Chapter 3

Some Useful Distributions

The two stage trimmed means of Chapter 2 are asymptotically equivalent to a classical trimmed mean provided that $A_n = \text{MED}(n) - k_1\text{MAD}(n) \xrightarrow{D} a$, $B_n = \text{MED}(n) + k_2\text{MAD}(n) \xrightarrow{D} b$ and if $100F(a-)$ and $100F(b)$ are not integers. This result will also hold if k_1 and k_2 depend on n. For example take $k_1 = k_2 = c_1 + c_2/n$. Then $\text{MED}(n) \pm k_1\text{MAD}(n) \xrightarrow{D} \text{MED}(Y) \pm c_1\text{MAD}(Y)$. A *trimming rule* suggests values for c_1 and c_2 and depends on the distribution of Y. Sometimes the rule is obtained by transforming the random variable Y into another random variable W (eg transform a lognormal into a normal) and then using the rule for W. These rules may not be as resistant to outliers as rules that do not use a transformation. For example, an observation which does not seem to be an outlier on the log scale may appear as an outlier on the original scale.

Several of the trimming rules in this chapter have been tailored so that the probability is high that none of the observations are trimmed when the sample size is moderate. Robust (but perhaps ad hoc) analogs of classical procedures can be obtained by applying the classical procedure to the data that remains after trimming.

Relationships between the distribution's parameters and MED(Y) and MAD(Y) are emphasized. Note that for location–scale families, highly outlier resistant estimates for the two parameters can be obtained by replacing MED(Y) by MED(n) and MAD(Y) by MAD(n).

Definition 3.1. The *moment generating function* (mgf) of a random variable Y is

$$m(t) = E(e^{tY})$$

provided that the expectation exists for t in some neighborhood of 0.

Definition 3.2. The *characteristic function* (chf) of a random variable Y is

$$c(t) = E(e^{itY})$$

where the complex number $i = \sqrt{-1}$.

Definition 3.3. The *indicator function* $I_A(x) \equiv I(x \in A) = 1$ if $x \in A$ and 0, otherwise. Sometimes an indicator function such as $I_{(0,\infty)}(y)$ will be denoted by $I(y > 0)$.

3.1 The Binomial Distribution

If Y has a binomial distribution, $Y \sim \text{BIN}(\text{k}, \rho)$, then the probability mass function (pmf) of Y is

$$P(Y = y) = \binom{k}{y} \rho^y (1-\rho)^{k-y}$$

for $0 < \rho < 1$ and $y = 0, 1, \ldots, k$.
The moment generating function $m(t) = ((1-\rho) + \rho e^t)^k$, and the characteristic function $c(t) = ((1-\rho) + \rho e^{it})^k$.
$E(Y) = k\rho$, and
$\text{VAR}(Y) = k\rho(1-\rho)$.

The following normal approximation is often used.

$$Y \approx N(k\rho, k\rho(1-\rho))$$

when $k\rho(1-\rho) > 9$. Hence

$$P(Y \leq y) \approx \Phi\left(\frac{y + 0.5 - k\rho}{\sqrt{k\rho(1-\rho)}}\right).$$

Also

$$P(Y = y) \approx \frac{1}{\sqrt{k\rho(1-\rho)}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(y-k\rho)^2}{k\rho(1-\rho)}\right).$$

See Johnson, Kotz and Kemp (1992, p. 115). This normal approximation suggests that $\text{MED}(Y) \approx k\rho$, and $\text{MAD}(Y) \approx 0.6745\sqrt{k\rho(1-\rho)}$. Hamza (1995) states that $|E(Y) - \text{MED}(Y)| \leq \max(\rho, 1-\rho)$ and shows that

$$|E(Y) - \text{MED}(Y)| \leq \log(2).$$

Following Olive (2008, ch. 9), let $W = \sum_{i=1}^n Y_i \sim \text{bin}(\sum_{\text{i=1}}^{\text{n}} \text{k}_\text{i}, \rho)$ and let $n_w = \sum_{i=1}^n k_i$. Often $k_i \equiv 1$ and then $n_w = n$. Let $P(F_{d_1,d_2} \leq F_{d_1,d_2}(\alpha)) = \alpha$ where F_{d_1,d_2} has an F distribution with d_1 and d_2 degrees of freedom. Then the Clopper Pearson "exact" 100 $(1-\alpha)$% CI for ρ is

$$\left(0, \frac{1}{1 \ + \ n_w \ F_{2n_w,2}(\alpha)}\right) \quad \text{for} \quad \text{W} = 0,$$

$$\left(\frac{n_w}{n_w \ + \ F_{2,2n_w}(1-\alpha)}, 1\right) \quad \text{for} \quad \text{W} = \text{n}_\text{w},$$

and (ρ_L, ρ_U) for $0 < W < n_w$ with

$$\rho_L = \frac{W}{W + (n_w - W + 1)F_{2(n_w - W+1),2W}(1-\alpha/2)}$$

and

$$\rho_U = \frac{W+1}{W+1+(n_w - W)F_{2(n_w-W),2(W+1)}(\alpha/2)}.$$

Suppose $Y_1, ..., Y_n$ are iid bin$(1, \rho)$. Let $\hat{\rho}$ = number of "successes"$/n$ and let $P(Z \leq z_{1-\alpha/2}) = 1 - \alpha/2$ if $Z \sim N(0,1)$. Let $\tilde{n} = n + z_{1-\alpha/2}^2$ and

$$\tilde{\rho} = \frac{n\hat{\rho} + 0.5 z_{1-\alpha/2}^2}{n + z_{1-\alpha/2}^2}.$$

Then the large sample 100 $(1-\alpha)$% Agresti Coull CI for ρ is

$$\tilde{p} \pm z_{1-\alpha/2} \sqrt{\frac{\tilde{\rho}(1-\tilde{\rho})}{\tilde{n}}}.$$

Given a random sample of size n, the classical estimate of ρ is $\hat{\rho} = \bar{y}_n/k$. If each y_i is a nonnegative integer between 0 and k, then a trimming rule is keep y_i if

$$\text{med}(n) - 5.2(1 + \frac{4}{n})\text{mad}(n) \leq y_i \leq \text{med}(n) + 5.2(1 + \frac{4}{n})\text{mad}(n).$$

(This rule can be very bad if the normal approximation is not good.)

3.2 The Burr Distribution

If Y has a Burr distribution, $Y \sim \text{Burr}(\phi, \lambda)$, then the probability density function (pdf) of Y is

$$f(y) = \frac{1}{\lambda} \frac{\phi y^{\phi-1}}{(1+y^{\phi})^{\frac{1}{\lambda}+1}}$$

where y, ϕ, and λ are all positive. The cumulative distribution function (cdf) of Y is

$$F(y) = 1 - \exp\left[\frac{-\log(1+y^{\phi})}{\lambda}\right] = 1 - (1+y^{\phi})^{-1/\lambda} \quad \text{for} \quad \text{y} > 0.$$

$\text{MED}(Y) = [e^{\lambda \log(2)} \ - \ 1]^{1/\phi}$. See Patel, Kapadia and Owen (1976, p. 195).

Assume that ϕ is known. Since $W = \log(1+Y^{\phi})$ is $EXP(\lambda)$,

$$\hat{\lambda} = \frac{\text{MED}(W_1, ..., W_n)}{\log(2)}$$

is a robust estimator. If all the $y_i \geq 0$ then a trimming rule is keep y_i if

$$0.0 \leq w_i \leq 9.0(1 + \frac{2}{n})\text{med}(n)$$

where med(n) is applied to $w_1, \ldots, w_n$ with $w_i = \log(1 + y_i^{\phi})$.

3.3 The Cauchy Distribution

If Y has a Cauchy distribution, $Y \sim C(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{\sigma}{\pi} \frac{1}{\sigma^2 + (y-\mu)^2} = \frac{1}{\pi\sigma[1 + (\frac{y-\mu}{\sigma})^2]}$$

where y and μ are real numbers and $\sigma > 0$.
The cdf of Y is $F(y) = \frac{1}{\pi}[\arctan(\frac{y-\mu}{\sigma}) + \pi/2]$. See Ferguson (1967, p. 102). This family is a location–scale family that is symmetric about μ. The moments of Y do not exist, but the chf of Y is $c(t) = \exp(it\mu - |t|\sigma)$.
$\text{MED}(Y) = \mu$, the upper quartile $= \mu + \sigma$, and the lower quartile $= \mu - \sigma$.
$\text{MAD}(Y) = F^{-1}(3/4) - \text{MED}(Y) = \sigma$. For a standard normal random variable, 99% of the mass is between -2.58 and 2.58 while for a standard Cauchy $C(0,1)$ random variable 99% of the mass is between -63.66 and 63.66. Hence a rule which gives weight one to almost all of the observations of a Cauchy sample will be more susceptible to outliers than rules which do a large amount of trimming.

