

Classical Location Model

To introduce the main ideas we consider first the simple location-dispersion model

$$y_i = \mu + \sigma \varepsilon_i \quad (1)$$

where μ and σ are the location and dispersion parameters and $\varepsilon_1, \dots, \varepsilon_n$ are independent and identically distributed (i.i.d.) with common distribution F . Suppose that we wish to estimate the location parameter μ where σ is an unknown (nuisance) dispersion parameter.

Let $\boldsymbol{\theta} = (\mu, \sigma)$ and $F_{\boldsymbol{\theta}}$ be the distribution of y_i under model (1).

If we believe that *all* the data points have been independently “generated” by the distribution F which is known and completely specified except for its location and dispersion parameters $\boldsymbol{\theta} = (\mu, \sigma)$, then we can use the *parametric family*

$$\mathcal{F} = \{F_{\boldsymbol{\theta}} : \mu \in \mathcal{R}, \sigma > 0\} \quad (2)$$

If in addition F has a density $f = F'$ we can estimate $\boldsymbol{\theta}$ by the method of *maximum likelihood* (ML): define $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})$ by the property of maximizing the *likelihood function*

$$L(\mu, \sigma) = \prod \frac{1}{\sigma} f\left(\frac{y_i - \mu}{\sigma}\right).$$

That is, $L(\mu, \sigma) \leq L(\hat{\mu}, \hat{\sigma})$, for all μ and for all $\sigma > 0$. If the density f is continuously differentiable, the ML estimates $\hat{\mu}$ and $\hat{\sigma}$ satisfy the ML equations

$$\frac{1}{n} \sum \left(-\frac{f'((y_i - \mu)/\sigma)}{f((y_i - \mu)/\sigma)} \right) = 0 \quad (3)$$

and

$$\frac{1}{n} \sum \left(-\frac{f'((y_i - \mu)/\sigma) ((y_i - \mu)/\sigma)}{f((y_i - \mu)/\sigma)} \right) = 1. \quad (4)$$

The function

$$\psi(y) = -\frac{f'(y)}{f(y)} \quad (5)$$

is called the *location score function* and the function

$$\chi(y) = -\frac{f'(y)y}{f(y)} \quad (6)$$

is called the *dispersion score function*. With this notation (3) and (4) can be rewritten as

$$\frac{1}{n} \sum \psi\left(\frac{y_i - \mu}{\sigma}\right) = 0 \quad (7)$$

$$\frac{1}{n} \sum \chi\left(\frac{y_i - \mu}{\sigma}\right) = 1 \quad (8)$$

Example 1 Let $F = \Phi$, the standard normal distribution function. Then $\psi(y) = y$, $\chi(y) = y^2$ and the ML estimates $\hat{\mu}$ and $\hat{\sigma}$ satisfy the equations

$$\begin{aligned}\frac{1}{n} \sum \left(\frac{y_i - \mu}{\sigma} \right) &= 0 \\ \frac{1}{n} \sum \left(\frac{y_i - \mu}{\sigma} \right)^2 &= 1.\end{aligned}$$

Therefore, $\hat{\mu} = \bar{y}$, the sample mean, and $\hat{\sigma} = \sqrt{\frac{1}{n} \sum (y_i - \bar{y})^2} = s$, the sample standard deviation.

