Inference for the Pareto, half normal and related distributions

Hassan Abuhassan and David J. Olive

Department of Mathematics, Southern Illinois University, Mailcode 4408, Carbondale, IL 62901-4408, USA

July 24, 2008

Abstract

A simple technique for deriving exact confidence intervals for the Pareto and power distributions is given, and an improved confidence interval for the location parameter of the half normal distribution is given. A competitor for the Pareto distribution is obtained by transforming the half normal distribution.

Keywords: Exponential distribution, Half normal distribution, Pareto distribution, Power distibution, Robust estimators

1 Introduction

The Pareto distribution is used to model economic data such as national incomes. See Arnold (1983) for additional applications. If Y has a Pareto distribution, $Y \sim \text{PAR}(\sigma, \lambda)$, then the probability density function (pdf) of Y is

$$f(y) = \frac{\sigma^{1/\lambda}}{\lambda \, y^{1+1/\lambda}}$$

where $y \geq \sigma > 0$ and $\lambda > 0$. The distribution function of Y is $F(y) = 1 - (\sigma/y)^{1/\lambda}$ for $y > \sigma$. This family is a scale family when λ is fixed.

Exact $100(1-\alpha)\%$ confidence intervals (CIs) for σ and λ that are based on the maximum likelihood estimators (MLEs) will be developed in Section 3 using the fact that $W = \log(Y)$ has an exponential distribution. Before reviewing inference for the exponential distribution, the following notation will be useful. Suppose that $X \sim \chi^2_d$ has a chi–square distribution with d degrees of freedom. Let $\chi^2_{d,\alpha}$ denote the α percentile of the χ^2_d distribution: $P(X \leq \chi^2_{d,\alpha}) = \alpha$ for $0 < \alpha < 1$. Let $X \sim G(\nu, \lambda)$ indicate that X has a gamma distribution where $\nu > 0$ and $\lambda > 0$. If $X \sim \chi^2_d$, then $X \sim G(d/2, 2)$. If $Z \sim N(0,1)$ has a standard normal distribution, let $P(Z < z_\alpha) = \alpha$.

If W has an exponential distribution, $W \sim EXP(\theta, \lambda)$, then the pdf of W is

$$f(w) = \frac{1}{\lambda} \exp\left(\frac{-(w-\theta)}{\lambda} \right)$$

where $\lambda > 0$, $w \geq \theta$ and θ is real. This is a location–scale family. If $X \sim EXP(\lambda)$, then $X \sim EXP(0, \lambda)$ has a one parameter exponential distribution and $X \sim G(1, \lambda)$.

Inference for this distribution is discussed, for example, in Johnson and Kotz (1970, p. 219) and Mann, Schafer, and Singpurwalla (1974, p. 176). Let $W_1, ..., W_n$ be independent and identically distributed (iid) $EXP(\theta, \lambda)$ random variables. Let $W_{1:n} = \min(W_1, ..., W_n)$. Then the MLE

$$(\hat{\theta}, \hat{\lambda}) = \left(W_{1:n}, \frac{1}{n} \sum_{i=1}^{n} (W_i - W_{1:n}) \right) = (W_{1:n}, \overline{W} - W_{1:n}).$$

Let $D_n = n\hat{\lambda}$. For $n > 1$, an exact $100(1-\alpha)\%$ confidence interval for θ is

$$(W_{1:n} - \hat{\lambda}[(\alpha)^{-1/(n-1)} - 1], W_{1:n}) \tag{1.1}$$

while an exact $100(1-\alpha)\%$ CI for λ is

$$\left(\frac{2D_n}{\chi^2_{2(n-1),1-\alpha/2}}, \frac{2D_n}{\chi^2_{2(n-1),\alpha/2}} \right). \tag{1.2}$$

It can be shown that n length of CI (1.1) converges in probability to $-\lambda \log(\alpha)$. By the central limit theorem, $\chi^2_n \approx \sqrt{2n}Z + n \sim N(n, 2n)$. Thus $\chi^2_{n,\alpha}/\sqrt{n} \approx \sqrt{2}z_\alpha + \sqrt{n}$, and it can be shown that $\sqrt{n}$ length of CI (1.2) converges in probability to $\lambda(z_{1-\alpha/2} - z_{\alpha/2})$.

