\psi(a)/\mathrm{d}a$ are, respectively, the digamma and trigamma function.

Computation of elements $I_{33}(\Theta)$, $I_{34}(\Theta)$, and $I_{44}(\Theta)$ is trivial because the corresponding second derivatives are constants. Thus,

$$I_{33}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma_1^2}\right] f_0(x)\,\mathrm{d}x = \psi'(\gamma_1) - \psi'(\gamma_1 + \gamma_2),$$

$$I_{34}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma_1\,\partial\gamma_2}\right] f_0(x)\,\mathrm{d}x = -\psi'(\gamma_1 + \gamma_2),$$

$$I_{44}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma_2^2}\right] f_0(x)\,\mathrm{d}x = \psi'(\gamma_2) - \psi'(\gamma_1 + \gamma_2).$$

Derivation of the remaining elements is based on the following strategy: first, we express each $I_{ij}(\Theta)$ in terms of integrals $A1$–$A6$ which are defined (and their explicit formulas are presented) in the Appendix; then, straightforward (but in some cases tedious) algebraic simplifications yield the following formulas:

$$I_{11}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\sigma^2}\right] f_0(x)\,\mathrm{d}x$$
$$= \frac{\gamma_1}{\gamma\sigma^2} - \frac{\gamma_1+\gamma_2}{\gamma\sigma^2}[(1-1/\gamma)A1 + (1/\gamma)A2] = \frac{\gamma_1\gamma_2}{\gamma^2\sigma^2(\gamma_1+\gamma_2+1)},$$

$$I_{12}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma}\right] f_0(x)\,\mathrm{d}x$$
$$= \frac{\gamma_1}{\gamma^2\sigma} - \frac{\gamma_1+\gamma_2}{\gamma^2\sigma}[A1-A5] = \frac{\gamma_1\gamma_2[\psi(\gamma_2)-\psi(\gamma_1)]+\gamma_1-\gamma_2}{\gamma^2\sigma(\gamma_1+\gamma_2+1)},$$

$$I_{13}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma_1}\right] f_0(x)\,\mathrm{d}x = -\frac{1}{\gamma\sigma}[1-A1] = \frac{-\gamma_2}{\gamma\sigma(\gamma_1+\gamma_2)},$$

$$I_{14}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\sigma\,\partial\gamma_2}\right] f_0(x)\,\mathrm{d}x = \frac{1}{\gamma\sigma}A1 = \frac{\gamma_1}{\gamma\sigma(\gamma_1+\gamma_2)},$$

$$I_{22}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma^2}\right] f_0(x)\,\mathrm{d}x$$
$$= -\frac{1}{\gamma^2} + \frac{2\gamma_1}{\gamma^2}A3 - \frac{2(\gamma_1+\gamma_2)}{\gamma^2}A4 + \frac{\gamma_1+\gamma_2}{\gamma^2}A6 = \frac{1}{\gamma^2} + \frac{\gamma_1\gamma_2}{\gamma^2(\gamma_1+\gamma_2+1)}$$
$$\times\left[\psi'(\gamma_1)+\psi'(\gamma_2)+\left(\psi(\gamma_1)-\psi(\gamma_2)+\frac{\gamma_2-\gamma_1}{\gamma_1\gamma_2}\right)^2 - \frac{\gamma_1^2+\gamma_2^2}{\gamma_1^2\gamma_2^2}\right],$$

$$I_{23}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma\,\partial\gamma_1}\right] f_0(x)\,\mathrm{d}x = -\frac{1}{\gamma}[A3-A4] = \frac{\gamma_2[\psi(\gamma_1)-\psi(\gamma_2)]-1}{\gamma(\gamma_1+\gamma_2)},$$

$$I_{24}(\Theta) = -\int_0^\infty \left[\frac{\partial^2 \log f_0(x)}{\partial\gamma\,\partial\gamma_2}\right] f_0(x)\,\mathrm{d}x = \frac{1}{\gamma}A4 = \frac{\gamma_1[\psi(\gamma_2)-\psi(\gamma_1)]-1}{\gamma(\gamma_1+\gamma_2)}.$$

3.1. Special cases

In this section, we provide several examples illustrating how to obtain Fisher information matrix for distributions which are special cases of $\mathrm{FP}(0,\sigma,\gamma,\gamma_1,\gamma_2)$. Other choices of more specialized families can be treated in a similar manner.

