# 22

# L-Statistics

*In this chapter we prove the asymptotic normality of linear combinations of order statistics, particularly those used for robust estimation or testing, such as trimmed means. We present two methods: The projection method presumes knowledge of Chapter 11 only; the second method is based on the functional delta method of Chapter 20.*

## 22.1 Introduction

Let $X_{n(1)}, \dots, X_{n(n)}$ be the order statistics of a sample of real-valued random variables. A linear combination of (transformed) order statistics, or *$L$-statistic*, is a statistic of the form

$$\sum_{i=1}^{n} c_{ni} a(X_{n(i)}).$$

The coefficients $c_{ni}$ are a triangular array of constants and $a$ is some fixed function. This "score function" can without much loss of generality be taken equal to the identity function, for an $L$-statistic with monotone function $a$ can be viewed as a linear combination of the order statistics of the variables $a(X_1), \dots, a(X_n)$, and an $L$-statistic with a function $a$ of bounded variation can be dealt with similarly, by splitting the $L$-statistic into two parts.

**22.1** ***Example (Trimmed and Winsorized means).*** The simplest example of an $L$-statistic is the sample mean. More interesting are the *$\alpha$-trimmed means*†

$$\frac{1}{n - 2\lfloor \alpha n \rfloor} \sum_{i=\lfloor \alpha n \rfloor + 1}^{n - \lfloor \alpha n \rfloor} X_{n(i)},$$

and the *$\alpha$-Winsorized means*

$$\frac{1}{n}\left[ \lfloor \alpha n \rfloor X_{n(\lfloor \alpha n \rfloor)} + \sum_{i=\lfloor \alpha n \rfloor + 1}^{n - \lfloor \alpha n \rfloor} X_{n(i)} + \lfloor \alpha n \rfloor X_{n(n - \lfloor \alpha n \rfloor + 1)} \right].$$

† The notation $\lfloor x \rfloor$ is used for the greatest integer that is less than or equal to $x$. Also $\lceil x \rceil$ denotes the smallest integer greater than or equal to $x$. For a natural number $n$ and a real number $0 \le x \le n$ one has $\lfloor n - x \rfloor = n - \lceil x \rceil$ and $\lceil n - x \rceil = n - \lfloor x \rfloor$.

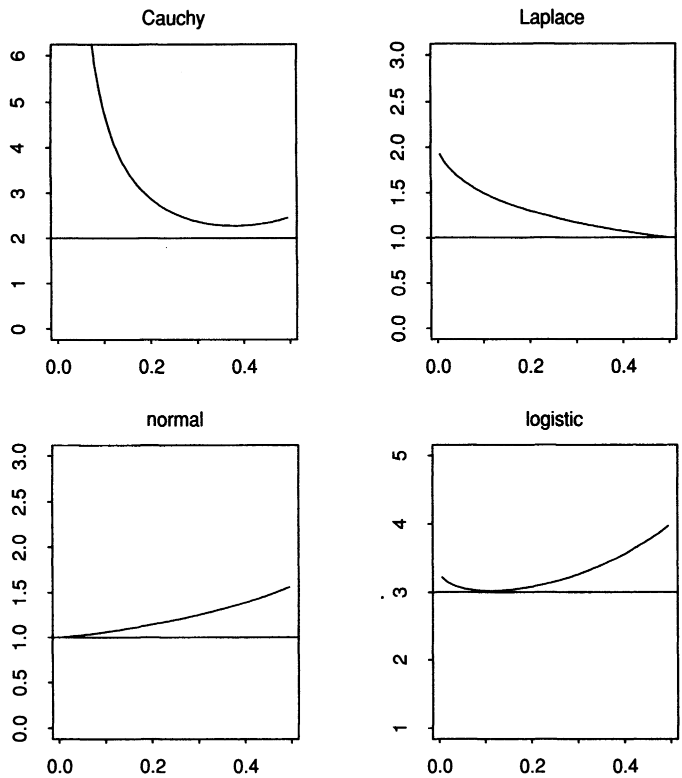


**Figure 22.1.** Asymptotic variance of the $\alpha$-trimmed mean of a sample from a distribution $F$ as function of $\alpha$ for four distributions $F$.