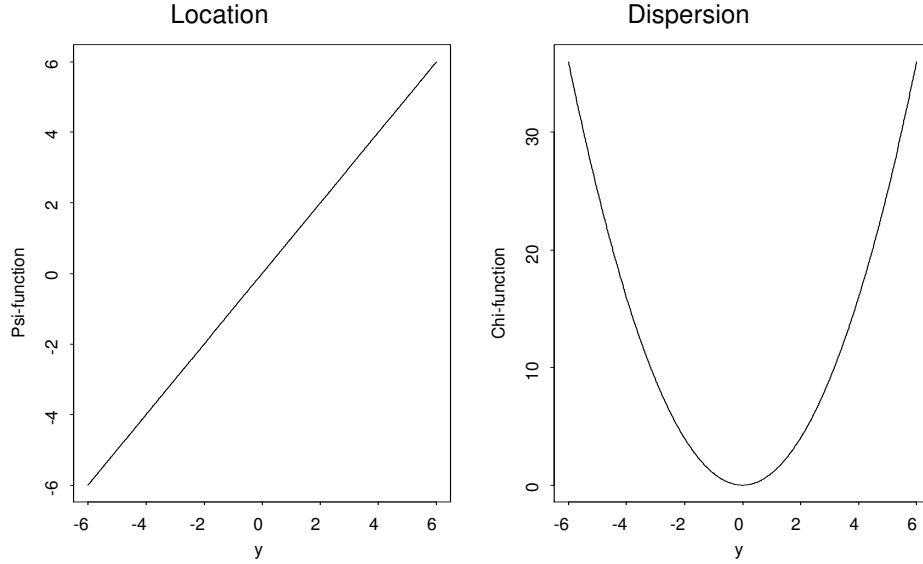


Figure 1: Location and dispersion score functions for the normal model.

Example 2 Let $f(y) = e^{-|y|}$, be the double exponential density function. Then $\psi(y) = \text{sign}(y)$, $\chi(y) = |y|$ (set $\psi(0) = 0$ and $\chi(0) = 0$) and so $\hat{\mu} = \text{Med}(y_i)$, the sample median, and $\hat{\sigma} = \frac{1}{n} \sum |y_i - \hat{\mu}|$, the sample median absolute deviation.

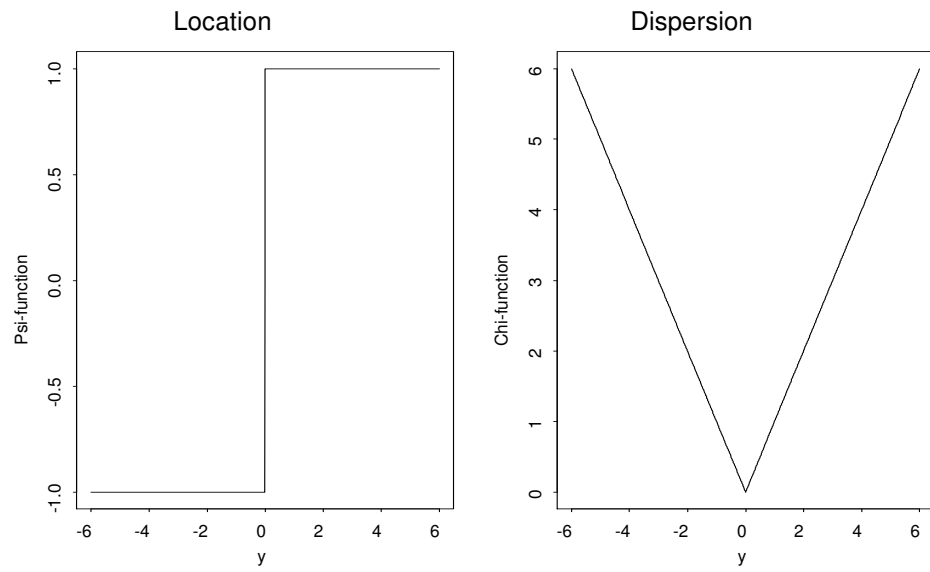


Figure 2: Location and dispersion score functions for the double exponential model.

Example 3 Let $F(y) = 0.5 + (1/\pi) \tan^{-1}(y)$ be the Cauchy distribution function. Then $f(y) = 1/(\pi(1+y^2))$, $\psi(y) = 2y/(1+y^2)$ and $\chi(y) = 2y^2/(1+y^2)$. The location and dispersion estimates cannot be expressed in closed form.

