Chapter 2
Order Statistics

2.1 Order Statistics and Their Distributions

Let $X_1, X_2, \ldots, X_n, \ldots$ be some random variables and let $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ denote the corresponding variational series based on random variables $X_1, X_2, \ldots, X_n$. Elements $X_{k,n}$, $1 \le k \le n$, are called *order statistics*. Observed values of $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$. we denote $x_{1,n}, x_{2,n}, \ldots, x_{n,n}$ and call a realization of order statistics. Let us note that

$$X_{1,n} = m(n) = \min\{X_1, X_2, \ldots, X_n\}$$

and

$$X_{n,n} = M(n) = \max\{X_1, X_2, \ldots, X_n\},\ n = 1, 2, \ldots.$$

In the classical theory of order statistics (see, for example, David and Nagaraja (2003), Ahsanullah et al. (2013)) the initial X's present independent and identically distributed random variables. Below we in the most part of situations also will deal with such types of random variables.

Hence consider now independent random variables $X_1, X_2, \ldots, X_n$ with some common distribution function (d.f.) F. It enables us to consider the set $\{X_1, X_2, \ldots, X_n\}$ as a sample of size n taken from the population distribution F. The set of the observed values $\{x_1, x_2, \ldots, x_n\}$ of random variables $X_1, X_2, \ldots, X_n$ is called a realization of the sample. In this case elements $X_{k,n}$, $1 \le k \le n$, that is order statistics based on sample

$X_{1,n} \le X_{2,n} \le \ldots \prec X_{n,n}$ can be identified as order statistics from this d.f. F.

It follows from the definition of order statistics that

$$P\{X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}\} = 1. \tag{2.1.1}$$

© Atlantis Press and the author(s) 2015

M. Ahsanullah and V.B. Nevzorov, *Records via Probability Theory*,
Atlantis Studies in Probability and Statistics 6,
DOI 10.2991/978-94-6239-136-9_2

Moreover it is easy to show that if F is a continuous d.f. then equality (2.1.1) can be sharpened and written as

$$P\{X_{1,n} < X_{2,n} < \cdots < X_{n,n}\} = 1. \tag{2.1.2}$$

Indeed, if X's have a discrete d.f. then (2.1.2) is not true. Consider the corresponding example in Exercises 2.1.1 and 2.1.2.

Exercise 2.1.1 Let $X_1, X_2, \ldots, X_n$ be independent random variables taking two values 0 and 1 with probabilities ½. Find

$$p_n = P\{X_{1,n} = X_{2,n} = \cdots = X_{n,n}\}, n = 1, 2, \ldots,$$

to convince yourself that $p_n \neq 0$.

Exercise 2.1.2 Let X_1, X_2, X_3 be independent random variables, having the geometric distribution, that is

$$P\{X = m\} = (1-p)p^m, m = 0, 1, 2, \ldots.$$

Find

$$P\{X_{1,3} < X_{2,3} < X_{3,3}\}.$$

As one sees there are two cases (F is continuous and F is discrete) which need different approach for writing distributions of order statistics.

Let us suppose below that X's are independent random variables having a common continuous d.f. F.

There are rather simple formulae for order statistics $X_{1,n}$ and $X_{n,n}$. Really,

$$\begin{aligned} F_{n,n}(x) = P\{X_{n,n} \leq x\} &= P\{X_1 \leq x, \ldots, X_n \leq x\} \\ &= P\{X_1 \leq x\} \ldots P\{X_n \leq x\} = F^n(x), \end{aligned} \tag{2.1.3}$$

where

$$F(x) = P\{X_k \leq x\}, k = 1, 2, \ldots, n.$$

Analogously we obtain that

$$\begin{aligned} F_{1,n}(x) = P\{X_{1,n} \leq x\} &= 1 - P\{X_{1,n} > x\} \\ &= 1 - P\{X_{1n} > x, \ldots, X_n > x\} = 1 - (1 - F(x))^n. \end{aligned} \tag{2.1.4}$$

Let us note that (2.1.3) and (2.1.4) are valid for any d.f. F (continuous and discrete).

Exercise 2.1.3 Find the joint d.f.

$$F_{1,n,n}(x,y) = P\{X_{1,n} \le x, X_{n,n} \le y\}$$

of order statistics $X_{1,n}$ and $X_{n,n}$.

Exercise 2.1.4 Consider the case *when* $X_1, X_2, \ldots, X_n$ are independent random variables with probably different d.f.'s $F_1, F_2, \ldots, F_n$ and find in this case the d.f.'s of $m(n) = \min\{X_1, X_2, \ldots, X_n\}$ and $M(n) = \max\{X_1, X_2, \ldots, X_n\}$.

If our initial X's have probability density function (pdf) $f(x)$ then it is the following method to write the pdf $f_{k,n}(x)$ and d.f. $F_{k,n}(x)$ of $X_{k,n}$, $1 \le k \le n$. To find pdf we must consider the situation, when one of our n X's takes value x and exactly $k-1$ of the rest X's are less than x. Thus we get easily that

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!}(F(x))^{k-1}(1-F(x))^{n-k}f(x). \tag{2.1.5}$$

Then

$$\begin{aligned} F_{k,n}(x) &= \int_0^x f_{k,n}(y)dy = \frac{n!}{(k-1)!(n-k)!}\int_0^x (F(y))^{k-1}(1-F(y))^{n-k}f(y)dy \\ &= I_{F(x)}(k, n-k+1), \end{aligned} \tag{2.1.6}$$

where

$$I_x(a,b) = \frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}dt$$

denotes the incomplete beta function.

Note that equality

$$F_{k,n}(x) = I_{F(x)}(k, n-k+1), \tag{2.1.7}$$

which is given here, was proved only for X's having some pdf $f(x)$.

What shall we do in the general situation? How to prove that (2.1.6) is valid for any d.f. F? It appears that it is enough to take into account the following two facts.

(a) If we consider the uniform $U(0,1)$ sample $U_1, U_2, \ldots, U_n$ with

$$\mathrm{P}\{U_k \le x\} = x,\ 0 < x < 1,\ k = 1, 2, \ldots, n,$$

and the corresponding order statistics $U_{1,n} \le U_{2,n} \le \cdots \le U_{n,n}$, then it follows from (2.1.7) that

$$\mathrm{P}\{U_{k,n} \le x\} = I_x(k, n-k+1), 0<x<1. \tag{2.1.8}$$

(b) Now if we compare the uniform $U_{k,n}$ order statistics and the order statistics $X_{k,n}$ from any d.f. $F(x)$, then it is known that $X_{k,n}$ can be presented as follows:

$$G(U_{k,n}) = X_{k,n}(\text{in distribution}), \tag{2.1.9}$$

where G is the inverse of F.

Hence it follows immediately from (a) and (b) that relation (2.1.7) is valid for any d.f. F.

When we want to consider joint distributions of some order statistics we must work with complicate expressions even in the case of small values of the sample size n. There are rather simple formulae in the situations when the initial random variables have some density function $f(x)$. In this case the joint density function of all order statistics $X_{1,n}$, $X_{2,n}$,..., $X_{n,n}$ has the following form:

$$f_{1,2,\ldots,n:n}(x_1, x_2, \cdots, x_n) = n! \prod_{k=1}^{n} f(x_k),\ -\infty < x_1 < x_2 < \ldots < x_n < \infty \tag{2.1.10}$$

and

$$f_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = 0, \text{otherwise}.$$

It can be found from (2.1.10) that the joint pdf

$$f_{k(1),k(2),\ldots,k(r):n}(x_1, x_2, \ldots, x_r)$$

of order statistics

$$X_{k(1),n}, X_{k(2),n}, \ldots, X_{k(r),n},$$

where

$$1 \le k(1) < k(2) < \cdots < k(r) \le n,$$

is of the form

$$f_{k(1),k(2),\ldots,k(r):n}(x_1, x_2, \ldots, x_r) = \frac{n!}{\prod_{m=1}^{r+1} (k(m) - k(m-1) - 1)!} \prod_{m=1}^{r+1} (F(x_m) - F(x_{m-1}))^{k(m)-k(m-1)-1} \prod_{m=1}^{r} f(x_m), \tag{2.1.11}$$

if $x_1 < x_2 < \cdots < x_r$, and $f_{k(1),k(2),\ldots,\,k(r):n}\,(x_1,\, x_2,\ldots,\, x_r) = 0$, otherwise.

In particular, if $r = 2$, $1 \le i < j \le n$, and $x_1 < x_2$, then the joint pdf $f_{i,j:n}\,(x_1,\, x_2)$ of two order statistics is given by the following equality:

$$f_{i,j:n}(x_1, x_2) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (F(x_1))^{i-1}(F(x_2) - F(x_1))^{j-i-1}(1 - F(x_2))^{n-j} f(x_1) f(x_2). \tag{2.1.12}$$

To understand the level of complication of the formulae for the joint distributions in the general case we suggest our reader the corresponding expression for the joint d.f. of two order statistics $X_{r,n}$ and $X_{s,n}$, $1 \le r < s \le n$ (below $x_1 < x_2$):

$$F_{r,s:n}(x_1, x_2) = \sum_{i=r}^{n} \sum_{j=\max\{0,\mathrm{s-i}\}}^{n-i} \frac{n!}{i!j!(n-i-j)!} (F(x_1))^i (F(x_2) - F(x_1))^j (1 - F(x_2))^{n-i-j}. \tag{2.1.13}$$

There are some situations when it is possible to get rather simple expressions for joint distributions of order statistics. For example, if one considers probabilities

$$P\{y_1 < X_{1,n} \le x_1,\, y_2 < X_{2,n} \le x_2, \cdots,\, y_n < X_{n,n} \le x_n\}$$

for any values

$$-\infty < y_1 < x_1 \le y_2 < x_2 \le \cdots \le y_n < x_n < \infty,$$

then it can be obtained that the corresponding equality has the form

$$P\{y_1 < X_{1,n} \le x_1,\, y_2 < X_{2,n} \le x_2, \ldots,\, y_n < X_{n,n} \le x_n\} = n! \prod_{k=1}^{n} (F(x_k) - F(y_k)). \tag{2.1.14}$$

Exercise 2.1.5 Let $X_1,\, X_2,\ldots,\, X_n$ be a sample of size n from a geometrically distributed random variable X, such that

$$P\{X = m\} = (1-p)p^m,\ m = 0, 1, 2, \ldots.$$

Find

$$P\{\, X_{1,n} \ge r,\, X_{n,n} < s\},\ r < s.$$

2.2 The Classical Representations for Order Statistics

It was mentioned above that ordered random variables $X_{1,n} \le X_{2,n} \le \cdots \le X_{n,n}$ are dependent. Indeed there exist essentially more methods in the probability theory to work with independent variables than with dependent ones. Hence different possibilities to express dependent random variables in terms of independent values are convenient for the future investigation. There are corresponding expressions for some types of order statistics, which will be presented below. We give these results without their proof, which can be found, for example in Ahsanullah et al. (2013).

