Estimates Viewed as Functionals

The calculation of some estimates does not depend on the order in which we enter the observations. If T_n is one such estimate, then its value doesn't change when applied to any arbitrary permutation of the data. Hence we can write

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

where $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$ is the so called *order statistic*.

When (1) holds we can move one step forward and view T_n as a *functional* of the empirical distribution function

$$H_n(t) = \frac{1}{n} \sum I(y_i \leq t),$$

where

$$I(y_i \leq t) = \begin{cases} 1 & \text{if} \quad y_i \leq t \\ 0 & \text{if} \quad y_i > t. \end{cases}$$

In other words, when (1) holds we can write the functional representation

$$T_n = T(H_n).$$

We use the word *functional* instead of *function* to indicate that the domain of T is no longer a subset R^n but a subset of functions (the empirical distribution functions in the present case).

Example 1 *The sample mean and in general, any location M-estimate with general dispersion, are functionals of the empirical distribution function:*

$$\bar{y} = \frac{1}{n} \sum y_i = E_{H_n}(y)$$

and

$$\hat{\mu}_n = \sup \left\{ t : \frac{1}{n} \sum \psi(\frac{y_i - t}{\hat{\sigma}}) > 0 \right\} = \sup \left\{ t : E_{H_n} \psi(\frac{y - t}{\hat{\sigma}(H_n)}) > 0 \right\} = \hat{\mu}(H_n).$$

provided that the dispersion estimate $\hat{\sigma}$ can also be computed from H_n.

The definition of robust estimates $T(H_n)$ can be extended to all the distribution functions in a neighborhood $\mathcal{F}_\epsilon$ by setting $T(H)$ equal to the almost sure limit of $T(H_n)$ under H. That is

$$T(H) = \lim_{n \to \infty} T(H_n) \qquad \text{a.s. } [H].$$

Notice that

$$\lim_{n\to\infty} H_n(t) = H(t) \qquad \text{a.s. } [H]$$

and therefore robust estimates are "continuous" in the sense that $T(H_n) \to T(H)$ when $H_n \to H$.

In general, the definition of $T(H)$ is obtained by replacing H_n by H in the expression that defines $T(H_n)$. For example, in the case of M-estimate with general dispersion (with continuous ψ) $\hat{\mu}(H)$ is the unique root of the equation

$$E_H \left\{ \psi \left(\frac{y-t}{\hat{\sigma}(H)} \right) \right\} = 0, \tag{2}$$

Moreover, if ψ is discontinuous but there is a unique change of sign point t_0 for the function

$$g\,(t) = E_H \left\{ \psi \left(\frac{y-t}{\hat{\sigma}(H)} \right) \right\},$$

it can be shown that the corresponding location estimate $\hat{\mu}(H_n)$ still converges a.s. $[H]$ to t_0. Notice that in this case the functional $\hat{\mu}(H)$ can be defined as $\hat{\mu}(H) = t_0$ and satisfies equation (2).

Linear Approximation for a Functional

Suppose that the real function $f(x)$ is differentiable at x_0. Then

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + o(x - x_0) \\ &= f(x_0) + L(x) + o(x - x_0) \end{aligned}$$

where $o(x - x_0)/(x - x_0) \to 0$ as $x \to x_0$ and $L(x)$ is an "approximating linear function" satisfying $L(x_0) = 0$.

We wish to define the concept of *linear approximation for a functional.* For that we need to introduce the concepts of *linear functional* and that of *distance between two distribution functions.*

The functional $T\,(H)$ is linear if said to be linear and only if

$$T\,[(1-\alpha)H_1 + \alpha H_2] \;=\; (1-\alpha)T\,(H_1) + \alpha T(H_2),$$

for all $0 \le \alpha \le 1$ and for all distributions H_1 and H_2 in its domain.

The concept of distance $d(H_1, H_2)$ between two distributions functions H_1 and H_2 can be defined in many different ways. For example, we can use the *total variation distance*

$$d(H_1, H_2) = \sup_y |H_1(y) - H_2(y)|.$$

Another example of distance between distributions is Levy's distance

$$d(H_1, H_2) = \sup_g \left| \int g(y)\, d[H_1 - H_2](y) \right|$$

where the supremum is over the set of all continuous functions with finite support and $\sup_y |g(y)| \leq 1$.

