1 Location-scale estimation

Let Y be a real valued stochastic variable with at least fourth moment, i.e. $\mathbb{E}Y^4 < \infty$. Let $\Theta = \mathbb{R} \times (0, \infty)$ and introduce for $\theta = (\xi, \psi) \in \Theta$ the location-scale transformed stochastic variable

$$X = \psi Y + \xi. \tag{1}$$

If $X_1, \ldots, X_n$ are iid stochastic variables all with the same distribution as X, we will study the estimation of the parameters ξ and ψ from $X_1, \ldots, X_n$. In particular we will study the asymptotic distributional behaviour of the estimator based on the empirical mean and variance.

We introduce some notation. Define

$$\begin{aligned} \mu_1 &= \mathbb{E}Y \\ \sigma_0^2 &= \mathbb{V}Y = \mathbb{E}(Y - \mu_1)^2 \\ \mu_{0,3} &= \mathbb{E}(Y - \mu_1)^3 \\ \mu_{0,4} &= \mathbb{E}(Y - \mu_1)^4, \end{aligned}$$

then

$$\begin{aligned} \mu &= \mathbb{E}X = \xi + \psi\mu_1 \\ \sigma^2 &= \mathbb{V}X = \psi^2\sigma_0^2 \\ \mu_3 &= \mathbb{E}(X - \mu_1)^3 = \psi^3\mu_{0,3} \\ \mu_4 &= \mathbb{E}(X - \mu_1)^4 = \psi^4\mu_{0,4}. \end{aligned}$$

The unbiased empirical estimators of μ and σ^2 are

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2}$$

$$\hat{\sigma^2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \hat{\mu})^2. \tag{3}$$

We can find the variance and covariance of these estimators

$$\begin{aligned} \mathbb{V}\hat{\mu} &= \frac{1}{n}\sigma^2 \\ \mathbb{V}\hat{\sigma^2} &= \frac{1}{n}(\mu_4 - \sigma^4) + \frac{2\sigma^4}{n(n-1)} = \frac{1}{n}(\mu_4 - \sigma^4) + o(n^{-1}) \\ \operatorname{cov}(\hat{\mu}, \hat{\sigma^2}) &= \frac{1}{n}\mu_3. \end{aligned}$$

With

$$\Sigma_1 = \left\{ \begin{array}{cc} \sigma^2 & \mu_3 \\ \mu_3 & \mu_4 - \sigma^4 \end{array} \right\} = \psi^2 \left\{ \begin{array}{cc} \sigma_0^2 & \psi\mu_{0,3} \\ \psi\mu_{0,3} & \psi^2(\mu_{0,4} - \sigma_0^4) \end{array} \right\}$$

it holds that asymptotically as $n \to \infty$

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma^2} \end{pmatrix} \overset{\text{as}}{\sim} N\left(\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \frac{1}{n}\Sigma_1 \right)$$

If we introduce the function $f : \mathbb{R} \times (0, \infty) \to \mathbb{R} \times (0, \infty)$ given by

$$f\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x - \mu_1\sigma_0^{-1}\sqrt{y} \\ \sigma_0^{-1}\sqrt{y} \end{pmatrix} \tag{4}$$

then $f(\mu, \sigma^2) = (\xi, \psi)$ so a natural estimator for the location-scale parameter $\theta = (\xi, \psi)$ is

$$\begin{pmatrix} \tilde{\xi} \\ \tilde{\psi} \end{pmatrix} = f\begin{pmatrix} \hat{\mu} \\ \hat{\sigma^2} \end{pmatrix}. \tag{5}$$

To find the asymptotic normal distribution of this estimator we need to compute the derivative of f:

$$Df\begin{pmatrix} x \\ y \end{pmatrix} = \left\{ \begin{array}{cc} 1 & -\frac{\mu_1}{2\sigma_0\sqrt{y}} \\ 0 & \frac{1}{2\sigma_0\sqrt{y}} \end{array} \right\}. \tag{6}$$

Observe that if Λ denotes the derivative evaluated in (μ, σ^2) then

$$\Lambda = Df\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix} = \left\{ \begin{array}{cc} 1 & -\frac{\mu_1}{2\sigma_0\sigma} \\ 0 & \frac{1}{2\sigma_0\sigma} \end{array} \right\} = \left\{ \begin{array}{cc} 1 & -\frac{\mu_1}{2\psi\sigma_0^2} \\ 0 & \frac{1}{2\psi\sigma_0^2} \end{array} \right\}. \tag{7}$$

