Chapter 6
Finite Sample Theory of Order Statistics and Extremes

The ordered values of a sample of observations are called the order statistics of the sample, and the smallest and the largest are called the extremes. Order statistics and extremes are among the most important functions of a set of random variables that we study in probability and statistics. There is natural interest in studying the highs and lows of a sequence, and the other order statistics help in understanding the concentration of probability in a distribution, or equivalently, the diversity in the population represented by the distribution. Order statistics are also useful in statistical inference, where estimates of parameters are often based on some suitable functions of the order statistics. In particular, the median is of very special importance. There is a well-developed theory of the order statistics of a fixed number n of observations from a fixed distribution, as also an asymptotic theory where n goes to infinity. We discuss the case of fixed n in this chapter. A distribution theory for order statistics when the observations are from a discrete distribution is complex, both notationally and algebraically, because of the fact that there could be several observations which are actually equal. These ties among the sample values make the distribution theory cumbersome. We therefore concentrate on the continuous case.

Principal references for this chapter are the books by David (1980), Reiss (1989), Galambos (1987), Resnick (2007), and Leadbetter et al. (1983). Specific other references are given in the sections.

6.1 Basic Distribution Theory

Definition 6.1. Let $X_1, X_2, \ldots, X_n$ be any n real-valued random variables. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the ordered values of $X_1, X_2, \cdots, X_n$. Then, $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are called the *order statistics* of $X_1, X_2, \ldots, X_n$.

Remark. Thus, the minimum among $X_1, X_2, \ldots, X_n$ is the first-order statistic, and the maximum the nth-order statistic. The middle value among $X_1, X_2, \ldots, X_n$ is called the median. But it needs to be defined precisely, because there is really no middle value when n is an even integer. Here is our definition.

A. DasGupta, *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*, Springer Texts in Statistics, DOI 10.1007/978-1-4419-9634-3_6,

Definition 6.2. Let $X_1, X_2, \ldots, X_n$ be any n real-valued random variables. Then, the *median* of $X_1, X_2, \ldots, X_n$ is defined to be $M_n = X_{(m+1)}$ if $n = 2m + 1$ (an odd integer), and $M_n = X_{(m)}$ if $n = 2m$ (an even integer). That is, in either case, the median is the order statistic $X_{(k)}$ where k is the smallest integer $\geq \frac{n}{2}$.

Example 6.1. Suppose .3, .53, .68, .06, .73, .48, .87, .42, .89, .44 are ten independent observations from the $U[0, 1]$ distribution. Then, the order statistics are .06, .3, .42, .44, .48, .53, .68, .73, .87, .89. Thus, $X_{(1)} = .06$, $X_{(n)} = .89$, and because $\frac{n}{2} = 5$, $M_n = X_{(5)} = .48$.

An important connection to understand is the connection order statistics have with the *empirical CDF*, a function of immense theoretical and methodological importance in both probability and statistics.

Definition 6.3. Let $X_1, X_2, \ldots, X_n$ be any n real-valued random variables. The *empirical CDF* of $X_1, X_2, \ldots, X_n$, also called the empirical CDF of the sample, is the function

$$F_n(x) = \frac{\#\{X_i : X_i \leq x\}}{n};$$

that is, $F_n(x)$ measures the proportion of sample values that are $\leq x$ for a given x.

Remark. Therefore, by its definition, $F_n(x) = 0$ whenever $x < X_{(1)}$, and $F_n(x) = 1$ whenever $x \geq X_{(n)}$. It is also a constant, namely, $\frac{k}{n}$, for all x-values in the interval $[X_{(k)}, X_{(k+1)})$. So F_n satisfies all the properties of being a valid CDF. Indeed, it is the CDF of a discrete distribution, which puts an equal probability of $\frac{1}{n}$ at the sample values $X_1, X_2, \ldots, X_n$. This calls for another definition.

Definition 6.4. Let P_n denote the discrete distribution that assigns probability $\frac{1}{n}$ to each X_i. Then, P_n is called the *empirical measure of the sample*.

Definition 6.5. Let $Q_n(p) = F_n^{-1}(p)$ be the quantile function corresponding to F_n. Then, $Q_n = F_n^{-1}$ is called the *quantile function* of $X_1, X_2, \ldots, X_n$, or the empirical quantile function.

We can now relate the median and the order statistics to the quantile function F_n^{-1}.

Proposition. *Let $X_1, X_2, \ldots, X_n$ be n random variables. Then,*

(a) $X_{(i)} = F_n^{-1}(\frac{i}{n})$;
(b) $M_n = F_n^{-1}(\frac{1}{2})$.

We now specialize to the case where $X_1, X_2, \ldots, X_n$ are independent random variables with a common density function $f(x)$ and CDF $F(x)$, and work out the fundamental distribution theory of the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$.

Theorem 6.1 (Joint Density of All the Order Statistics). *Let $X_1, X_2, \ldots, X_n$ be independent random variables with a common density function $f(x)$. Then, the joint density function of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is given by*

$$f_{1,2,\ldots,n}(y_1, y_2, \ldots, y_n) = n!\, f(y_1) f(y_2) \cdots f(y_n) I_{\{y_1 < y_2 < \cdots < y_n\}}.$$

Proof. A verbal heuristic argument is easy to understand. If $X_{(1)} = y_1, X_{(2)} = y_2, \ldots, X_{(n)} = y_n$, then exactly one of the sample values $X_1, X_2, \ldots, X_n$ is y_1, exactly one is y_2, and so on, but we can put any of the n observations at y_1, any of the other $n-1$ observations at y_2, and so on, and so the density of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is $f(y_1)f(y_2)\cdots f(y_n) \times n(n-1)\cdots 1 = n!f(y_1)f(y_2)\cdots f(y_n)$, and obviously if the inequality $y_1 < y_2 < \cdots < y_n$ is not satisfied, then at such a point the joint density of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ must be zero.

Here is a formal proof. The multivariate transformation $(X_1, X_2, \ldots, X_n) \to (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ is not one-to-one, as any permutation of a fixed $(X_1, X_2, \ldots, X_n)$ vector has exactly the same set of order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$. However, fix a specific permutation $\{\pi(1), \pi(2), \ldots, \pi(n)\}$ of $\{1, 2, \ldots, n\}$ and consider the subset $A_\pi = \{(x_1, x_2, \ldots, x_n) : x_{\pi(1)} < x_{\pi(2)} < \cdots < x_{\pi(n)}\}$. Then, the transformation $(x_1, x_2, \ldots, x_n) \to (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is one-to-one on each such A_π, and indeed, then $x_{(i)} = x_{\pi(i)}, i = 1, 2, \ldots, n$. The Jacobian matrix of the transformation is 1, for each such A_π. A particular vector $(x_1, x_2, \ldots, x_n)$ falls in exactly one A_π, and there are $n!$ such regions A_π, as we exhaust all the $n!$ permutations $\{\pi(1), \pi(2), \ldots, \pi(n)\}$ of $\{1, 2, \ldots, n\}$. By a modification of the Jacobian density theorem, we then get

$$\begin{aligned} f_{1,2,\ldots,n}(y_1, y_2, \ldots, y_n) &= \sum_\pi f(x_1)f(x_2)\cdots f(x_n) \\ &= \sum_\pi f(x_{\pi(1)})f(x_{\pi(2)})\cdots f(x_{\pi(n)}) \\ &= \sum_\pi f(y_1)f(y_2)\cdots f(y_n) \\ &= n!f(y_1)f(y_2)\cdots f(y_n). \end{aligned}$$

□

Example 6.2 (Uniform Order Statistics). Let $U_1, U_2, \ldots, U_n$ be independent $U[0,1]$ variables, and $U_{(i)}, 1 \le i \le n$, their order statistics. Then, by our theorem above, the joint density of $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ is

$$f_{1,2,\ldots,n}(u_1, u_2, \ldots, u_n) = n!I_{0<u_1<u_2<\cdots<u_n<1}.$$

Once we know the joint density of all the order statistics, we can find the marginal density of any subset by simply integrating out the rest of the coordinates, but being extremely careful in using the correct domain over which to integrate the rest of the coordinates. For example, if we want the marginal density of just $U_{(1)}$, that is, of the sample minimum, then we will want to integrate out $u_2, \ldots, u_n$, and the correct domain of integration would be, for a given u_1, a value of $U_{(1)}$, in (0,1),

$$u_1 < u_2 < u_3 < \cdots < u_n < 1.$$

So, we integrate down in the order $u_n, u_{n-1}, \ldots, u_2$, to obtain

$$\begin{aligned} f_1(u_1) &= n!\int_{u_1}^1 \int_{u_2}^1 \cdots \int_{u_{n-1}}^1 du_n du_{n-1} \cdots du_3 du_2 \\ &= n(1-u_1)^{n-1},\ 0 < u_1 < 1. \end{aligned}$$

Likewise, if we want the marginal density of just $U_{(n)}$, that is, of the sample maximum, then we want to integrate out $u_1, u_2, \ldots, u_{n-1}$, and now the answer is

$$\begin{aligned} f_n(u_n) &= n! \int_0^{u_n} \int_0^{u_{n-1}} \cdots \int_0^{u_2} du_1 du_2 \cdots du_{n-1} \\ &= n u_n^{n-1}, \quad 0 < u_n < 1. \end{aligned}$$

However, it is useful to note that for the special case of the minimum and the maximum, we could have obtained the densities much more easily and directly. Here is why. First take the maximum. Consider its CDF; for $0 < u < 1$:

$$\begin{aligned} P(U_{(n)} \le u) &= P(\cap_{i=1}^n \{X_i \le u\}) = \prod_{i=1}^n P(X_i \le u) \\ &= u^n, \end{aligned}$$

and hence, the density of $U_{(n)}$ is $f_n(u) = \frac{d}{du}[u^n] = nu^{n-1}, 0 < u < 1$.

