Chapter 2. Order Statistics

1 The Order Statistics

For a sample of independent observations $X_1, X_2, \ldots, X_n$ on a distribution F, the ordered sample values

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)},$$

or, in more explicit notation,

$$X_{(1:n)} \le X_{(2:n)} \le \cdots \le X_{(n:n)},$$

are called the order statistics. If F is continuous, then with probability 1 the order statistics of the sample take distinct values (and conversely).

There is an alternative way to visualize order statistics that, although it does not necessarily yield simple expressions for the joint density, does allow simple derivation of many important properties of order statistics. It can be called the quantile function representation. The *quantile function* (or *inverse distribution function*, if you wish) is defined by

$$F^{-1}(y) = \inf\{x : F(x) \ge y\}. \tag{1}$$

Now it is well known that if U is a Uniform(0,1) random variable, then $F^{-1}(U)$ has distribution function F. Moreover, if we envision $U_1, \ldots, U_n$ as being iid Uniform(0,1) random variables and $X_1, \ldots, X_n$ as being iid random variables with common distribution F, then

$$(X_{(1)}, \ldots, X_{(n)}) \stackrel{d}{=} (F^{-1}(U_{(1)}), \ldots, F^{-1}(U_{(n)})), \tag{2}$$

where $\stackrel{d}{=}$ is to be read as "has the same distribution as."

1.1 The Quantiles and Sample Quantiles

Let F be a distribution function (continuous from the right, as usual). The proof of F is right continuous can be obtained from the following fact:

$$F(x + h_n) - F(x) = P(x < X \le x + h_n),$$

where $\{h_n\}$ is a sequence of real numbers such that $0 < h_n \downarrow 0$ as $n \to \infty$. It follows from the continuity property of probabaility ($P(\lim_n A_n) = \lim_n P(A_n)$ if $\lim A_n$ exists.) that

$$\lim_{n\to\infty} [F(x + h_n) - F(x)] = 0,$$

and hence that F is right-continuous. Let D be the set of all discontinuity points of F and n be a positive integer. Set

$$D_n = \left\{ x \in D : P(X = x) \ge \frac{1}{n} \right\}.$$

Since $F(\infty)-F(-\infty)=1$, the number of elements in D_n cannot exceed n. Clearly $D=\cup_n D_n$, and it follows that D is countable. Or, the set of discontinuity points of a distribution function F is countable. We then conclude that every distribution function F admits the decomposition

$$F(x)=\alpha F_d(x)+(1-\alpha)F_c(x), \quad (0\leq\alpha\leq 1),$$

where F_d and F_c are both continuous function such that F_d is a step function and F_c is continuous. Moreover, the above decomposition is unique.

Let λ denote the Lebesgue measure on $\mathcal{B}$, the σ-field of Borel sets in R. It follows from the Lebesgue decomposition theorem that we can write $F_c(x)=\beta F_s(x)+(1-\beta)F_{ac}(x)$ where $0\leq\beta\leq 1$, F_s is singular with respect to λ, and F_{ac} is absolutely continuous with respect to λ. On the other hand, the Radon-Nikodym theorem implies that there exists a nonnegative Borel-measurable function on R such that

$$F_{ac}(x)=\int_{-\infty}^{x} f d\lambda,$$

where f is called the Radon-Nikodym derivative. This says that every distribution function F admits a unique decomposition

$$F(x)=\alpha_1 F_d(x)+\alpha_2 F_s(x)+\alpha_3 F_{ac}(x), \quad (x\in R),$$

where $\alpha_i\geq 0$ and $\sum_{i=1}^{3}\alpha_i=1$.

For $0<p<1$, the pth *quantile* or *fractile* of F is defined as

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

This definition is motivated by the following observation:

- If F is continuous and strictly increasing, F^{-1} is defined by

$$F^{-1}(y)=x \quad \text{when } y=F(x).$$

- If F has a discontinuity at x_0, suppose that $F(x_0-)<y<F(x_0)=F(x_0+)$. In this case, although there exists no x for which $y=F(x)$, $F^{-1}(y)$ is defined to be equal to x_0.

- Now consider the case that F is not strictly increasing. Suppose that

$$F(x)\begin{cases} <y & \text{for } x<a \\ =y & \text{for } a\leq x\leq b \\ >y & \text{for } x>b \end{cases}$$

  Then any value $a\leq x\leq b$ could be chosen for $x=F^{-1}(y)$. The convention in this case is to define $F^{-1}(y)=a$.

Now we prove that if U is uniformly distributed over the interval $(0,1)$, then $X = F_X^{-1}(U)$ has cumulative distribution function $F_X(x)$. The proof is straightforward:

$$P(X \leq x) = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)] = F_X(x).$$

Note that discontinuities of F become converted into flat stretches of F^{-1} and flat stretches of F into discontinuities of F^{-1}.

In particular, $\xi_{1/2} = F^{-1}(1/2)$ is called the *median* of F. Note that ξ_p satisfies

$$F(\xi(p)-) \leq p \leq F(\xi(p)).$$

The function $F^{-1}(t)$, $0 < t < 1$, is called the *inverse* function of F. The following proposition, giving useful properties of F and F^{-1}, is easily checked.

Lemma 1 *Let F be a distribution function. The function $F^{-1}(t)$, $0 < t < 1$, is nondecreasing and left-continuous, and satisfies*

(i) $F^{-1}(F(x)) \leq x$, $-\infty < x < \infty$,

(ii) $F(F^{-1}(t)) \geq t$, $0 < t < 1$.

Hence

(iii) $F(x) \geq t$ *if and only if* $x \geq F^{-1}(t)$.

Corresponding to a sample $\{X_1, X_2, \ldots, X_n\}$ of observations on F, the sample pth *quantile* is defined as the pth quantile of the sample distribution function F_n, that is, as $F_n^{-1}(p)$. Regarding the sample pth quantile as an estimator of ξ_p, we denote it by $\hat{\xi}_{pn}$, or simply by $\hat{\xi}_p$ when convenient.

Since the order stastistics is equivalent to the sample distribution function F_n, its role is fundamental even if not always explicit. Thus, for example, the sample mean may be regarded as the mean of the order statistics, and the sample pth quantile may be expressed as

$$\hat{\xi}_{pn} = \begin{cases} X_{n,np} & \text{if } np \text{ is an integer} \\ X_{n,[np]+1} & \text{if } np \text{ is not an integer.} \end{cases}$$

1.2 Functions of Order Statistics

Here we consider statistics which may be expressed as *functions* of order statistics. A variety of short-cut procedures for quick estimates of location or scale parameters, or for quick tests of related hypotheses, are provided in the form of *linear* functions of order statistics, that is statistics of the form

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

We term such statistics "L-estimates." For example, the *sample range* $X_{(n:n)} - X_{(1:n)}$ belongs to this class. Another example is given by the α-*trimmed mean*.

$$\frac{1}{n-2[n\alpha]} \sum_{i=[n\alpha]+1}^{n-[n\alpha]} X_{(i:n)},$$

which is a popular competitor of $\bar{X}$ for *robust* estimation of location. The asymptotic distribution theory of L-statistics takes quite different forms, depending on the character of the coefficients $\{c_{ni}\}$.

The representations of $\bar{X}$ and $\hat{\xi}_{pn}$ in terms of order statistics are a bit artificial. On the other hand, for many useful statistics, the most natural and efficient representations are in terms of order statistics. Examples are the *extreme values* $X_{1:n}$ and $X_{n:n}$ and the *sample range* $X_{n:n} - X_{1:n}$.

1.3 General Properties

Theorem 1 **(1)** *$P(X_{(k)} \leq x) = \sum_{i=k}^{n} C(n,i)[F(x)]^i[1-F(x)]^{n-i}$ for $-\infty < x < \infty$.*

(2) *The density of $X_{(k)}$ is given by $nC(n-1,k-1)F^{k-1}(x)[1-F(x)]^{n-k}f(x)$.*

(3) *The joint density of $X_{(k_1)}$ and $X_{(k_2)}$ is given by*

$$\frac{n!}{(k_1-1)!(k_2-k_1-1)!(n-k_2)!}[F(x_{(k_1)})]^{k_1-1}[F(x_{(k_2)})-F(x_{(k_1)})]^{k_2-k_1-1}$$
$$[1-F(x_{(k_2)})]^{n-k_2}f(x_{(k_1)})f(x_{(k_2)})$$

for $k_1 < k_2$ and $x_{(k_1)} < x_{(k_2)}$.

(4) *The joint pdf of all the order statistics is $n!f(z_1)f(z_2)\cdots f(z_n)$ for $-\infty < z_1 < \cdots < z_n < \infty$.*

(5) *Define $V = F(X)$. Then V is uniformly distributed over $(0,1)$.*

Proof. (1) The event $\{X_{(k)} \leq x\}$ occurs if and only if at least k out of $X_1, X_2, \ldots, X_n$ are less than or equal to x.
(2) The density of $X_{(k)}$ is given by $nC(n-1,k-1)F^{k-1}(x)[1-F(x)]^{n-k}f(x)$. It can be shown by the fact that

$$\frac{d}{dp}\sum_{i=k}^{n} C(n,i)p^i(1-p)^{n-i} = nC(n-1,k-1)p^{k-1}(1-p)^{n-k}.$$

Heuristically, $k-1$ smallest observations are $\leq x$ and $n-k$ largest are $> x$. $X_{(k)}$ falls into a small interval of length dx about x is $f(x)dx$.

1.4 Conditional Distribution of Order Statistics

In the following two theorems, we relate the conditional distribution of order statistics (conditioned on another order statistic) to the distribution of order statistics from a population whose distribution is a truncated form of the original population distribution function $F(x)$.