It follows by the delta-method that

$$\begin{pmatrix} \tilde{\xi} \\ \tilde{\psi} \end{pmatrix} \overset{\text{as}}{\sim} N\left(f\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \frac{1}{n}\Lambda\Sigma_1\Lambda^T \right) = N\left(\begin{pmatrix} \xi \\ \psi \end{pmatrix}, \frac{1}{n}\Sigma \right)$$

where

$$\begin{aligned} \Sigma &= \Lambda\Sigma_1\Lambda^T \\ &= \psi^2 \left\{ \begin{array}{cc} 1 & -\frac{\mu_1}{2\psi\sigma_0^2} \\ 0 & \frac{1}{2\psi\sigma_0^2} \end{array} \right\} \left\{ \begin{array}{cc} \sigma_0^2 & \psi\mu_{0,3} \\ \psi\mu_{0,3} & \psi^2(\mu_{0,4} - \sigma_0^4) \end{array} \right\} \left\{ \begin{array}{cc} 1 & 0 \\ -\frac{\mu_1}{2\psi\sigma_0^2} & \frac{1}{2\psi\sigma_0^2} \end{array} \right\} \\ &= \psi^2 \left\{ \begin{array}{cc} \sigma_0^2 - \frac{\mu_1^2\mu_{0,3}}{\sigma_0^2} + \frac{\mu_1(\mu_{0,4}-\sigma_0^4)}{4\sigma_0^4} & \frac{\mu_{0,3}}{2\sigma_0^2} - \frac{\mu_1(\mu_{0,4}-\sigma_0^4)}{4\sigma_0^4} \\ \frac{\mu_{0,3}}{2\sigma_0^2} - \frac{\mu_1(\mu_{0,4}-\sigma_0^4)}{4\sigma_0^4} & \frac{\mu_{0,4}-\sigma_0^4}{4\sigma_0^4} \end{array} \right\}. \end{aligned}$$

One should note that besides the factor ψ^2 this matrix does not depend on the parameters and can be computed from the knowledge of the first four moments of the distribution of Y. A particular simple case is when the distribution of Y is normalised so that $\mu_1 = 0$ and $\sigma_0^2 = 1$, since then

$$\Sigma = \psi^2 \left\{ \begin{array}{cc} 1 & \frac{\mu_{0,3}}{2} \\ \frac{\mu_{0,3}}{2} & \frac{\mu_{0,4}-1}{4} \end{array} \right\}$$

If Y is $N(0,1)$-distributed it holds in addition that $\mu_{0,3} = 0$ and $\mu_{0,4} = 3$ so

$$\Sigma = \psi^2 \left\{ \begin{array}{cc} 1 & 0 \\ 0 & \frac{1}{2} \end{array} \right\},$$

which tells us that the empirical mean and variance are asymptotically independent. In this case it actually holds that $\hat{\mu}$ and $\hat{\sigma^2}$ are independent, $\hat{\mu} \sim N(\xi, \psi^2)$ and $\hat{\sigma^2} \sim \psi^2(n-1)^{-1}\chi^2(n-1)$ where $\rho\chi^2(n-1)$ denotes a χ^2-distribution with $n-1$ degrees of freedom and scale parameter ρ.

2 The Gumbel distribution

The (standard) Gumbel distribution has distribution function

$$\Lambda(x) = \exp(-\exp(-x)) \tag{8}$$

and density

$$\lambda(x) = \exp(-x - \exp(-x)). \tag{9}$$

The raw moments can be found as follows:

$$\begin{aligned} \mu'_n &= \int_{-\infty}^{\infty} x^n \lambda(x) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} x^n \exp(-x - \exp(-x)) \mathrm{d}x \\ &= \int_0^{\infty} (-\log z)^n \exp(-z) \mathrm{d}z \\ &= (-1)^n \int_0^{\infty} (\log z)^n \exp(-z) \mathrm{d}z, \end{aligned}$$

where we use the substitution $z = \exp(-x)$ for the third equality. This sequence of numbers, in particular using the last representation, is known as the Euler-Mascheroni integrals. For $n = 0$ the integral necessarily equals 1 (λ is a density), but it is also easy to check directly that $\int_0^\infty \exp(-z)\mathrm{d}z = 1$. Furthermore,

$$\mu'_1 = -\int_0^{\infty} \log z \exp(-z) \mathrm{d}z = \gamma = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \log n \simeq 0.57721566$$

is the Euler (or Euler-Mascheroni) constant γ. For $n \geq 2$ explicit expressions for μ'_n seems to become more and more complicated, and at least for $n = 2, 3, 4$ the expressions for the central moments, which we need, are simpler:

$$\begin{aligned}
\sigma_0^2 &= \int_0^\infty (\log z + \gamma)^2 \exp(-z)\mathrm{d}z = \zeta(2) = \frac{\pi^2}{6} \\
\mu_{0,3} &= -\int_0^\infty (\log z + \gamma)^3 \exp(-z)\mathrm{d}z = 2\zeta(3) \\
\mu_{0,4} &= \int_0^\infty (\log z + \gamma)^3 \exp(-z)\mathrm{d}z = \frac{27\zeta(4)}{2} = \frac{3\pi^4}{20}.
\end{aligned}$$