In the sequel we will use the special notation $U_{1,n} \le \cdots \le U_{n,n}$, $n = 1, 2, \ldots$, for the standard uniform order statistics (the case when X's have d.f. $F(x) = x$, $0 \le x \le 1$) and the notation $Z_{1,n} \le \cdots \le Z_{n,n}$, $n = 1, 2, \ldots$, for the standard exponential order statistics (here $F(x) = 1 - \exp(-x)$, $x \ge 0$).

Representation 1 *The following expression is valid for the uniform order statistics under any* $n = 1, 2, \ldots$:

$$\left(U_{1,n}, \ldots, U_{n,n}\right) \stackrel{d}{=} \left(\frac{S_1}{S_{n+1}} \ldots, \frac{S_n}{S_{n+1}}\right), \tag{2.2.1}$$

where

$$S_n = v_1 + v_2 + \cdots + v_n,\ n = 1, 2, \ldots,$$

and $v_1, v_2, \ldots$ *be independent random variables having the standard* $E(1)$ *exponential distribution.*

Remark 2.2.1 Moreover, sometimes it can come in useful that for any $n = 1, 2, \ldots$ the vector

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

and the sum S_{n+1} are independent.

Remark 2.2.2 Representation 1 can be also rewritten in the following useful form:

$$\left(U_{1,n},\ U_{2,n} - U_{1,n}, \ldots, U_{n,n} - U_{n-1,n}\right) \stackrel{d}{=} \left(\frac{v_1}{v_1 + \cdots + v_{n+1}}, \ldots, \frac{v_n}{v_1 + \cdots + v_{n+1}}\right) \tag{2.2.2}$$

Exercise 2.2.1 Show that the range

$$T(1,n) = M(n) - m(n) = U_{n,n} - U_{1,n},$$

has the same distribution as the order statistic $U_{n-1,n}$.

Representation 2 *Using the same terms as above we can write that the equality*

$$\left(U_{1,n},\ldots,U_{n,n}\right) \stackrel{d}{=} (S_1,\ldots,S_n|\, S_{n+1}=1), \tag{2.2.3}$$

also holds, i.e., the distribution of the vector of uniform order statistics coincides with the conditional distribution of the vector of sums $S_1,\ldots, S_n$ given that $S_{n+1} = 1$.

Representation 3 *Let W_1, $W_2,\ldots$ be independent uniformly distributed on* [0,1] *random variables. Then the next equality is true for any $n = 1, 2,\ldots$:*

$$\left(U_{1,n}, U_{2,n},\ldots,U_{n,n}\right) \stackrel{d}{=} \left(W_1 W_2^{1/2}\ldots W_{n-1}^{1/(n-1)} W_n^{1/n},\ W_2^{1/2}\ldots W_{n-1}^{1/(n-1)} W_n^{1/n},\ldots,\ W_n^{1/n}\right). \tag{2.2.4}$$

Exercise 2.2.2 Show that for any $n = 2, 3,\ldots$, ratios

$$V_k = \left(U_{k,n}/U_{k+1,n}\right)^k,\ k = 1,2,\ldots,n,$$

where $U_{n+1,\,n} = 1$, are independent and have the same uniform distribution on [0,1].

Now we will show that the exponential order statistics $Z_{k,n}$ also can be presented in very convenient form.

Representation 4 *Let again v_1, $v_2,\ldots$, be independent random variables having the standard $E(1)$ exponential distribution. Then the following result holds for the exponential order statistics and any $n = 1, 2,\ldots$:*

$$\left(Z_{1,n},\, Z_{2,n},\ldots,Z_{n,n}\right) \stackrel{d}{=} \left(\frac{v_1}{n},\ \frac{v_1}{n}+\frac{v_2}{n-1},\ldots,\frac{v_1}{n}+\frac{v_2}{n-1}+\cdots+\frac{v_{n-1}}{2}+v_n\right). \tag{2.2.5}$$

Remark 2.2.3 It follows from (2.2.5) that normalized differences

$$(n-k+1)\left(Z_{k,n}-\ Z_{k-1,n}\right),\ k = 1,2,\ldots,n,$$

where $Z_{0,n} = 0$, are independent and have the same exponential $E(1)$ distribution.

Exercise 2.2.3 For exponential order statistics $Z_{1,n}$, $Z_{2,n},\ldots$, $Z_{n,n}$ show that statistics $Z_{1,n}$ and

$$L = (n-1)Z_{n,n} - Z_{1,n} - Z_{2,n} - \cdots - Z_{n-1,n}$$

are independent.

Our reader can ask if the useful representations exist for these two distributions (standard uniform and standard exponential) only. Indeed, it is not difficult to rewrite the corresponding relations for any representative of these families of

distributions. The matter is that if X's have some standard distribution (uniform or exponential, as in our situations) then any Y's which belong to the same family of distributions can be expressed via X's in the form $Y = aX + b$, where $a > 0$ and $-\infty < b < \infty$, and the corresponding order statistics $Y_{k,n}$ can be presented as

$$Y_{k,n} = aX_{k,n} + b.$$

It means that all these representations can be rewritten easily for any uniform or exponential distributions.

Moreover the given results for $U_{k,n}$ and $Z_{k,n}$ can be rewritten for order statistics from arbitrary distribution.

For any d.f. F we determine the inverse function

$$G(s) = inf\{x : F(x) \geq s\},\ 0 < s < 1. \tag{2.2.6}$$

Exercise 2.2.4 Let $F(x)$ be a continuous d.f. of a random variable X. Show that in this case

$$F(G(x)) = x,\ 0 < x < 1, \tag{2.2.7}$$

and it means that $U = F(X)$ has the uniform distribution on interval [0,1].

Remark 2.2.4 Relation (2.2.7) above is true if F is any continuous d.f. Indeed, this equality fails if F has jump points, since then the values of $F(X)$, unlike U, do not cover all interval [0,1]. But it is not difficult to show that the dual equality

$$G(F(x)) = x,\ 0 < x < 1, \tag{2.2.8}$$

holds for any x, where $F(x)$ strongly increases. Thus, presentation

$$X \stackrel{d}{=} G(U), \tag{2.2.9}$$

holds for any random variable, while the dual equality

$$F(X) \stackrel{d}{=} U \tag{2.2.10}$$

is valid for random variables with continuous distribution functions only.

Taking into account equalities (2.2.9) and (2.2.10) one can obtain the corresponding relations which allow to represent any order statistics $X_{1,n} \leq \cdots \leq X_{n,n}$ via the uniform order statistics $U_{1,n} \leq \cdots \leq U_{n,n}$.

Since d.f. F and its inverse G are monotone they don't disturb the ordering of the initial X's. It enables us to get the following presentations which correspond to equalities (2.2.9) and (2.2.10).

Representation 5 *If F is a continuous d.f. then*

$$\big(F\big(X_{1,n}\big),\ldots,\,F(X_{n,n})\big)\overset{d}{=}\big(U_{1,n},\ldots,U_{n,n}\big). \tag{2.2.11}$$

Representation 6 *Let F be any d.f. and G be its inverse. Then* the following equality

$$\big(X_{1,n},\ldots,X_{n,n}\big)\overset{d}{=}\big(G\big(U_{1,n}\big),\ldots,G\big(U_{n,n}\big)\big), \tag{2.2.12}$$

is true.

Let now $X_{1,n}\le\cdots\le X_{n,n}$ and $Y_{1,n}\le\cdots\le Y_{n,n}$ be order statistics corresponding to an arbitrary d.f. F and a continuous d.f. H. Let also G be as above the inverse of F. Combining relations (2.2.11) for Y's and (2.2.12) for X's, one gets one more important result.

Representation 7 *The following equality ties two sets of order statistics*:

$$\big(X_{1,n},\ldots,X_{n,n}\big)\overset{d}{=}\big(G\big(H\big(Y_{1,n}\big)\big),\ldots,G\big(H\big(Y_{n,n}\big)\big)\big). \tag{2.2.13}$$

Example 2.2.1 For instance, if we compare arbitrary order statistics $X_{1,n},\ldots,X_{n,n}$ and exponential order statistics $Z_{1,n},\ldots,Z_{n,n}$, then

$$H(x)=1-\exp(-x),\ x>0,$$

and (2.2.13) can be rewritten as

$$\big(X_{1,n},\ldots,X_{n,n}\big)\overset{d}{=}\big(G\big(1-\exp\big(-Z_{1,n}\big)\big),\ldots,G\big(1-\exp\big(-Z_{n,n}\big)\big)\big). \tag{2.2.14}$$

Remark 2.2.5 Indeed, the results analogous to (2.2.13) are valid for any monotone increasing function $R(x)$ (no necessity to suppose that R is a distribution function). Namely, if

$$X_k=R(Y_k),\ k=1,2,\ldots,n,$$

then the following relation is true:

$$\big(X_{1,n},\ldots,X_{n,n}\big)\overset{d}{=}\big(R\big(Y_{1,n}\big),\ldots,R\big(Y_{n,n}\big)\big). \tag{2.2.15}$$

If R is a monotone decreasing function, then the transformation $R(Y)$ changes the ordering of the original Y's and we have the following equality:

$$\big(X_{1,n},\ldots,X_{n,n}\big)\overset{d}{=}\big(R\big(Y_{n,n}\big),\ldots,R\big(Y_{1,n}\big)\big). \tag{2.2.16}$$

Thus, equalities (2.2.12) and (2.2.14) give opportunity to express any order statistics via independent terms. The following example gives one of the possible presentations.