Theorem 2 *Let $X_1, X_2, \ldots, X_n$ be a random sample from an absolutely continuous population with cdf $F(x)$ and density function $f(x)$, and let $X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}$ denote the order statistics obtained from this sample. Then the conditional distribution of $X_{(j:n)}$, given that $X_{(i:n)} = x_i$ for $i < j$, is the same as the distribution of the $(j-i)$th order statistic obtained from a sample of size $n-i$ from a population whose distribution is simply $F(x)$ truncated on the left at x_i.*

Proof. From the marginal density function of $X_{(i:n)}$ and the joint density function of $X_{(i:n)}$ and $X_{(j:n)}$, we have the conditional density function of $X_{(j:n)}$, given that $X_{(i:n)} = x_i$, as

$$\begin{aligned} f_{X_{(j:n)}}(x_j|X_{(i:n)} = x_i) &= f_{X_{(i:n)},X_{(j:n)}}(x_i, x_j)/f_{X_{(i:n)}}(x_i) \\ &= \frac{(n-i)!}{(j-i-1)!(n-j)!}\left\{\frac{F(x_j)-F(x_i)}{1-F(x_i)}\right\}^{j-i-1} \\ &\quad \times \left\{\frac{1-F(x_j)}{1-F(x_i)}\right\}^{n-j} \frac{f(x_j)}{1-F(x_i)}. \end{aligned}$$

Here $i < j \leq n$ and $x_i \leq x_j < \infty$. The result follows easily by realizing that $\{F(x_j) - F(x_i)\}/\{1-F(x_i)\}$ and $f(x_j)/\{1-F(x_i)\}$ are the cdf and density function of the population whose distribution is obtained by truncating the distribution $F(x)$ on the left at x_i.

Theorem 3 *Let $X_1, X_2, \ldots, X_n$ be a random sample from an absolutely continuous population with cdf $F(x)$ and density function $f(x)$, and let $X_{(1:n)} \leq X_{(2:n)} \leq \cdots \leq X_{(n:n)}$ denote the order statistics obtained from this sample. Then the conditional distribution of $X_{(i:n)}$, given that $X_{(j:n)} = x_j$ for $j > i$, is the same as the distribution of the ith order statistic in a sample of size $j-1$ from a population whose distribution is simply $F(x)$ truncated on the right at x_j.*

Proof. From the marginal density function of $X_{(i:n)}$ and the joint density function of $X_{(i:n)}$ and $X_{(j:n)}$, we have the conditional density function of $X_{(i:n)}$, given that $X_{(j:n)} = x_j$, as

$$\begin{aligned} f_{X_{(i:n)}}(x_i|X_{(j:n)} = x_j) &= f_{X_{(i:n)},X_{(j:n)}}(x_i, x_j)/f_{X_{(j:n)}}(x_j) \\ &= \frac{(j-1)!}{(i-1)!(j-i-1)!}\left\{\frac{F(x_i)}{F(x_j)}\right\}^{i-1} \\ &\quad \times \left\{\frac{F(x_j)-F(x_i)}{F(x_j)}\right\}^{j-i-1} \frac{f(x_i)}{F(x_j)}. \end{aligned}$$

Here $1 \le i < j$ and $-\infty < x_i \le x_j$. The proof is completed by noting that $F(x_i)/F(x_j)$ and $f(x_i)/F(x_j)$ are the cdf and density function of the population whose distribution is obtained by truncating the distribution $F(x)$ on the right at x_j

1.5 Computer Simulation of Order Statistics

In this section, we will discuss some methods of simulating order statistics from a distribution $F(x)$. First of all, it should be mentioned that a straightforward way of simulating order statistics is to generate a pseudorandom sample from the distribution $F(x)$ and then sort the sample through an efficient algorithm like quick-sort. This general method (being time-consuming and expensive) may be avoided in many instances by making use of some of the distributional properties to be established now.

For example, if we wish to generate the complete sample $(x_{(1)}, \ldots, x_{(n)})$ or even a Type II censored sample $(x_{(1)}, \ldots, x_{(r)})$ from the standard exponential distribution. This may be done simply by generating a pseudorandom sample $y_1, \ldots, y_r$ from the standard exponential distribution first, and then setting

$$x_{(i)} = \sum_{j=1}^{i} y_j/(n-j+1), \quad i = 1, 2, \ldots, r.$$

The reason is as follows:

Theorem 4 *Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics from the standard exponential distribution. Then, the random variables $Z_1, Z_2, \ldots, Z_n$, where*

$$Z_i = (n-i+1)(X_{(i)} - X_{(i-1)}),$$

with $X_{(0)} \equiv 0$, are statistically independent and also have standard exponential distributions.

Proof. Note that the joint density function of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \exp\left(-\sum_{i=1}^{n} x_i\right), \quad 0 \le x_1 < x_2 < \cdots < x_n < \infty.$$

Now let us consider the transformation

$$Z_1 = nX_{(1)}, Z_2 = (n-1)(X_{(2)} - X_{(1)}), \ldots, Z_n = X_{(n)} - X_{(n-1)},$$

or the equivalent transformation

$$X_{(1)} = Z_1/n, X_{(2)} = \frac{Z_1}{n} + \frac{Z_2}{n-1}, \ldots, X_{(n)} = \frac{Z_1}{n} + \frac{Z_2}{n-1} + \cdots + Z_n.$$

After noting the Jacobian of this transformation is $1/n!$ and that $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} z_i$, we immediately obtain the joint density function of $Z_1, Z_2, \ldots, Z_n$ to be

$$f_{Z_1,Z_2,\ldots,Z_n}(z_1, z_2, \ldots, z_n) = \exp\left(-\sum_{i=1}^{n} z_i\right), \quad 0 \le z_1, \ldots, z_n < \infty.$$

If we wish to generate order statistics from the Uniform$(0,1)$ distribution, we may use the following two Theorems and avoid sorting once again. For example, if we only need the ith order statistic $u_{(i)}$, it may simply be generated as a pseudorandom observation from $Beta(i, n-i+1)$ distribution.

Theorem 5 *For the Uniform$(0,1)$ distribution, the random variables $V_1 = U_{(i)}/U_{(j)}$ and $V_2 = U_{(j)}$, $1 \le i < j \le n$, are statistically independent, with V_1 and V_2 having Beta$(i, j-i)$ and Beta$(j, n-j+1)$ distributions, respectively.*

Proof. From Theorem 1(3), we have the joint density function of $U_{(i)}$ and $U_{(j)}$ $(1 \le i < j \le n)$ to be

$$f_{i,j:n}(u_i, u_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} u_i^{i-1}(u_j - u_i)^{j-i-1}(1-u_j)^{n-j}, \;\; 0 < u_i < u_j < 1.$$

Now upon makin the transformation $V_1 = U_{(i)}/U_{(j)}$ and $V_2 = U_{(j)}$ and noting that the Jacobian of this transformation is v_2, we derive the joint density function of V_1 and V_2 to be

$$\begin{aligned} f_{V_1,V_2}(v_1, v_2) &= \frac{(j-1)!}{(i-1)!(j-i-1)!} v_1^{i-1}(1-v_1)^{j-i-1} \\ &= \frac{n!}{(j-1)!(n-j)!} v_2^{j-1}(1-v_2)^{n-j}, \end{aligned}$$

$0 < v_1 < 1$, $0 < v_2 < 1$. From the above equation it is clear that the random variables V_1 and V_2 are statistically independent, and also that they are distributed as Beta$(i, j-i)$ and Beta$(j, n-j+1)$, respectively.

Theorem 6 *For the Uniform$(0,1)$ distribution, the random variables*

$$V_1^* = \frac{U_{(1)}}{U_{(2)}}, V_2^* = \left(\frac{U_{(2)}}{U_{(3)}}\right)^2, \cdots, V_{(n-1)}^* = \left(\frac{U_{(n-1)}}{U_{(n)}}\right)^{n-1}$$

and $V_n^* = U_{(n)}^n$ are all independent Uniform$(0,1)$ random variables.

Proof. Let $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$ denote the order statistics from the standard exponential distribution. Then upon making use of the facts that $X = -\log U$ has a standard exponential distribution and that $-\log u$ is a monotonically decreasing function in u, we immediately have $X_{(i)} \stackrel{d}{=} -\log U_{(n-i+1)}$. The above equation yields

$$V_i^* = \left(\frac{U_{(i)}}{U_{(i+1)}}\right)^I \stackrel{d}{=} \left(\frac{e^{-X_{(n-i+1)}}}{e^{-X_{(n-i)}}}\right) = \exp[-i(X_{(n-i+1)} - X_{(n-i)})] \stackrel{d}{=} \exp(-Y_{n-i+1})$$

upon using the above theorem, where Y_i' are independent standard exponential random variables.

The just-described methods of simulating uniform order statistics may also be used easily to generate order statistics from any known distribution $F(x)$ for which $F^{-1}(\cdot)$ is relatively easy to compute. We may simply obtain the order statistics $x_{(1)}, \ldots, x_{(n)}$ from the required distribution $F(\cdot)$ by setting $x_{(i)} = F^{-1}(u_{(i)})$.

2 Large Sample Properties of Sample Quantile

2.1 An Elementary Proof

Consider the sample pth quantile, $\hat{\xi}_{pn}$, which is $X_{([np])}$ or $X_{([np]+1)}$ depending on whether np is an integer (here $[np]$ denotes the integer part of np). For simplicity, we discuss the properties of $X_{([np])}$ where $p \in (0,1)$ and n is large. This will in turn inform us of the properties of $\hat{\xi}_{pn}$.