3.4 The Chi Distribution

If Y has a chi distribution, $Y \sim \chi_p$, then the pdf of Y is

$$f(y) = \frac{y^{p-1} e^{-y^2/2}}{2^{\frac{p}{2}-1}\Gamma(p/2)}$$

where $y \geq 0$ and p is a positive integer.
MED$(Y) \approx \sqrt{p - 2/3}$.
See Patel, Kapadia and Owen (1976, p. 38). Since $W = Y^2$ is χ_p^2, a trimming rule is keep y_i if $w_i = y_i^2$ would be kept by the trimming rule for χ_p^2.

3.5 The Chi–square Distribution

If Y has a chi–square distribution, $Y \sim \chi_p^2$, then the pdf of Y is

$$f(y) = \frac{y^{\frac{p}{2}-1} e^{-\frac{y}{2}}}{2^{\frac{p}{2}}\Gamma(\frac{p}{2})}$$

where $y \geq 0$ and p is a positive integer.
$E(Y) = p$.
VAR$(Y) = 2p$.
Since Y is gamma $G(\nu = p/2, \lambda = 2)$,

$$E(Y^r) = \frac{2^r \Gamma(r + p/2)}{\Gamma(p/2)}, \ r > -p/2.$$

MED$(Y) \approx p - 2/3$. See Pratt (1968, p. 1470) for more terms in the expansion of MED(Y). Empirically,

$$\text{MAD}(Y) \approx \frac{\sqrt{2p}}{1.483}(1 - \frac{2}{9p})^2 \approx 0.9536\sqrt{p}.$$

Note that $p \approx$ MED$(Y) + 2/3$, and VAR$(Y) \approx 2$MED$(Y) + 4/3$. Let i be an integer such that $i \leq w < i + 1$. Then define $rnd(w) = i$ if $i \leq w \leq i + 0.5$ and $rnd(w) = i + 1$ if $i + 0.5 < w < i + 1$. Then $p \approx rnd(\text{MED}(Y) + 2/3)$, and the approximation can be replaced by equality for $p = 1, \ldots, 100$.

There are several normal approximations for this distribution. For p large, $Y \approx N(p, 2p)$, and

$$\sqrt{2Y} \approx N(\sqrt{2p}, 1).$$

Let
$$\alpha = P(Y \le \chi^2_{p,\alpha}) = \Phi(z_\alpha)$$
where Φ is the standard normal cdf. Then
$$\chi^2_{p,\alpha} \approx \frac{1}{2}(z_\alpha + \sqrt{2p})^2.$$
The Wilson–Hilferty approximation is
$$\left(\frac{Y}{p}\right)^{\frac{1}{3}} \approx N(1 - \frac{2}{9p}, \frac{2}{9p}).$$
See Bowman and Shenton (1992, p. 6). This approximation gives
$$P(Y \le x) \approx \Phi[((\frac{x}{p})^{1/3} - 1 + 2/9p)\sqrt{9p/2}],$$
and
$$\chi^2_{p,\alpha} \approx p(z_\alpha\sqrt{\frac{2}{9p}} + 1 - \frac{2}{9p})^3.$$
The last approximation is good if $p > -1.24\log(\alpha)$. See Kennedy and Gentle (1980, p. 118).

Assume all $y_i > 0$. Let $\hat{p} = rnd(\text{med}(n) + 2/3)$. Then a trimming rule is keep y_i if
$$\frac{1}{2}(-3.5 + \sqrt{2\hat{p}})^2 I(\hat{p} \ge 15) \le y_i \le \hat{p}[(3.5 + 2.0/n)\sqrt{\frac{2}{9\hat{p}}} + 1 - \frac{2}{9\hat{p}}]^3.$$
Another trimming rule would be to let
$$w_i = \left(\frac{y_i}{\hat{p}}\right)^{1/3}.$$
Then keep y_i if the trimming rule for the normal distribution keeps the w_i.

3.6 The Double Exponential Distribution

If Y has a double exponential distribution (or Laplace distribution), $Y \sim \text{DE}(\theta, \lambda)$, then the pdf of Y is
$$f(y) = \frac{1}{2\lambda}\exp\left(\frac{-|y-\theta|}{\lambda}\right)$$

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

$$F(y) = 0.5 \exp\left(\frac{y-\theta}{\lambda}\right) \quad \text{if} \quad y \le \theta,$$

and

$$F(y) = 1 - 0.5 \exp\left(\frac{-(y-\theta)}{\lambda}\right) \quad \text{if} \quad y \ge \theta.$$

This family is a location–scale family which is symmetric about θ.
The mgf $m(t) = \exp(\theta t)/(1-\lambda^2 t^2), \ |t| < 1/\lambda$ and
the chf $c(t) = \exp(\theta i t)/(1+\lambda^2 t^2)$.
$E(Y) = \theta$, and
$\text{MED}(Y) = \theta$.
$\text{VAR}(Y) = 2\lambda^2$, and
$\text{MAD}(Y) = \log(2)\lambda \approx 0.693\lambda$.
Hence $\lambda = \text{MAD}(Y)/\log(2) \approx 1.443\text{MAD}(Y)$.
To see that $\text{MAD}(Y) = \lambda \log(2)$, note that $F(\theta + \lambda \log(2)) = 1 - 0.25 = 0.75$.

The maximum likelihood estimators are $\hat{\theta}_{MLE} = \text{MED}(n)$ and

$$\hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \text{MED}(n)|.$$

A $100(1-\alpha)\%$ confidence interval (CI) for λ is

$$\left(\frac{2\sum_{i=1}^n |Y_i - \text{MED}(n)|}{\chi^2_{2n-1,1-\frac{\alpha}{2}}}, \frac{2\sum_{i=1}^n |Y_i - \text{MED}(n)|}{\chi^2_{2n-1,\frac{\alpha}{2}}} \right),$$

and a $100(1-\alpha)\%$ CI for θ is

$$\left(\text{MED}(n) \pm \frac{z_{1-\alpha/2} \sum_{i=1}^n |Y_i - \text{MED}(n)|}{n\sqrt{n - z^2_{1-\alpha/2}}} \right)$$

where $\chi^2_{p,\alpha}$ and z_α are the α percentiles of the χ^2_p and standard normal distributions, respectively. See Patel, Kapadia and Owen (1976, p. 194).

A trimming rule is keep y_i if

$$y_i \in [\text{med}(n) \pm 10.0(1 + \frac{2.0}{n})\text{mad}(n)].$$

Note that $F(\theta + \lambda \log(1000)) = 0.9995 \approx F(\text{MED}(Y) + 10.0\text{MAD}(Y))$.

3.7 The Exponential Distribution

If Y has an exponential distribution, $Y \sim \text{EXP}(\lambda)$, then the pdf of Y is

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

where $\lambda > 0$ and the indicator $I(y \geq 0)$ is one if $y \geq 0$ and zero otherwise. The cdf of Y is

$$F(y) = 1 - \exp(-y/\lambda), \ y \geq 0.$$

The mgf $m(t) = 1/(1 - \lambda t)$, $t < 1/\lambda$ and the chf $c(t) = 1/(1 - i\lambda t)$.
$E(Y) = \lambda$,
and $\text{VAR}(Y) = \lambda^2$.

Since Y is gamma $G(\nu = 1, \lambda)$, $E(Y^r) = \lambda \Gamma(r + 1)$ for $r > -1$.
$\text{MED}(Y) = \log(2)\lambda$ and
$\text{MAD}(Y) \approx \lambda/2.0781$ since it can be shown that

$$\exp(\text{MAD}(Y)/\lambda) = 1 + \exp(-\text{MAD}(Y)/\lambda).$$

Hence $2.0781 \ \text{MAD}(Y) \approx \lambda$.
A robust estimator is $\hat{\lambda} = \text{MED}(n)/\log(2)$.