Suppose that $Y = t(W)$ and $W = t^{-1}(Y)$ where W has a pdf f_W with parameters $\boldsymbol{\theta}$, and the transformation t does not depend on any unknown parameters. The pdf of Y is

$$f_Y(y) = f_W(t^{-1}(y)) \left| \frac{dt^{-1}(y)}{dy} \right|.$$

If $W_1, ..., W_n$ are iid with pdf $f_W(w)$, assume that the MLE of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}}_W(\boldsymbol{w})$ where the w_i are the observed values of W_i and $\boldsymbol{w} = (w_1, ..., w_n)^T$. Following Brownstein and Pensky (2008), if $Y_1, ..., Y_n$ are iid and the y_i are the observed values of Y_i, then maximizing the log likelihood $\log(L_Y(\boldsymbol{\theta}))$ is equivalent to maximizing $\log(L_W(\boldsymbol{\theta}))$ and the MLE

$$\hat{\boldsymbol{\theta}}_Y(\boldsymbol{y}) = \hat{\boldsymbol{\theta}}_W(\boldsymbol{w}) = \hat{\boldsymbol{\theta}}_W(t^{-1}(y_1), ..., t^{-1}(y_n)). \tag{1.3}$$

This result is useful since if the MLE based on the W_i has simple inference, then the MLE based on the Y_i will also have simple inference. For example, if $W_1, ..., W_n$ are iid

$EXP(\theta = \log(\sigma), \lambda)$ and $Y_1, ..., Y_n$ are iid Pareto $(\sigma = e^\theta, \lambda)$, then $Y = e^W = t(W)$ and $W = \log(Y) = t^{-1}(Y)$. The MLE of (θ, λ) based on the W_i is $(\hat{\theta}, \hat{\lambda}) = (W_{1:n}, \overline{W} - W_{1:n})$. Hence by (1.3) and invariance, the MLE of (σ, λ) based on the Y_i is $(\hat{\sigma}, \hat{\lambda})$ where $\hat{\sigma} = \exp(\hat{\theta}) = \exp(W_{1:n}) = Y_{1:n}$ and $\hat{\lambda} = \overline{W} - W_{1:n} = \frac{1}{n} \sum_{i=1}^n \log(Y_i) - \log(Y_{1:n})$.

A competitor for the Pareto distribution, called the hpar distribution, can be created by transforming the half normal distribution. If W has a half normal distribution, $W \sim HN(\mu, \sigma)$, then the pdf of W is

$$f(w) = \frac{2}{\sqrt{2\pi}\ \sigma} \exp\left(\frac{-(w-\mu)^2}{2\sigma^2}\right)$$

where $\sigma > 0$, $w \geq \mu$ and μ is real. This is a location–scale family that is very similar to the exponential distribution in shape. Section 2 shows that inference for the half normal distribution is similar to that of the exponential distribution.

If $Y \sim hpar(\theta, \lambda)$, then $W = \log(Y) \sim HN(\mu = \log(\theta), \sigma = \lambda)$. If $W \sim HN(\mu, \sigma)$, then $Y = e^W \sim hpar(\theta = e^\mu, \lambda = \sigma)$. The pdf of Y is

$$f(y) = \frac{2}{\sqrt{2\pi}\lambda} \frac{1}{y} \exp\left[\frac{-(\log(y) - \log(\theta))^2}{2\lambda^2}\right]$$

where $y \geq \theta > 0$ and $\lambda > 0$. Using (1.3) and Section 2, the MLE is $(\hat{\theta}, \hat{\lambda})$ where $\hat{\theta} = Y_{1:n}$ and

$$\hat{\lambda} = \sqrt{\frac{\sum_{i=1}^n [\log(Y_i) - \log(Y_{1:n})]^2}{n}}.$$

Section 3 derives exact CIs for the Pareto and power distributions, large sample CIs for the hpar distribution, and presents a small simulation study. Section 4 gives two additional transformed distributions.