Likewise, for the minimum, for $0 < u < 1$, the tail CDF is:

$$P(U_{(1)} > u) = P(\cap_{i=1}^n \{X_i > u\}) = (1-u)^n,$$

and so the density of $U_{(1)}$ is

$$f_1(u) = \frac{d}{du}[1 - (1-u)^n] = n(1-u)^{n-1}, \quad 0 < u < 1.$$

For a general r, $1 \le r \le n$, the density of $U_{(r)}$ works out to a Beta density:

$$f_r(u) = \frac{n!}{(r-1)!(n-r)!} u^{r-1}(1-u)^{n-r}, \quad 0 < u < 1,$$

which is the Be $(r, n-r+1)$ density.

As a rule, if the underlying CDF F is symmetric about its median, then the sample median will also have a density symmetric about the median of F; see the exercises. When n is even, one has to be careful about this, because there is no universal definition of a sample median when n is even. In addition, the density of the sample maximum will generally be skewed to the right, and that of the sample minimum skewed to the left. For general CDFs, the density of the order statistics usually will not have a simple formula in terms of elementary functions; but approximations for large n are often possible. This is treated in a later chapter. Although such approximations for large n are often available, they may not be very accurate unless n is very large; see Hall (1979).

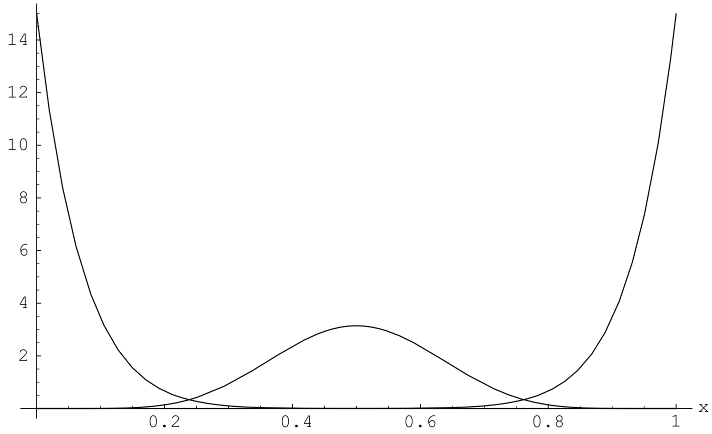

Fig. 6.1 Density of minimum, median, and maximum of U[0, 1] variables; n = 15

Above we have plotted in Fig. 6.1 the density of the minimum, median, and maximum in the $U[0, 1]$ case when $n = 15$. The minimum and the maximum clearly have skewed densities, whereas the density of the median is symmetric about .5.

6.2 More Advanced Distribution Theory

Example 6.3 (Density of One and Two Order Statistics). The joint density of any subset of the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ can be worked out from their joint density, which we derived in the preceding section. The most important case in applications is the joint density of two specific order statistics, say $X_{(r)}$ and $X_{(s)}, 1 \leq r < s \leq n$, or the density of a specific one, say $X_{(r)}$. A verbal heuristic argument helps in understanding the formula for the joint density of $X_{(r)}$ and $X_{(s)}$, and also the density of a specific one $X_{(r)}$.

First consider the density of just $X_{(r)}$. Fix u. To have $X_{(r)} = u$, we must have exactly one observation at u, another $r-1$ below u, and $n-r$ above u. This will suggest that the density of $X_{(r)}$ is

$$f_r(u) = nf(u)\binom{n-1}{r-1}(F(u))^{r-1}(1-F(u))^{n-r}$$
$$= \frac{n!}{(r-1)!(n-r)!}(F(u))^{r-1}(1-F(u))^{n-r}f(u),$$

$-\infty < u < \infty$. This is in fact the correct formula for the density of $X_{(r)}$.

Next, consider the case of the joint density of two order statistics, $X_{(r)}$ and $X_{(s)}$. Fix $0 < u < v < 1$. Then, to have $X_{(r)} = u, X_{(s)} = v$, we must have exactly one observation at u, another $r-1$ below u, one at v, $n-s$ above v, and $s-r-1$ between u and v. This will suggest that the joint density of $X_{(r)}$ and $X_{(s)}$ is

$$f_{r,s}(u,v) = nf(u)\binom{n-1}{r-1}(F(u))^{r-1}(n-r)f(v)\binom{n-r-1}{n-s}(1-F(v))^{n-s}$$
$$(F(v)-F(u))^{s-r-1}$$
$$= \frac{n!}{(r-1)!(n-s)!(s-r-1)!}(F(u))^{r-1}(1-F(v))^{n-s}$$
$$(F(v)-F(u))^{s-r-1}f(u)f(v),$$

$-\infty < u < v < \infty$.

Again, this is indeed the joint density of two specific order statistics $X_{(r)}$ and $X_{(s)}$.

The arguments used in this example lead to the following theorem.

Theorem 6.2 (Density of One and Two Order Statistics and Range). *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF $F(x)$ with density function $f(x)$. Then,*

(a) $X_{(n)}$ has the density $f_n(u) = nF^{n-1}(u)f(u), -\infty < u < \infty$.
(b) $X_{(1)}$ has the density $f_1(u) = n(1-F(u))^{n-1}f(u), -\infty < u < \infty$.
(c) For a general $r, 1 \le r \le n, X_{(r)}$ has the density

$$f_r(u) = \frac{n!}{(r-1)!(n-r)!}F^{r-1}(u)(1-F(u))^{n-r}f(u), \quad -\infty < u < \infty.$$

(d) For general $1 \le r < s \le n, (X_{(r)}, X_{(s)})$ have the joint density

$$= \frac{n!}{(r-1)!(n-s)!(s-r-1)!}(F(u))^{r-1}(1-F(v))^{n-s}(F(v)-F(u))^{s-r-1}$$
$$f(u)f(v), \quad -\infty < u < v < \infty.$$

(e) The minimum and the maximum, $X_{(1)}$ and$X_{(n)}$ have the joint density

$$f_{1,n}(u,v) = n(n-1)(F(v)-F(u))^{n-2}f(u)f(v), \quad -\infty < u < v < \infty.$$

(f) **(CDF of Range).** *$W = W_n = X_{(n)} - X_{(1)}$ has the CDF*

$$F_W(w) = n\int_{-\infty}^{\infty}[F(x+w)-F(x)]^{n-1}f(x)dx.$$

(g) **(Density of Range).** *$W = W_n = X_{(n)} - X_{(1)}$ has the density*

$$f_W(w) = n(n-1)\int_{-\infty}^{\infty}[F(x+w)-F(x)]^{n-2}f(x)f(x+w)dx.$$

Example 6.4 (Moments of Uniform Order Statistics). The general formulas in the above theorem lead to the following moment formulas in the uniform case.

In the $U[0,1]$ case,

$$E(U_{(1)}) = \frac{1}{n+1}, \quad E(U_{(n)}) = \frac{n}{n+1},$$
$$\mathrm{Var}(U_{(1)}) = \mathrm{Var}(U_{(n)}) = \frac{n}{(n+1)^2(n+2)}; \quad 1 - U_{(n)} \stackrel{\mathcal{L}}{=} U_{(1)};$$
$$\mathrm{Cov}(U_{(1)}, (U_{(n)}) = \frac{1}{(n+1)^2(n+2)},$$
$$E(W_n) = \frac{n-1}{n+1}, \quad \mathrm{Var}(W_n) = \frac{2(n-1)}{(n+1)^2(n+2)}.$$

For a general order statistic, it follows from the fact that $U_{(r)} \sim Be(r, n-r+1)$, that

$$E(U_{(r)}) = \frac{r}{n+1}; \quad \mathrm{Var}(U_{(r)}) = \frac{r(n-r+1)}{(n+1)^2(n+2)}.$$

Furthermore, it follows from the formula for their joint density that

$$\mathrm{Cov}(U_{(r)}, U_{(s)}) = \frac{r(n-s+1)}{(n+1)^2(n+2)}.$$

Example 6.5 (Exponential Order Statistics). A second distribution of importance in the theory of order statistics is the exponential distribution. The mean λ essentially arises as just a multiplier in the calculations. So, we treat only the standard exponential case.