We first consider the case that X is uniformly distributed over $[0,1]$. Let $U_{([np])}$ denote the sample pth quantile. If $i = [np]$, we have

$$nC(n-1, i-1) = \frac{n!}{(i-1)!(n-i)!} = \frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} = B(i, n-i+1).$$

Elementary computations beginning with Theorem 1(2) yields $U_{([np])} \sim Beta(i_0, n+1-i_0)$ where $i_0 = [np]$. Then $U_{(i_0)} \sim Beta(i_0, n-i_0+1)$. Note that

$$\begin{aligned} EU_{([np])} &= \frac{[np]}{n+1} \to p \\ nCov\left(U_{(np_1)}, U_{(np_2)}\right) &= n\frac{[np_1](n+1-[np_2])}{(n+1)^2(n+2)} \to p_1(1-p_2). \end{aligned}$$

Use these facts and Chebyschev inequality, we can show easily that $U_{([np])} \xrightarrow{P} p$ with rate $n^{-1/2}$. This generates the question whether we can claim that

$$\hat{\xi}_{pn} \xrightarrow{P} \xi_p.$$

Recall that $U = F(X)$. If F is absolutely continuous with finite positive density f at ξ_p, it is expected that the above claim holds.

Recall that $U_{([np])} \xrightarrow{P} p$ with rate $n^{-1/2}$. The next question would be what the distribution of $\sqrt{n}(U_{([np])} - p)$ is? Note that $U_{([np])}$ is a $Beta([np], n-[np]+1)$ random variable. Thus, it can be expressed as

$$U_{[np]} = \frac{\sum_{i=1}^{i_0} V_i}{\sum_{i=1}^{i_0} V_i + \sum_{i=i_0+1}^{n+1} V_i},$$

where the V_i's are iid $Exp(1)$ random variables. Observe that

$$\begin{aligned} &\sqrt{n}\left(\frac{\sum_{i=1}^{[np]} V_i}{\sum_{i=1}^{[np]} V_i + \sum_{i=[np]+1}^{n+1} V_i} - p\right) \\ &= \frac{\frac{1}{\sqrt{n}}\left\{(1-p)\left(\sum_{i=1}^{[np]} V_i - [np]\right) - p\left(\sum_{[np]+1}^{n+1} V_i - (n-[np]+1)\right) + [(1-p)[np] - p(n-[np]+1)]\right\}}{\sum_{i=1}^{n+1} V_i/n} \end{aligned}$$

and $(\sqrt{n})^{-1}\left\{(1-p)[np] - p(n-[np]+1)\right\} \to 0$. Since $E(V_i) = 1$ and $Var(V_i) = 1$, from the central limit theorem it follows that

$$\frac{\sum_{i=1}^{i_0} V_i - i_0}{\sqrt{i_0}} \xrightarrow{d} N(0,1)$$

and

$$\frac{\sum_{i=i_0+1}^{n+1} V_i - (n - i_0 + 1)}{\sqrt{n - i_0 + 1}} \xrightarrow{d} N(0,1).$$

Consequently,

$$\frac{\sum_{i=1}^{i_0} V_i - i_0}{\sqrt{n}} \xrightarrow{d} N(0,p)$$

and

$$\frac{\sum_{i=i_0+1}^{n+1} V_i - (n - i_0 + 1)}{\sqrt{n}} \xrightarrow{d} N(0,1-p).$$

Since $\sum_{i=1}^{i_0} V_i$ and $\sum_{i=i_0+1}^{n} V_i$ are independent for all n,

$$(1-p)\frac{\sum_{i=1}^{[np]} V_i - [np]}{\sqrt{n}} - p\frac{\sum_{[np]+1}^{n+1} V_i - (n - [np] + 1)}{\sqrt{n}} \xrightarrow{d} N(0, p(1-p)).$$

Now, using the weak law of large numbers, we have

$$\frac{1}{n+1}\sum_{i=1}^{n+1} V_i \xrightarrow{P} 1.$$

Hence, by Slutsky's Theorem,

$$\sqrt{n}\left(U_{([np])} - p\right) \xrightarrow{d} N(0, p(1-p)).$$

For an arbitrary cdf F, we have $X_{([np])} = F^{-1}(U_{([np])})$. Upon expanding $F^{-1}(U_{([np])})$ in a Taylor-series around the point $E(U_{([np])}) = [np]/(n+1)$, we get

$$X_{([np])} \stackrel{d}{=} F^{-1}(p) + (U_{([np])} - p)\{f(F^{-1}(D_n))\}^{-1}$$

where the random variable D_n is betwen $U_{([np])}$ and p. This can be rearranged as

$$\sqrt{n}\left\{X_{([np])} - F^{-1}(p)\right\} \stackrel{d}{=} \sqrt{n}\left(U_{([np])} - p\right)\{f(F^{-1}(D_n)\}^{-1}.$$

When f is continuous at $F^{-1}(p)$, it follows that as $n \to \infty$, $f(F^{-1}(D_n)) \xrightarrow{P} f(F^{-1}(p))$. Use the delta method, it yields

$$\sqrt{n}\left(X_{([np])} - F^{-1}(p)\right) \xrightarrow{d} N\left(0, \frac{p(1-p)}{[f(F^{-1}(p))]^2}\right).$$

2.2 A Probability Inequality for $|\hat{\xi}_p - \xi_p|$

We shall use the following result of Hoeffding (1963) to show that $P(|\hat{\xi}_{pn} - \xi_p| > \epsilon) \to 0$ *exponentially* fast.

Lemma 2 *Let $Y_1, \ldots, Y_n$ be independent random variables satisfying $P(a \le Y_i \le b) = 1$, each i, where $a < b$. Then, for $t > 0$,*

$$P\left(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} E(Y_i) \ge nt\right) \le \exp\left[-2nt^2/(b-a)^2\right].$$

Remark. Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed random variables. Use Bery-Esseen theorem, we have

$$P\left(\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} E(Y_i) \geq nt\right) = \Phi(t\sqrt{Var(Y_1)}/\sqrt{n}) + Error.$$

Here

$$|Error| \leq \frac{c}{\sqrt{n}} \frac{E|Y_1 - EY_1|^3}{[Var(Y_1)]^{3/2}}.$$

Theorem 7 *Let $0 < p < 1$. Suppose that ξ_p is the unique solution x of $F(x-) \leq p \leq F(x)$. Then, for every $\epsilon > 0$,*

$$P\left(|\hat{\xi}_{pn} - \xi_p| > \epsilon\right) \leq 2\exp\left(-2n\delta_\epsilon^2\right), \qquad \textit{all } n,$$

where $\delta_\epsilon = \min\{F(\xi_p + \epsilon) - p, p - F(\xi_p - \epsilon)\}$.

Proof. Let $\epsilon > 0$. Write

$$P\left(|\hat{\xi}_{pn} - \xi_p| > \epsilon\right) = P\left(\hat{\xi}_{pn} > \xi_p + \epsilon\right) + P\left(\hat{\xi}_{pn} < \xi_p - \epsilon\right).$$

By Lemma **??**,

$$\begin{aligned} P\left(\hat{\xi}_{pn} > \xi_p + \epsilon\right) &= P(p > F_n(\xi_p + \epsilon)) \\ &= P\left(\sum_{i=1}^{n} I(X_i > \xi_p + \epsilon) > n(1-p)\right) \\ &= P\left(\sum_{i=1}^{n} V_i - \sum_{i=1}^{n} E(V_i) > n\delta_1\right), \end{aligned}$$

where $V_i = I(X_i > \xi_p + \epsilon)$ and $\delta_1 = F(\xi_p + \epsilon) - p$. Likewise,

$$\begin{aligned} P\left(\hat{\xi}_{pn} < \xi_p - \epsilon\right) &= P(p > F_n(\xi_p - \epsilon)) \\ &= P\left(\sum_{i=1}^{n} W_i - \sum_{i=1}^{n} E(W_i) > n\delta_2\right), \end{aligned}$$

where $W_i = I(X_i < \xi_p - \epsilon)$ and $\delta_2 = p - F(\xi_p - \epsilon)$. Therefore, utilizing Hoeffding's lemma, we have

$$P\left(\hat{\xi}_{pn} > \xi_p + \epsilon\right) \leq \exp\left(-2n\delta_1^2\right)$$

and

$$P\left(\hat{\xi}_{pn} < \xi_p - \epsilon\right) \leq \exp\left(-2n\delta_2^2\right).$$

Putting $\delta_\epsilon = \min\{\delta_1, \delta_2\}$, the proof is completed.

