21

Quantiles and Order Statistics

In this chapter we derive the asymptotic distribution of estimators of quantiles from the asymptotic distribution of the corresponding estimators of a distribution function. Empirical quantiles are an example, and hence we also discuss some results concerning order statistics. Furthermore, we discuss the asymptotics of the median absolute deviation, which is the empirical 1/2-quantile of the observations centered at their 1/2-quantile.

21.1 Weak Consistency

The *quantile function* of a cumulative distribution function F is the generalized inverse $F^{-1}: (0, 1) \mapsto \mathbb{R}$ given by

$$F^{-1}(p) = \inf\bigl\{x: F(x) \geq p\bigr\}.$$

It is a left-continuous function with range equal to the support of F and hence is often unbounded. The following lemma records some useful properties.

21.1 ***Lemma.*** *For every $0 < p < 1$ and $x \in \mathbb{R}$,*

- *(i) $F^{-1}(p) \leq x$ iff $p \leq F(x)$;*
- *(ii) $F \circ F^{-1}(p) \geq p$ with equality iff p is in the range of F; equality can fail only if F is discontinuous at $F^{-1}(p)$;*
- *(iii) $F_- \circ F^{-1}(p) \leq p$;*
- *(iv) $F^{-1} \circ F(x) \leq x$; equality fails iff x is in the interior or at the right end of a "flat" of F;*
- *(v) $F^{-1} \circ F \circ F^{-1} = F^{-1}$; $F \circ F^{-1} \circ F = F$;*
- *(vi) $(F \circ G)^{-1} = G^{-1} \circ F^{-1}$.*

Proof. The proofs of the inequalities in (i) through (iv) are best given by a picture. The equalities (v) follow from (ii) and (iv) and the monotonicity of F and F^{-1}. If $p = F(x)$ for some x, then, by (ii) $p \leq F \circ F^{-1}(p) = F \circ F^{-1} \circ F(x) = F(x) = p$, by (iv). This proves the first statement in (ii); the second is immediate from the inequalities in (ii) and (iii). Statement (vi) follows from (i) and the definition of $(F \circ G)^{-1}$. ∎

Consequences of (ii) and (iv) are that $F \circ F^{-1}(p) \equiv p$ on $(0, 1)$ if and only if F is continuous (i.e., has range $[0, 1]$), and $F^{-1} \circ F(x) \equiv x$ on $\mathbb{R}$ if and only if F is strictly increasing (i.e., has no "flats"). Thus F^{-1} is a proper inverse if and only if F is both continuous and strictly increasing, as one would expect.

By (i) the random variable $F^{-1}(U)$ has distribution function F if U is uniformly distributed on $[0, 1]$. This is called the *quantile transformation*. On the other hand, by (i) and (ii) the variable $F(X)$ is uniformly distributed on $[0, 1]$ if and only if X has a continuous distribution function F. This is called the *probability integral transformation*.

A sequence of quantile functions is defined to *converge weakly* to a limit quantile function, denoted $F_n^{-1} \rightsquigarrow F^{-1}$, if and only if $F_n^{-1}(t) \to F^{-1}(t)$ at every t where F^{-1} is continuous. This type of convergence is not only analogous in form to the weak convergence of distribution functions, it is the same.

21.2 Lemma. *For any sequence of cumulative distribution functions, $F_n^{-1} \rightsquigarrow F^{-1}$ if and only if $F_n \rightsquigarrow F$.*

Proof. Let U be uniformly distributed on $[0, 1]$. Because F^{-1} has at most countably many discontinuity points, $F_n^{-1} \rightsquigarrow F^{-1}$ implies that $F_n^{-1}(U) \to F^{-1}(U)$ almost surely. Consequently, $F_n^{-1}(U)$ converges in law to $F^{-1}(U)$, which is exactly $F_n \rightsquigarrow F$ by the quantile transformation.

For a proof the converse, let V be a normally distributed random variable. If $F_n \rightsquigarrow F$, then $F_n(V) \xrightarrow{\text{as}} F(V)$, because convergence can fail only at discontinuity points of F. Thus $\Phi\big(F_n^{-1}(t)\big) = \mathrm{P}\big(F_n(V) < t\big)$ (by (i) of the preceding lemma) converges to $\mathrm{P}\big(F(V) < t\big) = \Phi\big(F^{-1}(t)\big)$ at every t at which the limit function is continuous. This includes every t at which F^{-1} is continuous. By the continuity of Φ^{-1}, $F_n^{-1}(t) \to F^{-1}(t)$ for every such t. ∎

A statistical application of the preceding lemma is as follows. If a sequence of estimators $\hat{F}_n$ of a distribution function F is weakly consistent, then the sequence of estimators $\hat{F}_n^{-1}$ is weakly consistent for the quantile function F^{-1}.

21.2 Asymptotic Normality

In the absence of information concerning the underlying distribution function F of a sample, the empirical distribution function $\mathbb{F}_n$ and empirical quantile function $\mathbb{F}_n^{-1}$ are reasonable estimators for F and F^{-1}, respectively. The empirical quantile function is related to the order statistics $X_{n(1)}, \dots, X_{n(n)}$ of the sample through

$$\mathbb{F}_n^{-1}(p) = X_{n(i)}, \qquad \text{for } p \in \left(\frac{i-1}{n}, \frac{i}{n}\right].$$

One method to prove the asymptotic normality of empirical quantiles is to view them as M-estimators and apply the theorems given in Chapter 5. Another possibility is to express the distribution function $\mathrm{P}(X_{n(i)} \le x)$ into binomial probabilities and apply approximations to these. The method that we follow in this chapter is to deduce the asymptotic normality of quantiles from the asymptotic normality of the distribution function, using the delta method.