The classical estimator is $\hat{\lambda} = \overline{Y}_n$ and the $100(1 - \alpha)\%$ CI for $E(Y) = \lambda$ is

$$\left(\frac{2\sum_{i=1}^n Y_i}{\chi^2_{2n,1-\frac{\alpha}{2}}}, \frac{2\sum_{i=1}^n Y_i}{\chi^2_{2n,\frac{\alpha}{2}}} \right)$$

where $P(Y \leq \chi^2_{2n,\frac{\alpha}{2}}) = \alpha/2$ if Y is χ^2_{2n}. See Patel, Kapadia and Owen (1976, p. 188).

If all the $y_i \geq 0$, then the trimming rule is keep y_i if

$$0.0 \leq y_i \leq 9.0(1 + \frac{c_2}{n})\text{med}(n)$$

where $c_2 = 2.0$ seems to work well. Note that $P(Y \leq 9.0\text{MED}(Y)) \approx 0.998$.

3.8 The Two Parameter Exponential Distribution

If Y has a two parameter exponential distribution, $Y \sim \text{EXP}(\theta, \lambda)$, then the pdf of Y is

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

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

$$F(y) = 1 - \exp[-(y-\theta)/\lambda)], \; y \geq \theta.$$

This family is an asymmetric location-scale family.
The mgf $m(t) = \exp(t\theta)/(1 - \lambda t)$, $t < 1/\lambda$ and
the chf $c(t) = \exp(it\theta)/(1 - i\lambda t)$.
$E(Y) = \theta + \lambda$,
and $\text{VAR}(Y) = \lambda^2$.

$$\text{MED}(Y) = \theta + \lambda \log(2)$$

and

$$\text{MAD}(Y) \approx \lambda/2.0781.$$

Hence $\theta \approx \text{MED}(Y) - 2.0781 \log(2) \text{MAD}(Y)$. See Rousseeuw and Croux (1993) for similar results. Note that $2.0781 \log(2) \approx 1.44$.

Let $D_n = \sum_{i=1}^n (Y_i - Y_{(1)}) = n\hat{\lambda}$. Then for $n \geq 2$,

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

is a $100(1-\alpha)\%$ CI for λ, while

$$(Y_{(1)} - \hat{\lambda}[(\alpha)^{-1/(n-1)} - 1], Y_{(1)})$$

is a 100 $(1-\alpha)\%$ CI for θ.

If θ is known and $T_n = \sum_{i=1}^n (Y_i - \theta)$, then a $100(1-\alpha)\%$ CI for λ is

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

A trimming rule is keep y_i if

$$\text{med}(n) - 1.44(1.0 + \frac{c_4}{n})\text{mad}(n) \le y_i \le$$

$$\text{med}(n) - 1.44\text{mad}(n) + 9.0(1 + \frac{c_2}{n})\text{med}(n)$$

where $c_2 = 2.0$ and $c_4 = 2.0$ may be good choices.

To see that 2.0781 MAD$(Y) \approx \lambda$, note that

$$0.5 = \int_{\theta+\lambda\log(2)-\text{MAD}}^{\theta+\lambda\log(2)+\text{MAD}} \frac{1}{\lambda}\exp(-(y-\theta)/\lambda)dy$$

$$= 0.5[-e^{-\text{MAD}/\lambda} + e^{\text{MAD}/\lambda}]$$

assuming $\lambda\log(2) > \text{MAD}$. Plug in MAD $= \lambda/2.0781$ to get the result.

3.9 The Gamma Distribution

If Y has a gamma distribution, $Y \sim G(\nu, \lambda)$, then the pdf of Y is

$$f(y) = \frac{y^{\nu-1}e^{-y/\lambda}}{\lambda^\nu \Gamma(\nu)}$$

where ν, λ, and y are positive. The mgf of Y is

$$m(t) = \left(\frac{1/\lambda}{\frac{1}{\lambda} - t}\right)^\nu = \left(\frac{1}{1-\lambda t}\right)^\nu$$

for $t < 1/\lambda$. The chf

$$c(t) = \left(\frac{1}{1-i\lambda t}\right)^\nu.$$

$E(Y) = \nu\lambda$.
VAR$(Y) = \nu\lambda^2$.

$$E(Y^r) = \frac{\lambda^r\Gamma(r+\nu)}{\Gamma(\nu)} \text{ if } r > -\nu.$$

Chen and Rubin (1986) show that $\lambda(\nu - 1/3) < \text{MED}(Y) < \lambda\nu = E(Y)$. Empirically, for $\nu > 3/2$,

$$\text{MED}(Y) \approx \lambda(\nu - 1/3),$$

and

$$\mathrm{MAD}(Y) \approx \frac{\lambda\sqrt{\nu}}{1.483}.$$

This family is a scale family for fixed ν, so if Y is $G(\nu,\lambda)$ then cY is $G(\nu, c\lambda)$ for $c > 0$. If W is $\mathrm{EXP}(\lambda)$ then W is $G(1,\lambda)$. If W is χ^2_p, then W is $G(p/2, 2)$. If Y and W are independent and Y is $G(\nu,\lambda)$ and W is $G(\phi,\lambda)$, then $Y+W$ is $G(\nu+\phi,\lambda)$.

Some classical estimators are given next. Let

$$w = \log\left[\frac{\overline{y}_n}{\text{geometric mean}(n)}\right]$$

where geometric mean$(n) = (y_1 y_2 \ldots y_n)^{1/n} = \exp[\frac{1}{n}\sum_{i=1}^n \log(y_i)]$. Then Thom's estimator (Johnson and Kotz 1970a, p. 188) is

$$\hat{\nu} \approx \frac{0.25(1+\sqrt{1+4w/3}\ )}{w}.$$

Also

$$\hat{\nu}_{MLE} \approx \frac{0.5000876 + 0.1648852w - 0.0544274w^2}{w}$$

for $0 < w \le 0.5772$, and

$$\hat{\nu}_{MLE} \approx \frac{8.898919 + 9.059950w + 0.9775374w^2}{w(17.79728 + 11.968477w + w^2)}$$

for $0.5772 < w \le 17$. If $w > 17$ then estimation is much more difficult, but a rough approximation is $\hat{\nu} \approx 1/w$ for $w > 17$. See Bowman and Shenton (1988, p. 46) and Greenwood and Durand (1960). Finally, $\hat{\lambda} = \overline{y}_n/\hat{\nu}$. Notice that $\hat{\lambda}$ may not be very good if $\hat{\nu} < 1/17$. For some M–estimators, see Marazzi and Ruffieux (1996).

Several normal approximations are available. For large ν, $Y \approx N(\nu\lambda, \nu\lambda^2)$. The Wilson–Hilferty approximation says that for $\nu \ge 0.5$,

$$Y^{1/3} \approx N\left((\nu\lambda)^{1/3}(1-\frac{1}{9\nu}), (\nu\lambda)^{2/3}\frac{1}{9\nu}\right).$$

Hence if Y is $G(\nu,\lambda)$ and

$$\alpha = P[Y \le G_\alpha],$$

then

$$G_\alpha \approx \nu\lambda \left[z_\alpha \sqrt{\frac{1}{9\nu}} + 1 - \frac{1}{9\nu} \right]^3$$

where z_α is the standard normal percentile, $\alpha = \Phi(z_\alpha)$. Bowman and Shenton (1988, p. 101) include higher order terms.

Next we give some trimming rules. Assume each $y_i > 0$. Assume $\nu \geq 0.5$. Rule 1. Assume λ is known. Let $\hat{\nu} = (\text{med}(n)/\lambda) + (1/3)$. Keep y_i if $y_i \in [lo, hi]$ where

$$lo = \max(0, \hat{\nu}\lambda \; [-(3.5 + 2/n)\sqrt{\frac{1}{9\hat{\nu}}} + 1 - \frac{1}{9\hat{\nu}}]^3),$$

and

$$hi = \hat{\nu}\lambda \; [(3.5 + 2/n)\sqrt{\frac{1}{9\hat{\nu}}} + 1 - \frac{1}{9\hat{\nu}}]^3.$$

Rule 2. Assume ν is known. Let $\hat{\lambda} = \text{med}(n)/(\nu - (1/3))$. Keep y_i if $y_i \in [lo, hi]$ where

$$lo = \max(0, \nu\hat{\lambda} \; [-(3.5 + 2/n)\sqrt{\frac{1}{9\nu}} + 1 - \frac{1}{9\nu}]^3),$$

and

$$hi = \nu\hat{\lambda} \; \left[(3.5 + 2/n)\sqrt{\frac{1}{9\nu}} + 1 - \frac{1}{9\nu} \right]^3.$$

Rule 3. Let $d = \text{med}(n) - c \; \text{mad}(n)$. Keep y_i if

$$dI[d \geq 0] \leq y_i \leq \text{med}(n) + c \; \text{mad}(n)$$

where

$$c \in [9, 15].$$

3.10 The Half Cauchy Distribution

If Y has a half Cauchy distribution, $Y \sim \text{HC}(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{2}{\pi\sigma[1 + (\frac{y-\mu}{\sigma})^2]}$$

where $y \geq \mu$, μ is a real number and $\sigma > 0$. The cdf of Y is

$$F(y) = \frac{2}{\pi} \arctan(\frac{y - \mu}{\sigma})$$

for $y \geq \mu$ and is 0, otherwise. This distribution is a right skewed location-scale family.