2.3 Asymptotic Normality

Theorem 8 *Let $0 < p < 1$. Suppose that F possesses a density f in a neighborhood of ξ_p and f is positive and continuous at ξ_p, then*

$$\sqrt{n}(\hat{\xi}_{pn} - \xi_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{[f(\xi_p)]^2}\right).$$

Proof. We only consider $p = 1/2$. Note that ξ_p is the unique median since $f(\xi_p) > 0$. First, we consider that n is odd (i.e., $n = 2m - 1$).

$$\begin{aligned} P\left[\sqrt{n}\left(X_{(m)} - F^{-1}\left(\frac{1}{2}\right)\right) \le t\right] &= P\left(\sqrt{n}X_{(m)} \le t \,\middle|\, F^{-1}\left(\frac{1}{2}\right) = 0\right) \\ &= P\left(X_{(m)} \le t/\sqrt{n} \,\middle|\, F^{-1}\left(\frac{1}{2}\right) = 0\right). \end{aligned}$$

Let S_n be the number of X's that exceed $t/\sqrt{n}$. Then

$$X_{(m)} \le \frac{t}{\sqrt{n}} \text{ if and only if } S_n \le m - 1 = \frac{n-1}{2}.$$

Or, S_n is a $\mathrm{Bin}(n, 1 - F(F^{-1}(1/2) + tn^{-1/2}))$ random variable. Set $F^{-1}(1/2) = 0$ and $p_n = 1 - F(n^{1/2}t)$. Note that

$$\begin{aligned} &P\left[\sqrt{n}\left(X_{(m)} - F^{-1}\left(\frac{1}{2}\right)\right) \le t\right] = P\left(S_n \le \frac{n-1}{2}\right) \\ &= P\left(\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right). \end{aligned}$$

Recall that p_n depends on n. Write

$$\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} = \sum_{i=1}^{n} \frac{Y_i - p_n}{\sqrt{np_n(1-p_n)}} = \sum_{i=1}^{n} Y_{in}.$$

Again, they can be expressed as a double array with $Y_i \sim Bin(1, p_n)$

Now utilize the Berry-Esseen Theorem to have

$$P\left(S_n \le \frac{n-1}{2}\right) - \Phi\left[\frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right] \to 0$$

as $n \to \infty$. Writing

$$\begin{aligned} &\frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}} \approx \frac{\sqrt{n}\left(\frac{1}{2} - p_n\right)}{1/2} \\ &= \frac{\sqrt{n}\left(-\frac{1}{2} + F(n^{-1/2}t)\right)}{1/2} = 2t\frac{F(n^{-1/2}t) - F(0)}{n^{-1/2}t} \to 2tf(0). \end{aligned}$$

Thus,

$$\Phi\left[\frac{\frac{1}{2}(n-1) - np_n}{\sqrt{np_n(1-p_n)}}\right] \approx \Phi\left(2f(0) \cdot t\right)$$

or

$$\sqrt{n}\left(X_{(m)} - F^{-1}(1/2)\right) \xrightarrow{d} N\left(0, \frac{1}{4f^2(F^{-1}(1/2))}\right).$$

When n is even (i.e., $n = 2m$), both $P[\sqrt{n}(X_{(m)} - F^{-1}(1/2)) \leq t]$ and $P[\sqrt{n}(X_{(m+1)} - F^{-1}(1/2)) \leq t]$ tend to $\Phi\left(2f(F^{-1}(1/2)) \cdot t\right)$.

We just prove asymptotic normality of $\hat{\xi}_p$ in the case that F possesses derivative at the point ξ_p. So far we have discussed in detail the asymptotic normality of a single quantile. This discussion extends in a natural manner to the asymptotic joint normality of a fixed number of quantiles. This is made precise in the following result.

Theorem 9 *Let $0 < p_1 < \cdots < p_k < 1$. Suppose that F has a density f in neighborhoods of $\xi_{p_1}, \ldots, \xi_{p_k}$ and that f is positive and continuous at $\xi_{p_1}, \ldots, \xi_{p_k}$. Then $(\hat{\xi}_{p_1}, \ldots, \hat{\xi}_{p_k})$ is asymptotically normal with mean vector $(\xi_{p_1}, \ldots, \xi_{p_k})$ and covariance σ_{ij}/n, where*

$$\sigma_{ij} = \frac{p_i(1-p_j)}{f(\xi_{p_i})f(\xi_{p_j})} \qquad \textit{for} \qquad 1 \leq i \leq j \leq k$$

and $\sigma_{ij} = \sigma_{ji}$ for $i > j$.

Suppose that we have a sample of size n from a normal distribution $N(\mu, \sigma^2)$. Let m_n represent the median of this sample. Then because $f(\mu) = (\sqrt{2\pi}\sigma)^{-1}$,

$$\sqrt{n}(m_n - \mu) \xrightarrow{d} N(0, (1/4)/f^2(\mu)) = N(0, \pi\sigma^2/2).$$

Compare m_n with $\bar{X}_n$ as an estimator of μ. We conclude immediately that $\bar{X}_n$ is better than m_n since the latter has a much larger variance. Now consider the above problem again with Cauchy distribution $\mathcal{C}(\mu, \sigma)$ with density function

$$f(x) = \frac{1}{\pi\sigma}\frac{1}{1 + [(x-\mu)/\sigma]^2}.$$

What is your conclusion? (Exercise)

2.4 A Measure of Dispersion Based on Quantiles

The joint normality of a fixed number of central order statistics can be used to construct simultaneous confidence regions for two or more population quantiles. As an illustration, we now consider the *semi-interquantile range*, $R = \frac{1}{2}(\xi_{3/4} - \xi_{1/4})$. (Note that the parameter σ in $\mathcal{C}(\mu, \sigma)$ is the semi-interquantile range.) It is used as an alternative to σ to measure the spread of the data. A natural estimate of R is $\hat{R}_n = \frac{1}{2}(\hat{\xi}_{3/4} - \hat{\xi}_{1/4})$. Theorem 4 gives the joint distribution of $\hat{\xi}_{1/4}$ and $\hat{\xi}_{3/4}$. We can use the following result, due to Cramer and Wold (1936), which reduces the convergence of multivariate distribution functions to the convergence of univariate distribution functions.

Theorem 10 *In R^k, the random vectors $\mathbf{X}_n$ converge in distribution to the random vector $\mathbf{X}$ if and only if each linear combination of the components of $\mathbf{X}_n$ converges in distribution to the same linear combination of the components of $\mathbf{X}$.*

By Theorem 9 and the Cramer-Wold device, we have

$$\hat{R} \sim AN\left(R, \frac{1}{64n}\left(\frac{3}{[f(\xi_{1/4})]^2} - \frac{2}{f(\xi_{1/4})f(\xi_{3/4})} + \frac{3}{[f(\xi_{3/4})]^2}\right)\right).$$

For $F = N(\mu, \sigma^2)$, we have

$$\hat{R} \sim AN\left(0.6745\sigma, \frac{(0.7867)^2\sigma^2}{n}\right).$$

2.5 Confidence Intervals for Population Quantiles

Assume that F posseses a density f in a neighborhood of ξ_p and f is positive and continuous at ξ_p. For simplicity, consider $p = 1/2$. Then

$$\sqrt{n}\left(\hat{\xi}_{1/2} - \xi_{1/2}\right) \xrightarrow{d} N\left(0, \frac{1}{4f^2(\xi_{1/2})}\right).$$

Therefore, we can derive a confidence interval of $\xi_{1/2}$ if either $f(\xi_{1/2})$ is known or a good estimator of $f(\xi_{1/2})$ is available. A natural question to ask then is how do we estimate $f(\xi_{1/2})$. Here we propose two estimates. The first estimate is

$$\frac{\text{number of observatiobns falling in } \left(\xi_{1/2} - h_n, \xi_{1/2} + h_n\right)}{2h_n}$$

which is motivated by

$$\frac{F\left(\xi_{1/2} + h_n\right) - F\left(\xi_{1/2} - h_n\right)}{2h_n} \approx f(\xi_{1/2}).$$

The second one is

$$\frac{X_{[n/2+\ell]} - X_{[n/2-\ell]}}{2\ell/n} \quad \text{where } \ell = O(n^d) \text{ for } 0 < d < 1.$$

2.6 Distribution-Free Confidence Interval

When F is absolutely continuous, $F(F^{-1}(p)) = p$ and, hence, we have

$$\begin{aligned} P(X_{(i:n)} \le F^{-1}(p)) &= P(F(X_{(i:n)}) \le p) = P(U_{(i:n)} \le p) \\ &= \sum_{r=i}^{n} C(n,r)p^r(1-p)^{n-r}. \end{aligned} \tag{3}$$

Now for $i < j$, consider

$$\begin{aligned} P(X_{(i:n)} \le F^{-1}(p)) &= P(X_{(i:n)} \le F^{-1}(p), X_{(j:n)} < F^{-1}(p)) \\ &\quad + P(X_{(i:n)} \le F^{-1}(p), X_{(j:n)} \ge F^{-1}(p)) \\ &= P(X_{(j:n)} < F^{-1}(p)) + P(X_{(i:n)} \le F^{-1}(p) \le X_{(j:n)}). \end{aligned}$$

Since $X_{(j:n)}$ is absolutely continuous, this equation can be written as

$$\begin{aligned} P\left(X_{(i:n)} \le F^{-1}(p) \le X_{(j:n)}\right) &= P\left(X_{(i:n)} \le F^{-1}(p)\right) - P\left(X_{(j:n)} \le F^{-1}(p)\right) \\ &= \sum_{r=i}^{j-1} C(n,r)p^r(1-p)^{n-r}, \end{aligned} \tag{4}$$

where the last equality follows from (**??**). Thus, we have a confidence interval $[X_{(i:n)}, X_{(j:n)}]$ for $F^{-1}(p)$ whose confidence coefficient $\alpha(i,j)$ given by (**??**), is free of F and can be read from the table of binomial probabilities. If p and the desired confidence level α_0 are specified, we choose i and j so that $\alpha(i,j)$ exceeds α_0. Because of the fact that $\alpha(i,j)$ is a step function, usually the interval we obtain tends to be conservative. Further, the choice of i and j is not unique, and the choice which makes $(j-i)$ small appear reasonable. For a given n and p, the binomial pmf $C(n,r)p^r(1-p)^{n-r}$ increases as r increases up to around $[np]$, and then decreases. So if we want to make $(j-i)$ small, we have to start with i and j close to $[np]$ and gradually increase $(j-i)$ until $\alpha(i,j)$ exceeds α_0.

2.7 Q-Q Plot

Wilk and Gnanadesikan (1968) proposed a graphical, rather informal, method of testing the goodness-of-fit of a hypothesized distribution to given data. It essentially plots the quantile function of one cdf against that of another cdf. When the latter cdf is the empirical cdf defined below, order statistics come into the picture. The empirical cdf, to be denoted by $F_n(x)$ for all real x, represents the proportion of sample values that do not exceed x. It has jumps of magnitude $1/n$ at $X_{(i:n)}$, $1 \le i \le n$. Thus, the order statistics represent the values taken by $F_n^{-1}(p)$, the sample quantile function.

