Robustness - Module 2

Ruben Zamar
Dept. of Stats, UBC

September 20, 2015

Introduction

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

$$y_i = \mu + \sigma\varepsilon_i$$

- $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. random variables with common **symmetric** distribution F_0.

- Wish to estimate μ.

- σ is an unknown nuisance parameter.

Dispersion-Location Model

- The parametric model:

$$F_{\mu,\sigma}(y) = F_0\left(\frac{y-\mu}{\sigma}\right)$$

- The likelihood function

$$L(\mu,\sigma) = \prod \frac{1}{\sigma} f_0\left(\frac{y_i-\mu}{\sigma}\right),$$

- The log likelihood function

$$l(\mu,\sigma) = \log\left[L(\mu,\sigma)\right] = -n\log(\sigma) + \sum \log\left[f_0\left(\frac{y_i-\mu}{\sigma}\right)\right]$$

ML-Estimate

- $(\widehat{\mu}, \widehat{\sigma})$ minimizes the log-likelihood

$$l(\mu, \sigma) \leq l(\widehat{\mu}, \widehat{\sigma}), \quad \text{for all } \mu \text{ and } \sigma > 0.$$

- The ML estimating equations:

$$\frac{1}{n} \sum \left(-\frac{f_0' \left((y_i - \mu) / \sigma \right)}{f_0 \left((y_i - \mu) / \sigma \right)} \right) = 0$$

$$\frac{1}{n} \sum \left(-\frac{f_0' \left[(y_i - \mu) / \sigma \right] \left((y_i - \mu) / \sigma \right)}{f_0 \left[(y_i - \mu) / \sigma \right]} \right) = 1$$

ML-Estimate (continued)

- The score functions:

$$\psi(y) = -\frac{f_0'(y)}{f_0(y)} \qquad \text{(location score)}$$

$$\chi(y) = -\frac{f_0'(y)\,y}{f_0(y)} = \psi(y)y \qquad \text{(dispersion score)}$$

- The ML estimating equations:

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

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

Example (normal case)

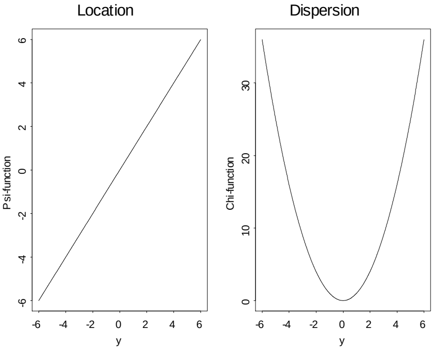
$$f_0(y) = \varphi(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

$$\psi(y) = -\frac{d}{dy} \log \varphi(y)$$

$$= -\frac{d}{dy}\left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}y^2\right] = y$$

$$\chi(y) = \psi(y)y = y^2$$

Normal Score Functions



$\hat{\mu}$ and $\hat{\sigma}$ satisfy the equations:

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

$$\frac{1}{n}\sum \chi\left(\frac{y_i-\mu}{\sigma}\right) = \frac{1}{n}\sum \left(\frac{y_i-\mu}{\sigma}\right)^2 = 1.$$

Hence

$$\widehat{\mu} = \bar{y}, \qquad \text{the sample mean,}$$

$$\widehat{\sigma} = \sqrt{\frac{1}{n}\sum (y_i-\bar{y})^2} = s, \qquad \text{the sample standard deviation.}$$

Example: the Double Exponential Model

$$f_0(y) = e^{|-y|}$$

$$\log\left[f_0(y)\right] = -\left|y\right|$$

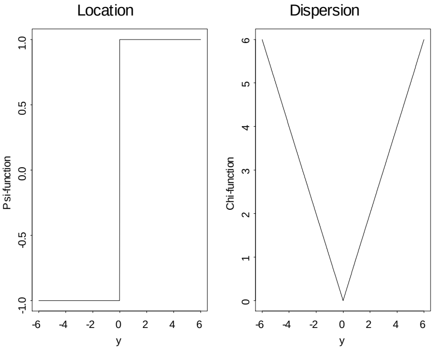
$$\psi(0) = \chi(0) = 0$$

and for $y \neq 0$,

$$\psi(y) = \frac{d}{dy}\left|y\right| = \operatorname{sign}(y),$$

$$\chi(y) = \operatorname{sign}(y)y = \left|y\right|$$

Double Exponential Score Functions



The Estimating Equations

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

$$\Rightarrow \quad \hat{\mu}=Med(y_i) \quad \text{(the median)}$$

$$\frac{1}{n}\sum\left|\frac{y_i-\widehat{\mu}}{\sigma}\right| = 1$$

$$\Rightarrow \quad \widehat{\sigma}=\frac{1}{n}\sum\left|y_i-\widehat{\mu}\right| \quad \text{(median absolute deviation)}$$