Let $X_1, X_2, \ldots, X_n$ be independent standard exponential variables. Then, by the general theorem on the joint density of the order statistics, in this case the joint density of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is

$$f_{1,2,\ldots,n}(u_1, u_2, \ldots, u_n) = n!e^{-\sum_{i=1}^n u_i},$$

$0 < u_1 < u_2 < \cdots < u_n < \infty$. Also, in particular, the minimum $X_{(1)}$ has the density

$$f_1(u) = n(1-F(u))^{n-1} f(u) = ne^{-(n-1)u}e^{-u} = ne^{-nu},$$

$0 < u < \infty$. In other words, we have the quite remarkable result that the minimum of n independent standard exponentials is itself an exponential with mean $\frac{1}{n}$. Also, from the general formula, the maximum $X_{(n)}$ has the density

$$f_n(u) = n(1-e^{-u})^{n-1}e^{-u} = n\sum_{i=0}^{n-1}(-1)^i \binom{n-1}{i} e^{-(i+1)u}, \quad 0 < u < \infty.$$

As a result,

$$E(X_{(n)}) = n\sum_{i=0}^{n-1}(-1)^i\binom{n-1}{i}\frac{1}{(i+1)^2} = \sum_{i=1}^{n}(-1)^{i-1}\binom{n}{i}\frac{1}{i},$$

which also equals $1+\frac{1}{2}+\cdots+\frac{1}{n}$. We show later in the section on spacings that by another argument, it also follows that in the standard exponential case, $E(X_{(n)}) = 1+\frac{1}{2}+\cdots+\frac{1}{n}$.

Example 6.6 (Normal Order Statistics). Another clearly important case is that of the order statistics of a normal distribution. Because the general $N(\mu,\sigma^2)$ random variable is a location-scale transformation of a standard normal variable, we have the distributional equivalence that $(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$ have the same joint distribution as $(\mu+\sigma Z_{(1)}, \mu+\sigma Z_{(2)}, \ldots, \mu+\sigma Z_{(n)})$. So, we consider just the standard normal case.

Because of the symmetry of the standard normal distribution around zero, for any r, $Z_{(r)}$ has the same distribution as $-Z_{(n-r+1)}$. In particular, $Z_{(1)}$ has the same distribution as $-Z_{(n)}$. From our general formula, the density of $Z_{(n)}$ is

$$f_n(x) = n\Phi^{n-1}(x)\phi(x), \quad -\infty < x < \infty.$$

Again, this is a skewed density. It can be shown, either directly, or by making use of the general theorem on existence of moments of order statistics (see the next section) that every moment, and in particular the mean and the variance of $Z_{(n)}$, exists. Except for very small n, closed-form formulas for the mean or variance are not possible. For small n, integration tricks do produce exact formulas. For example,

$$E(Z_{(n)}) = \frac{1}{\sqrt{\pi}}, \quad \text{if } n=2; \quad E(Z_{(n)}) = \frac{3}{2\sqrt{\pi}}, \quad \text{if } n=3.$$

Such formulas are available for $n \le 5$; see David (1980).

We tabulate the expected value of the maximum for some values of n to illustrate the slow growth.

n	$E(Z_{(n)})$
2	.56
5	1.16
10	1.54
20	1.87
30	2.04
50	2.25
100	2.51
500	3.04
1000	3.24
10000	3.85

More about the expected value of $Z_{(n)}$ is discussed later.

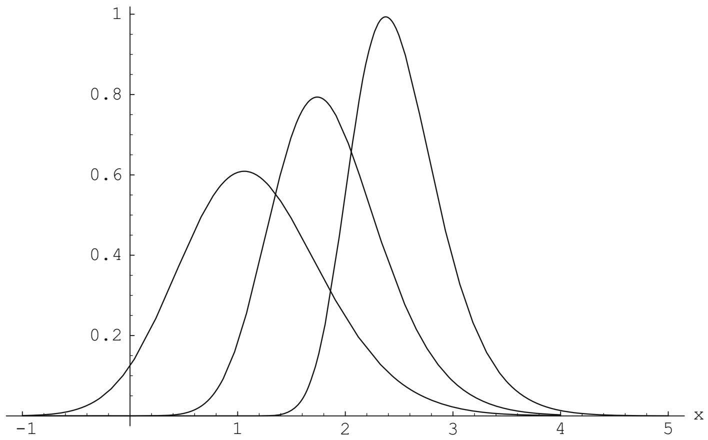

Fig. 6.2 Density of maximum of standard normals; n = 5, 20, 100

The density of $Z_{(n)}$ is plotted in Fig. 6.2 for three values of n. We can see that the density is shifting to the right, and at the same time getting more peaked. Theoretical asymptotic (i.e., as $n \to \infty$) justifications for these visual findings are possible, and we show some of them in a later chapter.

6.3 Quantile Transformation and Existence of Moments

Uniform order statistics play a very special role in the theory of order statistics, because many problems about order statistics of samples from a general density can be reduced, by a simple and common technique, to the case of uniform order statistics. It is thus especially important to understand and study uniform order statistics. The technique that makes helpful reductions of problems in the general continuous case to the case of a uniform distribution on [0,1] is one of making just the quantile transformation. We describe the exact correspondence below.

Theorem 6.3 (Quantile Transformation Theorem). *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF $F(x)$ on the real line, and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote their order statistics. Let $F^{-1}(p)$ denote the quantile function of F. Let $U_1, U_2, \ldots, U_n$ be independent observations from the $U[0,1]$ distribution, and let $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ denote their order statistics. Also let $g(x)$ be any nondecreasing function and let $Y_i = g(X_i), 1 \le i \le n$, with $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ be the order statistics of $Y_1, Y_2, \ldots, Y_n$. Then, the following equalities in distributions hold:*

(a) $F(X_1) \sim U[0,1]$, *that is,* $F(X_1)$ *and* U_1 *have the same distribution.*
We write this equality in distribution as $F(X_1) \stackrel{\mathcal{L}}{=} U_1$.

(b) $F^{-1}(U_1) \stackrel{\mathcal{L}}{=} X_1$.

(c) $F(X_{(i)}) \stackrel{\mathcal{L}}{=} U_{(i)}$.
(d) $F^{-1}(U_{(i)}) \stackrel{\mathcal{L}}{=} X_{(i)}$.
(e) $(F(X_{(1)}), F(X_{(2)}), \ldots, F(X_{(n)})) \stackrel{\mathcal{L}}{=} (U_{(1)}, U_{(2)}, \ldots, U_{(n)})$.
(f) $(F^{-1}(U_{(1)}), F^{-1}(U_{(2)}), \ldots, F^{-1}(U_{(n)})) \stackrel{\mathcal{L}}{=} (X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.
(g) $(Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}) \stackrel{\mathcal{L}}{=} (g(F^{-1}(U_{(1)})), g(F^{-1}(U_{(2)})), \ldots, g(F^{-1}(U_{(n)})))$.

Remark. We are already familiar with parts (a) and (b); they are restated here only to provide the context. The parts that we need to focus on are the last two parts. They say that any question about the set of order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ of a sample from a general continuous distribution can be rephrased in terms of the set of order statistics from the $U[0, 1]$ distribution. For this, all we need to do is to substitute $F^{-1}(U_{(i)})$ in place of $X_{(i)}$, where F^{-1} is the quantile function of F.

So, at least in principle, as long as we know how to work skillfully with the joint distribution of the uniform order statistics, we can answer questions about any set of order statistics from a general continuous distribution, because the latter is simply a transformation of the set of order statistics of the uniform. This has proved to be a very useful technique in the theory of order statistics.

As a corollary of part (f) of the above theorem, we have the following connection between order statistics of a general continuous CDF and uniform order statistics.

Corollary. *Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics of a sample from a general continuous CDF F, and $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ the uniform order statistics. Then, for any $1 \le r_1 < r_2 < \cdots < r_k \le n$,*

$$P(X_{(r_1)} \le u_1, \ldots, X_{(r_k)} \le u_k) = P(U_{(r_1)} \le F(u_1), \ldots, U_{(r_k)} \le F(u_k)),$$

$\forall u_1, \ldots, u_k$.

Several important applications of this quantile transformation method are given below.

Proposition. *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F. Then, for any r, s, $Cov(X_{(r)}, X_{(s)}) \ge 0$.*

Proof. We use the fact that if $g(x_1, x_2, \ldots, x_n), h(x_1, x_2, \ldots, x_n)$ are two functions such that they are coordinatewise nondecreasing in each x_i, then

$\text{Cov}(g(U_{(1)}, \ldots, U_{(n)}), h(U_{(1)}, \ldots, U_{(n)})) \ge 0$. By the quantile transformation theorem, $\text{Cov}(X_{(r)}, X_{(s)}) = \text{Cov}(F^{-1}(U_{(r)}), F^{-1}(U_{(s)})) \ge 0$, as $F^{-1}(U_{(r)})$ is a nondecreasing function of $U_{(r)}$, and $F^{-1}(U_{(s)})$ is a nondecreasing function of $U_{(s)}$, and hence they are also coordinatewise nondecreasing in each order statistic $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$. □

This proposition was first proved in Bickel (1967), but by using a different method. The next application is to existence of moments of order statistics.

Theorem 6.4 (On the Existence of Moments). *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F, and let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$*

be the order statistics. Let $g(x_1, x_2, \ldots, x_n)$ be a general function. Suppose $E[|g(X_1, X_2, \ldots, X_n)|] < \infty$. Then, $E[|g(X_{(1)}, X_{(2)}, \ldots, X_{(n)})|] < \infty$.