Here

$$\zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n}$$

is the Riemann zeta function evaluated in n. For odd n, in particular $n = 3$, we have no completely explicit formula for $\zeta(n)$. The constant $\zeta(3)$, also known as Apéry's constant, can be approximated to e.g. eight digits by

$$\zeta(3) \simeq 1.20205690.$$

The asymptotic covariance matrix for the estimators $\tilde{\xi}$ and $\tilde{\psi}$ within the framework of the Gumbel distribution is

$$\Sigma = \psi^2 \left\{ \begin{array}{ll} 1.168 & 0.09583 \\ 0.09583 & 1.100 \end{array} \right\}.$$

2.1 Maximum likelihood estimation

Instead of using an estimator based on the empirical mean and variance we will consider the maximum likelihood estimator.

The density for $X = \psi Y + \xi$ is

$$\lambda_{\xi,\psi}(x) = \frac{1}{\psi}\lambda\left(\frac{x-\xi}{\psi}\right) = \frac{1}{\psi}\exp\left(-\frac{x-\xi}{\psi} - \exp\left(-\frac{x-\xi}{\psi}\right)\right).$$

With

$$l_x(\xi,\psi) = -\log\lambda_{\xi,\psi}(x) = \log\psi + \frac{x-\xi}{\psi} + \exp\left(-\frac{x-\xi}{\psi}\right) \tag{10}$$

we find that the derivatives of l_x are

$$\begin{aligned}
\frac{\partial l_x}{\partial \xi}(\xi,\psi) &= \frac{1}{\psi}\exp\left(-\frac{x-\xi}{\psi}\right) - \frac{1}{\psi} \\
\frac{\partial l_x}{\partial \psi}(\xi,\psi) &= \frac{1}{\psi} - \frac{x-\xi}{\psi^2} + \frac{x-\xi}{\psi^2}\exp\left(-\frac{x-\xi}{\psi}\right)
\end{aligned}$$

and the second derivatives

$$\begin{aligned}
\frac{\partial^2 l_x}{\partial \xi^2}(\xi,\psi) &= \frac{1}{\psi^2}\exp\left(-\frac{x-\xi}{\psi}\right)\\
\frac{\partial^2 l_x}{\partial \psi \partial \xi}(\xi,\psi) &= \frac{\partial^2 l_x}{\partial \xi \partial \psi}(\xi,\psi) = \frac{1}{\psi^2} - \frac{1}{\psi^2}\exp\left(-\frac{x-\xi}{\psi}\right) + \frac{x-\xi}{\psi^3}\exp\left(-\frac{x-\xi}{\psi}\right)\\
\frac{\partial^2 l_x}{\partial \psi^2}(\xi,\psi) &= -\frac{1}{\psi^2} + \frac{2(x-\xi)}{\psi^3} - \frac{2(x-\xi)}{\psi^3}\exp\left(-\frac{x-\xi}{\psi}\right) + \left(\frac{x-\xi}{\psi^2}\right)^2\exp\left(-\frac{x-\xi}{\psi}\right).
\end{aligned}$$

We can then define the following matrix of stochastic variables

$$D^2 l_X(\xi,\psi) = \left\{ \begin{array}{cc} \frac{\partial^2 l_X}{\partial \xi^2}(\xi,\psi) & \frac{\partial^2 l_X}{\partial \psi \partial \xi}(\xi,\psi) \\ \frac{\partial^2 l_X}{\partial \xi \partial \psi}(\xi,\psi) & \frac{\partial^2 l_X}{\partial \psi^2}(\xi,\psi) \end{array} \right\}.$$

The expectation of this matrix is the *expected information* – or just the information – and it is

$$i(\xi,\psi) = \mathbb{E}D^2 l_X(\xi,\psi) = \frac{1}{\psi^2}\left\{ \begin{array}{cc} 1 & \gamma-1 \\ \gamma-1 & 1+\pi^2/6+\gamma^2-2\gamma \end{array} \right\} \tag{11}$$

where γ is the Euler constant as defined above. One derives this by first observing that

$$\frac{X-\xi}{\psi} = Y$$

so that e.g.