MED$(Y) = \mu + \sigma$.

MAD$(Y) = 0.73205\sigma$.

3.11 The Half Logistic Distribution

If Y has a half logistic distribution, $Y \sim \text{HL}(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{2 \exp(-(y - \mu)/\sigma)}{\sigma[1 + \exp(-(y - \mu)/\sigma)]^2}$$

where $\sigma > 0$, $y \geq \mu$ and μ are real. The cdf of Y is

$$F(y) = \frac{\exp[(y - \mu)/\sigma] \ - \ 1}{1 + \exp[(y - \mu)/\sigma)]}$$

for $y \geq \mu$ and 0 otherwise. This family is a right skewed location–scale family.

MED$(Y) = \mu + \log(3)\sigma$.

MAD$(Y) = 0.67346\sigma$.

3.12 The Half Normal Distribution

If Y has a half normal distribution, $Y \sim \text{HN}(\mu, \sigma)$, then the pdf of Y is

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

where $\sigma > 0$ and $y \geq \mu$ and μ is real. Let $\Phi(y)$ denote the standard normal cdf. Then the cdf of Y is

$$F(y) = 2\Phi(\frac{y - \mu}{\sigma}) - 1$$

for $y > \mu$ and $F(y) = 0$, otherwise. This is an asymmetric location–scale family that has the same distribution as $\mu + \sigma|Z|$ where $Z \sim N(0, 1)$.

$E(Y) = \mu + \sigma\sqrt{2/\pi} \approx \mu + 0.797885\sigma.$

$\text{VAR}(Y) = \frac{\sigma^2(\pi-2)}{\pi} \approx 0.363380\sigma^2.$

Note that $Z^2 \sim \chi^2_1$. Hence the formula for the rth moment of the χ^2_1 random variable can be used to find the moments of Y.

$\text{MED}(Y) = \mu + 0.6745\sigma.$

$\text{MAD}(Y) = 0.3990916\sigma.$

Thus $\hat{\mu} \approx \text{MED}(n) - 1.6901\text{MAD}(n)$ and $\hat{\sigma} \approx 2.5057\text{MAD}(n)$.

Pewsey (2002) shows that classical inference for this distribution is simple. The MLE of (μ, σ^2) is

$$(\hat{\mu}, \hat{\sigma}^2) = (Y_{(1)}, \frac{1}{n}\sum_{i=1}^{n}(Y_i - Y_{(1)})^2).$$

A large sample $100(1-\alpha)\%$ confidence interval for σ^2 is

$$\left(\frac{n\hat{\sigma}^2}{\chi^2_{n-1}(1-\alpha/2)}, \frac{n\hat{\sigma}^2}{\chi^2_{n-1}(\alpha/2)}\right),$$

while a large sample $100(1-\alpha)\%$ CI for μ is

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

Let $T_n = \sum(Y_i - \mu)^2$. If μ is known, then a $100(1-\alpha)\%$ CI for σ^2 is

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

3.13 The Largest Extreme Value Distribution

If Y has a largest extreme value distribution (or extreme value distribution for the max, or Gumbel distribution), $Y \sim \text{LEV}(\theta, \sigma)$, then the pdf of Y is

$$f(y) = \frac{1}{\sigma}\exp(-(\frac{y-\theta}{\sigma}))\exp[-\exp(-(\frac{y-\theta}{\sigma}))]$$

where y and θ are real and $\sigma > 0$. (Then $-Y$ has the smallest extreme value distribution or the log–Weibull distribution, see Section 3.24.) The cdf of Y is

$$F(y) = \exp[-\exp(-(\frac{y-\theta}{\sigma}))].$$

This family is an asymmetric location–scale family with a mode at θ.
The mgf $m(t) = \exp(t\theta)\Gamma(1 - \sigma t)$ for $|t| < 1/\sigma$.
$E(Y) \approx \theta + 0.57721\sigma$, and
$\text{VAR}(Y) = \sigma^2\pi^2/6 \approx 1.64493\sigma^2$.

$$\text{MED}(Y) = \theta - \sigma \log(\log(2)) \approx \theta + 0.36651\sigma$$

and

$$\text{MAD}(Y) \approx 0.767049\sigma.$$

$W = \exp(-(Y - \theta)/\sigma) \sim \text{EXP}(1)$.

A trimming rule is keep y_i if

$$\text{med}(n) - 2.5\text{mad}(n) \leq y_i \leq \text{med}(n) + 7\text{mad}(n).$$

3.14 The Logistic Distribution

If Y has a logistic distribution, $Y \sim L(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{\exp\left(-(y - \mu)/\sigma\right)}{\sigma[1 + \exp\left(-(y - \mu)/\sigma\right)]^2}$$

where $\sigma > 0$ and y and μ are real. The cdf of Y is

$$F(y) = \frac{1}{1 + \exp\left(-(y - \mu)/\sigma\right)} = \frac{\exp\left((y - \mu)/\sigma\right)}{1 + \exp\left((y - \mu)/\sigma\right)}.$$

This family is a symmetric location–scale family.
The mgf of Y is $m(t) = \pi\sigma t e^{\mu t} \csc(\pi\sigma t)$ for $|t| < 1/\sigma$, and
the chf is $c(t) = \pi i \sigma t e^{i\mu t} \csc(\pi i \sigma t)$ where $\csc(t)$ is the cosecant of t.
$E(Y) = \mu$, and
$\text{MED}(Y) = \mu$.
$\text{VAR}(Y) = \sigma^2\pi^2/3$, and
$\text{MAD}(Y) = \log(3)\sigma \approx 1.0986\ \sigma$.
Hence $\sigma = \text{MAD}(Y)/\log(3)$.

The estimators $\hat{\mu} = \overline{Y}_n$ and $\hat{\sigma}^2 = 3S^2/\pi^2$ where $S^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$ are sometimes used. A trimming rule is keep y_i if

$$\text{med}(n) - 7.6(1 + \frac{c_2}{n})\text{mad}(n) \leq y_i \leq \text{med}(n) + 7.6(1 + \frac{c_2}{n})\text{mad}(n)$$

where c_2 is between 0.0 and 7.0. Note that if

$$q = F_{L(0,1)}(c) = \frac{e^c}{1+e^c} \quad \text{then} \quad c = \log(\frac{q}{1-q}).$$

Taking $q = .9995$ gives $c = \log(1999) \approx 7.6$. To see that $\text{MAD}(Y) = \log(3)\sigma$, note that $F(\mu + \log(3)\sigma) = 0.75$, while $F(\mu - \log(3)\sigma) = 0.25$ and $0.75 = \exp(\log(3))/(1 + \exp(\log(3)))$.

3.15 The Log-Cauchy Distribution

If Y has a log–Cauchy distribution, $Y \sim LC(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{1}{\pi \sigma y [1 + (\frac{\log(y)-\mu}{\sigma})^2]}$$

where $y > 0$, $\sigma > 0$ and μ is a real number. This family is a scale family with scale parameter $\tau = e^\mu$ if σ is known.

$W = \log(Y)$ has a Cauchy(μ, σ) distribution.

Robust estimators are $\hat{\mu} = \text{MED}(W_1, ..., W_n)$ and $\hat{\sigma} = \text{MAD}(W_1, ..., W_n)$.

3.16 The Log-Logistic Distribution

If Y has a log–logistic distribution, $Y \sim LL(\phi, \tau)$, then the pdf of Y is

$$f(y) = \frac{\phi\tau(\phi y)^{\tau-1}}{[1 + (\phi y)^\tau]^2}$$

where $y > 0$, $\phi > 0$ and $\tau > 0$. The cdf of Y is

$$F(y) = 1 - \frac{1}{1 + (\phi y)^\tau}$$

for $y > 0$. This family is a scale family with scale parameter ϕ^{-1} if τ is known.