Proof. By the quantile transformation theorem above,

$$
\begin{aligned}
&E[|g(X_{(1)}, X_{(2)}, \ldots, X_{(n)})|] \\
&\quad = E[|g(F^{-1}(U_{(1)}), F^{-1}(U_{(2)}), \ldots, F^{-1}(U_{(n)}))|] \\
&\quad = n! \int_{0<u_1<u_2<\cdots<u_n<1} |g(F^{-1}(u_1), F^{-1}(u_2), \ldots, F^{-1}(u_n))| du_1 du_2 \cdots du_n \\
&\quad \le n! \int_{(0,1)^n} |g(F^{-1}(u_1), F^{-1}(u_2), \ldots, F^{-1}(u_n))| du_1 du_2 \cdots du_n \\
&\quad = n! \int_{(0,1)^n} |g(u_1, u_2, \ldots, u_n)| f(u_1) f(u_2) \cdots f(u_n) du_1 du_2 \cdots du_n \\
&\quad = n! E[|g(X_1, X_2, \ldots, X_n)|] < \infty. \qquad \square
\end{aligned}
$$

Corollary. *Suppose F is a continuous CDF such that $E_F(|X|^k) < \infty$, for some given k. Then, $E(|X_{(i)}|^k) < \infty \; \forall i \; 1 \le i \le n$.* $\square$

Aside from just the existence of the moment, explicit bounds are always useful. Here is a concrete bound (see Reiss (1989)); approximations for moments of order statistics for certain distributions are derived in Hall (1978).

Proposition. *(a)* $\forall r \le n, E(|X_{(r)}|^k) \le \frac{n!}{((r-1)!(n-r)!)} E_F(|X|^k)$;
(b) $E(|X_{(r)}|^k) < \infty \Rightarrow |F^{-1}(p)|^k p^r (1-p)^{n-r+1} \le C < \infty \; \forall p$;
(c) $|F^{-1}(p)|^k p^r (1-p)^{n-r+1} \le C < \infty \; \forall p \Rightarrow E(|X_{(s)}|^m) < \infty$, *if* $1 + \frac{mr}{k} \le s \le n - \frac{(n-r+1)m}{k}$.

Example 6.7 (Nonexistence of Every Moment of Every Order Statistic). Consider the continuous CDF $F(x) = 1 - \frac{1}{\log x}, x \ge e$. Setting $1 - \frac{1}{\log x} = p$, we get the quantile function $F^{-1}(p) = e^{\frac{1}{1-p}}$. Fix any n, k, and $r \le n$. Consider what happens when $p \to 1$.

$$
\begin{aligned}
|F^{-1}(p)|^k p^r (1-p)^{n-r+1} &= e^{\frac{k}{(}1-p)} p^r (1-p)^{n-r+1} \\
\ge C e^{\frac{k}{1-p}} (1-p)^{n-r+1} &= \quad C e^{ky} y^{-(n-r+1)},
\end{aligned}
$$

writing y for $\frac{1}{1-p}$. For any $k > 0$, as $y \to \infty (\Leftrightarrow p \to 1), e^{ky} y^{-(n-r+1)} \to \infty$. Thus, the necessary condition of the proposition above is violated, and it follows that for any r, n, k, $E(|X_{(r)}|^k) = \infty$.

Remark. The preceding example and the proposition show that existence of moments of order statistics depends on the tail of the underlying CDF (or, equivalently, the tail of the density). If the tail is so thin that the density has a finite mgf in some neighborhood of zero, then all order statistics will have all moments finite. Evaluating them in closed form is generally impossible, however. If the tail of the underlying density is heavy, then existence of moments of the order statistics, and especially the minimum and the maximum, may be a problem. It is possible for some central order statistics to have a few finite moments, and the minimum or the maximum to have none. In other words, depending on the tail, anything can happen. An especially interesting case is the case of a Cauchy density, notorious for its troublesome tail. The next result describes what happens in that case.

Proposition. *Let $X_1, X_2, \ldots, X_n$ be independent $C(\mu, \sigma)$ variables. Then,*

(a) $\forall n, E(|X_{(n)}|) = E(|X_{(1)}|) = \infty$;
(b) For $n \geq 3$, $E(|X_{(n-1)}|)$ and $E(|X_{(2)}|)$ are finite;
(c) For $n \geq 5$, $E(|X_{(n-2)}|^2)$ and $E(|X_{(3)}|^2)$ are finite;
(d) In general, $E(|X_{(r)}|^k) < \infty$ if and only if $k < \min\{r, n+1-r\}$.

Example 6.8 (Cauchy Order Statistics). From the above proposition we see that the truly problematic order statistics in the Cauchy case are the two extreme ones, the minimum and the maximum. Every other order statistic has a finite expectation for $n \geq 3$, and all but the two most extremes from each tail even have a finite variance, as long as $n \geq 5$. The table below lists the mean of $X_{(n-1)}$ and $X_{(n-2)}$ for some values of n.

n	$E(X_{(n-1)})$	$E(X_{(n-2)})$
5	1.17	.08
10	2.98	1.28
20	6.26	3.03
30	9.48	4.67
50	15.87	7.90
100	31.81	15.88
250	79.56	39.78
500	159.15	79.57

Example 6.9 (Mode of Cauchy Sample Maximum). Although the sample maximum $X_{(n)}$ never has a finite expectation in the Cauchy case, it always has a unimodal density (see a general result in the exercises). So it is interesting to see what the modal values are for various n. The table below lists the mode of $X_{(n)}$ for some values of n.

n	Mode of $X_{(n)}$
5	.87
10	1.72
20	3.33
30	4.93
50	8.12
100	16.07
250	39.98
500	79.76

By comparing the entries in this table with the previous table, we see that the mode of $X_{(n)}$ is quite close to the mean of $X_{(n-2)}$. It would be interesting to find a theoretical result in this regard.

6.4 Spacings

Another set of statistics helpful in understanding the distribution of probability are the *spacings*, which are the gaps between successive order statistics. They are useful in discerning tail behavior. At the same time, for some particular underlying distributions, their mathematical properties are extraordinarily structured, and in turn lead to results for other distributions. Two instances are the spacings of uniform and exponential order statistics. Some basic facts about spacings are discussed in this section.

Definition 6.6. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics of a sample of n observations $X_1, X_2, \ldots, X_n$. Then, $W_i = X_{(i+1)} - X_{(i)}, 1 \leq i \leq n-1$ are called the *spacings* of the sample, or the spacings of the order statistics.

6.4.1 Exponential Spacings and Réyni's Representation

The spacings of an exponential sample have the characteristic property that the spacings are all independent exponentials as well. Here is the precise result.

Theorem 6.5. *Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics from an* $\mathrm{Exp}(\lambda)$ *distribution. Then $W_0 = X_{(1)}, W_1, \ldots, W_{n-1}$ are independent, with $W_i \sim \mathrm{Exp}(\frac{\lambda}{n-i}), i =$* $0, 1, \ldots, n-1$.

Proof. The proof follows on transforming to the set of spacings from the set of order statistics, and by applying the Jacobian density theorem. The transformation $(u_1, u_2, \ldots, u_n) \rightarrow (w_0, w_1, \ldots, w_{n-1})$, where $w_0 = u_1, w_1 = u_2 - u_1, \ldots, w_{n-1} = u_n - u_{n-1}$ is one to one, with the inverse transformation $u_1 = w_0, u_2 = w_0 + w_1, u_3 = w_0 + w_1 + w_2, \ldots, u_n = w_0 + w_1 + \cdots + w_{n-1}$. The Jacobian matrix is triangular, and has determinant one. From our general theorem, the order statistics $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ have the joint density

$$f_{1,2,\ldots,n}(u_1, u_2, \ldots, u_n) = n! f(u_1) f(u_2) \cdots f(u_n) I_{\{0<u_1<u_2<\cdots<u_n<\infty\}}.$$

Therefore, the spacings have the joint density

$$\begin{aligned} &g_{0,1,\ldots,n-1}(w_0, w_1, \ldots, w_{n-1}) \\ &\quad = n! f(w_0) f(w_0 + w_1) \cdots f(w_0 + w_1 + \cdots + w_{n-1}) I_{\{w_i>0 \forall i\}}. \end{aligned}$$

This is completely general for any underlying nonnegative variable. Specializing to the standard exponential case, we get

$$\begin{aligned} g_{0,1,\ldots,n-1}(w_0, w_1, \ldots, w_{n-1}) &= n! e^{-\sum_{j=0}^{n-1} \sum_{i=0}^{j} w_i} I_{\{w_i>0 \forall i\}} \\ &= n! e^{-n w_0 - (n-1) w_1 - \cdots - w_{n-1}} I_{\{w_i>0 \forall i\}}. \end{aligned}$$

It therefore follows that $W_0, W_1, \ldots, W_{n-1}$ are independent, and also that $W_i \sim \text{Exp}(\frac{1}{n-i})$. The case of a general λ follows from the standard exponential case. $\square$

Corollary (Réyni). *The joint distribution of the order statistics of an exponential distribution with mean λ have the representation*

$$(X_{(r)})|_{r=1}^{n} \stackrel{\mathcal{L}}{=} \left(\sum_{i=1}^{r} \frac{X_i}{n-i+1}\right)|_{r=1}^{n},$$

where $X_1, \ldots, X_n$ are themselves independent exponentials with mean λ. $\square$

Remark. Verbally, the order statistics of an exponential distribution are linear combinations of independent exponentials, with a very special sequence of coefficients. In an obvious way, the representation can be extended to the order statistics of a general continuous CDF by simply using the quantile transformation.