3.1.1. Pareto (IV) $(0,\sigma,\gamma,\alpha)$ distribution

This is a special case of $\mathrm{FP}(0,\sigma,\gamma,\gamma_1,\gamma_2)$ with $\gamma_1=\alpha$ and $\gamma_2=1$. Therefore, elements that represent information about parameter γ_2 (i.e., $I_{4j}(\Theta)$ and $I_{j4}(\Theta), 1\leqslant j\leqslant 4$) vanish. And into the expressions of the remaining elements we substitute $\gamma_2=1$ and $\gamma_1=\alpha$. This yields

$$\begin{pmatrix} \frac{\alpha}{(\gamma\sigma)^2(\alpha+2)} & \frac{\alpha[\psi(1)-\psi(\alpha)+1]-1}{\gamma^2\sigma(\alpha+2)} & -\frac{1}{\gamma\sigma(\alpha+1)} \\ \frac{\alpha[\psi(1)-\psi(\alpha)+1]-1}{\gamma^2\sigma(\alpha+2)} & \frac{\alpha[(\psi(\alpha)-\psi(1)-1)^2+\psi'(\alpha)+\psi'(1)]+2(\psi(\alpha)-\psi(1))}{\gamma^2(\alpha+2)} & \frac{\psi(\alpha)-\psi(1)-1}{\gamma(\alpha+1)} \\ -\frac{1}{\gamma\sigma(\alpha+1)} & \frac{\psi(\alpha)-\psi(1)-1}{\gamma(\alpha+1)} & \frac{1}{\alpha^2} \end{pmatrix}.$$

Remark. We note here that information matrix for Pareto (IV) is readily derived in Brazauskas (2001). It is quite surprising that computations presented there are much messier and based on more complicated integration results than in the case of more general FP family.

3.1.2. Inverse Burr (θ,γ,τ) distribution

Information matrix for the Inverse Burr distribution is derived via a two-step procedure. First step, we find information matrix for the transformed beta family. Second step, we use the fact that Inverse Burr is a special case of the transformed beta family.

Since the transformed beta family is a reparametrization of $\mathrm{FP}(0,\sigma,\gamma,\gamma_1,\gamma_2)$ it follows from Lehmann (1983, Section 2.7) that its information matrix can be derived from formula $J\mathbf{I}_{\mathrm{FP}}(\Theta)J'$, where J is the Jacobian matrix of the transformation of variables, and $\mathbf{I}_{\mathrm{FP}}(\Theta)$ is the FP information matrix. Next, Inverse Burr (θ,γ,τ) is a special case of the transformed beta distribution with $\alpha=1$, therefore elements that represent information about parameter α vanish, and into expressions of the remaining elements we substitute $\alpha=1$. This leads to

$$\begin{pmatrix} \frac{\tau}{(\gamma\theta)^2(\tau+2)} & \frac{\tau[\psi(1)-\psi(\tau)+1]-1}{\theta(\tau+2)} & \frac{1}{\gamma\theta(\tau+1)} \\ \frac{\tau[\psi(1)-\psi(\tau)+1]-1}{\theta(\tau+2)} & \frac{\gamma^2[\tau[(\psi(\tau)-\psi(1)-1)^2+\psi'(\tau)+\psi'(1)]+2(\psi(\tau)-\psi(1))]}{\tau+2} & \frac{\gamma[\psi(\tau)-\psi(1)-1]}{\tau+1} \\ \frac{1}{\gamma\theta(\tau+1)} & \frac{\gamma[\psi(\tau)-\psi(1)-1]}{\tau+1} & \frac{1}{\tau^2} \end{pmatrix}.$$

3.1.3. Generalized Pareto (θ,α,τ) distribution

Information matrix for Generalized Pareto (θ,α,τ) is derived by following same steps as in the case of Inverse Burr (θ,γ,τ). This yields

$$\begin{pmatrix} \frac{\alpha\tau}{\theta^2(\alpha+\tau+1)} & -\frac{\tau}{\theta(\alpha+\tau)} & \frac{\alpha}{\theta(\alpha+\tau)} \\ -\frac{\tau}{\theta(\alpha+\tau)} & \psi'(\alpha)-\psi'(\alpha+\tau) & -\psi'(\alpha+\tau) \\ \frac{\alpha}{\theta(\alpha+\tau)} & -\psi'(\alpha+\tau) & \psi'(\tau)-\psi'(\alpha+\tau) \end{pmatrix}.$$

Appendix

In all expressions below function $f_0(x)$ is the FP $(0, \alpha, \gamma, \gamma_1, \gamma_2)$ density function given by (2). Integrals $A1$ and $A2$ are derived via straightforward integration.