An advantage of this method is that it is not restricted to empirical quantiles but applies to the quantiles of any estimator of the distribution function.

For a nondecreasing function $F \in D[a, b]$, $[a, b] \subset [-\infty, \infty]$, and a fixed $p \in \mathbb{R}$, let $\phi(F) \in [a, b]$ be an arbitrary point in $[a, b]$ such that

$$F\big(\phi(F)-\big) \le p \le F\big(\phi(F)\big).$$

The natural domain $\mathbb{D}_\phi$ of the resulting map ϕ is the set of all nondecreasing F such that there exists a solution to the pair of inequalities. If there exists more than one solution, then the precise choice of $\phi(F)$ is irrelevant. In particular, $\phi(F)$ may be taken equal to the pth quantile $F^{-1}(p)$.

21.3 ***Lemma.*** *Let $F \in \mathbb{D}_\phi$ be differentiable at a point $\xi_p \in (a, b)$ such that $F(\xi_p) = p$, with positive derivative. Then $\phi : \mathbb{D}_\phi \subset D[a, b] \mapsto \mathbb{R}$ is Hadamard-differentiable at F tangentially to the set of functions $h \in D[a, b]$ that are continuous at ξ_p, with derivative $\phi'_F(h) = -h(\xi_p)/F'(\xi_p)$.*

Proof. Let $h_t \to h$ uniformly on $[a, b]$ for a function h that is continuous at ξ_p. Write ξ_{pt} for $\phi(F + th_t)$. By the definition of ϕ, for every $\varepsilon_t > 0$,

$$(F + th_t)(\xi_{pt} - \varepsilon_t) \le p \le (F + th_t)(\xi_{pt}).$$

Choose ε_t positive and such that $\varepsilon_t = o(t)$. Because the sequence h_t converges uniformly to a bounded function, it is uniformly bounded. Conclude that $F(\xi_{pt} - \varepsilon_t) + O(t) \le p \le F(\xi_{pt}) + O(t)$. By assumption, the function F is monotone and bounded away from p outside any interval $(\xi_p - \varepsilon, \xi_p + \varepsilon)$ around ξ_p. To satisfy the preceding inequalities the numbers ξ_{pt} must be to the right of $\xi_p - \varepsilon$ eventually, and the numbers $\xi_{pt} - \varepsilon_t$ must be to the left of $\xi_p + \varepsilon$ eventually. In other words, $\xi_{pt} \to \xi_p$.

By the uniform convergence of h_t and the continuity of the limit, $h_t(\xi_{pt} - \varepsilon_t) \to h(\xi_p)$ for every $\varepsilon_t \to 0$. Using this and Taylor's formula on the preceding display yields

$$\begin{aligned} p + (\xi_{pt} - \xi_p)F'(\xi_p) - o(\xi_{pt} - \xi_p) + O(\varepsilon_t) + th(\xi_p) - o(t) \\ \le p \le p + (\xi_{pt} - \xi_p)F'(\xi_p) + o(\xi_{pt} - \xi_p) + O(\varepsilon_t) + th(\xi_p) + o(t). \end{aligned}$$

Conclude first that $\xi_{pt} - \xi_p = O(t)$. Next, use this to replace the $o(\xi_{pt} - \xi_p)$ terms in the display by $o(t)$ terms and conclude that $(\xi_{pt} - \xi_p)/t \to -(h/F')(\xi_p)$. ■

Instead of a single quantile we can consider the *quantile function* $F \mapsto \big(F^{-1}(p)\big)_{p_1 < p < p_2}$, for fixed numbers $0 \le p_1 < p_2 \le 1$. Because any quantile function is bounded on an interval $[p_1, p_2]$ strictly contained in $(0, 1)$, we may hope that a quantile estimator converges in distribution in $\ell^\infty(p_1, p_2)$ for such an interval. The quantile function of a distribution with compact support is bounded on the whole interval $(0, 1)$, and then we may hope to strengthen the result to weak convergence in $\ell^\infty(0, 1)$.

Given an interval $[a, b] \subset \mathbb{R}$, let $\mathbb{D}_1$ be the set of all restrictions of distribution functions on $\mathbb{R}$ to $[a, b]$, and let $\mathbb{D}_2$ be the subset of $\mathbb{D}_1$ of distribution functions of measures that give mass 1 to $(a, b]$.

21.4 Lemma.
(i) *Let $0 < p_1 < p_2 < 1$, and let F be continuously differentiable on the interval $[a, b] = \big[F^{-1}(p_1) - \varepsilon, F^{-1}(p_2) + \varepsilon\big]$ for some $\varepsilon > 0$, with strictly positive derivative f. Then the inverse map $G \mapsto G^{-1}$ as a map $\mathbb{D}_1 \subset D[a, b] \mapsto \ell^\infty[p_1, p_2]$ is Hadamard differentiable at F tangentially to $C[a, b]$.*
(ii) *Let F have compact support $[a, b]$ and be continuously differentiable on its support with strictly positive derivative f. Then the inverse map $G \mapsto G^{-1}$ as a map $\mathbb{D}_2 \subset D[a, b] \mapsto \ell^\infty(0, 1)$ is Hadamard differentiable at F tangentially to $C[a, b]$.*
In both cases the derivative is the map $h \mapsto -(h/f) \circ F^{-1}$.