2 Inference for the half normal distribution

Suppose $W_1, ..., W_n$ are iid $HN(\mu, \sigma)$. Let

$$D_n = \sum_{i=1}^{n} (W_i - W_{1:n})^2. \tag{2.1}$$

The MLE of (μ, σ^2) is $(\hat{\mu}, \hat{\sigma}^2) = (W_{1:n}, \frac{1}{n} D_n)$. Pewsey (2002) showed that

$$\frac{W_{1:n} - \mu}{\sigma \Phi^{-1}(\frac{1}{2} + \frac{1}{2n})} \xrightarrow{D} EXP(1),$$

(where $\xrightarrow{D}$ denotes convergence in distribution) and noted that $(\sqrt{\pi/2})/n$ is an approximation to $\Phi^{-1}(\frac{1}{2} + \frac{1}{2n})$ based on a first order Taylor series expansion such that

$$\frac{\Phi^{-1}(\frac{1}{2} + \frac{1}{2n})}{(\sqrt{\pi/2})/n} \to 1.$$

Thus

$$\frac{n(W_{1:n} - \mu)}{\sigma \sqrt{\frac{\pi}{2}}} \xrightarrow{D} EXP(1). \tag{2.2}$$

Using this fact, Pewsey (2002) noted that a large sample $100(1-\alpha)\%$ CI for μ is

$$(\hat{\mu} + \hat{\sigma} \log(\alpha/2) \Phi^{-1}(\frac{1}{2} + \frac{1}{2n}), \ \ \hat{\mu} + \hat{\sigma} \log(1 - \alpha/2) \Phi^{-1}(\frac{1}{2} + \frac{1}{2n})). \tag{2.3}$$

Let (L_n, U_n) be an exact or large sample $100(1-\alpha)\%$ CI for θ. If

$$n^{\delta}(U_n - L_n) \xrightarrow{P} A_{\alpha},$$

then A_α is the *scaled asymptotic length* of the CI. Typically $\delta = 0.5$ but superefficient CIs have $\delta = 1$. The CIs (1.1), (2.3) and (2.4) below are superefficient. For fixed δ and fixed coverage $1 - \alpha$, a CI with smaller A_α is "better" than a CI with larger A_α. The scaled expected CI length often converges to the scaled asymptotic length, which is often easier to compute. If $A_{1,\alpha}$ and $A_{2,\alpha}$ are for two competing CIs with the same δ, then $(A_{2,\alpha}/A_{1,\alpha})^{1/\delta}$ is a measure of "asymptotic relative efficiency."

Consider CIs for μ of form (2.3) that use cutoffs α_1 and $1-\alpha_2$ where $\alpha_1+\alpha_2 = \alpha$. Since the exponential distribution is decreasing, the asymptotically shortest CI is obtained by using as little of the right tail as possible as in (1.1). Thus the large sample $100(1-\alpha)\%$ CI for μ with the shortest asymptotic scaled length is

$$(\hat{\mu} + \hat{\sigma}\log(\alpha)\ \Phi^{-1}(\frac{1}{2} + \frac{1}{2n})\ (1 + 13/n^2), \quad \hat{\mu}). \tag{2.4}$$

The term $(1 + 13/n^2)$ is a small sample correction factor chosen so that the coverage of CI (2.4) is close to 0.95 for $n \geq 5$. Similar correction factors were used by Olive (2007) for prediction intervals.

The new CI (2.4) has $A_\alpha = -\sigma\log(\alpha)\sqrt{\pi/2}$ while the CI (2.3) has $A_\alpha = -\sigma[\log(1-\alpha/2) - \log(\alpha/2)]\sqrt{\pi/2}$. For a 95% CI, the CI (2.3) has $A_{0.05} = 3.6636\sigma\sqrt{\pi/2}$ while the CI (2.4) has $A_{0.05} = 2.9957\sigma\sqrt{\pi/2}$. Since $\delta = 1$, the 95% CI (2.3) has about 82% "asymptotic relative efficiency" compared to the 95% CI (2.4).

Pewsey (2002) suggested that a large sample $100(1-\alpha)\%$ CI for σ^2 is

$$\left(\frac{D_n}{\chi^2_{n-1,1-\alpha/2}}, \frac{D_n}{\chi^2_{n-1,\alpha/2}}\right). \tag{2.5}$$

It can be shown that $\sqrt{n}$ length of CI (2.5) converges in probability to $\sigma^2\sqrt{2}(z_{1-\alpha/2} - z_{\alpha/2})$. This "equal tail" CI has shortest asymptotic length of CIs of form (2.5) that use cutoffs α_1 and $1-\alpha_2$ where $\alpha_1+\alpha_2 = \alpha$. For fixed $n > 25$, the CI (2.5) is short since the χ^2_{n-1} distribution is approximately symmetric for $n > 25$. Pewsey (2002) claimed that $D_n \xrightarrow{D} \sigma^2\chi^2_{n-1}$, which is impossible since the limiting distribution can not depend on the sample size n. The appendix gives a correct justification of the CI.