$\text{MED}(Y) = 1/\phi$.

$W = \log(Y)$ has a logistic$(\mu = -\log(\phi), \sigma = 1/\tau)$ distribution. Hence $\phi = e^{-\mu}$ and $\tau = 1/\sigma$.

Robust estimators are $\hat{\tau} = \log(3)/\text{MAD}(W_1, ..., W_n)$ and
$\hat{\phi} = 1/\text{MED}(Y_1, ..., Y_n)$ since $\text{MED}(Y) = 1/\phi$.

3.17 The Lognormal Distribution

If Y has a lognormal distribution, $Y \sim \text{LN}(\mu, \sigma^2)$, then the pdf of Y is

$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(\log(y)-\mu)^2}{2\sigma^2}\right)$$

where $y > 0$ and $\sigma > 0$ and μ is real. The cdf of Y is

$$F(y) = \Phi\left(\frac{\log(y)-\mu}{\sigma}\right) \quad \text{for} \quad y > 0$$

where $\Phi(y)$ is the standard normal N(0,1) cdf. This family is a scale family with scale parameter $\tau = e^{\mu}$ if σ^2 is known.
$E(Y) = \exp(\mu + \sigma^2/2)$ and
$\text{VAR}(Y) = \exp(\sigma^2)(\exp(\sigma^2) - 1)\exp(2\mu)$.
For any r, $E(Y^r) = \exp(r\mu + r^2\sigma^2/2)$.
$\text{MED}(Y) = \exp(\mu)$ and
$\exp(\mu)[1 - \exp(-0.6744\sigma)] \leq \text{MAD}(Y) \leq \exp(\mu)[1 + \exp(0.6744\sigma)]$.

Inference for μ and σ is simple. Use the fact that $W_i = \log(Y_i) \sim N(\mu, \sigma^2)$ and then perform the corresponding normal based inference on the W_i. For example, a the classical $(1-\alpha)100\%$ CI for μ when σ is unknown is

$$(\overline{W}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_W}{\sqrt{n}}, \overline{W}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_W}{\sqrt{n}})$$

where

$$S_W = \frac{n}{n-1}\hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(W_i - \overline{W})^2},$$

and $P(t \leq t_{n-1,1-\frac{\alpha}{2}}) = 1 - \alpha/2$ when t is from a t distribution with $n-1$ degrees of freedom.

Robust estimators are

$$\hat{\mu} = \text{MED}(W_1, ..., W_n) \quad \text{and} \quad \hat{\sigma} = 1.483\text{MAD}(W_1, ..., W_n).$$

Assume all $y_i \geq 0$. Then a trimming rule is keep y_i if

$$\text{med}(n) - 5.2(1 + \frac{c_2}{n})\text{mad}(n) \leq w_i \leq \text{med}(n) + 5.2(1 + \frac{c_2}{n})\text{mad}(n)$$

where c_2 is between 0.0 and 7.0. Here $\text{med}(n)$ and $\text{mad}(n)$ are applied to $w_1, \ldots, w_n$ where $w_i = \log(y_i)$.

3.18 The Maxwell-Boltzmann Distribution

If Y has a Maxwell–Boltzmann distribution, $Y \sim MB(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{\sqrt{2}(y-\mu)^2 e^{\frac{-1}{2\sigma^2}(y-\mu)^2}}{\sigma^3\sqrt{\pi}}$$

where μ is real, $y \geq \mu$ and $\sigma > 0$. This is a location–scale family.

$$E(Y) = \mu + \sigma\sqrt{2}\frac{1}{\Gamma(3/2)}.$$

$$\text{VAR}(Y) = 2\sigma^2 \left[\frac{\Gamma(\frac{5}{2})}{\Gamma(3/2)} - \left(\frac{1}{\Gamma(3/2)} \right)^2 \right].$$

MED$(Y) = \mu + 1.5381722\sigma$ and MAD$(Y) = 0.460244\sigma$.
Note that $W = (Y-\mu)^2 \sim G(3/2, 2\sigma^2)$.

3.19 The Normal Distribution

If Y has a normal distribution (or Gaussian distribution), $Y \sim N(\mu, \sigma^2)$, then the pdf of Y is

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y-\mu)^2}{2\sigma^2} \right)$$

where $\sigma > 0$ and μ and y are real. Let $\Phi(y)$ denote the standard normal cdf. Recall that $\Phi(y) = 1 - \Phi(-y)$. The cdf $F(y)$ of Y does not have a closed form, but

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma} \right),$$

and

$$\Phi(y) \approx 0.5(1 + \sqrt{1 - \exp(-2y^2/\pi)}\ )$$

for $y \geq 0$. See Johnson and Kotz (1970a, p. 57).
The moment generating function is $m(t) = \exp(t\mu + t^2\sigma^2/2)$.
The characteristic function is $c(t) = \exp(it\mu - t^2\sigma^2/2)$.
$E(Y) = \mu$ and
$\text{VAR}(Y) = \sigma^2$.

$$E[|Y-\mu|^r] = \sigma^r\ \frac{2^{r/2}\Gamma((r+1)/2)}{\sqrt{\pi}} \quad \text{for} \ \ r > -1.$$

If $k \geq 2$ is an integer, then $E(Y^k) = (k-1)\sigma^2 E(Y^{k-2}) + \mu E(Y^{k-1})$.
MED$(Y) = \mu$ and

$$\text{MAD}(Y) = \Phi^{-1}(0.75)\sigma \approx 0.6745\sigma.$$

Hence $\sigma = [\Phi^{-1}(0.75)]^{-1}\text{MAD}(Y) \approx 1.483\text{MAD}(Y)$.
This family is a location–scale family which is symmetric about μ.

Suggested estimators are

$$\overline{Y}_n = \hat{\mu} = \frac{1}{n}\sum_{i=1}^n Y_i \text{ and } S^2 = S_Y^2 = \hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2.$$

The classical $(1-\alpha)100\%$ CI for μ when σ is unknown is

$$(\overline{Y}_n - t_{n-1,1-\frac{\alpha}{2}}\frac{S_Y}{\sqrt{n}}, \overline{Y}_n + t_{n-1,1-\frac{\alpha}{2}}\frac{S_Y}{\sqrt{n}})$$

where $P(t \leq t_{n-1,1-\frac{\alpha}{2}}) = 1 - \alpha/2$ when t is from a t distribution with $n-1$ degrees of freedom.

If $\alpha = \Phi(z_\alpha)$, then

$$z_\alpha \approx m - \frac{c_o + c_1 m + c_2 m^2}{1 + d_1 m + d_2 m^2 + d_3 m^3}$$

where

$$m = [-2\log(1-\alpha)]^{1/2},$$

$c_0 = 2.515517$, $c_1 = 0.802853$, $c_2 = 0.010328$, $d_1 = 1.432788$, $d_2 = 0.189269$, $d_3 = 0.001308$, and $0.5 \leq \alpha$. For $0 < \alpha < 0.5$,

$$z_\alpha = -z_{1-\alpha}.$$

See Kennedy and Gentle (1980, p. 95).

A trimming rule is keep y_i if

$$\text{med}(n) - 5.2(1 + \frac{c_2}{n})\text{mad}(n) \leq y_i \leq \text{med}(n) + 5.2(1 + \frac{c_2}{n})\text{mad}(n)$$

where c_2 is between 0.0 and 7.0. Using $c_2 = 4.0$ seems to be a good choice.

Note that

$$P(\mu - 3.5\sigma \leq Y \leq \mu + 3.5\sigma) = 0.9996.$$

To see that $\text{MAD}(Y) = \Phi^{-1}(0.75)\sigma$, note that $3/4 = F(\mu + \text{MAD})$ since Y is symmetric about μ. However,

$$F(y) = \Phi\left(\frac{y-\mu}{\sigma}\right)$$

and

$$\frac{3}{4} = \Phi\left(\frac{\mu + \Phi^{-1}(3/4)\sigma - \mu}{\sigma}\right).$$

So $\mu + \text{MAD} = \mu + \Phi^{-1}(3/4)\sigma$. Cancel μ from both sides to get the result.