Proof. It suffices to make the proof of the preceding lemma uniform in p. We use the same notation.

(i). Because the function F has a positive density, it is strictly increasing on an interval $[\xi_{p_1'}, \xi_{p_2'}]$ that strictly contains $[\xi_{p_1}, \xi_{p_2}]$. Then on $[p_1', p_2']$ the quantile function F^{-1} is the ordinary inverse of F and is (uniformly) continuous and strictly increasing. Let $h_t \to h$ uniformly on $[\xi_{p_1'}, \xi_{p_2'}]$ for a continuous function h. By the proof of the preceding lemma, $\xi_{p_i t} \to \xi_{p_i}$ and hence every ξ_{pt} for $p_1 \le p \le p_2$ is contained in $[\xi_{p_1'}, \xi_{p_2'}]$ eventually. The remainder of the proof is the same as the proof of the preceding lemma.

(ii). Let $h_t \to h$ uniformly in $D[a, b]$, where h is continuous and $F + th_t$ is contained in $\mathbb{D}_2$ for all t. Abbreviate $F^{-1}(p)$ and $(F + th_t)^{-1}(p)$ to ξ_p and ξ_{pt}, respectively. Because F and $F + th_t$ are concentrated on $(a, b]$ by assumption, we have $a < \xi_{pt}, \xi_p \le b$ for all $0 < p < 1$. Thus the numbers $\varepsilon_{pt} = t^2 \wedge (\xi_{pt} - a)$ are positive, whence, by definition,

$$(F + th_t)(\xi_{pt} - \varepsilon_{pt}) \le p \le (F + th_t)(\xi_{pt}).$$

By the smoothness of F we have $F(\xi_p) = p$ and $F(\xi_{pt} - \varepsilon_{pt}) = F(\xi_{pt}) + O(\varepsilon_{pt})$, uniformly in $0 < p < 1$. It follows that

$$-th(\xi_{pt}) + o(t) \le F(\xi_{pt}) - F(\xi_p) \le -th(\xi_{pt} - \varepsilon_{pt}) + o(t).$$

The $o(t)$ terms are uniform in $0 < p < 1$. The far left side and the far right side are $O(t)$; the expression in the middle is bounded above and below by a constant times $|\xi_{pt} - \xi_p|$. Conclude that $|\xi_{pt} - \xi_p| = O(t)$, uniformly in p. Next, the lemma follows by the uniform differentiability of F. ∎

Thus, the asymptotic normality of an estimator of a distribution function (or another nondecreasing function) automatically entails the asymptotic normality of the corresponding quantile estimators. More precisely, to derive the asymptotic normality of even a single quantile estimator $\hat{F}_n^{-1}(p)$, we need to know that the estimators $\hat{F}_n$ are asymptotically normal as a process, in a neighborhood of $F^{-1}(p)$. The standardized empirical distribution function is asymptotically normal as a process indexed by $\mathbb{R}$, and hence the empirical quantiles are asymptotically normal.

21.5 Corollary. *Fix $0 < p < 1$. If F is differentiable at $F^{-1}(p)$ with positive derivative $f\big(F^{-1}(p)\big)$, then*

$$\sqrt{n}\big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\big) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \frac{1\big\{X_i \le F^{-1}(p)\big\} - p}{f\big(F^{-1}(p)\big)} + o_P(1).$$

Consequently, the sequence $\sqrt{n}\big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\big)$ is asymptotically normal with mean 0 and variance $p(1-p)/f^2\big(F^{-1}(p)\big)$. Furthermore, if F satisfies the conditions (i) or (ii) of the preceding lemma, then $\sqrt{n}(\mathbb{F}_n^{-1} - F^{-1})$ converges in distribution in $\ell^\infty[p_1, p_2]$ or $\ell^\infty(0,1)$, respectively, to the process $\mathbb{G}_\lambda/f\big(F^{-1}(p)\big)$, where $\mathbb{G}_\lambda$ is a standard Brownian bridge.

Proof. By Theorem 19.3, the empirical process $\mathbb{G}_{n,F} = \sqrt{n}(\mathbb{F}_n - F)$ converges in distribution in $D[-\infty,\infty]$ to an F-Brownian bridge process $\mathbb{G}_F = \mathbb{G}_\lambda \circ F$. The sample paths of the limit process are continuous at the points at which F is continuous. By Lemma 21.3, the quantile function $F \mapsto F^{-1}(p)$ is Hadamard-differentiable tangentially to the range of the limit process. By the functional delta method, the sequence $\sqrt{n}\big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\big)$ is asymptotically equivalent to the derivative of the quantile function evaluated at $\mathbb{G}_{n,F}$, that is, to $-\mathbb{G}_{n,F}\big(F^{-1}(p)\big)/f\big(F^{-1}(p)\big)$. This is the first assertion. Next, the asymptotic normality of the sequence $\sqrt{n}\big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\big)$ follows by the central limit theorem.

The convergence of the quantile process follows similarly, this time using Lemma 21.4. ■

21.6 ***Example.*** The uniform distribution function has derivative 1 on its compact support. Thus, the uniform empirical quantile process converges weakly in $\ell^\infty(0,1)$. The limiting process is a standard Brownian bridge.