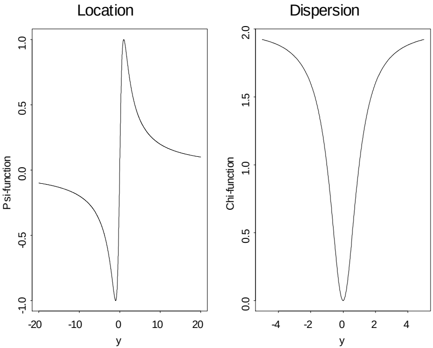
Example: The Cauchy Model

$$f_0(y) = 1/\left(\pi(1+y^2)\right),$$

$$\log\left(f_0(y)\right) = c - \log\left[(1+y^2)\right]$$

$$\psi(y) = 2y/(1+y^2) \;\; \text{and} \;\; \chi(y) = 2y^2/(1+y).$$

The Cauchy Score Functions



Asymptotic Properties of ML-Estimates

Under mild regularity conditions:

- Consistent:

$$\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} \rightarrow \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \quad \text{a.s.}$$

- Asymptotically normal

$$\sqrt{n}\left[\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma} \end{pmatrix} - \begin{pmatrix} \mu \\ \sigma \end{pmatrix}\right] \rightarrow N\left(\mathbf{0},\Sigma\right)$$

- Asymptotically efficient

$$\Sigma = \left[I\left(\mu,\sigma\right)\right]^{-1},$$

$$I\left(\mu,\sigma\right) = \text{Fisher information matrix}$$

$$I\left(\mu,\sigma\right) = \frac{1}{\sigma^2}\begin{pmatrix} E\left[\psi'\left(\frac{y-\mu}{\sigma}\right)\right] & 0 \\ 0 & E\left[\chi'\left(\frac{y-\mu}{\sigma}\right)\left(\frac{y-\mu}{\sigma}\right)\right] \end{pmatrix}$$

Robust Location-Dispersion Model

- In some applications it is not reasonable to assume that all the observations come from $F_{\mu,\sigma}(y)$
- Contamination neighborhood:

$$F(y) = (1-\epsilon)F_0\left(\frac{y-\mu}{\sigma}\right) + \epsilon G(y)$$

$$0 < \epsilon < 1/2, \qquad G \text{ unknown}$$

- $(1-\epsilon)100\%$ of the observations come from $F_{\mu,\sigma}(y)$
- The remaining $\epsilon 100\%$ observations come from an unknown (contamination) distribution $G(y)$

The Robustness Paradigm

- The goal of robust estimates is to estimate **the parameters of the central distribution**, $F_{\mu,\sigma}(y)$
- Since ϵ is usually small, there are very few observations coming from G
- Robust methods do not attempt to estimate *all* the components of the mixture distribution F.

Location-Dispersion M-Estimates

- Introduced by Huber in his influential 1964 paper.
- Similar to ML-estimates, M-estimates satisfy estimating equations.
- The key difference: the estimating equation for the M-estimate is not related to the data density and likelihood function.
- Separation between "data assumptions" and "estimation method": the main idea behind M-estimates.

M-Estimates Estimating Equations

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

$$\frac{1}{n}\sum \chi\left(\frac{y_i-\mu}{\sigma}\right)=E_{F_0}\left[\chi\left(y\right)\right]$$

- ψ is odd and non-decreasing
- χ is even and non-decreasing on $[0,\infty)$

Computational Issues

Unfortunately, the following questions do not have a positive/easy answer.

- Does a solution to the simultaneous estimating equation exist?
- Is the solution unique (provided it exists)?
- Is there a computing algorithm to efficiently and safely solve the estimating equation?
- Does the computing algorithm converge?

M-Estimates of Location with Fixed Dispersion

- Solve the equation:

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

where $\hat{\sigma}_n = \hat{\sigma}(y_1, y_2, \ldots, y_n)$ is a given dispersion estimate

- To achieve robustness $\psi(y)$ must be bounded

- $\hat{\sigma}_n$ must be a robust dispersion estimate, used to calibrate the size of the location residuals.

- A popular choice for $\hat{\sigma}_n$ is the MAD, defined as

$$\hat{\sigma}_n = \frac{\text{Med}\left\{\left|y_1 - m\right|, \ \left|y_2 - m\right|, \ \ldots, \ \left|y_n - m\right|\right\}}{\Phi^{-1}(3/4)}$$

$$m = Med\left\{y_i\right\}$$

- $1/\Phi^{-1}(3/4) \approx 1.5$ (for consistency under Normal model)

- Calculated by the function mad in R

```
> x=rnorm(1000); sd(x); mad(x)
[1] 1.030051
[1] 1.051008

> x=c(rnorm(900),rt(100,1)); sd(x); mad(x)
[1] 2.782721
[1] 0.9926197
```

Equivariance Considerations

In general, dispersion estimates must have the following properties:

S1. **Scale Equivariance:** $\hat{\sigma}_n(ay_1, ay_2, \ldots, ay_n) = |a|\,\hat{\sigma}_n(y_1, y_2, \ldots, y_n)$, for all a in R

S2. **Location Invariance:**
$\hat{\sigma}_n(y_1 + b, y_2 + b, \ldots, y_n + b) = \hat{\sigma}_n(y_1, y_2, \ldots, y_n)$,
for all b in R

Equivariance Considerations

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

L1. **Scale Equivariance:** $\hat{\mu}(ay_1, ay_2, \ldots, ay_n) = a\hat{\mu}(y_1, y_2, \ldots, y_n)$, for all a in R.