3.20 The Pareto Distribution

If Y has a Pareto distribution, $Y \sim \text{PAR}(\sigma, \lambda)$, then the pdf of Y is

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

where $y \geq \sigma$, $\sigma > 0$, and $\lambda > 0$. The cdf of Y is $F(y) = 1 - (\sigma/y)^{1/\lambda}$ for $y > \sigma$.

This family is a scale family when λ is fixed. $E(Y) = \frac{\sigma}{1-\lambda}$ for $\lambda < 1$.

$$E(Y^r) = \frac{\sigma^r}{1-\lambda r} \quad \text{for } r < 1/\lambda.$$

$\text{MED}(Y) = \sigma 2^{\lambda}$.

$X = \log(Y/\sigma)$ is $\text{EXP}(\lambda)$ and $W = \log(Y)$ is $\text{EXP}(\theta = \log(\sigma), \lambda)$.

Let $D_n = \sum_{i=1}^n (W_i - W_{1:n}) = n\hat{\lambda}$ where $W_{(1)} = W_{1:n}$. For $n > 1$, a $100(1-\alpha)\%$ CI for θ is

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

Exponentiate the endpoints for a $100(1-\alpha)\%$ CI for σ. A $100(1-\alpha)\%$ CI for λ is

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

Let $\hat{\theta} = \text{MED}(W_1, ..., W_n) - 1.440\text{MAD}(W_1, ..., W_n)$. Then robust estimators are

$$\hat{\sigma} = e^{\hat{\theta}} \quad \text{and} \quad \hat{\lambda} = 2.0781\text{MAD}(W_1, ..., W_n).$$

A trimming rule is keep y_i if

$$\text{med}(n) - 1.44\text{mad}(n) \le w_i \le 10\text{med}(n) - 1.44\text{mad}(n)$$

where med(n) and mad(n) are applied to $w_1, \ldots, w_n$ with $w_i = \log(y_i)$.

3.21 The Poisson Distribution

If Y has a Poisson distribution, $Y \sim \text{POIS}(\theta)$, then the pmf of Y is

$$P(Y = y) = \frac{e^{-\theta}\theta^y}{y!}$$

for $y = 0, 1, \ldots,$ where $\theta > 0$. The mgf of Y is $m(t) = \exp(\theta(e^t - 1))$, and the chf of Y is $c(t) = \exp(\theta(e^{it} - 1))$.
$E(Y) = \theta$, and Chen and Rubin (1986) and Adell and Jodrá (2005) show that $-1 < \text{MED}(Y) - E(Y) < 1/3$.
$\text{VAR}(Y) = \theta$.

The classical estimator of θ is $\hat{\theta} = \overline{Y}_n$. Let $W = \sum_{i=1}^n Y_i$ and suppose that $W = w$ is observed. Let $P(T < \chi^2_d(\alpha)) = \alpha$ if $T \sim \chi^2_d$. Then an "exact" 100 $(1-\alpha)$% CI for θ is

$$\left(\frac{\chi^2_{2w}(\frac{\alpha}{2})}{2n}, \frac{\chi^2_{2w+2}(1 - \frac{\alpha}{2})}{2n} \right)$$

for $w \neq 0$ and

$$\left(0, \frac{\chi^2_2(1-\alpha)}{2n} \right)$$

for $w = 0$.

The approximations $Y \approx N(\theta, \theta)$ and $2\sqrt{Y} \approx N(2\sqrt{\theta}, 1)$ are sometimes used.

Suppose each y_i is a nonnegative integer. Then a trimming rule is keep y_i if $w_i = 2\sqrt{y_i}$ is kept when a normal trimming rule is applied to the $w_i's$. (This rule can be very bad if the normal approximation is not good.)

3.22 The Power Distribution

If Y has a power distribution, $Y \sim \text{POW}(\lambda)$, then the pdf of Y is

$$f(y) = \frac{1}{\lambda} y^{\frac{1}{\lambda} - 1},$$

where $\lambda > 0$ and $0 < y \leq 1$. The cdf of Y is $F(y) = y^{1/\lambda}$ for $0 < y \leq 1$.
MED$(Y) = (1/2)^\lambda$.

$W = -\log(Y)$ is EXP(λ).

Let $T_n = -\sum \log(Y_i)$. A $100(1-\alpha)\%$ CI for λ is

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

If all the $y_i \in [0, 1]$, then a cleaning rule is keep y_i if

$$0.0 \leq w_i \leq 9.0(1 + \frac{2}{n})\text{med}(n)$$

where med(n) is applied to $w_1, \ldots, w_n$ with $w_i = -\log(y_i)$. See Problem 3.7 for robust estimators.

3.23 The Rayleigh Distribution

If Y has a Rayleigh distribution, $Y \sim R(\mu, \sigma)$, then the pdf of Y is

$$f(y) = \frac{y-\mu}{\sigma^2} \exp\left[-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right]$$

where $\sigma > 0$, μ is real, and $y \geq \mu$. See Cohen and Whitten (1988, Ch. 10). This is an asymmetric location–scale family. The cdf of Y is

$$F(y) = 1 - \exp\left[-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right]$$

for $y \geq \mu$, and $F(y) = 0$, otherwise.
$E(Y) = \mu + \sigma\sqrt{\pi/2} \approx \mu + 1.253314\sigma$.
VAR$(Y) = \sigma^2(4-\pi)/2 \approx 0.429204\sigma^2$.
MED$(Y) = \mu + \sigma\sqrt{\log(4)} \approx \mu + 1.17741\sigma$.
Hence $\mu \approx$ MED$(Y) - 2.6255$MAD(Y) and $\sigma \approx 2.230$MAD(Y).
Let $\sigma D =$ MAD(Y). If $\mu = 0$, and $\sigma = 1$, then

$$0.5 = \exp[-0.5(\sqrt{\log(4)} - D)^2] - \exp[-0.5(\sqrt{\log(4)} + D)^2].$$

Hence $D \approx 0.448453$ and MAD$(Y) \approx 0.448453\sigma$.
It can be shown that $W = (Y-\mu)^2 \sim$ EXP$(2\sigma^2)$.

Other parameterizations for the Rayleigh distribution are possible. See Problem 3.9.

3.24 The Smallest Extreme Value Distribution

If Y has a smallest extreme value distribution (or log-Weibull distribution), $Y \sim SEV(\theta, \sigma)$, then the pdf of Y is

$$f(y) = \frac{1}{\sigma} \exp(\frac{y-\theta}{\sigma}) \exp[-\exp(\frac{y-\theta}{\sigma})]$$

where y and θ are real and $\sigma > 0$. The cdf of Y is

$$F(y) = 1 - \exp[-\exp(\frac{y-\theta}{\sigma})].$$

This family is an asymmetric location-scale family with a longer left tail than right.

$E(Y) \approx \theta - 0.57721\sigma$, and
VAR$(Y) = \sigma^2\pi^2/6 \approx 1.64493\sigma^2$.
MED$(Y) = \theta - \sigma \log(\log(2))$.
MAD$(Y) \approx 0.767049\sigma$.

If Y has a SEV(θ, σ) distribution, then $W = -Y$ has an LEV$(-\theta, \sigma)$ distribution.

3.25 The Student's t Distribution

If Y has a Student's t distribution, $Y \sim t_p$, then the pdf of Y is

$$f(y) = \frac{\Gamma(\frac{p+1}{2})}{(p\pi)^{1/2}\Gamma(p/2)}(1 + \frac{y^2}{p})^{-(\frac{p+1}{2})}$$

where p is a positive integer and y is real. This family is symmetric about 0. The t_1 distribution is the Cauchy$(0,1)$ distribution. If Z is $N(0,1)$ and is independent of $W \sim \chi^2_p$, then

$$\frac{Z}{(\frac{W}{p})^{1/2}}$$

is t_p.
$E(Y) = 0$ for $p \geq 2$.
MED$(Y) = 0$.

VAR(Y) $= p/(p-2)$ for $p \geq 3$, and
MAD(Y) $= t_{p,0.75}$ where $P(t_p \leq t_{p,0.75}) = 0.75$.

If $\alpha = P(t_p \leq t_{p,\alpha})$, then Cooke, Craven, and Clarke (1982, p. 84) suggest the approximation

$$t_{p,\alpha} \approx \sqrt{p[\exp(\frac{w_\alpha^2}{p}) - 1)]}$$

where

$$w_\alpha = \frac{z_\alpha(8p+3)}{8p+1},$$

z_α is the standard normal cutoff: $\alpha = \Phi(z_\alpha)$, and $0.5 \leq \alpha$. If $0 < \alpha < 0.5$, then

$$t_{p,\alpha} = -t_{p,1-\alpha}.$$

This approximation seems to get better as the degrees of freedom increase.