The Q-Q plot is the graphical representation of the points $(F^{-1}(p), X_{(i:n)})$, where population quantiles are recorded along the horizontal axis and the sample quantiles on the vertical axis. If the sample is in fact from F, we expect the Q-Q plot to be close to a straight line. If not, the plot may show nonlinearity at the upper or lower ends, which may be an indication of the presence of outliers. If the nonlinearity shows up at other points as well, one could question the validity of the assumption that the parent cdf is F.

3 Empirical Distribution Function

Consider an i.i.d. sequence $\{X_i\}$ with distribution function F. For each sample of size n, $\{X_1, \dots, X_n\}$, a corresponding *sample (empirical) distribution function* F_n is constructed by placing at each observation X_i a mass $1/n$. Thus F_n may be represented as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x), \qquad -\infty < x < \infty.$$

(The definition of F defined on R^k is completely analogous.)

For each fixed sample, $\{X_1, \dots, X_n\}$, $F_n(\cdot)$ is a distribution function, considered as a function of x. On the other hand, for each fixed value of x, $F_n(x)$ is a *random variable*, considered as a function of the sample. In a view encompassing both features, $F_n(\cdot)$ is a *random distribution function* and thus may be treated as a particular *stochastic process* (a random element of a suitable space.)

Note that the *exact* distribution of $nF_n(x)$ is simply binomial$(n, F(x))$. It follows immediately that

Theorem 11 *1.* $E[F_n(x)] = F(x)$.

2. $Var\{F_n(x)\} = \frac{F(x)[1-F(x)]}{n} \to 0,$ *as* $n \to \infty$.

3. For each fixed x, $-\infty < x < \infty$,

$$F_n(x) \ \ is \ \ AN\left(F(x), \frac{F(x)[1-F(x)]}{n}\right).$$

The sample distribution function is utilized in statistical inference in several ways. Firstly, its most direct application is for estimation of the population distribution function F. Besides *pointwise* estimation of $F(x)$, each x, it is also of interest to characterize *globally* the estimation of F by F_n. For each fixed x, the strong law of large numbers implies that

$$F_n(x) \stackrel{a.s.}{\to} F(x).$$

We now describe the *Glivenko-Cantelli Theorem* which ensures that the ecdf converges uniformly almost surely to the true distribution function.

Theorem 12

$$P\{\sup_x |F_n(x) - F(x)| \to 0\} = 1.$$

Proof. Let $\epsilon > 0$. Find an integer $k > 1/\epsilon$ and numbers

$$-\infty = x_0 < x_1 \le x_2 \le \cdots \le x_{k-1} < x_k = \infty,$$

such that $F(x_j^-) \le j/k \le F(x_j)$ for $j = 1, \ldots, k-1$. Note that if $x_{j-1} < x_j$, then $F(x_j^-) - F(x_{j-1}) \le \epsilon$. From the strong law of large numbers,

$$F_n(x_j) \stackrel{a.s.}{\to} F(x_j) \quad \text{and} \quad F_n(x_j^-) \stackrel{a.s.}{\to} F(x_j^-)$$

for $j = 1, \ldots, k-1$. Hence,

$$\triangle_n = \max(|F_n(x_j) - F(x_j)|, |F_n(x_j^-) - F(x_j^-)|, j = 1, \ldots, k-1) \stackrel{a.s.}{\to} 0.$$

Let x be arbitrary and find j such that $x_{j-1} < x \le x_j$. Then,

$$F_n(x) - F(x) \le F_n(x_j^-) - F(x_{j-1}) \le F_n(x_j^-) - F(x_j^-) + \epsilon,$$

and

$$F_n(x) - F(x) \ge F_n(x_{j-1}) - F(x_j^-) \ge F_n(x_{j-1}) - F(x_{j-1}) - \epsilon.$$

This implies that

$$\sup_x |F_n(x) - F(x)| \le \triangle_n + \epsilon \stackrel{a.s.}{\to} \epsilon.$$

Since this holds for all $\epsilon > 0$, the theorem follows.

3.1 Kolmogorov-Smirnov Test

To this effect, a very useful measure of closeness of F_n to F is the *Kolmogorov-Smirnov* distance

$$D_n = \sup_{-\infty<x<\infty} |F_n(x) - F_0(x)|.$$

A related problem is to express *confidence bands* for $F(x)$, $-\infty < x < \infty$. Thus, for selected functions $a(x)$ and $b(x)$, it is of interest to compute probabilities of the form

$$P(F_n(x) - a(x) \leq F(x) \leq F_n(x) + b(x), -\infty < x < \infty).$$

The general problem is quite difficult. However, in the simplest case, namely $a(x) = b(x) = d$, the problem reduces to computation of $P(D_n < d)$.

For the case of F 1-dimensional, an exponential-type probability inequality for D_n was established by Dvoretzky, Kiefer, and Wolfowitz (1956).

Theorem 13 *The distribution of D_n under H_0 is the same for all continuous distribution.*

Proof. For simplicity we give the proof for F_0 strictly increasing. Then F_0 has inverse F_0^{-1} and as u ranges over $(0,1)$, $F_0^{-1}(u)$ ranges over a;; the possible values of X. Thus

$$D_n = \sup_{0<u<1} |F_n(F_0^{-1}(u))F_0(F_0^{-1}(u))| = \sup_{0<u<1} |F_n(F_0^{-1}(u)) - u|.$$

Next note that

$$F_n(F_0^{-1}(u)) = [\text{number of } X_i \leq F_0^{-1}(u)]/n = [\text{number of } F_0(X_i) \leq u]/\text{n}.$$

Let $U_i = F_0(X_i)|$. Then $U_1, \ldots, U_n$ are a sample from $UNIF(0,1)$, since

$$P[F_0(X_i) \leq u] = P[X_i \leq F_0^{-1}(u)] = F_0(F_0^{-1}(u)) = u, \quad 0 < u < 1.$$

Thus,

$$D_n = \sup_{0<u<1} |F_n^*(u) - u|$$

where $F_n^*(u)$ is the empirical distribution of the uniform sample $U_1, \ldots, U_n$ and the distribution of D_n does not depend on F_0.

We now give an important fact which is used in Donsker (1952) to give a rigorous proof of the Kolmogorov-Smirnov Theorems.

Theorem 14 *The distribution of the order statistic $(Y_{(1)}, \ldots, Y_{(n)})$ of n iid random variables $Y_1, Y_2, \ldots$ from the uniform distribution on $[0,1]$ can also be obtained as the distribution of the ratios*

$$\left(\frac{S_1}{S_{n+1}}, \frac{S_2}{S_{n+1}}, \ldots, \frac{S_n}{S_{n+1}}\right),$$

where $S_k = T_1 + \cdots + T_k$, $k \geq 1$, and $T_1, T_2, \ldots$ is an iid sequence of (mean 1) exponentially distributed random variables.

Intuitively, if the T_i are regarded as the successive times between occurrence of some phenomena, then S_{n+1} is the time to the $(n+1)$st occurrence and, in units of S_{n+1}, the occurrence times should be randomly distributed because of lack of memory and independence properties.

Recall that $D_n = \sup_{0<u<1} |F_n^*(u) - u|$ where $F_n^*(u)$ is the empirical distribution of the uniform sample $U_1, \ldots, U_n$. We then have

$$\begin{aligned} D_n &= \sqrt{n} \sup_{0<u<1} |F_n^*(u) - u| = \sqrt{n} \max_{k \le n} \left| Y_{(k)} - \frac{k}{n} \right| \\ &\stackrel{d}{=} \sqrt{n} \max_{k \le n} \left| \frac{S_k}{S_{n+1}} - \frac{k}{n} \right| = \frac{n}{S_{n+1}} \max_{k \le n} \left| \frac{S_k - k}{\sqrt{n}} - \frac{k}{n} \frac{S_{n+1} - n}{\sqrt{n}} \right| \end{aligned}$$

Theorem 15 *Let F be defined on R. There exists a finite positive constant C (not depending on F) such that*

$$P(D_n < d) \le C \exp(-2nd^2), \quad d > 0,$$

for all $n = 1, 2, \ldots$.

Moreover,

Theorem 16 *Let F be 1-dimensional and continuous.. Then*

$$\lim_{n \to \infty} P(n^{1/2} D_n < d) = 1 - 2 \sum_{j=1}^{\infty} (-1)^{j+1} \exp(-2j^2 d^2), \quad d > 0,$$

for all $n = 1, 2, \ldots$.

Secondly, we consider *goodness of fit* test statistics based on the sample distribution function. The null hypotheseis in the simple case is $H_0 : F = F_0$, where F_0 is specified. A useful procedure is the *Kolmogorov-Smirnov test statistic*

$$\triangle_n = \sup_{-\infty < x < \infty} |F_n(x) - F_0(x)|,$$

which reduces to D_n under the null hypothesis. More broadly, a class of such statistics is obtained by introducing weight functions:

$$\sup_{-\infty < x < \infty} |w(x)[F_n(x) - F_0(x)]|.$$

Another important class of statistics is based on the *Cramer-von Mises test statistic*

$$C_n = n \int_{-\infty}^{\infty} [F_n(x) - F_0(x)]^2 dF_0(x)$$

and takes the general form $n \int w(F_0(x))[F_n(x) - F_0(x)]^2 dF_0(x)$. For example, for $w(t) = [t(1-t)]^{-1}$, each duscrepancy $F_n(x) - F(x)$ becomes weighted by the reciprocal of its standard deviation (under H_0), yielding the Anderson-Darling statistic.