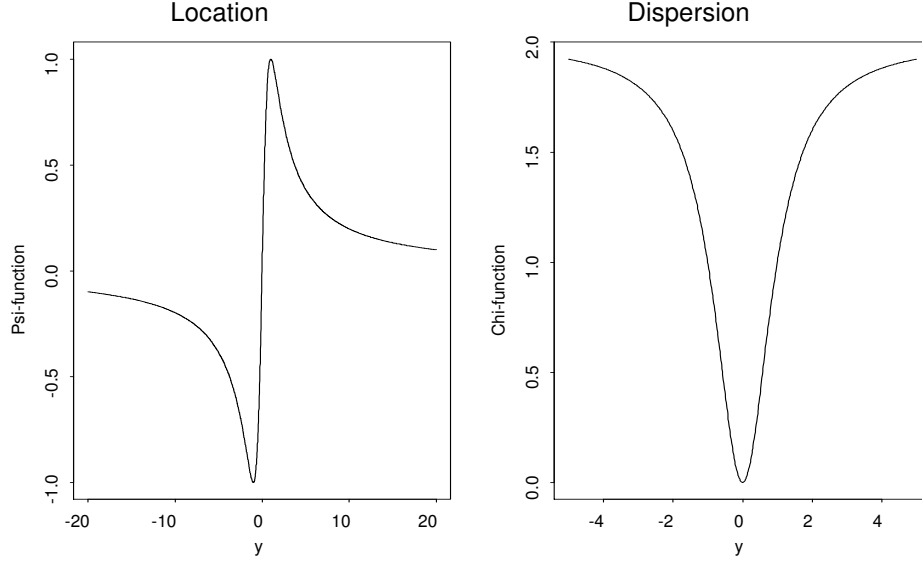


Figure 3: Location and dispersion score functions for the Cauchy model.

Note: It can be shown that under some regularity assumptions

$$\hat{\theta} \rightarrow \theta \quad \text{a.s.} \quad [F_{\theta}] \quad \text{and} \quad (9)$$

$$\sqrt{n}(\hat{\theta} - \theta) \rightarrow_d N(\mathbf{0}, \Sigma). \quad (10)$$

1 Robust Location Model

In many practical situations it is not reasonable to assume that all the data come from F which is then only partially specified. In these cases the set $\mathcal{F}$ of all possible distribution functions for y_i becomes much larger. The set of all possible partially specified distribution functions for y_i under model (1)) is called a *robustness family*.

For example one may have

$$F = (1 - \epsilon)\Phi + \epsilon G \quad (11)$$

where $0 < \epsilon < 1/2$, and G is unknown. According to this model, $(1 - \epsilon)100\%$ of the errors are normal with mean zero and constant variance and the remaining

$\epsilon 100\%$ may be outliers generated by an unknown and unspecified distribution function G .

The corresponding family

$$\mathcal{F}_\epsilon = \{H : H(y) = (1 - \epsilon)\Phi((y - \mu)/\sigma) + \epsilon G((y - \mu)/\sigma), \mu \in \mathcal{R}, \sigma > 0\} \quad (12)$$

of all the possible distribution functions of y_i under model (1) is called *normal contamination neighborhood*.

The family (12) constitutes a simple and intuitive way to model data that deviate from normality because of heavy tailness (a larger than expected fraction of relatively large errors) and/or the presence of outliers (a certain fraction of aberrant observations or gross errors). This model is indeed very flexible and the reader is encouraged to simulate data using distributions from (12) for different choices of ϵ and G .

The goal of robust methods is to estimate the parameters of the central distribution (the normal part of the mixture distribution H in (12)). Therefore one wishes to filter out the perturbation introduced by the contamination distribution G . It is important to understand at this point that robust methods do not attempt to estimate all the components of the mixture distribution H . In fact these methods are designed to work well without specific knowledge of ϵ and G . Since ϵ is usually small, there will normally be too few observations from which to base an attempt to estimate G . There could also be situations in which G changes from case to case in the sample.

Unless other robustness family is specified, it will be given by (12).

1.1 Location M-Estimates

The class of location M-estimates was first introduced by Huber in his influential 1964 paper. The location estimate is defined by an estimating equation, similar to usual Maximum Likelihood Estimates (MLE). The name M-estimates then is due to remark the connection with MLE estimation. The important difference is that the equation used to define M-estimates is not related to the data density and likelihood function. The separation between data assumptions and the estimation method is then the most important idea regarding these estimates.