The $\alpha$-trimmed mean is the average of the middle $(1-2\alpha)$-th fraction of the observations, the $\alpha$-Winsorized mean replaces the $\alpha$th fractions of smallest and largest data by $X_{n(\lfloor \alpha n \rfloor)}$ and $X_{n(n-\lfloor \alpha n \rfloor+1)}$, respectively, and next takes the average. Both estimators were already used in the early days of statistics as location estimators in situations in which the data were suspected to contain outliers. Their properties were studied systematically in the context of robust estimation in the 1960s and 1970s. The estimators were shown to have good properties in situations in which the data follows a heavier tailed distribution than the normal one. Figure 22.1 shows the asymptotic variances of the trimmed means as a function of $\alpha$ for four distributions. (A formula for the asymptotic variance is given in Example 22.11.) The four graphs suggest that 10% to 15% trimming may give an improvement over the sample mean in some cases and does not cost much even for the normal distribution. □

**22.2** ***Example (Ranges).*** Two estimators of dispersion are the *interquartile range* $X_{n(\lceil 3n/4 \rceil)} - X_{n(\lceil n/4 \rceil)}$ and the *range* $X_{n(n)} - X_{n(1)}$. Of these, the range does not have a normal limit distribution and is not within the scope of the results of this chapter. □

We present two methods to prove the asymptotic normality of $L$-statistics. The first method is based on the Hájek projection; the second uses the delta method. The second method is preferable in that it applies to more general statistics, but it necessitates the study of empirical processes and does not cover the simplest $L$-statistic: the sample mean.

## 22.2 Hájek Projection

The Hájek projection of a general statistic is discussed in section 11.3. Because a projection is linear and an $L$-statistic is linear in the order statistics, the Hájek projection of an $L$-statistic can be found from the Hájek projections of the individual order statistics. Up to centering at mean zero, these are the sums of the conditional expectations $\mathrm{E}(X_{n(i)} \mid X_k)$ over $k$. Some thought shows that the conditional distribution of $X_{n(i)}$ given $X_k$ is given by

$$\mathrm{P}(X_{n(i)} \leq y \mid X_k = x) = \begin{cases} \mathrm{P}(X_{n-1(i)} \leq y) & \text{if } y < x, \\ \mathrm{P}(X_{n-1(i-1)} \leq y) & \text{if } y \geq x. \end{cases}$$

This is correct for the extreme cases $i = 1$ and $i = n$ provided that we define $X_{n-1(0)} = -\infty$ and $X_{n-1(n)} = \infty$. Thus, we obtain, by the partial integration formula for an expectation, for $x \geq 0$,

$$\begin{aligned} E(X_{n(i)} \mid X_k = x) &= \int_0^x \mathrm{P}(X_{n-1(i)} > y)\,dy + \int_x^\infty \mathrm{P}(X_{n-1(i-1)} > y)\,dy \\ &\quad - \int_{-\infty}^0 \mathrm{P}(X_{n-1(i)} \leq y)\,dy \\ &= -\int_x^\infty \big(\mathrm{P}(X_{n-1(i)} > y) - \mathrm{P}(X_{n-1(i-1)} > y)\big)\,dy + \mathrm{E}X_{n-1(i)}. \end{aligned}$$