$$\mathbb{E}\left(\frac{X-\xi}{\psi}\right) = \mathbb{E}Y = -\int_0^\infty \log z \exp(-z)\mathrm{d}z = \gamma$$

and

$$\begin{aligned}
\mathbb{E}\exp\left(-\frac{X-\xi}{\psi}\right) &= \mathbb{E}\exp(-Y)\\
&= \int_{-\infty}^{\infty} \exp(-y)\lambda(y)\mathrm{d}y\\
&= \int_{-\infty}^{\infty} \exp(-2y-\exp(-y))\mathrm{d}y\\
&= \int_0^\infty z\exp(-z)\mathrm{d}z = 1
\end{aligned}$$

where the fourth equality is based on the substitution $z = \exp(-y)$ and the integral can be computed either by integration by parts or simply by observing that this is the mean of a standard exponentially distributed variable. By similar methods we can also derive that

$$\begin{aligned}
\mathbb{E}\left(\frac{X-\xi}{\psi}\right)\exp\left(\frac{X-\xi}{\psi}\right) &= \mathbb{E}Y\exp(-Y) \\
&= -\int_0^\infty z\log z\exp(-z)\mathrm{d}z \\
&= \gamma - 1
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(\frac{X-\xi}{\psi}\right)^2\exp\left(\frac{X-\xi}{\psi}\right) &= \mathbb{E}Y^2\exp(-Y) \\
&= \int_0^\infty z(\log z)^2\exp(-z)\mathrm{d}z \\
&= \gamma^2 + \frac{\pi^2}{6} - 2\gamma.
\end{aligned}$$

For the computation of the two resulting integrals we use, once again, integration by parts, and for the last case we use that $\int_0^\infty (\log z)^2 \exp(-z)\mathrm{d}z = \gamma^2 + \pi^2/6$, cf. the expression for σ_0^2 above.

Plugging in the numerical value of γ in (11) we find that the information is

$$i(\xi,\psi) \simeq \frac{1}{\psi^2}\left\{\begin{array}{cc} 1 & -0.4229 \\ -0.4229 & 1.824 \end{array}\right\}$$

and the inverse information is

$$i(\xi,\psi)^{-1} \simeq \psi^2\left\{\begin{array}{cc} 1.109 & 0.2570 \\ 0.2570 & 0.6079 \end{array}\right\}.$$

If $X_1,\ldots,X_n$ are iid stochastic variables with the same distribution as X we find the minus-log-likelihood function

$$l_n(\xi,\psi) = \sum_{i=1}^n -\log\lambda_{\xi,\psi}(X_i) = \sum_{i=1}^n l_{X_i}(\xi,\psi) \tag{12}$$

so the likelihood equations amounts to

$$\begin{aligned}
\sum_{i=1}^n \frac{\partial l_{X_i}}{\partial\xi}(\xi,\psi) &= \sum_{i=1}^n \frac{1}{\psi}\exp\left(-\frac{X_i-\xi}{\psi}\right) - \frac{1}{\psi} = 0 \\
\sum_{i=1}^n \frac{\partial l_{X_i}}{\partial\psi}(\xi,\psi) &= \sum_{i=1}^n \frac{1}{\psi} - \frac{X_i-\xi}{\psi^2} + \frac{X_i-\xi}{\psi^2}\exp\left(-\frac{X_i-\xi}{\psi}\right) = 0
\end{aligned}$$

or by rearraging the terms a little

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\exp\left(-\frac{X_i-\xi}{\psi}\right) &= 1\\ \frac{1}{n}\sum_{i=1}^{n}\frac{X_i-\xi}{\psi} &= 1+\frac{1}{n}\sum_{i=1}^{n}\frac{X_i-\xi}{\psi}\exp\left(-\frac{X_i-\xi}{\psi}\right)\end{aligned}$$

There seems to be no explicit solution to these equations and one must rely on numerical procedures.

The resulting maximum likelihood estimators $\hat{\xi}$ and $\hat{\psi}$ follows asymptotically a normal distribution

$$\begin{pmatrix}\hat{\xi}\\ \hat{\psi}\end{pmatrix} \overset{\text{as}}{\sim} N\left(\begin{pmatrix}\xi\\ \psi\end{pmatrix}, \frac{1}{n} i(\xi,\psi)^{-1}\right)$$

where the inverse information is given above. We should especially notice that the maximum likelihood estimator $\hat{\psi}$ of the scale parameter has close to half the asymptotic variance as compared to $\tilde{\psi}$. For the location parameter the difference in variance is on the other hand small. One should perhaps also notice that whereas the asymptotic correlation of $\tilde{\xi}$ and $\tilde{\psi}$ is low (0.0845) the asymptotic correlation of the maximum-likelihood estimators $\hat{\xi}$ and $\hat{\psi}$ turns out to be somewhat larger (0.313).