The normal and Cauchy distribution functions have continuous derivatives that are bounded away from zero on any compact interval. Thus, the normal and Cauchy empirical quantile processes converge in $\ell^\infty[p_1, p_2]$, for every $0 < p_1 < p_2 < 1$. □

The empirical quantile function at a point is equal to an order statistic of the sample. In estimating a quantile, we could also use the order statistics directly, not necessarily in the way that $\mathbb{F}_n^{-1}$ picks them. For the k_n-th order statistic $X_{n(k_n)}$ to be a consistent estimator for $F^{-1}(p)$, we need minimally that $k_n/n \to p$ as $n \to \infty$. For mean-zero asymptotic normality, we also need that $k_n/n \to p$ faster than $1/\sqrt{n}$, which is necessary to ensure that $X_{n(k_n)}$ and $\mathbb{F}_n^{-1}(p)$ are asymptotically equivalent. This still allows considerable freedom for choosing k_n.

21.7 ***Lemma.*** *Let F be differentiable at $F^{-1}(p)$ with positive derivative and let $k_n/n = p + c/\sqrt{n} + o(1/\sqrt{n})$. Then*

$$\sqrt{n}\big(X_{n(k_n)} - \mathbb{F}_n^{-1}(p)\big) \xrightarrow{\mathrm{P}} \frac{c}{f\big(F^{-1}(p)\big)}.$$

Proof. First assume that F is the uniform distribution function. Denote the observations by U_i, rather than X_i. Define a function $g_n\colon \ell^\infty(0,1) \mapsto \mathbb{R}$ by $g_n(z) = z(k_n/n) - z(p)$. Then $g_n(z_n) \to z(p) - z(p) = 0$, whenever $z_n \to z$ for a function z that is continuous at p. Because the uniform quantile process $\sqrt{n}(\mathbb{G}_n^{-1} - G^{-1})$ converges in distribution in $\ell^\infty(0,1)$, the extended continuous-mapping theorem, Theorem 18.11, yields $g_n\big(\sqrt{n}(\mathbb{G}_n^{-1} - G^{-1})\big) = \sqrt{n}\big(U_{n(k_n)} - \mathbb{G}_n^{-1}(p)\big) - \sqrt{n}(k_n/n - p) \rightsquigarrow 0$. This is the result in the uniform case.

A sample from a general distribution function F can be generated as $F^{-1}(U_i)$, by the quantile transformation. Then $\sqrt{n}\big(X_{n(k_n)} - \mathbb{F}_n^{-1}(p)\big)$ is equal to

$$\sqrt{n}\big[F^{-1}(U_{n(k_n)}) - F^{-1}(p)\big] - \sqrt{n}\Big[F^{-1}\big(\mathbb{G}_n^{-1}(p)\big) - F^{-1}(p)\Big].$$

Apply the delta method to the two terms to see that $f\big(F^{-1}(p)\big)$ times their difference is asymptotically equivalent to $\sqrt{n}(U_{n(k_n)} - p) - \sqrt{n}\big(\mathbb{G}_n^{-1}(p) - p\big)$. ■

21.8 ***Example (Confidence intervals for quantiles).*** If $X_1, \ldots, X_n$ is a random sample from a continuous distribution function F, then $U_1 = F(X_1), \ldots, U_n = F(X_n)$ are a random sample from the uniform distribution, by the probability integral transformation. This can be used to construct confidence intervals for quantiles that are distribution-free over the class of continuous distribution functions. For any given natural numbers k and l, the interval $(X_{n(k)}, X_{n(l)}]$ has coverage probability

$$\mathrm{P}_F\big(X_{n(k)} < F^{-1}(p) \leq X_{n(l)}\big) = \mathrm{P}(U_{n(k)} < p \leq U_{n(l)}).$$

Because this is independent of F, it is possible to obtain an exact confidence interval for every fixed n, by determining k and l to achieve the desired confidence level. (Here we have some freedom in choosing k and l but can obtain only finitely many confidence levels.) For large n, the values k and l can be chosen equal to

$$\frac{k, l}{n} = p \pm z_\alpha \sqrt{\frac{p(1-p)}{n}}.$$

To see this, note that, by the preceding lemma,

$$U_{n(k)}, U_{n(l)} = \frac{\mathbb{G}_n^{-1}(p)}{\sqrt{n}} \pm z_\alpha \sqrt{\frac{p(1-p)}{n}} + o_P\Big(\frac{1}{\sqrt{n}}\Big).$$

Thus the event $U_{n(k)} < p \leq U_{n(l)}$ is asymptotically equivalent to the event $\sqrt{n}\big|\mathbb{G}_n^{-1}(p) - p\big| \leq z_\alpha \sqrt{p(1-p)}$. Its probability converges to $1 - 2\alpha$.

An alternative is to use the asymptotic normality of the empirical quantiles $\mathbb{F}_n^{-1}$, but this has the unattractive feature of having to estimate the density $f\big(F^{-1}(p)\big)$, because this appears in the denominator of the asymptotic variance. If using the distribution-free method, we do not even have to assume that the density exists. □

The application of the Hadamard differentiability of the quantile transformation is not limited to empirical quantiles. For instance, we can also immediately obtain the asymptotic normality of the quantiles of the product limit estimator, or any other estimator of a distribution function in semiparametric models. On the other hand, the results on empirical quantiles can be considerably strengthened by taking special properties of the empirical distribution into account. We discuss a few extensions, mostly for curiosity value.†

Corollary 21.5 asserts that $R_n(p) \xrightarrow{\mathrm{P}} 0$, for, with $\xi_p = F^{-1}(p)$,

$$R_n(p) = f(\xi_p)\sqrt{n}\big(\mathbb{F}_n^{-1}(p) - F^{-1}(p)\big) + \sqrt{n}\big(\mathbb{F}_n(\xi_p) - F(\xi_p)\big).$$