3 Inference for the Pareto, power and hpar distributions

Suppose $Y_1, ..., Y_n$ are iid PAR(σ, λ). Arnold (1983) notes that $\log(Y/\sigma) \sim EXP(0, \lambda)$ can be used to simplify many of the results in the literature. The parameter free transformation $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$ is even more useful, but this result seems to be nearly unknown with an exception of Brownstein and Pensky (2008).

Let $\theta = \log(\sigma)$, so $\sigma = e^\theta$. By (1.3) the MLE $(\hat{\theta}, \hat{\lambda}) = (W_{1:n}, \overline{W} - W_{1:n})$, and by invariance, the MLE $(\hat{\sigma}, \hat{\lambda}) = (Y_{1:n}, \overline{W} - W_{1:n})$. Inference is simple. An exact $100(1-\alpha)\%$ CI for θ is (1.1). A $100(1-\alpha)\%$ CI for σ is obtained by exponentiating the endpoints of (1.1), and an exact $100(1-\alpha)\%$ CI for λ is (1.2).

It is well known that a parameter free one to one transformation makes interval estimation simple. See, for example, Brownstein and Pensky (2008). The interval for λ can also be derived using the pivotal $n\hat{\lambda}/\lambda \sim G(n-1, 1)$ given in Arnold (1983, p. 217) and Muniruzzaman (1957). The interval for σ seems to be new. Arnold (1983, pp. 195, 216-217) states that $n \log(\hat{\sigma}/\sigma)/\lambda \sim G(1, 1)$ and derives a joint confidence region for σ and $1/\lambda$. Arnold (1983, p. 217) and Grimshaw (1993) suggest using large sample confidence intervals of the form MLE $\pm z_{1-\alpha/2}$ se(MLE) where se is the standard error. Also see references in Kuş and Kaya (2007).

Now suppose that $Y_1, ..., Y_n$ are iid hpar(θ, λ), then $W = \log(Y) \sim HN(\mu = \log(\theta), \sigma = \lambda)$. Inference is again simple. A large sample $100(1-\alpha)\%$ CI for μ is (2.4). A large sample $100(1-\alpha)\%$ CI for $\theta = e^\mu$ is obtained by exponentiating the endpoints of (2.4), and a large sample $100(1-\alpha)\%$ CI for λ^2 is (2.5). Taking square roots of the endpoints gives a large sample $100(1-\alpha)\%$ CI for λ.

Inference for the power distribution is very similar to that of the Pareto distribution. If Y has a power distribution, $Y \sim power(\tau, \lambda)$, then the pdf

$$f(y) = \frac{1}{\tau\lambda}\left(\frac{y}{\tau}\right)^{\frac{1}{\lambda}-1} I(0 < y \leq \tau).$$

Then $W = -\log(Y) \sim EXP(-\log(\tau), \lambda)$. Thus (1.2) is an exact $100(1-\alpha)\%$ CI for λ, and (1.1) $= (L_n, U_n)$ is an exact $100(1-\alpha)\%$ CI for $\theta = -\log(\tau)$. Hence (e^{L_n}, e^{U_n}) is a $100(1-\alpha)\%$ CI for $1/\tau$, and (e^{-U_n}, e^{-L_n}) is a $100(1-\alpha)\%$ CI for τ.

A small simulation study used 5000 runs, and `cov` was the proportion of times the 95% CI contained the parameter. For 5000 runs, an observed coverage between 0.94 and 0.96 gives little evidence that the true coverage differs from the nominal coverage of 0.95. The scaled length `slen` multiplied the length of the CI by $\sqrt{n}$ for CIs (1.2) and (2.5) and by n for CIs (1.1) and (2.4). The $n = \infty$ lines give the asymptotic values. Programs `hnsim` and `expsim` for `R/Splus` are available from (www.math.siu.edu/olive/sipack.txt).

