ПОРЯДКОВЫЕ СТАТИСТИКИ

ORDER STATISTICS

(электронная версия спецкурса В.Б. Невзорова)

Оглавление

Introduction Введение 3

Chapter 1 . Basic definitions Основные определения 5

Chapter 2. Distributions of order statistics Распределения порядковых статистик 18

Chapter 3. Sample quantiles and ranges Выборочные квантили и ранги 27

Chapter 4. Representations for order statistics Представления для порядковых статистик 41

Chapter 5. Conditional distributions of order statistics Условные распределения порядковых статистик 57

Chapter 6. Order statistics for discrete distributions Порядковые статистики для дискретных распределений 67

Chapter 7. Moments of order statistics: general relations Моменты порядковых статистик: общие соотношения 81

Chapter 8. Moments of uniform and exponential order statistics Моменты равномерных и экспоненциальных порядковых статистик 90

Chapter 9. Moment relations for order statistics: normal distribution Моментные соотнощения для порядковых статистик: нормальное распределение 105

Chapter 10. Asymptotic behavior of the middle and intermediate order statistics Асимптотическое поведение средних и промежуточных порядковых статистик 116

Chapter 11. Asymptotic behavior of the extreme order statistics Асимптотическое поведение экстремальных порядковых статистик 126

REFERENCES Библиография 140

Introduction
Введение

Вашему вниманию предлагается материал, посвященный изучению широко используемых в различных теоретических и прикладных задачах математической статистики ***ПОРЯДКОВЫХ СТАТИСТИК*** (элементов вариационных рядов , построенных по исходным выборкам). Приведенные ниже 11 глав представляют собой первую часть курса “Порядковые статистики и рекордные величины”. Рекордам (они тесно связаны с порядковыми статистиками) будет посвящена вторая часть.

Обычно в математической статистике мы имеем дело с набором из n независимых одинаково распределенных случайных величин $X_1,X_2,\ldots,X_n$, представляющих собой выборку объема n из генеральной совокупности, характеризуемой некоторой известной или неизвестной нам функцией распределения F(x). Множество $\{x_1, x_2,\ldots,x_n\}$ наблюдаемых значений этих X-ов представляет собой реализацию данной выборки. Эти наблюдаемые значения позволяют исследовать различные свойства элементов генеральной совокупности, оценивать неизвестные параметры, проверять статистические гипотезы. Можно рассмотреть ситуацию, когда проверяются некоторые гипотезы о каких-то характеристиках (скажем, долговечности) некоторой продукции. Из данной большой партии случайно выбираются ***n*** элементов (скажем, некоторых деталей), выставляемых на испытательный стенд. По мере выхода этих деталей из строя фиксируются их времена жизни, образующие некоторую монотонную последовательность наблюдений

$x_{1,n} \le x_{2,n} \le \ldots \le x_{n-1,n} \le x_{n,n}$. Отметим, что эти величины представляют собой наблюдаемые значения уже потерявших свойство независимости элементов вариационного ряда $X_{1,n}$, $X_{2,n}$, …, $X_{n-1,n}$, $X_{n,n}$. Естественно, что делать статистические выводы по наблюдениям, соответствующим исходным независимым $X_1,X_2,\ldots,X_n$, существенно проще, чем по наблюдаемым значениям заведомо зависимых порядковых статистик $X_{1,n}$, $X_{2,n}$, …, $X_{n-1,n}$, $X_{n,n}$, но часто по ряду причин (например, нет времени дождаться, когда последние из находящихся на испытательном стенде деталей выйдут из строя) приходится иметь дело не с полным набором порядковых статистик, а лишь с наблюдениями, представляющими некоторый отрезок вариационного ряда. Поэтому исследования распределений, характеристик и свойств порядковых статистик вызывают несомненный интерес, что подтверждается, например, большим числом публикаций, приведенных в списке библиографии.

Представленный ниже материал предназначен, в первую очередь, для студентов-математиков, обучающихся на кафедре теории вероятностей и математической статистики. Многие из этих студентов выбирают темы курсовых и выпускных работ, связанных с порядковыми статистиками и некоторыми другими типами упорядоченных случайных величин. Из предлагаемого ниже списка литературы, связанной с порядковыми статистиками и другими вариантами таких величин (например, рекордными величинами), видно, что подавляющее большинство публикаций по данной тематике представлено на английском языке. Предлагается, чтобы нашим студентам (тем, в первую очередь, кто пишет курсовые и дипломные работы, посвященные различным упорядоченным случайным величинам) было легче разбираться с такого рода статьями, следующий подход к построению данного сайта. Материал приводится на английском языке с необходимыми пояснениями на русском. По ходу изучения свойств порядковых статистик читатель может проверить свои знания, решая приводимые в тексте задачи. Решения (или ответы) этих задач приводятся в конце каждого параграфа.

Welcome to study ORDER STATISTICS !

Chapter 1 . Basic definitions
Основные определения

random variables (r. v.'s) = случайные величины

distribution function (d. f.) = функция распределения

sample of size n = выборка объема n

variational series = вариационный ряд

order statistics = порядковые статистики

empirical (sample) distribution function = эмпирическая (выборочная) функция распределения

vector of ranks = вектор рангов,   antiranks = антиранги,

sequential ranks= последовательные ранги

==

We introduce the basic definitions. They are as follows:

$X_1, X_2,\ldots, X_n$ – initial random variables .

As a rule in the sequel we will suppose that these random variables are independent and have a 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. We can simply say also that $X_1, X_2,\ldots, X_n$ present n independent observations on X, where X is a random variable, having a d. f. F.

$X_{1,n}\le X_{2,n}\le\ldots\le X_{n, n}$ denotes variational series based on random variables $X_1, X_2,\ldots, X_n$. If X's are independent and identically distributed one can say that it is the variational series based on a sample $X_1, X_2,\ldots, X_n$. Elements $X_{k,n}$, $1\le k\le n$, are called order statistics (order statistics based on a sample $X_1, X_2,\ldots, X_n$; order statistics from a d.f. F; ordered observations on X). 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 realizations 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,….

$F_n^*(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k \le x\}}$ denotes the empirical (or sample) distribution function.

Here $1_{\{X\le x\}}$ is a random indicator, which equals 1, if $X\le x$ and equals 0, if $X>x$.

Let us mention that

$F_n^*(x) =0$, if $x<X_{1,n}$, $F_n^*(x) =k/n$, if $X_{k,n}\le x<X_{k+1,n}$, $1\le k\le n-1$, and $F_n^*(x) =1$, if $x\ge X_{n,n}$.

(Заметим, что в отечественной русскоязычной литературе в качестве функций распределения обычно рассматриваются функции $F(x) = P\{X<x\}$- непрерывные слева. В англоязычной литературе чаще можно встретить варианты одномерных или многомерных непрерывных справа функций распределения, имеющих вид $F(x) = P\{X\leq x\}$ или $F(x_1,x_2,\ldots,x_n) = P\{X_1\leq x, X_2 \leq x_2,\ldots,X_n\leq x_n\}$).

Together with a random sample $X_1, X_2,\ldots, X_n$ we can consider a vector of ranks

$(R(1),R(2),\ldots,R(n))$, where

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

Ranks provide the following equalities for events:

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

Symmetrically antiranks $\Delta(1),\Delta(2),\ldots,\Delta(n)$ are defined by equalities

$$\{\Delta(k)=m\}=\{X_{k,n}=X_m\}, \quad m=1,2,\ldots,n, \ k=1,2,\ldots,n.$$

One more type of ranks is presented by sequential ranks. For any sequence of random variables $X_1, X_2,\ldots$ we introduce sequential ranks $\rho(1),\rho(2),\ldots$ as follows:

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

Sequential rank $\rho(m)$ shows a position of a new coming observation X_m among its predecessors $X_1, X_2,\ldots,X_{m-1.}$

Sometimes we need to investigate ordered random variables. Indeed, we always can order a sequence of values. For example, if $a_1=3$, $a_2=1$, $a_3=3$, $a_4=2$ and $a_5=8$, then ordered values can be presented as 1,2,3,3,8 (non-decreasing order) or 8,3,3,2,1 (non-increasing order).Let us have now some random variables $X_1, X_2,\ldots, X_n$ defined on a common probability space $(\Omega,\Im, P)$. Each random variable maps Ω into $\Re$, the real line. It means that for any elementary event $\omega\in\Omega$ we have n real values $X_1(\omega),X_2(\omega),\ldots,X_n(\omega)$, which can be arranged in nondecreasing order. It enables us to introduce new random variables $X_{1,n}= X_{1,n}(\omega)$, $X_{2,n}= X_{2,n}(\omega),\ldots,X_{n,n}= X_{n,n}(\omega)$ defined on the same probability space $(\Omega,\Im, P)$ as follows: for each $\omega\in\Omega$, $X_{1,n}(\omega)$ coincides with the smallest of the values $X_1(\omega),X_2(\omega),\ldots,X_n(\omega)$, $X_{2,n}(\omega)$ is the second smallest of these values, $X_{3,n}(\omega)$ is the third smallest,…, and $X_{n,n}(\omega)$ is assigned the largest of the values $X_1(\omega),X_2(\omega),\ldots,X_n(\omega)$. Thus, a set of n arbitrary random variables $X_1, X_2,\ldots, X_n$ generates another set of random variables $X_{1,n}, X_{2,n},\ldots,X_{n,n}$, already ordered. The used construction provides two important equalities:

$$P\{X_{1,n} \leq X_{2,n}\leq\ldots\leq X_{n,n}\}=1 \qquad (1.1)$$

and

$$X_{1,n}+X_{2,n}+\ldots+X_{n,n}= X_1+ X_2+\ldots+ X_n. \qquad (1.2)$$

Exercise 1.1. Let a set of elementary events Ω consist of two elements ω_1 and ω_2 and random variables X_1 and X_2 be defined as follows: $X_1(\omega_1)=0$, $X_1(\omega_2)=3$, $X_2(\omega_1)=1$, $X_2(\omega_2)=2$. Describe ordered random variables $X_{1,2}$ and $X_{2,2}$ as functions on Ω.

Exercise1.2. Let now $\Omega=[0,1]$ and three random variables X_1, X_2 and X_3 are defined as follows: $X_1(\omega)=\omega$,$X_2(\omega)=1-\omega$, $X_3(\omega)=1/4$, $\omega\in[0,1]$. What is the structure of functions $X_{k,3}(\omega)$, $k=1,2,3$?

Definition1.1. We say that

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

is the variational series based on random variables $X_1, X_2,\ldots, X_n$. Elements $X_{k,n}$, $k=1,2,\ldots,n$, of variational series are said to be order statistics.

Very often in mathematical statistics we deal with sequences of independent random variables having a common d.f. F. Then a collection $X_1, X_2,\ldots, X_n$ can be interpreted as a random sample of size n from the d.f. F. We can say also that $X_1, X_2,\ldots, X_n$ present n independent observations on X, where X is a random variable, having a d.f. F. Hence in this situation we deal with the variational series and order statistics based on a sample $X_1, X_2,\ldots, X_n$. We can also say that $X_{k,n}$, $1\le k\le n$, are order statistics from a d.f. F; or, for example, ordered observations on X. As a result of a statistical experiment we get a set of the observed values $\{x_1,x_2,\ldots,x_n\}$ of random variables $X_1,X_2,\ldots,X_n$. This set is called a realization of the sample. Analogously 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.

In the sequel we will consider, in general, sequences of independent random variables. Here we must distinguish two important situations, namely, the case of continuous

d. f.'s F and the case, when d. f.'s F have some discontinuity points.

The structure of order statistics essentially differs in these two situations. Let us try to show this difference.

Exercise1.3. Let X_1 and X_2 be independent random variables with continuous d. f.' s F_1 and F_2. Prove that $P\{ X_1= X_2\}=0$.

Exercise1.4. Let $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$

be the variational series based on independent random variables $X_1, X_2,\ldots, X_n$ with continuous

d. f.' s $F_1,F_2,\ldots,F_n$. Show that in this case

$$P\{X_{1,n} < X_{2,n} < \ldots < X_{n,n}\}=1.$$

Exercise1.5. Let X_1, X_2 ,…, X_n be independent random variables having the uniform distribution on the set {1,2,3,4,5,6}. This situation corresponds, for instance, to the case, when a die is rolled n times. Find

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

Exercise1.6. Let X_1, X_2,…, X_n be n independent observations on random variable X, having the geometric distribution, that is

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

and $X_{k, n}$ be the corresponding order statistics. Find

$$p_n= P\{ X_{1,n} <X_{2,n}<\ldots<X_{n, n}\},\ n=2,3,\ldots.$$

Unless otherwise is proposed, in the sequel we suppose that X's are independent random variables having a common continuous d.f. F. In this situation order statistics satisfy inequalities

$$X_{1,n}<X_{2,n}<\ldots<X_{n, n}$$

with probability one.

There are different types of rank statistics, which help us to investigate ordered random variables. Together with a sample X_1, X_2,…, X_n we will consider a vector of ranks

$$(R(1),R(2),\ldots,R(n)),$$

elements of which show the location of X's in the variational series

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

Definition 1.2. Random variables R(1),…, R(n) given by equalities

$$R(m)= \sum_{k=1}^{n} 1_{\{X_m \ge X_k\}}\ ,\ m=1,2,\ldots,n. \qquad (1.3)$$

are said to be ranks corresponding to the sample X_1, X_2,…, X_n.

Since we consider the situation, when different X's can coincide with zero probability, (1.3) can be rewritten in the following form:

$$R\ (m) =1+ \sum_{k=1}^{n} 1_{\{X_m > X_k\}}\ . \qquad (1.4)$$

The following equality for events is a corollary of (1.4):

$$\{R(m)=k\}=\{X_{k, n}=X_{m}\},\ m=1,2,\ldots,n,\ k=1,2,\ldots,n. \qquad (1.5)$$

Another form of (1.5) is

$$X_{\ m}=X_{R(m),n},\ m=1,2,\ldots,n. \qquad (1.6)$$

Exercise1.7. For any m=1,2,…,n, prove that R(m) has the discrete uniform distribution on set {1,2,…,n}.

We know that i.i.d. random variables X_1, X_2,…, X_n taken from a continuous distribution have no coincidences with probability one. Hence, realizations (r(1),…,r(n)) of the corresponding vector of ranks (R(1),R(2),…,R(n)) represent all permutations of values 1,2,…,n. Any realization (r(1),…,r(n)) corresponds to the event

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

where δ(r(k))=k. For example, the event

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

is equivalent to the event

$$\{X_n<X_{n-1}<\ldots<X_1\}.$$

Taking into account the symmetry of the sample x_1, x_2,…, $x_{\ n}$, we obtain that events

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

have the same probabilities for any permutations (δ(1),…,δ(n)) of numbers 1,2,..,n. Hence

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

$$P\{(X_{\delta(1)}<X_{\delta(2)}<\ldots< X_{\delta(n)})\}=1/n! \qquad (1.7)$$

for any permutation (r(1),…,r(n)) of numbers 1,2,…,n.

Exercise 1.8. For fixed n and k<n, find

$$P\{R(1) = r(1),R(2) = r(2),\ldots,R(k) = r(k)\},$$

where r(1),r(2),…, r(k) are different numbers taken from the set {1,2,…,n}.

Exercise1.9. Show that ranks R(1),R(2),…,R(n) are dependent random variables for any n=2,3,….

The dependence of ranks is also approved by the evident equality

$$R(1)+R(2)+\ldots+R(n)=1+2+\ldots+n=n(n+1)/2, \qquad (1.8)$$

which is valid with probability one always, when X's are independent and have a common continuous distribution

It follows from (1.7) that ranks are exchangeable random variables: for any permutation (α(1),α(2),…,α(n)) of numbers 1,2,…,n, vectors (R(α(1)),…,R(α(n))) and (R(1),…,R(n)) have the same distributions.

Exercise1.10. Find expectations and variances of R (k), $1\le k\le n$.

Exercise1.11. Find covariance Cov (R (k), R (m)) and correlation coefficients

ρ(R (k), R (m)) between R (k) and R(m), $1\le k, m\le n$.

Exercise 1.12. Find ER(1)R(2)…R(n-1) and ER(1)R(2)…R(n).

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

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

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),…,R(n)), $\delta(k)$ can take on different values from the set {1,2,…,n} and we really deal with new random variables, which realizations are $\delta(1), \delta(2),\ldots, \delta(n)$.

Definition 1.3. Let $X_1, X_2,\ldots, X_n$ 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. Random variables $\Delta(1)$, $\Delta(2),\ldots,\Delta(n)$, which satisfy the following equalities:

$$\{\Delta(m)=k\}=\{X_{m,n}=X_k\},\ m=1,2,\ldots,n,\ k=1,2,\ldots,n, \tag{1.9}$$

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,…,n) and all n! 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

$$\{\Delta(k)=m\}=\{X_{k,n}=X_m\}=\{R(m)=k\} \tag{1.10}$$

hold with probability one. We can write also the following identities for ranks and antiranks:

$$\Delta(R(m))=m \tag{1.11}$$

and

$$R(\Delta(m))=m, \tag{1.12}$$

which hold with probability one for any m=1,2,…,n.

Exercise 1.13. Find the joint distribution of $\Delta(1)$ and R(1).

While ranks and antiranks are associated with some random sample and its size n, there are rank statistics, which characterize a sequence of random variables $X_1, X_2, \ldots$.

Definition 1.4. Let X_1, X_2, … be independent random variables, having continuous (not necessary identical) distributions. Random variables ρ(1), ρ(2),… given by equalities:

$$\rho(m) = \sum_{k=1}^{m} 1\{X_m \geq X_k\} \text{ , m=1,2,….} \qquad (1.13)$$

are said to be *sequential ranks.*

Sequential ranks are not associated with some sample of a fixed size n. In fact,

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

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

In general, ρ(m)=k implies that

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

It is not difficult to see that ρ(m) takes on the values 1, 2,…,m. If independent random variables

X_1, X_2,…, X_m have the same continuous distribution then the standard arguments used above enable us to see that for any m=1,2,…,

$$P\{\rho(m)=k\}=P\{X_m=X_{k,m}\}=1/m, \text{ k=1,2,…,m.} \qquad (1.14)$$

Exercise1.14. Let X_1, X_2, … be independent random variables with a common continuous d. f. F. Prove that the corresponding sequential ranks ρ(1), ρ(2),… are independent.

In example 1.1 we will deal with different types of ranks.

Example 1.1. Let the following data represent the lifetimes (hours) of 15 batteries (realization of some sample of size 15):

x_1=20.3 x_2=17.2 x_3=15.4 x_4=16.8 x_5=24.1

x_6 =12.6 x_7=15.0 x_8=18.1 x_9=19.1 x_{10}=21.3

x_{11}=22.3 x_{12}=16.4 x_{13}=13.5 x_{14}=25.8 x_{15}=16.9

Being ordered these observations give us realizations of order statistics:

$x_{1,15}$=12.6 $x_{2,15}$=13.5 $x_{3,15}$=15.0 $x_{4,15}$=15.4 $x_{5,15}$=16.4

$x_{6,15}$=16.8 $x_{7,15}$=16.9 $x_{8,15}$=17.2 $x_{9,15}$=18.1 $x_{10,15}$=19.1

$x_{11,15}$=20.3 $x_{12,15}$=21.3 $x_{13,15}$=22.3 $x_{14,15}$=24.1 $x_{15,15}$=25.8

Realizations of ranks are given as follows:

r(1)=11, r(2)=8, r(3)= 4, r(4)=6, r(5)=14, r(6)=1, r(7)=3, r(8)=9,

r(9)=10, r(10)=12, r(11)=13, r(12)=5, r(13)=2, r(14)=15, r(15)=7

Antiranks are presented by the sequence:

δ(1) =6, δ(2) =13, δ(3) =7, δ(4) =3, δ(5) = 12, δ(6) =4, δ(7) =15, δ(8) =2,

δ(9) =8, δ(10) =9, δ(11) =1, δ(12) =10, δ(13) =11, δ(14) =5, δ(15) =14

The sequential ranks are:

1; 1; 1; 2; 5; 1; 2; 6; 7; 9; 10; 4; 2; 14; 7 .

Very often to estimate an unknown population distribution function F a statistician uses the so-called empirical (sample) distribution function

$$F_n^*(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k \le x\}} . \qquad (1.15)$$

Empirical distribution functions are closely tied with order statistics so far as

$$F_n^*(x) = 0, \text{ if } x < X_{1,n}, \; F_n^*(x) = 1, \text{ if } x \ge X_{n,n},$$

and

$$F_n^*(x) = k/n, \text{ if } X_{k,n} \le x < X_{k+1,n}, \; 1 \le k \le n-1. \qquad (1.16)$$

Exercise1.15. Find the expectation and the variance of $F_n^*(x)$.

Check your solutions

Вернитесь к предложенным выше задачам. Сделайте попытку решить их, а после этого посмотрите представленные ниже варианты решений или убедитесь, что полученные самостоятельно ответы совпадают с правильными.

***Exercise 1.1* (solution).** On comparing values of X_1 and X_2 one can see that

$$X_{1,2}(\omega_1)=0 , X_{1,2}(\omega_2)=2 \text{ and } X_{2,2}(\omega_1)=1 , X_{2,2}(\omega_2)=3.$$

Thus, $X_{1,2}$ partially coincides with X_1 (on ω_1) and partially with X_2 (on ω_2) as well as $X_{2,2}$, which coincides with X_1 (on ω_2) and X_2 (on ω_2).