When we assume that F belongs to a robustness family, the distribution function F and its density f are no longer fully specified and so the ML estimates (3) and (4) cannot be defined. However, equations (7) and (8) can still be used with the score functions ψ and χ chosen in some appropriate way.

Huber (1964) defined the class of location-dispersion *M-estimates* by (5) and (6) with ψ chosen from a large class

$$\mathcal{C}_\psi^* = \{\psi \text{ is odd and non-decreasing}\} \quad (13)$$

of *location score functions* and χ chosen from a large class

$$\mathcal{C}_\chi = \{\chi \text{ is even and non-decreasing on } [0, \infty)\} \quad (14)$$

of *scale score functions*. For technical reasons to be discussed later on, Huber replaced the right side of (8) by

$$b = E_{\Phi}\chi(y). \quad (15)$$

Given that in general χ and ψ are non-linear functions the location-dispersion M-estimates are calculated by iterative algorithms. Unfortunately, the definition of M-estimates as simultaneous solutions of (7) and (8) poses serious numerical and theoretical problems.

In general the following questions do not have an easy (let alone positive) answer:

- Does a solution to (7) and (8) exist?
- If a solution to (7) and (8) exists, is it unique?
- Does the computing algorithm used to simultaneously solve (7) and (8) converge?

Because of that, and other technical reasons to be discussed later on, estimates defined as the simultaneous solution of equations (7) and (8) are no longer used in practice. There is a simpler approach that splits the calculations of the dispersion and location estimates thus avoiding the numerical and theoretical problems of equations (7) and (8). This is the topic of the next section.

2 M-estimates with general dispersion

The class of M-estimates with general dispersion was first introduced by Martin, Yohai and Zamar (1989) in the regression setup to avoid the computational and the theoretical difficulties of simultaneous regression and dispersion estimates.

Location M-estimates with general dispersion are defined by

$$\hat{\mu} = \sup \left\{ t : \frac{1}{n} \sum \psi \left(\frac{y_i - t}{\hat{\sigma}} \right) > 0 \right\} \quad (16)$$

where $\hat{\sigma} = \hat{\sigma}(y_1, y_2, \dots, y_n)$ is a dispersion M-estimate. We will assume that the function ψ satisfies the following smoothness properties:

- (1) ψ is continuous and odd
- (2) ψ is non-decreasing, with $\psi(y) > 0$ for $y > 0$
- (3) $\hat{\sigma}_n$ is a smooth dispersion estimate (as described above).

When ψ is continuous then $\hat{\mu}_n$ satisfies the estimating equation

$$\frac{1}{n} \sum \psi \left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n} \right) = 0 \quad (17)$$

Moreover, it suffices that

$$\sup \left\{ \psi' \left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n} \right), i = 1, \dots, n \right\} > 0$$

for the solution to (17) to be unique.

An easier to compute dispersion estimate is the median absolute deviation about the median (MAD):

$$\begin{aligned} \hat{\sigma} &= \text{Med} \{ |y_1 - m|, |y_2 - m|, \dots, |y_n - m| \} / \Phi^{-1}(3/4) \\ &\approx 1.5 \text{Med} \{ |y_1 - m|, |y_2 - m|, \dots, |y_n - m| \} \end{aligned} \quad (18)$$

where Med stands for median. Because of its simple calculation and intuitive appeal the MAD is commonly used in practice. However, we cannot fully recommend the use of MAD due to its lack of smoothness. In particular, no one was able to prove yet the asymptotic normality of the MAD-based location M-estimate for asymmetric distributions. Notice that asymmetric distributions are the only type of relevant distributions in the robustness context.

The reader can easily verify that dispersion M-estimates have the following properties:

- S1. *Scale Equivariance*: $\hat{\sigma}(ay_1, ay_2, \dots, ay_n) = |a| \hat{\sigma}(y_1, y_2, \dots, y_n)$, for all a in $\mathcal{R}$.
- S2. *Location Invariance*: $\hat{\sigma}(y_1 + b, y_2 + b, \dots, y_n + b) = \hat{\sigma}(y_1, y_2, \dots, y_n)$, for all b in $\mathcal{R}$.