Table 1 provides results for CIs (1.1), (1.2), (2.4) and (2.5). The CI (1.1) is for θ if $Y \sim EXP(\theta, \lambda)$, for $\theta = \log(\sigma)$ if $Y \sim$ Pareto $(\sigma = e^\theta, \lambda)$, and for $\theta = -\log(\tau)$ if $Y \sim power(\tau, \lambda)$. The CI (1.2) is for λ for all three distributions. The CI (2.4) is for μ if $Y \sim HN(\mu, \sigma^2)$ or if $Y \sim hpar(\theta = e^\mu, \lambda = \sigma)$. The CI (2.5) is for σ^2 if $Y \sim HN(\mu, \sigma^2)$ and for λ^2 if $Y \sim hpar(\theta = e^\mu, \lambda = \sigma)$. The simulations used $\lambda = 1$. Using alternative values would simply change the scaled CI length by a factor of λ. The coverage and scaled length do not depend on the location parameter.

Table 1 shows that the scaled asymptotic length is a good approximation for the scaled average length for $n = 1000$, and not too bad for $n = 50$. Observed coverages are close to 0.95 for all entries. Of course the coverage for the exact intervals (1.1) and (1.2) is 0.95 for $n > 1$.

Arnold (1983, pp. 280-281) describes two data sets. The golf data is lifetime earnings in thousands of dollars for 50 golfers who had earned more that \$700,000 through 1980. The Pareto distribution with $\sigma = 700$ and $\hat{\lambda} = 0.4396$ was used to model this data. Treating both λ and σ as unknown gave $\hat{\lambda} = 0.428$ with 95% CI (0.336,0.591) and $\hat{\sigma} = 708$ with 95% CI (689.144,708).

The county data consists of the 157 of 254 Texas counties in which total personal income in millions of dollars exceeded \$20 million in 1969. The Pareto distribution with $\sigma = 20$ and $\hat{\lambda} = 1.179$ was used to model this data. Treating both λ and σ as unknown gave $\hat{\lambda} = 1.184$ with 95% CI (1.025,1.404) and $\hat{\sigma} = 20.2$ with 95% CI (19.741,20.2). Notice that for both data sets the CI for σ was short and contained the "true value" of σ. The data sets are available from (www.math.siu.edu/olive/sidata.txt).

4 Some Transformed Distributions

Section 3 gave the Pareto and hpar distributions. If $Y \sim \text{PAR}(\sigma, \lambda)$, then $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$. If $Y \sim hpar(\theta, \lambda)$, then $W = \log(Y) \sim HN(\mu = \log(\theta), \sigma = \lambda)$. Hence the Pareto distribution is obtained by transforming the exponential distribution and the hpar distribution is obtained by applying the same transformation to the half normal distribution.

Several other distributions, including the power distribution, are obtained by transforming the exponential distribution. Competitors for these distributions can be obtained by transforming the half normal distribution. Before giving two examples, theory for the one parameter exponential and half normal distributions is needed.

If the W_i are iid $EXP(\theta, \lambda)$ with θ known, then

$$\hat{\lambda}_\theta = \frac{\sum_{i=1}^n (W_i - \theta)}{n} = \overline{W} - \theta$$

is the UMVUE and MLE of λ, and a $100(1-\alpha)\%$ CI for λ is

$$\left(\frac{2T_n}{\chi^2_{2n,1-\alpha/2}}, \frac{2T_n}{\chi^2_{2n,\alpha/2}} \right). \tag{4.1}$$

It can be shown that $\sqrt{n}$ CI length converges to $\lambda(z_{1-\alpha/2} - z_{\alpha/2})$ in probability.

If the W_i are iid $HN(\mu, \sigma)$ with μ known, then $T_n = \sum_{i=1}^n (W_i - \mu)^2 \sim G(n/2, 2\sigma^2)$, and a $100(1-\alpha)\%$ CI for σ^2 is

$$\left(\frac{T_n}{\chi^2_n(1-\alpha/2)}, \frac{T_n}{\chi^2_n(\alpha/2)} \right). \tag{4.2}$$

It can be shown that $\sqrt{n}$ CI length converges to $\sigma^2\sqrt{2}(z_{1-\alpha/2} - z_{\alpha/2})$ in probability.

Example 1: One parameter Power vs. hpow.