A trimming rule for $p \geq 3$ is keep y_i if $y_i \in [\pm 5.2(1 + 10/n)\text{mad}(n)]$.

3.26 The Truncated Extreme Value Distribution

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

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

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

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

for $y > 0$.
MED(Y) $= \log(1 + \lambda \log(2))$.

$W = e^Y - 1$ is EXP(λ).

Let $T_n = \sum(e^{Y_i} - 1)$. A $100(1-\alpha)\%$ CI for λ is

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

If all the $y_i > 0$, then a trimming rule is keep y_i if

$$0.0 \leq w_i \leq 9.0(1 + \frac{2}{n})\text{med}(n)$$

where med(n) is applied to $w_1, \ldots, w_n$ with $w_i = e^{y_i} - 1$. See Problem 3.8 for robust estimators.

3.27 The Uniform Distribution

If Y has a uniform distribution, $Y \sim U(\theta_1, \theta_2)$, then the pdf of Y is

$$f(y) = \frac{1}{\theta_2 - \theta_1} I(\theta_1 \leq y \leq \theta_2).$$

The cdf of Y is $F(y) = (y - \theta_1)/(\theta_2 - \theta_1)$ for $\theta_1 \leq y \leq \theta_2$.
This family is a location-scale family which is symmetric about $(\theta_1 + \theta_2)/2$.
By definition, $m(0) = c(0) = 1$. For $t \neq 0$, the mgf of Y is

$$m(t) = \frac{e^{t\theta_2} - e^{t\theta_1}}{(\theta_2 - \theta_1)t},$$

and the chf of Y is

$$c(t) = \frac{e^{it\theta_2} - e^{it\theta_1}}{(\theta_2 - \theta_1)it}.$$

$E(Y) = (\theta_1 + \theta_2)/2$, and
MED$(Y) = (\theta_1 + \theta_2)/2$.
VAR$(Y) = (\theta_2 - \theta_1)^2/12$, and
MAD$(Y) = (\theta_2 - \theta_1)/4$.
Note that $\theta_1 = \text{MED}(Y) - 2\text{MAD}(Y)$ and $\theta_2 = \text{MED}(Y) + 2\text{MAD}(Y)$.

Some classical estimators are $\hat{\theta}_1 = Y_{(1)}$ and $\hat{\theta}_2 = Y_{(n)}$. A trimming rule is keep y_i if

$$\text{med}(n) - 2.0(1 + \frac{c_2}{n})\text{mad}(n) \leq y_i \leq \text{med}(n) + 2.0(1 + \frac{c_2}{n})\text{mad}(n)$$

where c_2 is between 0.0 and 5.0. Replacing 2.0 by 2.00001 yields a rule for which the cleaned data will equal the actual data for large enough n (with probability increasing to one).

3.28 The Weibull Distribution

If Y has a Weibull distribution, $Y \sim W(\phi, \lambda)$, then the pdf of Y is

$$f(y) = \frac{\phi}{\lambda} y^{\phi-1} e^{-\frac{y^\phi}{\lambda}}$$

where $\lambda, y,$ and ϕ are all positive. For fixed ϕ, this is a scale family in $\sigma = \lambda^{1/\phi}$. The cdf of Y is $F(y) = 1 - \exp(-y^\phi/\lambda)$ for $y > 0$.
$E(Y) = \lambda^{1/\phi}\ \Gamma(1 + 1/\phi)$.
$\text{VAR}(Y) = \lambda^{2/\phi}\Gamma(1 + 2/\phi) \ - \ (E(Y))^2$.

$$E(Y^r) = \lambda^{r/\phi}\ \Gamma(1 + \frac{r}{\phi}) \quad \text{for} \;\; r > -\phi.$$

$\text{MED}(Y) = (\lambda \log(2))^{1/\phi}$. Note that

$$\lambda = \frac{(\text{MED}(Y))^\phi}{\log(2)}.$$

Since $W = Y^\phi$ is EXP(λ), if all the $y_i > 0$ and if ϕ is known, then a cleaning rule is keep y_i if

$$0.0 \le w_i \le 9.0(1 + \frac{2}{n})\text{med}(n)$$

where med(n) is applied to $w_1, \ldots, w_n$ with $w_i = y_i^\phi$.

$W = \log(Y)$ has a smallest extreme value SEV($\theta = \log(\lambda^{1/\phi}), \sigma = 1/\phi$) distribution.

See Olive (2006) and Problem 3.10c for robust estimators of ϕ and λ.

3.29 Complements

Many of the distribution results used in this chapter came from Johnson and Kotz (1970ab) and Patel, Kapadia and Owen (1976). Bickel and Doksum (2007), Castillo (1988), Cohen and Whitten (1988), Cramér (1946), DeGroot and Schervish (2001), Ferguson (1967), Hastings and Peacock (1975) Kennedy and Gentle (1980), Leemis and McQuestion (2008), Lehmann (1983), Meeker and Escobar (1998), Abuhassan and Olive (2008) and Olive (2008) also have useful results on distributions. Also see articles in Kotz and Johnson

(1982ab,1983ab, 1985ab, 1986, 1988ab) and Armitrage and Colton (1998a-f). Often an entire book is devoted to a single distribution, see for example, Bowman and Shenton (1988).

Many of the robust point estimators in this chapter are due to Olive (2006). These robust estimators are usually inefficient, but can be used as starting values for iterative procedures such as maximum likelihood and as a quick check for outliers. These estimators can also be used to create a robust fully efficient cross checking estimator.

If no outliers are present and the sample size is large, then the robust and classical methods should give similar estimates. If the estimates differ, then outliers may be present or the assumed distribution may be incorrect. Although a plot is the best way to check for univariate outliers, many users of statistics plug in data and then take the result from the computer without checking assumptions. If the software would print the robust estimates besides the classical estimates and warn that the assumptions might be invalid if the robust and classical estimates disagree, more users of statistics would use plots and other diagnostics to check model assumptions.

3.30 Problems

PROBLEMS WITH AN ASTERISK * ARE ESPECIALLY USEFUL.

3.1. Verify the formula for the cdf F for the following distributions.
a) Cauchy (μ, σ).
b) Double exponential (θ, λ).
c) Exponential (λ).
d) Logistic (μ, σ).
e) Pareto (σ, λ).
f) Power (λ).
g) Uniform (θ_1, θ_2).
h) Weibull $W(\phi, \lambda)$.

3.2*. Verify the formula for MED(Y) for the following distributions.
a) Exponential (λ).
b) Lognormal (μ, σ^2). (Hint: $\Phi(0) = 0.5$.)
c) Pareto (σ, λ).
d) Power (λ).

e) Uniform (θ_1, θ_2).
f) Weibull (ϕ, λ).

3.3*. Verify the formula for MAD(Y) for the following distributions. (Hint: Some of the formulas may need to be verified numerically. Find the cdf in the appropriate section of Chapter 3. Then find the population median MED(Y) $= M$. The following trick can be used except for part c). If the distribution is symmetric, find $U = y_{0.75}$. Then $D = \text{MAD}(Y) = U - M$.)
a) Cauchy (μ, σ).
b) Double exponential (θ, λ).
c) Exponential (λ).
d) Logistic (μ, σ).
e) Normal (μ, σ^2).
f) Uniform (θ_1, θ_2).

3.4. Verify the formula for the expected value $E(Y)$ for the following distributions.
a) Binomial (k, ρ).
b) Double exponential (θ, λ).
c) Exponential (λ).
d) gamma (ν, λ).
e) Logistic (μ, σ). (Hint from deCani and Stine (1986): Let $Y = [\mu + \sigma W]$ so $E(Y) = \mu + \sigma E(W)$ where $W \sim L(0, 1)$. Hence

$$E(W) = \int_{-\infty}^{\infty} y \frac{e^y}{[1 + e^y]^2} dy.$$

Use substitution with

$$u = \frac{e^y}{1 + e^y}.$$

Then

$$E(W^k) = \int_0^1 [\log(u) - \log(1 - u)]^k du.$$

Also use the fact that

$$\lim_{v \to 0} v \log(v) = 0$$

to show $E(W) = 0$.)
f) Lognormal (μ, σ^2).
g) Normal (μ, σ^2).

h) Pareto (σ, λ).
i) Poisson (θ).
j) Uniform (θ_1, θ_2).
k) Weibull (ϕ, λ).