Location estimates $\hat{\mu} = \hat{\mu}(y_1, y_2, \dots, y_n)$ must satisfy the following equivariance properties:

- L1. *Scale Equivariance*: $\hat{\mu}(ay_1, ay_2, \dots, ay_n) = a\hat{\mu}(y_1, y_2, \dots, y_n)$, for all a in $\mathcal{R}$.
- L2. *Location Equivariance*: $\hat{\mu}(y_1 + b, y_2 + b, \dots, y_n + b) = \hat{\mu}(y_1, y_2, \dots, y_n) + b$, for all b in $\mathcal{R}$.

Computational Issues

As in the dispersion case, there are two main algorithms that can be used to compute smooth location estimates. The first one (re-weighting) is easy to code and doesn't require the calculation of the derivative of the score function. The second one (Newton-Raphson) converges faster and therefore is usually preferred.

The Newton-Raphson algorithm can also be used to find a solution to the equation $g(t) = 0$. This method consists of solving the approximate linear equation

$$g(t^{(k)}) + g'(t^{(k)})(t - t^{(k)}) = 0, \quad (19)$$

where $t^{(k)}$ is the current “approximate” solution. Solving (19) gives

$$t = t^{(k)} - \frac{g(t^{(k)})}{g'(t^{(k)})}. \quad (20)$$

Re-weighting Let

$$a = \lim_{y \rightarrow 0} \frac{\psi(y)}{y}$$

and define

$$w(x) = \begin{cases} \psi(x)/x & x \neq 0 \\ a & x = 0 \end{cases} \quad (21)$$

From (17) and (21) we can write (provided $y_i - \hat{\mu}_n \neq 0$)

$$\sum w\left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n}\right) \left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n}\right) = 0,$$

or equivalently

$$\hat{\mu}_n = \frac{\sum w\left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n}\right) y_i}{\sum w\left(\frac{y_i - \hat{\mu}_n}{\hat{\sigma}_n}\right)}. \quad (22)$$

Equation (22) suggests the following computing algorithm for $\hat{\mu}_n$:

Step 1: Set $\hat{\mu}_n^{(0)} = \text{Median}(y_i)$.

Step 2: Let

$$w_i^{(m)} = w\left(\frac{y_i - \hat{\mu}_n^{(m)}}{\hat{\sigma}_n}\right), \quad i = 1, \dots, n.$$

Step 3: Set

$$\hat{\mu}_n^{(m+1)} = \frac{\sum y_i w_i^{(m)}}{\sum w_i^{(m)}}$$

Step 4: Stop when $\left|\hat{\mu}_n^{(m+1)} - \hat{\mu}_n^{(m)}\right|/\hat{\sigma}_n \leq \delta$.

Newton-Raphson In the location case we have

$$g(m) = \frac{1}{n} \sum \psi\left(\frac{y_i - m}{\hat{\sigma}_n}\right) \quad \text{and} \quad g'(m) = -\frac{1}{n\hat{\sigma}_n} \sum \psi'\left(\frac{y_i - m}{\hat{\sigma}_n}\right)$$

and the algorithms follows the following steps:

Step 1: Set $\hat{\mu}_n^{(0)} = \text{Median}(y_i)$.

Step 2: Set

$$\hat{\mu}^{(k+1)} = \hat{\mu}^{(k)} - \frac{\sum \psi\left(\frac{y_i - \hat{\mu}^{(k)}}{\hat{\sigma}_n}\right)}{\sum \psi'\left(\frac{y_i - \hat{\mu}^{(k)}}{\hat{\sigma}_n}\right)} \hat{\sigma}_n.$$

Step 3: Stop when $\left|\hat{\mu}_n^{(m+1)} - \hat{\mu}_n^{(m)}\right| / \hat{\sigma}_n \leq \delta$.

Asymptotic Properties of Location M-Estimates

We will assume first that H is symmetric and show that the asymptotic distribution of the location M-estimate $\hat{\mu}$ is given by (16). When H is asymmetric the asymptotic normality still holds (under certain regularity conditions) but the derivation of the asymptotic distribution and the formula for the asymptotic variance become more involved.