Example 2.2.2 One can express distributions of arbitrary order statistics $X_{k,n}$ (related to some d.f. F) as follows. If G is the inverse of F and $v_1, v_2, \ldots$ are independent exponentially $E(1)$ distributed random variables, then

$$X_{k,n} \stackrel{d}{=} G\left(\frac{v_1 + \cdots + v_k}{v_1 + \cdots + v_{n+1}}\right) \stackrel{d}{=} G\left(1 - \exp\left(-\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_k}{n-k+1}\right)\right)\right), \quad k = 1, 2, \ldots, n. \tag{2.2.17}$$

Exercise 2.2.5 Let $X_{1,n}, \ldots, X_{n,n}$ be order statistics corresponding to the distribution with the density

$$f(x) = ax^{a-1}, 0 < x < 1, \ a > 0. \tag{2.2.18}$$

Express the product $X_{r,n}$, $X_{s,n}$, $1 \le r < s \le n$, in terms of independent uniformly distributed random variables.

2.3 Moment Characteristics of Order Statistics

Let us recall equality (2.1.7), which gives the expression for distribution functions $F_{k,n}(x)$ of order statistics $X_{k,n}$:

$$F_{k,n}(x) = I_{F(x)}(k, n-k+1),$$

where

$$I_x(a,b) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt.$$

Taking into account this expression one can write immediately that formula for moments

$$\mu_{k:n}^{(r)} = E\left(X_{k,n}\right)^r$$

of order statistics $X_{k,n}$, related to a population with a d.f. F, is given as follows:

$$\begin{aligned}\mu_{k:n}^{(r)} &= \int_{-\infty}^{\infty} x^r dF_{k:n}(x) \\ &= \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^r (F(x))^{k-1} (1-F(x))^{n-k} dF(x)\end{aligned} . \tag{2.3.1}$$

If F is continuous then this equality can be expressed as

$$\mu_{k:n}^{(r)} = \frac{n!}{(k-1)!(n-k)!} \int_0^1 (G(u))^r u^{k-1} (1-u)^{n-k} du, \tag{2.3.2}$$

where $G(u)$ is the inverse of F. For distributions with probability density function f the RHS of (2.3.2) coincides with

$$\frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^r (F(x))^{k-1} (1-F(x))^{n-k} f(x) dx. \tag{2.3.3}$$

Similar relations are valid for joint (product) moments of order statistics. For the sake of simplicity we consider joint moments

$$\mu_{k(1),k(2):n}^{(r(1),r(2))} = E\left(X_{k(1),n}\right)^{r(1)} \left(X_{k(2),n}\right)^{r(2)}, 1 \le k(1) < k(2) \le n,$$

of two order statistics only.

For absolutely continuous distributions one gets that

$$\begin{aligned}\mu_{k(1),k(2):n}^{(r(1),r(2))} = c(r(1), r(2), n) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{r(1)} y^{r(2)} (F(x))^{r(1)-1} (F(y) - F(x))^{r(2)-r(1)-1} \\ (1-F(y))^{n-r(2)} f(x) f(y) dx dy,\end{aligned} \tag{2.3.4}$$

where

$$c(r(1), r(2), n) = \frac{n!}{(r(1)-1)!(r(2)-r(1)-1)!(n-r(2))!}. \tag{2.3.5}$$

In the general case (for any d.f. F) we have the following expression:

$$\mu_{k(1),k(2):n}^{(r(1),r(2))} = c(r(1),r(2),n) \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^{r(1)} y^{r(2)} (F(x))^{r(1)-1} (F(y)-F(x))^{r(2)-r(1)-1} (1-F(y))^{n-r(2)} dF(x) dF(y), \tag{2.3.6}$$

where coefficients $c(r(1), r(2),n)$ are defined in (2.3.5). Below we will use the following notations also:

$$\mu_{k:n} = EX_{k.n}$$

will be applied for the sake of simplicity instead of $\mu_{k:n}^{(1)}$; $\mu_{k(1),k(2):n} = EX_{k(1),n}\, X_{k(2),n}$ will change $\mu_{k(1),k(2):n}^{(1,1)}$;

$$Var(X_{k,n}) = \mu_{k:n}^{(2)} - (\mu_{k:n})^2$$

will denote the variance of $X_{k,n}$;

$$cov(X_{r,n}, X_{s,n}) = \mu_{r,s:n} - \mu_{r:n}\mu_{s:n}$$

will be used for the covariance between $X_{r,n}$ and $X_{s,n}$.

It is interesting to find the corresponding conditions, which provide the existence of different moments of order statistics.

Example 2.3.1 Let there exist the population moment $\alpha_r = EX^r$, i.e.,

$$E|X|^r = \int_{-\infty}^{\infty} |x|^r dF(x) < \infty. \tag{2.3.7}$$

Then due to (2.3.1) we obtain that

$$\begin{aligned} E|X_{k,n}|^r &\le \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} |x|^r (F(x))^{k-1} (1-F(x))^{n-k} dF(x) \\ &\le \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} |x|^r dF(x) = \frac{n!}{(k-1)!(n-k)!} E|X|^r < \infty. \end{aligned} \tag{2.3.8}$$

It follows from (2.3.8) that the existence of the moment α_r implies the existence of all moments

$$E|X_{k,n}|^r, 1 \le k \le n, n = 1, 2, \ldots$$

of order statistics $X_{k,n}$.

Remark 2.3.1 It is not difficult to prove that if

$$E|X|^r = \infty$$

for some r, then for any $n = 1, 2, \ldots$ there exists such order statistic $X_{k,n}$ that

$$E|X_{k,n}|^r = \infty.$$

Remark 2.3.2 Since

$$P\{X_{s,n} \le X_{k,n} \le X_{m,n}\} = 1$$

for any $s < k < m$, one has the evident inequality

$$E|X_{k,n}|^r \le E\left(|X_{s,n}|^r + |X_{m,n}|^r\right).$$

Hence, if $E|X_{k,n}|^r = \infty$, then at least one of equalities $E|X_{s,n}|^r = \infty$ or $E|X_{m,n}|^r = \infty$ is valid. It means that if $E|X_{k,n}|^r = \infty$, then at least one of the following relations hold:

$$E|X_{1,n}|^r = E|X_{2,n}|^r = \ldots = E|X_{k-1,n}|^r = \infty$$

or

$$E|X_{k+1,n}|^r = E|X_{k+2,n}|^r = \ldots = E|X_{n,n}|^r = \infty.$$

Exercise 2.3.1 Let X's have the Cauchy distribution with the density function

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Show that for any $r = 1, 2, \ldots$, relation

$$E|X_{k,n}|^r < \infty$$

holds if $r < k < n - r + 1$.

Remark 2.3.3 The more general situation than one given in Exercise 2.3.1 was considered by Sen (1959).

Let $E|X|^{\alpha} < \infty$. It appears that then moments $\mu_{k:n}^{(r)}$ exist for all k such that

$$r/\alpha \le k \le (n - r + 1)/\alpha.$$

Some useful relations for moments one can obtain from the evident identity

$$X_{1,n} + \cdots + X_{n,n} = X_1 + \cdots + X_n \tag{2.3.9}$$

and related equalities. For instance, the simplest corollary of (2.3.9) is as follows:

$$E\big(X_{1,n} + \cdots + X_{n,n}\big) = E(X_1 + \cdots + X_n) = nEX. \tag{2.3.10}$$

Natural generalizations of (2.3.10) have the form

$$g\left(\sum_{k=1}^{n} h\big(X_{k,n}\big)\right) = g\left(\sum_{k=1}^{n} h(X_k)\right), \tag{2.3.11}$$

where $g(x)$ and $h(x)$ are arbitrary functions.

Example 2.3.2 The corresponding equalities based on (2.3.11) can be useful in some situations:

$$E\left(\sum_{k=1}^{n} X_{k,n}^{m}\right)^{r} = E\left(\sum_{k=1}^{n} X_k^{m}\right)^{r},\ m = 1, 2, \ldots,\ r = 1, 2, \ldots. \tag{2.3.12}$$

If $r = 1$ we get from (2.3.12) that the equality

$$\sum_{k=1}^{n} EX_{k,n}^{m} = \sum_{k=1}^{n} EX_k^{m} = nEX^{m} \tag{2.3.13}$$

holds for any m provided that the corresponding moment EX^m exists.

It is not difficult to understand that the representations of the uniform and exponential order statistics via sums or products of independent random variables given above allow us to find rather easily single and joint moments of these statistics.

Consider the case of the uniform $U([0,1])$ distribution and the corresponding order statistics $U_{k,n}$. From Representation 1 and Remark 2.2.1 we know that

$$U_{k,n} \stackrel{d}{=} \frac{S_k}{S_{n+1}}$$

and S_k/S_{n+1} is independent on the sum $S_{n+1} = v_1 + v_2 + \ldots + v_{n+1}$,where $v_1, v_2,\ldots$ be independent random variables having the standard $E(1)$ exponential distribution.