Example 6.10 (Moments and Correlations of Exponential Order Statistics). From the representation in the above corollary, we immediately have

$$E(X_{(r)}) = \lambda \sum_{i=1}^{r} \frac{1}{n-i+1}; \quad \text{Var}(X_{(r)}) = \lambda^2 \sum_{i=1}^{r} \frac{1}{(n-i+1)^2}.$$

Furthermore, by using the same representation, for $1 \le r < s \le n$, $\text{Cov}(X_{(r)}, X_{(s)}) = \lambda^2 \sum_{i=1}^{r} \frac{1}{n-i+1^2}$, and therefore the correlation

$$\rho_{X_{(r)}, X_{(s)}} = \sqrt{\frac{\sum_{i=1}^{r} \frac{1}{(n-i+1)^2}}{\sum_{i=1}^{s} \frac{1}{(n-i+1)^2}}}.$$

In particular,

$$\rho_{X_{(1)}, X_{(n)}} = \frac{\frac{1}{n}}{\sqrt{\sum_{i=1}^{n} \frac{1}{i^2}}} \approx \frac{\sqrt{6}}{n\pi},$$

for large n. In particular, $\rho_{X_{(1)}, X_{(n)}} \to 0$, as $n \to \infty$. In fact, in large samples the minimum and the maximum are in general approximately independent.

6.4.2 Uniform Spacings

The results on exponential spacings lead to some highly useful and neat representations for the spacings and the order statistics of a uniform distribution. The next result describes the most important properties of uniform spacings and order statistics. Numerous other properties of a more special nature are known. David (1980) and Reiss (1989) are the best references for such additional properties of the uniform order statistics.

Theorem 6.6. *Let $U_1, U_2, \ldots, U_n$ be independent $U[0,1]$ variables, and $U_{(1)}, U_{(2)}, \ldots, U_{(n)}$ the order statistics. Let $W_0 = U_{(1)}, W_i = U_{(i+1)} - U_{(i)}, 1 \le i \le n-1$, and $V_i = \frac{U_{(i)}}{U_{(i+1)}}, 1 \le i \le n-1, V_n = U_{(n)}$. Let also $X_1, X_2, \ldots, X_{n+1}$ be $(n+1)$ independent standard exponentials, independent of the $U_i, i = 1, 2, \ldots, n$. Then,*

(a) $V_1, V_2^2, \ldots, V_{n-1}^{n-1}, V_n^n$ *are independent* $U[0,1]$ *variables, and* $(V_1, V_2, \ldots V_{n-1})$ *are independent of* V_n.
(b) $(W_0, W_1, \ldots, W_{n-1}) \sim \mathcal{D}(\alpha)$, *a Dirichlet distribution with parameter vector* $\alpha_{n+1 \times 1} = (1, 1, \ldots, 1)$. *That is,* $(W_0, W_1, \ldots, W_{n-1})$ *is uniformly distributed in the* n*-dimensional simplex.*
(c) $(W_0, W_1, \ldots, W_{n-1}) \stackrel{\mathcal{L}}{=} \left(\frac{X_1}{\sum_{j=1}^{n+1} X_j}, \frac{X_2}{\sum_{j=1}^{n+1} X_j}, \ldots, \frac{X_n}{\sum_{j=1}^{n+1} X_j} \right)$.
(d) $(U_{(1)}, U_{(2)}, \ldots, U_{(n)}) \stackrel{\mathcal{L}}{=} \left(\frac{X_1}{\sum_{j=1}^{n+1} X_j}, \frac{X_1 + X_2}{\sum_{j=1}^{n+1} X_j}, \ldots, \frac{X_1 + X_2 + \cdots + X_n}{\sum_{j=1}^{n+1} X_j} \right)$.

Proof. For part (a), use the fact that the negative of the logarithm of a $U[0,1]$ variable is standard exponential, and then use the result that the exponential spacings are themselves independent exponentials. That $V_1, V_2^2, \ldots, V_{n-1}^{n-1}$ are also uniformly distributed follows from looking at the joint density of $U_{(i)}, U_{(i+1)}$ for any given i. It follows trivially from the density of V_n that $V_n^n \sim U[0,1]$.

For parts (b) and (c), first consider the joint density of the uniform order statistics, and then transform to the variables $W_i, i = 0, \ldots, n-1$. This is a one-to-one transformation, and so we can apply the Jacobian density theorem. The Jacobian theorem easily gives the joint density of the $W_i, i = 0, \ldots, n-1$, and we simply recognize it to be the density of a Dirichlet with the parameter vector having each coordinate equal to one. Finally, use the representation of a Dirichlet random vector in the form of ratios of Gammas (see Chapter 4). □

Part (d) is just a restatement of part (c).

Remark. Part (d) of this theorem, representing uniform order statistics in terms of independent exponentials is one of the most useful results in the theory of order statistics.

6.5 Conditional Distributions and Markov Property

The conditional distributions of a subset of the order statistics given another subset satisfy some really structured properties. An illustration of such a result is that if we know that the sample maximum $X_{(n)} = x$, then the rest of the order statistics would act like the order statistics of a sample of size $n-1$ from the original CDF, but truncated on the right at that specific value x. Another prominent property of the conditional distributions is the *Markov property*. Again, a lot is known about the conditional distributions of order statistics, but we present the most significant and easy to state results. The best references for reading more about the conditional distributions are still David (1980) and Reiss (1989). Each of the following theorems follows on straightforward calculations, and we omit the calculations.

Theorem 6.7. *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Fix $1 \le i < j \le n$. Then, the conditional distribution of $X_{(i)}$ given $X_{(j)} = x$ is the same as the unconditional distribution of the ith order statistic in a sample of size $j-1$ from a new distribution, namely the original F truncated at the right at x. In notation,*

$$f_{X_{(j)}|X_{(i)}=x}(u) = \frac{(j-1)!}{(i-1)!(j-1-i)!}\left(\frac{F(u)}{F(x)}\right)^{i-1} \left(1-\frac{F(u)}{F(x)}\right)^{j-1-i} \frac{f(u)}{F(x)}, \quad u < x.$$

Theorem 6.8. *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Fix $1 \le i < j \le n$. Then, the conditional distribution of $X_{(j)}$ given $X_{(i)} = x$ is the same as the unconditional distribution of the $(j-i)$th order statistic in a sample of size $n-i$ from a new distribution, namely the original F truncated at the left at x. In notation,*

$$f_{X_{(j)}|X_{(i)}=x}(u) = \frac{(n-i)!}{(j-i-1)!(n-j)!}\left(\frac{F(u)-F(x)}{1-F(x)}\right)^{j-i-1} \left(\frac{1-F(u)}{1-F(x)}\right)^{n-j} \frac{f(u)}{1-F(x)}, \quad u > x.$$

Theorem 6.9 (The Markov Property). *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Fix $1 \le i < j \le n$. Then, the conditional distribution of $X_{(j)}$ given $X_{(1)} = x_1, X_{(2)} = x_2, \ldots, X_{(i)} = x_i$ is the same as the conditional distribution of $X_{(j)}$ given $X_{(i)} = x_i$. That is, given $X_{(i)}$, $X_{(j)}$ is independent of $X_{(1)}, X_{(2)}, \ldots, X_{(i-1)}$.*

Theorem 6.10. *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Then, the conditional distribution of $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$ given $X_{(n)} = x$ is the same as the unconditional distribution of the order statistics in a sample of size $n-1$ from a new distribution, namely the original F truncated at the right at x. In notation,*

$$f_{X_{(1)},\ldots,X_{(n-1)}|X_{(n)}=x}(u_1, \ldots, u_{n-1}) = (n-1)! \prod_{i=1}^{n-1} \frac{f(u_i)}{F(x)}, u_1 < \cdots < u_{n-1} < x.$$

Remark. A similar and transparent result holds about the conditional distribution of $X_{(2)}, X_{(3)}, \ldots, X_{(n)}$ given $X_{(1)} = x$.

Theorem 6.11. *Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Then, the conditional distribution of $X_{(2)}, \ldots, X_{(n-1)}$ given $X_{(1)} = x, X_{(n)} = y$ is the same as the unconditional distribution of the order statistics in a sample of size $n-2$ from a new distribution, namely the original F truncated at the left at x, and at the right at y. In notation,*

$$f_{X_{(2)},\ldots,X_{(n-1)}|X_{(1)}=x,X_{(n)}=y}(u_2, \ldots, u_{n-1}) = (n-2)! \prod_{i=2}^{n-1} \frac{f(u_i)}{F(y)-F(x)},$$
$$x < u_2 < \cdots < u_{n-1} < y.$$

Example 6.11 (Mean Given the Maximum). Suppose $X_1, X_2, \ldots, X_n$ are independent $U[0,1]$ variables. We want to find $E(\bar{X}|X_{(n)} = x)$. We use the theorem above about the conditional distribution of $X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$ given $X_{(n)} = x$.

$$\begin{aligned} E(n\bar{X}|X_{(n)} = x) &= E\left(\sum_{i=1}^{n} X_i | X_{(n)} = x\right) \\ &= E\left(\sum_{i=1}^{n} X_{(i)} | X_{(n)} = x\right) = x + E\left(\sum_{i=1}^{n-1} X_{(i)} | X_{(n)} = x\right) \\ &= x + \sum_{i=1}^{n-1} \frac{ix}{n}, \end{aligned}$$

because, given $X_{(n)} = x, X_{(1)}, X_{(2)}, \ldots, X_{(n-1)}$ act like the order statistics of a sample of size $n-1$ from the $U[0,x]$ distribution. Now summing the series, we get,

$$\begin{aligned} E(n\bar{X}|X_{(n)} = x) &= x + \frac{(n-1)x}{2} = \frac{n+1}{2}x, \\ \Rightarrow E(\bar{X}|X_{(n)} = x) &= \frac{n+1}{2n}x. \end{aligned}$$