Now that we have the notions of distance and linearity for a functional we can introduce the concept of derivative of a functional. The functional $T(H)$ is "differentiable" at H_0 if there exists a *continuous linear functional* $L(H)$ such that

$$T(H) = T(H_0) + L(H) + o[d(H, H_0)],$$

with

$$\frac{o[d(H, H_0)]}{d(H, H_0)} \to 0 \quad \text{as} \quad d(H, H_0) \to 0.$$

The continuous linear approximating functional can be expressed as an integral (Riesz Representation Theorem, see for instance Ash, pg. 130)

$$L(H) = \int a_0(y) dH(y).$$

The function $a_0(y)$ is a is called the *kernel* of the linear functional $L(H)$. Notice that for all $0 \leq \alpha \leq 1$,

$$\begin{aligned} L[(1-\alpha)H_1 + \alpha H_2] &= \int a_0(y) d[(1-\alpha)H_1 + \alpha H_2](y) \\ &= (1-\alpha)\int a_0(y) dH_1(y) + \alpha \int a_0(y) dH_2(y) \\ &= (1-\alpha)L(H_1) + \alpha L(H_2). \end{aligned}$$

Since

$$L(H) = T(H) - T(H_0) - o[d(H, H_0)],$$

and $o[d(H_0, H_0)] = 0$ it follows that

$$L(H_0) = \int a_0(y) dH_0(y) = 0.$$

Gateaux Derivatives

To fix ideas, suppose that $d(H, H_0)$ is the total variation distance

$$d(H, H_0) = \sup_y |H(y) - H_0(y)|$$

and let

$$H_{t,G} = (1-t)H_0 + tG.$$

In this case,

$$\begin{aligned} d(H_{t,G}, H_0) &= d[(1-t)H_0 + tG, H_0] \\ &= \sup_y |(1-t)H_0(y) + tG(y) - H_0(y)| \\ &= \sup_y |t[G(y) - H_0(y)]| \\ &= |t| \sup_y |G(y) - H_0(y)| \\ &= |t|\, k(G) = O(t), \end{aligned}$$

since $k = k(G)$ is a constant that only depends on G. Moreover,

$$\left| \frac{o[d(H_{t,G}, H_0)]}{t} \right| = \frac{o[|t|k(G)]}{|t|k(G)} k(G) \to 0,$$

and so

$$o[d(H_{t,G}, H_0)] = o(t).$$

The functional $T(H)$ is said to be Gateaux differentiable at H_0 in the direction of G if

$$T(H_{t,G}) = T(H_0) + \int a_0(y) dH_{t,G}(y) + o(t).$$

Notice that,

$$\int a_0(y) dH_{t,G}(y) = (1-t) \int a_0(y) dH_0(y) + t \int a_0(y) dG(y) = t \int a_0(y) dG(y).$$

The Influence Function

If $T(H)$ is Gateaux differentiable at H_0 in the direction of G then,

$$T(H_{t,G}) = T(H_0) + t \int a_0(y) dG(y) + o(t),$$

and so

$$\lim_{t\to 0}\frac{T(H_{t,G})-T(H_0)}{t}=\int a_0(y)dG(y)+\lim_{t\to 0}\frac{o(t)}{t}=\int a_0(y)dG(y).$$

Taking $G=\delta_x$, a point mass distribution at x, we have

$$\int a_0(y)d\delta_x(y)=a_0(x).$$

In other words, the kernel $a_0(x)$ can be obtained by calculating the limit

$$\lim_{t\to 0}\frac{T((1-t)H_0+t\delta_x)-T(H_0)}{t}=a_0(x). \tag{3}$$

If this limit in (3)exists, it is called the influence function of T at x and H_0, denoted $IF(T,x,H_0)$. That is,

$$IF(T,x,H_0) = \lim_{t\to 0}\frac{T((1-t)H_0+t\delta_x)-T(H_0)}{t}, \text{ provided the limit exist.} \tag{4}$$

Notice that the limit in (3) does exist and equals the kernel $a_0(x)$ of $T(H)$ at H_0 when the functional $T(H)$ is Gateaux differentiable at H_0. However, the limit (4) may exist for a functional $T(H)$ which is not Gateaux differentiable. Therefore, the influence function is a weaker (more general) concept than the concept of Gateaux derivative.

To derive the influence function of an estimate T_n at a "central distribution" H_0 and at the point x we may follow the following steps:

1. Derive the functional representation $T_n = T(H_n)$. The representation may be explicit or implicit .
2. Derive the functional $T(H)$ by replacing H_n in 1. by a general H.
3. Evaluate (implicitly or explicitly) $g(t)=T((1-t)H_0+t\delta_x)$
4. Calculate (directly or indirectly) $\frac{d}{dt}g(t)=\frac{d}{dt}T((1-t)H_0+t\delta_x)=h(t)$
5. Finally, $IF(T,x,H_0)=h(0)=\frac{d}{dt}T((1-t)H_0+t\delta_x)|_{t=0}$.

Steps 1-5 are summarized in the following expression:

$$IF(T,x,H_0) = \frac{d}{dt}\left[T((1-t)H_0+t\delta_x)\right]_{t=0}=\frac{d}{dt}\left[T(H_{t,\delta_x})\right]_{t=0}$$

.