Then

$$E\left(U_{k,n}\right)^{\alpha} = E\left(\frac{S_k}{S_{n+1}}\right)^{\alpha}.$$

Now due to the independence of S_k/S_{n+1} and S_{n+1} we have the following relation:

$$E(S_k)^{\alpha} = E\left(\frac{S_k}{S_{n+1}}S_{n+1}\right)^{\alpha} = E\left(\frac{S_k}{S_{n+1}}\right)^{\alpha} E(S_{n+1})^{\alpha}.$$

Thus,

$$E\left(\frac{S_k}{S_{n+1}}\right)^{\alpha} = E(S_k)^{\alpha}/E(S_{n+1})^{\alpha}.$$

Now we must recall that S_m has gamma distribution with parameter m and hence

$$E(S_m)^{\alpha} = \frac{1}{(m-1)!}\int_0^{\infty} x^{a+m-1}e^{-x}dx = \Gamma(\alpha+m)/\Gamma(m), m = 1,2,\ldots.$$

Here $\Gamma(s)$ denotes gamma function with parameter s.

Finally,

$$E\left(U_{k,n}\right)^{\alpha} = \frac{\Gamma(\alpha+k)\Gamma(n+1)}{\Gamma(k)\Gamma(\alpha+n+1)} = \frac{n!\Gamma(\alpha+k)}{(k-1)!\Gamma(\alpha+n+1)} \tag{2.3.14}$$

Note that (2.3.14) is true for any $\alpha > -k$.

Now we can use Representation 3 to get joint (product) moments of the uniform order statistics. It follows from this representation, where $W_1, W_2,\ldots$ are independent uniformly $U([0,1])$ distributed random variables, that

$$\begin{aligned} EU_{r,n}U_{s,n} &= E\left(W_r^{1/r}W_{r+1}^{1/(r+1)}\ldots W_n^{1/n}W_s^{1/s}W_{s+1}^{1/(s+1)}\ldots W_n^{1/n}\right) \\ &= E\left(W_r^{1/r}\right)E\left(W_{r+1}^{1/(r+1)}\right)\ldots E\left(W_s^{2/s}\right)E(W_{s+1})^{2/(s+1)}\ldots E\left(W_n^{2/n}\right) \\ &= \prod_{k=r}^{s-1}\frac{1}{(1+1/k)}\prod_{k=s}^{n}\frac{1}{(1+2/k)} = \frac{r(s+1)}{(n+1)(n+2)}. \end{aligned} \tag{2.3.15}$$

for any $1 \le r < s \le n$.

Remark 2.3.4 Indeed, Representation 3 allows us to get also more complicate joint moments of any order for any number of the uniform order statistics. Taking into account equality (2.3.14) we can fix the following important partial cases:

$$EU_{k,n} = \frac{k}{n+1},\ 1 \le k \le n; \tag{2.3.16}$$

$$E\left(1/U_{k,n}\right) = \frac{n}{k-1},\ 2 \le k \le n; \tag{2.3.17}$$

$$E\left(U_{k,n}\right)^2 = \frac{k(k+1)}{(n+1)(n+2)},\ 1 \le k \le n. \tag{2.3.18}$$

In general form, for $r = 1, 2, \ldots$, we have

$$E\left(U_{k,n}\right)^r = \frac{k(k+1)\ldots(k+r-1)}{(n+1)(n+2)\ldots(n+r)},\ 1 \le k \le n. \tag{2.3.19}$$

It follows from (2.3.16) and (2.3.18) that

$$Var\left(U_{k,n}\right) = \frac{k(n-k+1)}{(n+1)^2(n+2)},\ 1 \le k \le n. \tag{2.3.20}$$

Taking into account equality (2.3.15) one gets that

$$Cov\left(U_{r,n}, U_{s,n}\right) = EU_{r,n}U_{s,n} - EU_{r,n}EU_{s,n} = \frac{r(n-s+1)}{(n+1)^2(n+2)},\ r \le s. \tag{2.3.21}$$

Now from (2.3.20) and (2.3.21) we obtain the corresponding expression for correlation coefficients of the uniform order statistics:

$$\rho\left(U_{r,n}, U_{s,n}\right) = \left(\frac{r(n-s+1)}{s(n-r+1)}\right)^{1/2},\ r \le s. \tag{2.3.22}$$

Exercise 2.3.2 Find the variance of $1/U_{k,n}$.

Exercise 2.3.3 Find product moments

$$E(U_{r,n}^{\alpha}U_{s,n}^{\beta}),\ \alpha \ge 0, \beta \ge 0.$$

Let us consider now order statistics $Z_{1,n} \le Z_{2,n} \le \cdots \le Z_{n,n}$, $n = 1, 2, \ldots$, which correspond to the standard $E(1)$ exponential distribution with d.f.

$$H(x) = 1 - \exp(-x),\ x > 0.$$

Let us recall Representation 4:

$$\left(Z_{1,n},\ Z_{2,n}, \ldots, Z_{n,n}\right) \stackrel{d}{=} \left(\frac{v_1}{n}, \frac{v_1}{n} + \frac{v_2}{n-1}, \ldots, \frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_{n-1}}{2} + v_n\right),$$

where $v_1,\ v_2, \ldots,\ v_n$ are independent random variables having the standard $E(1)$ exponential distribution.

One immediately gets applying this result that

$$EZ_{k,n} = E\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_k}{n-k+1}\right) = \sum_{r=1}^{k} \frac{1}{n-r+1} \tag{2.3.23}$$

and

$$Var\left(Z_{k,n}\right) = \sum_{r=1}^{k} \mathrm{Var}\left(\frac{v_r}{n-r+1}\right) = \sum_{r=1}^{k} \frac{1}{(n-r+1)^2}. \tag{2.3.24}$$

Exercise 2.3.4 Applying Representation 4 find now covariances between order statistics $Z_{r,n}$ and $Z_{s,n}$.

Calculations of the integral

$$\frac{n!}{(k-1)!(n-k)!} \int_0^\infty x^\alpha (1 - e^{-x})^{k-1} e^{-x(n-k+1)} dx$$

also allow us to get moments

$$E\left(Z_{k,n}\right)^\alpha,\ k = 1, 2, \ldots, n.$$

We obtain after the corresponding calculations that

$$E\left(Z_{k,n}\right)^\alpha = \frac{n!}{(k-1)!(n-k)!} \sum_{r=0}^{k-1} (-1)^r \binom{k-1}{r} \Gamma(\alpha+1)/(n-k+r+1)^{(\alpha+1)}. \tag{2.3.25}$$

For instance, if $k = 1$, then

$$E\left(Z_{1,n}\right)^\alpha = \Gamma(\alpha+1)/\mathrm{n}^\alpha, \alpha > -1. \tag{2.3.26}$$

For $k = 2$ and $\alpha > -1$ we have

$$E\left(Z_{2,n}\right)^{\alpha} = n(n-1)\Gamma(\alpha+1)\{(n-1)^{-(\alpha+1)} - n^{-(\alpha+1)}\}. \tag{2.3.27}$$

Indeed, one of the most important distributions in the probability theory is normal. Let now $X_1, X_2, \ldots$ be independent random variables having the standard $N(0,1)$ normal distribution and $X_{1,n} \le \cdots \le X_{n,n}$ be the corresponding normal order statistics.

One can write immediately that

$$EX_{k,n}^{r} = \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^{r} \Phi^{k-1}(x)(1-\Phi(x))^{n-k}\varphi(x)dx, \tag{2.3.28}$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$$

and

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt.$$

There are some effective numerical methods to compute integrals (2.3.28). Unlike the uniform and exponential cases moments of the normal order statistics have the explicit expressions for small sample sizes n only.

Example 2.3.3 Consider the case $n = 2$. We get that

$$EX_{2,2} = 2\int_{-\infty}^{\infty} x\Phi(x)\varphi(x)dx = -2\int_{-\infty}^{\infty} \Phi(x)d(\varphi(x))$$
$$= 2\int_{-\infty}^{\infty} \varphi^2(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \exp\left(-x^2\right)dx = \frac{1}{\sqrt{\pi}}.$$

From the identity

$$E\left(X_{1,2} + X_{2,2}\right) = E(X_1 + X_2) = 0$$

we obtain now that

$$EX_{1,2} = -EX_{2,2} = -\frac{1}{\sqrt{\pi}}.$$

Remark 2.3.5 If we have two samples $X_1, X_2, \ldots, X_n$ (from the standard $N(0,1)$ normal distribution) and $Y_1, Y_2, \ldots, Y_n$ (from the normal $N(a,\sigma^2)$ distribution with expectation a and variance σ^2, $\sigma > 0$), then evidently

$$EY_{k,n} = a + \sigma X_{k,n}.$$

Exercise 2.3.5 Let $X_{1,3} \le X_{2,3} \le X_{3,3}$ be the order statistics corresponding to the standard normal distribution. Find expectations $EX_{k,3}$ and variances $Var(X_{k,3})$, $k = 1, 2, 3$.

The explicit expressions for moments of the normal order statistics are rather complicated, although the normal distribution possesses a number of useful properties, which can simplify the computational schemes in some situations.

Example 2.3.4 A lot of statistical procedures for the normal distribution are based on the independence property of vector

$$(X_1 - \bar{X}, X_2 - \bar{X}, \ldots, X_n - \bar{X})$$

and the sample mean

$$\bar{X} = (X_1 + X_2 + \cdots + X_n)/n.$$

This property yields also the independence of vector

$$(X_{1,n} - \bar{X}, X_{2,n} - \bar{X}, \ldots, X_{n,n} - \bar{X})$$

and the sample mean $\bar{X}$.

Let $X_1, X_2, \ldots, X_n$ be a sample from the standard normal distribution. We see then that

$$E(X_{k,n} - \bar{X})\bar{X} = E(X_{k,n} - \bar{X})E\bar{X} = 0 \tag{2.3.29}$$

and we obtain the following results:

$$EX_{k,n}\bar{X} = E(\bar{X}^2) = Var\,\bar{X} = 1/n,\ k = 1, 2, \ldots, n, \tag{2.3.30}$$

and hence

$$\sum_{m=1}^{n} EX_{k,n}X_m = 1, \tag{2.3.31}$$

as well as

$$\sum_{m=1}^{n} EX_{k,n}X_{m,n} = 1. \tag{2.3.32}$$

As corollaries of (2.3.31) one gets that

$$E\big(X_{k,n}X_m\big) = 1/n \tag{2.3.33}$$

and

$$cov\big(X_{k,n}, X_m\big) = E\big(X_{k,n}X_m\big) - EX_{k,n}EX_m = 1/n \tag{2.3.34}$$

for any $k = 1, 2,\ldots, n$, $m = 1, 2,\ldots, n$ and $n = 1, 2,\ldots$.