Example 6.12 (Maximum Given the First Half). Suppose $X_1, X_2, \ldots, X_{2n}$ are independent standard exponentials. We want to find $E(X_{(2n)}|X_{(1)} = x_1, \ldots, X_{(n)} = x_n)$. By the theorem on the *Markov property*, this conditional expectation equals $E(X_{(2n)}|X_{(n)} = x_n)$. Now, we further use the representation that

$$(X_{(n)}, X_{(2n)}) \stackrel{\mathcal{L}}{=} \left(\sum_{i=1}^{n} \frac{X_i}{2n-i+1}, \sum_{i=1}^{2n} \frac{X_i}{2n-i+1}\right).$$

Therefore,

$$\begin{aligned} E(X_{(2n)}|X_{(n)} = x_n) = E\Bigg(&\sum_{i=1}^{n} \frac{X_i}{2n-i+1} \\ &+ \sum_{i=n+1}^{2n} \frac{X_i}{2n-i+1} \Bigg| \sum_{i=1}^{n} \frac{X_i}{2n-i+1} = x_n\Bigg) \end{aligned}$$

$$= x_n + E\left(\sum_{i=n+1}^{2n} \frac{X_i}{2n-i+1} | \sum_{i=1}^{n} \frac{X_i}{2n-i+1} = x_n\right)$$

$$= x_n + E\left(\sum_{i=n+1}^{2n} \frac{X_i}{2n-i+1}\right)$$

because the X_i are all independent

$$= x_n + \sum_{i=n+1}^{2n} \frac{1}{2n-i+1}.$$

For example, in a sample of size 4, $E(X_{(4)}|X_{(1)}{=}x, X_{(2)}{=}y){=}E(X_{(4)}|X_{(2)}{=}y) = y + \sum_{i=3}^{4} \frac{1}{5-i} = y + \frac{3}{2}$.

6.6 Some Applications

Order statistics and the related theory have many interesting and important applications in statistics, in modeling of empirical phenomena, for example, climate characteristics, and in probability theory itself. We touch on a small number of applications in this section for purposes of reference. For further reading on the vast literature on applications of order statistics, we recommend, among numerous possibilities, Lehmann (1975), Shorack and Wellner (1986), David (1980), Reiss (1989), Martynov (1992), Galambos (1987), Falk et al. (1994), Coles (2001), Embrechts et al. (2008), and DasGupta (2008).

6.6.1 ∗ Records

Record values and their timings are used for the purposes of tracking changes in some process, such as temperature, and for preparation for extremal events, such as protection against floods. They are also interesting on their own right.

Let $X_1, X_2, \ldots,$ be an infinite sequence of independent observations from a continuous CDF F. We first give some essential definitions.

Definition 6.7. We say that a *record* occurs at time i if $X_i > X_j \ \forall j < i$. By convention, we say that X_1 is a record value, and $i = 1$ is a record time.

Let Z_i be the indicator of the event that a record occurs at time i. The sequence $T_1, T_2, \ldots$ defined as $T_1 = 1; T_j = \min\{i > T_{j-1} : Z_i = 1\}$ is called the sequence of *record times*. The differences $D_{i+1} = T_{i+1} - T_i$ are called the *interarrival times*.

The sequence $X_{T_1}, X_{T_2}, \ldots,$ is called the sequence of *record values*.

Example 6.13. The values 1.46, .28, 2.20, .72, 2.33, .67, .42, .85, .66, .67, 1.54, .76, 1.22, 1.72, .33 are 15 simulated values from a standard exponential distribution. The record values are 1.46, 2.20, 2.33, and the record times are $T_1 = 1, T_2 = 3$, $T_3 = 5$. Thus, there are three records at time $n = 15$. We notice that no records were observed after the fifth observation in the sequence. In fact, in general, it becomes increasingly more difficult to obtain a record as time passes; justification for this statement is shown in the following theorem.

The following theorem summarizes a number of key results about record values, times, and number of records; this theorem is a superb example of the power of the quantile transformation method, because the results for a general continuous CDF F can be obtained from the $U[0,1]$ case by making a quantile transformation. The details are worked out in Port (1993, pp. 502–509).

Theorem 6.12. *Let $X_1, X_2, \ldots$ be an infinite sequence of independent observations from a CDF F, and assume that F has the density f. Then,*

(a) The sequence $Z_1, Z_2, \ldots$ is an infinite sequence of independent Bernoulli variables, with $E(Z_i) = P(Z_i = 1) = \frac{1}{i}, i \geq 1$.
(b) Let $N = N_n = \sum_{i=1}^{n} Z_i$ be the number of records at time n. Then,

$$E(N) = \sum_{i=1}^{n} \frac{1}{i}; \quad Var(N) = \sum_{i=1}^{n} \frac{i-1}{i^2}.$$

(c) Fix $r \geq 2$. Then D_r has the pmf

$$P(D_r = k) = \sum_{i=0}^{k-1} (-1)^i \binom{k-1}{i} (i+2)^{-r}, k \geq 1.$$

(d) The rth record value X_{T_r} has the density

$$f_r(x) = \frac{[-\log(1-F(x))]^{r-1}}{(r-1)!} f(x), \quad -\infty < x < \infty.$$

(e) The first n record values, $X_{T_1}, X_{T_2}, \ldots, X_{T_n}$ have the joint density

$$f_{12\cdots n}(x_1, x_2, \ldots, x_n) = f(x_n) \prod_{r=1}^{n-1} \frac{f(x_i)}{1-F(x_i)} I_{\{x_1 < x_2 < \cdots < x_n\}}.$$

(f) Fix a sequence of reals $t_1 < t_2 < t_3 < \cdots < t_k$, and let for any given real t, $M(t)$ be the total number of record values that are $\leq t$:

$$M(t) = \#\{i : X_i \leq t \text{ and } X_i \text{ is a record value}\}.$$

Then, $M(t_i) - M(t_{i-1}), 2 \le i \le k$ *are independently distributed, and*

$$M(t_i) - M(t_{i-1}) \sim \text{Poi}\left(\log \frac{1 - F(t_{i-1})}{1 - F(t_i)}\right).$$

Remark. From part (a) of the theorem, we learn that if indeed the sequence of observations keeps coming from the same CDF, then obtaining a record becomes harder as time passes; $P(Z_i = 1) \to 0$. We learn from part (b) that both the mean and the variance of the number of records observed until time n are of the order of $\log n$. The number of records observed until time n is well approximated by a Poisson distribution with mean $\log n$, or a normal distribution with mean and variance equal to $\log n$. We learn from parts (c) and (d) that the interarrival times of the record values do not depend on F, but the *magnitudes* of the record values do. Part (f) is another example of the Poisson distribution providing an approximation in an interesting problem. It is interesting to note the connection between part (b) and part (f). In part (f), if we take $t = F^{-1}(1 - \frac{1}{n})$, then *heuristically*, N_n, the number of records observed up to time n, satisfies $N_n \approx M(X_{(n)}) \approx M(F^{-1}(1 - \frac{1}{n})) \approx \text{Poi}(-\log(1 - F(F^{-1}(1 - \frac{1}{n})))) = \text{Poi}(\log n)$, which is what we mentioned in the paragraph above.

Example 6.14 (Density of Record Values and Times). It is instructive to look at the effect of the tail of the underlying CDF F on the magnitude of the record values. Figure 6.3 gives the density of the third record value for three choices of $F, F = N(0,1), \text{DoubleExp}(0,1), C(0,1)$. Although the modal values are not very different, the effect of the tail of F on the tail of the record density is clear. In particular, for the standard Cauchy case, record values do not have a finite expectation.

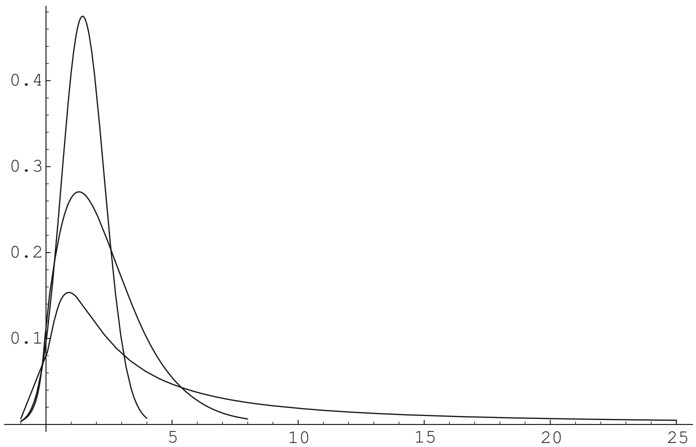

Fig. 6.3 Density of the third record value for, top to bottom, N(0, 1), double exp (0,1), C(0,1) case

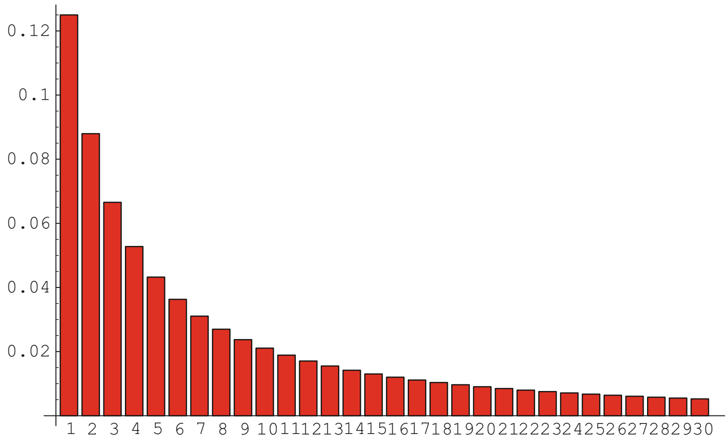

Fig. 6.4 PMF of interarrival time between second and third record

Next, consider the distribution of the gap between the arrival times of the second and the third record. Note the long right tail, akin to an exponential density in Fig. 6.4.