Example 2 *(Influence Function for the Mean) In this example we will compute the influence function for $T_n = \bar{x}$, the sample mean, at $H_0 = N(0,1)$. In this case*

$$\begin{aligned} T(H_n) &= \frac{1}{n}\sum x_i = E_{H_n}\{x\} \\ T(H) &= E_H\{x\} = \int_{-\infty}^{\infty} x\ dH(x) \qquad \textit{(provided the integral exists)} \end{aligned}$$

$$\begin{aligned} g(t) &= T(H_{t,\delta_y}) \\ &= \int_{-\infty}^{\infty} x\ dH_{t,\delta_y}(x) \\ &= \int_{-\infty}^{\infty} x\ d\left[(1-t)H_0 + t\delta_y\right](x) \\ &= (1-t)\int_{-\infty}^{\infty} x\ dH_0(x) + t\int_{-\infty}^{\infty} x\delta_y(x) \\ &= (1-t)\int_{-\infty}^{\infty} y\varphi(y)dy + ty \\ &= ty \end{aligned}$$

Now

$$h(t) = \frac{d}{dt}g(t) = \frac{d}{dt}ty = y$$

Therefore,

$$IF(Mean, y, N(0,1)) \quad = \quad = y.$$

Example 3 *(Influence function for the median) Let*

$$T(H_n) = Median(x_i),$$

be the sample median and set

$$T(H_{t,\delta_x}) \equiv T(x,t)$$

to simplify the notations. To fix ideas suppose that $H_0 = \Phi$ is the standard normal distribution.

We have the following results:

(a) $T(0,t)=0$, *for all* t.

(b) $T(x,t)\to 0$ *as* $t\to 0$, *for all* x

(c) $IF(T,x,\Phi)=(\partial/\partial t)\,T(x,t)\,|_{t=0}\;=\; sign(x)/[2\varphi(0)] = sign(x)\sqrt{\pi/2}$.

To prove (a) notice that by definition

$$\delta_0(x)=\begin{cases} 0 & \forall\ x<0 \\ 1 & \forall\ x\geq 0\end{cases}$$

and so

$$H_{t,\delta_0}(y)=(1-t)\Phi(y)+t\delta_0(y)=\begin{cases} (1-t)\Phi(y)<(1-t)/2 & \forall\ y<0 \\ (1-t)\Phi(y)+t>(1-t)/2+t & \forall\ y>0.\end{cases}$$

Therefore,

$$H_{t,\delta_0}(y)=\begin{cases} <1/2 & \forall\ y<0 \\ >1/2 & \forall\ y>0,\end{cases}$$

and (a) follows. Part (b) follows directly from the fact that for all $H_{t,G}$,

$$|T(H_{t,G})|\leq \Phi^{-1}\left[\frac{1}{2(1-t)}\right]\to 0,\ \text{as } t\to 0 \text{ (why?)}. \tag{5}$$

To prove (c), let $x>0$ *be given. Since by (b)* $T(x,t)\to 0$, $|T(x,t)|<x$, *for small enough* t. *Then, the single jump of* H_{t,δ_x} *occurs after* $T(x,t)$ *and so* $T(x,t)$ *satisfies the equation*

$$\frac{1}{2}=(1-t)\Phi\left(T(x,t)\right)+t\delta_x\left(T(x,t)\right)=(1-t)\Phi\left(T(x,t)\right).$$

Differentiating the two sides of the equation above and evaluating at $t=0$, *we obtain*

$$0=-\Phi\left(0\right)+\varphi(0)IF(T,x,\Phi)$$

and

$$IF(T,x,\Phi)=\frac{\Phi\left(0\right)}{\varphi(0)}=\frac{1}{2\varphi(0)}.$$

The case $x<0$ *is proved similarly. Finally, the case* $x=0$ *follows trivially from part (a).*

The Influence Function of Dispersion M-Estimates

Consider the smooth M-dispersion functional $s(H)$ with "centering" location estimate $m(H)$,

$$E_H \left\{ \chi \left(\frac{y - m(H)}{s(H)} \right) \right\} = b. \tag{6}$$

Suppose that:

1) H_0 is symmetric about m_0 and therefore $m(H_0) = m_0$.
2) $IF\,(m, x, H_0)$ is well defined.

We wish to calculate the influence function $IF\,(m, x, H_0)$ of $s(H)$ at H_0 and x. When H in (6) is given by

$$H_{t,\delta_x} \quad = \quad (1-t)H_0 + t\delta_x$$

this estimating equation becomes

$$(1-t)E_{H_0} \left\{ \chi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\} + t\chi \left(\frac{x - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) = b. \tag{7}$$