If Y has a one parameter power distribution, $Y \sim POW(\lambda)$, then the pdf of Y is

$$f(y) = \frac{1}{\lambda} y^{\frac{1}{\lambda}-1} I_{(0,1)}(y) = \frac{1}{\lambda}\frac{1}{y} I_{(0,1)}(y) \exp\left[\frac{-1}{\lambda}(-\log(y))\right]$$

where $\lambda > 0$. The cdf of Y is $F(y) = y^{1/\lambda}$ for $0 \le y \le 1$. $W = -\log(Y)$ is $EXP(0, \lambda)$. $T_n = -\sum \log(Y_i) \sim G(n, \lambda)$, and

$$\hat{\lambda} = \frac{-\sum_{i=1}^n \log(Y_i)}{n}$$

is the UMVUE and MLE of λ. A robust estimator is $\hat{\lambda}_R = \log(\text{MED}(n))/\log(0.5)$. A 100 $(1-\alpha)\%$ CI for λ is (4.1).

If $Y \sim hpow(\lambda)$, then $W = -\log(Y) \sim HN(0, \sigma = \lambda)$. If $W \sim HN(0, \sigma)$, then $Y = e^{-W} \sim hpow(\lambda = \sigma)$. The pdf of Y is

$$f(y) = \frac{2}{\sqrt{2\pi}\lambda} \frac{1}{y} I_{(0,1)}(y) \exp\left[\frac{-(\log(y))^2}{2\lambda^2}\right]$$

where $\lambda > 0$, and $T_n = \sum_{i=1}^n [\log(Y_i)]^2 \sim G(n/2, 2\lambda^2)$. The MLE $\hat{\lambda} = \sqrt{\sum_{i=1}^n W_i^2/n}$ and a large sample 100 $(1-\alpha)$% CI for λ^2 is (4.2). Taking square roots of the endpoints gives a large sample 100 $(1-\alpha)$% CI for λ. A two parameter hpow distribution can be handled similarly.

Example 2: Truncated extreme value vs. htev.

If Y has a truncated extreme value distribution, $Y \sim TEV(\lambda)$, then the pdf of Y is

$$f(y) = \frac{1}{\lambda} \exp\left(y - \frac{e^y - 1}{\lambda}\right) I(y \geq 0) = \frac{1}{\lambda} e^y I(y \geq 0) \exp\left[\frac{-1}{\lambda}(e^y - 1)\right]$$

where $\lambda > 0$. The cdf of Y is

$$F(y) = 1 - \exp\left[\frac{-(e^y - 1)}{\lambda}\right]$$

for $y > 0$. $W = e^Y - 1$ is $EXP(0, \lambda)$, and $T_n = \sum(e^{Y_i} - 1) \sim G(n, \lambda)$.

$$\hat{\lambda} = \frac{\sum(e^{Y_i} - 1)}{n}$$

is the UMVUE and MLE of λ. A robust point estimator is $\hat{\lambda}_R = [\exp(\text{MED}(n)) - 1]/\log(2)$. A 100 $(1-\alpha)$% CI for λ is (4.1).

If $Y \sim htev(\lambda)$, then $W = e^Y - 1 \sim HN(0, \sigma = \lambda)$. If $W \sim HN(0, \sigma)$, then $Y = \log(W + 1) \sim htev(\lambda = \sigma)$. The pdf of Y is

$$f(y) = \frac{2}{\sqrt{2\pi}\lambda} \exp\left(y - \frac{(e^y - 1)^2}{2\lambda^2}\right) I(y > 0) = \frac{2}{\sqrt{2\pi}\lambda} e^y \exp\left(\frac{-(e^y - 1)^2}{2\lambda^2}\right) I(y > 0)$$

where $\lambda > 0$. $T_n = \sum_{i=1}^n (e^{Y_i} - 1)^2 \sim G(n/2, 2\lambda^2)$. The MLE $\hat{\lambda} = \sqrt{\sum_{i=1}^n W_i^2/n}$ and a large sample 100 $(1-\alpha)$% CI for λ^2 is (4.2). Taking square roots of the endpoints gives a large sample 100 $(1-\alpha)$% CI for λ.

5 Conclusions

The results in this paper are based on Abuhassan (2007), who derives the asymptotic lengths of the CIs. If $Y \sim PAR(\sigma, \lambda)$, then $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$ and the MLE $(\hat{\theta}, \hat{\lambda}) = (W_{1:n}, \overline{W} - W_{1:n})$ for both the Pareto and exponential distributions. Hence the exact CIs for the exponential distribution are exact CIs for λ and $\theta = \log(\lambda)$ for the Pareto distribution.