L2. **Location Equivariance:**
$\hat{\mu}(y_1 + b, y_2 + b, \ldots, y_n + b) = \hat{\mu}(y_1, y_2, \ldots, y_n) + b$,
for all b in R.

Computation of Location M-Estimates

- To compute robust location estimate we must solve the non-linear equation by numerical means.
- We will discuss two algorithms:

Re-weighting: it is easy to code and doesn't require the calculation of the derivative of the location-score function.

Newthon-Raphson: converges faster and therefore is usually preferred.

Reweighting

Set

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

and define

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

We can write

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

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

or equivalently

$$\hat{\mu}_n=\frac{\sum w\left(\frac{y_i-\hat{\mu}_n}{\hat{\sigma}_n}\right)y_i}{\sum w\left(\frac{y_i-\hat{\mu}_n}{\hat{\sigma}_n}\right)}. \quad (1)$$

Reweighting Algorithm

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

Step 2 Let

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

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

We wish to find a solution t^* for the equation

$$g(t) = 0, \tag{2}$$

Suppose we have an approximate solution, $t^{(k)}$, at step k.

Instead of directly solving (2) we solve the linear approximation

$$g(t^{(k)}) + g'(t^{(k)})(t - t^{(k)}) = 0. \tag{3}$$

- We get

$$t^{(k+1)} = t^{(k)} - \frac{g\left(t^{(k)}\right)}{g'\left(t^{(k)}\right)}. \tag{4}$$

- This procedure critically depends on an initial estimate (approximation) $t^{(0)}$

- $t^{(0)}$ should be close to the solution t^*.

In the location case we have

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

and

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

Newton-Raphson Algorithm

The algorithm has the following steps:

Step 1 Set $\hat{\mu}_n^{(0)}$ =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 RESULTS

Notation

Let

$$\lambda_F(t) = \mathrm{E}_F\left\{-\psi\left(\frac{y-t}{\sigma(F)}\right)\right\} \quad , \quad \sigma_F^2(t) = \mathrm{Var}_F\left\{\psi\left(\frac{y-t}{\sigma(F)}\right)\right\}$$

and

$$\lambda_n(t) = -\frac{1}{n}\sum\psi\left(\frac{y_i-t}{\hat{\sigma}_n}\right). \tag{5}$$

Assumptions

A.0 The distribution function F is symmetric about μ

A.1 $\hat{\sigma}_n$ is consistent, that is $\hat{\sigma}_n \to \sigma(F) \quad \text{a.s.} \quad [F]$.

A.2 $\sqrt{n}\,(\hat{\sigma}_n - \sigma(F))$ is bounded in probability, that is, $\sqrt{n}\,(\hat{\sigma}_n - \sigma(F)) = 0_p(1)$.

A.3 $\sigma_F^2(t)$ is continuous on a neighborhood of μ and $0 < \sigma_F^2(\mu) < \infty$.

A.4 $\lambda_F(\mu) = 0$, $\lambda_F(t)$ is continuously differentiable on a neighborhood of μ and $\lambda_F'(\mu) > 0$.

Asymptotic Results

Suppose that A.0 - A.4 hold. Then:

$$\hat{\mu}_n \to \mu \quad \text{a.s.} \quad [F]$$

and

$$\sqrt{n}(\hat{\mu}_n - \mu) \to_d \text{ N}\left(0, \text{AV}(\psi, F)\right),$$

with

$$\text{AV}(\psi, F) = \frac{\text{E}_F\left\{\psi^2\left(\frac{y-\mu}{\sigma(F)}\right)\right\}}{\left(\lambda_F'(\mu)\right)^2}.$$

Suppose that

$$F(y) = F_0\left(\frac{y-\mu}{\sigma}\right)$$

and that

$$\hat{\sigma}_n \to \sigma(F) = \sigma$$

In this case

$$\begin{aligned} E_F\left\{\psi^2\left(\frac{y-\mu}{\sigma}\right)\right\} &= \frac{1}{\sigma}\int_{-\infty}^{\infty}\psi^2\left(\frac{y-\mu}{\sigma}\right) f_0\left(\frac{y-\mu}{\sigma}\right) dy \\ &= \int_{-\infty}^{\infty}\psi^2(y)\, f_0(y)\, dy = E_{F_0}\left\{\psi^2(y)\right\} \end{aligned}$$

Example 1: Sample Median

$$\psi(y) = sign(y)$$

$$\begin{aligned}
\lambda_F(t) &= -E_F\left\{sign\left(\frac{y-t}{\sigma}\right)\right\} \\
&= P_F(y \le t) - P_F(y > t) \\
&= P_F(y \le t) - [1 - P_F(y \le t)] = 2F(t) - 1 \\
&= 2F(t) - 1
\end{aligned}$$

Hence

$$\lambda_F'(\mu) = \frac{d}{dt}\left[2F(t) - 1\right]\Big|_{t=\mu} = 2f(\mu)$$

Since

$$E_F\left\{sign^2\left(\frac{y-\mu}{\sigma}\right)\right\} = 1$$

we have

$$AV\left(\text{Median}, F_{\mu,\sigma}\right) = \frac{1}{4f^2(\mu)} = \frac{1}{4\left(1/\sigma^2\right) f_0^2\left(\frac{\mu-\mu}{\sigma}\right)} = \frac{\sigma^2}{4f_0^2(0)}$$