Differentiating with respect to t we get,

$$\frac{d}{dt} E_{H_{t,\delta_x}} \left\{ \chi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\} = -E_{H_0} \left\{ \chi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\} +$$

$$(1-t) \qquad E_{H_0} \left\{ \chi' \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \frac{-\dot{m}(H_{t,\delta_x}) s(H_{t,\delta_x}) - \dot{s}(H_{t,\delta_x})\left(x - m(H_{t,\delta_x})\right)}{s(H_{t,\delta_x})^2} \right\} +$$

$$\chi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) + t\frac{d}{dt}\chi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \quad = \quad 0$$

where

$$\dot{m}(H_{t,\delta_x}) = \frac{d}{dt} m(H_{t,\delta_x})$$

and

$$\dot{s}(H_{t,\delta_x}) = \frac{d}{dt} s(H_{t,\delta_x}).$$

Evaluating the derivative at $t = 0$ gives

$$\frac{d}{dt}E_{H_{t,\delta_x}}\left\{\chi\left(\frac{y-m(H_{t,\delta_x})}{s(H_{t,\delta_x})}\right)\right\}|_{t=0} = -b-\frac{IF(m,x,H_0)}{s(H_0)}E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\right\}$$

$$-\frac{IF\left(s,x,H_0\right)}{s(H_0)}E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\left(\frac{y-m_0}{s(H_0)}\right)\right\}+\chi\left(\frac{x-m_0}{s(H_0)}\right)$$

$$= \left[\chi\left(\frac{x-m_0}{s(H_0)}\right)-b\right]-\frac{IF\left(s,x,H_0\right)}{s(H_0)}E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\left(\frac{y-m_0}{s(H_0)}\right)\right\}=0$$

Therefore,

$$IF\left(s,x,H_0\right) = s(H_0)\frac{\chi\left(\frac{x-m_0}{s(H_0)}\right)-b}{E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\left(\frac{y-m_0}{s(H_0)}\right)\right\}}. \tag{8}$$

The Influence Function of Location M-Estimates

Consider the smooth M-location functional $m(H)$ with dispersion estimate $s(H)$,

$$E_H \left\{ \psi \left(\frac{y - m(H)}{s(H)} \right) \right\} = 0. \tag{9}$$

Suppose that:

1) H_0 is symmetric about m_0 and suppose that $m(H_0) = m_0$
2) $IF(s, x, H_0)$ is well defined.

Set $s(H_0) = s_0$.

When $H(y) = H_{t,\delta_x}(y) = (1-t)H_0(y) + t\delta_x(y)$ equation (9) becomes

$$(1-t)E_{H_0} \left\{ \psi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\} + t\psi \left(\frac{x - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) = 0. \tag{10}$$

Notice that

$$\frac{d}{dt} E_{H_0} \left\{ \psi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\}_{t=0} =$$

$$E_{H_0} \left\{ \psi' \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \frac{-s(H_{t,\delta_x})\dot{m}(H_{t,\delta_x}) - \dot{s}(H_{t,\delta_x})\left(y - m(H_{t,\delta_x})\right)}{s(H_{t,\delta_x})^2} \right\}_{t=0} =$$

$$E_{H_0} \left\{ \psi' \left(\frac{y - m_0}{s_0} \right) \frac{-s_0 IF(m, x, H_0) - IF(s, x, H_0)(y - m_0)}{s_0^2} \right\}. \tag{11}$$

Notice that

$$E_{H_0} \left\{ \psi' \left(\frac{y - m_0}{s_0} \right) \left(\frac{y - m_0}{s_0} \right) \right\} = 0$$

because $\psi'(-y)(-y) = -\psi'(y)\, y$ and H_0 is symmetric about m_0.

From (11),

$$\frac{d}{dt} E_{H_0} \left\{ \psi \left(\frac{y - m(H_{t,\delta_x})}{s(H_{t,\delta_x})} \right) \right\}_{t=0} = -\frac{1}{s_0} IF(m, x, H_0)\, E_{H_0} \left\{ \psi' \left(\frac{y - m_0}{s_0} \right) \right\} \tag{12}$$

Moreover

$$\frac{d}{dt}\left[t\psi\left(\frac{y-\hat{\mu}(H_{t,\delta_x})}{s(H_{t,\delta_x})}\right)\right]|_{t=0} = \psi\left(\frac{x-m_0}{s_0}\right). \tag{13}$$

Differentiating (10) with respect to t, at $t=0$ and using (12) and (13) gives:

$$IF(m,x,H_0) = s_0\frac{\psi\left(\frac{x-m_0}{s_0}\right)}{E_{H_0}\left\{\psi'\left(\frac{x-m_0}{s_0}\right)\right\}}. \tag{14}$$

Influence function and the asymptotic variance

There is a heuristic relation between the influence function and the asymptotic variance of an estimate. This relation can be used to quickly obtain an expression for the asymptotic variance of the estimate. Unfortunately, the conditions under which this heuristic relation can be formalized are very involved and therefore it is generally more convenient to formally derive the asymptotic distribution of the estimate by other means.