Let

$$\lambda_H(t) = E_H \left\{ -\psi \left(\frac{y-t}{\sigma(H)} \right) \right\} \quad , \quad \sigma_H^2(t) = \text{Var}_H \left\{ \psi \left(\frac{y-t}{\sigma(H)} \right) \right\}$$

and

$$\lambda_n(t) = -\frac{1}{n} \sum \psi \left(\frac{y_i - t}{\hat{\sigma}} \right). \quad (23)$$

We will assume that the dispersion estimate $\hat{\sigma}$ is “regular”, that is, that $\hat{\sigma}$ satisfies the following:

- A.1 $\hat{\sigma}$ is consistent, that is $\hat{\sigma} \rightarrow \sigma(H)$ a.s. $[H]$.
- A.2 $\sqrt{n}(\hat{\sigma} - \sigma(H))$ is bounded in probability, that is, $\sqrt{n}(\hat{\sigma} - \sigma(H)) = O_p(1)$.

We will assume that the location score function ψ is “regular”, that is,

- A.3 $\sigma_H^2(t)$ is continuous on a neighborhood of μ and $0 < \sigma_H^2(\mu) < \infty$.
- A.4 $\lambda_H(\mu) = 0$, $\lambda_H(t)$ is continuously differentiable on a neighborhood of μ and $\lambda'_H(\mu) > 0$.

Theorem 1: *Suppose that A.1-A.4 hold and that H is symmetric. Then*

$$\hat{\mu} \rightarrow \mu \quad \text{a.s.} \quad [H] \quad (24)$$

and

$$\sqrt{n}(\hat{\mu} - \mu) \rightarrow_d N(0, \text{AV}(\psi, H)), \quad (25)$$

where

$$\text{AV}(\psi, H) = \frac{\mathbb{E}_H \left\{ \psi^2 \left(\frac{y-\mu}{\sigma(H)} \right) \right\}}{(\lambda'_H(\mu))^2}. \quad (26)$$

Proof: For simplicity we will assume that ψ is continuously differentiable.

First we prove equation (24). It is not difficult to see that

$$\{\hat{\mu} < t\} \subseteq \{\lambda_n(t) \geq 0\} \subseteq \{\hat{\mu} \leq t\} \quad (27)$$

By Lemma A in Serffling (1980), *page* 253, we have that

$$\lambda_n(t) \rightarrow \lambda_H(t) \quad \text{a.s.} \quad [H]. \quad (28)$$

Let $\epsilon > 0$ be given. By (28) and A.4 there exists n_0 such that, outside a set of probability zero,

$$\lambda_n(\mu + \epsilon) > 0, \quad \text{for all } n \geq n_0 \quad (29)$$

and

$$\lambda_n(\mu - \epsilon) < 0, \quad \text{for all } n \geq n_0. \quad (30)$$

By (29) and the second part of (27) we have that outside a set of probability zero

$$\hat{\mu} \leq \mu + \epsilon \quad \text{for all } n \geq n_0.$$

Finally, by (30) and the first part of (27) we have that outside a set of probability zero [notice that from the first part of (27) $\lambda_n(\mu - \epsilon) < 0$ implies $\hat{\mu} \geq \mu - \epsilon$]

$$\hat{\mu} \geq \mu - \epsilon \quad \text{for all } n \geq n_0.$$

proving (24).

To prove equation (25) notice that by continuity of ψ and ψ' we can write

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum \psi \left(\frac{y_i - \hat{\mu}}{\hat{\sigma}} \right) = \frac{1}{\sqrt{n}} \sum \psi \left(\frac{y_i - \mu}{\sigma(H)} \right) - \\ &\quad \frac{1}{n\tilde{\sigma}} \sum \psi' \left(\frac{y_i - \tilde{\mu}}{\tilde{\sigma}} \right) [\sqrt{n}(\hat{\mu} - \mu)] - \\ &\quad \frac{1}{n\tilde{\sigma}} \sum \psi' \left(\frac{y_i - \tilde{\mu}}{\tilde{\sigma}} \right) \left(\frac{y_i - \tilde{\mu}}{\tilde{\sigma}} \right) [\sqrt{n}(\hat{\sigma} - \sigma(H))] \end{aligned}$$