Exercise 1.2 **(answers).** $X_{1,3}(\omega)=X_1(\omega)=\omega$, if $0\leq\omega\leq 1/4$; $X_{1,3}(\omega)=X_3(\omega)=1/4$, if $1/4<x<3/4$, $X_{1,3}(\omega)=X_2(\omega)$, if $3/4\leq x\leq 1$;

$X_{2,3}(\omega)=X_3(\omega)=1/4$, if $0\leq\omega\leq 1/4$ and if $3/4 \leq\omega\leq 1$, $X_{2,3}(\omega)=X_1(\omega)=\omega$, if $1/4<\omega\leq 1/2$,

$X_{2,3}(\omega) = X_3(\omega)=\omega$, if $1/2<\omega<3/4$;

$X_{3,3}(\omega)=X_3(\omega)=1-\omega$, if $0\leq\omega\leq 1/2$, $X_{3,3}(\omega)=X_1(\omega)=\omega$, if $1/2<x\leq 1$.

Exercise 1.3 **(solution)**. Since X_1 has a continuous d. f. we have for any x that

$$P\{X_1=x\}=F_1(x)-F_1(x-0)=0.$$

Then, taking into account the independence of for any X_1 and X_2 we obtain that

$$P\{ X_1= X_2\}= \int_{-\infty}^{\infty} P\{ X_1= X_2| X_2=x\}dF_2(x)= \int_{-\infty}^{\infty} P\{ X_1=x| X_2=x\}dF_2(x)$$

$$= \int_{-\infty}^{\infty} P\{ X_1=x\}dF_2(x)=0.$$

Exercise 1.4 **(solution).** Let A be an event such that there are at least two coincidences among $X_1, X_2,\ldots, X_n$ and

$$A_{ij}= \{X_i=X_j\}.$$

We know from Exercise 1.3 that $P\{A_{ij}\}=0$ if $i\neq j$.

The assertion of exercise 1.4 holds since

$$1-P\{ X_{1,n} <X_{2,n}<\ldots<X_{n,n}\}=P\{A\}\leq\sum_{i\neq j} P\{A_{ij}\}=0.$$

Exercise 1.5 **(solution)**. It is evident that $p_1=1$ and $p_n=0$, if $n>6$. One can see that p_n is equal to the probability that the die shows n different values. Hence

$$p_n=6!/(6-n)!6^n, n=2,\ldots,6,$$

and, in particular, $p_2=5/6$, $p_3=5/9$, $p_4=5/18$, $p_5=5/54$ and $p_6=5/324$.

Exercise 1.6 **(solution)**. In this case

$$p_n= n!P\{ X_1< X_2<\ldots< X_n\}=$$

$$n! \sum_{k_1=0}^{\infty}(1-p)p^{k_1} \sum_{k_2=k_1+1}^{\infty}(1-p)p^{k_2} \ldots \sum_{k_n=k_{n-1}+1}^{\infty}(1-p)p^{k_n} .$$

Sequential simplifying of this series gives us the following expression for p_n:

$$p_n=n!(1-p)^n p^{n(n-1)/2}/\prod_{k=1}^{n}(1-p^k)=n!\, p^{n(n-1)/2}/(1+p)(1+p+p^2)\ldots(1+p+\ldots+p^{n-1}).$$

In particular, $p_2=2p/(1+p)$ and $p_3=6p^3/(1+p)(1+p+p^2)$.

***Exercise 1.7* (solution)**.

a) The symmetry argument enables us to prove the necessary statement for one rank only, say for R(1). From (1.4), on using the total probability rule and independence of X's, one has that

$$P\{R(1)=k\}=P\{\sum_{s=2}^{n}1_{\{X_1>X_s\}}=k-1\} =$$

$$\int_{-\infty}^{\infty} P\{\sum_{s=2}^{n}1\{X_1>X_s\}=k-1|X_1=x\}dF(x)=\int_{-\infty}^{\infty} P\{\sum_{s=2}^{n}1\{X_s<x\}=k-1|X_1=x\}dF(x)$$

$$=\int_{-\infty}^{\infty} P\{\sum_{s=2}^{n}1\{X_s<x\}=k-1\}dF(x).$$

Indeed, the sum under probability sign has the binomial distribution with parameters n-1 and F(x) and hence we finally have that

$$P\{\sum_{s=2}^{n}1\{X_s<x\}=k-1\}=\binom{n-1}{k-1}(F(x))^{k-1}(1-F(x))^{n-k}$$

and

$$\int_{-\infty}^{\infty} P\{\sum_{s=2}^{n}1\{X_s<x\}=k-1\}d\,F(x)=\binom{n-1}{k-1}\int_{-\infty}^{\infty}(F(x))^{k-1}(1-F(x))^{n-k}d\,F(x)$$

$$=\binom{n-1}{k-1}\int_{0}^{1}u^{k-1}(1-u)^{n-k}d\,u=\binom{n-1}{k-1}B(k,n-k+1)=\binom{n-1}{k-1}(k-1)!(n-k)!/n!=1/n.$$

Above we used the following formula for beta function:

$$B\,(m, n) = \int_0^1 u^{m-1}(1-u)^{n-1}du=(m-1)!(n-1)!/(m+n-1)!.$$

b) A less rigorous proof also is based on symmetry argument. We can note that for any k=1,2,…,n, events $\{X_{k.n}=X_1\}$, $\{X_{k.n}=X_2\}$,…,$\{X_{k.n}=X_n\}$ must have equal probabilities. Their sum equals one. Hence $P\{X_{k,n}=X_m\}=1/n$ for any $1\le m, k\le n$.

Exercise 1.8 **(solution).** The event {R (1)=r(1),…,R(k)=r(k)} is the union of (n-k)! events

$$\{R(1)=r(1),\dots,R(k)=r(k),\ R(k+1)=s(1),\dots,R(n)=s(n-k)\},$$

where (s(1),s(2),…,s(n-k)) are permutations of (n-k) numbers taken from the set

$$\{1, 2,\dots,n\}\backslash\{r(1),r(2),\dots,r(k)\}.$$

Since each of events {R(1)=r(1),…,R(k)=r(k), R(k+1)=s(1),…,R(n)=s(n-k)}

has probability 1/n!, we get that

$$P\{R(1)=r(1),\dots,R(k)=r(k)\}=(n-k)!/n!.$$

Exercise 1.9 **(solution).** On comparing equalities

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

and

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

we obtain that

$$1/n!= P\{R(1)=r(1),R(2)=r(2),\dots,R(n)=r(n)\} \neq P\{R(1)=r(1)\}\dots P\{R(n)=r(n)\}=1/n^n.$$

It means that ranks R(1), R(2),…,R(n) are dependent.

Exercise 1.10 **(answers)** . $ER(k)=(n+1)/2$, $\mathrm{Var}\,R(k)=(n-1)^2/12$, $1\le k\le n$.

Exercise 1.11 **(solution)**. Indeed,

$$\mathrm{Cov}(R(k),R(k))=\mathrm{Var}\,R(k)=(n-1)^2/12,\ k=1,2,\dots,n,$$

as it follows from Exercise 1.11, and $\rho(R(k),R(k))=1$ for any k. Exchangeability of ranks implies that

$$\mathrm{Cov}(R(k),R(m))=\mathrm{cov}(R(1),R(2))$$

and

$$\rho(R(k),R(m))= \rho(R(1),R(2))$$

for any $1\le k\neq m\le n$. Then, we obtain from (1.8) that

$$0=\mathrm{Var}\ (R(1)+R(2)+\ldots+R(n))$$

$$=\sum_{k=1}^{n}\ \mathrm{Var}\ R(k)\ +2\sum_{1\le k<m\le n}\ \mathrm{Cov}(R(k),R(m))$$

$$=\mathrm{nVar}\ R(1)+2\binom{n}{2}\mathrm{Cov}(R(1),R(2)).$$

Hence,

$$\rho(R(1),R(2))=\mathrm{cov}(R(1),R(2))/(\mathrm{Var}\ R(1))^{1/2}(\mathrm{Var}R(2))^{1/2}$$

$$=\mathrm{cov}(R(1),R(2))/(\mathrm{Var}\ R(1))=-n/2\binom{n}{2}=-1/(n-1)$$

and

$$\mathrm{cov}(R(1),R(2))=-n\mathrm{Var}\ R(1)/n(n-1)=\ -(n+1)/12.$$

***Exercise 1.12* (solution)**. Since (R(1),R(2),…,R(n)) is a random permutation of numbers 1,2,…,n, we have that R(1)R(2)…R(n) = n! with probability one. It implies that

$$E\ R\ (1)\ R\ (2)\ldots R(n\ )=n!.$$

Analogously we can write that

$$E\ R(1)\ R(2)\ldots R(n-1)\ =E(n!/R(n)) = n!\ E\ (1/R\ (n)).$$

Since R (n) takes on the values 1,2,…,n with probabilities 1/n, we have now that

$$E\ (1/R\ (n)) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{k}$$

and

$$E\ R(1)R(2)\ldots R(n-1)=(n-1)!\ \sum_{k=1}^{n}\frac{1}{k}\ .$$

***Exercise 1.13* (answers)**. p(m ,s) = P{Δ(1)=m, R(1)=s}=1/n, if m=s=1;

p(m,s)=1/n(n-1), if s≠1, m≠1, and p(m, s)=0 otherwise.

***Exercise 1.14* (solution).** Since P{ρ(m)=k}=1/m, k=1,2,…m, it needs to show that

for any n=1,2,…, and any *a*(k), taking on values 1,2,…,k, 1≤k≤n,

P {ρ(1) =*a*(1), ρ(2)=*a*(2),…, ρ(n)=*a*(n)}=

P{ρ(1)=*a*(1)}P{ρ(2)=*a*(2)}…P{ρ(n)=*a*(n)}=1/n!.

Fix n and consider ranks R(1),R(2),…,R(n). It is not difficult to see that a set

{a(1), a(2),…, a(n)} uniquely determines values r(1),r(2),…, r(n) of R(1),R(2),…,R(n). In fact, r(n)= a(n). Further, r(n-1)=a(n-1), if a(n) > a(n-1), and r(n-1)= a(n-1)+1, if a(n) ≤ a(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

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

coincides with one of n! events

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

For instance,

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

Since

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

for any permutation (r(1),r(2),…,r(n)) of the numbers 1,2,…,n, we have that

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

for any set {a(1), a(2),…,a(n)}, where 1≤ a(k) ≤ k, k=1,2,…,n.

Exercise 1.15 **(answers)**. E F_n^*(x) =F(x), Var F_n^*(x) =F(x)(1-F(x))/n.

Chapter 2. Distributions of order statistics
Распределения порядковых статистик

incomplete beta function = неполная бета- функция

joint distribution function (joint d.f.) = совместная функция распределения

joint probability density function (joint p.d.f.) = совместная плотность распределения

We give some some important formulae for distributions of order statistics.

Let us begin from simple formulae for distributions of maxima and minima.

Example 2.1. Let random variables $X_1, X_2, \dots, X_n$ have a joint d.f.

$$H(x_1, x_2, \dots, x_n) = P\{X_1 \le x_1, X_2 \le x_2, \dots, X_n \le x_n\}.$$

Then d.f. of

$$M(n) = \max\{X_1, X_2, \dots, X_n\}$$

has the form

$$P\{M(n) \le x\} = P\{X_1 \le x, X_2 \le x, \dots, X_n \le x\} = H(x, x, \dots, x). \qquad (2.1)$$

Similarly we can get the distribution of

$$m(n) = \min\{X_1, X_2, \dots, X_n\}.$$

One has

$$P\{m(n) \le x\} = 1 - P\{m(n) > x\} = 1 - P\{X_1 > x, X_2 > x, \dots, X_n > x\}. \qquad (2.2)$$

Exercise 2.1. Find the joint distribution function of M(n-1) and M(n).

Exercise 2.2. Express d.f. of $Y=\min\{X_1,X_2\}$ in terms of joint d.f.

$$H(x_1,x_2)=P\{X_1\le x_1, X_2 \le x_2\}.$$

Exercise 2.3. Let $H(x_1,x_2)$ be the joint d.f. of X_1 and X_2. Find the joint d.f. of

$$Y=\min\{X_1,X_2\} \text{ and } Z=\max\{X_1,X_2\}.$$

From (2.1) and (2.2) one obtains the following elementary expressions for the case, when X_1, $X_2,\dots, X_n$ present a sample from a population d.f. F (not necessary continuous) :

$$P\{M(n) \le x\}=F^n(x) \qquad (2.3)$$

and

$$P\{m(n) \le x\}=1-(1-F(x))^n. \qquad (2.4)$$

Exercise 2.4. Let $X_1, X_2,\dots, X_n$ be a sample of size n from a geometrically distributed random variable X, such that

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

Find

$$P\{Y\ge r, Z<s\},\ r<s,$$

where

$$Y=\min\{ X_1, X_2,\dots, X_n\}$$

and

$$Z=\max\{ X_1, X_2,\dots, X_n\}.$$

There is no difficulty to obtain d.f.'s for single order statistics $X_{k,n}$. Let

$$F_{k:n}(x) = P\{X_{k,n}\le x\}.$$

One can see immediately from (2.3) and (2.4) that

$$F_{n:n}(x)=F^n(x)$$

and

$$F_{1:n}(x)=1-(1-F(x))^n.$$

The general formula for $F_{k:n}(x)$ is not much more complicated. In fact,

$$F_{k:n}(x) = P\{\text{at least } k \text{ variables among } X_1, X_2,\dots, X_n \text{ are less or equal x}\} =$$

$$\sum_{m=k}^{n} P\{\text{exactly m variables among } X_1, X_2,\dots, X_n \text{ are less or equal x}\}=$$

$$\sum_{m=k}^{n} \binom{n}{m} (F(x))^m(1-F(x))^{n-m},\ 1\le k\le n. \tag{2.5}$$

Exercise 2.5. Prove that identity

$$\sum_{m=k}^{n} \binom{n}{m} x^m(1-x)^{n-m} = I_x\,(k,n-k+1) \tag{2.6}$$

holds for any $0\le x\le 1$, where

$$I_x\,(a,b)=\frac{1}{B(a,b)}\int_0^x t^{a-1}(1-t)^{b-1}dt \tag{2.7}$$

is the incomplete beta function with parameters *a* and *b*, *B(a,b)* being the classical beta function. By comparing (2.5) and (2.6) one obtains that

$$F_{k:n}(x)= I_{F(x)}\,(k,n-k+1). \tag{2.8}$$

Remark 2.1. It follows from (2.8) that $X_{k,n}$ has the beta distribution with parameters k and n-k+1, if X has the uniform on [0,1] distribution.

Remark 2.2. Equality (2.8) is valid for any distribution function F.

Remark 2.3. If one has any table of the function $I_x\,(k,n-k+1)$, it is possible to obtain d.f. $F_{k:n}(x)$ for arbitrary d.f. F.

Exercise 2.6. Find the joint distribution of two order statistics $X_{r,n}$ and $X_{s,n.}$

Example 2.2. Let us try to find the joint distribution of all elements of the variational series $X_{1,n}, X_{2,n},\dots,X_{n,n}$. It seems that the joint d.f.

$$F_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n) = P\{X_{1,n}\le x_1,\ X_{2,n}\le x_2,\dots,\ X_{n,n}\le x_n\}$$

promises to be very complicated. Hence, we consider probabilities

$$P(y_1,x_1,y_2,x_2,\dots,y_n,x_n) = P\{y_1<X_{1,n}\le x_1,\ y_2< X_{2,n}\le x_2,\dots,\ y_n< X_{n,n}\le x_n\}$$

for any values

$$-\infty \le y_1<x_1 \le y_2<x_2 \le\ldots \le y_n<x_n \le \infty.$$

It is evident, that the event

$$A=\{y_1<X_{1,n} \le x_1,\ y_2< X_{2,n} \le x_2,\ldots,\ y_n< X_{n,n} \le x_n\}$$

is a union of n! disjoint events

$$A(\alpha(1),\alpha(2),\ldots,\alpha(n)) = \{y_1<X_{\alpha(1)}\le x_1,\ y_2< X_{\alpha(2)}\le x_2,\ldots,\ y_n< X_{\alpha(n)} \le x_n\},$$

where the vector $(\alpha(1),\alpha(2),\ldots,\alpha(n))$ runs all permutations of numbers 1,2,…,n. The symmetry argument shows that all events $A(\alpha(1),\alpha(2),\ldots,\alpha(n))$ have the same probability. Note that this probability is equal to

$$\prod_{k=1}^{n} (F(x_k)-F(y_k)).$$

Finally, we obtain that

$$P\{y_1<X_{1,n} \le x_1,\ y_2< X_{2,n} \le x_2,\ldots,\ y_n< X_{n,n} \le x_n\}=n!\prod_{k=1}^{n} (F(x_k)-F(y_k)). \tag{2.9}$$

Let us now consider the case when our population distribution has a density function *f*. It means that for almost all x (i.e., except, possibly, a set of zero Lebesque measure) $F'(x)=f(x)$. In this situation (2.9) enables us to find an expression for the joint probability density function (p.d.f.) (denote it

$f_{1,2,\ldots,n:n}(x_1, x_2,\ldots,x_n)$) of order statistics $X_{1,n}, X_{2,n},\ldots, X_{n,n}$. In fact, differentiating both sides of (2.9) with respect to $x_1,x_2,\ldots,x_n$, we get the important equality

$$f_{1,2,\ldots,n:n}(x_1, x_2,\ldots,x_n)=n!\prod_{k=1}^{n} f(x_k),\ -\infty<x_1<x_2<\ldots<x_n <\infty. \tag{2.10}$$

Otherwise (if inequalities $x_1< x_2<\ldots< x_n$ fail) we naturally put

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

Taking (2.10) as a starting point one can obtain different results for joint distributions of arbitrary sets of order statistics.

Example 2.3. Let $f_{m:n}$ denote the p.d.f. of $X_{m,n}$. We get from (2.10) that

$$f_{m:n}(x)=\int \ldots\int f_{1,2,\ldots,n:n}(x_1,\ldots,x_{k-1},x,x_{k+1},\ldots,x_n)dx_1\ldots dx_{k-1}dx_{k+1}\ldots dx_n =$$

$$\text{n!f(x)} \int \ldots \int \prod_{k=1}^{m-1} \text{f(x}_k\text{)} \prod_{k=m+1}^{n} \text{f(x}_k\text{) dx}_1\text{...dx}_{k-1}\text{dx}_{k+1}\text{...dx}_n\text{,} \tag{2.11}$$

where the integration is over the domain

$$-\infty<x_1<\ldots<x_{k-1}<x<x_{k+1}<\ldots<x_n <\infty.$$

The symmetry of

$$\prod_{k=1}^{m-1} \text{f(x}_k\text{)}$$

with respect to $x_1,\ldots,x_{m-1}$, as well as the symmetry of

$$\prod_{k=m+1}^{n} \text{f(x}_k\text{)}$$

with respect to $x_{m+1},\ldots,x_n$, helps us to evaluate the integral on the RHS of (2.11) as follows:

$$\int \ldots \int \prod_{k=1}^{m-1} \text{f(x}_k\text{)} \prod_{k=m+1}^{n} \text{f(x}_k\text{) dx}_1\text{...dx}_{k-1}\text{dx}_{k+1}\text{...dx}_n\text{=}$$

$$\frac{1}{(m-1)!} \prod_{k=1}^{m-1} \int_{-\infty}^{x} f(x_k)dx_k \frac{1}{(n-m)!} \prod_{k=m+1}^{n} \int_{x}^{\infty} f(x_k)dx_k =$$

$$(F(x))^{m-1}(1-F(x))^{n-m}/(m-1)!(n-m)!. \tag{2.12}$$

Combining (2.11) and (2.12), one gets that

$$f_{m:n}(x)= \frac{n!}{(m-1)!(n-m)!}(F(x))^{m-1}(1-F(x))^{n-m} f(x). \tag{2.13}$$

Indeed, equality (2.13) is immediately follows from the corresponding formula for d.f.'s of single order statistics (see (2.8), for example), but the technique, which we used to prove (2.13), is applicable for more complicated situations. The following exercise can illustrate this statement.

Exercise 2.7. Find the joint p.d.f.

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

of order statistics

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

where $1\le k(1)<k(2)<\dots<k(r)\le n$.

Remark 2.4. In the sequel we will often use the particular case of the joint probability density functions from exercise 2.7, which corresponds to the case r=2. It turns out that

$$f_{i,j:n}\,(x_1,x_2)=$$

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!}(F(x_1))^{i-1}(F(x_2)-F(x_1))^{j-i-1}(1-F(x_2))^{n-j}f(x_1)f(x_2), \qquad (2.14)$$

if $1\le i<j\le n$ and $x_1<x_2$.

Expression (2.10) enables us also to get the joint d.f.

$$F_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n) = P\{X_{1,n}\le x_1,\ X_{2,n}\le x_2,\dots,\ X_{n,n}\le x_n\}.$$

One has

$$F_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n)= n!\iiint\limits_{D}\prod_{k=1}^{n} f(u_k)du_1\dots du_n\ , \qquad (2.15)$$

where

$$D=\{u_1,\dots,u_n:\ u_1<u_2<\dots<u_n;\ u_1<x_1,u_2<x_2,\dots,u_n<x_n\}.$$

Note that (2.15) is equivalent to the expression

$$F_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n)= n!\iiint\limits_{\widehat{D}} du_1\dots du_n, \qquad (2.16)$$

where integration is over

$$\widehat{D}=\{u_1,\dots,u_n:\ u_1<u_2<\dots<u_n;\ u_1<F(x_1),u_2<F(x_2),\dots,u_n<F(x_n)\}.$$

Remark 2.5. It can be proved that unlike (2.15), which needs the existence of population density function f, expression (2.16), as well as (2.9), is valid for arbitrary distribution function F.