Improvements on the Pewsey (2002) confidence intervals and corresponding theory for the half normal distribution are also given. Several important distributions are obtained by transforming the exponential distribution. Competitors for these distributions can be obtained by applying the same transformation to the half normal distribution. Abuhassan (2007) gives several examples, including the hpar, hpow and htev distributions.

The exponential and half normal distributions are interesting location scale families because superefficient (rate n instead of rate $\sqrt{n}$) estimators of the location parameter exist. For example, for the half normal distribution, an approximate α level test of $H_0 : \sigma^2 \leq \sigma_o^2$ versus $H_A : \sigma^2 > \sigma_0^2$ that rejects H_0 if and only if

$$D_n > \sigma_0^2 \chi^2_{n-1}(1-\alpha) \tag{5.1}$$

has nearly as much power as the α level uniformly most powerful test when μ is known if n is large.

Since $W = \log(Y) \sim EXP(\theta = \log(\sigma), \lambda)$ if $Y \sim PAR(\sigma, \lambda)$, the large literature on the exponential distribution can be used to derive methods such as CIs, tests of hypotheses and ways for handling censored data for the Pareto distribution. See Fernandez (2008) and Wu (2008) for alternative approaches.

For example, to derive robust estimators, follow Olive (2008, Ch. 3), and let MED(W) and MAD(W) = MED($|W - \text{MED}(W)|$) be the population median and median absolute deviation. Let MED(n) and MAD(n) be the sample versions. For the exponential distribution, MED(W) $= \theta + \lambda \log(2)$ and MAD(W) $= \lambda/2.0781$. Hence robust point estimators for both the exponential and Pareto distributions are $\hat{\theta}_R = \text{MED}(n) - 1.440\ \text{MAD}(n)$, $\hat{\lambda}_R = 2.0781\ \text{MAD}(n)$ and $\hat{\sigma}_R = \exp(\hat{\theta}_R)$. For the golf data, $\hat{\lambda}_R = 0.449$ and $\hat{\sigma}_R = 741.435$ while for the county data $\hat{\lambda}_R = 1.118$ and $\hat{\sigma}_R = 21.410$

6 Appendix

For iid half normal data, note that $T_n = \sum_{i=1}^n (W_i - \mu)^2 \sim G(n/2, 2\sigma^2)$, and

$$D_n = \sum_{i=1}^n (W_i - W_{1:n})^2 = \sum_{i=1}^n (W_i - \mu + \mu - W_{1:n})^2 =$$

$$\sum_{i=1}^n (W_i - \mu)^2 + n(\mu - W_{1:n})^2 + 2(\mu - W_{1:n}) \sum_{i=1}^n (W_i - \mu).$$

Hence

$$D_n = T_n + \frac{1}{n}[n(W_{1:n} - \mu)]^2 - 2[n(W_{1:n} - \mu)] \frac{\sum_{i=1}^n (W_i - \mu)}{n},$$

or

$$\frac{D_n - T_n}{\sigma^2} = \frac{1}{n}\frac{1}{\sigma^2}[n(W_{1:n} - \mu)]^2 - 2[\frac{n(W_{1:n} - \mu)}{\sigma}] \frac{\sum_{i=1}^n (W_i - \mu)}{n\sigma} \tag{6.1}$$

$$\text{or} \quad \frac{\mathrm{D_n} - \mathrm{T_n}}{\sigma^2} \xrightarrow{\mathrm{D}} -\chi^2_2.$$

To see the convergence, consider the last two terms of (6.1). By (2.2) the first term converges to 0 in distribution while the second term converges in distribution to a $-2EXP(1)$ or $-\chi^2_2$ distribution since $\sum_{i=1}^n (W_i - \mu)/(\sigma n)$ is the sample mean of HN(0,1) random variables and $E(X) = \sqrt{2/\pi}$ when $X \sim HN(0,1)$.