The expression on the left is known as the standardized *empirical difference process*. "Standardized" refers to the leading factor $f(\xi_p)$. That a sum is called a difference is curious but stems from the fact that minus the second term is approximately equal to the first term. The identity shows an interesting symmetry between the empirical distribution and quantile

† See [134, pp. 586–587] for further information.

processes, particularly in the case that F is uniform, if $f(\xi_p) \equiv 1$ and $\xi_p \equiv p$. The result that $R_n(p) \xrightarrow{\mathrm{P}} 0$ can be refined considerably. If F is twice-differentiable at ξ_p with positive first derivative, then, by the *Bahadur-Kiefer theorems*,

$$\limsup_{n\to\infty} \frac{n^{1/4}}{(\log\log n)^{3/4}} \big|R_n(p)\big| = \left[\frac{32}{27}\, p(1-p)\right]^{1/4}, \qquad \text{a.s.},$$
$$n^{1/4} R_n(p) \rightsquigarrow \frac{2}{\sqrt{p(1-p)}} \int_0^\infty \Phi\Big(\frac{x}{\sqrt{y}}\Big)\, \phi\Big(\frac{y}{\sqrt{p(1-p)}}\Big)\, dy.$$

The right side in the last display is a distribution function as a function of the argument x. Thus, the magnitude of the empirical difference process is $O_P(n^{-1/4})$, with the rate of its fluctuations being equal to $n^{-1/4}(\log\log n)^{3/4}$. Under some regularity conditions on F, which are satisfied by, for instance, the uniform, the normal, the exponential, and the logistic distribution, versions of the preceding results are also valid in supremum norm,

$$\limsup_{n\to\infty} \frac{n^{1/4}}{(\log n)^{1/2}(2\log\log n)^{1/4}} \|R_n\|_\infty = \frac{1}{\sqrt{2}}, \qquad \text{a.s.},$$
$$\frac{n^{1/4}}{(\log n)^{1/2}} \|R_n\|_\infty \rightsquigarrow \sqrt{\|\mathbb{Z}_\lambda\|_\infty}.$$

Here $\mathbb{Z}_\lambda$ is a standard Brownian motion indexed by the interval $[0, 1]$.

21.3 Median Absolute Deviation

The *median absolute deviation* of a sample $X_1, \ldots, X_n$ is the robust estimator of scale defined by

$$\mathrm{MAD}_n = \operatorname*{med}_{1\le i\le n} \Big| X_i - \operatorname*{med}_{1\le i\le n} X_i \Big|.$$

It is the median of the deviations of the observations from their median and is often recommended for reducing the observations to a standard scale as a first step in a robust procedure. Because the median is a quantile, we can prove the asymptotic normality of the median absolute deviation by the delta method for quantiles, applied twice.

If a variable X has distribution function F, then the variable $|X-\theta|$ has the distribution function $x \mapsto F(\theta + x) - F_-(\theta - x)$. Let $(\theta, F) \mapsto \phi_2(\theta, F)$ be the map that assigns to a given number θ and a given distribution function F the distribution function $F(\theta+x) - F_-(\theta - x)$, and consider the function $\phi = \phi_3 \circ \phi_2 \circ \phi_1$ defined by

$$F \overset{\phi_1}{\mapsto} \big(\theta\colon= F^{-1}(1/2), F\big) \overset{\phi_2}{\mapsto} G\colon= F(\theta + \cdot) - F_-(\theta - \cdot) \overset{\phi_3}{\mapsto} G^{-1}(1/2).$$

If we identify the median with the $1/2$-quantile, then the median absolute deviation is exactly $\phi(\mathbb{F}_n)$. Its asymptotic normality follows by the delta method under a regularity condition on the underlying distribution.

21.9 ***Lemma.*** *Let the numbers m_F and m_G satisfy $F(m_F) = \frac{1}{2} = F(m_F + m_G) - F(m_F - m_G)$. Suppose that F is differentiable at m_F with positive derivative and is continuously differentiable on neighborhoods of $m_F + m_G$ and $m_F - m_G$ with positive derivative at*

$m_F + m_G$ and/or $m_F - m_G$. Then the map ϕ: $D[-\infty,\infty] \mapsto \mathbb{R}$, with as domain the distribution functions, is Hadamard-differentiable at F, tangentially to the set of functions that are continuous both at m_F and on neighborhoods of $m_F + m_G$ and $m_F - m_G$. The derivative $\phi'_F(H)$ is given by

$$\frac{H(m_F)}{f(m_F)}\frac{f(m_F+m_G)-f(m_F-m_G)}{f(m_F+m_G)+f(m_F-m_G)} - \frac{H(m_F+m_G)-H(m_F-m_G)}{f(m_F+m_G)+f(m_F-m_G)}.$$

Proof. Define the maps ϕ_i as indicated previously.

By Lemma 21.3, the map ϕ_1: $D[-\infty,\infty] \mapsto \mathbb{R} \times D[-\infty,\infty]$ is Hadamard-differentiable at F tangentially to the set of functions H that are continuous at m_F.