Assuming that the functional $T(H)$ is differentiable we can write

$$\begin{aligned} T(H) &= T(H_0) + \int a_0(x) dH(x) + o(t) \\ &= T(H_0) + \int IF\left(T, x, H_0\right) dH(x) + o(t) \\ &\approx T(H_0) + \int IF\left(T, x, H_0\right) dH(x). \end{aligned}$$

Applying this approximation when $H = H_n$ and $H_0 = H$, we obtain

$$\begin{aligned} T_n &= T(H_n) \approx T(H) + \int IF\left(T, x, H\right) dH_n(x) \\ &\approx T(H) + \frac{1}{n} \sum IF\left(T, y_i, H\right), \end{aligned}$$

and so

$$\sqrt{n}\left[T(H_n) - T(H)\right] \simeq \frac{1}{\sqrt{n}} \sum IF\left(T, y_i, H\right) \to_d N(0, E_H\left\{IF^2\left(T, y_i, H\right)\right\}).$$

Therefore,

$$AV(T,H) = \int IF^2\left(T, x, H\right) dH(x) \tag{15}$$

and the asymptotic variance of $T(H_n)$ under H can then be quickly obtained by integrating the square of $IF(T, x, H)$.

For example, in the case of the median,

$$IF(T, x, H) = \frac{\operatorname{sign}(x - m_0)}{2h(m_0)},$$

provided that H has a symmetric density $h(x) = H'(x)$. According to the heuristic relationship (15) we have

$$AV(Median, H) = \int \left(\frac{\operatorname{sign}(x - m_0)}{2h(m_0)}\right)^2 h(x)dx = \frac{1}{4h^2(m_0)}.$$

Influence function and the jackknife

Let $\bar{y}$ be the sample mean, that is the average of the data set $\{y_1, \ldots, y_n\}$ and let

$$\bar{y}^{(i)} = \frac{1}{n-1} \sum_{j \neq i} y_j,$$

be the average of $\{y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n\}$, the original data set without the i^{th} observation.

It is well known that the standard error of $\bar{y}$ (as an estimate of the population mean) is estimated by

$$\hat{\sigma}_{\bar{y}} = \sqrt{\frac{1}{n(n-1)} \sum (y_i - \bar{y})^2} = \frac{\hat{\sigma}}{\sqrt{n}}. \qquad (16)$$

A problem with formula (16) is that it doesn't naturally generalize for other estimates $T_n = T(H_n)$ (where H_n is the empirical distribution function). Let

$$\begin{aligned} \bar{y} &= M(H_n) \\ H_n^{(i)}(y) &= \frac{1}{n-1} \sum_{j \neq i} I(y_j \leq y) \\ \bar{y}^{(i)} &= M(H_n^{(i)}) \\ (y_i - \bar{y})^2 &= [(n-1)(\bar{y}^{(i)} - \bar{y})]^2 = [(n-1)(M(H_n^{(i)}) - M(H_n))]^2 \end{aligned}$$

Using the above notation we can obtain a more suggestive formula for $\hat{\sigma}_{\bar{y}}$:

$$\hat{\sigma}_{\bar{y}} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} r_i^2}, \qquad (17)$$

with

$$r_i = (n-1)[M(H_n^{(i)}) - M(H_n)]. \qquad (18)$$

The estimate for the standard error of $T_n = T(H_n)$:

$$\hat{\sigma}_J = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^{n} r_i^2} \quad \text{with} \quad r_i = (n-1)[T(H_n) - T(H_n^{(i)})]$$

is now clearly suggested by (17) and (18).

Notice that

$$\begin{aligned} H_n^{(i)}(y) &= \frac{1}{n-1}\sum_{j\neq i} I(y_j \le y) \\ &= \frac{1}{n-1}[nH_n(y) - I(y_i \le y)] \\ &= \frac{n}{n-1}H_n(y) - \frac{1}{n-1}I(y_i \le y) \\ &= \frac{n}{n-1}H_n(y) - \frac{1}{n-1}\delta_{y_i}(y), \end{aligned}$$

where $\delta_{y_i}(y)$ is a point mass distribution at y_i [$\delta_{y_i}(y) = 1$ iff $y \ge y_i$]. Setting

$$t_n = -\frac{1}{n-1}$$

and noticing that

$$1 - t_n = \frac{n}{n-1}$$

we can write

$$H_n^{(i)}(y) = \frac{n}{n-1}H_n(y) - \frac{1}{n-1}\delta_{y_i}(y) = (1-t_n)H_n(y) + t_n\delta_{y_i}(y).$$

Moreover,

$$T(H_n) - T(H_n^{(i)}) = -[T(H_n^{(i)}) - T(H_n)] = -\{T[(1-t_n)H_n(y) + t_n\delta_{y_i}(y)] - T(H_n)\}$$

and so

$$r_i = (n-1)[T(H_n) - T(H_n^{(i)})] = \frac{T[(1-t_n)H_n(y) + t_n\delta_{y_i}(y)] - T(H_n)}{t_n} \equiv IF_n(y_i).$$