Check your solutions

Exercise 2.1 **(solution).** If $x\ge y$, then

$$P\{M(n-1)\le x,\ M(n)\le y\}= P\{M(n)\le y\}= H(y,y,\dots,y).$$

Otherwise,

$$P\{M(n-1)\leq x, M(n)\leq y\}=P\{ M(n-1)\leq x, X_n\leq y\}=H(x,\ldots,x,y).$$

Exercise 2.2 **(answer).** $P\{Y\leq x\}=1-H(x,\infty) - H(\infty,x) + H(x,x)$,

where

$$H(x, \infty)=P\{X_1\leq x, X_2<\infty\}=P\{X_1\leq x\}$$

and

$$H(\infty,x)=P\{X_2\leq x\}.$$

Exercise 2.3 **(solution**). If $x\geq y$, then

$$P\{Y\leq x, Z\leq y\}=P\{Z\leq y\}=H(y,y).$$

If $x<y$, then

$$P\{Y\leq x, Z\leq y\}=P\{X_1 \leq x, X_2 \leq y\}+P\{X_1\leq y, X_2\leq x\}-P\{X_1\leq x, X_2\leq x\}= H(x,y) + H(y,x) - H(x,x).$$

Exercise 2.4 **(solution)**. One can see that

$$P\{Y\geq r, Z<s\}=P\{r\leq X_k<s, k=1,2,\ldots,n\}=$$

$$(P\{r\leq X<s\})^n =(P\{X\geq r\}- P\{X\geq s\})^n=(p^r-p^s)^n.$$

Exercise 2.5 **(solution)**. It is easy to see that (2.6) is valid for x=0. Now it suffices to prove that both sides of (2.6) have equal derivatives. The derivative of the RHS is naturally equal to

$$x^{k-1}(1-x)^{n-k} /B(k,n-k+1) = n!x^{k-1}(1-x)^{n-k}/(k-1)!(n-k)!,$$

because

$$B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$$

and the gamma function satisfies equality

$$\Gamma(k) = (k-1)! \text{ for } k=1,2,\ldots.$$

It turns out (after some simple calculations) that the derivative of the LHS also equals

$$n!x^{k-1}(1-x)^{n-k}/(k-1)!(n-k)! .$$

Exercise 2.6 **(solution)**. Let r<s. Denote

$$F_{r,s:n}(x_1,x_2)=P\{X_{r,n}\le x_1,\ X_{s,n}\le x_2\}.$$

If $x_2 \le x_1$, then evidently

$$P\{X_{r,n}\le x_1,\ X_{s,n}\le x_2\}=P\{X_{s,n}\le x_2\}$$

and

$$F_{r,s:n}(x_1,x_2)=\sum_{m=s}^{n}\binom{n}{m}(F(x_2))^m(1-F(x_2))^{n-m}=\boldsymbol{\mathit{1}}_{F(x_2)}\,(s,n-s+1).$$

Consider now the case $x_2 > x_1$. To find $F_{r,s:n}(x_1,x_2)$ let us mention that any X from the sample $X_1,X_2,\ldots,X_n$ (independently on other X's) with probabilities $F(x_1)$, $F(x_2)-F(x_1)$ and $1-F(x_2)$ can fall into intervals $(-\infty,x_1]$, $(x_1,x_2]$, (x_2,∞), respectively. One sees that the event $A=\{X_{r,n}\le x_1,\ X_{s,n}\le x_2\}$ is a union of some disjoint events

$A_{i,j,n-i-j}$={ i elements of the sample fall into $(-\infty,x_1]$, j elements fall into

Interval $(x_1,x_2]$ and (n-i-j) elements lie to the right of x_2}.

Recalling the polynomial distribution we obtain that

$$P\{A_{i,j,n-i-j}\}=\frac{n!}{i!\,j!\,(n-i-j)!}(F(x_1))^i(F(x_2)-F(x_1))^j(1-F(x_2))^{n-i-j}.$$

To construct A one has to take all $A_{i,j,n-i-j}$ such that $r\le i\le n$, $j\ge 0$ and $s\le i+j\le n$.

Hence,

$$F_{r,s:n}(x_1,x_2)=P\{A\}=\sum_{i=r}^{n}\ \sum_{j=\max\{0,s-i\}}^{n-i} P\{A_{i,j,n-i-j}\}=$$

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

Exercise 2.7 **(answer).** Denote for convenience, k(0) =0, k(r+1) =n+1, $x_0 = -\infty$ and $x_{r+1} = \infty$. Then

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

$$\frac{n!}{\prod_{m=1}^{r+1}(k(m)-k(m-1)-1)!}\prod_{m=1}^{r+1}(F(x_m)-F(x_{m-1}))^{k(m)-k(m-1)-1}\prod_{m=1}^{r}f(x_m),$$

if $x_1<x_2<\ldots<x_r$, 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

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

Chapter 3. Sample quantiles and ranges
Выборочные квантили и ранги

sample quantile = выборочная квантиль

sample range = размах выборки (выборочный размах)

quasi- range = квази-размах

midrange = середина размаха

quasi- midranges = середина квази-размаха

sufficient statistic =достаточная статистика

likelihood function = функция правдоподобия

maximum likelihood estimate = оценка максимального правдоподобия

location and scale parameters = параметры сдвига (положения) и масштаба

sample distribution function = выборочная функция распределения

sample median= выборочная медиана

unbiased estimate = несмещенная оценка

It turns out very often that in statistical inference estimates of some unknown parameters, which are the best in some sense (efficient, robust) or satisfy useful properties (sufficient, simple and convenient for applications), have the form of order statistics or can be expressed as functions of order statistics.

Example 3.1. Let us have a sample $X_1,\ldots,X_n$ of size n from a population d.f. $F(x,\theta)$, θ being an unknown parameter, which we need to estimate using some statistic $T=T(X_1,\ldots,X_n)$. Statistic T is sufficient if it contains as much information about θ as all sample $X_1,\ldots,X_n$. Rigorously saying , T is a sufficient for θ if the conditional distribution of the vector $(X_1,\ldots,X_n)$ given T= t does not depend on θ. There are some useful criteria to determine sufficient statistics. For instance, consider the case, when our population has a probability density function $f(x,\theta)$. Then T is sufficient for θ if

the equality

$$f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta) = h(x_1,x_2,\ldots,x_n)g(\theta, T(x_1,x_2,\ldots,x_n)) \qquad (3.1)$$

holds for some nonnegative functions ***h*** (which does not depend on θ) and ***g*** (which depends on θ and $T(x_1,x_2,\ldots,x_n)$ only).

Let now X have the uniform distribution on $[0, \theta]$, $\theta>0$ being an unknown parameter. In this case $f(x, \theta) =1/\theta$, if $0\le x\le\theta$, and $f(x, \theta) =0$, otherwise. Thus,

$$f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta)=1/\theta^n,\ 0 \le x_1,x_2,\ldots,x_n \le \theta, \qquad (3.2)$$

and

$$f(x_1,\theta)f(x_2,\theta)\ldots f(x_n,\theta)=0, \text{ otherwise.}$$

The RHS of (3.2) can be expressed as

$$h(x_1,x_2,\ldots,x_n)g(\theta, T(x_1,x_2,\ldots,x_n)),$$

where

$$h(x_1,x_2,\ldots,x_n) =1, \quad \text{if } x_k\ge0,\ k=1,2,\ldots,n, \text{ and } h(x_1,x_2,\ldots,x_n) =0, \quad \text{otherwise;}$$

$$T(x_1,x_2,\ldots,x_n)=\max\{x_1,x_2,\ldots,x_n\}$$

and

$$g(\theta, T(x_1,x_2,\ldots,x_n)) =1_{\{T(x_1,x_2,\ldots,x_n)\le\theta\}} /\theta^n.$$

One sees that in this case the sufficient statistic has the form

$$T=\max\{X_1,\ldots,X_n\}=X_{n,n}.$$

Example 3.2. Let us again consider a sample $X_1,\ldots,X_n$ from a population, having a probability density function $f(x, \theta)$, where θ is an unknown parameter. In this situation the likelihood function is defined as

$$L(x_1,x_2,\ldots,x_n,\theta)=f(x_1, \theta) f(x_2, \theta)\ldots f(x_n, \theta). \qquad (3.3)$$

To construct the maximum likelihood estimate of θ one must find such

$$\theta^*= \theta^*(x_1,\ldots, x_n),$$

which maximizes the RHS of (3.3), and take $\theta^*(X_1,\ldots,X_n)$ as the estimate of θ.

Consider the case, when a sample is taken from the Laplace distribution with p.d.f.

$$f(x, \theta)=\exp(-|x-\theta|)/2 .$$

What is the maximum likelihood estimate of θ? We see that the likelihood function in this case has the form

$$L(x_1,x_2,\dots,x_n,\theta)=\exp\{-\sum_{k=1}^{n}|x_k-\theta|\}/2^n. \qquad (3.4)$$

To maximize $L(x_1,x_2,\dots,x_n,\theta)$ it suffices to minimize

$$\sigma(\theta)=\sum_{k=1}^{n}|x_k-\theta|.$$

Let us arrange observations $x_1,\dots,x_n$ in the non-decreasing order: $x_{1,n}\le\dots\le x_{n,n}$. Here we must distinguish two situations. At first, we consider odd values of n. Let n=2k+1, k=0,1,2,…. It is not difficult to show that $\sigma(\theta)$ decreases with respect to θ in the interval $(-\infty, x_{k+1,2k+1})$ and $\sigma(\theta)$ increases for $\theta\in(x_{k+1,2k+1}, \infty)$. It implies that $\sigma(\theta)$ attains its minimal value if θ coincides with $x_{k+1,2k+1}$. Thus, in this case order statistic $X_{k+1,2k+1}$ is the maximum likelihood estimate of the location parameter θ. If n=2k, k=1,2,…, is even, then $\sigma(\theta)$ decreases in $(-\infty,x_{k,2k})$, increases in $(x_{k+1,2k}, \infty)$ and has a constant value in the interval $(x_{k,2k}, x_{k+1,2k})$. Hence $\sigma(\theta)$ attains minimal values for any $\theta\in[x_{k,2k}, x_{k+1,2k}]$. Any point of this interval can be presented as

$$\alpha x_{k,2k}+(1-\alpha)x_{k+1,2k},$$

where $0\le\alpha\le1$. It means that any statistic of the form

$$\alpha X_{k,2k}+(1-\alpha)X_{k+1,2k},\ 0\le\alpha\le1,$$

is the maximum likelihood estimate of θ. If $\alpha=1/2$, we get the statistics

$$(X_{k,2k}+X_{k+1,2k})/2.$$

Example 3.3. Consider a sample from the normal $N(\theta, \sigma^2)$ population. Let

$$\overline{X} = (X_1+\dots+X_n)/n$$

denote the sample mean. It is known that for normal distributions the vector

$$(X_1-\overline{X}, X_2-\overline{X},\dots, X_n-\overline{X})$$

and the sample mean $\overline{X}$ are independent. Then the vector

$$(X_{1,n}-\overline{X}, X_{2,n}-\overline{X},\dots,X_{n,n}-\overline{X})$$

and $\overline{X}$ are also independent. In statistical inference, based on the normal samples, statisticians very often need to use independent estimates of location (θ) and scale (σ) parameters. The best in any respects for their purposes are independent estimates $\overline{X}$ and

$$S = \left(\frac{1}{(n-1)}\sum_{k=1}^{n} (X_k - \overline{X})^2\right)^{1/2}.$$

For the sake of simplicity we can change S by another estimate of σ. Convenient analogues of S are presented by statistics

$$d(r,k,n)(X_{r,n}-X_{k,n}),$$

where d(r,k,n) are correspondingly chosen normalizing constants, which provide unbiased estimation of parameter σ. Since differences $(X_{r,n}-\overline{X})$ and $(X_{s,n}-\overline{X})$ do not depend on $\overline{X}$, the statistic

$$(X_{r,n}-X_{k,n}) = (X_{r,n}-\overline{X}) - (X_{s,n}-\overline{X})$$

does not depend on $\overline{X}$ also. Thus, we see that statistics of the form $X_{r,n} - X_{k,n}$, $1 \le k < r \le n$ (статистики, которые мы называем квази-размахами выборки!), have good properties and can be used in the estimation theory.

The suggested examples show that order statistics naturally arise in statistical inference. The most popular are extreme order statistics, such as $X_{1,n}$, $X_{n,n}$, and the so-called sample quantiles. To determine sample quantiles we must recall the definition of the quantiles for random variables (or quantiles of d.f. F).

Definition 3.1. A value x_p is called a *quantile of order p*, $0<p<1$, if

$$P\{X<x_p\} \le p \le P\{X \le x_p\}. \tag{3.5}$$

If F is a d.f. of X, then (3.5) is equivalent to the relation

$$F(x_p-0) \le p \le F(x_p). \tag{3.6}$$

For continuous F, x_p is any solution of the equation

$$F(x_p) = p. \tag{3.7}$$

Note that (3.7) has a unique solution, if F is strictly increasing. Otherwise, any point of the interval $[\delta,\gamma]$, where

$$\delta = \inf\{x: F(x) = p\}$$

and

$$\gamma = \sup\{x: F(x) = p\},$$

satisfies (3.7) and may be called a quantile of order p. Very often one uses the sample distribution function

$$F_n^*(x) = \frac{1}{n}\sum_{k=1}^{n} 1_{\{X_k \le x\}}$$

to estimate the population d.f. F. It is natural to take quantiles of $F_n^*(x)$ as estimates of quantiles of F. Substituting $F_n^*(x)$ instead of F to (3.6), we get the relation

$$F_n^*(x-0) \le p \le F_n^*(x). \tag{3.8}$$

Let us recall now (see (1.16)) that

$$F_n^*(x) = k/n, \text{ if } X_{k,n} \le x < X_{k+1,n},\ 1 \le k \le n-1.$$

Comparing (1.16) and (3.8) one obtains that the only solution of (3.8) is $X_{k,n}$ if

$(k-1)/n < p < k/n$, $k=1,2,\ldots,n$. If $p=k/n$, $k=1,2,\ldots,n-1$, then any $x \in [X_{k,n}, X_{k+1,n}]$ satisfies (3.8). Hence,

*i*f ***pn*** is an integer, then any statistics of the form

$$\alpha X_{pn,n} + (1-\alpha)X_{pn+1,n},\ 0 \le \alpha \le 1,$$

including

$$X_{pn,n},\quad X_{pn+1,n},\quad (X_{pn,n} + X_{pn+1,n})/2$$

as possible options, can be regarded as a sample quantile. Otherwise, (3.8) has the unique solution $X_{[pn]+1,n}$. Thus, the evident simple definition of the sample quantile of order p, 0<p<1, which covers both situations, is given as $\boldsymbol{X_{[pn]+1,n}}$.

Example 3.4. We return to the case, when ***np*** is an integer. Indeed, if a size of the sample is large, there is no essential difference between all possible versions of sample quantiles and we can take the simplest of them, say $X_{pn+1,n}$. For small sizes of samples different definitions can give distinguishable results of statistical procedures. Hence in any concrete situation we must choose the best (in some sense) of the statistics

$$\alpha X_{pn,n} + (1-\alpha)X_{pn+1,n},\ 0 \le \alpha \le 1.$$

One of the possible criterions of the optimal choice is the unbiasedness of the statistic.

Consider the case, when the sample quantile is used to estimate the quantile of order ***pn*** for the uniform on the interval [0,1] distribution. The statistic

$$\alpha X_{pn,n} + (1-\alpha)X_{pn+1,n}$$

is unbiased in this case, if the following equality holds:

$$E(\alpha X_{pn,n} + (1-\alpha)X_{pn+1,n}) = p. \tag{3.9}$$

It appears (some later it will be proved) that

$$EX_{k,n}= k/(n+1) , 1\leq k\leq n, \qquad (3.10)$$

for the uniform U([0,1]) distribution. From (3.9) and (3.10) we obtain that α=1-p.

The most important of sample quantiles is the sample median, which corresponds to the case p=1/2. If n=2k+1, k=1,2,…, then the sample median is defined as $X_{k+1,2k+1}$. For even (n=2k, k=1,2,…) size of a sample any statistics of the form

$$\alpha X_{k,2k} +(1-\alpha)X_{k+1,2k}, 0\leq\alpha\leq 1,$$

may be regarded as the sample median. Sample medians are especially good for estimation of the location parameter in the case, when the population distribution is symmetric.

We say that X is symmetric random variable if X and –X have the same distribution. Analogously, X is symmetric with respect to some location parameter θ, if X-θ and θ-X have the same distribution. If X is symmetric with respect to some value θ, then X-θ is simply symmetric. If X is symmetric with respect to θ and there exists the expectation of X, then EX equals θ, as well as the median of X. Hence, if θ is an unknown parameter, we can use different estimates for θ, the sample mean and the sample median among them. Moreover, there are situations (see, for instance, exercise 3.2), when the sample median is the best estimate in some sense.

***Exercise 3.1**.* Show that if X has some symmetric distribution then equality

$$x_p = -x_{1-p},\ 0<p<1,$$

holds for quantiles of order p and 1-p.

Exercise 3.2. Let $X_{1,n}, X_{2,n},…, X_{n,n}$ be order statistics , corresponding to a continuous d.f. F, and let x_p and x_{1-p} denote quantiles of order p and 1-p, respectively, for F. Show that then the following equality

$$P\{X_{[\alpha n],n}\leq x_p\}+P\{X_{[(1-\alpha)n],n}\leq x_{1-p}\}=1$$

holds for any 0<α<1 in the case, when ***αn*** is not an integer, while the relation

$$P\{X_{\alpha n,n}\leq x_p\}+P\{X_{(1-\alpha)n+1,n}\leq x_{1-p}\} =1$$

is valid for the case, when ***αn*** is an integer.

The statement of the next exercise arises to van der Vaart's (1961) paper.

Exercise 3.3. Let $X_{k+1,2k+1}$ be a sample median based on a sample of odd size from a distribution with a continuous d.f. F. Show that the median of $X_{k+1,2k+1}$ coincides with the median of the population distribution.

***Exercise 3.4**.* Let $X_{k+1,2k+1}$ be a sample median, corresponding to a sample of odd size. Show that $X_{k+1,2k+1}$ has a symmetric distribution if and only if the population distribution is symmetric.

As we mentioned above, sample medians may be good estimates of the location parameter. One more type of statistics, which are used for estimation of the location parameter, is presented by different midranges. The classical midrange is defined as

$$(X_{1,n}+X_{n,n})/2,$$

while the so-called quasi-midranges are given as

$$(X_{k,n}+X_{n-k+1,n})/2,\ k=2,\dots,[n/2].$$

We can see that the sample median

$$(X_{k,2k}+X_{k+1,2k})/2$$

also presents one of quasi-midranges. As a measure of the population spread, a statistician can use ranges $X_{n,n}-X_{1,n}$ and quasi-ranges $X_{n-k+1,n}-X_{k,n}$, $k=2,\dots,[n/2]$.

In example 3.3 we found that any quasi-range $X_{n-k+1,n}-X_{k,n}$ (as well as range $X_{n,n}-X_{1,n}$) and the sample mean $\overline{X}=(X_1+\dots+X_n)/n$ are independent for the normal distribution. To use ranges and midranges in statistical inference we need to know distributions of these statistics.

Example 3.5. We suppose that our population has p.d.f. ***f*** and will try to find probability density functions of ranges

$$W_n=X_{n,n}-X_{1,n},\quad n=2,3,\dots.$$

Substituting i=1 and j=n to (2.14), one can get the joint pdf of order statistics $X_{1,n}$ and $X_{n,n}$ as follows :

$$f_{1,n:n}(x,y)=\ n(n-1)(F(y)-F(x))^{n-2}f(x)f(y)\text{ , if }x<y, \tag{3.11}$$

and

$$f_{1,n:n}(x,y)=0,\text{ if }x\geq y.$$

Consider the linear change of variables (u,v) = (x,y-x) with the unit Jacobian, which corresponds to the passage to random variables $U=X_{1,n}$ and $V=X_{n,n}-X_{1,n}>0$. Now (3.11) implies that random variables U and V have the joint density function

$$f_{U,V}(u,v)=n(n-1)(F(u+v)-F(u))^{n-2}f(u)f(u+v),\ -\infty<u<\infty,\ v>0, \tag{3.12}$$

and $f_{U,V}(u,v)=0$, otherwise.

Integrating (3.12) with respect to u, one obtains that the range $X_{n,n}-X_{1,n}$ has the density

$$f_V(v)=n(n-1)\int_{-\infty}^{\infty}(F(u+v)-F(u))^{n-2}f(u)f(u+v)du,\ v>0. \tag{3.13}$$

One more integration (now with respect to v) enables us to get the distribution function of the range:

$$F_V(x)=P\{X_{n,n}-X_{1,n}\leq x\}=\int_0^x f_V(v)dv=$$

$$n\int_{-\infty}^{\infty} f(u)(\int_0^x d((F(u+v)-F(u))^{n-1}))du=$$

$$n\int_{-\infty}^{\infty} (F(u+x)-F(u))^{n-1} f(u)du,\ x>0. \tag{3.14}$$

Exercise 3.5. Find the distribution of quasi-range

$$X_{n-r+1,n}-X_{r,n}\ ,\ r=1,2,\ldots,[n/2],$$

When $f(x)=1,\ a\leq x\leq a+1$.

Exercise 3.6. Let

$$F(x)=\max\{0,1-\exp(-x)\}.$$

Show that $U=X_{1,n}$ and the range $V=X_{n,n}-X_{1,n}$ are independent.

Exercise 3.7. Let f(x) be a population density function and

$$V=(X_{r,n}+X_{n-r+1,n})/2$$

be the corresponding quasi-midrange. Find the p.d.f. of V.