The second expression is valid for $x < 0$ as well, as can be seen by a similar argument. Because $X_{n-1(i-1)} \leq X_{n-1(i)}$, the difference between the two probabilities in the last integral is equal to the probability of the event $\{X_{n-1(i-1)} \leq y < X_{n-1(i)}\}$. This is precisely the probability that a binomial $\big(n-1, F(y)\big)$-variable is equal to $i-1$. If we write this probability as $B_{n-1,F(y)}(i-1)$, then the Hájek projection $\hat{X}_{n(i)}$ of $X_{n(i)}$ satisfies, with $\mathbb{F}_n$ the empirical distribution function of $X_1, \ldots, X_n$,

$$\begin{aligned} \hat{X}_{n(i)} - \mathrm{E}\hat{X}_{n(i)} &= -\sum_{k=1}^n \int_{X_k}^\infty B_{n-1,F(y)}(i-1)\,dy + C_n \\ &= -\int n(\mathbb{F}_n - F)(y)\, B_{n-1,F(y)}(i-1)\,dy. \end{aligned}$$

For the projection of the $L$-statistic $T_n = \sum_{i=1}^n c_{ni} X_{n(i)}$ we find

$$\hat{T}_n - \mathrm{E}\hat{T}_n = -\int n(\mathbb{F}_n - F)(y) \sum_{i=1}^n c_{ni}\, B_{n-1,F(y)}(i-1)\,dy.$$

Under some conditions on the coefficients $c_{ni}$, this sum (divided by $\sqrt{n}$) is asymptotically normal by the central limit theorem. Furthermore, the projection $\hat{T}_n$ can be shown to be asymptotically equivalent to the $L$-statistic $T_n$ by Theorem 11.2. Sufficient conditions on the $c_{ni}$ can take a simple appearance for coefficients that are "generated" by a function $\phi$ as in (13.4).

**22.3 Theorem.** *Suppose that $\mathrm{E}X_1^2 < \infty$ and that $c_{ni} = \phi\big(i/(n+1)\big)$ for a bounded function $\phi$ that is continuous at $F(y)$ for Lebesgue almost-every $y$. Then the sequence $n^{-1/2}(T_n - \mathrm{E}T_n)$*

*converges in distribution to a normal distribution with mean zero and variance*

$$\sigma^2(\phi, F) = \iint \phi\big(F(x)\big)\phi\big(F(y)\big)\,\big(F(x \wedge y) - F(x)F(y)\big)\,dx\,dy.$$

***Proof.*** Define functions $e(y) = \phi\big(F(y)\big)$ and

$$e_n(y) = \sum_{i=1}^{n} c_{ni}\, B_{n-1,F(y)}(i-1) = \mathrm{E}\phi\Big(\frac{B_n + 1}{n+1}\Big),$$

for $B_n$ binomially distributed with parameters $\big(n-1, F(y)\big)$. By the law of large numbers $(B_n+1)/(n+1) \xrightarrow{\mathrm{P}} F(y)$. Because $\phi$ is bounded, $e_n(y) \to e(y)$ for every $y$ such that $\phi$ is continuous at $F(y)$, by the dominated-convergence theorem. By assumption, this includes almost every $y$.

By Theorem 11.2, the sequence $n^{-1/2}(T_n - \hat{T}_n)$ converges in second mean to zero if the variances of $n^{-1/2}T_n$ and $n^{-1/2}\hat{T}_n$ converge to the same number. Because $n^{-1/2}(\hat{T}_n - \mathrm{E}\hat{T}_n) = -\int \mathbb{G}_n(y)\, e_n(y)\, dy$, the second variance is easily computed to be

$$\frac{1}{n}\operatorname{var}\hat{T}_n = \iint \big(F(x \wedge y) - F(x)F(y)\big)\, e_n(x)e_n(y)\,dx\,dy.$$

