So, the influence function of the MLE T is

$$IF(x;T,P)=\begin{pmatrix} x-\mu \\ \frac{[x-\mu]^2}{2\sigma}-\frac{\sigma}{2}\end{pmatrix} \tag{5.45}$$

if P is the $N(\mu,\sigma^2)$ distribution. In fact, one can verify directly that if P is any distribution with finite variance and $T'(P)$ is the standard deviation, then $IF(x;T',P)$ is the second coordinate in (5.45). (See Problem 32 on page 342.)

Example 5.46. For an example that does not meet the smoothness criteria, consider $X_1,\ldots,X_n$ conditionally IID with $U(0,\theta)$ distribution given $\Theta=\theta$. Let $\mathcal{P}_0$ be the class of distributions on $(\mathbb{R},\mathcal{B}^1)$ with bounded support, and let $T(P)$ be the supremum of the support. The MLE is the maximum of the sample, which is the supremum of the support of the empirical CDF, so the MLE is T of the empirical CDF. The influence function for T is $IF(x;T,P)=0$ if $x\le T(P)$ and ∞ if $x>T(p)$. (See Problem 33 on page 342.)

One famous M-estimator of a one-dimensional location parameter is based on the function $\psi(x,\theta)=h(x-\theta)$, where

$$h(t)=\begin{cases} -b & \text{if } t<-b,\\ t & \text{if } -b\le t\le b,\\ b & \text{if } t>b.\end{cases}$$

If P is a continuous distribution, then $\psi(y,\theta)$ is differentiable at $\theta=T(P)$ with probability 1. The influence function will be

$$IF(x;T,P)=\frac{\psi(x,T(P))}{P([T(P)-b,T(P)+b])}.$$

Notice that as $b\to 0$, the influence function approaches the influence function of the median. The finite sample version is the estimator T_n which solves $\sum_{i=1}^n \psi(X_i,T_n)=0$. As with all M-estimators, one can rewrite this equation as $\sum_{i=1}^n w_{n,i}(X_i-T_n)=0$, where $w_{n,i}=\psi(X_i,T_n)/(X_i-T_n)$. It is not difficult to see that $T_n=\sum_{i=1}^n w_{n,i}X_i/\sum_{i=1}^n w_{n,i}$ solves the equation. Since T_n appears on both sides of this equation a solution must be found iteratively. For example, make an initial guess $T_n^{(0)}$ and define

$$\begin{aligned} w_{n,i}^{(k)} &= \frac{\psi(X_i,T_n^{(k-1)})}{X_i-T_n^{(k-1)}} \\ T_n^{(k)} &= \frac{\sum_{i=1}^n w_{n,i}^{(k)}X_i}{\sum_{i=1}^n w_{n,i}^{(k)}}, \end{aligned}$$

for $k\ge 1$ until convergence occurs. In the special case we are considering, the weights $w_{n,i}$ have a nice form: $w_{n,i}=1$ if $|X_i-T_n|\le b$ and $w_{n,i}=b/|X_i-T_n|$ otherwise.

Other types of robust estimators include trimmed means. The $100\alpha\%$ *trimmed mean* of a distribution is the conditional mean given that the observation falls between the α and $1-\alpha$ quantiles of the distribution. That is, if F is the CDF corresponding to P,

$$T(P) = \frac{\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} x dP(x)}{1-2\alpha}$$

for $\alpha < 1/2$. For $\alpha = 1/2$, the tradition is to call the median the 50% trimmed mean. The influence function of a trimmed mean at a continuous distribution is bounded and it has a shape similar to that of the previous estimator. (See Problem 34 on page 342.)

In Section 7.3.6, we will give some results concerning large sample properties of M-estimators. More detailed discussion of robust estimators can be found in the books by Huber (1977, 1981) and Hampel et al. (1986). In Section 8.6.3, we discuss robustness considerations that are peculiar to the Bayesian perspective.

5.2 Set Estimation

A set estimator of a function g of a parameter Θ is a function from the data space $\mathcal{X}$ to a collection of subsets of the space in which $g(\Theta)$ lies.

5.2.1 Confidence Sets

In Section 4.6, we introduced the P-value as an alternative to testing a hypothesis at a fixed level. In nice problems, the P-value gave us the set of all levels at which we could accept the hypothesis. Another alternative is to fix the level and ask for the set of all hypotheses that we could accept at that level. This leads to the concept of a *confidence set.*

Definition 5.47. Let $g : \Omega \to G$ be a function, let η be the collection of all subsets of G, and let $R : \mathcal{X} \to \eta$ be a function. The function R is a *coefficient* γ *confidence set for* $g(\Theta)$ if for every $\theta \in \Omega$,

- $\{x : g(\theta) \in R(x)\}$ is measurable, and
- $P'_\theta(g(\theta) \in R(X)) \geq \gamma$.

The confidence set R is *exact* if $P'_\theta(g(\theta) \in R(X)) = \gamma$ for all $\theta \in \Omega$. If $\inf_{\theta\in\Omega} P'_\theta(g(\theta) \in R(X)) > \gamma$, the confidence set is called *conservative.*[3]

The following result shows how confidence sets relate to nonrandomized tests in general. Its proof is left to the reader.

[3]Some authors require that $P'_\theta(R(X) = \emptyset) = 0$ for all θ before calling R a confidence set. Some require that $\inf_{\theta\in\Omega} P'_\theta(g(\theta) \in R(X)) = \gamma$ before saying that the coefficient is γ.