Example 3.6. We will find now the joint distribution of $W=X_{n,n}-X_{1,n}$ and $V=(X_{1,n}+X_{n,n})/2$. Consider (3.11) and make the linear change of variables

$$(w,v)=(y-x,(x+y)/2)$$

with the unit Jacobian, which corresponds to the passage to random variables W>0 and V. After noting that

$$x=v-\frac{w}{2} \text{ and } y=v+\frac{w}{2},$$

one gets the joint probability density function of W and V:

$$f_{W,V}(w,v)=$$

$$n(n-1)(F(v+\frac{w}{2})-F(v-\frac{w}{2}))^{n-2}f(v-\frac{w}{2})f(v+\frac{w}{2}),\ -\infty<v<\infty,\ w>0. \qquad (3.15)$$

Remark 3.1. Consider the joint distribution of range W and midrange V in the case, when n=2 and

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2).$$

From (3.15) we have

$$f_{W,V}(w,v)=\frac{1}{\pi}\exp(-v^2)\exp(-w^2/4),\ -\infty<v<\infty,\ w>0. \qquad (3.16)$$

Equality (3.16) means that W and V are independent. This fact is not surprising so far as for n=2 the midrange coincides with the sample mean

$$\overline{X}=(X_1+X_2)/2$$

and we know from example 3.3 that the ranges and sample means

$$\overline{X}=(X_1+\ldots+X_n)/n$$

are independent for normal distributions.

Check your solutions

***Exercise 3.1* (solution)**. If X is symmetrically distributed, then

$$P\{X<x\}= P\{X>-x\}$$

and

$$P\{X\le x\}= P\{X\ge -x\}.$$

Let x_p be a quantile of order p. It means (see (3.6)) that

$$P\{X<x_p\}\le p \le P\{X \le x_p\}.$$

It follows now from the relations given above that

$$P\{X >-x_p\}\le p \le P\{X \ge -x_p\}.$$

The latter inequalities can be rewritten in the form

$$P\{X< -x_p\}\le 1\text{-}p \le P\{X\le -x_p\},$$

Which confirms that $-x_p$ satisfies the definition of the quantile of order 1-p. Note that if any point of some interval [a,b] is a quantile of order p for symmetric distribution, then any point of interval [-b,-a] is a quantile of order 1-p.

Exercise 3.2 **(solution)**. Recalling that $F(x_p) = p$ for continuous d.f., we have from (2.5) that

$$P\{X_{[\alpha n],n}\le x_p\}= \sum_{m=[\alpha n]}^{n} \binom{n}{m} (F(x_p))^m(1-F(x_p))^{n-m}=$$

$$\sum_{m=[\alpha n]}^{n} \binom{n}{m} p^m(1-p)^{n-m}$$

and similarly

$$P\{X_{[(1-\alpha)n],n}\le x_{1-p}\}= \sum_{m=[(1-\alpha)n]}^{n} \binom{n}{m} (1-p)^m p^{n-m}=$$

$$\sum_{m=0}^{n-[(1-\alpha)n]} \binom{n}{n-m} (1-p)^{n-m}p^m = \sum_{m=0}^{n-[(1-\alpha)n]} \binom{n}{m} p^m(1-p)^{n-m},$$

so far as

$$\binom{n}{n-m} = \binom{n}{m}.$$

Consider values r= [αn] and s=[(1-α)n]. It is easy to see that r+s =n, if ***αn*** is an integer, and r+s=n-1 in the opposite case. Thus, if ***αn*** is not an integer, we obtain that

$$n-[(1-\alpha)n]=r-1=[\alpha n]-1$$

and

$$P\{X_{[\alpha n],n}\le x_p\}+ P\{X_{[(1-\alpha)n],n}\le x_{1-p}\}= \sum_{m=0}^{n} \binom{n}{m} p^m(1-p)^{n-m}=(p+(1-p))^n=1.$$

If ***αn*** is an integer, then n-[(1-α)n]=r=[αn] and we get one more term in our sum. In this case

$$P\{X_{\alpha n,n} \le x_p\} + P\{X_{(1-\alpha)n,n} \le x_{1-p}\} = 1 + \binom{n}{\alpha n} p^{\alpha n}(1-p)^{(1-\alpha)n} > 1.$$

Some understandable change gives us the following equality:

$$P\{X_{\alpha n,n} \le x_p\} + P\{X_{(1-\alpha)n+1,n} \le x_{1-p}\} = 1.$$

Exercise 3.3 **(solution)**. Let μ be a population median. It means that

$$F(\mu)=1/2.$$

It follows now from (2.5) that

$$P\{X_{k+1,2k+1} \le \mu\} = \sum_{m=k+1}^{2k+1} \binom{n}{m} (F(x_p))^m (1-F(x_p))^{n-m} =$$

$$\sum_{m=k+1}^{2k+1} \binom{n}{m} (\frac{1}{2})^m (\frac{1}{2})^{n-m} = (\frac{1}{2})^{2k+1} \sum_{m=k+1}^{2k+1} \binom{n}{m} = 1/2,$$

so far as

$$\sum_{m=k+1}^{2k+1} \binom{n}{m} = \sum_{m=0}^{k} \binom{n}{m}$$

and hence,

$$\sum_{m=k+1}^{2k+1} \binom{n}{m} = \frac{1}{2} \sum_{m=0}^{2k+1} \binom{n}{m} = \frac{1}{2} (1+1)^{2k+1} = 2^{2k}.$$

Exercise 3.4 **(solution)**. We can use equality (2.8), which in our situation has the form

$$F_{k+1,2k+1}(x) = I_{F(x)}\,(k+1,k+1) = \frac{1}{B(k+1,k+1)} \int_0^{F(x)} t^k(1-t)^k dt,$$

Where

$$B(k+1,k+1) = \int_0^1 t^k(1-t)^k dt.$$

If the population distribution is symmetric, F(-x)=1-F(x-0)

and then for any x we have

$$1-F_{k+1,2k+1}(x-0)=1-\frac{1}{B(k+1,k+1)}\int_{0}^{F(x-0)} t^k(1-t)^k dt=$$

$$\frac{1}{B(k+1,k+1)}\int_{F(x-0)}^{1} t^k(1-t)^k dt=\frac{1}{B(k+1,k+1)}\int_{1-F(-x)}^{1} t^k(1-t)^k dt=$$

$$\frac{1}{B(k+1,k+1)}\int_{0}^{F(-x)} t^k(1-t)^k dt=F_{k+1,2k+1}(-x).$$

This means that $X_{k+1,2k+1}$ has a symmetric distribution.

Let now admit that $X_{k+1,2k+1}$ is a symmetric random variable. Then, for any x,

$$1-F_{k+1,2k+1}(x-0) = F_{k+1,2k+1}(-x)$$

and this is equivalent to the relation

$$I_{F(-x)}(k+1,k+1)+I_{F(x-0)}(k+1,k+1)=1.$$

It is not difficult to see that the latter equality, in its turn, is equivalent to the relation

$$\int_{0}^{F(-x)} t^k(1-t)^k dt+\int_{1-F(x-0)}^{1} t^k(1-t)^k dt=B(k+1,k+1)=\int_{0}^{1} t^k(1-t)^k dt,$$

which immediately implies that

F(-x)=1-F(x-0).

Hence, the population distribution is symmetric.

***Exercise* 3.5 (hint and answer).** Consider (2.14) with

i=r, j=n-r+1, f(x)=1, a≤x≤a+1, F(x)=x-a, a≤x≤a+1 and use the linear change of variables (u,v) =(x,y-x). It will give you the joint density function of $X_{r,n}$ and $W(r)=X_{n-r+1,n}-X_{r,n}$. Now the integration with respect to u enables you to get the density function of W(r).

It turns out, that the density function of W(r) does not depend on ***a*** and has the form

$$f_{W(r)}(x)= x^{n-2r}(1-x)^{2r-1}/B(n-2r+1,2r),\ 0\le x\le 1,$$

i.e. W(r) has the beta distribution with parameters n-2r+1 and 2r.

Exercise 3.6 **(solution).** In this case, recalling (3.12), we have that for any u>0 and v>0, the joint density function is given as follows:

$$f_{U,V}(u,v) =n(n-1)(\exp(-u)-\exp(-(u+v))))^{n-2}\exp(-u)\exp(-u-v) =$$

$$n(n-1)\exp(-nu)(1-\exp(-v))^{n-2}\exp(-v)=h(u)g(v),$$

where

$$h(u) = n\exp(-nu)$$

and

$$g(v) = (n-1)\ (1-\exp(-v))^{n-2}\exp(-v).$$

The existence of the factorization

$$f_{U,V}(u,v) = h(u)g(v)$$

suffices to state that U and V are independent random variables. Indeed, one can check that in fact h(u) and g(v) are densities of U and V respectively.

Exercise 3.7 (solution). Substituting i=r and j=n-r+1 to (2.14), we get the joint pdf of order statistics $X_{r,n}$ and $X_{n-r+1,n}$:

$$f_{r,n-r+1:n}(x,y)= c(r,n)(F(x))^{r-1}(F(y)-F(x))^{n-2r}(1-F(y))^{r-1} f(x)f(y),\ x<y,$$

and

$$f_{r,n-r+1:n}(x,y)=0,$$

otherwise, where

$$c(r,n)=n!/(r-1)!(n-2r)!(r-1)!.$$

Consider the linear change of variables (u,v) = (x,(x+y)/2), which corresponds to the passage to random variables

$$U=X_{1,n} \text{ and } V=(X_{r,n}+X_{n-r+1,n})/2.$$

After noting that the Jacobian of this transformation is ½, $x=u$, $y=2v-u$ and the inequality $y>x$ means that $v>u$, one obtains that the joint p.d.f. of U and V is given as follows:

$$f_{U,V}(u,v)=2c(r,n)\,(F(u))^{r-1}(F(2v-u)-F(u))^{n-2r}(1-F(2v-u))^{r-1}\,f(u)f(2v-u),\ u<v,$$

and

$$f_{U,V}(u,v)=0, \text{ if } u\geq v.$$

Integration with respect to ***u*** enables us to get the density function of V:

$$f_V(v)=2c(r,n)\int_{-\infty}^{v}(F(u))^{r-1}(F(2v-u)-F(u))^{n-2r}(1-F(2v-u))^{r-1}\,f(u)f(2v-u)du, \quad -\infty<v<\infty.$$

In particular, the p.d.f. of midrange

$$M=(X_{1,n}+X_{n,n})/2$$

has the form

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

while the p.d.f.'s of the quasi-midranges

$$(X_{k,2k}+X_{k+1,2k})/2\ ,\ k=1,2,\dots,$$

which coincide with sample medians for samples of even size, are given as follows:

$$f_V(v)=\frac{2(2k)!}{((k-1)!)^2}\int_{-\infty}^{v}(F(u))^{k-1}(1-F(2v-u))^{k-1}\,f(u)f(2v-u)du, \qquad -\infty<v<\infty.$$

Chapter 4. Representations for order statistics
Представления для порядковых статистик

Очень важная глава. Для заведомо зависимых случайных величин, которыми являются порядковые статистики, получены соотношения, позволяющие распределения этих статистик выражать через соответствующие распределения независимых случайных величин.

Exponential order statistics = экспоненциальные порядковые статистики

uniform order statistics = равномерные порядковые статистики

Random variables $X_1, X_2,\ldots,{}_n$ lose their original independence property being arranged in nondecreasing order. It is evident that order statistics tied by inequalities

$X_{1,n} \le X_2 \le \ldots \le X_n$ are dependent. Hence all situations, when we can express order statistics as functions of sums of independent terms, are very important for statisticians. In the sequel we will use the special notation $U_{1,n} \le \ldots \le U_{n,n}$ for the uniform order statistics (corresponding to the d.f. $F(x)=x, 0\le x\le 1$) and the notation $Z_{1,n} \le \ldots \le Z_{n,n}$ for the exponential order statistics ($F(x)=1-\exp(-x), x\ge 0$). We will prove some useful representations for exponential and uniform order statistics. At first we will show how different 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

$$G(s)=\inf\{x:F(x)\ge s\},\ 0<s<1. \qquad (4.1)$$

Exercise 4.1. Let F(x) be a continuous d.f. of a random variable X. Show that

in this case

$$F(G(x))=x,\ 0<x<1,$$

and $Y=F(X)$ has the uniform distribution on interval [0,1].

Remark 4.1. In exercise 4.1 it is proved, in particular, that the inverse function G(s) provides the equality

$$F(G(s))=s,\ 0<s<1,$$

if F is a continuous d.f. Moreover, it is not difficult to see that the dual equality

$$G(F(s))=s,\ 0<s<1,$$

olds for any s , where F(s) strongly increases.

Remark 4.2. The second statement of exercise 4.1 is very important. In fact, for any random variable with a continuous d.f. F we have equality

$$F(X) \overset{d}{=} U, \tag{4.2}$$

where any relation of the type

$$Y \overset{d}{=} Z$$

denotes that random variables (or random vectors) Y and Z have the same distribution. In (4.2) U is a random variable, which has the uniform on [0,1] distribution.

Indeed, (4.2) fails if F has jump points, since then the values of F(X), unlike U, do not cover all interval [0,1].

Exercise 4.2. Let X take on values 1, 2 and 3 with equal probabilities 1/3. Find the corresponding inverse function G(x), 0<x<1, and the distribution of G(U), where U has the uniform on [0,1] distribution.

Example 4.1. Now consider a more general case than in exercise 4.2. Take an arbitrary random variable X with a d.f. F. Let G be inverse of F . It is not difficult to see that inequality

$$G(s)\le z,\ 0<s<1,$$

Is equivalent to the inequality $s\le F(z)$. Hence, the events $\{G(U) \le z\}$ and $\{U\le F(z)\}$ have the same probability. Thus,

$$P\{G(U) \le z\}=P\{U\le F(z)\}=F(z). \tag{4.3}$$

Remark 4.3. It follows from (4.3) that relation

$$X \overset{d}{=} G(U), \tag{4.4}$$

where G is the inverse of d.f.F, holds for any random variable, while the dual equality

$$F(X) \overset{d}{=} U$$

is valid for random variables with continuous distribution functions only.

Let us take a sample $X_1, X_2,\dots, X_n$ and order statistics $X_{1,n} \le\dots\le X_{n,n}$ corresponding to a d.f. F and consider random variables

$$Y_k=F(X_k),\ k=1,2,\dots,n.$$

Let now $Y_{1,n},\dots,Y_{n,n}$ be order statistics based on $Y_1,\dots,Y_n$. Since F is a monotone function, it does not disturb the ordering of X's and hence the vector $(Y_{1,n},\dots,Y_{n,n})$ coincides with the vector $(F(X_{1,n}),\dots,F(X_{n,n}))$. If F is a continuous d.f., then, as we know from (4.2), independent random variables $Y_1,\dots,Y_n$ are uniformly distributed on interval [0,1] and hence the vector $(Y_{1,n},\dots,Y_{n,n})$ has the same distribution as the vector of uniform order statistics $(U_{1,n},\dots,U_{n,n})$. All the saying enables us to write that

$$(F(X_{1,n}),\dots, F(X_{n,n})) \overset{d}{=} (U_{1,n},\dots,U_{n,n}). \qquad (4.5)$$

Taking into account (4.4), we similarly have the following dual equality

$$(X_{1,n},\dots,X_{n,n}) \overset{d}{=} (G(U_{1,n}),\dots,G(U_{n,n})), \qquad (4.6)$$

which is valid (unlike (4.5)) for any distribution.

Example 4.2. Let now $X_{1,n} \le\dots\le X_{n,n}$ and $Y_{1,n} \le\dots\le Y_{n,n}$ be order statistics corresponding to an arbitrary d.f. F and a continuous d.f. H correspondingly. Let also G be the inverse of F. Combining relations (4.5) for Y's and (4.6) for X's, one gets the following equality, which ties two sets of order statistics:

$$(X_{1,n},\dots,X_{n,n}) \overset{d}{=} (G(H(Y_{1,n})),\dots,G(H(Y_{n,n}))). \qquad (4.7)$$

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

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

and (4.7) can be rewritten as

$$(X_{1,n},\dots,X_{n,n}) \overset{d}{=} (G(1-\exp(-Z_{1,n})),\dots,G(1-\exp(-Z_{n,n}))). \qquad (4.8)$$

Remark 4.4. Indeed, analogous results are valid for any monotone increasing function R(x) (no necessity to suppose that R is a distribution function). Namely, if

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

then the corresponding order statistics based on Y's and X's satisfy the relation

$$(Y_{1,n},\dots,Y_{n,n}) \overset{d}{=} (R(X_{1,n}),\dots,R(X_{n,n})). \qquad (4.9)$$

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

$$(Y_{1,n},\ldots,Y_{n,n}) \overset{d}{=} (R(X_{n,n}),\ldots,R(X_{1,n})). \tag{4.10}$$

Now we consider some special distributions, exponential and uniform among them. At first we consider order statistics based on exponential distributions.

Exercise 4.3. Let Z_1 and Z_2 be independent and have exponential distributions with parameters λ and μ respectively. Show that random variables

$$V= \min\{Z_1,Z_2\}$$

and

$$W= \max\{Z_1, Z_2\}- \min\{Z_1,Z_2\}$$

are independent.

Example 4.3. From the result of exercise 4.3 we see that if $Z_{1,2}$ and $Z_{2,2}$ are exponential order statistics based on a sample Z_1 and Z_2 from the standard $E(1)$ exponential distribution, then random variables

$$Z_{1,2} = \min \{Z_1,Z_2\}$$

and

$$Z_{2,2} - Z_{1,2} = \max\{Z_1,Z_2\}- \min\{Z_1,Z_2\}$$

are independent. It turns out that one can prove a more general result for differences of exponential order statistics.

We consider exponential order statistics $Z_{0,n}, Z_{1,n},\ldots,Z_{n,n}$, where $Z_{0,n}= 0$ is introduced for our convenience, and differences

$$V_k = Z_{k.n}-Z_{k-1,n},\ \ k=1,2,\ldots,n.$$

It appears that $V_1, V_2,\ldots,V_n$ are mutually independent random variables. To prove this, we recall (see (2.10)) that the joint probability density function of order statistics $X_{1,n},\ldots,X_{n,n}$, corresponding to a distribution with a density function f, has the form

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

otherwise it equals zero. In our case

$$f(x) = \exp(-x),\ x\geq 0,$$

and

$$f_{1,2,\ldots,n:n}(x_1,x_2,\ldots,x_n) = n!\exp(-(x_1+x_2+\ldots+x_n)),\ 0\leq x_1<x_2<\ldots<x_n <\infty. \tag{4.11}$$

By the linear change

$$(v_1,v_2,\dots,v_n) = (x_1,x_2-x_1,\dots,x_n-x_{n-1})$$

with the unit Jacobian, taking into account that

$$(x_1+x_2+\dots+x_n) = nv_1+(n-1)v_2+\dots+2v_{n-1}+v_n,$$

one obtains that the joint p.d.f. of differences $V_1, V_2,\dots,V_n$ is of the form

$$f(v_1,v_2,\dots,v_n) = \prod_{k=1}^{n} g_k(v_k),\ v_1>0,v_2>0,\dots,v_n>0, \tag{4.12}$$

where

$$g_k(v)=(n-k+1)\exp(-(n-k+1)v),\ v>0,$$

is the density function of the exponential $E(1/(n-k+1))$ distribution. Evidently, (4.12) means that $V_1, V_2,\dots,V_n$ are independent. Moreover, in fact, we obtained that the vector $(V_1, V_2,\dots,V_n)$ has the same distribution as the vector

$$(\frac{\nu_1}{n}, \frac{\nu_2}{n-1},\dots,\frac{\nu_{n-1}}{2},\nu_n),$$

where $\nu_1,\nu_2,\dots, \nu_n$ are independent random variables having the standard E(1) exponential distribution. Let us write this fact as

$$(V_1, V_2,\dots,V_n) \overset{d}{=} (\frac{\nu_1}{n}, \frac{\nu_2}{n-1},\dots,\nu_n). \tag{4.13}$$

Remark 4.5. The following important equalities are evident corollaries of (4.13):

$$(nV_1, (n-1)V_2,\dots,V_n) \overset{d}{=} (\nu_1, \nu_2 ,\dots,\nu_n) \tag{4.14}$$

(i.e., normalized differences $(n-k+1)V_k$, $k=1,2,\dots$, are independent and have the same exponential E(1) distribution);

$$(Z_{1,n}, Z_{2,n},\dots,Z_{n,n}) \overset{d}{=} (\frac{\nu_1}{n}, \frac{\nu_1}{n}+\frac{\nu_2}{n-1},\dots,\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+\dots+\frac{\nu_{n-1}}{2}+\nu_n). \tag{4.15}$$

(Еще раз подчеркнем, что соотношение (4.15) позволяет к экспоненциальным порядковым статистикам и их линейным комбинациям применять классические методы работы с независимыми случайными величинами).

Remark 4.6. Note that an analogous (but essentially more complicate) representation (via mixtures of sums of independent exponential values) of the order statistics based on exponential random variables with different scale parameters is given in Nevzorov (1984) .

Exercise 4.4. For exponential order statistics $Z_{1,n}, Z_{2,n},\ldots,Z_{n,n}$ show that if $c_1+\ldots+c_n= 0$, then $Z_{1,n}$ and any linear combination of order statistics

$$L=c_1 Z_{1,n}+c_2Z_{2,n}+\ldots+c_n Z_{n,n}$$

are independent.

Now we will study the structure of the uniform order statistics $U_{k,n}$, $1\leq k\leq n$.