3.2 Stieltjes Integral

If $[a, b]$ is a compact interval, a set of points $\{x_0, x_1, \ldots, x_n\}$, satisfying the inequalities

$$a = x_0 < x_1 < \cdots < x_n = b,$$

is called a partition of $[a, b]$. Write $\triangle f_k = f(x_k) - f(x_{k-1})$ for $k = 1, 2 \ldots, n$. If there exists a positive number M such that

$$\sum_{k=1}^{n} | \triangle f_k| \leq M$$

for all partitions of $[a, b]$, then f is said to be of bounded variation on $[a, b]$. Let $F(x)$ be a function of bounded variation and continuous from the left such as a distribution function. Given a finite interval (a, b) and a function $f(x)$ we can form the sum

$$J_n = \sum_{i=1}^{n} f(x_i^{'})[F(x_i) - F(x_{i-1})]$$

for a division of (a, b) by points x_i such that $a < x_1 < \cdots < x_n < b$ and arbitrary $x_i^{'} \in (x_{i-1}, x_i)$. It may be noted that in the Riemann integration a similar sum is considered with the length of the interval $(x_i - x_{i-1})$ instead of $F(x_i) - F(x_{i-1})$. If $J = \lim_{n \to \infty} J_n$ as the length of each interval $\to 0$, then J is called the Stieltjes integral of $f(x)$ with respect to $F(x)$ and is denoted by

$$J = \int_a^b f(x) dF(x).$$

The improper integral is defined by

$$\lim_{a \to -\infty, b \to \infty} \int_a^b f(x) dF(x) = \int f(x) dF(x).$$

One point of departure from Riemann integration is that it is necessary to specify whether the end points are included in the integration or not. From the definition it is easily shown that

$$\int_a^b f(x) dF(x) = \int_{a^+0}^b f(x) dF(x) + f(a)[F(a^+0) - F(a)]$$

where a^+0 indicates that the end point a is not included. If $F(x)$ jumps at a, then

$$\int_a^b f(x) dF(x) - \int_{a^+0}^b f(x) dF(x) + f(a)[F(a^+0) - F(a)],$$

so that the integral taken over an interval that reduces to zero need not be zero. We shall follow the convention that the lower end point is always included but not the upper end point. With this convention, we see that $\int_a^b dF(x) = F(b) - F(a)$. If there exists a function $p(x)$ such that $F(x) = \int_{-\infty}^x p(x) dx$, the Stieltjes integral reduces to a Riemann integral

$$\int f(x) dF(x) = \int f(x) p(x) dx.$$

Theorem 17 *Let α be a step function defined on $[a, b]$ with jumps α_k at x_k. Then $\int_a^b f(x) d\alpha(x) = \sum_{k=1}^{n} f(x_k) \alpha_k$.*

4 Sample Density Functions

Recall that $F^{'} = f$. This suggests that we can use the *derivative* of F_n to estimate f. In particular, we consider

$$f_n(x) = \frac{F_n(x+b_n) - F_n(x-b_n)}{2b_n}.$$

Observe that $2nb_n f_n(x) \sim Bin(n, F(x+b_n) - F(x-b_n))$ and we have

$$\begin{aligned}
Ef_n(x) &= \frac{1}{2nb_n} n\left[F(x+b_n) - F(x-b_n)\right] \\
&\approx \frac{1}{2nb_n} n \cdot 2b_n f(x) \text{ if } b_n \to 0, \\
var(f_n(x)) &= \frac{1}{(2nb_n)^2} n\left[F(x+b_n) - F(x-b_n)\right]\left[1 - F(x+b_n) + F(x-b_n)\right] \\
&\approx \frac{1}{4nb_n^2} 2b_n f(x) = \frac{f(x)}{2nb_n} \text{ if } nb_n \to \infty.
\end{aligned}$$

A natural question to ask is to find an optimal choice of b_n. To do so, we need to find the magnitude of bias, $Ef_n(x) - f(x)$.

Since the above estimate can be viewed as the widely used kernel estimate with kernel $W(z) = 1/2$ if $|z| \le 1$ and $= 0$ otherwise, we will find the the magnitude of bias for the following kernel estimate of $f(x)$ instead.

$$f_n(x) = \frac{1}{nb_n} \sum_{i=1}^n W\left(\frac{x - X_i}{b_n}\right),$$

where W is an integrable nonnegative weight function. Typically, W are chosen to be a density function, $\int tW(t)dt = 0$, and $\int t^2 W(t)dt = \alpha \neq 0$. We have

$$\begin{aligned}
Ef_n(x_0) &= \frac{1}{b_n} EW\left(\frac{x_0 - X}{b_n}\right) \\
&= \frac{1}{b_n} \int W\left(\frac{x_0 - x}{b_n}\right) f(x)dx = \int W(t) f(x_0 - b_n t)dt \\
&= \int W(t)\left[f(x_0) - b_n t f^{'}(x_0 - \theta_t b_n t)\right] dt \\
&= f(x_0) - b_n \int tW(t) f^{'}(x_0 - \theta_t b_n t)dt.
\end{aligned}$$

When $\int tW(t)dt \neq 0$, we have

$$Ef_n(x_0) = f(x_0) - b_n f^{'}(x_0) \int tW(t)dt + o(b_n).$$

When $\int tW(t)dt = 0$ and $\int t^2 W(t)dt \neq 0$, we have

$$\begin{aligned}
Ef_n(x_0) &= \int W(t)\left[f(x_0) - b_n t f^{'}(x_0) + \frac{b_n^2}{2} t^2 f^{"}(x_0 - \theta_t b_n t)\right] dt \\
&= f(x_0) + \frac{b_n^2}{2} f^{"}(x_0) \int t^2 W(t)dt + o(b_n^2).
\end{aligned}$$

Therefore, $b_n = O(n^{-1/3})$ when $\int tW(t)dt \neq 0$ (i.e., Assume that $f^{'}$ exists.), and $b_n = O(n^{-1/5})$ when $\int tW(t)dt = 0$, $\int t^2 W(t)dt \neq 0$ (i.e., Assume that $f^{"}$ exists.)

5 Applications of Order Statistics

The following is a brief list of settings in which order stastistics might have a significant role.

1. **Robust Location Estimates.** Suppose that n independent measurements are available, and we wish to estimate their assumed common mean. It has long been recognized that the sample mean, though attractive from many viewpoints, suffers from an extreme sensitivity to outliers and model violations. Estimates based on the median or the average of central order statistics are less sensitive to model assumptions. A particularly well-known application of this observation is the accepted practice of using trimmed means (ignoring highest and lowest scores) in evaluating Olympic figure skating performances.

2. **Detection of Outliers.** If one is confronted with a set of measurements and is concerned with determining whether some have been incorrectly made or reported, attention naturally focuses on certain order statistics of the sample. Usually the largest one or two and/or the smallest one or two are deemed most likely to be outliers. Typically we ask questions like the following: If the observations really were iid, what is the probability that the latgest order statistic would be as large as the suspiciously large value we have observed?

3. **Censored Sampling.** Considser *life-testing* experiments, in which a fixed number n of items are placed on test and the experiment is terminated as soon as a prescribed number r have failed. The observed lifetimes are thus $X_{1:n} \leq \cdots \leq X_{r:n}$, whereas the lifetimes $X_{r+1:n} \leq \cdots \leq X_{n:n}$ remain unobserved.

4. **Waiting for the Big One.** Disastrous floods and destructive earthquakes recur throughout history. Dam construction has long focused on so called 100-year flood. Presumably the dams are built big enough and strong enough to handle any water flow to be encountered except for a level expected to occur only once every 100 years. Architects in California are particularly concerned with construction designed to withstand "the big one," presumably an earthquake of enormous strength, perhaps a "100-year quake." Whether one agrees or not with the 100-year diaster philosophy, it is obvious that designers of dams and skycrapers, and even doghouses, should be concerned with the distribution of large order statistics from a possibly dependent, possibly not identically distributed sequence.

   After the disastrous flood of February 1st, 1953, in which the sea-dikes broke in several parts of the Netherlands and nearly two thousand people were killed, the Dutch government appointed a committee (so-called Delta-committee) to recommend on an appropriate level for the dikes (called Delta-level since) since no specific statistical study had been done to fix a safer level for the sea-dikes before 1953. The Dutch government set as the standard for the sea-dikes that at any time in a given year the sea level exceeds the level of the dikes with probability $1/10,000$. A statistical group from the Mathematical Centre in Amsterdam headed by D. van Dantzig showed that high tides occurring during certain dangerous windstorms (to ensure independence) within the dangerous wintermonths December, January and February (for homogenity) follow closely an exponential distribution if the smaller high tides are neglected.

If we model the annual maximum flood by a random variable Z, the Dutch government wanted to determine therefore the $(1-q)$-quantile

$$F^{-1}(1-q) = \inf\{t \in R : F(t) \geq 1-q\}$$

of Z, where F denotes the distribution of Z and q has the value 10^{-4}.

5. **Strength of Materials.** The adage that a chain is no longer than its weakest link underlies much of the theory of strength of materials, whether they are threads, sheets, or blocks. By considering failure potential in infinitesimally small sections of the material, on quickly is led to strength distributions associated with limits of distributions of sample minima. Of course, if we stick to the finite chain with n links, its strength would be the minimum of the strengths of its n component links, again an order statistic.