The function $\mathrm{IF}_n(y_i)$ is called the *empirical influence function.* Notice that, under mild regularity conditions

$$IF_n(y_i) = \frac{T[(1-t_n)H_n(y) + t_n\delta_{y_i}(y)] - T(H_n)}{t_n} \to IF(T, y_i, H).$$

Finally,

$$(n-1)^2[T(H_n) - T(H_n^{(i)})]^2 = IF_n^2(y_i)$$

and, in view of (15),

$$\begin{aligned} \hat{\sigma}_J^2 &= \frac{(n-1)}{n}\sum_{i=1}^{n}[T(H_n) - T(H_n^{(i)})]^2 = \frac{1}{(n-1)n}\sum_{i=1}^{n} IF_n^2(y_i) \\ &\approx \frac{1}{n}AV\,(T, H)\,. \end{aligned}$$

The Gross Error Sensitivity

An estimate with uniformly small influence function will tend to be robust because it will be only mildly affected by the addition of an arbitrary data point to the sample (provided the sample is large). This idea can be expressed through the following heuristic argument.

Let H_n and H_{n+1} be the empirical distribution functions of the original sample

$$\{y_1, y_2, ..., y_n\}$$

and the augmented sample

$$\{y_1, y_2, ..., y_n, y\}$$

where y is an arbitrary number. That is,

$$H_n(x) = \frac{1}{n} \sum I_{(-\infty, y_i]}(x)$$

and

$$\begin{aligned} H_{n+1}(x) &= \left(1 - \frac{1}{n+1}\right) H_n(x) + \frac{1}{n+1} I_{(-\infty, y]}(x) \\ &= \left(1 - \frac{1}{n+1}\right) H_n(x) + \frac{1}{n+1} \delta_y(x) \end{aligned}$$

If the functional is differentiable then

$$\begin{aligned} T(H_{n+1}) &\approx T(H_n) + \int IF(H_n, x)\, dH_{n+1}(x) \\ &= T(H_n) + \frac{1}{n+1} \left[\sum_{i=1}^{n} IF(H_n, y_i) + IF(H_n, y) \right] \\ &= T(H_n) + \frac{1}{n+1} \left[\frac{n+1}{n} \int IF(H_n, x)\, dH_n(x) + IF(H_n, y) \right] \\ &= T(H_n) + \frac{1}{n+1} IF(H_n, y) \end{aligned}$$

because $\int IF\left(H_n, y\right) dH_n\left(y\right) = 0$. Therefore,

$$\sup_y \left|T\left(H_{n+1}\right) - T\left(H_n\right)\right| \quad \approx \quad \frac{\sup_y \left|IF\left(H_n, y\right)\right|}{n+1}.$$

If the sample size n is large and the sample comes from the distribution H_0 then $IF\left(H_n, y\right) \approx IF\left(H_0, y\right)$ and so

$$\sup_y \left|T\left(H_{n+1}\right) - T\left(H_n\right)\right| \quad \approx \quad \frac{\sup_y \left|IF\left(H_0, y\right)\right|}{n+1}. \tag{19}$$

The heuristic reasoning above suggests that the robustness of T_n can be measured by Hampels's *gross-error sensitivity* which is defined as the numerator on the right hand side of (19):

$$GES\left(T, H_0\right) \quad = \quad \sup_y \left|IF\left(H_0, y\right)\right|$$

Example 4 *The gross error sensitivity of location and dispersion M-estimates can be easily obtained using the formulas (8)*

$$IF(m, x, H_0) = s_0 \frac{\psi\left(\frac{x-m_0}{s_0}\right)}{E_{H_0}\left\{\psi'\left(\frac{y-m_0}{s_0}\right)\right\}}$$

and (14)

$$IF\left(s, x, H_0\right) = s(H_0) \frac{\chi\left(\frac{x-m_0}{s(H_0)}\right) - b}{E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\left(\frac{y-m_0}{s(H_0)}\right)\right\}}.$$

In fact,

$$\begin{aligned} GES\left(m, H_0\right) &= \sup_x \left|IF(m, x, H_0)\right| \\ &= \sup_x \left| s_0 \frac{\psi\left(\frac{x-m_0}{s_0}\right)}{E_{H_0}\left\{\psi'\left(\frac{y-m_0}{s_0}\right)\right\}} \right| \\ &= s_0 \frac{\psi\left(\infty\right)}{E_{H_0}\left\{\psi'\left(\frac{y-m_0}{s_0}\right)\right\}} \end{aligned}$$

and

$$
\begin{aligned}
GES\,(s, H_0) &= \sup_x |IF(m, x, H_0)| \\
&= \sup_x \left| s(H_0) \frac{\chi\left(\frac{x-m_0}{s(H_0)}\right) - b}{E_{H_0}\left\{\chi'\left(\frac{y-m_0}{s(H_0)}\right)\left(\frac{y-m_0}{s(H_0)}\right)\right\}} \right| \\
&= s_0 \frac{\max\{\chi\,(\infty) - b, b - \chi\,(0)\}}{E_{H_0}\left\{\chi'\left(\frac{y\,-\,m_0}{s_0}\right)\left(\frac{y\,-\,m_0}{s(H_0)}\right)\right\}} \\
&= s_0 \frac{\max\{\chi\,(\infty) - b, b\}}{E_{H_0}\left\{\chi'\left(\frac{y\,-\,m_0}{s_0}\right)\left(\frac{y\,-\,m_0}{s(H_0)}\right)\right\}}
\end{aligned}
$$