The map ϕ_2: $\mathbb{R} \times D[-\infty,\infty] \mapsto D[m_G - \varepsilon, m_G + \varepsilon]$ is Hadamard-differentiable at the point (m_F, F) tangentially to the set of points (g, H) such that H is continuous on the intervals $[m_F \pm m_G - 2\varepsilon, m_F \pm m_G + 2\varepsilon]$, for sufficiently small $\varepsilon > 0$. This follows because, if $a_t \to a$ and $H_t \to H$ uniformly,

$$\frac{(F+tH_t)(m_F+ta_t+x) - F(m_F+x)}{t} \to af(m_F+x) + H(m_F+x),$$

uniformly in $x \approx m_G$, and because a similar statement is valid for the differences $(F + tH_t)_-(m_F + ta_t - x) - F_-(m_F - x)$. The range of the derivative is contained in $C[m_G - \varepsilon, m_G + \varepsilon]$.

Finally, by Lemma 21.3, the map $\phi_3 : D[m_G - \varepsilon, m_G + \varepsilon] \mapsto \mathbb{R}$ is Hadamard-differentiable at $G = \phi_2(m_F, F)$, tangentially to the set of functions that are continuous at m_G, because G has a positive derivative at its median, by assumption.

The lemma follows by the chain rule, where we ascertain that the tangent spaces match up properly. ■

The F-Brownian bridge process $\mathbb{G}_F$ has sample paths that are continuous everywhere that F is continuous. Under the conditions of the lemma, they are continuous at the point m_F and in neighborhoods of the points $m_F + m_G$ and $m_F - m_G$. Thus, in view of the lemma and the delta method, the sequence $\sqrt{n}\big(\phi(\mathbb{F}_n) - \phi(F)\big)$ converges in distribution to the variable $\phi'_F(\mathbb{G}_F)$.

21.10 ***Example (Symmetric F).*** If F has a density that is symmetric about 0, then its median m_F is 0 and the median absolute deviation m_G is equal to $F^{-1}(3/4)$. Then the first term in the definition of the derivative vanishes, and the derivative $\phi'_F(\mathbb{G}_F)$ at the F-Brownian bridge reduces to $-\big(\mathbb{G}_\lambda(3/4) - \mathbb{G}_\lambda(1/4)\big)/2f\big(F^{-1}(3/4)\big)$ for a standard Brownian bridge $\mathbb{G}_\lambda$. Then the asymptotic variance of $\sqrt{n}(\mathrm{MAD}_n - m_G)$ is equal to $(1/16)/f \circ F^{-1}(3/4)^2$. □

21.11 ***Example (Normal distribution).*** If F is equal to the normal distribution with mean zero and variance σ^2, then $m_F = 0$ and $m_G = \sigma\Phi^{-1}(3/4)$. We find an asymptotic variance $(\sigma^2/16)\phi \circ \Phi^{-1}(3/4)^{-2}$. As an estimator for the standard deviation σ, we use the estimator $\mathrm{MAD}_n/\Phi^{-1}(3/4)$, and as an estimator for σ^2 the square of this. By the delta method, the latter estimator has asymptotic variance equal to $(1/4)\sigma^4\phi \circ \Phi^{-1}(3/4)^{-2}\Phi^{-1}(3/4)^{-2}$, which is approximately equal to $5.44\sigma^4$. The relative efficiency, relative to the sample variance, is approximately equal to 37%, and hence we should not use this estimator without a good reason. □

21.4 Extreme Values

The asymptotic behavior of order statistics $X_{n(k_n)}$ such that $k_n/n \to 0$ or 1 is, of course, different from that of central-order statistics. Because $X_{n(k_n)} \le x_n$ means that at most $n - k_n$ of the X_i can be bigger than x_n, it follows that, with $p_n = \mathrm{P}(X_i > x_n)$,

$$\Pr(X_{n(k_n)} \le x_n) = \mathrm{P}\big(\mathrm{bin}(n, p_n) \le n - k_n\big).$$

Therefore, limit distributions of general-order statistics can be derived from approximations to the binomial distribution. In this section we consider the most extreme cases, in which $k_n = n - k$ for a fixed number k, starting with the maximum $X_{n(n)}$. We write $\overline{F}(t) = \mathrm{P}(X_i > t)$ for the survival distribution of the observations, a random sample of size n from F.

The distribution function of the maximum can be derived from the preceding display, or directly, and satisfies

$$\mathrm{P}(X_{n(n)} \le x_n) = F(x_n)^n = \left(1 - \frac{n\overline{F}(x_n)}{n}\right)^n.$$

This representation readily yields the following lemma.

21.12 ***Lemma.*** *For any sequence of numbers x_n and any $\tau \in [0, \infty]$, we have $\mathrm{P}(X_{n(n)} \le x_n) \to e^{-\tau}$ if and only if $n\overline{F}(x_n) \to \tau$.*

In view of the lemma we can find "interesting limits" for the probabilities $\mathrm{P}(X_{n(n)} \le x_n)$ only for sequences x_n such that $\overline{F}(x_n) = O(1/n)$. Depending on F this may mean that x_n is bounded or converges to infinity.

Suppose that we wish to find constants a_n and $b_n > 0$ such that $b_n^{-1}(X_{n(n)} - a_n)$ converges in distribution to a nontrivial limit. Then we must choose a_n and b_n such that $\overline{F}(a_n + b_n x) = O(1/n)$ for a nontrivial set of x. Depending on F such constants may or may not exist. It is a bit surprising that the set of possible limit distributions is extremely small.†

21.13 ***Theorem (Extremal types).*** *If $b_n^{-1}(X_{n(n)} - a_n) \rightsquigarrow G$ for a nondegenerate cumulative distribution function G, then G belongs to the location-scale family of a distribution of one of the following forms:*
(i) $e^{-e^{-x}}$ with support $\mathbb{R}$;
(ii) $e^{-(1/x^\alpha)}$ with support $[0, \infty)$ and $\alpha > 0$;
(iii) $e^{-(-x)^\alpha}$ with support $(-\infty, 0]$ and $\alpha > 0$.