Since ψ' and the error density f are even

$$\mathbb{E}_F \left\{ \psi' \left(\frac{y - \mu}{\sigma(H)} \right) \left(\frac{y - \mu}{\sigma(H)} \right) \right\} = \int \psi' \left(\frac{y - \mu}{\sigma(H)} \right) \left(\frac{y - \mu}{\sigma(H)} \right) f \left(\frac{y - \mu}{\sigma(H)} \right) dy = 0$$

and so (see Serffling 1980, *page* 253)

$$\frac{1}{n\tilde{\sigma}} \sum \psi' \left(\frac{y_i - \tilde{\mu}}{\tilde{\sigma}} \right) \left(\frac{y_i - \tilde{\mu}}{\tilde{\sigma}} \right) \rightarrow_p 0, \quad \text{as } n \rightarrow \infty.$$

Moreover, since $\sqrt{n}(\hat{\sigma} - \sigma(H))$ is bounded in probability

$$\frac{1}{n\sigma} \sum \psi' \left(\frac{y_i - \mu}{\sigma(H)} \right) \left(\frac{y_i - \mu}{\sigma(H)} \right) [\sqrt{n}(\hat{\sigma} - \sigma(H))] \rightarrow_p 0, \quad \text{as } n \rightarrow \infty.$$

Therefore,

$$0 = \frac{1}{\sqrt{n}} \sum \psi \left(\frac{y_i - \mu}{\sigma(H)} \right) - \frac{1}{n\sigma} \sum \psi' \left(\frac{y_i - \mu}{\sigma(H)} \right) [\sqrt{n}(\hat{\mu} - \mu)] + o_p(1).$$

Finally, by the Central Limit and Slutsky's Theorems

$$\sqrt{n}(\hat{\mu} - \mu) = \tilde{\sigma} \frac{\frac{1}{\sqrt{n}} \sum \psi \left(\frac{y_i - \mu}{\sigma(H)} \right)}{\frac{1}{n} \sum \psi' \left(\frac{y_i - \mu}{\sigma(H)} \right)} \rightarrow N(0, \text{AV}(\psi, H)),$$

which is the desired result.

ML Estimates (particular case)

We can assume without loss of generality that $\mu = 0$ and $\sigma(F) = 1$. Since in this case

$$\psi_F(y) = -\frac{f'(y)}{f(y)},$$

we have

$$\mathbb{E}_F \{ \psi_F^2(y) \} = \int \psi_F^2(y) f(y) dy = \int \left(-\frac{f'(y)}{f(y)} \right)^2 f(y) dy = I_\mu(F)$$

and

$$\begin{aligned} \lambda'_F(0) &= \int \psi'_F(y) f(y) dy = - \int \frac{f''(y)f(y) + f'(y)^2}{f(y)^2} f(y) dy \\ &= - \int f''(y) dy - I_\mu(F) = -I_\mu(F) \end{aligned}$$

because $-\int f''(y) dy = 0$. Therefore,

$$\text{AV}(\psi_F, F) = \frac{I_\mu(F)}{(-I_\mu(F))^2} = \frac{1}{I_\mu(F)}.$$

Relative Efficiency

The asymptotic relative efficiency, under H , of two asymptotically normal point estimates T_1 and T_2 with asymptotic variances $AV(T_1, H)$ and $AV(T_2, H)$ is defined as

$$EFF(T_1, T_2, H) = \frac{AV(T_2, H)}{AV(T_1, H)}. \quad (31)$$

Asymptotic confidence intervals of level α constructed using these estimates have lengths

$$L_i = z_\alpha \frac{v_i}{\sqrt{n_i}}, \quad i = 1, 2 \quad (32)$$

where $z_\alpha = \Phi^{-1}(1 - \alpha/2)$, $v_i = \sqrt{AV(T_i, H)}$, and n_i is the corresponding sample size ($i = 1, 2$). Therefore the sample sizes n_1 and n_2 needed to obtain confidence intervals of the same length are

$$n_2 = \left(\frac{v_2^2}{v_1^2} \right) n_1 = EFF(T_1, T_2, H) n_1. \quad (33)$$

For example, if $EFF(T_1, T_2, H) = 1.3$ then T_2 requires 30% more observations to produce a confidence interval of the same length and level as T_1 .