Problems

Problem 1 *Let H be any given distribution function. Define*

$$H^{-1}(u) \quad = \quad \inf \{y : H(y) \geq u\}.$$

(a) Show that $H^{-1}(H(y)) \leq y$, for all y, with equality if H is strictly increasing on its support.

(b) Show that $H(H^{-1}(u)) \geq u$, for all $0 \leq u \leq 1$, with equality if H is continuous.

Problem 2 *(a) Show that for all H,*

$$\left\{u : H^{-1}(u) \leq y\right\} \quad = \quad \{u : u \leq H(y)\}.$$

(b) Therefore, if

$$U \sim U(0,1)$$

(uniform random variable), then

$$H^{-1}(U) \sim H.$$

(c) Show that $H(Y) \sim U(0,1)$ if and only if H is continuous.

Problem 3 *Let*

$$H_{t,x}(y) \quad = \quad (1-t)\Phi(y) + t\delta_x(y)$$

and let $0 < \alpha < 1$. Show that for all $0 \leq t < \alpha$

$$H_{t,x}^{-1}(\alpha) = \begin{cases} \Phi^{-1}\left(\frac{\alpha-t}{1-t}\right) & x < \Phi^{-1}\left(\frac{\alpha-t}{1-t}\right) \\ x & \Phi^{-1}\left(\frac{\alpha-t}{1-t}\right) \leq x < \Phi^{-1}\left(\frac{\alpha}{1-t}\right) \\ \Phi^{-1}(\frac{\alpha}{1-t}) & x \geq \Phi^{-1}\left(\frac{\alpha}{1-t}\right) \end{cases} \qquad (20)$$

Problem 4 *Write down the functional representation for the following estimates.*

1) The sample standard deviation.

2) The sample Inter Quartile Range R_n.
3) The α-trimmed mean

$$\bar{y}_\alpha = 1/[n(1-\alpha)] \sum_{[n\alpha/2]}^{[n(1-\alpha/2)]} y_{(i)},$$

where $[x]$ represents the largest integer smaller than or equal to x.
4) The sample median.
Give conditions under which the corresponding asymptotic functionals are well defined.

Problem 5 *Assuming for simplicity that H_0 is normal, derive the influence function for the following estimates.*
1) The standard deviation.
2) The Inter Quartile Range.
*[**Hint**: use (20) and the fact that the corresponding asymptotic functional is*

$$R(H) = H^{-1}(3/4) - H^{-1}(1/4)].$$

3) The α-trimmed mean

$$\bar{y}_\alpha = 1/[n(1-\alpha)] \sum_{[n\alpha/2]}^{[n(1-\alpha/2)]} y_{(i)}.$$

*[**Hint:** use (20) and the fact that the corresponding asymptotic functional is*

$$\bar{\mu}_\alpha = \frac{\int_{H^{-1}(\alpha/2)}^{H^{-1}(1-\alpha/2)} y dH(y)}{1-\alpha}.]$$

Problem 6 *Use the influence function to obtain the asymptotic variance under Gaussian and Laplace distributions for the following estimates:*
1) The samole mean.
2) The sample standard deviation.
2) The sample Inter Quartile Range.
3) The α-trimmed mean

$$\bar{y}_\alpha = 1/[n(1-\alpha)] \sum_{[n\alpha/2]}^{[n(1-\alpha/2)]} y_{(i)},$$

where $[x]$ represents the largest integer smaller than or equal to x.
4) The sample median.
5) The dispersion S-estimate with score function χ.
6) The location M-estimate with score function ψ and dispersion $\sigma_n(H)$.

Problem 7 *Write a computer program to estimate the standard error of a Huber's location M-estimate (with* c = 1.345*, say) and dispersion estimate equal to the MAD:*

$$s = 1.5 \times Med\,|y_i - Med\,(y_i)|$$

using (a) the empirical asymptotic variance

$$SE\,(m) = s\sqrt{\frac{\sum \psi_c^2\left(\frac{y_i - m}{s}\right)}{\left[\sum \psi_c'\left(\frac{y_i - m}{s}\right)\right]^2}}$$

and (b) the Jakknife. Perform a small Monte Carlo study to compare these two standard error estimates.