Note that there are different tables (see, for example, Teichroew (1956) or Tietjen et al. (1977)) which give expected values and some other moments of order statistics for samples of large sizes from the normal distribution.

2.4 Extremes

For any person who wants to study records it is necessary to know the theory of the extreme order statistics.

Order statistics $X_{k(n),n}$ are said to be extreme if $k = k(n)$ or $n - k(n) + 1$ is fixed, as $n \to \infty$. The most popular are maximal order statistics $X_{n,n}$ and minimal order statistics $X_{1,n}$.

It is not difficult to see that if we have two sequences $Y_1 = -X_1$, $Y_2 = -X_2,\ldots, Y_n = -X_n$ then the following equalities hold for any $k = 1, 2,\ldots, n$:

$$X_{n-k+1,n} \stackrel{d}{=} -Y_{k,n}, \tag{2.4.1}$$

and, in particular, in this case

$$X_{n,n} \stackrel{d}{=} -Y_{1,n}. \tag{2.4.2}$$

Due to these relationships between maximal and minimal order statistics we can study one type of them only, say, maximal ones.

Very often we need to know asymptotic distributions of $X_{1,n}$ and $X_{n,n}$, as $n \to \infty$.

Consider a sequence of order statistics $M(n) = X_{n,n}$, $n = 1, 2,\ldots$. Let $F(x)$ be the distribution function of X and

$$\beta = \sup\{x : F(x) < 1\}$$

be the right end point of the support of X. If $\beta = \infty$, then for any finite x one gets that $F(x) < 1$ and hence

$$P\{M(n) \leq x\} = (F(x))^n \to 0, \text{as } n \to \infty.$$

It means that $M(n)$ converges to infinity. In the case, when $\beta < \infty$, we need to distinguish two situations. If

$$P\{X = \beta\} = p > 0,$$

then

$$P\{M(n) = \beta\} = 1 - P\{M(n) < \beta\} = 1 - P^n(X < \beta) = 1 - (1-p)^n$$

and

$$P\{M(n) = \beta\} \to 1, \text{ as } n \to \infty. \tag{2.4.3}$$

If $P\{X = \beta\} = 0$, then we get that $P\{M(n) < \beta\} = 1$ for any n and $M(n) \to \beta$ in distribution. Thus, we see that in all situations $M(n) \to \beta$ in distribution. This result can be sharpened if we consider the asymptotic distributions of the centered and normalized order statistics $X_{n,n}$. Indeed, if $\beta < \infty$ and $P\{X = \beta\} > 0$, then relation (2.4.3) gives completed information on $M(n)$ and in this case any centering and normalizing can not improve our knowledge about the asymptotic behavior of $M(n)$. We have another situation if $P\{X = \beta\} = 0$. In this case one can try to solve the following problem: if there exist any centering (a_n) and norming (b_n) constants, such that the sequence $(M(n) - a_n)/b_n$ converges to some nondegenerate distribution? Let us solve firstly the following exercises.

Exercise 2.4.1 Consider the exponential distribution with d.f. $F(x) = 1 - \exp(-x)$, $x \geq 0$. Show that in this case $M(n) - \log n$ *converges* to some nondegenerate distribution with d.f.

$$H_0(x) = \exp(-\exp(-x)). \tag{2.4.4}$$

Exercise 2.4.2 Consider X's with d.f.

$$F(x) = 1 - (-x)^{\alpha}, -1 < x < 0, \alpha > 0,$$

and prove that the asymptotic distribution of $n^{1/\alpha}M(n)$ has the following form:

$$H_{1,\alpha}(x) = \exp\left(-(-x)^{\alpha}\right), -\infty < x \leq 0,$$

and

$$H_{1,\alpha}(x) = 1,\ x > 0. \tag{2.4.5}$$

Exercise 2.4.3 Let X's have Pareto distribution with d.f.

$$F(x) = 1 - x^{-\alpha},\ x > 1, \alpha > 0.$$

Prove that the asymptotic d.f. of $M(n)/n^{1/\alpha}$ is of the form:

$$H_{2,\alpha}(x) = 0,\ x<0,$$

and

$$H_{2,\alpha}(x) = \exp\{-x^{-\alpha}\},\ x \geq 0. \tag{2.4.6}$$

Remark 2.4.1 Changing suitably the normalized constants a_n and b_n for maximal values considered in Exercises 2.4.1–2.4.3, one gets that any d.f. of the form $H_0(a + bx)$, $H_{1,\alpha}(a + bx)$, $H_{2,\alpha}(a + bx)$, where $\alpha > 0$, $b > 0$ and a are arbitrary constants, can serve as the limit distribution for

$$(M(n)-a_n)/b_n.$$

Considering two d.f.'s, $H(d + cx)$ and $H(a + bx)$, where $b > 0$ and $c > 0$, we say that these d.f.'s belong to the same type of distributions. Any distribution of the given type can be obtained from other distribution of the same type by some linear transformation.

Usually one of distributions, say $H(x)$, having the most simplest (or convenient) form, is chosen to represent all the distributions of the given type, which we call then H-type. As basic for their own types, we suggested above the following distributions:

$$\begin{aligned} H_0(x) &= \exp(-\exp(-x)); \\ H_{1,\alpha}(x) &= \exp\left(-(-x)^{\alpha}\right),\ -\infty < x \leq 0, \text{ and } H_{1,\alpha}(x) = 1, x > 0; \\ H_{2,\alpha}(x) &= 0, x < 0, \text{ and } H_{2,\alpha}(x) = \exp\{-x^{-\alpha}\},\ x \geq 0, \end{aligned}$$

where $\alpha > 0$.

Very often one can find that the types of distributions based on $H_0(x)$, $H_{1,\alpha}(x)$ and $H_{2,\alpha}(x)$ are named correspondingly as *Gumbel*, *Frechet* and *Weibull* types of the limiting extreme value distributions.

Note also that any two of d.f.'s $H_{1,\alpha}$ and $H_{1,\beta}$, $\alpha \neq \beta$, do not belong to the same type, as well as d.f.'s $H_{2,\alpha}$ and $H_{2,\beta}$, $\alpha \neq \beta$.

There are considered above three situations when we can get d.f.'s (2.4.4)–(2.4.6) as the limit d.f.'s for the normalized order statistics $X_{n,n}$.

It is surprising that there are none of other non-degenerate distributions, besides of $H_0(x)$-, $H_{1,\alpha}(x)$- and $H_{2,\alpha}(x)$—types, which would be limit for the suitably centering and norming maximal values.

Remark 2.4.2 It appeared that the set of all possible non-degenerate limit distributions for maximal values includes only the types generated by d.f.'s H_0, $H_{1,\alpha}$ and $H_{2,\alpha}$ only. Now it is important for us to know what d.f.'s F belong to the domains of attraction ($D(H_0)$, $D(H_{1,\alpha})$ and $D(H_{2,\alpha})$) of the corresponding limit laws. We write that $F \in D(H)$, if the suitably normalized maximal values $M(n)$, based on X's with a common d.f. F, have the limit d.f. H. For instance, it follows (from the results of Exercises 2.4.1–2.4.3 and Remark 2.4.1) that if

$$F(x) = 1 - \exp(-x),\ x > 0,$$

then $F \in D(H_0)$;
if X's are the uniformly $U([a,b])$ distributed random variables with

$$F(x) = (x-a)/(b-a),\ a < x < b,$$

then $F \in D(H_{1,1})$
and if

$$F(x) = 1 - x^{-\alpha},\ x > 1$$

(Pareto distribution), then $F \in D(H_{2,\alpha})$.

There are necessary and sufficient conditions for F to belong $D(H_0)$, $D(H_{1,\alpha})$ and $D(H_{2,\alpha})$ but the form of these conditions is rather cumbersome. Hence, simple sufficient conditions are more interesting for us. We present below some of them.

Theorem 2.4.1 *Let d.f. F have positive derivative F' for all $x > x_0$. If the following relation is valid for some $\alpha > 0$:*

$$xF'(x)/(1 - F(x)) \to \alpha, \tag{2.4.7}$$

as $x \to \infty$, then $F \in D(H_{2,\alpha})$. The centering, a_n, and normalizing, b_n, constants can be taken to satisfy relations $a_n = 0$ and $F(b_n) = 1 - 1/n$.

Theorem 2.4.2 *Let d.f. F have positive derivative F' for x in some interval (x_1, x_0) and $F'(x) = 0$ for $x > x_0$. If*

$$(x_0 - x)F'(x)/(1 - F(x)) \to \alpha, x \to x_0, \tag{2.4.8}$$

then $F \in D(H_{1,\alpha})$. The centering, a_n, and normalizing, b_n, constants can be taken to satisfy relations $a_n = x_0$ and $F(x_0 - b_n) = 1 - 1/n$.

Theorem 2.4.3 *Let d.f. F have negative second derivative $F''(x)$ for x in some interval (x_1, x_0), and let $F'(x) = 0$ for $x > x_0$. If*

$$F''(x)\,(1 - F(x))/(F'(x))^2 = -1, \tag{2.4.9}$$

then $F \in D(H_0)$. The centering, a_n, and normalizing, b_n, constants can be taken to satisfy relations $F(a_n) = 1 - 1/n$ and $b_n = h(a_n)$, where

$$h(x) = (1 - F(x))/F'(x).$$

Exercise 2.4.4 Let

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\, x$$

(the Cauchy distribution). Prove that in this situation

$$F \in D\big(H_{2,1}\big).$$

What normalizing constants, a_n and b_n, can be taken in this case?