Example 4.4. Taking into account relations (4.5) and (4.15) we get the following equalities:

$$(U_{1,n},\ldots,U_{n,n}) \overset{d}{=} (1-\exp(-Z_{1,n}),\ldots,1-\exp(-Z_{n,n})) \overset{d}{=}$$

$$(1-\exp(-\frac{\nu_1}{n}),\ldots,1-\exp(-(\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+\ldots+\frac{\nu_{n-1}}{2}+\nu_n))), \tag{4.16}$$

where $\nu_1,\ldots,\nu_n$ are independent and have the standard E(1) exponential distribution.

Recalling (4.2) we see that

$$(\exp(-\nu_1),\ldots, \exp(-\nu_n)) \overset{d}{=} (1-U_1,\ldots,1-U_n) \overset{d}{=}$$

$$(W_1,\ldots,W_n) \overset{d}{=} (W_n,\ldots,W_1), \tag{4.17}$$

where $U_1,U_2,\ldots,U_n$ as well as $W_1,W_2,\ldots,W_n$ are independent uniformly distributed on [0,1] random variables. Then (4.16) can be rewritten as

$$(1-U_{1,n},\ 1-U_{2,n},\ldots,\ 1-U_{n,n}) \overset{d}{=}$$

$$(W_n^{1/n}, W_n^{1/n} W_{n-1}^{1/(n-1)},\ldots, W_n^{1/n}W_{n-1}^{1/(n-1)}\ldots W_2^{1/2}W_1). \tag{4.18}$$

The standard uniform distribution on [0,1] is symmetric with respect to the point ½ . This enables us to state that

$$(1-U_{1,n},1-U_{2,n},\ldots,1-U_{n,n}) \overset{d}{=} (U_{n,n},U_{n-1,n},\ldots,U_{1,n}) . \tag{4.19}$$

More strongly one can use (4.10) with the function $R(x) =1-x$ and get (4.19). Combining now (4.18) and (4.19) we obtain that

$$(U_{1,n},U_{2,n},\ldots,U_{n,n}) \overset{d}{=}$$

$$(W_1W_2^{1/2}\ldots W_{n-1}^{1/(n-1)}\,W_n^{1/n},\ W_2^{1/2}\ldots W_{n-1}^{1/(n-1)}\,W_n^{1/n},\ldots,\ W_n^{1/n}). \qquad (4.20)$$

(Приведенное представление (4.20) не такое удобное, как соответствующее представление для экспоненциальных порядковых статистик, но оно играет важную роль при работе с произведениями равномерных порядковых статистик).

Thus, we see that any uniform order statistics $U_{k,n}$ is represented in (4.20) as the product of powers of independent uniformly distributed random variables as follows:

$$U_{k,n} \overset{d}{=} W_k^{1/k}W_{k+1}^{1/(k+1)}\ldots W_n^{1/n},\ k=1,2,\ldots,n. \qquad (4.21)$$

Exercise 4.5. Show that for any n=2,3,…, ratios

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

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

Exercise 4.6. Let $U_{k,n}$, $1\le k\le n$, n=1,2,… denote order statistics based on the sequence of independent, uniformly on the interval [0,1] distributed random variables $U_1,U_2,\ldots$ and $V_{k,n}$, $1\le k\le n$, be order statistics corresponding to the uniformly distributed random variables $V_1,V_2,\ldots,V_n$, where V's and U's are also independent. Show that the following equality holds for any $1\le m\le n$:

$$(U_{1,n},\ldots,U_{m,n}) \overset{d}{=} (U_{1,m}V_{m+1,n},\ldots,U_{m,m}V_{m+1,n}). \qquad (4.22)$$

Example 4.5. Taking into account (4.20) we observe that any uniform order statistic $U_{k,n}$ satisfies the following equality:

$$U_{k,n} \overset{d}{=} U_{k,k}^{(1)}\,U_{k+1,k+1}^{(2)}\cdots U_{n,n}^{(n-k+1)}, \qquad (4.23)$$

where the multipliers

$$U_{r,r}^{(r-k+1)},\ r=k,\ldots,n,$$

on the RHS of (4.23) are the maximum order statistics of (n-k+1) independent samples from the uniform distributions, which sizes are k, k+1,…,n respectively.

Example 4.6. There is one more representation for the uniform order statistics.

Let again $\nu_1,\nu_2,\dots$ be independent random variables having the standard E(1) exponential distribution and

$$S_n=\nu_1+\nu_2+\dots+\nu_n,\ n=1,2,\dots.$$

It turns out that the following representation is valid for the uniform order statistics:

$$(U_{1,n},\dots,U_{n,n}) \overset{d}{=} (\frac{S_1}{S_{n+1}},\dots,\frac{S_n}{S_{n+1}}). \qquad (4.24)$$

(В зависимости от решаемой задачи для равномерных порядковых статистик можно применять одно из этих двух соотношений – (4.21) или (4.24)).

To prove (4.24) one must recall (see (2.10) with $f(x)=1$, $0<x<1$) that the joint pdf of order statistics $(U_{1,n},\dots,U_{n,n})$ has the form:

$$f_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n)=n!,\ 0<x_1<x_2<\dots<x_n, \qquad (4.25)$$

and it equals zero otherwise.

Now we will prove that the RHS of (4.24) coincides with (4.25). Consider the joint pdf of random variables $\nu_1,\nu_2,\dots,\nu_{n+1}$, which naturally is of the form:

$$g(v_1,\dots,v_{n+1})= \exp\{-(v_1+\dots+v_{n+1})\},\ v_1\ge 0,\dots,v_{n+1}\ge 0. \qquad (4.26)$$

Using the linear transformation

$$(y_1,y_2,\dots,y_{n+1}) = (\frac{v_1}{v_1+\dots+v_{n+1}},\dots,\frac{v_1+v_2+\dots+v_n}{v_1+\dots+v_{n+1}},\ v_1+\dots+v_{n+1}),$$

one gets the joint distribution density $h(y_1,y_2,\dots,y_n,y_{n+1})$ of a set of random variables

$$\frac{S_1}{S_{n+1}},\dots,\frac{S_n}{S_{n+1}}$$

and S_{n+1} as follows:

$$h(y_1,y_2,\dots,y_n,y_{n+1})=(y_{n+1})^n\exp(-y_{n+1}), \qquad (4.27)$$

if $0<y_1<\dots<y_n,\ y_{n+1}\ge 0$.

The next step, integration over y_{n+1}, gives us the final formula for the joint p.d.f. of ratios

$$\frac{S_1}{S_{n+1}},\dots,\frac{S_n}{S_{n+1}}.$$

It is not difficult to see that this expression coincides with (4.25). Moreover, from (4.27) we get also the independence of the vector

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

and the sum S_{n+1}.

Representation (4.24) can be rewritten in the following useful form:

$$(U_{1,n},\ U_{2,n}-U_{1,n},\dots,U_{n,n}-U_{n-1,n}) \overset{d}{=} \left(\frac{\nu_1}{\nu_1+\dots+\nu_{n+1}},\dots,\frac{\nu_n}{\nu_1+\dots+\nu_{n+1}}\right). \qquad (4.28).$$

Exercise 4.7. Show that the uniform quasi-midrange

$$U_{n-k+1,n}-U_{k,n}\ ,\ k\leq n/2$$

has the same distribution as $U_{n-2k+1,n}$.

Exercise 4.8. We know from (4.24) that

$$(U_{1,n},\dots,U_{n,n}) \overset{d}{=} \left(\frac{S_1}{S_{n+1}},\dots,\frac{S_n}{S_{n+1}}\right),$$

where

$$S_k=\nu_1+\nu_2+\dots+\nu_k,\ k=1,2,\dots\ ,$$

and $\nu_1,\nu_2,\dots$ are independent random variables having the standard E(1) exponential distribution. There is one more representation of the uniform order statistics in terms of the sums S_k, namely,

$$(U_{1,n},\dots,U_{n,n}) \overset{d}{=} (S_1,\dots,S_n \mid S_{n+1}=1), \qquad (4.29)$$

i.e., the distribution of the vector of uniform order statistics coincides with the conditional distribution of the vector of sums $S_1,\dots,S_n$ given that $S_{n+1}=1$. Prove representation (4.29).

Example 4.7. Let a stick of the unit length be randomly broken in n places. We get (n+1) pieces of the stick. Let us arrange these pieces of the stick in increasing order with respect to their lengths. What is the distribution of the k-th item in this variational series? The given construction deals with two orderings. Coordinates of the break points coincide with the uniform order statistics $U_{1,n},U_{2,n},\dots,U_{n,n}$. Hence, the lengths of random pieces are

$$\delta_1= U_{1,n},\ \delta_2=U_{2,n}-U_{1,n},\dots,\ \delta_n= U_{n,n}-U_{n-1,n}\ ,\ \delta_{n+1}=1-U_{n,n}.$$

Let

$$\delta_{1,n+1}\le\delta_{2,n+1}\le\ldots\le\delta_{n+1,n+1}$$

denote order statistics based on $\delta_1, \delta_2,\ldots, \delta_{n+1}$. From (4.28) we find that

$$(\delta_1,\delta_2,\ldots,\delta_{n+1}) \overset{d}{=} \left(\frac{v_1}{v_1+\ldots+v_{n+1}},\ldots,\frac{v_{n+1}}{v_1+\ldots+v_{n+1}}\right). \quad . \qquad (4.30)$$

The ordering of $\delta_1,\delta_2,\ldots,\delta_{n+1}$ is equivalent to ordering of the exponential random variables $\nu_1, \nu_2,\ldots,\nu_{n+1}$. Hence, we obtain that

$$(\delta_{1,n+1},\delta_{2,n+1},\ldots,\delta_{n+1,n+1}) \overset{d}{=} \left(\frac{v_{1,n+1}}{v_1+\ldots+v_{n+1}},\ldots,\frac{v_{n+1,n+1}}{v_1+\ldots+v_{n+1}}\right). \qquad (4.31)$$

Recalling that

$$\nu_{1,n+1} + \nu_{2,n+1}+\ldots+ \nu_{n+1,n+1}= \nu_1+\nu_2+\ldots+\nu_{n+1},$$

one can express the RHS of (4.31) as

$$\left(\frac{v_{1,n+1}}{v_{1,n+1}+\ldots+v_{n+1,n+1}},\ldots,\frac{v_{n+1,n+1}}{v_{1,n+1}+\ldots+v_{n+1,n+1}}\right).$$

Now we can apply representation (4.15) to exponential order statistics

$$\nu_{1,n+1}, \nu_{2,n+1},\ldots,\nu_{n+1,n+1} .$$

Making some natural changing in (4.15) one comes to the appropriate result for ordered lengths of the pieces of the broken stick. It turns out that

$$\delta_{k,n+1} \overset{d}{=} \left(\frac{v_1}{n+1}+\frac{v_2}{n}+\ldots+\frac{v_k}{n-k+2}\right)/(\nu_1+\ldots+\nu_{n+1}),\ k=1,2,\ldots,n+1, \qquad (4.32)$$

where $\nu_1,\ldots,\nu_{n+1}$ are independent random variables having the standard E(1) exponential distribution. In particular,

$$\delta_{1,n+1} \overset{d}{=} \frac{v_1}{(n+1)(v_1+\ldots+v_{n+1})}. \qquad (4.33)$$

This means that $(n+1)\delta_{1,n+1}$ has the same distribution as $U_{1,n}$.

Remark 4.7. Combining representations (4.6), (4.8), (4.15), (4.21), (4.24), (4.29) one can successfully express distributions of arbitrary order statistics $X_{k,n}$ (related to a some d.f. F) in terms of distributions for sums or products of independent random variables. For instance, if G is the inverse of F and ν_1, ν_2,… are independent exponentially E(1) distributed random variables then

$$X_{k,n} \overset{d}{=} G\left(\frac{v_1+\ldots+v_k}{v_1+\ldots+v_{n+1}}\right) \overset{d}{=}$$

$$G\left(1-\exp\left(-\left(\frac{v_1}{n}+\frac{v_2}{n-1}+\ldots+\frac{v_k}{n-k+1}\right)\right)\right), k=1,2,\ldots,n. \qquad (4.34)$$

Exercise 4.9. Let $X_{1,n},\ldots,X_{n,n}$ be order statistics corresponding to the distribution with the density

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

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

Check your solutions

Exercise 4.1 **(solution)**. It is clear that $0\le Y\le 1$. Fix any $z\in(0,1)$. We have

$$P\{Y\le z\}=P\{F(X)\le z\}.$$

Since F is continuous, the equality F(x) = z has one solution (denote it α_z) at least. There are two possible options:

a) F strongly increases at α_z. Then $G(z)=\alpha_z$. Evidently, in this case events

$$\{\omega:F(X(\omega))\le z\}$$

and

$$\{\omega:X(\omega)\le G(z)\}=\{\omega:X(\omega)\le\alpha_z\}$$

have the same probability. Noting that

$$P\{\omega:X(\omega)\le\alpha_z\}=F(\alpha_z)=z,$$

one gets that

$$P\{Y\le z\}=P\{X\le G(z)\} = F(G(z))=z.$$

b) α_z belongs to some constancy interval [*a,b*] of d.f. F, where

$$a= \inf\{x:F(x)\geq z\}$$

and

$$b=\sup\{x:F(x) \leq z\}.$$

Then

$$F(a)=F(\alpha_z)=F(b)=z$$

and

$$G(z)= a.$$

Indeed, events $\{F(X) \leq z\}$ and $\{X\leq G(z)\}=\{X\leq a\}$ also have the same probability and

$$P\{Y\leq z\}=P\{X\leq G(z)\}=F(G(z))=F(a)= F(\alpha_z)=z.$$

Thus, both necessary statements are proved.

***Exercise 4.2* (answer)**. $G(x)=1$, if $0<x\leq 1/3$, $G(x)=2$, if $1/3<x\leq 2/3$, and $G(x)=3$, if $2/3<x<1$; random variable $G(U)$ has the same distribution as X.

***Exercise 4.3* (solution)**. Note that Z_1 and Z_2 have d.f.'s

$$F_1(x)=1-e^{-x/\lambda}, x\geq 0,$$

and

$$F_2(x)=1-e^{-x/\mu}, x\geq 0,$$

correspondingly. One can see that

$$P\{\min\{ Z_1,Z_2\}>x, \max\{ Z_1,Z_2\}\leq y\} = P\{x< Z_1\leq y, x< Z_2\leq y\}=$$

$$(F_1(y)-F_1(x)) (F_2(y)-F_2(x))= (e^{-x/\lambda} -e^{-y/\lambda}) (e^{-x/\mu} -e^{-y/\mu}), \quad 0<x<y.$$

By differentiating over x and y we get that that the joint probability density function $g(x,y)$ of $\min\{ Z_1,Z_2\}$ and $\max\{ Z_1,Z_2\}$ has the form

$$g(x,y)=(\exp(-\frac{x}{\lambda}-\frac{y}{\mu})+\exp(-\frac{x}{\mu}-\frac{y}{\lambda}))/\lambda\mu, \quad 0<x<y.$$

Consider the linear change of variables (v,w) = (x,y-x) with the unit Jacobian, which corresponds to the passage to random variables V and W>0. This transformation gives us the joint pdf of random variables V and W:

$$f_{V,W}(v,w) = (\exp(-\frac{v}{\lambda} - \frac{v+w}{\mu}) + \exp(-\frac{v}{\mu} - \frac{v+w}{\lambda}))/\lambda\mu,\ v>0, w>0.$$

One sees that

$$f_{V,W}(v,w) = h_1(v)h_2(w)/\lambda\mu,\ v>0, w>0,$$

where

$$h_1(v)=\exp(-v(\frac{1}{\lambda}+\frac{1}{\mu}))$$

and

$$h_2(w)=(\exp(-\frac{w}{\mu})+\exp(-\frac{w}{\lambda})).$$

The existence of the factorization

$$f_{V,W}(v,w) = h_1(v)h_2(w)/\lambda\mu$$

enables us to state that random variables V and W are independent. Moreover, we can see that the functions

$$r_1(v)=\frac{\lambda+\mu}{\lambda\mu} h_1(v) = \frac{\lambda+\mu}{\lambda\mu} \exp(-\frac{\lambda+\mu}{\lambda\mu} v),\ v>0,$$

and

$$r_2(w)=\frac{1}{\lambda+\mu}\ h_2(w)= \frac{1}{\lambda+\mu} (\exp(-\frac{w}{\mu})+\exp(-\frac{w}{\lambda})),\ w>0$$

present here the probability density functions of random variables V and W respectively.

***Exercise 4.4* (solution)**. It follows from (4.15) that

$$(Z_{1,n}, L) \overset{d}{=} (\frac{\nu_1}{n}, \sum_{k=2}^{n} b_k\nu_k),$$

where

$$b_k = \frac{1}{n-k+1} \sum_{j=k}^{n} c_j.$$

Since

$$\nu_k,\ k=1,2,\dots,n,$$

are independent random variables,

$$\frac{\nu_1}{n} \quad \text{and} \quad \sum_{k=2}^{n} b_k \nu_k$$

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

Exercise 4.5 **(solution)**. The statement of exercise 4.5 immediately follows from (4.20), so far as simple transformations enable us to get the equality

$$(V_1,V_2,\dots,V_n) \overset{d}{=} (W_1,W_2,\dots,W_n),$$

where W's are independent uniformly distributed random variables.

Exercise 4.6 **(solution)**. It suffices to understand that *m* components

$$(W_1 W_2^{1/2} \dots W_{n-1}^{1/(n-1)} W_n^{1/n}, \dots, W_m^{1/m} \dots W_{n-1}^{1/(n-1)} W_n^{1/n})$$

of vector (4.20), which correspond to order statistics $(U_{1,n},\dots,U_{m,n})$, can be given in the form

$$(W_1 W_2^{1/2} \dots W_{n-1}^{1/(n-1)} W_m^{1/m} T, \dots, W_m^{1/m} T),$$

where

$$T = W_{m+1}^{1/(m+1)} \dots W_{n-1}^{1/(n-1)} W_n^{1/n}$$

does not depend on $W_1,W_2,\dots,W_m$. Noting that vector

$$(W_1 W_2^{1/2} \dots W_{n-1}^{1/(n-1)} W_m^{1/m}, \dots, W_m^{1/m})$$

has the same distribution as $(U_{1,n},\dots,U_{m,n})$, while T is distributed as $V_{m+1,n}$, we complete the solution of the exercise.

Exercise 4.7 **(solution)**. From (4.24) we obtain that $U_{n-k+1,n} - U_{k,n}$ has the same distribution as the ratio

$$\frac{v_{k+1}+v_{k+2}+\ldots+v_{n-k+1}}{v_1+\ldots+v_{n+1}}.$$

It is evident that this ratio and ratio

$$\frac{v_1+v_2+\ldots+v_{n-2k+1}}{v_1+\ldots+v_{n+1}},$$

which corresponds to $U_{n-2k+1,n}$, also have the same distribution.

***Exercise 4.8* (solution)**. The joint pdf of random variables $\nu_1,\nu_2,\ldots\ \nu_{n+1}$ is given as

$$p_{n+1}(x_1,x_2,\ldots,x_{n+1}) = \exp(-(x_1+x_2+\ldots+x_{n+1})),\ x_1>0,x_2>0,\ldots,x_{n+1}>0.$$

By means of the linear transformation

$$(y_1,y_2,\ldots,y_{n+1}) = (x_1,x_1+x_2,\ldots,x_1+x_2+\ldots+x_{n+1})$$

with a unit Jacobian one can get the following joint pdf for sums $S_1,S_2,\ldots,S_{n+1}$:

$$g_{n+1}(y_1,y_2,\ldots,y_{n+1}) = \exp(-y_{n+1}),\ 0<y_1<y_2<\ldots<y_{n+1}.$$

It is well known that any sum S_{n+1} of (n+1) independent exponentially E(1) distributed terms has the gamma distribution with parameter n+1:

$$h_{n+1}(v) = v^n\exp(-v)/n!,\ v>0.$$

To obtain the conditional density function of sums $S_1,S_2,\ldots,S_n$, given that $S_{n+1}=1$, one has to calculate the expression

$$g_{n+1}(y_1,y_2,\ldots,y_n,1)/h_{n+1}(1),$$

which evidently equals n!, if $0<y_1<y_2<\ldots<y_n$, and equals zero , otherwise. This expression coincides with the joint pdf of the uniform order statistics given in (4.25). It means that representation (4.29) is true.

***Exercise 4.9* (solution)**. One obtains that the corresponding d.f. in this exercise has the form

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

Hence, the inverse function is given as

$$G(x) = x^{1/\alpha},\ 0<x<1,$$

and relations

$$X_{r,n} \overset{d}{=} (U_{r,n})^{1/\alpha}$$

and

$$X_{r,n}X_{s,n} \overset{d}{=} (U_{r,n}\, U_{s,n})^{1/a}$$

are valid. Now we use representation (4.21) and have the following equalities:

$$X_{r,n} \overset{d}{=} W_r^{1/ra} Wr_{+1}^{1/(r+1)a} \ldots W_n^{1/na},\ r=1,2,\ldots,n,$$

and

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

where $W_1, W_2, \ldots, W_n$ are independent random variables with a common uniform on [0,1] distribution.

Chapter 5. Conditional distributions of order statistics Условные распределения порядковых статистик

conditional distributions = условные распределения

Markov property = марковское свойство

There are some useful relations related to conditional distributions of order statistics.