If $\psi(y)$ is continuous, almost everywhere differentiable,

$$
\begin{aligned}
\lambda_F'(\mu) &= \left.\frac{d}{dt}\lambda_F(t)\right|_{t=\mu} \\
& \left.\frac{d}{dt}E_F\left\{-\psi\left(\frac{y-t}{\sigma}\right)\right\}\right|_{t=\mu} \\
&= \left.-E_F\left\{\frac{d}{dt}\psi\left(\frac{y-t}{\sigma}\right)\right\}\right|_{t=\mu} \\
&= \frac{1}{\sigma}E_F\left\{\psi'\left(\frac{y-\mu}{\sigma}\right)\right\} = \frac{1}{\sigma}E_{F_0}\left\{\psi'(y)\right\}
\end{aligned}
$$

Hence, under regularity assumptions (smooth ψ function)

$$\mathrm{AV}(\psi, F) \quad = \quad \sigma^2 \frac{E_{F_0}\left\{\psi^2(y)\right\}}{\left[E_{F_0}\left\{\psi'(y)\right\}\right]^2}$$

NOTE: $\mathrm{AV}(\psi, F_{\mu,\sigma})$ doesn't depend on μ. It depends on σ^2 only as a multiplicative factor.

To fix ideas, we will consider the Gaussian case:

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

$$f_{\mu,\sigma}(y) = \frac{1}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right)$$

Example 2: Sample Mean

$$\psi(y) = y \quad \text{(very smooth)}$$

Use the formula

$$\mathrm{AV}(\psi, F) = \sigma^2 \frac{E_{F_0}\left\{\psi^2(y)\right\}}{\left[E_{F_0}\left\{\psi'(y)\right\}\right]^2}$$

$$E_{\Phi}\left\{\psi'(y)\right\} = 1$$

$$E_{\Phi}\left\{\psi^2(y)\right\} = E_{\Phi}\left\{y^2\right\} = 1$$

$$AV\left(\text{Mean,Normal}\right) = \frac{1}{\left(1/\sigma\right)^2} = \sigma^2$$

Example 3: Huber's Optimal Psi-Function

$$\psi_c(y) = \begin{cases} y & -c < y < c \\ c \;\; sign(y) & \text{otherwise} \end{cases}$$

Since $\psi_c(y)$ is continuous and almost everywhere differentiable we can use the formula

$$AV(\psi, F) = \sigma^2 \frac{E_{F_0}\{\psi^2(y)\}}{\left[E_{F_0}\{\psi'(y)\}\right]^2}$$

$$
\begin{aligned}
E_{\Phi}\left\{\psi_c'(y)\right\} &= \int_{-c}^{c} \varphi(y)\,dy \\
&= \Phi(c) - \Phi(-c) \\
&= 2\Phi(c) - 1
\end{aligned}
$$

$$E_{\Phi}\left\{\psi_c^2(y)\right\} = 2c^2(1-\Phi(c)) + 2\int_0^c y^2\varphi(y)\,dy$$

$$\begin{aligned}\int_0^c y^2\varphi(y)\,dy &= -\int_0^c y\varphi'(y)\,dy \\ &= -\left[y\varphi(y)|_0^c - \int_0^c \varphi(y)\,dy\right] \\ &= -c\varphi(c) + \Phi(c) - \frac{1}{2}\end{aligned}$$

In summary

$$E_{\Phi}\left\{\psi_c'(y)\right\} = 2\Phi(c) - 1$$

$$E_{\Phi}\left\{\psi_c^2(y)\right\} = 2\left[c^2(1-\Phi(c)) - c\varphi(c) + \Phi(c) - \frac{1}{2}\right]$$

Therefore,

$$\begin{aligned} AV(\psi_c, \text{Normal}) &= \sigma^2 \frac{E_{\Phi}\left\{\psi_c^2(y)\right\}}{\left[E_{\Phi}\left\{\psi_c'(y)\right\}\right]^2} \\ &= \sigma^2 \frac{2\left[c^2(1-\Phi(c)) - c\varphi(c) + \Phi(c) - \frac{1}{2}\right]}{\left[2\Phi(c) - 1\right]^2} \end{aligned}$$

ML-Estimate

In this case

$$\psi_{F_{\mu,\sigma}}(y) = -\frac{f_0'\left(\frac{y-\mu}{\sigma}\right)}{f_0\left(\frac{y-\mu}{\sigma}\right)},$$

We will use the formula:

$$AV(\psi_{F_{\mu,\sigma}}, F_{\mu,\sigma}) = \sigma^2 \frac{E_{F_{\mu,\sigma}}\left\{\psi^2_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}}{\left(E_{F_{\mu,\sigma}}\left\{\psi'_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}\right)^2}.$$

Recall the (ML theory) result

$$I_\mu\left(F_{\mu,\sigma}\right) = E_{F_{\mu,\sigma}}\left\{\left(\frac{\partial}{\partial\mu}\log\left[\frac{1}{\sigma}f_0\left(\frac{y-\mu}{\sigma}\right)\right]\right)^2\right\}$$

$$= -E_{F_{\mu,\sigma}}\left\{\frac{\partial^2}{\partial\mu\partial\mu}\log\left[\frac{1}{\sigma}f_0\left(\frac{y-\mu}{\sigma}\right)\right]\right\}$$