Let $T_{n-p} = \sum_{i=1}^{n-p} (W_i - \mu)^2 \sim \sigma^2 \chi^2_{n-p}$. Then

$$D_n = T_{n-p} + \sum_{i=n-p+1}^{n} (W_i - \mu)^2 - V_n \tag{6.2}$$

where $\frac{V_n}{\sigma^2} \xrightarrow{D} \chi^2_2$. Hence $\frac{D_n}{T_{n-p}} \xrightarrow{D} 1$. Thus a large sample $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{D_n}{\chi^2_{n-p,1-\alpha/2}}, \frac{D_n}{\chi^2_{n-p,\alpha/2}} \right). \tag{6.3}$$

For finite samples, the choice of p is important. Since $W_{1:n} > \mu$, notice that $T_n = \sum (W_i - \mu)^2 > \sum (W_i - W_{1:n})^2 = D_n$, and $D_n/\sigma^2 \approx \chi^2_{n-p} + \chi^2_p - \chi^2_2$ where χ^2_p is replaced by 0 for $p = 0$. Hence $p = 0$ is too small while $p > 2$ is too large. Thus $p = 1$ or $p = 2$ should be used. Pewsey (2002) used $p = 1$ and the simulations showed that the confidence interval coverage was good for n as small as 20. For $n = 5$ a nominal 95% interval appears to have coverage between 0.94 and 0.95 with a length that is about half that of the CI that uses $p = 2$. The $p = 2$ CI appears to have coverage between 0.95 and 0.96. We followed Pewsey (2002) and used $p = 1$.

Acknowledgement

This research was supported by NSF grant DMS-0600933.

References

Abuhassan, H., 2007. Some Transformed Distributions. Ph.D. Thesis, Southern Illinois University, online at (www.math.siu.edu/olive/shassan.pdf).

Arnold, B.C., 1983. Pareto Distributions. International Co-operative Publishing House, Fairland, MD.

Brownstein, N., Pensky, M., 2008. Application of transformations in parametric inference. J. Statist. Educat., 16, online at (www.amstat.org/publications/jse/v16n1/brownstein.pdf).

Fernandez, A.J., 2008. Reliability inference and sample-size determination under double censoring for some two-parameter models. Computat. Statist. Data Analys. 52, 3426-3440.

Grimshaw, S.D., 1993. Computing maximum likelihood estimates for the generalized Pareto distribution. Technometrics 35, 185-191.

Johnson, N.L., Kotz, S., 1970. Continuous Univariate Distributions, Vol. 1. Houghton Mifflin Company, Boston, MA.

Kuş, C., Kaya, M.F., 2007. Estimating the parameters of the Pareto distribution under progressive censoring. Commun. Statist. Theory and Methods 36, 1359-1365.

Mann, N.R., Schafer, R.E., Singpurwalla, N.D., 1974. Methods for Statistical Analysis of Reliability and Life Data. Wiley, New York.

Muniruzzaman, A.N.M., 1957. On measures of location and dispersion and tests of hypotheses in a Pareto population,. Bull. Calcutta Statist. Assoc. 7, 115-123.

Olive, D.J., 2007. Prediction intervals for regression models. Computat. Statist. Data Analys. 51, 3115-3122.

Olive, D.J., 2008. Applied Robust Statistics. Unpublished text online at (www.math.siu.edu/olive/).

Pewsey, A., 2002. Large-sample inference for the half-normal distribution. Commun. Statist. Theory and Methods 31, 1045-1054.

Wu, S.-F. (2008), Interval estimation for a Pareto distribution based on a doubly type II censored sample. Computat. Statist. Data Analys. 52, 3779-3788.

Table 1: CIs for Exponential, Pareto, Power, HN and Hpar Distributions

	(1.2)	(1.2)	(1.1)	(1.1)	(2.5)	(2.5)	(2.4)	(2.4)
n	cov	slen	cov	slen	cov	slen	cov	slen
5	0.9532	7.2453	0.9526	4.4954	0.9554	15.1766	0.9534	4.4727
10	0.9492	5.1485	0.9538	3.5812	0.9450	8.2714	0.9440	3.7436
20	0.9534	4.5400	0.9456	3.2543	0.9494	6.6917	0.9436	3.6591
50	0.9534	4.1096	0.9498	3.0861	0.9460	5.9051	0.9458	3.6732
100	0.9494	4.0112	0.9482	3.0390	0.9508	5.7201	0.9436	3.7099
1000	0.9504	3.9361	0.9492	3.0055	0.9480	5.5633	0.9524	3.7506
∞	0.95	3.9199	0.95	2.9957	0.95	5.5437	0.95	3.7546