6.6.2 The Empirical CDF

The empirical CDF $F_n(x)$, defined in Section 6.1.2, is a tool of tremendous importance in statistics and probability. The reason for its effectiveness as a tool is that if sample observations arise from some CDF F, then the empirical CDF F_n will be very close to F for large n. So, we can get a very good idea of what the true F is by looking at F_n. Furthermore, because $F_n \approx F$, it can be expected that if $T(F)$ is a nice functional of F, then the empirical version $T(F_n)$ would be close to $T(F)$. Perhaps the simplest example of this is the mean $T(F) = E_F(X)$. The empirical version then is $T(F_n) = E_{F_n}(X) = \frac{\sum_{i=1}^n X_i}{n}$, because F_n assigns the equal probability $\frac{1}{n}$ to just the observation values $X_1, \dots, X_n$. This means that the mean of the sample values should be close to the expected value under the true F. And, this is indeed true under simple conditions, and we have already seen some evidence for it in the form of the central limit theorem. We provide some basic properties and applications of the empirical CDF in this section.

Theorem 6.13. *Let $X_1, X_2, \dots$ be independent observations from a CDF F. Then,*

(a) *For any fixed x, $nF_n(x) \sim \text{Bin}(n, F(x))$.*
(b) **(DKW Inequality).** *Let $\Delta_n = \sup_{x \in \mathcal{R}} |F_n(x) - F(x)|$. Then, for all $n, \epsilon > 0$, and all F,*

$$P(\Delta_n > \epsilon) \le 2e^{-2n\epsilon^2}.$$

(c) *Assume that F is continuous. For any given n, and $\alpha, 0 < \alpha < 1$, there exist positive constants D_n, independent of F, such that whatever be F,*

$$P(\forall x \in \mathcal{R}, F_n(x) - D_n \le F(x) \le F_n(x) + D_n) \ge 1 - \alpha.$$

Remark. Part (b), the DKW inequality, was first proved in Dvoretzky et al. (1956), but in a weaker form. The inequality stated here is proved in Massart (1990). Furthermore, the constant 2 in the inequality is the best possible choice of the constant; that is, the inequality is false with any other constant $C < 2$. The inequality says that *uniformly in x*, for large n, the empirical CDF is arbitrarily close to the true CDF with a very high probability, and the probability of the contrary is *sub-Gaussian*. We show more precise consequences of this in a later chapter. Part (c) is important for statisticians, as we show in our next example.

Example 6.15 (Confidence Band for a Continuous CDF). This example is another important application of the quantile transformation method. Imagine a hypothetical sequence of independent $U[0,1]$ variables, $U_1, U_2, \ldots$, and let G_n denote the empirical CDF of this sequence of uniform random variables; that is,

$$G_n(t) = \frac{\#\{i : U_i \le t\}}{n}.$$

By the quantile transformation,

$$\begin{aligned}\Delta_n &= \sup_{x \in \mathcal{R}} |F_n(x) - F(x)| \stackrel{\mathcal{L}}{=} \sup_{x \in \mathcal{R}} |G_n(F(x)) - F(x)| \\ &= \sup_{0<t<1} |G_n(t) - t|,\end{aligned}$$

which shows that as long as F is a continuous CDF, so that the quantile transformation can be applied, for any n, the distribution of Δ_n is the same for all F. This common distribution is just the distribution of $\sup_{0<t<1}|G_n(t) - t|$. Consequently, if D_n is such that $P(\sup_{0<t<1}|G_n(t) - t| > D_n) \le \alpha$, then D_n also satisfies (the apparently stronger statement)

$$P(\forall x \in \mathcal{R}, F_n(x) - D_n \le F(x) \le F_n(x) + D_n) \ge 1 - \alpha.$$

The probability statement above provides the assurance that with probability $1 - \alpha$ or more, the true CDF $F(x)$, as a function, is caught between the pair of functions $F_n(x) \pm D_n$. As a consequence, the band $F_n(x) - D_n \le F(x) \le F_n(x) + D_n, x \in \mathcal{R}$, is called a $100(1-\alpha)\%$ *confidence band* for F. This is of great use in statistics, because statisticians often consider the true CDF F to be not known.

The constants D_n have been computed and tabulated for small and moderate n. We tabulate the values of D_n for some selected n for easy reference and use.

n	95th Percentile	99th Percentile
20	.294	.352
21	.287	.344
22	.281	.337
23	.275	.330
24	.269	.323
25	.264	.317
26	.259	.311
27	.254	.305
28	.250	.300
29	.246	.295
30	.242	.290
35	.224	.269
40	.210	.252
>40	$\frac{1.36}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

6.7 * Distribution of the Multinomial Maximum

The maximum cell frequency in a multinomial distribution is of current interest in several areas of probability and statistics. It is of wide interest in cluster detection, data mining, goodness of fit, and in occupancy problems in probability. It also arises in sequential clinical trials. It turns out that the technique of *Poissonization* can be used to establish, in principle, the exact distribution of the multinomial maximum cell frequency. This can be of substantial practical use. Precisely, if $N \sim \text{Poisson}(\lambda)$, and given $N = n, (f_1, f_2, \ldots, f_k)$ has a multinomial distribution with parameters $(n, p_1, p_2, \ldots, p_k)$, then unconditionally, $f_1, f_2, \ldots, f_k$ are independent and $f_i \sim \text{Poisson}(\lambda p_i)$. It follows that with any given fixed value n, and any given fixed set A in the k-dimensional Euclidean space $\mathcal{R}^k$, the multinomial probability that $(f_1, f_2, \ldots, f_k)$ belongs to A equals $n!c(n)$, with $c(n)$ being the coefficient of λ^n in the power series expansion of $e^{\lambda} P((X_1, X_2, \ldots, X_k) \in A)$, where now X_i are independent Poisson(λp_i). In the *equiprobable case* (i.e., when the p_i are all equal to $\frac{1}{k}$), this leads to the equality that

$$P(\max\{f_1, f_2, \ldots, f_k\} \geq x) = \frac{n!}{k^n} \times \text{the coefficient of } \lambda^n \text{ in } \left(\sum_{j=0}^{x-1} \frac{\lambda^j}{j!}\right)^k;$$

see Chapter 2.

As a result, we can compute $P(\max\{f_1, f_2, \ldots, f_k\} \geq x)$ exactly whenever we can compute the coefficient of λ^n in the expansion of $(\sum_{j=0}^{x-1} \frac{\lambda^j}{j!})^K$. This is possible to do by using symbolic software; see Ethier (1982) and DasGupta (2009).

Example 6.16 (Maximum Frequency in Die Throws). Suppose a fair six-sided die is rolled 30 times. Should we be surprised if one of the six faces appears 10 times? The usual probability calculation to quantify the surprise is to calculate $P(\max\{f_1, f_2, \ldots, f_6\} \geq 10)$, namely the P-value, where $f_1, f_2, \ldots, f_6$ are the frequencies of the six faces in the 30 rolls. Because of our Poissonization result, we can compute this probability. From the table of exact probabilities below, we can see that it would not be very surprising if some face appeared 10 times in 30 rolls of a fair die; after all $P(\max\{f_1, f_2, \ldots, f_6\} \geq 10) = .1176$, not a very small number, for 30 rolls of a fair die. Similarly, it would not be very surprising if some face appeared 15 times in 50 rolls of a fair die, as can be seen in the table below.

$P(\max\{f_1, f_2, \ldots, f_k\} \geq x)(k = 6)$		
x	$n = 30$	$n = 50$
8	.6014	1
9	.2942	1
10	.1176	.9888
11	.0404	.8663
12	.0122	.6122
13	.0032	.3578
14	.00076	.1816
15	.00016	.0827
16	.00003	.0344

Exercises

Exercise 6.1. Suppose X, Y, Z are three independent $U[0, 1]$ variables. Let U, V, W denote the minimum, median, and the maximum of X, Y, Z.

(a) Find the densities of U, V, W.
(b) Find the densities of $\frac{U}{V}$ and $\frac{V}{W}$, and their joint density.
(c) Find $E(\frac{U}{V})$ and $E(\frac{V}{W})$,

Exercise 6.2. Suppose $X_1, \ldots, X_5$ are independent $U[0, 1]$ variables. Find the joint density of $X_{(2)}, X_{(3)}, X_{(4)}$, and $E(X_{(4)} + X_{(2)} - 2X_{(3)})$.

Exercise 6.3. * Suppose $X_1, \ldots, X_n$ are independent $U[0, 1]$ variables.

(a) Find the probability that all n observations fall within some interval of length at most .9.
(b) Find the smallest n such that $P(X_{(n)} \geq .99, X_{(1)} \leq .01) \geq .99$.