That is:

$$
\begin{aligned}
I_{\mu}\left(F_{\mu,\sigma}\right) &= \frac{1}{\sigma^2} E_{F_{\mu,\sigma}}\left\{\psi^2_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\} \\
&= \frac{1}{\sigma^2} E_{F_{\mu,\sigma}}\left\{\psi'_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}
\end{aligned}
$$

Hence

$$
\begin{aligned}
AV(\psi_{F_{\mu,\sigma}}, F_{\mu,\sigma}) &= \sigma^2 \frac{E_{F_{\mu,\sigma}}\left\{\psi^2_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}}{\left(E_{F_{\mu,\sigma}}\left\{\psi'_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}\right)^2} \\
&= \frac{\frac{1}{\sigma^2} E_{F_{\mu,\sigma}}\left\{\psi^2_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}}{\left(\frac{1}{\sigma^2} E_{F_{\mu,\sigma}}\left\{\psi'_{F_{\mu,\sigma}}\left(\frac{y-\mu}{\sigma}\right)\right\}\right)^2} \\
&= \frac{I_\mu(F_{\mu,\sigma})}{\left(I_\mu(F_{\mu,\sigma})\right)^2} = \frac{1}{I_\mu(F_{\mu,\sigma})}
\end{aligned}
$$

Relative Asymptotic Efficiency

- Suppose that $T_1$and T_2 are asymptotically normal under F :

$$\sqrt{n}\left(T_1-\mu\right) \rightarrow {}_d N\left(0, v_1^2\left(F\right)\right)$$

$$\sqrt{n}\left(T_2-\mu\right) \rightarrow {}_d N\left(0, v_2^2\left(F\right)\right)$$

- The asymptotic efficiency of T_1 relative to T_2 is given by:

$$EFF\left(T_1, T_2, F\right)=\frac{v_2^2\left(F\right)}{v_1^2\left(F\right)}$$

- For example, if $v_1^2\left(F\right)=1.052632$ and $v_2^2\left(F\right)=1$ we have

$$EFF\left(T_1, T_2, F\right)=\frac{1}{1.052632}=0.95$$

Comparison of the Mean and Median

- We consider

$$F(y) = \frac{1}{\sigma} f_0 \left(\frac{y-\mu}{\sigma} \right)$$

- $f_0(y) =$ scaled student t-distribution with 3, 5, 10, 20, 100 degrees of freedom.
- Scaled so that

$$Var_{F_0}(y) = 1$$

- Since

$$Var\left(t_{(v)}\right) = \frac{v}{v-2}, \quad v = \quad \text{degrees of freedom}$$

We take

$$y = \sqrt{\frac{v-2}{v}} t_{(v)}$$

$$f(t) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v}\right)^{-(v+1)/2} \qquad (\ t_{(v)}\ )$$

$$f_0(y) = \sqrt{\frac{v}{v-2}}\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\left(1+\frac{t^2}{v-2}\right)^{-(v+1)/2} \qquad (\text{scaled } t_{(v)}\ )$$

$$4\left[f_0'(0)\right]^2 = 4\frac{v}{v-2}\left(\frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\,\Gamma\left(\frac{v}{2}\right)}\right)^2$$

We have shown that

$$AV(Median, F) = \frac{\sigma^2}{4\left[f_0'(0)\right]^2}$$

$$AV(Mean, F) = \sigma^2$$

Therefore

$$EFF(Median, Mean, F) = 4\left[f_0'(0)\right]^2$$

Relative efficiency of the median relative to the mean

Degrees of Freedom	1	2	3	4	5	10	20	1000
Relative Efficiency	∞	∞	1.62	1.12	0.96	0.76	0.69	0.64

ESTIMATES VIEWED AS FUNCTIONALS

Empirical Distribution

- Empirical distribution function

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(y_i \leq y)$$

- If $y_1, y_2, ..., y_n$ are iid F then

$$\lim_{n\to\infty} \sup_{y\in R} |F_n(y) - F(y)| = 0 \quad \text{(Glivenko-Cantelli)}$$

Estimating Functional

- Many estimates do not depend on the order in which the observations are inputted

$$T_n = T(y_1, y_2, ..., y_n) = T\left(y_{(1)}, y_{(2)}, ..., y_{(n)}\right)$$

$$y_{(1)} < y_{(2)} < \cdots < y_{(n)} \qquad \text{(order statistic)}$$

- Examples: mean, median, M-estimates, regression LS estimates, etc.
- In this case we can write

$$T_n = T(F_n)$$

Examples

- Sample mean

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = E_{F_n}(y) = Ave(F_n)$$

- Sample median

$$m = F_n^{-1}\left(\frac{1}{2}\right) \qquad \text{(informally)}$$

$$\begin{aligned} m_1 &= \sup\left\{y : F_n(y) \leq \frac{1}{2}\right\} \qquad \text{(upper median)} \\ m_2 &= \inf\left\{y : F_n(y) \geq \frac{1}{2}\right\} \qquad \text{(lower median)} \end{aligned}$$

$$m = \frac{m_1 + m_2}{2} = Med(F_n)$$

Examples

- Location M-estimates

$$T_n \quad \text{solves} \quad : \quad \frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{y_i - t}{\widehat{\sigma}\left(F_n\right)}\right) = 0$$

$$T\left(F_n\right) \quad \text{solves} \quad : \quad E_{F_n}\left(\psi\left(\frac{y - t}{\widehat{\sigma}\left(F_n\right)}\right)\right) = 0$$