$$A1 = \int_0^\infty \frac{f_0(x)}{1+(x/\sigma)^{1/\gamma}}\,\mathrm{d}x = \frac{\gamma_1}{\gamma_1+\gamma_2},$$

$$A2 = \int_0^\infty \frac{f_0(x)}{[1+(x/\sigma)^{1/\gamma}]^2}\,\mathrm{d}x = \frac{\gamma_1(\gamma_1+1)}{(\gamma_1+\gamma_2)(\gamma_1+\gamma_2+1)}.$$

Initial simplifications of integrals $A3$–$A6$ lead to the following types of integrals: $\int_0^1 t^{a-1}(1-t)^{b-1}\log^k(t)\,\mathrm{d}t$, $\int_0^1 t^{a-1}(1-t)^{b-1}\log^k(1-t)\,\mathrm{d}t$, for $k=1$ or 2, and $\int_0^1 t^{a-1}(1-t)^{b-1}\log(t)\log(1-t)\,\mathrm{d}t$. It is easily seen that these are either first- or second-order partial derivatives of the beta function $B(a,b)$. Thus, differentiation of $B(a,b)$ and further algebraic simplifications yield

$$A3 = \int_0^\infty \log(x/\sigma)^{1/\gamma} f_0(x)\,\mathrm{d}x = \psi(\gamma_2) - \psi(\gamma_1),$$

$$A4 = \int_0^\infty \frac{\log(x/\sigma)^{1/\gamma}}{1+(x/\sigma)^{1/\gamma}} f_0(x)\,\mathrm{d}x = \frac{\gamma_1}{\gamma_1+\gamma_2}[\psi(\gamma_2) - \psi(\gamma_1) - 1/\gamma_1],$$

$$A5 = \int_0^\infty \frac{(x/\sigma)^{1/\gamma}\log(x/\sigma)^{1/\gamma}}{[1+(x/\sigma)^{1/\gamma}]^2} f_0(x)\,\mathrm{d}x$$

$$= \frac{\gamma_1\gamma_2}{(\gamma_1+\gamma_2)(\gamma_1+\gamma_2+1)}[\psi(\gamma_1) - \psi(\gamma_2) + 1/\gamma_2 - 1/\gamma_1],$$

$$A6 = \int_0^\infty \frac{(x/\sigma)^{1/\gamma}\log^2(x/\sigma)^{1/\gamma}}{[1+(x/\sigma)^{1/\gamma}]^2} f_0(x)\,\mathrm{d}x = \frac{\gamma_1\gamma_2}{(\gamma_1+\gamma_2)(\gamma_1+\gamma_2+1)}$$

$$\times\left[\psi'(\gamma_1) + \psi'(\gamma_2) + \left(\psi(\gamma_1) - \psi(\gamma_2) + \frac{\gamma_2-\gamma_1}{\gamma_1\gamma_2}\right)^2 - \frac{\gamma_1^2+\gamma_2^2}{\gamma_1^2\gamma_2^2}\right].$$

References

Arnold, B.C., 1983. Pareto Distributions. International Cooperative Publishing House, Fairland, MD.

Arnold, B.C., Laguna, L., 1977. On Generalized Pareto Distributions with Applications to Income Data. International Studies in Economics, Monograph #10, Department of Economics, Iowa State University, Ames, IA.

Brazauskas, V., 2001. Information matrix for Pareto (IV), Burr, and related distributions. Preprint.

Brazauskas, V., Serfling, R., 2000a. Robust estimation of tail parameters for two-parameter Pareto and exponential models via generalized quantile statistics. Extremes 3, 231–249.

Brazauskas, V., Serfling, R., 2000b. Robust and efficient estimation of the tail index of a single-parameter Pareto distribution. North American Actuarial Journal 4, 12–27.

Daykin, C.D., Pentikäinen, T., Pesonen, M., 1994. Practical Risk Theory for Actuaries. Chapman & Hall, London.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer, Berlin.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications, Vol. 2, 2nd Edition. Wiley, New York.
Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics: The Approach Based on Influence Functions. Wiley, New York.
Huber, P.J., 1981. Robust Statistics. Wiley, New York.
Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. 1, 2nd Edition. Wiley, New York.
Kimber, A.C., 1983a. Trimming in gamma samples. Applied Statistics 32, 7–14.
Kimber, A.C., 1983b. Comparison of some robust estimators of scale in gamma samples with known shape. Journal of Statistical Computation and Simulation 18, 273–286.
Klein, J.P., Moeschberger, M.L., 1997. Survival Analysis: Techniques for Censored and Truncated Data. Springer, New York.
Klugman, S.A., Panjer, H.H., Willmot, G.E., 1998. Loss Models: From Data to Decisions. Wiley, New York.
Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York. (transferred to Wadsworth, 1991)
Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.