21.14 ***Example (Uniform).*** If the distribution has finite support $[0, 1]$ with $\overline{F}(t) = (1 - t)^\alpha$, then $n\overline{F}(1 + n^{-1/\alpha}x) \to (-x)^\alpha$ for every $x \le 0$. In view of Lemma 21.12, the sequence $n^{1/\alpha}(X_{n(n)} - 1)$ converges in distribution to a limit of type (iii). The uniform distribution is the special case with $\alpha = 1$, for which the limit distribution is the negative of an exponential distribution. □

† For a proof of the following theorem, see [66] or Theorem 1.4.2 in [90].

21.15 ***Example (Pareto).*** The survival distribution of the Pareto distribution satisfies $\overline{F}(t) = (\mu/t)^\alpha$ for $t \geq \mu$. Thus $n\overline{F}(n^{1/\alpha}\mu x) \to 1/x^\alpha$ for every $x > 0$. In view of Lemma 21.12, the sequence $n^{-1/\alpha} X_{n(n)}/\mu$ converges in distribution to a limit of type (ii). □

21.16 ***Example (Normal).*** For the normal distribution the calculations are similar, but more delicate. We choose

$$a_n = \sqrt{2\log n} - \frac{1}{2}\frac{\log\log n + \log 4\pi}{\sqrt{2\log n}}, \qquad b_n = 1/\sqrt{2\log n}.$$

Using *Mill's ratio*, which asserts that $\overline{\Phi}(t) \sim \phi(t)/t$ as $t \to \infty$, it is straightforward to see that $n\overline{\Phi}(a_n + b_n x) \to e^{-x}$ for every x. In view of Lemma 21.12, the sequence $\sqrt{2\log n}(X_{n(n)} - a_n)$ converges in distribution to a limit of type (i). □

The problem of convergence in distribution of suitably normalized maxima is solved in general by the following theorem. Let $\tau_F = \sup\{t: F(t) < 1\}$ be the right endpoint of F (possibly ∞).

21.17 ***Theorem.*** *There exist constants a_n and b_n such that the sequence $b_n^{-1}(X_{n(n)} - a_n)$ converges in distribution if and only if, as $t \to \tau_F$,*

(i) There exists a strictly positive function g on $\mathbb{R}$ such that $\overline{F}(t + g(t)x)/\overline{F}(t) \to e^{-x}$, for every $x \in \mathbb{R}$;

(ii) $\tau_F = \infty$ and $\overline{F}(tx)/\overline{F}(t) \to 1/x^\alpha$, for every $x > 0$;

(iii) $\tau_F < \infty$ and $\overline{F}\big(\tau_F - (\tau_F - t)x\big)/\overline{F}(t) \to x^\alpha$, for every $x > 0$.

The constants (a_n, b_n) can be taken equal to $\big(u_n, g(u_n)\big)$, $(0, u_n)$ and $(\tau_F, \tau_F - u_n)$, respectively, for $u_n = F^{-1}(1 - 1/n)$.

Proof. We only give the proof for the "only if" part, which follows the same lines as the preceding examples. In every of the three cases, $n\overline{F}(u_n) \to 1$. To see this it suffices to show that the jump $F(u_n) - F(u_n-) = o(1/n)$. In case (i) this follows because, for every $x < 0$, the jump is smaller than $\overline{F}\big(u_n + g(u_n)x\big) - \overline{F}(u_n)$, which is of the order $\overline{F}(u_n)(e^{-x} - 1) \leq (1/n)(e^{-x} - 1)$. The right side can be made smaller than $\varepsilon(1/n)$ for any $\varepsilon > 0$, by choosing x close to 0. In case (ii), we can bound the jump at u_n by $\overline{F}(xu_n) - \overline{F}(u_n)$ for every $x < 1$, which is of the order $\overline{F}(u_n)(1/x^\alpha - 1) \leq (1/n)(1/x^\alpha - 1)$. In case (iii) we argue similarly.

We conclude the proof by applying Lemma 21.12. For instance, in case (i) we have $n\overline{F}\big(u_n + g(u_n)x\big) \sim n\overline{F}(u_n)e^{-x} \to e^{-x}$ for every x, and the result follows. The argument under the assumptions (ii) or (iii) is similar. ■

If the maximum converges in distribution, then the $(k + 1)$-th largest-order statistics $X_{n(n-k)}$ converge in distribution as well, with the same centering and scaling, but a different limit distribution. This follows by combining the preceding results and the Poisson approximation to the binomial distribution.

21.18 ***Theorem.*** *If $b_n^{-1}(X_{n(n)} - a_n) \rightsquigarrow G$, then $b_n^{-1}(X_{n(n-k)} - a_n) \rightsquigarrow H$ for the distribution function $H(x) = G(x)\sum_{i=0}^{k}\big(-\log G(x)\big)^i/i!$.*

Proof. If $p_n = \overline{F}(a_n + b_n x)$, then $np_n \to -\log G(x)$ for every x where G is continuous (all x), by Lemma 21.12. Furthermore,

$$\mathrm{P}\bigl(b_n^{-1}(X_{n(n-k)} - a_n) \le x\bigr) = \mathrm{P}\bigl(\mathrm{bin}(n, p_n) \le k\bigr).$$

This converges to the probability that a Poisson variable with mean $-\log G(x)$ is less than or equal to k. (See problem 2.21.) ∎

By the same, but more complicated, arguments, the sample extremes can be seen to converge jointly in distribution also, but we omit a discussion.