Using (26) for the cases $\psi(y) = y$ (sample mean) and $\psi(y) = \text{sign}(y)$ (sample median) we have

$$AV(\text{Mean}, H) = \sigma_H^2 \quad \text{and} \quad AV(\text{Median}, H) = \frac{1}{4 h^2(0)}.$$

It can be shown that (26) holds for the median even though its score function ψ has a jump-discontinuity at 0. Also notice, by Exercise () we can assume without loss of generality that $\mu = 0$. Therefore,

$$EFF(\text{Median}, \text{Mean}, H) = \frac{AV(\text{Mean}, H)}{AV(\text{Median}, H)} = 4 \sigma_H^2 h^2(0).$$

Some numerical results for the Student's t-distribution with several degrees of freedom are given on Table 1.

Degrees of Freedom	< 3	3	4	5	10	20	100	∞
$AV(\text{Mean})/AV(\text{Median})$	∞	1.64	1.13	0.96	0.76	0.69	0.65	0.63

Table 1: Approximated relative asymptotic efficiency of the median relative to the mean, for Student t-distributions.

Problems

1. Show that (27) and (28) hold in the case of Examples 1 and 2.

2. Let $g(\mu, \sigma) = \log(\sigma) - E_F \{ \log f((y_i - \mu)/\sigma) \}$. Suppose that $g(\mu, \sigma)$ is continuous and that $g(\mu, \sigma) > g(\mu_0, \sigma_0)$ for all $\mu \neq \mu_0$ and $\sigma \neq \sigma_0$. Show that if

$$\lim_{\theta \rightarrow \infty} g(\mu, \sigma) > g(\mu_0, \sigma_0)$$

then (9) holds. Find additional sufficient conditions for (10) also to hold in this case.

3. Show that the sample mean and the sample median are location estimates, that is, satisfy the equivariance conditions L1 and L2.
4. Show that the sample standard deviation and the MAD are dispersion estimates, that is, satisfy the equivariance-invariance conditions S1 and S2.
5. Show that $\hat{\mu}$ defined by (16) satisfies conditions L1 and L2 above, for any location score function ψ in (13).
6. Verify (27) holds. Verify that the assumptions of Lemma A (in Serfling(1980), pg. 253) are satisfied and so (28) holds.
7. Show that (24) also holds in probability (without using the fact that a.s. convergence implies convergence in probability).
8. Show that the asymptotic variance formula (26) doesn't depend on the value of the "true" location μ and therefore μ can be taken equal to 0, for simplicity.
9. Calculate the relative asymptotic efficiency of the median relative to the mean for normal-normal contaminated distributions $H = 0.90N(0, 1) + 0.10N(0, \sigma^2)$, for a range of values of σ^2 .

10. Let

$$\psi_c(y) = \begin{cases} y/c & \text{if } |y| < c \\ \text{sign}(y) & \text{if } |y| \geq c \end{cases} \quad (34)$$

be Huber's location score function. Determine the value of c_0 such that the corresponding location M-estimate is 95% efficient (compared to the mean) for normal errors.

11. Compare the relative efficiency of the *mean* and the 95% efficient Huber's estimate found in the previous problem for normal-normal contaminated distributions $H = 0.90N(0, 1) + 0.10N(0, \sigma^2)$, for a range of values of σ^2 . Assume that the dispersion estimate is $\hat{\sigma}_n = \text{MAD}(y_i)$.
12. Compare the relative efficiency of the *median* and the 95% efficient Huber's estimate found in the previous problem for normal-normal contaminated distributions $H = 0.90N(0, 1) + 0.10N(0, \sigma^2)$, for a range of values of σ^2 . Assume that the dispersion estimate is $\hat{\sigma}_n = \text{MAD}(y_i)$.

13. Show that (under suitable conditions)

$$AV(\psi, H) \geq I_{\mu}^{-1}(H).$$