6. **Reliability.** The example of a cord composed of n threads can be extended to lead us to reliability applications of order statistics. It may be that failiure of one thread will cause the cord to break (the weakest link), but more likely the cord will function as long as k (a number less than n) of the threads remains unbroken. as such it is an example of a k out of n system commonly discussed in reliability settings. With regard to tire failure, the automobile is often an example of a 4 out of 5 system (remember the spare). Borrowing on terminology from electrical systems, the n out of n system is also known as a series system. Any component failure is disastrous. The 1 out of n system is known as a parallel system; it will function as long as any of the components survives. The life of the k out of n system is clearly $X_{n-k+1:n}$, the $(n-k+1)$st largest of the component lifetimes, or, equivalently, the time until less than k components are functioning. But, in fact, they can be regarded as perhaps complicated hierarchies of parallel and series subsystems, and the study of system lifetime will necessarily involve distributions of order statistics.

7. **Quality Control.** Take a comfortable chair and watch the daily production of Snickers candy bars pass by on the conveyor belt. Each candy bar should weigh

2.1 ounces; just a smidgen over the weight stated on the wrapper. No matter how well the candy pouring machine was just adjusted at the beginning of the shift, minor fluctuations will occur, and potentially major aberrations might be encountered (if a peanut gets stuck in the control valve). We must be alert for correctable malfunctions causing unreasonable variation in the candy bar weight. Enter the quality control man with his $\bar{X}$ *and R charts or his median and R charts.* A sample of candy bars is weighed every hour, and close attention is paid to the order statistics of the weights so obtained. If the median (or perhaps the mean) is far from the target value, we must shut down the line. Either we are turning out skinny bars and will hear from disgruntled six-year-olds, or we are turning out overweight bars and wasting money (so we will hear from disgruntled management). Attention is also focused on the sample range, the largest minus the smallest weight. If it is too large, the process is out of control, and the widely fluctuating candy bar weights will probably cause problems further down the line. So again we stop and seek to identify a correctable cause before restarting the Snickers line.

8. **Selecting the Best.** Field trials of corn varieties involved carefully balanced experiments to determine which of several varieties is most productive. Obviously we are concerned with the maximum of a set of probability not identically distributed variables in such a setting. The situation is not unlike the one discussed earlier in the context of identification of outliers. In the present situation, the outlier (the best variety) is, however, good and merits retention (rather than being discarded or discounted as would be the case in the usual outliers setting). Another instance in biology in which order statistics play a clear role involves *selective breeding by culling.* Here perhaps the best 10% with respect to meatiness of the animals in each generation are raised for breeding purposes.

   Geneticists and breeders measure the effectiveness of a selection program by comparing the average of the selected group with the population average. This difference, expressed in standard deviation units, is known as the *selection differential.* Usually, the selected group consists of top or bottom order statistics. Without loss of generality let us assume the top k order statistics are selected. Then the selection differential is

$$D_{k,n}(\mu,\sigma) = \frac{1}{\sigma}\left\{\left(\sum_{i=n-k+1}^{n} X_{i:n}\right)\frac{1}{k} - \mu\right\},$$

   where μ and σ are the population mean and standard deviation, respectively. Breeders quite often use $E(D_{k,n}(\mu,\sigma))$ or $D_{k,n}(\mu,\sigma)$ as a measure of improvement due to selection. If $k = n - [np]$, then except for a change of location and scale,

$D_{k,n}(\mu, \sigma)$ is a trimmed mean with $p_1 = p$ and $p_2 = 1$ where p_1 and $1 - p_2$ represent the proportion of the sample trimmed at either ends.

9. **Inequality of Measurement.** The income distribution in Bolivia (where a few individuals earn most of the money) is clearly more unequal than that of Sweden (where progressive taxation has a leveling effect). How does one make such statements precise? The usual approach involves order statistics of the corresponding income distributions. The particular device used is call a *Lorenz curve.* It summarizes the percent of total income accruing to the poorest p percent of the population for various values of p. Mathematically this is just the scaled integral of the empirical quantile function, a function with jump $X_{(i:n)}$ at the point i/n; $i = 1, 2, \ldots, n$ (where n is the number of individual incomes in the population). A high degree of convexity in the Lorenz curve signals a high degree of inequality in the income distribution.

10. **Olympic Records.** Bob Beamon's 1968 long jump remains on the Olympic record book. Few other records last that long. If the best performances in each Olympic Games were modeled as independent identically distributed random variables, then records would become more and more scarce as time went by. Such is not the case. The simplest explanation involves improving and increasing populations Thus the 1964 high jumping champion was the best of, say, N_1 active international-caliber jumpers. In 1968 there were more high-caliber jumpers of probably higher caliber So we are looking, most likely, at a sequence of not identically distributed random variables. But in any case we are focusing on maximum.

11. **Alocation of Prize Money in Tournaments.**

12. **Characterizations and Goodness of Fit.** The exponential distribution is famous for its so-called lack of memory. The usual model involves a light bulb or other electronic device. The argument goes that a light bulb that has been in service 20 hours is no more and no less likely to fail in the next minute than one that has been in service for, say, 5 hours, or even, for that metter, than a brand new bulb. Such a curious distributional situation is reflected by the order statistics from exponential samples. For example, if $X_1, \ldots, X_n$ are iid exponential, then their spacings $(X_{(i)} - X_{(i-1)})$ are again exponential and, remarkably, are independent. It is only in the case of exponential random variables that such spacings properties are encountered. A vast literature of exponential characterizations and related goodness-of-fit tests has consequently developed. We remark in passing that most tests of goodness of fit for any parent distribution implicitly involve order statistics, since they often focus on deviations between the empirical quantile

function and the hypothesized quantile function.

6 Extreme Value Theory

At the beginning of this Chapter, we describe some asymptotic results concerning the pth sample quantile $X_{(r)}$, where as $n \to \infty$, $r = [np]$ and $0 < p < 1$. This is the so-called *central order statistic*. When F is absolutely continuous with finite positive pdf at $F^{-1}(p)$, $X_{(r)}$ is asymptotically normal after suitable normalization.

When either r or $n-r$ is fixed and the sample size $n \to \infty$, $X_{(r)}$ is called *extreme order statistic*. We treat the extreme case in Section 6.1. One important message about extreme order statistics is that if the limit distribution exists, it is nonnormal and depends on F only through its tail behavior.

When both r or $n-r$ approach infinity but $r/n \to 0$ or 1, $X_{(r)}$ is called *intermediate order statistic*. The limiting distribution of $X_{(r)}$ depends on the rate of growth of r, and it could be either normal or nonnormal. Refer to Falk (1989) for further details.

6.1 Possible Limiting Distributions for the Sample Maximum

Now we discuss the possible nondegenerate limit distributions for $X_{(n)}$. Denote the upper limit of the support of F by $F^{-1}(1)$. Suppose that $F^{-1}(1) < \infty$. Observe that $F_{X_{(n)}}(x) = [F(x)]^n$. Then $X_{(n)} \xrightarrow{P} F^{-1}(1)$ can be established easily. Recall an elementary result in Calculus:

$$\left(1 - \frac{c_n}{n}\right)^n \to \exp(-c) \qquad \text{if and only if } \lim_n c_n = c,$$

where $\{c_n\}$ be a sequence of real numbers. Consider the normalization $[X_{(n)} - F^{-1}(1)]/b_n$ and $n^{-1}c_n = 1 - F(F^{-1}(1) + b_n x)$. Depending on the tail of F, it is expected that different norming constants b_n and asymptotic distribution emerge. When $F^{-1}(1) = \infty$, it is not clear how to find the limiting distribution of $X_{(n)}$ in general. In order to hope for a nondegenerate limit distribution, we will have to appropriately normalize or standardize $X_{(n)}$. In other words, we look at the sequence $\{(X_{(n)} - a_n)/b_n, n \geq 1\}$ where a_n represents a shift in location and $b_n > 0$ represents a change in scale. The cdf of the normalized $X_{(n)}$ is $F^n(a_n + b_n x)$. We will now ask the following questions:

(i) Is it possible to find a_n and $b_n > 0$ such that $F^n(a_n + b_n x) \to G(x)$ at all continuity points of a nondegenerate cdf G?

(ii) What kind of cdf G can appear as the limiting cdf?

(iii) How is G is related to F; that is, given F can we identify G?

(iv) What are appropriate choices for a_n and b_n in (i)?

In order to answer these questions precisely and facilitate the ensuing discussion, we introduce two definitions

Definition (domain of maximal attraction). A cdf F (discrete or absolutely continuous) is said to belong to the *domain of maximal attraction* of a nondegenerate cdf G if there exist sequence $\{a_n\}$ and $\{b_n > 0\}$ such that

$$\lim_{n\to\infty} F^n(a_n + b_n x) = G(x)$$

at all continuity points of $G(x)$. If the above holds, we will write $F \in \mathcal{D}(G)$.

Definition Two cdfs F_1 and F_2 are said to be of the *same type* if there exist constants a_0 and $b_0 > 0$ such that $F_1(a_0 + b_0 x) = F_2(x)$.

If the random variable $(X_{(n)} - a_n)/b_n$ has a limit distribution for some choice of constants $\{a_n\}$, $\{b_n\}$, then the limit distribution must be of the form G_1, G_2, or G_3, where

$$\begin{aligned} G_1(x;\alpha) &= \begin{cases} 0, & x \le 0, \\ \exp\left(-x^{-\alpha}\right), & x > 0, \end{cases} \\ G_2(x;\alpha) &= \begin{cases} \exp\left(-(-x)^{\alpha}\right), & x < 0, \\ 1, & x \ge 0, \end{cases} \end{aligned}$$

and

$$G_3(x) = \exp\left(-e^{-x}\right), -\infty < x < \infty.$$

(In G_1 and G_2, α is a positive constant.) This result was established by Gnedenko (1943), following less rigorous treatments by earlier authors.

Note that the above three families of distributions may be related to the exponential distribution as follows. If Y follows an exponential distribution, then $G_1(x, \alpha)$ is the distribution function of $Y^{-1/\alpha}$, $G21(x, \alpha)$ is the distribution function of $-Y^{-1/\alpha}$, and $G_3(x)$ is the distribution function of $-\log(Y)$.