Any order statistic depends, by its definition, on all observations. However, asymptotically central and extreme order statistics depend on the observations in orthogonal ways and become stochastically independent. One way to prove this is to note that central-order statistics are asymptotically equivalent to means, and averages and extreme order statistics are asymptotically independent, which is a result of interest on its own.

21.19 ***Lemma.*** *Let g be a measurable function with $Fg = 0$ and $Fg^2 = 1$, and suppose that $b_n^{-1}(X_{n(n)} - a_n) \rightsquigarrow G$ for a nondegenerate distribution G. Then $\bigl(n^{-1/2}\sum_{i=1}^n g(X_i), b_n^{-1}(X_{n(n)} - a_n)\bigr) \rightsquigarrow (U, V)$ for independent random variables U and V with distributions $N(0, 1)$ and G.*

Proof. Let $U_n = n^{-1/2}\sum_{i=1}^{n-1} g(X_{n(i)})$ and $V_n = b_n^{-1}(X_{n(n)} - a_n)$. Because $Fg^2 < \infty$, it follows that $\max_{1 \le i \le n} |g(X_i)| = o_P(\sqrt{n})$. Hence $n^{-1/2}\bigl|g(X_{n(n)})\bigr| \xrightarrow{\mathrm{P}} 0$, whence the distance between $(\mathbb{G}_n g, V_n)$ and (U_n, V_n) converges to zero in probability. It suffices to show that $(U_n, V_n) \rightsquigarrow (U, V)$. Suppose that we can show that, for every u,

$$F_n(u \mid V_n) := \mathrm{P}(U_n \le u \mid V_n) \xrightarrow{\mathrm{P}} \Phi(u).$$

Then, by the dominated-convergence theorem, $\mathrm{E}F_n(u \mid V_n)1\{V_n \le v\} = \Phi(u)\mathrm{E}1\{V_n \le v\} + o(1)$, and hence the cumulative distribution function $\mathrm{E}F_n(u \mid V_n)1\{V_n \le v\}$ of (U_n, V_n) converges to $\Phi(u)G(v)$.

The conditional distribution of U_n given that $V_n = v_n$ is the same as the distribution of $n^{-1/2}\sum X_{ni}$ for i.i.d. random variables $X_{n,1}, \ldots, X_{n,n-1}$ distributed according to the conditional distribution of $g(X_1)$ given that $X_1 \le x_n := a_n + b_n v_n$. These variables have absolute mean

$$\bigl|\mathrm{E}X_{n1}\bigr| = \frac{\bigl|\int_{(-\infty, x_n]} g\,dF\bigr|}{F(x_n)} = \frac{\bigl|\int_{(x_n, \infty)} g\,dF\bigr|}{F(x_n)} \le \frac{\bigl(\int_{(x_n,\infty)} g^2\,dF\,\overline{F}(x_n)\bigr)^{1/2}}{F(x_n)}.$$

If $v_n \to v$, then $\mathrm{P}(V_n \le v_n) \to G(v)$ by the continuity of G, and, by Lemma 21.12, $n\overline{F}(x_n) = O(1)$ whenever $G(v) > 0$. We conclude that $\sqrt{n}\mathrm{E}X_{n1} \to 0$. Because we also have that $\mathrm{E}X_{n1}^2 \to Fg^2$ and $\mathrm{E}X_{n1}^2 1\bigl\{|X_{n1}| \ge \varepsilon\sqrt{n}\bigr\} \to 0$ for every $\varepsilon > 0$, the Lindeberg-Feller theorem yields that $F_n(u \mid v_n) \to \Phi(u)$. This implies $F_n(u \mid V_n) \rightsquigarrow \Phi(u)$ by Theorem 18.11 or a direct argument. ∎

By taking linear combinations, we readily see from the preceding lemma that the empirical process $\mathbb{G}_n$ and $b_n^{-1}(X_{n(n)} - a_n)$, if they converge, are asymptotically independent as well. This independence carries over onto statistics whose asymptotic distribution can be derived from the empirical process by the delta method, including central order statistics $X_{n(k_n/n)}$ with $k_n/n = p + O(1/\sqrt{n})$, because these are asymptotically equivalent to averages.

Notes

For more results concerning the empirical quantile function, the books [28]and [134] are good starting points. For results on extreme order statistics, see [66] or the book [90].

PROBLEMS

1. Suppose that $F_n \to F$ uniformly. Does this imply that $F_n^{-1} \to F^{-1}$ uniformly or pointwise? Give a counterexample.
2. Show that the asymptotic lengths of the two types of asymptotic confidence intervals for a quantile, discussed in Example 21.8, are within $o_P(1/\sqrt{n})$. Assume that the asymptotic variance of the sample quantile (involving $1/f \circ F^{-1}(p)$) can be estimated consistently.
3. Find the limit distribution of the median absolute deviation from the mean, $\text{med}_{1\leq i\leq n} |X_i - \overline{X}_n|$.
4. Find the limit distribution of the pth quantile of the absolute deviation from the median.
5. Prove that $\overline{X}_n$ and $X_{n(n-1)}$ are asymptotically independent.