This converges to $\sigma^2(\phi, F)$ by the dominated-convergence theorem. The variance of $n^{-1/2}T_n$ can be written in the form

$$\frac{1}{n}\operatorname{var} T_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ni}c_{nj}\operatorname{cov}(X_{n(i)}, X_{n(j)}) = \iint R_n(x,y)\,dx\,dy,$$

where, because $\operatorname{cov}(X,Y) = \iint \operatorname{cov}\big(\{X \le x\}, \{Y \le y\}\big)\,dx\,dy$ for any pair of variables $(X,Y)$,

$$R_n(x,y) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\phi\Big(\frac{i}{n+1}\Big)\phi\Big(\frac{j}{n+1}\Big)\operatorname{cov}\big(\{X_{n(i)} \le x\}, \{X_{n(j)} \le y\}\big).$$

Because the order statistics are positively correlated, all covariances in the double sum are nonnegative. Furthermore,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\operatorname{cov}\big(\{X_{n(i)} \le x\}, \{X_{n(j)} \le y\}\big) &= \operatorname{cov}\big(\mathbb{G}_n(x), \mathbb{G}_n(y)\big)\\ &= \big(F(x \wedge y) - F(x)F(y)\big).\end{aligned}$$

For pairs $(i,j)$ such that $i \approx nF(x)$ and $j \approx nF(y)$, the coefficient of the covariance is approximately $e(x)e(y)$ by the continuity of $\phi$. The covariances corresponding to other pairs $(i,j)$ are negligible. Indeed, for $i \ge nF(x) + n\varepsilon_n$,

$$\begin{aligned}0 \le \operatorname{cov}\big(\{X_{n(i)} \le x\}, \{X_{n(j)} \le y\}\big) &\le 2\mathrm{P}(X_{n(i)} \le x)\\ &\le 2\mathrm{P}\big(\mathrm{bin}(n, F(x)) \ge nF(x) + n\varepsilon_n\big)\\ &\le 2\exp -2n\varepsilon_n^2,\end{aligned}$$

by Hoeffding's inequality.† Thus, because $\phi$ is bounded, the terms with $i \geq nF(x) + n\varepsilon_n$ contribute exponentially little as $\varepsilon_n \to 0$ not too fast (e.g., $\varepsilon_n^2 = n^{-1/2}$). A similar argument applies to the terms with $i \leq nF(x) - n\varepsilon_n$ or $|j - nF(y)| \geq n\varepsilon_n$. Conclude that, for every $(x, y)$ such that $\phi$ is continuous at both $F(x)$ and $F(y)$,

$$R_n(x, y) \to e(x)e(y)\big(F(x \wedge y) - F(x)F(y)\big).$$

Finally, we apply the dominated convergence theorem to see that the double integral of this expression, which is equal to $n^{-1}\mathrm{var}\, T_n$, converges to $\sigma^2(\phi, F)$.

This concludes the proof that $T_n$ and $\hat{T}_n$ are asymptotically equivalent. To show that the sequence $n^{-1/2}(\hat{T}_n - \mathrm{E}\hat{T}_n)$ is asymptotically normal, define $S_n = -\int \mathbb{G}_n(y)\, e(y)\, dy$. Then, by the same arguments as before, $n^{-1}\mathrm{var}(S_n - \hat{T}_n) \to 0$. Furthermore, the sequence $n^{-1/2}S_n$ is asymptotically normal by the central limit theorem. ∎

## 22.3 Delta Method

The order statistics of a sample $X_1, \ldots, X_n$ can be expressed in their empirical distribution $\mathbb{F}_n$, or rather the empirical quantile function, through

$$\mathbb{F}_n^{-1}(s) = X_{n(\lceil sn \rceil)} = X_{n(i)}, \qquad \text{for } \frac{i-1}{n} < s \leq \frac{i}{n}.$$

Consequently, an $L$-statistic can be expressed in the empirical distribution function as well. Given a fixed function $a$ and a fixed signed measure $K$ on $(0, 1)$‡, consider the function

$$\phi(F) = \int_0^1 a(F^{-1})\, dK.$$

View $\phi$ as a map from the set of distribution functions into $\mathbb{R}$. Clearly,

$$\phi(\mathbb{F}_n) = \sum_{i=1}^{n} K\left(\frac{i-1}{n}, \frac{i}{n}\right] a(X_{n(i)}). \tag{22.4}$$