Example 5.1. Let $X_{1,n} \le \ldots \le X_{r-1,n} \le X_{r,n} \le X_{r+1,n} \le \ldots \le X_{n,n}$ be a variational series corresponding to a distribution with a density function f(x). Fix a value of $X_{r,n}$ and consider the conditional distribution of the rest elements of the variational series given that $X_{r,n}$=v. We suppose that $f_{r:n}(v)>0$ for this value v, where $f_{r:n}(v)$, as usual, denotes the pdf of $X_{r,n}$. Let

$$f_{1,\ldots,r-1,r+1,\ldots,n|r}(x_1,\ldots,x_{r-1,}x_{r+1,\ldots,}x_n|v)$$

denote the joint conditional density of order statistics $X_{1,n},\ldots, X_{r-1,n}, X_{r+1,n},\ldots,X_{n,n}$ given that $X_{r,n}$=v. The definition of the conditional densities requires to know the (unconditional) joint pdf of all random variables (including fixed) and separately the pdf of the set of fixed random variables. In our case these pdf's are $f_{1,2,\ldots,n:n,}$ the joint pdf of all order statistics $X_{1,n},\ldots,X_{n,n}$, and $f_{r:n}$, the pdf of $X_{r,n}$. The standard procedure gives us the required density function:

$$f_{1,\ldots,r-1,r+1,\ldots,n|r}(x_1,\ldots,x_{r-1,}x_{r+1,\ldots,}x_n|v) =$$

$$f_{1,2,\ldots,n:n}(x_1,\ldots,x_{r-1,}v,x_{r+1,\ldots,}x_n)/ f_{r:n}(v). \qquad (5.1)$$

Upon substituting (2.10) and (2.13) in (5.1), we get that

$$f_{1,\ldots,r-1,r+1,\ldots,n|r}(x_1,\ldots,x_{r-1,}x_{r+1,\ldots,}x_n|v) =$$

$$(r-1)!\prod_{k=1}^{r-1}\frac{f(x_k)}{F(v)}\ (n-r)!\prod_{k=r+1}^{n}\frac{f(x_k)}{1-F(v)}, \qquad (5.2)$$

if $x_1<\ldots<x_{r-1}<v<x_{r+1}<\ldots<x_n$, and the LHS of (5.2) equals zero otherwise.

For each value *v* we introduce d.f.'s

$$G(x,v)=P\{X\le x|X\le v\}=F(x)/F(v),\ \ x\le v, \qquad (5.3)$$

and

$$H(x,v)=P\{X\le x|X>v\}=(F(x)-F(v))/(1-F(v)),\ \ x>v. \qquad (5.4)$$

The corresponding densities have the form

$$g(x,v)=f(x)/F(v),\ x\leq v, \text{ and } g(x,v)=0, \text{ if } x>v; \qquad (5.5)$$

and

$$h(x,v)=f(x)/(1-F(v)),\ x>v, \text{ and } h(x,v)=0, \text{ if } x\leq v. \qquad (5.6)$$

Now (5.2) can be expressed as follows:

$$f_{1,\dots,r-1,r+1,\dots,n|r}(x_1,\dots,x_{r-1},x_{r+1,\dots,}x_n|v) = g(x_1,\dots,x_{r-1},v)h(x_{r+1},\dots,x_n,v), \qquad (5.7)$$

where

$$g(x_1,\dots,x_{r-1},v)=(r-1)!g(x_1,v)\dots g(x_{r-1},v), \text{ if } x_1<\dots<x_{r-1}<v,$$

and

$$g(x_1,\dots,x_{r-1},v)=0, \quad \text{otherwise},$$

while

$$h(x_{r+1},\dots,x_n) = (n-r)!g(x_{k+1},v)\dots g(x_n,v), \text{ if } v<x_{r+1}<\dots<x_n,$$

and

$$h(x_{r+1},\dots,x_n) = 0, \text{ otherwise}.$$

We immediately derive from (5.7) that two sets of order statistics, $(X_{1,n},\dots,X_{r-1,n})$ and $(X_{r+1,n},\dots,X_{n,n})$ are conditionally independent given any fixed value of $X_{r,n}$. Observing the form of $g(x_1,\dots,x_{r-1},v)$ one can see that this conditional joint pdf, corresponding to order statistics $X_{1,n},\dots,X_{r-1,n}$, coincides with unconditional joint pdf of order statistics, say, $Y_{1,r-1}\leq\dots\leq Y_{r-1,r-1}$, corresponding to a population with d.f. G(x,v) and density g(x,v). This assertion we will write in the following form, where F(x) and G(x,v) denote population distribution functions:

$$\{F(x);\ X_{1,n},\dots,X_{r-1,n}|X_{r,n}=v\} \overset{d}{=} \{G(x,v);\ Y_{1,r-1},\dots,Y_{r-1,r-1}\}. \qquad (5.8)$$

Similarly we obtain that the conditional distribution of order statistics $X_{r+1,n},\dots,X_{n,n}$, given that $X_{r,n}=v$, coincides with the unconditional distribution of order statistics, say, $W_{1,n-r}\leq\dots\leq W_{n-r,n-r}$ related to d.f. H(x,v) and pdf h(x,v):

$$\{F(x);\ X_{r+1,n},\dots,X_{n,n}|X_{r,n}=v\} \overset{d}{=} \{H(x,v);\ W_{1,n-r},\dots W_{n-r,n-r}\}. \qquad (5.9)$$

Let $Y_1,Y_2,\dots,Y_{r-1}$ be a sample of size (r-1) from a population with d.f. G(x,v). Then the following corollary of (5.8) is valid:

$$\{F(x);\ X_{1,n}+\dots+X_{r-1,n}\ |X_{r,n}=v\} \overset{d}{=} \{G(x,v);\ Y_{1,r-1}+\dots+Y_{r-1,r-1}\} \overset{d}{=}$$

$$\{G(x,v);\ Y_1+\dots+Y_{r-1}\}. \qquad (5.10)$$

Relation (5.10) enable us to express the distribution of the sum $X_{1,n}+\ldots+X_{r-1,n}$ as a mixture (taken over the parameter v) of (r-1)-fold convolutions of d.f.'s G(x,v). Indeed, the similar corollary is valid for the sum $X_{r+1,n}+\ldots+X_{n,n}$.

Exercise 5.1. What is the structure of the conditional distribution of order statistics

$$X_{1,n},\ldots,X_{r-1,n};\ X_{r+1,n},\ldots,X_{s-1,n} \text{ and } X_{s+1,n},\ldots,X_{n,n},$$

given that two order statistics $X_{r,n}=v$ and $X_{s,n}=z$, $r<s$, $v<z$. are fixed?

Exercise 5.2. Show that the conditional distribution of the uniform order statistics $(U_{1,n},\ldots,U_{k,n})$, given that $U_{k+1,n}=v$, coincides with the unconditional distribution of the vector $(vU_{1,k},\ldots,vU_{k,k})$.

Exercise 5.3. Show that the conditional distribution of the uniform quasi-range

$U_{n-k+1,n}-U_{k,n}$, given that $U_{n,n}-U_{1,n} = v$, coincides with the unconditional distribution of

$v(U_{n-k,n-2}-U_{k-1,n-2})$.

Exercise 5.4. Prove that the conditional distribution of the exponential order statistics $(Z_{r+1,n},\ldots,Z_{n.n})$, given that $Z_{r,n}=v$, coincides with the unconditional distribution of the vector

$(v+Z_{1,n-r},\ldots,v+Z_{n-r,n-r})$.

Remark 5.1. The assertion obtained in example 5.1 stays valid for any continuous distribution function *F*. The following exercise shows that the conditional independence property can fail if d.f. *F* has jump points.

Exercise 5.5. Let X have an atom $P\{X=a\}=p>0$ at some point a . Let also

$$P\{X< a\} =F(a) -p >0 \text{ and } P\{X>a\} = 1-F(a)>0.$$

Consider order statistics $X_{1,3}\le X_{2,3}\le X_{3,3}$ and show that $X_{1,3}$ and $X_{3,3}$ are conditionally dependent, given that $X_{2,3}=a$.

Remark 5.2. We see that conditional independence of order statistics

$$X_{1,n}\le\ldots\le X_{r-1,n} \text{ and } X_{r+1,n}\le\ldots\le X_{n,n},$$

given that $X_{r,n}$ is fixed, can fail if there is a positive probability that two neighboring elements of the variational series coincide. The conditional independence property can be saved under additional condition that $X_{1,n}< X_{2,n} < \ldots< X_{n,n}$. The next exercise illustrates this fact.

Exercise 5.6. Let X take on values x_1, x_2, … with non-zero probabilities and let $X_{1,n},X_{2,n}, \ldots,X_{n,n}$ be the corresponding order statistics. We suppose that a number of possible values of X is not less than n. Show that vectors

$$(X_{1,n},\ldots,X_{r-1,n}) \text{ and } (X_{r+1,n},\ldots,X_{n,n})$$

are conditionally independent, given that $X_{r,n}=v$ is fixed and the relation $X_{1,n}<X_{2,n}<\ldots<X_{n,n}$ holds. Indeed, the value *v* is chosen here to satisfy the condition

$$P\{X_{1,n}<\ldots<X_{r-1,n}<X_{r,n}=v<X_{r+1,n}<\ldots<X_{n,n}\}>0.$$

Now we will discuss the Markov property of order statistics. We consider a sequence of order statistics $X_{1,n},X_{2,n},\ldots,X_{n,n}$, corresponding to a population with a density function *f.*

Example 5.2. From (5.7) we found that under fixed value of $X_{r,n}$ order statistics $X_{r+1,n},\ldots,X_{n,n}$ are conditionally independent on random variables $X_{1,n},\ldots,X_{r-1,n}$. The same arguments are used to check the Markov property of order statistics. We need to prove that for any r=3,4,…,n, the conditional density

$$f_{r|1,2,\ldots,r-1}(u|x_1,\ldots,x_{r-1})$$

of $X_{r,n}$, given that $X_{1,n}=x_1,\ldots,X_{r-1,n}=x_{r-1}$, coincides with the conditional density $f_{r|r-1}(u|x_{r-1})$ of $X_{r,n}$, given only that $X_{r-1,n}=x_{r-1}$ is fixed. Recalling the definitions of conditional densities, one finds that

$$f_{r|1,2,\ldots,r-1}(u|x_1,\ldots,x_{r-1})=f_{1,2,\ldots,r-1,r:n}(x_1,x_2,\ldots,x_{r-1},u)/f_{1,2,\ldots,r-1::n}(x_1,x_2,\ldots,x_{r-1})=$$

$$f_{1,2,\ldots,r-1|r}(x_1,x_2,\ldots,x_{r-1}|u)f_{r:n}(u)/f_{1,2,\ldots,r-2|r-1}(x_1,x_2,\ldots,x_{r-1}|x_{r-1})f_{r-1,n}(x_{r-1}). \tag{5.10}$$

In (5.10) $f_{1,2,\ldots,r-1|r}(x_1,x_2,\ldots,x_{r-1}|u)$ is the joint conditional density of $X_{1,n},\ldots,X_{r-1,n}$, given that $X_{r,n}=u$, and it coincides, as we know, with

$$g(x_1,\ldots,x_{r-1},u)=(r-1)!g(x_1,u)\ldots g(x_{r-1},u),$$

determined in (5.7), where $g(x,u)=f(x)/F(u),\ x<u.$

From (2.13) we also know that

$$f_{r:n}(x)=\frac{n!}{(r-1)!(n-r)!}(F(x))^{r-1}(1-F(x))^{n-r}f(x).$$

The similar expressions (by changing *r* for *r-1*) can be written for the rest terms on the RHS of (5.10). Substituting all these expressions in (5.10) one obtains that

$$f_{r|1,2,\ldots,r-1}(u|x_1,\ldots,x_{r-1})=(n-r+1)(1-F(u))^{n-r}f(u)/(1-F(x_{r-1}))^{n-r+1},\ u>x_{r-1}. \tag{5.11}$$

It suffices to obtain that

$$f_{r|r-1}(u|x_{r-1})=f_{r-1,r:n}(x_{r-1},u)/f_{r-1:n}(x_{r-1})$$

coincides with the expression on the RHS of (5.11). Due to relations (2.13) and (2.14), which give us density functions $f_{r-1:n}(x_{r-1})$ and $f_{r-1,r:n}(x_{r-1},u)$, we easily prove the desired statement.

Exercise 5.7. Give another proof of (5.11), based on the equality

$$f_{r|1,2,\ldots,r-1}(u|x_1,\ldots,x_{r-1})=f_{1,2,\ldots,r-1,r:n}(x_1,x_2,\ldots,x_{r-1},u)/f_{1,2,\ldots,r-1,n}(x_1,x_2,\ldots,x_{r-1}).$$

Remark 5.3. It follows from example 5.2 that a sequence $X_{1,n},\dots,X_{n,n}$ forms a Markov chain and

$$P\{X_{k+1,n}>x\,|\,X_{k,n}=u\} = ((1-F(x))/(1-F(u)))^{n-k},\ x>u, \qquad (5.12)$$

for any k=1,2,…,n-1. Note also that we assumed that the underlying distribution has a density function only for the sake of simplicity. In fact, order statistics satisfies the Markov property in a more general situation, when a population has any continuous d.f. *F*, while this property can fail if *F* has some jump points.

Example 5.3. In exercise 5.5 we considered a distribution having an atom $P\{X=a\}=p>0$ at some point a and supposed that $b=P\{X< a\} = F(a)-p >0$ and $c=P\{X>a\}=1-F(a)>0$. It turns out in this situation that $X_{1,3}$ and $X_{3,3}$ are conditionally dependent, given that $X_{2,3}=a$. We can propose that order statistics $X_{1,3}$, $X_{2,3}$ and $X_{3,3}$ can not possess the Markov structure for such distributions. In fact, in this case

$$P\{X_{1,3}= a, X_{2,3}=a\}=p^3+3p^2c, \qquad (5.13)$$

$$P\{ X_{1,3}= a, X_{2,3}=a, X_{3,3}= a\}=p^3, \qquad (5.14)$$

$$P\{X_{2,3}=a, X_{3,3}= a\}=p^3+3p^2b \qquad (5.15)$$

and

$$P\{X_{2,3}=a\}=p^3+3p^2(1-p)+6pbc. \qquad (5.16)$$

From (5.13) and (5.14) we obtain that

$$P\{ X_{3,3}= a\,|\, X_{2,3}=a, X_{1,3}= a\}=p/(p+3c), \qquad (5.17)$$

while , due to (5.15) and (5.16),

$$P\{ X_{3,3}= a\,|\, X_{2,3}=a\}=(p^2+3pb)/(p^2+3p(1-p)+6bc). \qquad (5.18)$$

It is not difficult to check now that equality

$$P\{ X_{3,3}= a\,|\, X_{2,3}=a, X_{1,3}= a\}= P\{ X_{3,3}= a\,|\, X_{2,3}=a\}$$

is equivalent to the relation $bc = 0$, which fails so far as $b>0$ and $c>0$ for considered distributions.

Remark 5.4. It follows from example 5.3 that the Markov property is not valid for order statistics if the underlying d.f. has three jump points or more, because in this situation there exists one point a at least such that $P\{X= a\}>0$, $P\{X<a\}>0$ and $P\{X>a\}>0$. Hence it remains to discuss distributions with one or two atoms, which coincide with end-points α=inf{x:F(x)>0} or/and β=sup{x:F(x)<1} of the distribution support.

Example 5.5. Let X have degenerate distribution, say P{X=*a*}=1. Then it is evident that order statistics form the Markov property.

Exercise 5.8. Let X take on two values *a<b* with probabilities

0<p=P{X= *a*}=1-P{X=*b*}<1.

Show that order statistics $X_{1,n},\ldots,X_{n,n}$ form a Markov chain.

Example 5.6. Let $X_{1,n}\le\ldots\le X_{r-1,n}\le X_{r,n}\le X_{r+1,n}\le\ldots\le X_{n,n}$ be a variational series corresponding to a distribution with the density f(x)= $e^{-x}, x\ge 0.$ Then the conditional pdf of $(X_{r+1,n}|X_{1,n},\ldots,X_{r,n})$ is given as

$$f_{r+1|1,2,\ldots,r}(x_{r+1}|\ x_{1,n}=x_1,\ldots,x_{r,n}=x_r) = (n-r)e^{-(n-r)(x_{r+1}-x_r)}, \quad 0\le x_r\le x_{r+1}<\infty.$$

If D_r = (n-r+1) $(X_{r,n}-X_{r-1,n})$, then D_r, r=1, 2,…,n, with $X_{0,n}=0$, are independent and identically distributed as exponential with cdf F(x) = $1-e^{-x}, x\ge 0.$

Example 5.7. Let $X_{1,n}$ be the smallest order statistics in a sample of size n from Weibull distribution with cdf F(x) $=1-e^{-x^c}, c>0, x\ge 0.$ The cdf $F_{1,n}(x)$ of $X_{1,n}$ can be written as

$$F_{1,n}(x)=1-(1-F(x))^n$$
$$=1-e^{-nx^c}, x\ge 0.$$

Thus $X_{1,n}$ is also distributed as Weibull distribution.

Remark 5.5. If c=1, then Weibull distribution coincides with the exponential distribution. Thus order statistic $X_{1,n}$ for the exponential distribution is also distributed as exponential.

Check your solutions

***Exercise 5.1* (answer)**. We have three conditionally independent sets of order statistics. Analogously to the case from example 5.1 the following relations (similar to (5.9) and (5.10)) are valid:

$$\{F(x); X_{1,n},\dots,X_{r-1,n} \mid X_{r,n}=v, X_{s,n}=z\} \overset{d}{=} \{G(x,v); Y_{1,r-1},\dots,Y_{r-1,r-1}\},$$

$$\{F(x); X_{r+1,n},\dots,X_{s-1,n} \mid X_{r,n}=v, X_{s,n}=z\} \overset{d}{=} \{T(x,v,z); V_{1,s-r-1},\dots, V_{s-r-1,s-r-1}\}$$

and

$$\{F(x); X_{s+1,n},\dots,X_{n-s,n} \mid X_{r,n}=v, X_{s,n}=z\} \overset{d}{=} \{H(x,z); W_{1,n-s},\dots,W_{n-s,n-s}\},$$

where $Y_{1,r-1},\dots,Y_{r-1,r-1}$ correspond to d.f.

$$G(x,v)=P\{X\le x \mid X\le v\}=F(x)/F(v),\ x\le v;$$

$W_{1,n-s},\dots,W_{n-s,n-s}$ are order statistics related to d.f.

$$H(x,z)=P\{X\le x \mid X>z\}=(F(x)-F(z))/(1-F(z)),\ x>z,$$

and order statistics $V_{1,s-r-1}\le\dots\le V_{s-r-1,s-r-1}$ correspond to d.f.

$$T(x,v,z)= P\{X\le x \mid v<X\le z\}=(F(x)-F(v))/(F(z)-F(v)),\ v<x<z.$$

Exercise 5.2 **(hint).** It suffices to use relation (5.8) with the d.f.

$$G(x,v)=x/v,\ 0<x<v,$$

which corresponds to the random variable vU, where U has the standard uniform distribution.

Exercise 5.3 **(hint).** Use the statements and results of exercises 5.1 and 5.2. Consider the conditional distribution of the uniform order statistics $U_{n-k+1,n}$ and $U_{k,n}$, given that $U_{1,n}=v$ and $U_{n,n}=z$. You will find that

$$(U_{k,n},U_{n-k+1,n} \mid U_{1,n}=v , U_{n,n}=z) \overset{d}{=} ((z-v)\, U_{k-1,n-2}, (z-v)\, U_{n-k,n-2})$$

and hence

$$(U_{n-k+1,n}-U_{k,n} \mid U_{1,n}=v , U_{n,n}=z) \overset{d}{=} (z-v)\, (U_{n-k,n-2}- U_{k-1,n-2}).$$

Note that really the desired conditional distribution depends on the difference $z-v$ rather than on z and v. Due to this fact one can show that

$$(U_{n-k+1,n}-U_{k,n} \mid U_{n,n} - U_{1,n}=u) \overset{d}{=} u(U_{n-k,n-2}- U_{k-1,n-2}).$$

Exercise 5.4 **(hint).** Show that in this situation the d.f. $H(x,v)$ from (5.9) corresponds to the random variable $Z+v$, where Z has the exponential $E(1)$ distribution.

Exercise 5.5 **(solution)**. Consider conditional probabilities

$$p_1=P\{X_{1,3}=a,\ X_{3,3}=a\,|\,X_{2,3}=a\}=P\{X_{1,3}=a,\ X_{2,3}=a, X_{3,3}=a\}/\ P\{X_{2,3}=a\}=$$

$$P\{X_1=a, X_2=a,\ X_3=a\}/\ P\{X_{2,3}=a\}=p^3/P\{X_{2,3}=a\},$$

$$p_2=P\{X_{1,3}=a\ |X_{2,3}=a\} = P\{\ X_{1,3}=a\ ,\ X_{2,3}=a\}/\ P\{X_{2,3}=a\}=$$

$$(p^3+3p^2(1-F(a)))/\ P\{X_{2,3}=a\}$$

and

$$p_3=\ P\{X_{3,3}=a\ |X_{2,3}=a\} = P\{\ X_{2,3}=a\ ,\ X_{3,3}=a\}/\ P\{X_{2,3}=a\}=$$

$$(p^3+3p^2(F(a)-p))/\ P\{X_{2,3}=a\}.$$

To provide the conditional independence of $X_{1,3}$ and $X_{3,3}$ one needs at least to have the equality $p_1=p_2p_3$, which is equivalent to the following relation:

$$(p^3+3p^2(1-F(a)))\ (p^3+3p^2(F(a)-p))=\ p^3\ P\{X_{2,3}=a\}.$$

Note also that

$$P\{X_{2,3}=a\}=p^3+3p^2F(a-0)+3p^2(1-F(a))+6pF(a-0)(1-F(a))=$$

$$p^3+3p^2(1-p)+6p(F(a)-p)(1-F(a))=3p^2-2p^3+6p(F(a)-p)(1-F(a))$$

and hence, the desired equality must have the form

$$(p^3+3p^2(1-F(a)))\ (p^3+3p^2(F(a)-p))=\ p^3\ (3p^2-2p^3+6p(F(a)-p)(1-F(a))).$$

After some natural simplifications one sees that the latter equality is valid only if

$$(F(a)-p)(1-F(a))=0,$$

but both possible solutions $F(a)-p=0$ and $1-F(a)=0$ are rejected by restrictions of the exercise. Hence, $X_{1,3}$ and $X_{3,3}$ are conditionally dependent, given that $X_{2,3}=a$.