6.2 Asymptotic Theory

As a motivated example, consider that $X_1, \ldots, X_n$ are uniformly distributed over $[0, 1]$. Because the uniform distribution has an upper terminal 1, it follows easily that $X_{(n)} \xrightarrow{P} 1$. Now we would like to know *how fast* $X_{(n)}$ tends to 1. Alternatively, we attempt to choose appropriate constants $\{b_n\}$ such that $Z_n = (1 - X_{(n)})/b_n$ will have a nondegenerate distribution. Note that

$$P(Z_n \le z) = \begin{cases} 0 & \text{if } z < 0 \\ 1 - (1 - b_n z)^n & \text{if } b_n^{-1} > z > 0 \end{cases}.$$

Choose $b_n = n^{-1}$ and then

$$(1 - z/n)^n \to \exp(-z).$$

This concludes that $n(1 - X_{(n)})$ has a limiting exponential distribution with unit mean.

We now consider the case that X has an upper terminal θ and $1 - F(x) \sim a(\theta - x)^c$ for some $c > 0$ as $x \to \theta$. Again, consider $Z_n = (\theta - X_{(n)})/b_n$ where b_n is to be chosen. We have the cumulative distribution function

$$1 - [F(\theta - b_n z)]^n \sim 1 - [1 - a(b_n z)^c]^n$$

for $z \geq 0$ and $b_n z = O(1)$. We take $ab_n^c = n^{-1}$ to show that $(\theta - X_{(n)})(na)^{1/c}$ has the limiting p.d.f. $cz^{c-1}\exp(-z^c)$.

Next, suppose that X does not have an upper terminal and that as $x \to \infty$

$$1 - F(x) \sim ax^{-c} \text{ for some } c > 0.$$

Note that F is a Pareto(θ) cdf when $1 - F(x) = x^\theta$, $x \geq 1$, $\theta > 0$, where θ is the shape parameter. Consider $Z_n = X_{(n)}/b_n$, we have that

$$Pr(Z_n \leq z) = [1 - \{1 - F(b_n z)\}]^n \sim \{1 - a(b_n z)^{-c}\}^n.$$

The choice $b_n = (an)^{1/c}$ then gives the limiting result

$$Pr(Z_n \leq z) \sim \exp(-z^{-c}).$$

This covers as a special case the Cauchy distribution with $c = 1$.

Finally, suppose that $1 - F(x)$ tends to zero exponentially fast as $x \to \infty$. We return to the more general standardization $Z_n = (X_{(n)} - a_n)/b_n$, when

$$Pr(Z_n \leq z) = (1 - \exp[\log\{1 - F(a_n + b_n z)\}])^n . \tag{5}$$

The crucial values of $X_{(n)}$ are those close to $F^{-1}(1 - 1/n)$, which we take to be a_n. Observe that

$$1 - F(a_n + b_n z) \approx 1 - \{F(a_n) + b_n z f(a_n))\} = n^{-1}(1 - nb_n z f(a_n))$$

and $nb_n z f(a_n) \approx 0$. Then, expanding (**??**) in a Taylor series, we obtain

$$Pr(Z_n \leq z) \sim \left[1 - n^{-1} exp\{-b_n z n f(a_n)\}\right]^n,$$

so that with $b_n^{-1} = nf(a_n)$, we have

$$Pr(Z_n \leq z) \sim \exp(-e^{-z}),$$

the corresponding p.d.f. being $\exp(-z - e^{-z})$.

We now use two examples to show how to find a_n and b_n.

Example 1 (Beta Distribution) Let F be a Beta(α, β) cdf with density

$$f(x) = cx^{\alpha-1}(1 - x)^{\beta-1} I(0 < x < 1),$$

where $c = \Gamma(\alpha+\beta)/(\Gamma(\alpha)+\Gamma(\beta))$. As $x \to 1^-$, $f(x) \approx c(1-x)^{\beta-1}$, and

$$1 - F(x) \approx c\int_x^1 (1-u)^{\beta-1}du = c(1-x)^\beta/\beta.$$

Find appropriate choices of a_n and b_n.

Sol. Note that $a_n = 1$ and $1 - F(1-b_n) = n^{-1}$. This leads to $b_n^\beta \approx \beta/(nc)$, so we may take

$$b_n = \left(\frac{\Gamma(\alpha)\Gamma(\beta+1)}{n\Gamma(\alpha+\beta)}\right)^{1/\beta}.$$

For $UNIF(0,1)$, $\alpha = \beta = 1$. Hence $b_n = n$.

Example 2 (Weibull Distribution) Let F be a Weibull(α) cdf where α is the shape parameter, that is $F(x) = 1 - \exp(-x^\alpha)$ for $x > 0$ and $\alpha > 0$. Find appropriate choices of a_n and b_n.

Sol. Note that $F(a_n) = 1 - n^{-1}$ which leads to $a_n = (\log n)^{1/\alpha}$. Hence, $b_n^{-1} = nf(a_n) = \alpha(\log n)^{(\alpha-1)/\alpha}$.

For exponential distribution, $\alpha = 1$ and hence $a_n = \log n$ and $b_n = 1$.

Example 3 (Standard Normal Distribution) Let F be a standard normal distribution. Find appropriate choices of a_n and b_n.

Sol. Using L'Hospital's rule, we obtain

$$1 - F(x) \approx \frac{1}{x}f(x)$$

when x is large. Note that $F(a_n) = 1-n^{-1}$ which leads to $a_n = \sqrt{2\log n}-(1/2)\log(4\pi\log n)/\sqrt{2\log n}$. Hence, $b_n^{-1} = nf(a_n) = 1/\sqrt{2\log n}$. Refer to p.99 of Ferguson (1996) for the detail of calculation.

As we know, Cauchy distribution is a special case of t-distribution. For t_v distribution, it has density

$$f(x) = \frac{c}{(v+x^2)^{(v+1)/2}} \approx cx^{-(v+1)}.$$

Can you figure out a_n and b_n? Refer to p.95 of Ferguson (1996) for the detail of calculation.

7 Linear Estimation of Location

Let $X_1, X_2, \ldots, X_n$ be a random sample from the absolutely continuous cdf $F(y;\theta)$ where θ is the location parameter. For several distributions, linear functions of order statistics provide good estimator of θ. Suppose a_{in}'s form a (double) sequence of constants. The statistic

$$L_n = \sum_{i=1}^n a_{in}X_{(i:n)}$$

is called an L statistic. When used as an estimator, it is often referred to as an L estimator.

A wide variety of limit distributions are possible for L_n. For example, when a_{in} is zero for all but one i, $1 \le i \le n$, L_n is a function of a single order statistic. We have seen in previous section the possible limit distributions for $X_{(i:n)}$ which depend on how i is related to n. If L_n is a function of a finite number of central order statistics, the limit distribution is normal, under mild conditions on F.

Even when the a_{in} are nonzero for many i's, L_n turns out to be asymptotically normal when the weights are reasonably smooth. In order to make this requirement more precise, let us suppose a_{in} is of the form $J(i/(n+1))/n$, where $J(u)$, $0 \leq u \leq 1$, is the associated *weight function*. In other words, we assume now that L_n can be expressed as

$$L_n = \frac{1}{n} \sum_{i=1}^{n} J\left(\frac{i}{n+1}\right) X_{(i:n)}.$$

The asymptotic normality of L_n has been established either by putting conditions on the weights or the weight function.

Suppose that μ and the population median $(= F^{-1}(1/2))$ coincide. Let us assume that the variance σ^2 is finite and $f(\mu)$ is finite and positive. For simplicity, let us take the sample size n to be odd. While $\bar{X}_n$ is an uubiased, asymptotically normal estimator of μ with variance $Var(\bar{X}_n) = \sigma^2/n$, $\tilde{X}_n$ is asymptotically unbiased and normal. If the population pdf is symmetric (around μ), $\tilde{X}_n$ is also unbiased. Further, $Var(\tilde{X}_n) \approx \{4n[f(\mu)]^2\}^{-1}$. Thus, as an estimator μ, the sample median would be more efficient than the sample mean, at least asymptotically, whenever $[2f(\mu)]^{-1} < \sigma$. This condition is satisfied, for example, for the Laplace distribution with pdf $f(x;\mu) = \frac{1}{2}\exp(-|x-\mu|)$, $-\infty < x < \infty$. For this distribution, we know that $\tilde{X}_n$ is the maximum-likelihood estimator of μ, and that it is robust against outliers. Further, since $f(\mu) = 1/2$, we can construct confidence intervals for μ using the fact that $\sqrt{n}(\tilde{X}_n - \mu)$ is asymptotically standard normal.

References

[1] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1992). *A First Course in Order Statistics.* John Wiley & Sons, Inc.

[2] Chung, K.L. (1974). *A Course in Probability Theory.* 2nd ed., Academic Press, New York.

[3] Cramer, H. and Wold, H. (1936). Some theorems on distribution functions. *J. London Math. Soc.* **11** 290-295.

[4] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27** 642-669.

[5] Falk, M. (1989). A note on uniform asymptotic normality of intermediate order statistics. *Ann. Inst. Statist. Math.* **41** 19-29.

[6] Ferguson, T.S. (1996). *A Course in Large Sample Theory.* Chapman & Hall.

[7] Gnedenko, B.V. (1943). Sue la distribution limite du terme maximum d'un serie aleatoire, *Ann. Math.* **44** 423-453.

[8] Wilk, M.B. and Gnanadesikan, R. (1968). Probability plotting methods for the analysis of data. *Biometrika* **55** 1-17.