Problem 8 *Derive the influence function* $IF\left(\hat{\beta}, H_0, (x,y)\right)$ *of the least squares estimate of slope*

$$\hat{\beta} = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2},$$

assuming that (x_i, y_i) *are independent and identically distributed with common distribution* $H(x,y)$*. Take* $H_0 = N\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right)$*. Study the dependence of* $IF\left(\hat{\beta}, H_0, (x,y)\right)$ *on* (x,y)*. Study also the dependence of* $IF\left(\hat{\beta}, H_0, (x,y)\right)$ *on* $\left(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\right)$*.*

Problem 9 *Derive the influence function* $IF\left(\hat{\beta}, H_0, (x,y)\right)$ *of the of the (robustified) regression estimates that satisfies the the equation*

$$\sum \psi\left(\frac{y_i - \mathbf{x}_i'\boldsymbol{\beta}}{\hat{\sigma}_n}\right)\mathbf{x}_i = 0$$

where

$$\mathbf{x}_i = \begin{pmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{pmatrix}$$

and $\hat{\sigma}_n$ *is some robust ewstimate of the regression residuals variance. Assume that the influence function of* $\hat{\sigma}_n$ *is given. Briefly state your conclusions.*

Problem 10 *(a) Use the asymptotic variance for the sample stadard deviation derived in Problem ?? to propose an unbiased estimate for the standard error of the sample standard deviation. Try your estimate on several real and synthetic data sets.*

(b)Write a computer code (you may want to use Splus) to calculate the jackknife standard error estimate for the sample standard deviation. Try your code on several real and synthetic data sets.

(c) Perform a small Monte Carlo simulation study to compare the two standard error estimates proposed in parts (a) and (b) above. Try different symmetric and asymmetric population models. What are your conclusions?

(c) Repeat parts (a)-(c) for the standard error of the interquartile range

$$R_n = y_{([\frac{3n}{4}])} - y_{([\frac{n}{4}])}.$$

Problem 11 *Consider the one-step Newton-Raphson dispersion estimate*

$$\hat{\sigma}_1 = \hat{\sigma}_0 \left[1 + \frac{\sum \chi\left((y_i - \hat{\mu})/\hat{\sigma}_0\right) - nb}{\sum \chi'\left((y_i - \hat{\mu})/\hat{\sigma}_0\right)\left((y_i - \hat{\mu})/\hat{\sigma}_0\right)} \right]$$

where $\hat{\mu}_0 = Med(y_i)$ *and* $\hat{\sigma}_0 = MAD(y_i)$*. Show that the influence function of* $\hat{\sigma}_1$ *[take* $H_0 = N(0,1)$*] is equal to that of the fully iterated* $\hat{\sigma}$*. What do you conclude from this result? [**Hint:** use the fact that*

$$IF(y, \hat{\mu}_0, \Phi) = sign(y) / \left(2\varphi(0)\right) \tag{21}$$

and

$$IF(y, \hat{\sigma}_0, \Phi) = \left(sign(y) - 0.5\right) / \left(2\varphi(\Phi^{-1}(0.75))\right)] \tag{22}$$

Problem 12 *Consider the one-step Newton-Raphson location estimate*

$$\hat{\mu}_1 = \hat{\mu}_0 + \hat{\sigma}_0 \frac{\sum \psi\left((y_i - \hat{\mu}_0)/\hat{\sigma}_0\right)}{\sum \psi'\left((y_i - \hat{\mu}_0)/\hat{\sigma}_0\right)}$$

where $\hat{\mu}_0 = Med(y_i)$ *and* $\hat{\sigma}_0 = MAD(y_i)$*. Show that the influence function of* $\hat{\mu}_1$ *[take* $H_0 = N(0,1)$*] is equal to that of the fully iterated* $\hat{\mu}$*. What do you conclude from this result? [**Hint:** use (21) and (22)]*

Problem 13 *Consider again the one-step Newton-Raphson location estimate*

$$\hat{\mu}_1 = \hat{\mu}_0 + \hat{\sigma} \frac{\sum \psi\left((y_i - \hat{\mu}_0)/\hat{\sigma}\right)}{\sum \psi'\left((y_i - \hat{\mu}_0)/\hat{\sigma}\right)}$$

where $\hat{\mu}_0$ *=Med*(y_i) *and*

$$\hat{\sigma} = sd = \sqrt{\frac{1}{n-1} \sum (y_i - \bar{y})^2}$$

(a) Show that the influence function of $\hat{\mu}_1$ *[take* $H_0 = N(0,1)$*] is equal to that of the fully iterated* $\hat{\mu}$*.*

(b) Calculate $\hat{\mu}_1$ *for a symthetic data set made of 1000 independent Standard Normal random variables and one extra outlying ofservation y. Give y values from* 5 *to* 1000 *(steps of 1) and plot y versus* $\hat{\mu}_1$*.*

(c) Briefly state your conclusions from (a) and (b).