- Dispersion M-estimates

$$S_n \quad \text{solves} \quad : \quad \frac{1}{n}\sum_{i=1}^{n}\chi\left(\frac{y_i - m\left(F_n\right)}{s}\right) = b, \quad b = E_{F_0}\left\{\chi\left(y\right)\right\}$$

$$S\left(F_n\right) \quad \text{solves} \quad : \quad E_{F_n}\left\{\chi\left(\frac{y_i - m\left(F_n\right)}{s}\right)\right\} = b$$

The Functional

- The functional $T(F)$ is defined for all empirical distribution function F_n

- We extend the definition over a larger set of distribution functions by substituting F_n by F.
- If the functional T is continuous we have:

$$\lim_{n\to\infty} T(F_n) \;=\; T(F) \quad \text{a.s. } [F]$$

when the data are iid $[F]$.

Examples

The sample mean:

$$T(F) = E_F(y), \quad \text{provided } E_F(|y|) < \infty$$

The sample median:

$$T(F) = F^{-1}(1/2), \quad \text{provided upper med}(F) = \text{ lower med } (F)$$

Location M-Estimates:

$$T(F) \text{ solves } : E_F\left\{\psi\left(\frac{y-t}{S(F)}\right)\right\} = 0, \quad S(F) \text{ is a disp. functional.}$$

Measuring Robustness

- Let $T(F)$ be an estimating functional
- Suppose $T(F)$ is defined on a set of distributions including
  - **Empirical distributions** F_n [in this case $T_n = T(F_n)$]
  - The **robustness neighborhood**

$$\mathcal{F}_\epsilon = \{F : F(y) = (1-\epsilon) F_{\boldsymbol{\theta}}(y) + H(y)\}$$

- A robust estimate should satisfy $T(F) \approx T(F_{\boldsymbol{\theta}})$ when ϵ is small

Examples

Let F be a contaminated normal distribution:

$$F(y) = (1-\epsilon) N\left(\mu, \sigma^2\right) + \epsilon Unif\left(d-h, d+h\right), \qquad d > \mu, h > 0$$

The sample mean is not robust:

$$Ave\left[N\left(\mu, \sigma^2\right)\right] = \mu, \qquad Ave\left(F\right) = (1-\epsilon)\mu + \epsilon d$$

$$\left|Ave\left[N\left(\mu, \sigma^2\right)\right] - Ave\left(F\right)\right| = \left|\epsilon\left(\mu - d\right)\right| \to \infty, \quad \text{as } d \to \infty$$

The sample median is robust:

$$Med\left(N\left(\mu,\sigma^2\right)\right)=\mu,\qquad \mu\leq Med\left(F\right)\leq\mu+\sigma\Phi\left(\frac{1}{2\left(1-\epsilon\right)}\right)$$

In fact:

$$\left(1-\epsilon\right)\Phi\left(\frac{y-\mu}{\sigma}\right)+\epsilon H\left(y\right)=\frac{1}{2}$$

$$\Phi\left(\frac{y-\mu}{\sigma}\right)=\frac{1/2-\epsilon H\left(y\right)}{1-\epsilon}\Rightarrow y\leq\mu+\sigma\Phi^{-1}\left(\frac{1}{2\left(1-\epsilon\right)}\right)$$

So,

$$\left|Med\left[N\left(\mu,\sigma^2\right)\right]-Med\left(F\right)\right|\leq\sigma\Phi^{-1}\left(\frac{1}{2\left(1-\epsilon\right)}\right)$$

Maxbias and Related Measures

- Let $F(\mathbf{y};\boldsymbol{\theta})$ be a parametric family, $\boldsymbol{\theta} \in \Theta$.
- $\mathbf{y}$ and $\boldsymbol{\theta}$ can be vector valued
- Let

$$\mathcal{F}_{\epsilon} = \{F(\mathbf{y}) : F(\mathbf{y}) = (1-\epsilon)F(\mathbf{y};\boldsymbol{\theta}) + \epsilon H(\mathbf{y})\}$$

- Consider an appropriate distance d on the parameter space Θ
  - **Example**: $\boldsymbol{\theta} = (\mu, \sigma)$, the parameter of interest is μ:

$$d(\widehat{\mu}, \mu) = \frac{|\widehat{\mu} - \mu|}{\sigma} \quad \text{(location-scale invariant)}$$

  - **Example**: $\boldsymbol{\theta} = (\mu, \sigma)$, the parameter of interest is σ:

$$\begin{aligned} d(\widehat{\sigma}, \sigma) &= \left| \frac{\widehat{\sigma} - \sigma}{\sigma} \right| \quad \text{(location-scale invariant)} \\ &= \left| \frac{\widehat{\sigma}}{\sigma} - 1 \right| \end{aligned}$$