Exercise 2.4.5 Let

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\big(-t^2/2\big)dt.$$

Show that

$$F \in D(H_0)$$

and the normalizing constants can be taken as follows:

$$a_n = (2\log n - \log\log n - \log\, 4\pi)^{1/2} \text{and}\, b_n = (2\log n)^{1/2}.$$

Exercise 2.4.6 Use Theorem 2.4.2 and find the limit distribution and the corresponding normalizing constants for gamma distribution with p.d.f.

$$f(x) = x^{\alpha-1}\exp(-x)/\Gamma(\alpha), x > 0, \alpha > 0.$$

Returning to relations (2.4.1) and (2.4.2) one can find the possible types of limit distributions for minimal values

$$m(n) = \min\{X_1, \ldots, X_n\}.$$

It appears that the corresponding set of non-degenerate asymptotic d.f.'s for the suitably normalized minimal values are defined by the following basic d.f.'s:

$$L_0(x) = 1 - \exp(-\exp(x)); \tag{2.4.10}$$

$$L_{1,\alpha}(x) = 0, x<0, \text{ and } L_{1,\alpha}(x) = 1 - \exp\,(-x^{\alpha}), 0 \le x < \infty; \tag{2.4.11}$$

$$L_{2,\alpha}(x) = 1 - \exp\{-(-x)^{-\alpha}\}, x<0, \text{ and } L_{2,\alpha}(x) = 1,\ x \ge 0, \tag{2.4.12}$$

where $\alpha > 0$.

Above we considered the situation with the asymptotic behavior of extremes $X_{n,n}$ and $X_{1,n}$. Analogous methods can be applied to investigate the possible asymptotic distributions of the kth extremes—order statistics $X_{n-k+1,n}$ and $X_{k,n}$, when $k = 2, 3, \ldots$ is some fixed number and n tends to infinity. The following results are valid in these situations.

Theorem 2.4.4 *Let random variables $X_1, X_2, \ldots$ be independent and have a common d.f. F and $X_{n-k+1,n}$, $n = k, k + 1, \ldots$, be the $(n - k + 1)$th order statistics. If for some normalizing constants a_n and b_n,*

$$P\{(X_{n,n}-a_n)/b_n<x\} \to H(x)$$

in distribution, as $n \to \infty$, then the limit relation

$$P\{(X_{n-k+1,n}-a_n)/b_n<x\} \to H(x)\sum_{j=0}^{k-1}(-\log H(x))^j/j! \tag{2.4.13}$$

holds for any x, as $n \to \infty$.

Theorem 2.4.5 *Let random variables $X_1, X_2, \ldots$ be independent and have a common d.f. F and $X_{k,n}$, $n = k, k + 1, \ldots$, be the kth order statistics. If for some normalizing constants a_n and b_n,*

$$P\{(X_{1,n}-a_n)/b_n<x\} \to L(x),\ n \to \infty,$$

in distribution, then the limit relation

$$P\{(X_{k,n}-a_n)/b_n<x\} \to L(x)\sum_{j=0}^{k-1}(-\log L(x))^j/j! \tag{2.4.14}$$

is valid for any x, as $n \to \infty$.

2.5 Order Statistics and Ranks

In Chap. 1 we mentioned that a random sample $X_1, X_2, \ldots, X_n$ is accompanied by the corresponding vector of ranks $(R(1), R(2), \ldots, R(n))$, where

$$R(m) = \sum_{k=1}^{n} 1_{\{X_m \geq X_k\}}, m = 1, 2, \ldots, n. \tag{2.5.1}$$

The following relations are valid for ranks and order statistics:

$$\{R(m) = k\} = \{X_m = X_{k,n}\},\ m = 1, 2, \ldots, n,\ k = 1, 2, \ldots, n. \tag{2.5.2}$$

If X's are proposed to have a continuous distribution then different X's can coincide with zero probability and (2.5.1) can be rewritten in the following form:

$$R(m) = 1 + \sum_{k=1}^{n} 1_{\{X_m > X_k\}}. \tag{2.5.3}$$

These ranks show the location of X's among the elements of the variational series

$$X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}.$$

We can write also that

$$X_m = X_{R(m),n},\ m = 1, 2, \ldots, n. \tag{2.5.4}$$

Consider the case when X's have a continuous distribution. Then any rank $R(m)$, $m = 1, 2, \ldots, n$, has the discrete uniform distribution on set $\{1, 2, \ldots, n\}$. Really, we know that random variables $X_1, X_2, \ldots, X_n$ taken from a continuous distribution have no coincidences with probability one. Hence, realizations $(r(1), \ldots, r(n))$ of the corresponding vector of ranks $(R(1), R(2), \ldots, R(n))$ represent all permutations of values $1, 2, \ldots, n$. Any realization $(r(1), \ldots, r(n))$ corresponds to the event

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

where $\delta(r(k)) = k$. Taking into account the symmetry of the sample $X_1, X_2, \ldots, X_n$, one obtains that events

$$(X_{\delta(1)} < X_{\delta(2)} < \cdots < X_{\delta(n)})$$

have the same probabilities for any permutations $(\delta(1), \ldots, \delta(n))$ of numbers $1, 2, \ldots, n$. Hence

$$\begin{aligned} &P\{R(1) = r(1), R(2) = r(2), \ldots, R(n) = r(n)\} \\ &= \mathrm{P}\{(\mathrm{X}_{\delta(1)} < \mathrm{X}_{\delta(2)} < \cdots < \mathrm{X}_{\delta(\mathrm{n})})\} = 1/\mathrm{n}! \end{aligned} \tag{2.5.5}$$

for any permutation $(r(1), \ldots, r(n))$ of numbers $1, 2, \ldots, n$. It follows now from (2.5.5) that

$$P\{R(m) = k\} = 1/n$$

for any fixed $m = 1, 2,\ldots, n$ and $k = 1, 2,\ldots, n$.

Indeed, ranks $R(1), R(2),\ldots, R(n)$ are dependent random variables for any $n = 2, 3,\ldots$. This dependence is approved by the evident equality

$$R(1) + R(2) + \cdots + R(n) = 1 + 2 + \cdots + n = n(n+1)/2,$$

but it follows from (2.5.5) that ranks present exchangeable random variables: for any permutation $(\alpha(1), \alpha(2),\ldots, \alpha(n))$ of numbers $1, 2,\ldots, n$ vectors $(R(\alpha(1)),\ldots, R(\alpha(n)))$ and $(R(1),\ldots, R(n))$ have the same distributions.

Exercise 2.5.1 Find expectations and variances of $R(k)$, $1 \le k \le n$, covariances $Cov(R(k), R(m))$ and the correlation coefficients $\rho(R(k), R(m))$ between $R(k)$ and $R(m)$, $1 \le k < m \le n$.

Above we mentioned that any realization $(r(1),\ldots, r(n))$ of $(R(1), R(2),\ldots, R(n))$ corresponds to the event

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

where $\delta(r(k)) = k$. Here $\delta(k)$ denotes the index of X, the rank of which for this realization takes on the value k. For different realizations of the vector $(R(1), R(2),\ldots, R(n))$, $\delta(k)$ can take on different values from the set $\{1, 2,\ldots, n\}$ and we really deal with new random variables, with realizations as $\delta(r(1)), \delta(r(2)),\ldots, \delta(r(n))$.

Let as above $X_1, X_2,\ldots, X_n$ be a random sample of size n taken from a continuous distribution and $X_{1,n}, X_{2,n},\ldots, X_{n,n}$ be the corresponding order statistics. We introduce random variables $\Delta(1), \Delta(2),\ldots, \Delta(n)$, which satisfy the following equalities:

$$\{\Delta(m) = k\} = \{X_{m,n} = X_k\}, m = 1, 2,\ldots, n,\ k = 1, 2,\ldots, n. \tag{2.5.6}$$

These random variables are said to be *antiranks*.

The same arguments, which we used for ranks, show that any realization $(\delta(1), \delta(2),\ldots, \delta(n))$ of the vector $(\Delta(1), \Delta(2),\ldots, \Delta(n))$ is a permutation of numbers $1, 2,\ldots, n$ and all n! such realizations have equal probabilities, $1/n!$ each.

Indeed, vectors of antiranks are tied closely with the corresponding order statistics and vectors of ranks. In fact, for any k and m equalities

$$\{\Delta(k) = m\} = \{X_{k,n} = X_m\} = \{R(m) = k\} \tag{2.5.7}$$

hold with probability one. One can write also the following relations for ranks and antiranks:

$$\Delta(R(m)) = m \tag{2.5.8}$$

and

$$R(\Delta(m)) = m, \tag{2.5.9}$$

which hold with probability one for any $m = 1, 2, \ldots, n$.

Exercise 2.5.2 Find the joint distribution of $\Delta(1)$ and $R(1)$.

While ranks and antiranks are associated with some random sample $X_1, X_2, \ldots, X_n$ of some fixed size n, there are rank statistics (the so-called sequential ranks), which characterize a sequence of random variables $X_1, X_2, \ldots$.

Let $X_1, X_2, \ldots$ be independent random variables, having continuous (not necessary identical) distributions. Random variables $\rho(1), \rho(2), \ldots$ given by equalities:

$$\rho(m) = \sum_{k=1}^{m} 1_{\{X_m \geq X_k\}}, m = 1, 2, \ldots, \tag{2.5.10}$$

are said to be *sequential ranks*.

Any sequential rank $\rho(m)$ shows the position of a new coming observation X_m among its predecessors $X_1, X_2, \ldots, X_{m-1}$. For instance, if $\rho(m) = 1$, then X_m is less than $X_{1,\, m-1}$ and it means that

$$X_m = X_{1,m}.$$

In general, $\rho(m) = k$ implies that

$$X_m = X_{k,m}.$$

It is not difficult to see that $\rho(m)$ can take on the values $1, 2, \ldots, m$. If independent random variables $X_1, X_2, \ldots, X_m$ have the same continuous distribution then the standard arguments used above enable us to see that for any $m = 1, 2, \ldots,$

$$P\{\rho(m) = k\} = P\{X_m = X_{k,m}\} = 1/m,\ k = 1, 2, \ldots, m. \tag{2.5.11}$$

Exercise 2.5.3 Let $X_1, X_2, \ldots$ be independent random variables with a common continuous d.f. F. Prove that the corresponding sequential ranks $\rho(1), \rho(2), \ldots$ are independent.