Exercise 6.4 (Correlation Between Order Statistics). Suppose $X_1, \ldots, X_5$ are independent $U[0, 1]$ variables. Find the exact values of $\rho_{X_{(i)}, X_{(j)}}$ for all $1 \leq i < j \leq 5$.

Exercise 6.5 * (Correlation Between Order Statistics). Suppose $X_1, \ldots, X_n$ are independent $U[0,1]$ variables. Find the smallest n such that $\rho_{X_{(1)}, X_{(n)}} < \epsilon, \epsilon = .5, .25, .1$.

Exercise 6.6. Suppose X, Y, Z are three independent standard exponential variables, and let U, V, W be their minimum, median, and maximum. Find the densities of $U, V, W, W - U$.

Exercise 6.7 (Comparison of Mean, Median, and Midrange). Suppose $X_1, X_2, \ldots, X_{2m+1}$ are independent observations from $U[\mu - \sigma, \mu + \sigma]$.

(a) Show that the expectation of each of $\bar{X}, X_{(m+1)}, \frac{X_{(1)}+X_{(n)}}{2}$ is μ.
(b) Find the variance of each $\bar{X}, X_{(m+1)}, \frac{X_{(1)}+X_{(n)}}{2}$. Is there an ordering among their variances?

Exercise 6.8 * (Waiting Time). Peter, Paul, and Mary went to a bank to do some business. Two counters were open, and Peter and Paul went first. Each of Peter, Paul, and Mary will take, independently, an Exp(λ) amount of time to finish their business, from the moment they arrive at the counter.

(a) What is the density of the epoch of the last departure?
(b) What is the probability that Mary will be the last to finish?
(c) What is the density of the total time taken by Mary from arrival to finishing her business?

Exercise 6.9. Let $X_1, \ldots, X_n$ be independent standard exponential variables.

(a) Derive an expression for the CDF of the maximum of the spacings, $W_0 = X_{(1)}, W_i = X_{(i+1)} - X_{(i)}, i = 1, \ldots, n-1$.
(b) Use it to calculate the probability that among 20 independent standard Exponential observations, no two consecutive observations are more than .25 apart.

Exercise 6.10 * (A Characterization). Let X_1, X_2 be independent observations from a continuous CDF F. Suppose that $X_{(1)}$ and $X_{(2)} - X_{(1)}$ are independent. Show that F must be the CDF of an exponential distribution.

Exercise 6.11 * (Range and Midrange). Let $X_1, \ldots, X_n$ be independent $U[0,1]$ variables. Let $W_n = (X_{(n)} - X_{(1)}), Y_n = \frac{X_{(n)}+X_{(1)}}{2}$. Find the joint density of W_n, Y_n (be careful about where the joint density is positive). Use it to find the conditional expectation of Y_n given $W_n = w$.

Exercise 6.12 * (Density of Midrange). Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F with density f. Show that the density of $Y_n = \frac{X_{(n)}+X_{(1)}}{2}$ is given by

$$f_Y(y) = n \int_{-\infty}^{y} [F(2y - x) - F(x)]^{n-1} f(x) dx.$$

Exercise 6.13 * (Mean Given the Minimum and Maximum). Let $X_1, \ldots, X_n$ be independent $U[0, 1]$ variables. Derive a formula for $E(\bar{X} \mid X_{(1)}, X_{(n)})$.

Exercise 6.14 * (Mean Given the Minimum and Maximum). Let $X_1, \ldots, X_n$ be independent standard exponential variables. Derive a formula for $E(\bar{X} \mid X_{(1)}, X_{(n)})$.

Exercise 6.15 * (Distance Between Mean and Maximum). Let $X_1, \ldots, X_n$ be independent $U[0, 1]$ variables. Derive as clean a formula as possible for $E(|\bar{X} - X_{(n)}|)$.

Exercise 6.16 * (Distance Between Mean and Maximum). Let $X_1, \ldots, X_n$ be independent standard exponential variables. Derive as clean a formula as possible for $E(|\bar{X} - X_{(n)}|)$.

Exercise 6.17 * (Distance Between Mean and Maximum). Let $X_1, \ldots, X_n$ be independent standard normal variables. Derive as clean a formula as possible for $E(|\bar{X} - X_{(n)}|)$.

Exercise 6.18 * (Relation Between Uniform and Standard Normal). Let $Z_1, Z_2, \ldots$ be independent standard normal variables. Let $U_1, U_2, \ldots$ be independent $U[0, 1]$ variables. Prove the distributional equivalence:

$$(U_{(r)})|_{r=1}^{n} \stackrel{\mathcal{L}}{=} \left(\frac{\sum_{i=1}^{2r} Z_i^2}{\sum_{i=1}^{2(n+1)} Z_i^2} \right) |_{r=1}^{n}.$$

Exercise 6.19 * (Confidence Interval for a Quantile). Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F. Fix $0 < p < 1, 0 < \alpha < 1$, and let $F^{-1}(p)$ be the pth quantile of F. Show that for large enough n, there exist $1 \leq r < s \leq n$ such that $P(X_{(r)} \leq F^{-1}(p) \leq X_{(s)}) \geq 1 - \alpha$.

Do such r, s exist for all n?

Hint: Use the quantile transformation.

Exercise 6.20. Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F. Find the smallest value of n such that $P(X_{(2)} \leq F^{-1}(\frac{1}{2}) \leq X_{(n-1)}) \geq .95$.

Exercise 6.21. Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F with a density symmetric about some μ. Show that for all odd sample sizes $n = 2m + 1$, the median $X_{(m+1)}$ has a density symmetric about μ.

Exercise 6.22. Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F with a density symmetric about some μ. Show that for any r, $X_{(n-r+1)} - \mu \stackrel{\mathcal{L}}{=} \mu - X_{(r)}$.

Exercise 6.23 * (Unimodality of Order Statistics). Let $X_1, \ldots, X_n$ be independent observations from a continuous CDF F with a density f. Suppose $\frac{1}{f(x)}$ is convex on the support of f, namely, on $\mathcal{S} = \{x : f(x) > 0\}$. Show that for any r, the density of $X_{(r)}$ is unimodal. You may assume that $\mathcal{S}$ is an interval.

Exercise 6.24. Let $X_1, X_2, \ldots, X_n$ be independent standard normal variables. Prove that the mode of $X_{(n)}$ is the unique root of

$$(n-1)\phi(x) = x\Phi(x).$$

Exercise 6.25 (Conditional Expectation Given the Order Statistics). Let $g(x_1, x_2, \ldots, x_n)$ be a general real-valued function of n variables. Let $X_1, X_2, \ldots, X_n$ be independent observations from a common CDF F. Find as clean an expression as possible for $E(g(X_1, X_2, \ldots, X_n) \,|X_{(1)}, X_{(2)}, \ldots, X_{(n)})$.

Exercise 6.26. Derive a formula for the expected value of the rth record when the sample observations are from an exponential density.

Exercise 6.27 * (Record Values in Normal Case). Suppose $X_1, X_2, \ldots$ are independent observations from the standard normal distribution. Compute the expected values of the first ten records.

Exercise 6.28. Let $F_n(x)$ be the empirical CDF of n observations from a CDF F. Show that

$$\begin{aligned}\Delta_n &= \sup_{x\in\mathcal{R}}|F_n(x) - F(x)| \\ &= \max_{1\le i\le n} \max\left\{\frac{i}{n} - F(X_{(i)}), F(X_{(i)}) - \frac{i-1}{n}\right\}.\end{aligned}$$

References

Bickel, P. (1967). Some contributions to order statistics, *Proc. Fifth Berkeley Symp.*, I, 575–591, L. Le Cam and J. Neyman Eds., University of California Press, Berkeley.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer, New York.

DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability*, Springer, New York.

DasGupta, A. (2009). Exact tail probabilities and percentiles of the multinomial maximum, Technical Report, Purdue University.

David, H. (1980). *Order Statistics*, Wiley, New York.

Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1956). Asymptotic minimax character of the sample distribution function, *Ann. Math. Statist.*, 27, 3, 642–669.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (2008). *Modelling Extremal Events: For Insurance and Finance*, Springer, New York.

Ethier, S. (1982). Testing for favorable numbers on a roulette wheel, *J. Amer. Statist. Assoc.*, 77, 660–665.

Falk, M., Hüsler, J. and Reiss, R. (1994). *Laws of Small Numbers, Extremes, and Rare Events*, Birkhauser, Basel.

Galambos, J. (1987). *Asymptotic Theory of Extreme Order Statistics*, Academic Press, New York.

Hall, P. (1978). Some asymptotic expansions of moments of order statistics, *Stoch. Proc. Appl.*, 7, 265–275.

Hall, P. (1979). On the rate of convergence of normal extremes, *J. Appl. Prob.*, 16, 2, 433–439.

Leadbetter, M., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Lehmann, E. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, McGraw-Hill, Columbus, OH.

Martynov, G. (1992). Statistical tests based on empirical processes and related problems, *Soviet J. Math*, 61, 4, 2195–2271.

Massart, P. (1990). The tight constant in the DKW inequality, *Ann. Prob.*, 18, 1269–1283.

Port, S. (1993). *Theoretical Probability for Applications*, Wiley, New York.

Reiss, R. (1989). *Approximation Theorems of Order Statistics*, Springer-Verlag, New York.

Resnick, S. (2007). *Extreme Values, Regular Variation, and Point Processes*, Springer, New York.

Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*, Wiley, New York.