3.5. Verify the formula for the variance VAR(Y) for the following distributions.
a) Binomial (k, ρ).
b) Double exponential (θ, λ).
c) Exponential (λ).
d) gamma (ν, λ).
e) Logistic (μ, σ). (Hint from deCani and Stine (1986): Let $Y = [\mu + \sigma X]$ so $V(Y) = \sigma^2 V(X) = \sigma^2 E(X^2)$ where $X \sim L(0,1)$. Hence

$$E(X^2) = \int_{-\infty}^{\infty} y^2 \frac{e^y}{[1+e^y]^2} dy.$$

Use substitution with

$$v = \frac{e^y}{1+e^y}.$$

Then

$$E(X^2) = \int_0^1 [\log(v) - \log(1-v)]^2 dv.$$

Let $w = \log(v) - \log(1-v)$ and $du = [\log(v) - \log(1-v)]dv$. Then

$$E(X^2) = \int_0^1 w du = uw|_0^1 - \int_0^1 u dw.$$

Now

$$uw|_0^1 = [v\log(v) + (1-v)\log(1-v)]\ w|_0^1 = 0$$

since

$$\lim_{v \to 0} v \log(v) = 0.$$

Now

$$-\int_0^1 u dw = -\int_0^1 \frac{\log(v)}{1-v} dv - \int_0^1 \frac{\log(1-v)}{v} dv = 2\pi^2/6 = \pi^2/3$$

using

$$\int_0^1 \frac{\log(v)}{1-v} dv = \int_0^1 \frac{\log(1-v)}{v} dv = -\pi^2/6.)$$

f) Lognormal (μ, σ^2).
g) Normal (μ, σ^2).
h) Pareto (σ, λ).
i) Poisson (θ).
j) Uniform (θ_1, θ_2).
k) Weibull (ϕ, λ).

3.6. Assume that Y is gamma (ν, λ). Let

$$\alpha = P[Y \le G_\alpha].$$

Using

$$Y^{1/3} \approx N((\nu\lambda)^{1/3}(1 - \frac{1}{9\nu}), (\nu\lambda)^{2/3}\frac{1}{9\nu}),$$

show that

$$G_\alpha \approx \nu\lambda[z_\alpha \sqrt{\frac{1}{9\nu}} + 1 - \frac{1}{9\nu}]^3$$

where z_α is the standard normal percentile, $\alpha = \Phi(z_\alpha)$.

3.7. Suppose that $Y_1, ..., Y_n$ are iid from a power (λ) distribution. Suggest a robust estimator for λ

a) based on Y_i and

b) based on $W_i = -\log(Y_i)$.

3.8. Suppose that $Y_1, ..., Y_n$ are iid from a truncated extreme value TEV(λ) distribution. Find a robust estimator for λ

a) based on Y_i and

b) based on $W_i = e^{Y_i} - 1$.

3.9. Other parameterizations for the Rayleigh distribution are possible. For example, take $\mu = 0$ and $\lambda = 2\sigma^2$. Then W is Rayleigh RAY(λ), if the pdf of W is

$$f(w) = \frac{2w}{\lambda} \exp(-w^2/\lambda)$$

where λ and w are both positive.
The cdf of W is $F(w) = 1 - \exp(-w^2/\lambda)$ for $w > 0$.
$E(W) = \lambda^{1/2}\ \Gamma(1 + 1/2)$.

$\mathrm{VAR}(W) = \lambda\Gamma(2) \ - \ (E(W))^2.$

$$E(W^r) = \lambda^{r/2}\ \Gamma(1 + \frac{r}{2}) \quad \text{for} \quad r > -2.$$

$\mathrm{MED}(W) = \sqrt{\lambda \log(2)}.$
W is RAY(λ) if W is Weibull $W(\lambda, 2)$. Thus $W^2 \sim$ EXP(λ). If all $w_i > 0$, then a trimming rule is keep w_i if $0 \le w_i \le 3.0(1 + 2/n)\mathrm{MED}(n)$.

a) Find the median MED(W).

b) Suggest a robust estimator for λ.

3.10. Suppose Y has a smallest extreme value distribution, $Y \sim SEV(\theta, \sigma)$. See Section 3.24.

a) Find MED(Y).

b) Find MAD(Y).

c) If X has a Weibull distribution, $X \sim W(\phi, \lambda)$, then $Y = \log(X)$ is SEV(θ, σ) with parameters

$$\theta = \log(\lambda^{\frac{1}{\phi}}) \ \text{ and } \ \sigma = 1/\phi.$$

Use the results of a) and b) to suggest estimators for ϕ and λ.

3.11. Suppose that Y has a half normal distribution, $Y \sim HN(\mu, \sigma)$.

a) Show that MED(Y) $= \mu + 0.6745\sigma$.

b) Show that MAD(Y) $= 0.3990916\sigma$ numerically.

3.12. Suppose that Y has a half Cauchy distribution, $Y \sim$ HC(μ, σ). See Section 3.10 for $F(y)$.

a) Find MED(Y).

b) Find MAD(Y) numerically.

3.13. If Y has a log–Cauchy distribution, $Y \sim LC(\mu, \sigma)$, then $W = \log(Y)$ has a Cauchy(μ, σ) distribution. Suggest robust estimators for μ and σ based on an iid sample $Y_1, ..., Y_n$.

3.14. Suppose Y has a half logistic distribution, $Y \sim$ HL(μ, σ). See Section 3.11 for $F(y)$. Find MED(Y).

3.15. Suppose Y has a log–logistic distribution, $Y \sim LL(\phi, \tau)$, then $W = \log(Y)$ has a logistic($\mu = -\log(\phi), \sigma = 1/\tau$) distribution. Hence $\phi = e^{-\mu}$ and $\tau = 1/\sigma$. See Kalbfleisch and Prentice (1980, p. 27-28).

a) Using $F(y) = 1 - \dfrac{1}{1 + (\phi y)^\tau}$ for $y > 0$, find MED(Y).

b) Suggest robust estimators for τ and ϕ.

3.16. If Y has a geometric distribution, $Y \sim geom(p)$, then the pmf of Y is $P(Y = y) = p(1-p)^y$ for $y = 0, 1, 2, ...$ and $0 \le p \le 1$. The cdf for Y is $F(y) = 1 - (1-p)^{\lfloor y+1 \rfloor}$ for $y \ge 0$ and $F(y) = 0$ for $y < 0$. Use the cdf to find an approximation for MED(Y).

3.17. Suppose Y has a Maxwell–Boltzmann distribution, $Y \sim MB(\mu, \sigma)$. Show that MED(Y) $= \mu + 1.5381722\sigma$ and MAD(Y) $= 0.460244\sigma$.

3.18 If Y is Fréchet (μ, σ, ϕ), then the cdf of Y is

$$F(y) = \exp\left[-\left(\frac{y-\mu}{\sigma}\right)^{-\phi}\right]$$

for $y \ge \mu$ and 0 otherwise where $\sigma, \phi > 0$. Find MED(Y).

3.19. If Y has an F distribution with degrees of freedom p and $n - p$, then

$$Y \stackrel{D}{=} \frac{\chi^2_p/p}{\chi^2_{n-p}/(n-p)} \approx \chi^2_p/p$$

if n is much larger than p ($n >> p$). Find an approximation for MED(Y) if $n >> p$.

3.20. If Y has a Topp–Leone distribution, $Y \sim TL(\phi)$, then the cdf of Y is $F(y) = (2y - y^2)^\phi$ for $\phi > 0$ and $0 < y < 1$. Find MED(Y).

3.21. If Y has a one sided stable distribution (with index 1/2), then the cdf

$$F(y) = 2\left[1 - \Phi\left(\sqrt{\frac{\sigma}{y}}\right)\right]$$

for $y > 0$ where $\Phi(x)$ is the cdf of a $N(0, 1)$ random variable. Find MED(Y).

3.22. If Y has a two parameter power distribution, then the pdf

$$f(y) = \frac{1}{\tau\lambda}\left(\frac{y}{\tau}\right)^{\frac{1}{\lambda}-1}$$

for $0 < y \leq \tau$ where $\lambda > 0$ and $\tau > 0$. Suggest robust estimators for τ and λ using $W = -\log(Y) \sim EXP(-\log(\tau), \lambda)$.