Exercise 5.6 **(solution).** Let $v_1,v_2,\ldots,v_{r-1}, v, v_{r+1},\ldots,v_n$ be any n values taken from a set $\{x_1,x_2,\ldots\}$. Note that

$$I(v_1,\ldots,v_{r-1},v,v_{r+1,\ldots,}v_n) =$$

$$P\{\ X_{1,n}=v_1,\ldots,X_{r-1,n}=v_{r-1},\ X_{r+1,n}=v_{r+1},\ldots,X_{n,n}=v_n\,|\ X_{r,n}=v,\ X_{1,n}<X_{2,n}<\ldots<X_{n,n}\}=$$

$$P\{\ X_{1,n}=v_1,\ldots,X_{r-1,n}=v_{r-1},\ X_{r,n}=v,\ X_{r+1,n}=x_{r+1},\ldots,X_{n,n}=x_n,\ X_{1,n}<X_{2,n}<\ldots<X_{n,n}\}/\ P\{\ X_{r,n}=v,\ X_{1,n}<X_{2,n}<\ldots<X_{n,n}\}.$$

It is not difficult to see that

$$P\{ X_{1,n}= v_1,\ldots,X_{r-1,n}= v_{r-1}, X_{r,n}= v, X_{r+1,n}= x_{r+1},\ldots,X_{n,n}= x_n, X_{1,n}<X_{2,n} < \ldots<X_{n,n} \}=$$

$$n!P\{X=v\}\prod_{k=1}^{r-1} P\{X = v_k\} \quad \prod_{k=r+1}^{n} P\{X = v_k\},$$

if $v_1<v_2<\ldots<v_{r-1}<v<v_{r+1}<\ldots<v_n$, and this probability equals zero otherwise. Then,

$$P\{ X_{r,n} = v, X_{1,n}<X_{2,n} < \ldots<X_{n,n}\}= n!P\{X_1<\ldots<X_{r-1}<X_r= v<X_{r+1}<\ldots<X_n\}=$$

$$n!P\{X=v\}P\{ X_1<\ldots<X_{r-1}<X_r= v\}P\{ v<X_{r+1}<\ldots<X_n\}.$$

Thus,

$$I(v_1,\ldots,v_{r-1},v,v_{r+1,\ldots,}v_n) = G(v_1,\ldots,v_{r-1},v)H(v,v_{r+1,\ldots,}v_n),$$

where

$$G(v_1,\ldots,v_{r-1},v)= \left(\prod_{k=1}^{r-1} P\{X = v_k\} / P\{ X_1<\ldots<X_{r-1}<X_r=v\}\right),$$

If $v_1<v_2<\ldots<v_{r-1}<v$, and

$$H(v,v_{r+1,\ldots,}v_n) = \left(\prod_{k=r+1}^{n} P\{X = v_k\} / P\{ v<X_{r+1}<\ldots<X_n\}\right),$$

If $v<v_{r+1}<\ldots<v_n$.

The existence of the factorization

$$I(v_1,\ldots,v_{r-1},v,v_{r+1,\ldots,}v_n) = G(v_1,\ldots,v_{r-1},v)H(v,v_{r+1,\ldots,}v_n)$$

provides the conditional independence of order statistics.

Exercise 5.7 **(hint).** The expression for joint density functions

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

presents a particular case of the equalities, which were obtained in exercise 2.7.

Exercise 5.8 **(hint).** Fix any $2<r\leq n$ and compare probabilities

$$P\{ X_{r,n}= x_r \mid X_{r-1,n}= x_{r-1},\ldots, X_{1,n}= x_1\} \quad \text{and} \quad P\{ X_{r,n}= x_r \mid X_{r-1,n}=x_{r-1}\}.$$

Since $x_1,x_2,\ldots,x_r$ is a non-decreasing sequence of zeros and ones, we need to consider only two situations. If $x_{r-1}=0$, then the events

$$\{X_{r-1,n}=0\} \text{ and } \{ X_{r-1,n}=0,\ldots, X_{1,n}= 0\}$$

coincide and then both conditional probabilities are equal. If $x_{r-1}=1$, then $X_{r,n}$ is obliged to be equal 1 with probability one and both conditional probabilities are equal to one.

Chapter 6. Order statistics for discrete distributions
Порядковые статистики для дискретных распределений

Выше были рассмотрены методы и получены различные результаты для порядковых статистик, построенных по наборам случайных величин, имеющих непрерывные функции распределения. Оказывается, что эти методы не всегда позволяют работать в случае исходных дискретных распределений. В этом случае требуются какие-то новые идеи и построения, с которыми читатель может познакомиться в данной главе. Много внимания уделяется порядковым статистикам, соответствующим геометрически распределенным случайным величинам. Выясняется, что такого рода порядковые статистики являются некоторым дискретным аналогом экспоненциальных порядковых статистик.

Теоретический материал сопровождается большим числом упражнений, которые помогают понять, в чем разница между порядковыми статистиками для дискретных и непрерывных распределений.

discrete order statistics = дискретные порядковые статистики (порядковые статистики, соответствующие исходным дискретным распределениям)

samples without replacement = выборки без возвращения

We consider order statistics $X_{1,n} \le X_{2,n} \le \ldots \le X_{n,n}$ based on a sample $X_1,X_2,\ldots,X_n$ for the case, when a population distribution is discrete. Let $X_1,X_2,\ldots,X_n$ be *n* independent observations on random variable X taking on values $x_1,x_2,\ldots$ with probabilities $p_1,p_2,\ldots$.

Exercise 6.1. Show that

$$0 \le p^2 \le P\{X_{1,2} = X_{2,2}\} \le p\ ,$$

where $p=\max\{p_1,p_2,\ldots\}$.

Exercise 6.2. Let X take on values $1,2,\ldots,n$ with probabilities $p_r= 1/n$, $1\le r\le n$. Find probabilities

$$\pi_r= P\{X_{1,r}<X_{2,r}<\ldots<X_{r,r}\}$$

for any r, $2\le r\le n$.

Exercise 6.3. Let X take on two values, 0 and 1, with probabilities p and

g=1-p correspondingly. Find the distribution of $X_{k,n}$.

Exercise 6.4. In the previous exercise find the joint distribution of order statistics $X_{j,n}$ and $X_{k,n}$, $1 \le j<k \le n$, and the distribution of the difference $W_{j,k,n}= X_{k,n}- X_{j,n}$.

Exercise 6.5. Let X have a geometric distribution with probabilities

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

where $0<p<1$. Find distributions of order statistics $X_{1,n}$ and $X_{n,n}$.

Exercise 6.6. Under conditions of exercise 6.3 find the distribution of $X_{2,n}$.

Exercise 6.7. Find the distribution of $Y= \min\{Y_1,Y_2,\ldots,Y_n\}$,

where Y's are independent and have different geometric distributions with probabilities

$$P\{Y_r=k\}=(1-p_r)\, p_r^k, \ k=0,1,\ldots;\ r=1,2,\ldots,n.$$

Exercise 6.8. Consider the case, when X takes on values 0,1,2,… with probabilities $p_0,p_1,p_2 \ldots$. Find expressions for

$$P\{X_{k,n}=r\},\ k=1,2,\ldots,n;\ r=0,1,\ldots.$$

Exercise 6.9. Let X have a geometric distribution with probabilities

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

Find the common distribution of $X_{1,n}$ and $X_{n,n}$.

Exercise 6.10. Under conditions of exercise 6.9 find the distribution of sample ranges

$$W_n=X_{n,n}-X_{1,n},\ n=2,3,\ldots.$$

Exercise 6.11*.* We again consider the geometric distribution from exercises 6.9 and 6.10. Show that then $X_{1,n}$ and W_n are independent for n=2,3,….

Now we consider conditional distributions of discrete order statistics.

Exercise 6.12. Let X take on values $x_1<x_2<\dots$ with probabilities $p_k=P\{X=k\}$. Find conditional probabilities

$$P\{X_{1,3}=x_r,X_{3,3}=x_s|X_{2,3}=x_k\},\quad P\{X_{1,3}=x_r\ |X_{2,3}=x_k\}\quad \text{and}\quad P\{X_{3,3}=x_s|X_{2,3}=x_k\},\ x_r<x_k<x_s.$$

Remark 6.1. It is interesting to check when order statistics $X_{1,3}$ and $X_{3,3}$ are conditionally independent given that $X_{2,3}$ is fixed. In exercise 5.5 we have got that $X_{1,3}$ and $X_{3,3}$ are conditionally dependent for any distribution, which has an atom in some point a ($P\{X=a\}=p>0$), such that

$$P\{X< a\}= F(a)-p>0 \quad \text{and} \quad P\{X>a\}=1-F(a)>0.$$

It means that for any discrete distribution, which has three or more values, order statistics $X_{1,3}$ and $X_{3,3}$ are conditionally dependent. Indeed, if X is degenerate and $P\{X=a\}=1$, then

$$P\{X_{1,3}=a,\ X_{3,3}=a|X_{2,3}=a\} = 1= P\{X_{1,3}=a|X_{2,3}=a\}\ P\{X_{3,3}=a|X_{2,3}=a\}$$

and hence $X_{1,3}$ and $X_{3,3}$ are conditionally independent given that $X_{2,3}$ is fixed. The only case, which we need to investigate now is the situation, when X takes on two values.

Exercise 6.13. Let

$$P\{X=a\}=1-P\{X=b\}=p,$$

where $0<p<1$ and $a<b$. Prove that $X_{1,3}$ and $X_{3,3}$ are conditionally independent given that $X_{2,3}$ is fixed.

Remark 6.2. In lecture 5 we investigated Markov properties of order statistics and found that they do not possess the Markov structure if X has three jump points or more. It was also shown that order statistics form a Markov chain if X is degenerate or if it takes on two values only.

Very often in statistics one uses sampling without replacement from a finite population. Let us have a set of N ordered distinct population values $x_1<x_2<\dots<x_N$. If a sample of size **n** ($n\le N$) is drawn at random without replacement we deal with some dependent identically distributed random variables $(X_1,X_2,\dots,X_n)$. The common distribution of these random variables are given by the equality

$$P\{\ X_1 =x_{k(1)},\ X_2=x_{k(2)},\dots,X_n=x_{k(n)}\} =1/N(N-1)\dots(N-n+1), \tag{6.1}$$

which holds for any group of n distinct values $x_{k(1)},\ x_{k(2)},\dots,x_{k(n)}$ taken from the original set of values $x_1,x_2,\dots,x_N$. By arranging X's in increasing order we come to the corresponding order statistics

$$X_{1,n,N}<X_{2,n,N}<\dots<X_{n,n,N},$$

where **n** denotes the sample size and **N** is the population size. We see now from (6.1) that

$$P\{\ X_{1,n,N}=x_{r(1)},\ X_{2,n,N}=x_{r(2),}\ \dots X_{n,n,N}=x_{r(n)}\}=$$

$$n!/N(N-1)\ldots(N-n+1) = 1/\binom{N}{n}, \qquad (6.2\ )$$

where $x_{r(1)} < x_{r(2)} < \ldots < x_{r(n)}$ are the ordered values $x_{k(1)}, x_{k(2)}, \ldots, x_{k(n)}$.

Simple combinatorial methods enable us to find different distributions of these order statistics.

Example 6.1. Let us find probabilities $p_{k,r} = P\{X_{k,n,N} = x_r\}$ for r=1,2,…,N. It is evident that $p_{k,r}=0$, if $r<k$ or if $r>N-n+k$. Now we will consider the case, when $k\leq r\leq N-n+k$. The event $\{X_{k,n,N} = x_r\}$ assumes that order statistics

$$X_{1,n,N}, X_{2,n,N}, \ldots, X_{k-1,n,N}$$

take on (k-1) arbitrary distinct values, which are less than x_r, while order statistics

$$X_{k+1,n,N} < X_{k+2,n,N} < \ldots < X_{n,n,N}$$

take on (n-k) arbitrary values , which are greater than x_r. The combinatorial arguments show that there are $(r-1)(r-2)\ldots(r-k+1)/(k-1)!$ options to get appropriate values for

$$X_{1,n,N}, X_{2,n,N}, \ldots, X_{k-1,n,N}$$

and $(N-r)(N-r-1)\ldots(N-n-r+k+1)/(n-k)!$ variants for

$$X_{k+1,n,N}, X_{k+2,n,N}, \ldots, X_{n,n,N}.$$

Then it follows from (6.2) that

$$p_{k,r}=(r-1)(r-2)\ldots(r-k+1)\,(N-r)(N-r-1)\ldots(N-n-r+k+1)n!/\,N(N-1)\ldots(N-n+1)(k-1)!(n-k)!=$$

$$\binom{r-1}{k-1}\binom{N-r}{n-k}/\binom{N}{n}, \qquad (6.3)$$

for any $1\leq k\leq n$ and $k\leq r\leq N-n+k$.

Exercise 6.14. Find the joint distribution of order statistics $X_{i,n,N}$ and $X_{j,n,N}$ for $1\leq i<j\leq n\leq N$.

Exercise 6.15. Find the joint distribution of order statistics

$$X_{1,n,N},\ X_{2,n,N}, \ldots, X_{k,n,N}.$$

Some interesting results can be obtained for conditional distributions of order statistics $X_{i,n,N.}$.

Example 6.2. Consider the conditional probabilities

$$P\{X_{1,n,N}=x_{r(1)},\ X_{2,n,N}=x_{r(2),\ldots,}X_{k-1,n,N}=x_{r(k-1)}\,|\ X_{k,n,N}=x_{r(k)}\}.$$

Indeed, these probabilities are defined if $k \leq r(k) \leq N-n+k$, when

$$P\{X_{r,n,N}=x_{k(r)}\}>0,$$

and they are positive, if $1 \leq r(1)<r(2)<\ldots<r(k-1)<r(k)$. From example 6.1 we know that

$$P\{ X_{k,n,N}=x_{r(k)}\} = \binom{r(k)-1}{k-1}\binom{N-r(k)}{n-k}/\binom{N}{n} \qquad (6.4\ )$$

and it follows from exercise 6.15 that

$$P\{X_{1,n,N}=x_{r(1)}, X_{2,n,N}=x_{r(2),\ldots,}X_{k-1,n,N}=x_{r(k-1)}, X_{k,n,N}=x_{r(k)}\}=$$

$$\binom{N-r(k)}{n-k}/\binom{N}{n}.$$

Hence,

$$P\{X_{1,n,N}=x_{r(1)}, X_{2,n,N}=x_{r(2),\ldots,}X_{k-1,n,N}=x_{r(k-1)}\,|\, X_{k,n,N}=x_{r(k)}\}=$$

$$P\{X_{1,n,N}=x_{r(1)}, X_{2,n,N}=x_{r(2),\ldots,}X_{k-1,n,N}=x_{r(k-1)}, X_{k,n,N}=x_{r(k)}\}/ P\{ X_{r,n,N}=X_{r(k)}\}=$$

$$1/\binom{r(k)-1}{k-1}.$$

Comparing with (6.2), we see that the conditional distribution of order statistics

$$X_{1,n,N}, X_{2,n,N,\ldots,}X_{k-1,n,N},$$

given that $X_{k,n,N}=r$, coincides with the unconditional distribution of order statistics

$$X_{1,k-1,r-1},\ldots, X_{k-1,k-1,r-1}\,,$$

which correspond to sampling without replacement from the set of population values $\{x_1,x_2,\ldots,x_{r-1}\}$.

The similar arguments show that the conditional distribution of

$$X_{k+1,n,N}, X_{k+2,n,N,\ldots,}X_{n,n,N},$$

given that $X_{k,n,N}=r$, coincides with the unconditional distribution of order statistics

$$Y_{1,n-k,N-r},\ldots, Y_{n-k,n-k,N-r}.$$

Moreover, it can be shown that vectors

$$(X_{1,n,N}, X_{2,n,N,\ldots,}X_{k-1,n,N}) \text{ and } (X_{k+1,n,N}, X_{k+2,n,N,\ldots,}X_{n,n,N})$$

are conditionally independent given that $X_{k,n,N}$ is fixed. The simplest case of this statement is considered in the next exercise.

Exercise 6.16. Show that for any $2 \le r \le N-1$ order statistics $X_{1,3,N}$ and $X_{3,3,N}$ are conditionally independent given that $X_{2,3,N}=x_r$.

Example 6.3. Consider the binomial population with

$$P(X=x) = \binom{n}{x} p^x (1-p)^x, x = 0,1,\ldots,n, n \ge 1$$

and

$$F(x) = P(X \le x) = \sum_{j=0}^{x} \binom{n}{x} p^j (1-p)^{n-j}.$$

For the *r*th order statistic

$$\mathbf{P(X_{r,n} \le x)=} \sum_{j=r}^{n} \binom{n}{j} (F(x))^j (1-F(x))$$
$$-\{\{(F(x))^j - (1-F(x-1))^{n-j}\}, x = 0,1,\ldots,n$$

With $F(-1)=0$.

The above expression can be written as

$$P(X_{r,n}=x) = I_{F(x)}(r, n-r+1) - I_{F(x-1)}(r, n-r+1),$$

where $I_\alpha(a,b)$ is the incomplete beta function defined as

$$I_\alpha(a,b) = \frac{1}{B(a,b)} \int_0^\alpha x^{a-1}(1-x)^{b-1} dx.$$

Exercise 6.17. Consider the Poisson distribution with

$$P(X=x) = \frac{\lambda^x}{x!} e^{-x}, x = 0,1,2,\ldots,\lambda > 0.$$

Show that

$$P(X_{r,n} = x) = I_{F(x)}(r, n-r+1, n) - I_{F(x-1)}(r, n-r+1, n),$$

where

$$F(x) = P(X \le x) = e^{-\lambda} \sum_{j=0}^{x} \frac{\lambda^j}{j!}.$$

Check your solutions.

Exercise 6.1 **(solution).** In this case the sample size n=2 and

$$P\{X_{1,2} =X_{2,2}\}=P\{X_1=X_2\}= \sum_{r} P\{X_1=x_r,X_2=x_r\}=$$

$$\sum_{r} P\{X_1=x_r\}P\{X_2=x_r\}=\sum_{r} p_r^2 . \qquad (6.5)$$

It is evident now that

$$0\le p^2\le \sum_{r} p_r^2 \le p\sum_{r} p_r=p.$$

Exercise 6.2 **(solution).** In the general case it follows from (6.5) that

$$\pi_2=1-\sum_{r} p_r^2 .$$

Then, in our partial case $\pi_2=1-1/n$. Since

$P\{X_{\alpha(1)}=1,X_{\alpha(2)}=2,\ldots,X_{\alpha(n)}=n\} = P\{X_{\alpha(1)}=1\}P\{X_{\alpha(2)}=2\}\cdots P\{X_{\alpha(n)}=n\} = P\{X=1\}P\{X_2=2\}\cdots P\{X_n=n\} =1/n^n$,

for any of n! permutations $(\alpha(1),\alpha(2),\ldots,\alpha(n))$ of numbers $(1,2,\ldots,n)$, one evidently obtains that

$$\pi_n=n!\, P\{X_1=1,X_2=2,\ldots,X_n=n\}=n!/n^n.$$

One obtains analogously that

$$\pi_r =n(n-1)\cdots(n-r+1)/n^r, \quad \text{if } 2<r<n.$$

Exercise 6.3 **(solution).** Evidently,

$$P\{X_{k,n}=0\}=P\{N(n)\ge k\},$$

where N(n) denotes the number of X's among $X_1,X_2,\ldots,X_n$, which are zero. We have that

$$P\{N(n)=m\}=\binom{n}{m} p^m g^{n-m}$$

and

$$P\{X_{k,n}=0\}=\sum_{m=k}^{n}\binom{n}{m}p^m g^{n-m}.$$

Hence,

$$P\{X_{k,n}=1\}=1-P\{X_{k,n}=0\}=1-\sum_{m=k}^{n}\binom{n}{m}p^m g^{n-m}=\sum_{m=0}^{k-1}\binom{n}{m}p^m g^{n-m}.$$

Exercise 6.4 **(answers).**

$$P\{X_{j,n}=0, X_{k,n}=0\}=P\{X_{k,n}=0\}=\sum_{m=k}^{n}\binom{n}{m}p^m g^{n-m},$$

$$P\{X_{j,n}=0, X_{k,n}=1\}=\sum_{m=j}^{k-1}\binom{n}{m}p^m g^{n-m}$$

and

$$P\{X_{j,n}=1, X_{j,,n}=1\}=P\{X_{j,n}=1\}=\sum_{m=0}^{j-1}\binom{n}{m}p^m g^{n-m}.$$

Then,

$$P\{W_{j,k,n}=1\}=1-P\{W_{j,k,n}=0\}=P\{X_{j,n}=0, X_{k,n}=1\}=\sum_{m=j}^{k-1}\binom{n}{m}p^m g^{n-m}.$$

Exercise 6.5 **(solution).** We note that $P\{X\geq k\}=p^k$ and hence

$$P\{X\leq k\}=1-P\{X\geq k+1\}=1-p^{k+1},\ k=0,1,2,\ldots.$$

Then

$$P\{X_{1,n}\geq k\}=P\{X_1\geq k,X_2\geq k,\ldots,X_n\geq k\}=P\{X_1\geq k\}P\{X_2\geq k\}\cdots P\{X_n\geq k\}=p^{kn}$$

and

$$P\{X_{1,n}=k\}=P\{X_{1,n}\geq k\}-P\{X_{1,n}\geq k+1\}=p^{kn}(1-p^n).$$

It means that $X_{1,n}$ has the geometric distribution also. Analogously we obtain that

$$P\{X_{n,n}\le k\}=P\{X_1\le k,\dots, X_n \le k\}=P^n \{X\le k\}=(1-p^{k+1})^n$$

and

$$P\{X_{n,n}=k\}= P\{X_{n,n}\le k\} - P\{X_{n,n}\le k-1\} = (1-p^{k+1})^n - (1-p^k)^n,\ k=0,1,2,\dots .$$

Exercise 6.6 **(answer).** $P\{X_{2,n}\ge k\}=p^{k(n-1)}(n-(n-1)p^k)$

and

$$P\{X_{2,n}=k\} = np^{k(n-1)}(1-p^{(n-1)})-(n-1)p^{kn}(1-p^n),\ k=0,1,2,\dots.$$

Exercise 6.7 **(answer).** In this case

$$P\{Y =k\}=(1-p)p^k,\ k=0,1,2,\dots,$$

where $p=p_1p_2\dots p_n$.