The right side is an $L$-statistic with coefficients $c_{ni} = K\big((i-1)/n, i/n\big]$. Not all possible arrays of coefficients $c_{ni}$ can be "generated" through a measure $K$ in this manner. However, most $L$-statistics of interest are almost of the form (22.4), so that not much generality is lost by assuming this structure. An advantage is simplicity in the formulation of the asymptotic properties of the statistics, which can be derived with the help of the von Mises method. More importantly, the function $\phi(F)$ can also be applied to other estimators besides $\mathbb{F}_n$. The results of this section yield their asymptotic normality in general.

**22.5** ***Example.*** The $\alpha$-trimmed mean corresponds to the uniform distribution $K$ on the interval $(\alpha, 1-\alpha)$ and $a$ the identity function. More precisely, the $L$-statistic generated by

† See for example, the appendix of [117]. This inequality gives more than needed. For instance, it also works to apply Markov's inequality for fourth moments.

‡ A signed measure is a difference $K = K_1 - K_2$ of two finite measures $K_1$ and $K_2$.

this measure is

$$\frac{1}{1-2\alpha}\int_\alpha^{1-\alpha}\mathbb{F}_n^{-1}(s)\,ds = \frac{1}{n-2\alpha n}\Bigg[(\lceil\alpha n\rceil-\alpha n)X_{n(\lceil\alpha n\rceil)} + \sum_{i=\lceil\alpha n\rceil+1}^{n-\lceil\alpha n\rceil}X_{n(i)} + (\lceil\alpha n\rceil-\alpha n)X_{n(n-\lceil\alpha n\rceil+1)}\Bigg].$$

Except for the slightly different weight factor and the treatment of the two extremes in the averages, this agrees with the $\alpha$-trimmed mean as introduced before. Because $X_{n(k_n)}$ converges in probability to $F^{-1}(p)$ if $k_n/n \to p$ and $\big(n-2\lfloor\alpha n\rfloor\big)/(n-2\alpha n) = 1+O(1/n)$, the difference between the two versions of the trimmed mean can be seen to be $O_P(1/n)$. For the purpose of this chapter this is negligible.

The $\alpha$-Winsorized mean corresponds to the measure $K$ that is the sum of Lebesgue measure on $(\alpha, 1-\alpha)$ and the discrete measure with pointmasses of size $\alpha$ at each of the points $\alpha$ and $1-\alpha$. Again, the difference between the estimator generated by this $K$ and the Winsorized mean is negligible.

The interquartile range corresponds to the discrete, signed measure $K$ that has pointmasses of sizes 1 and $-1$ at the points $1/4$ and $3/4$, respectively. □

Before giving a proof of asymptotic normality, we derive the influence function of an (empirical) $L$-statistic in an informal way. If $F_t = (1-t)F + t\delta_x$, then, by definition, the influence function is the derivative of the map $t \mapsto \phi(F_t)$ at $t = 0$. Provided $a$ and $K$ are sufficiently regular,

$$\frac{d}{dt}\int_0^1 a(F_t^{-1})\,dK = \int_0^1 a'(F_t^{-1})\left[\frac{d}{dt}F_t^{-1}\right]dK.$$

Here the expression within square brackets if evaluated at $t = 0$ is the influence function of the quantile function and is derived in Example 20.5. Substituting the representation given there, we see that the influence function of the $L$-function $\phi(F) = \int a(F^{-1})\,dK$ takes the form

$$\begin{aligned}\phi_F'(\delta_x - F) &= -\int_0^1 a'\big(F^{-1}(u)\big)\,\frac{1_{[x,\infty)}\big(F^{-1}(u)\big)-u}{f\big(F^{-1}(u)\big)}\,dK(u)\\ &= -\int a'(y)\,\frac{1_{[x,\infty)}(y)-F(y)}{f(y)}\,dK\circ F(y).\end{aligned}\tag{22.6}$$

The second equality follows by (a generalization of) the quantile transformation.