Note that if $\rho(m) = m$ than it means that $X_m = X_{m,m}$ and

$$X_m > \max\{X_1, X_2, \ldots, X_{m-1}\}. \tag{2.5.12}$$

Some later it will stay clear that (2.5.12) defines that X_m is the upper record value in the sequence X_1, X_2, ….

2.6 Exercises (solutions)

Exercise 2.1.1 (*solution*) There are two situations here when $X_{1,n}$ and $X_{n,n}$ can be equal: all X's are equal 0 and all X's are equal 1. Hence

$$p_n = P\{X_1 = X_2 = \ldots = X_n = 0\} + P\{X_1 = X_2 = \ldots = X_n = 1\} = 1/2^n + 1/2^n = 1/2^{n-1}$$

Exercise 2.1.2 (*solution*) In this situation the necessary probability can be written as follows:

$$\begin{aligned} P\{X_{1,3} < X_{2,3} < X_{3,3}\} &= 1 - P\{X_1 = X_2 \neq X_3\} - P\{X_1 = X_3 \neq X_2\} - P\{X_2 = X_3 \neq X_1\} - P\{X_1 = X_2 = X_3\} \\ &= 1 - 3P\{X_1 = X_2 \neq X_3\} - P\{X_1 = X_2 = X_3\} \\ &= 1 - 3(P\{X_1 = X_2\} - P\{X_1 = X_2 = X_3\}) - P\{X_1 = X_2 = X_3\} \\ &= 1 - 3P\{X_1 = X_2\} + 2P\{X_1 = X_2 = X_3\}. \end{aligned}$$

It is not difficult to find that

$$P\{X_1 = X_2\} = (1-p)^2/(1-p^2)$$

and

$$P\{X_1 = X_2 = X_3\} = (1-p)^3/(1-p^3).$$

Thus one gets the final expression:

$$P\{X_{1,3} < X_{2,3} < X_{3,3}\} = 6p^3/(1+p)(1+p+p^2).$$

Exercise 2.1.3 (*solution*) Indeed, if x ≥ y, then

$$P\{X_{1,n} \le x, X_{n,n} \le y\} = P\{X_{n,n} \le y\} = F^n(y).$$

If x < y, then

$$P\{X_{1,n} \le x, X_{n,n} \le y\} = P\{X_{n,n} \le y\} - P\{X_{1,n} > x, X_{n,n} \le y\} \\ = F^n(y) - (F(y) - F(x))^n.$$

Exercise 2.1.4 (*solution*) It is easy to see that

$$P\{m(n) \le x\} = 1 - P\{m(n) > x\} = 1 - P\{X_1 > x\} \ldots P\{X_n > x\} \\ = 1 - (1 - F_1(x)) \ldots (1 - F_n(x))$$

and

$$P\{M(n) \le x\} = P\{X_1 \le x\} \ldots P\{X_n \le x\} = F_1(x) \ldots F_n(x).$$

Exercise 2.1.5 (*solution*) We see that

$$P\{Y \ge r, Z < s\} = P\{r \le X_k < s, k = 1, 2, \ldots, n\} \\ = (P\{r \le X < s\})^n = (P\{X \ge r\} - P\{X \ge s\})^n = (p^r - p^s)^n.$$

Exercise 2.2.1 (*solution*) It follows from (2.2.2) that

$$T(1, n) = U_{n,n} - U_{1,n} \stackrel{d}{=} \frac{v_2 + \cdots + v_n}{v_1 + \cdots + v_{n+1}}.$$

Indeed the ratio $(v_2 + \cdots + v_n)/(v_1 + \cdots + v_{n+1})$ coincides in distribution with the ratio $(v_1 + \cdots + v_{n-1})/(v_1 + \cdots + v_{n+1})$. Hence

$$T(1, n) \stackrel{d}{=} \frac{v_1 + \cdots + v_{n-1}}{v_1 + \cdots + v_{n+1}} \stackrel{d}{=} U_{n-1,n}$$

Exercise 2.2.2 (*solution*) The necessary statement follows immediately from relation (2.2.4), which allows us to get the equality

$$(V_1, V_2, \ldots, V_n) \stackrel{d}{=} (W_1, W_2, \ldots, W_n),$$

where W's are independent uniformly distributed random variables.

Exercise 2.2.3 (*solution*) It follows from (2.2.5) that

$$(Z_{1,n}, L) \stackrel{d}{=} \left(\frac{v_1}{n}, \sum_{k=2}^{n} b_k n_k\right),$$

where

$$b_k = (k-1)/(n-k+1), k = 2, 3, \ldots.$$

Since v_k, k = 1, 2,. . ., n, are independent random variables,

$$\frac{v_1}{n} \text{ and } \sum_{k=2}^{n} b_k n_k$$

are also independent. Hence, so are $Z_{1,n}$ and L.

Exercise 2.2.4 (*hint*) It is enough to use definitions of d.f. $F(x)$ and its inverse function $G(x)$.

Exercise 2.2.5 (*solution*) The corresponding d.f. and its inverse are given here by equalities

$$F(x) = x^a,\ 0 < x < 1.$$

and

$$G(x) = x^{1/a}, 0 < x < 1.$$

It is possible to express $X_{r,n}$ and $X_{s,n}$ via the uniform order statistics as follows:

$$X_{r,n} \stackrel{d}{=} \left(U_{r,n}\right)^{1/a} \text{ and } X_{s,n} \stackrel{d}{=} \left(U_{s,n}\right)^{1/a}.$$

Hence

$$X_{r,n}X_{s,n} \stackrel{d}{=} \left(U_{r,n}U_{s,n}\right)^{1/a}.$$

Now it follows from presentation (2.2.4) that

$$X_{r,n}X_{s,n} \stackrel{d}{=} W_r^{1/ra} W_{r+1}^{1/(r+1)a} \ldots W_{s-1}^{1/(s-1)a} W_s^{2/sa} W_{s+1}^{2/(s+1)a} \ldots W_n^{2/na},$$

$1 \le r < s \le n$, where W_1, W_2,. . ., W_n are independent random variables with the common uniform on [0,1] distribution.

Exercise 2.3.1 (*hint*) Prove that

$$G(x) \sim \frac{1}{\pi(1-x)}, x \to 1,$$
$$G(x) \sim -\frac{1}{\pi x}, x \to 0,$$

and use equality (2.3.2) to provide the necessary statement.

Exercise 2.3.2 (*answer*)

$$Var\left(1/U_{k,n}\right) = \frac{n(n-k+1)}{(k-1)^2(k-2)}, 3 \le k \le n.$$

Exercise 2.3.3 (*answer*)

$$E\left(U_{r,n}^{\alpha} U_{s,n}^{\beta}\right) = \frac{n!\Gamma(r+\alpha)\Gamma(s+\alpha+\beta)}{(r-1)!\Gamma(s+\alpha)\Gamma(n+1+\alpha+\beta)}.$$

Exercise 2.3.4 (*solution*) Let $r \le s$. Due to (2.2.5),

$$\begin{aligned} Cov\left(Z_{r,n}, Z_{s,n}\right) &= Cov\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}, \frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_s}{n-s+1}\right) \\ &= Cov\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}, \frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}\right) \\ &+ Cov\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}, \frac{v_{r+1}}{n-r} + \cdots + \frac{v_s}{n-s+1}\right). \end{aligned}$$

Since sums

$$\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}$$

and

$$\frac{v_{r+1}}{n-r} + \cdots + \frac{v_s}{n-s+1}$$

are independent, we get that

$$\begin{aligned} Cov\left(Z_{r,n}, Z_{s,n}\right) &= Cov\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}, \frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}\right) \\ &= Var\left(\frac{v_1}{n} + \frac{v_2}{n-1} + \cdots + \frac{v_r}{n-r+1}\right) = VarZ_{r,n} = \sum_{k=1}^{r} \frac{1}{(n-k+1)^2}. \end{aligned}$$

Exercise 2.3.5 (*solution*) From symmetry of $N(0,1)$ distribution (note that X has the same distribution as $-X$) it follows that $EX_{2,3} = 0$ and $EX_{1,3} = -EX_{3,3}$. Thus, we need to find $EX_{3,3}$ only. One sees that

$$EX_{3,3} = 3\int_{-\infty}^{\infty} x\Phi^2(x)\varphi(x)dx = 3\int_{-\infty}^{\infty} \Phi^2(x)d\varphi(x)$$
$$= 6\int_{-\infty}^{\infty} \Phi(x)\varphi^2(x)dx.$$

Consider

$$I(a) = \int_{-\infty}^{\infty} \Phi(ax)\varphi^2(x)dx.$$

We obtain that

$$I(0) = \frac{1}{2}\int_{-\infty}^{\infty} \varphi^2(x)dx = \frac{1}{4\pi}\int_{-\infty}^{\infty} \exp\left(-x^2\right)dx = 1/(4\sqrt{\pi})$$

and

$$I'(a) = \int_{-\infty}^{\infty} x\varphi(ax)\varphi^2(x)dx = \frac{1}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} x\exp\left\{-x^2\left(a^2+2\right)\right\}dx = 0.$$