- **Contamination bias**:

$$b_T(\epsilon, F) = d\left[T(F), T(F_{\boldsymbol{\theta}})\right], \qquad F \in \mathcal{F}_\epsilon$$

- **Contamination maxbias**

$$B_T(\epsilon) = \sup_{F \in \mathcal{F}_\epsilon} d\left[T(F), T(F_{\boldsymbol{\theta}})\right]$$

The Breakdown Point (BP)

- The BP of an estimating functional $T(F)$ is defined as follows

$$BP_T = \sup\{\epsilon : B_T(\epsilon) < \infty\}$$

- **Example:** $BP_{Mean} = 0$
- **Example:** $BP_{median} = 1/2$
- **Example:** $BP_{MAD} = 1/2$
- **Example:** T is a location estimate with bounded ψ and dispersion $\widehat{\sigma}$. Then

$$BP_T = BP_{\widehat{\sigma}}$$

If $\widehat{\sigma} = MAD$ then $BP_T = 1/2$

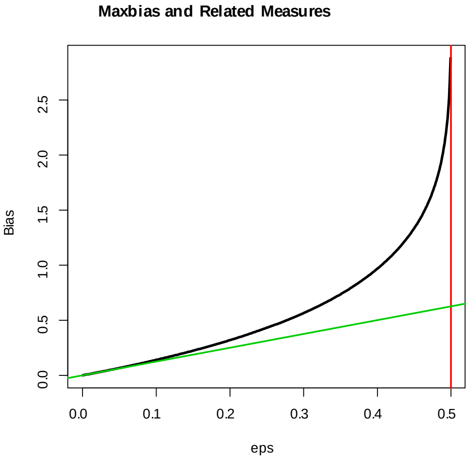
The Gross Error Sensitivity (GES)

The Gross-Error-Sensitivity (GES) is defined as follows:

$$GES_T = \frac{d}{d\epsilon} B_T(\epsilon) \Big|_{\epsilon=0} = B'_T(0)$$

Therefore

$$B_T(\epsilon) = \epsilon\, GES_T + o(\epsilon)$$


Maxbias and Related Measures
Bias
0.0
0.5
1.0
1.5
2.0
2.5
0.0
0.1
0.2
0.3
0.4
0.5
eps

The Gross-Error-Sensitivity (GES) is defined as follows:

$$GES_T = \frac{d}{d\epsilon} B_T(\epsilon)\Big|_{\epsilon=0} = B'_T(0)$$

Therefore

$$B_T(\epsilon) = \epsilon GES_T + o(\epsilon)$$

The Influence Function (IF)

- Introduced by Frank Hampel in his 1967 Ph.D. dissertation (together with the BP and GES)

- The IF is an asymptotic and infinitesimal concept

- Measures the asymptotic impact of a vanishingly small fraction of outliers located at a fixed position $\mathbf{y}$

Notation

- Consider a central parametric model

$$F_{\boldsymbol{\theta}}\left(\mathbf{x}\right)$$

and a Fisher Consistent estimate T:

$$T\left(F_{\boldsymbol{\theta}}\right)=\boldsymbol{\theta}$$

- Consider the "point mass contamination" distribution

$$F_{\epsilon}\left(\mathbf{x}\right)=\left(1-\epsilon\right)F_{\boldsymbol{\theta}}\left(\mathbf{x}\right)+\Delta_{\mathbf{y}}\left(\mathbf{x}\right)$$

Definition of IF

The IF of the estimate T at $\mathbf{y}$ and $F_{\boldsymbol{\theta}}$ is defined as:

$$IF_T(\mathbf{y}, F_{\boldsymbol{\theta}}) = \lim_{\epsilon \to 0} \frac{T(F_\epsilon) - T(F_{\boldsymbol{\theta}})}{\epsilon} = \frac{d}{d\epsilon} T(F_\epsilon)\Big|_{\epsilon=0}$$

provided the limit exists.

Example 1: The IF for the Sample Mean

- The central parametric model:

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- The estimate:

$$T(F) = AVE(F) = E_F(X)$$

  Clearly

$$AVE(F_{\mu,\sigma}) = \mu \qquad \text{(Fisher Consistent)}$$

- The point mass contamination distribution:

$$F_\epsilon(x) = (1-\epsilon)F_{\mu,\sigma}(x) + \Delta_y(x)$$

$$
\begin{aligned}
IF_{AVE}\left(y, F_{\mu,\sigma}\right) &= \lim_{\epsilon \to 0} \frac{AVE\left(F_{\epsilon}\right) - AVE\left(F_{\mu,\sigma}\right)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{AVE\left(\left(1-\epsilon\right) F_{\mu,\sigma}\left(x\right) + \Delta_{y}\left(x\right)\right) - AVE\left(F_{\mu,\sigma}\right)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{\left(1-\epsilon\right) E_{F_{\mu,\sigma}}\left(X\right) + \epsilon y - E_{F_{\mu,\sigma}}\left(X\right)}{\epsilon} \\
&= \lim_{\epsilon \to 0} \frac{\left(1-\epsilon\right) \mu + \epsilon y - \mu}{\epsilon} = \lim_{\epsilon \to 0} \frac{\epsilon\left(y-\mu\right)}{\epsilon} = y - \mu
\end{aligned}
$$