An alternative derivation of the influence function starts with rewriting $\phi(F)$ in the form

$$\phi(F) = \int a\,dK\circ F = \int_{(0,\infty)}\overline{(K\circ F)_-}\,da - \int_{(-\infty,0]}(K\circ F)_-\,da. \tag{22.7}$$

Here $\overline{K\circ F}(x) = K\circ F(\infty) - K\circ F(x)$ and the partial integration can be justified for $a$ a function of bounded variation with $a(0) = 0$ (see problem 22.6; the assumption that $a(0) = 0$ simplifies the formula, and is made for convenience). This formula for $\phi(F)$ suggests as influence function

$$\phi_F'(\delta_x - F) = -\int K'\big(F(y)\big)\big(1_{[x,\infty)}(y) - F(y)\big)\,da(y). \tag{22.8}$$

Under appropriate conditions each of the two formulas (22.6) and (22.8) for the influence function is valid. However, already for the defining expressions to make sense very different conditions are needed. Informally, for equation (22.6) it is necessary that $a$ and $F$ be differentiable with a positive derivative for $F$, (22.8) requires that $K$ be differentiable. For this reason both expressions are valuable, and they yield nonoverlapping results.

Corresponding to the two derivations of the influence function, there are two basic approaches towards proving asymptotic normality of $L$-statistics by the delta method, valid under different sets of conditions. Roughly, one approach requires that $F$ and $a$ be smooth, and the other that $K$ be smooth.

The simplest method is to view the $L$-statistic as a function of the empirical quantile function, through the map $\mathbb{F}_n^{-1} \mapsto \int a \circ \mathbb{F}_n^{-1}\, dK$, and next apply the functional delta method to the map $Q \mapsto \int a \circ Q\, dK$. The asymptotic normality of the empirical quantile function is obtained in Chapter 21.

**22.9 Lemma.** *Let $a : \mathbb{R} \mapsto \mathbb{R}$ be continuously differentiable with a bounded derivative. Let $K$ be a signed measure on the interval $(\alpha, \beta) \subset (0, 1)$. Then the map $Q \mapsto \int a(Q)\, dK$ from $\ell^\infty(\alpha, \beta)$ to $\mathbb{R}$ is Hadamard-differentiable at every $Q$. The derivative is the map $H \mapsto \int a'(Q)\, H\, dK$.*

***Proof.*** Let $H_t \to H$ in the uniform norm. Consider the difference

$$\int \left| \frac{a(Q + tH_t) - a(Q)}{t} - a'(Q)\, H \right| dK.$$

The integrand converges to zero everywhere and is bounded uniformly by $\|a'\|_\infty \big(\|H_t\|_\infty + \|H\|_\infty\big)$. Thus the integral converges to zero by the dominated-convergence theorem. ∎

If the underlying distribution has unbounded support, then its quantile function is unbounded on the domain $(0, 1)$, and no estimator can converge in $\ell^\infty(0, 1)$. Then the preceding lemma can apply only to generating measures $K$ with support $(\alpha, \beta)$ strictly within $(0, 1)$. Fortunately, such generating measures are the most interesting ones, as they yield bounded influence functions and hence robust $L$-statistics.

A more serious limitation of using the preceding lemma is that it could require unnecessary smoothness conditions on the distribution of the observations. For instance, the empirical quantile process converges in distribution in $\ell^\infty(\alpha, \beta)$ only if the underlying distribution has a positive density between its $\alpha$- and $\beta$-quantiles. This is true for most standard distributions, but unnecessary for the asymptotic normality of empirical $L$-statistics generated by smooth measures $K$. Thus we present a second lemma that applies to smooth measures $K$ and does not require that $F$ be smooth. Let $DF[-\infty, \infty]$ be the set of all distribution functions.