It means that

$$I(a) = 1/(4\sqrt{\pi})$$

and, in particular,

$$\int_{-\infty}^{\infty} \Phi(ax)\varphi^2(x)dx = I(1) = 1/(4\sqrt{\pi}).$$

Finally, we have that

$$EX_{3,3} = 6\,I(1) = \frac{3}{2\sqrt{\pi}}.$$

Due to the symmetry of N(0,1) distribution, one obtains also that

$$E\big(X_{1,3}\big)^2 = E\big(X_{3,3}\big)^2$$

and

$$VarX_{1,3} = Var\ X_{3,3}.$$

Now

$$\begin{aligned}
E(X_{3,3})^2 &= 3\int_{-\infty}^{\infty} x^2\Phi^2(x)\varphi(x)dx = -3\int_{-\infty}^{\infty} x\Phi^2(x)d\varphi(x)\\
&= 3\int_{-\infty}^{\infty} \varphi(x)d(x\Phi^2(x)) = 3\int_{-\infty}^{\infty} \varphi(x)\Phi^2(x)dx + 6\int_{-\infty}^{\infty} x\varphi^2(x)\Phi(x)dx\\
&= \int_{-\infty}^{\infty} d(\Phi^3(x)) + \frac{3}{\pi}\int_{-\infty}^{\infty} x\exp(-x^2)\Phi(x)dx = 1-\frac{3}{2\pi}\int_{-\infty}^{\infty} \Phi(x)d(\exp(-x^2))\\
&= 1+\frac{3}{2\pi}\int_{-\infty}^{\infty} \exp(-x^2)\varphi(x)dx = 1+\frac{3}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} \exp(-3x^2/2)dx = 1+\frac{\sqrt{3}}{2\pi}.
\end{aligned}$$

Taking into account that

$$EX_{3,3} = \frac{3}{2\sqrt{\pi}},$$

we obtain that

$$Var\big(X_{3,3}\big) = E\big(X_{3,3}\big)^2 - \big(EX_{3,3}\big)^2 = 1+\frac{\sqrt{3}}{2\pi}-\frac{9}{4\pi}.$$

Further,

$$\begin{aligned}
E\big(X_{2,3}\big)^2 &= 6\int_{-\infty}^{\infty} x^2\Phi(x)(1-\Phi(x))\varphi(x)dx\\
&= 6\int_{-\infty}^{\infty} x^2\Phi(x)\varphi(x)dx - 6\int_{-\infty}^{\infty} x^2\Phi^2(x)\varphi(x)dx\\
&= 6\int_{-\infty}^{\infty} x^2\Phi(x)\varphi(x)dx - 2E\big(X_{3,3}\big)^2.
\end{aligned}$$

Now we obtain that

$$\begin{aligned}\int_{-\infty}^{\infty} x^2\Phi(x)\varphi(x)dx &= -\int_{-\infty}^{\infty} x\Phi(x)d\varphi(x)\\ &= \int_{-\infty}^{\infty} \varphi(x)d(x\Phi(x)) = \int_{-\infty}^{\infty} \varphi(x)\Phi(x)dx + \int_{-\infty}^{\infty} x\varphi^2(x)dx\\ &= \frac{1}{2}\int_{-\infty}^{\infty} d\left(\Phi^2(x)\right) + \frac{1}{2\pi}\int_{-\infty}^{\infty} x\exp\left(-x^2\right)dx = \frac{1}{2}.\end{aligned}$$

Hence,

$$E\left(X_{2,3}\right)^2 = 3 - 2E\left(X_{3,3}\right)^2 = 3 - 2\left(1 + \frac{\sqrt{3}}{2\pi}\right) = 1 - \frac{\sqrt{3}}{\pi}$$

and

$$Var\left(X_{2,3}\right) = E\left(X_{2,3}\right)^2 - \left(EX_{2,3}\right)^2 = 1 - \frac{\sqrt{3}}{\pi},$$

so far as $EX_{2,3} = 0$. Thus,

$$\begin{aligned}E\left(X_{1,3}\right)^2 &= E\left(X_{3,3}\right)^2 = 1 + \frac{\sqrt{3}}{2\pi},\\ E\left(X_{2,3}\right)^2 &= 1 - \frac{\sqrt{3}}{\pi},\\ Var\left(X_{1,3}\right) &= Var\left(X_{3,3}\right) = 1 + \frac{\sqrt{3}}{2\pi} - \frac{9}{4\pi}\end{aligned}$$

and

$$Var\left(X_{2,3}\right) = E\left(X_{2,3}\right)^2 - \left(EX_{2,3}\right)^2 = 1 - \frac{\sqrt{3}}{\pi}.$$

Exercise 2.4.1 (*solution*) It is evident that

$$\begin{aligned}P(M(n) - \log n < x) &= (1 - \exp(-x - \log n))^n = (1 - \exp(-x)/n)^n\\ &\to \exp(-\exp(-x)),\end{aligned}$$

as $n \to \infty$.

Exercise 2.4.2 (*solution*) In this case

$$\begin{aligned} P\{n^{1/\alpha}M(n) < x\} &= (1 - (-x/\ n^{1/\alpha})^{\alpha})^{n} = (1 - (-x)^{\alpha}/n)^{n} \to \exp(-(-x)^{\alpha}) \\ &= H_{2,\alpha}(x), \end{aligned}$$

as $n \to \infty$, for any x, $-\infty < x \le 0$.

Exercise 2.4.3 (*solution*) One can see that

$$P\{M(n)/n^{1/\alpha} < x\} = (1 - (x\,n^{1/\alpha})^{-\alpha})^{n} = (1 - x^{-\alpha}/n)^{n} \to \exp\{-x^{-\alpha}\},$$

as $n \to \infty$, for any $x \ge 0$.

Exercise 2.4.4 (*hint and answer*) In this case

$$(1 - F(x)) \sim 1/\pi x \text{ and } F'(x) \sim 1/\pi x^2, \text{ as } x \to \infty.$$

Use the statement of Theorem 2.4.1 to see that $\alpha = 1$, centering constants $a_n = 0$, and normalizing constants b_n can be taken from the equality $F(b_n) = 1 - 1/n$, that is possible to take $b_n = n/\pi$.

Exercise 2.4.5 (*hint*) In this situation it is easy to show that

$$\begin{aligned} 1 - F(x) \sim F'(x)/x = \frac{1}{x\sqrt{2\pi}} \exp\left(-x^2/2\right), \\ F''(x) \sim -xF'(x),\ x \to \infty, \end{aligned}$$

and then to use the statement of Theorem 2.4.3. To find constants a_n one needs to solve the equation

$$1/n = 1 - F(a_n) \sim \frac{1}{a_n\sqrt{2\pi}} \exp\left(-(a_n)^2/2\right)$$

or simply to check that the sequence

$$a_n = (2\log n - \log\log n - \log\ 4\pi)^{1/2}$$

satisfies this equation. Theorem 2.4.3 recommends to take $b_n \sim h(a_n)$, where

$$h(x) = 1 - F(x)/F'(x).$$

In this case one can obtain that $b_n \sim (2\log n)^{1/2}$ suits us.

Exercise 2.4.6 (*answer*) It follows from Theorem 2.4.2 that in this case

$$F \in D(H_1),\ a_n = \log n, b_n = 1.$$

Exercise 2.5.1 (*answers*) Here

$$ER(k) = (n+1)/2, Var\, R(k) = (n-1)^2/12,\ 1 \le k \le n,$$
$$Cov(R(k), R(m)) = -(n-1)/12,\ 1 \le k < m \le n,$$

and

$$\rho(R(k), R(m)) = -1/(n-1),\ 1 \le k < m \le n.$$

Exercise 2.5.2 (*answers*) $P\{\Delta(1) = 1,\ R(1) = 1\} = 1/n$,

$$P\{\Delta(1) = 1, R(1) = s\} = P\{\Delta(1) = s, R(1) = 1\} = 0,\ if\ s \neq 1,$$

and

$$P\{\Delta(1) = m, R(1) = s\} = 1/n(n-1), if\ s \neq 1,\ m \neq 1.$$

Exercise 2.5.3 (*solution*) Since

$$P\{\rho(m) = k\} = 1/m, k = 1, 2, \ldots m,$$

it is enough to show that for any $n = 1, 2, \ldots$, and any a(k), taking on values $1, 2, \ldots, k$, $1 \le k \le n$,

$$P\{\rho(1) = a(1), \rho(2) = a(2), \ldots, \rho(n) = a(n)\} = 1/n!.$$

Fix n and consider ranks $R(1), R(2), \ldots, R(n)$. It is not difficult to see that a set $\{a(1), a(2), \ldots, a(n)\}$ uniquely determines values $r(1), r(2), \ldots, r(n)$ of $R(1), R(2), \ldots, R(n)$. In fact, $r(n) = a(n)$. Further,

$$r(n-1) = a(n-1), \text{if a(n)} > \text{a(n} - 1), \text{ and } r(n-1) = a(n-1) + 1, \text{if } a(\text{n}) \le a(\text{n} - 1).$$

The value of $R(n-2)$ is analogously determined by values $a(n)$, $a(n-1)$ and $a(n-2)$ and so on. Hence, each of $n!$ events

$$\{\rho(1) = a(1), \rho(2) = a(2), \ldots, \rho(n) = a(n)\}$$

coincides with one of $n!$ events

$$\{R(1) = r(1), R(2) = r(2), \ldots, R(n) = r(n)\}.$$

For instance,

$$\{\rho(1) = 1, \rho(2) = 1, \ldots, \rho(n) = 1\} = \{R(1) = n,\ R(2) = n-1, \ldots,\ R(n) = 1\}.$$

Since

$$P\{R(1) = r(1),\ R(2) = r(2), \ldots,\ R(n) = r(n)\} = 1/n!$$

for any permutation $(r(1),\ r(2), \ldots,\ r(n))$ of the numbers 1, 2,..., n, one gets that

$$P\{\rho(1) = a(1), \rho(2) = a(2), \ldots, \rho(n) = a(n)\} = 1/n!$$

for any set $\{a(1),\ a(2), \ldots,\ a(n)\}$, where $1 \le a(k) \le k$, $k = 1, 2, \ldots, n$.



http://www.springer.com/978-94-6239-135-2

Records via Probability Theory
Ahsanullah, M.; Nevzorov, V.B.
2015, XII, 255 p. 10 illus. in color., Hardcover
ISBN: 978-94-6239-135-2
A product of Atlantis Press