Example 2: The IF for Location M-Estimates

- The central parametric model:

$$F_{\mu,\sigma}(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- The estimate $T(F)$ satisfies the estimating equation:

$$E_F\left\{\psi\left(\frac{X-T(F)}{S(F)}\right)\right\} = 0$$

Fisher Consistency:

$$E_{F_{\mu,\sigma}}\left\{\psi\left(\frac{X-\mu}{S(F)}\right)\right\} = 0 \quad \text{(by symmetry)}$$

- The point mass contamination distribution:

$$F_\epsilon(x) = (1-\epsilon)F_{\mu,\sigma}(x) + \Delta_y(x)$$

Derivation of the IF for Location M-Estimates

$T\left(F_{\epsilon}\right)$ satisfies the equation:

$$E_{F_{\epsilon}}\left\{\psi\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\right\} \quad = \quad 0$$

$$E_{(1-\epsilon)F_{\mu,\sigma}+\epsilon\Delta_{y}}\left\{\psi\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\right\} \quad = \quad 0$$

$$(1-\epsilon)\,E_{F_{\mu,\sigma}}\left\{\psi\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\right\}+\epsilon\psi\left(\frac{y-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)=0$$

NOTATION

Set

$$\dot{T}(F_\epsilon) = \frac{d}{d\epsilon} T(F_\epsilon)$$

and

$$\dot{S}(F_\epsilon) = \frac{d}{d\epsilon} S(F_\epsilon)$$

Hence

$$IF_T(\mathbf{y}, F_{\boldsymbol{\theta}}) = \dot{T}(F_{\boldsymbol{\theta}}) \quad \text{and} \quad IF_S(\mathbf{y}, F_{\boldsymbol{\theta}}) = \dot{S}(F_{\boldsymbol{\theta}})$$

$$(1-\epsilon)\, E_{F_{\mu,\sigma}}\left\{\psi\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\right\}+\epsilon\psi\left(\frac{y-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right) \quad = \quad 0$$

Differentiating both sides with respect to ϵ :

$$-E_{F_{\mu,\sigma}}\left\{\psi\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\right\}-$$

$$(1-\epsilon)\, E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)\frac{\dot{T}\left(F_{\epsilon}\right) S\left(F_{\epsilon}\right)+\dot{S}\left(F_{\epsilon}\right)\left(X-T\left(F_{\epsilon}\right)\right)}{S^{2}\left(F_{\epsilon}\right)}\right\}$$

$$+\psi\left(\frac{y-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right)+\epsilon\frac{d}{d\epsilon}\psi\left(\frac{y-T\left(F_{\epsilon}\right)}{S\left(F_{\epsilon}\right)}\right) \quad = \quad 0$$

Evaluating the derivative at $\epsilon = 0$:

$$\overbrace{-E_{F_{\mu,\sigma}}\left\{\psi\left(\frac{X-\mu}{S\left(F_{\mu,\sigma}\right)}\right)\right\}}^{=0}$$

$$-\left(1-\epsilon\right)E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-\mu}{S\left(F_{\mu,\sigma}\right)}\right)\frac{IF_T\left(y,F_{\mu,\sigma}\right)}{S\left(F_{\mu,\sigma}\right)}\right\}$$

$$+\psi\left(\frac{y-\mu}{S\left(F_{\mu,\sigma}\right)}\right)=0$$

Therefore

$$IF_T\left(y, F_{\mu,\sigma}\right) = S\left(F_{\mu,\sigma}\right) \frac{\psi\left(\frac{y-\mu}{S\left(F_{\mu,\sigma}\right)}\right)}{E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-\mu}{S\left(F_{\mu,\sigma}\right)}\right)\right\}}$$

If the dispersion estimate $S\left(F\right)$ is Fisher consistent we have

$$IF_T\left(y, F_{\mu,\sigma}\right) = \sigma \frac{\psi\left(\frac{y-\mu}{\sigma}\right)}{E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-\mu}{\sigma}\right)\right\}}$$

Three Heuristic Results

1)

$$E_{F_{\mu,\sigma}}\left\{IF_T^2\left(Y, F_{\mu,\sigma}\right)\right\} = \sigma^2 \frac{E_{F_{\mu,\sigma}}\left\{\psi^2\left(\frac{y-\mu}{\sigma}\right)\right\}}{\left[E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-\mu}{\sigma}\right)\right\}\right]^2} = AV\left(T, F_{\mu,\sigma}\right)$$

2)

$$\sup_y IF_T\left(y, F_{\mu,\sigma}\right) = \frac{\sigma\psi\left(\infty\right)}{E_{F_{\mu,\sigma}}\left\{\psi'\left(\frac{X-\mu}{\sigma}\right)\right\}} = GES\left(T, F_{\mu,\sigma}\right)$$

3) If the Fisher consistent estimate $T(F)$ satisfies the estimating equation

$$E_F\left\{\Psi\left(\mathbf{X}, T(F)\right)\right\} = \mathbf{0}$$

then the IF of $T(F)$ at $F_{\boldsymbol{\theta}}$ and $\mathbf{y}$ is proportional to

$$\Psi\left(\mathbf{y}, T\left(F_{\boldsymbol{\theta}}\right)\right)$$