**22.10 Lemma.** *Let $a : \mathbb{R} \mapsto \mathbb{R}$ be of bounded variation on bounded intervals with $\int (a^+ + a^-)\, d|K \circ F| < \infty$ and $a(0) = 0$. Let $K$ be a signed measure on $(0, 1)$ whose distribution function $K$ is differentiable at $F(x)$ for a almost-every $x$ and satisfies $\big|K(u+h) - K(u)\big| \le M(u)h$ for every sufficiently small $|h|$, and some function $M$ such that $\int M(F_-)\, d|a| < \infty$. Then the map $F \mapsto \int a \circ F^{-1}\, dK$ from $DF[-\infty, \infty] \subset D[-\infty, \infty]$ to $\mathbb{R}$ is Hadamard-differentiable at $F$, with derivative $H \mapsto -\int (K' \circ F_-)\, H\, da$.*

***Proof.*** First rewrite the function in the form (22.7). Suppose that $H_t \to H$ uniformly and set $F_t = F + tH_t$. By continuity of $K$, $(K \circ F)_- = K(F_-)$. Because $K \circ F(\infty) = K(1)$ for all $F$, the difference $\phi(F_t) - \phi(F)$ can be rewritten as $-\int (K \circ F_{t-} - K \circ F_-)\, da$. Consider the integral

$$\int \left| \frac{K(F_- + tH_{t-}) - K(F_-)}{t} - K'(F_-)H \right| d|a|.$$

The integrand converges $a$-almost everywhere to zero and is bounded by $M(F_-)\big(\|H_t\|_\infty + \|H\|_\infty\big) \leq M(F_-)\big(2\|H\|_\infty + 1\big)$, for small $t$. Thus, the lemma follows by the dominated-convergence theorem. ■

Because the two lemmas apply to nonoverlapping situations, it is worthwhile to combine the two approaches. A given generating measure $K$ can be split in its discrete and continuous part. The corresponding two parts of the $L$-statistic can next be shown to be asymptotically linear by application of the two lemmas. Their sum is asymptotically linear as well and hence asymptotically normal.

**22.11** ***Example (Trimmed mean).*** The cumulative distribution function $K$ of the uniform distribution on $(\alpha, 1-\alpha)$ is uniformly Lipschitz and fails to be differentiable only at the points $\alpha$ and $1-\alpha$. Thus, the trimmed-mean function is Hadamard-differentiable at every $F$ such that the set $\{x : F(x) = \alpha, \text{ or } 1-\alpha\}$ has Lebesgue measure zero. (We assume that $\alpha > 0$.) In other words, $F$ should not have flats at height $\alpha$ or $1-\alpha$. For such $F$ the trimmed mean is asymptotically normal with asymptotic influence function $-\int_\alpha^{1-\alpha} \big(1_{x \leq y} - F(y)\big)\, dy$ (see (22.8)), and asymptotic variance

$$\int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \int_{F^{-1}(\alpha)}^{F^{-1}(1-\alpha)} \big(F(x \wedge y) - F(x)\,F(y)\big)\, dx\, dy.$$

Figure 22.1 shows this number as a function of $\alpha$ for a number of distributions. □

**22.12** ***Example (Winsorized mean).*** The generating measure of the Winsorized mean is the sum of a discrete measure on the two points $\alpha$ and $1-\alpha$, and Lebesgue measure on the interval $(\alpha, 1-\alpha)$. The Winsorized mean itself can be decomposed correspondingly. Suppose that the underlying distribution function $F$ has a positive derivative at the points $F^{-1}(\alpha)$ and $F^{-1}(1-\alpha)$. Then the first part of the decomposition is asymptotically linear in view of Lemma 22.9 and Lemma 21.3, the second part is asymptotically linear by Lemma 22.10 and Theorem 19.3. Combined, this yields the asymptotic linearity of the Winsorized mean and hence its asymptotic normality. □

## 22.4 $L$-Estimators for Location

The $\alpha$-trimmed mean and the $\alpha$-Winsorized mean were invented as estimators for location. The question in this section is whether there are still other attractive location estimators within the class of $L$-statistics.

One possible method of generating $L$-estimators for location is to find the best $L$-estimators for given location families $\{f(x-\theta) : \theta \in \mathbb{R}\}$, in which $f$ is some fixed density. For instance, for the $f$ equal to the normal shape this leads to the sample mean.

According to Chapter 8, an estimator sequence $T_n$ is asymptotically optimal for estimating the location of a density with finite Fisher information $I_f$ if

$$\sqrt{n}(T_n - \theta) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{I_f} \frac{f'}{f}(X_i - \theta) + o_P(1).$$

Comparison with equation (22.8) for the influence function of an $L$-statistic shows that the choices of generating measure $K$ and transformation $a$ such that

$$K'\big(F(x-\theta)\big)\, a'(x) = -\Big(\frac{1}{I_f} \frac{f'}{f}(x-\theta)\Big)'$$

lead to an $L$-statistic with the optimal asymptotic influence function. This can be accommodated by setting $a(x) \equiv x$ and

$$K'(u) = -\Big(\frac{1}{I_f} \frac{f'}{f}\Big)' \big(F^{-1}(u)\big).$$

The class of $L$-statistics is apparently large enough to contain an asymptotically efficient estimator sequence for the location parameter of any smooth shape. The $L$-statistics are not as simplistic as they may seem at first.

## Notes

This chapter gives only a few of the many results available on $L$-statistics. For instance, the results on Hadamard differentiability can be refined by using a weighted uniform norm combined with convergence of the weighted empirical process. This allows greater weights for the extreme-order statistics. For further results and references, see [74], [134], and [136].

## PROBLEMS

1. Find a formula for the asymptotic variance of the Winsorized mean.
2. Let $T(F) = \int F^{-1}(u)\, k(u)\, du$.
   (i) Show that $T(F) = 0$ for every distribution $F$ that is symmetric about zero if and only if $k$ is symmetric about $1/2$.
   (ii) Show that $T(F)$ is location equivariant if and only if $\int k(u)\, du = 1$.
   (iii) Show that "efficient" $L$-statistics obtained from symmetric densities possess both properties (i) and (ii).
3. Let $X_1, \dots, X_n$ be a random sample from a continuous distribution function. Show that conditionally on $(X_{n(k)}, X_{n(l)}) = (x, y)$, the variables $X_{n(k+1)}, \dots, X_{n(l-1)}$ are distributed as the order statistics of a random sample of size $l - k - 1$ from the conditional distribution of $X_1$ given that $x \le X_1 \le y$. How can you use this to study the properties of trimmed means?
4. Find an optimal $L$-statistic for estimating the location in the logistic and Laplace location families.
5. Does there exist a distribution for which the trimmed mean is asymptotically optimal for estimating location?

6. **(Partial Integration.)** If $a : \mathbb{R} \mapsto \mathbb{R}$ is right-continuous and nondecreasing with $a(0) = 0$, and $b : \mathbb{R} \mapsto \mathbb{R}$ is right-continuous, nondecreasing and bounded, then

$$\int a\,db = \int_{(0,\infty)} \big(b(\infty) - b_-\big)\,da + \int_{(-\infty,0]} \big(b(-\infty) - b_-\big)\,da.$$

Prove this. If $a$ is also bounded, then the righthand side can be written more succinctly as $ab|_{-\infty}^{\infty} - \int b_-\,da$. (Substitute $a(x) = \int_{(0,x]} da$ for $x > 0$ and $a(x) = -\int_{(x,0]} da$ for $x \leq 0$ into the left side of the equation, and use Fubini's theorem separately on the integral over the positive and negative part of the real line.)