Exercise 6.8 **(solution).** Denote

$$F(r)= p_0+p_1+\dots+p_r.$$

Then

$$P\{X_{k,n}\le r\}= \sum_{m=k}^{n-1} P\{X_{m,n}\le r,\ X_{m+1,n}>r\}+P\{X_{n,n}\le r\}= \sum_{m=k}^{n}\binom{n}{m}(F(r))^m(1-F(r))^{n-m}.$$

It was proved in exercise 2.5 that the following identity (2.6) is true:

$$\sum_{m=k}^{n} \binom{n}{m} x^m(1-x)^{n-m} = I_x (k,n-k+1),$$

where

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

is the incomplete beta function with parameters *a* and *b*, *B(a,b)* being the classical beta function. Hence

$$P\{X_{k,n}\le r\}= I_{F(r)} (k,n-k+1)$$

and

$$P\{X_{k,n}=r\}=P\{X_{k,n}\le r\}-P\{X_{k,n}\le r-1\}=I_{F(r)}(k,n-k+1)-I_{F(r-1)}(k,n-k+1). \tag{6.6}$$

Exercise 6.9 **(solution).** It is evident that

$$P\{X_{1,n}\ge i,X_{n,n}<j\}=(P\{i\le X<j\})^n=(P\{X\ge i\}-P\{X\ge j\})^n=(p^i-p^j)^n,$$

for any $0\le i<j$ and $P\{X_{1,n}\ge i,X_{n,n}<j\}=0$, if $i\ge j$. Hence,

$$P\{X_{1,n}=i, X_{n,n=}j\}=P\{X_{1,n}\ge i,X_{n,n}<j+1\}-P\{X_{1,n}\ge i,X_{n,n}<j\}-$$

$$P\{X_{1,n}\ge i+1,X_{n,n}<j+1\}+P\{X_{1,n}\ge i+1,X_{n,n}<j\}=$$

$$(p^i-p^{j+1})^n-(p^i-p^j)^n-(p^{i+1}-p^{j+1})^n+(p^{i+1}-p^j)^n=$$

$$p^{in}((1-p^{j-i+1})^n-(1-p^{j-i})^n-(p-p^{j-i+1})^n+(p-p^{j-i})^n),$$

if $i<j$,

$$P\{X_{1,n}=i, X_{n,n=}i\}=(1-p)^n p^{in},\quad i=0,1,2,\ldots.$$

and

$$P\{X_{1,n}=i, X_{n,n=}j\}=0,$$

if $j<i$.

Exercise 6.10 (solution). One sees that

$$P\{W_n=r\}=\sum_{i=0}^{\infty} P\{X_{1,n}=i, X_{n,n}=i+r\}.$$

On applying the result of exercise 6.9, we get that

$$P\{W_n=r\}=\sum_{i=0}^{\infty} p^{in}((1-p^{r+1})^n-(1-p^r)^n-(p-p^{r+1})^n+(p-p^r)^n)=$$

$$((1-p^{r+1})^n-(1-p^r)^n-(p-p^{r+1})^n+(p-p^r)^n)/(1-p^n),\quad r=1,2,\ldots,$$

and

$$P\{W_n=0\}=\sum_{i=0}^{\infty} P\{X_{1,n}=i, X_{n,n}=i\}=(1-p)^n/(1-p^n).$$

Exercise 6.11 **(hint).** Use the evident equality

$$P\{X_{1,n}=i, W_n=r\}=P\{X_{1,n}=i, X_{n,n}=i+r\}$$

separately for r=0 and r≥1. Then the joint distributions of $X_{1,n}$ and $X_{n,n}$,obtained in exercise 6.9 , distributions of $X_{1,n}$ and W_n ,given in exercises 6.5 and 6.10 correspondingly, help you to prove that

$$P\{X_{1,n}=i,\ W_n=r\}= P\{X_{1,n}=i\}P\{\ W_n=r\}$$

for any n≥2, i=0,1,2,… and r=0,1,2,….

Exercise 6.12 **(solution).** For the sake of simplicity we can suppose that x_k =k.

Let r<k<s. Then

$$P\{X_{1,3}=r,\ X_{3,3}=s\,|\,X_{2,3}=k\}= P\{X_{1,3}=r,\ X_{2,3}=k,\ X_{3,3}=s\}/P\{X_{2,3}=k\}=$$

$$6P\{X_1=r,X_2=k,X_3=s\}/P\{X_{2,3}=k\}=6p_rp_sg_k\ , \qquad (6.7)$$

where

$$g_k=p_k/\ P\{X_{2,3}=k\}=p_k/(\ p_k^3\ +3F(k-1)p_k^2\ +3(1-F(k))p_k^2\ +6F(k-1)(1-F(k)) \qquad (6.8)$$

and F(k)=p_1+…+p_k.

If r=k<s, then

$$P\{X_{1,3}=k,\ X_{3,3}=s\,|\,X_{2,3}=k\}= 3P\{X_1=X_2=k,\ X_3=s\}/P\{X_{2,3}=k\}=3\ p_kp_sg_k\ , \qquad (6.9)$$

where g_k is defined in (6.8).

If r<k=s, then analogously

$$P\{X_{1,3}=r,\ X_{3,3}=k\,|\,X_{2,3}=k\} = 3\ p_rp_sg_k, \qquad (6.10)$$

and for r=k=s one obtains that

$$P\{X_{1,3}=k,\ X_{3,3}=k\,|\,X_{2,3}=k\}=P\{\ X_1=X_2=\ X_3=k\}/P\{\ X_{2,3}=k\}=\ p_k^2\ g_k. \qquad (6.11)$$

Now if r<k we have that

$$P\{X_{1,3}=r\,|\,X_{2,3}=k\} = P\{X_{1,3}=r,X_{2,3}=k\}/P\{X_{2,3}=k\}=$$

$$(P\{X_{1,3}=r,X_{2,3}=k,\ X_{3,3}=k\}+ P\{X_{1,3}=r,X_{2,3}=k,\ X_{3,3}>k\})/\ P\{X_{2,3}=k\}=$$

$$(3p_r\ p_k^2\ +6p_rp_k(1-F(k)))/\ P\{X_{2,3}=k\}\ =3p_r(p_k+2-2F(k))g_k. \qquad (6.12)$$

If r=k, then

$$P\{X_{1,3}=k\,|\,X_{2,3}=k\}= (P\{X_{1,3}=X_{2,3}=\ X_{3,3}=k\}+\ P\{X_{1,3}=X_{2,3}=k,\ X_{3,3}>k\})/\ P\{X_{2,3}=k\}=$$

$$(p_k^2\ +3p_k(1-F(k)))g_k. \qquad (6.13)$$

Analogously,

$$P\{X_{3,3}=s|X_{2,3}=k\}=3p_s(p_k+2F(k-1))g_k, \qquad (6.14)$$

if s>k, and

$$P\{X_{3,3}=k|X_{2,3}=k\}=(p_k^2+3p_kF(k-1))g_k. \qquad (6.15)$$

Exercise 6.13 **(solution).** In fact, to solve this problem we need to check that the following four equalities are valid:

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}=a\} = P\{X_{1,3}=a|X_{2,3}=a\}\, P\{X_{3,3}=b|X_{2,3}=a\},$$

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}=b\} = P\{X_{1,3}=a|X_{2,3}=b\}\, P\{X_{3,3}=b|X_{2,3}=b\},$$

$$P\{X_{1,3}=a, X_{3,3}=a|X_{2,3}=a\} = P\{X_{1,3}=a|X_{2,3}=a\}\, P\{X_{3,3}=a|X_{2,3}=a\}$$

and

$$P\{X_{1,3}=b, X_{3,3}=b|X_{2,3}=b\} = P\{X_{1,3}=b|X_{2,3}=b\}\, P\{X_{3,3}=b|X_{2,3}=b\}.$$

It is evident that

$$P\{X_{1,3}=a|X_{2,3}=a\} =1, \quad P\{X_{3,3}=b|X_{2,3}=b\} =1,$$

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}= b\}= P\{X_{1,3}= a |X_{2,3}=b\},$$

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}= a\}=P\{X_{3,3}= b|X_{2,3}=a\}$$

and these relations immediately imply that

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}=b\} = P\{X_{1,3}=a|X_{2,3}=b\}\, P\{X_{3,3}= b|X_{2,3}= b\}$$

and

$$P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}=a\} =P\{X_{1,3}=a|X_{2,3}=a\}\, P\{X_{3,3}=b|X_{2,3}=a\}.$$

Then, we see that

$$P\{X_{1,3}=a, X_{3,3}=a|X_{2,3}= a\} = 1-P\{X_{1,3}=a, X_{3,3}=b|X_{2,3}=a\}=$$

$$P\{X_{3,3}=b|X_{2,3}=a\} = P\{X_{3,3}=a|X_{2,3}=a\}$$

and hence

$$P\{X_{1,3}=a, X_{3,3}=a|X_{2,3}=a\} = P\{X_{1,3}=a|X_{2,3}=a\}\, P\{X_{3,3}=a|X_{2,3}=a\}.$$

Analogously one obtains that

$$P\{X_{1,3}=b, X_{3,3}=b|X_{2,3}=b \}= P\{X_{1,3}=b|X_{2,3}=b\}\, P\{X_{3,3}=b|X_{2,3}=b\}.$$

Exercise 6.14 **(answer).**

$$\mathrm{P}\{ X_{i,n,N} = x_r,\ X_{j,n,N}=x_s\}= \binom{r-1}{i-1}\binom{s-r-1}{j-i-1}\binom{N-s}{n-j}/\binom{N}{n}, \qquad (6.16)$$

if $i \le r < s \le N\text{-}n\text{+}j$ and $s\text{-}r \ge j\text{-}i$, and

$$\mathrm{P}\{ X_{i,n,N} = x_r,\ X_{j,n,N}=x_s\}= 0, \text{ otherwise.}$$

Exercise 6.15 **(answer).**

$$\mathrm{P}\{ X_{1,n,N} = x_{r(1)},\ X_{2,n,N} = x_{r(2)},\dots,\ \ X_{k,n,N} = x_{r(k)}\}=$$

$$\binom{N-r(k)}{n-k}/\binom{N}{n},$$

if $1\le r(1)<r(2)<\dots<r(k)$ and $N\text{-}r(k)\ge n\text{-}k$, while

$$\mathrm{P}\{ X_{1,n,N} = x_{r(1)},\ X_{2,n,N}=x_{r(2)},\dots,\ \ X_{k,n,N}=x_{r(k)}\}= 0,$$

otherwise.

Exercise 6.16 **(solution).** One needs to show that for any $1\le i< r < j\le N$ the following relation holds:

$$\mathrm{P}\{X_{1,3,N} =x_i,\ X_{3,3,N}=x_j\,|\,X_{2,3,N}=x_r\}= \mathrm{P}\{X_{1,3,N} =x_i\,|\,X_{2,3,N}=x_r\}\mathrm{P}\{X_{3,3,N}=x_j\,|\,X_{2,3,N}=x_r\}.$$

This is equivalent to the equality

$$\mathrm{P}\{X_{1,3,N} =x_i,\ X_{2,3,N}=x_r,X_{3,3,N}=x_j\}\mathrm{P}\{ X_{2,3,N}=x_r\} = \mathrm{P}\{X_{1,3,N} =x_i,X_{2,3,N}=x_r\}\mathrm{P}\{X_{2,3,N}=x_r,\ X_{3,3,N}=x_j\}. \qquad (6.17)$$

To check this equality we must recall from (6.2), (6.3) and (6.16) that

$$\mathrm{P}\{X_{1,3,N} =x_i,\ X_{2,3,N}=x_r,X_{3,3,N}=x_j\}=1/\binom{N}{3},$$

$$\mathrm{P}\{X_{1,3,N} =x_i,X_{2,3,N}=x_r\}=(N-r)/\binom{N}{3},$$

$$\mathrm{P}\{X_{2,3,N}=x_r,\ X_{3,3,N}=x_j\}=(r-1)/\binom{N}{3}$$

and

$$\mathrm{P}\{ X_{2,3,N}=x_r\}=(r-1)(N-r)/\binom{N}{3}.$$

The rest part of the proof is evident.

Exercise 6.17 **(hint).** For the rth order statistic

$$P(X_{r,n} = x) = \sum_{j=r}^{n} \binom{n}{j} (F(x))^{j} (1-F(x))$$
$$-\{\{(F(x))^{j} - (1-F(x-1))^{n-j}\}, x = 0,1,\ldots,n$$

with F(-1) =0.

Chapter 7. Moments of order statistics: general relations
Моменты порядковых статистик: общие соотношения

Читатель уже знает, как находить распределения порядковых статистик и как с ними работать. Часто это не совсем приятная работа. Да и не всегда нам нужно знать подробности, связанные с этими вероятностными распределениями. Достаточно иметь представление о различных моментных характеристиках соответствующих порядковых статистик. В этой главе рассматриваются полезные соотношения для моментов. Один из любопытных и весьма полезных результатов приведен в Примере 7.2. Показано, в частности, что если даже у исходного распределения (скажем, распределения Коши) не существует математического ожидания, то у средних членов вариационного ряда можно уже найти дисперсию и моменты более высоких порядков.

Due to (2.8) we get the general formula for moments of order statistics $X_{k,n}$ related to a population with a distribution function F. In fact,

$$\mu_{k:n}^{(r)} = E(X_{k,n})^r = \int_{-\infty}^{\infty} x^r dF_{k:n}(x) =$$

$$\frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^r (F(x))^{k-1}(1-F(x))^{n-k} dF(x). \qquad (7.1)$$

In the case of continuous distribution functions *F* we can express (7.1) as

$$\mu_{k:n}^{(r)} = \frac{n!}{(k-1)!(n-k)!} \int_{0}^{1} (G(u))^r u^{k-1}(1-u)^{n-k} du, \qquad (7.2)$$

where G(u) is the inverse of F. For absolutely continuous distributions the RHS of (7.1) coincides with

$$\frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^r (F(x))^{k-1}(1-F(x))^{n-k} f(x)dx. \qquad (7.3)$$

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

$$\mu_{i,j:n}^{(r,s)} = E((X_{i,n})^r (X_{j,n})^s), \ 1 \le i < j \le n,$$

of two order statistics only. From (2.14) we obtain for absolutely continuous distributions that

$$\mu_{i,j:n}^{(r,s)} =$$

$$C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s \, (F(x))^{i-1}(F(y)-F(x))^{j-i-1}(1-F(y))^{n-j} f(x)f(y)dxdy, \qquad (7.4))$$

where

$$C=c(i,j,n)=\frac{n!}{(i-1)!(j-i-1)!(n-j)!} \; . \qquad (7.5)$$

In the general case

$$\mu_{i,j:n}^{(r,s)} =$$

$$C \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^r y^s \, (F(x))^{r-1}(F(y)-F(x))^{s-r-1}(1-F(y))^{n-s} dF(x)dF(y), \qquad (7.6)$$

where C=c(i,j,n) is given in (7.5). The following notations

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

will be used also for the sake of simplicity instead of $\mu_{k:n}^{(1)}$; then

$$\mu_{i,j:n} = E\,(X_{i,n}X_{j,.n})$$

will change $\mu_{i,j:n}^{(1,1)}$;

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

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

$$cov\,(X_{i,n}\,,\, X_{j,n}) = \mu_{i,j:n} - \mu_{i:n}\,\mu_{j:n}$$

will be used for the covariance between $X_{i,n}$ and $X_{j,n}$.

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

Example 7.1. Let there exist the population moment $\alpha_r=EX^r$, i.e.,

$$E|X|^r= \int_{-\infty}^{\infty} |x|^r dF(x) < \infty. \qquad (7.7)$$

Then due to (7.7) we easily derive that

$$E|X_{k,n}|^r \le \frac{n!}{(k-1)!(n-k)!}\int_{-\infty}^{\infty} |x|^r\,(F(x))^{k-1}(1-F(x))^{n-k}dF(x) \le$$

$$\frac{n!}{(k-1)!(n-k)!}\int_{-\infty}^{\infty} |x|^r\,dF(x) \le \frac{n!}{(k-1)!(n-k)!}\,E|X|^r<\infty. \qquad (7.8)$$

It follows from (7.8) that the existence of the moment $E|X|^r$ implies the existence of all moments

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

Exercise 7.1. Show that if

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

for some r, then for any n=1,2,…, you can find such order statistic $X_{k,n}$ that

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

Remark 7.1. Since

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

for any 1<k<n, one has the evident inequality

$$E|X_{k,n}|^r\le\ E(|X_{1,n}|^r+|X_{n,n}|^r).$$

Hence, if $E|X_{k,n}|^r = \infty$, then at least one of equalities $E|X_{1,n}|^r = \infty$ or $E|X_{n,n}|^r = \infty$ is valid.

Exercise 7.2. Let $E|X_{k,n}| = \infty$ and $E\ |X_{k+1,n}| <\infty$. Show that then $E|X_{r,n}| = \infty$, if r =1,2,…, k-1. Analogously, if $E|X_{k,n}| = \infty$ and $E|X_{k-1,n}|< \infty$, then $E|X_{r,n}| = \infty$, for r = k+1,…,n.

Exercise 7.3. Let X have the Cauchy distribution with the density function

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

Show that for any r=1,2,…, relation

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

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

Example 7.2. Consider a more general situation than one given in exercise 7.3. This result was proved by Sen (1959).

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

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

Due to the result given in example 7.1, this statement is evident if $r/\alpha \le 1$, since the existence of the moment $E|X|^{\alpha}$ implies the existence of moments

$$\mu_{k:n}^{(r)},\ k=1,2,\ldots,n,\ n=1,2,\ldots,$$

for $r \le \alpha$. Hence, we need to consider the case $r>\alpha$ only.

If $E|X|^{\alpha}<\infty$, then integrals

$$I_1(\alpha)=\int_0^{\infty} x^{\alpha-1} F(-x)dx$$

and

$$I_2(\alpha)=\int_0^{\infty} x^{\alpha-1} (1-F(x))dx$$

are finite and

$$E|X|^{\alpha} = \alpha(I_1(\alpha)+I_2(\alpha)). \tag{7.9}$$

Moreover, if it is known that $I_1(\alpha)$ and $I_2(\alpha)$ are finite, then (7.9) is also true.

Note, that if $E|X|^{\alpha} < \infty$ (or simply, if both integrals , $I_1(\alpha)$ and $I_2(\alpha)$ are finite) then

$$(1-F(x))=o(x^{-\alpha}),\ F(-x)= o(x^{-\alpha}),\ x\to\infty. \tag{7.10}$$

It suffices for us to prove that if

$$E|X|^{\alpha}<\infty \quad \text{and} \quad r/\alpha \le k \le (n-r+1)/\alpha,$$

then

$$\int_0^{\infty} x^{r-1} F_{k,n}(-x)dx<\infty,\ \int_0^{\infty} x^{r-1} (1-F_{k,n}(x))dx<\infty,$$

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

Let us recall (see relation (2.5)) that

$$F_{k,n}(x) = \sum_{m=k}^{n} \binom{n}{m} (F(x))^m(1-F(x))^{n-m} \qquad (7.11)$$

and

$$1-F_{k,n}(x) = \sum_{m=0}^{k-1} \binom{n}{m} (F(x))^m(1-F(x))^{n-m}. \qquad (7.12)$$

The following evident inequalities are valid for the LHS of (7.12):

$$1-F_{k,n}(x) \le (1-F(x))^{n-k+1} \sum_{m=0}^{k-1} \binom{n}{m} (F(x))^m(1-F(x))^{k-m-1} \le$$

$$(1-F(x))^{n-k+1} \sum_{m=0}^{k-1} \binom{n}{m} \le (1-F(x))^{n-k+1} \sum_{m=0}^{n} \binom{n}{m} = 2^n(1-F(x))^{n-k+1}.$$

Further,

$$0\le \int_0^\infty x^{r-1} (1-F_{k,n}(x))dx \le 2^n \int_0^\infty x^{r-1}(1-F(x))^{n-k+1} =$$

$$2^n \int_0^\infty x^{\alpha-1}(1-F(x))h_k(x)dx, \qquad (7.13)$$

where

$$h_k(x)=x^{r-\alpha}(1-F(x))^{n-k},$$

and due to (7.10),

$$h_k(x)=o(x^{r-\alpha-(n-k)\alpha})=o(1),\ x\to\infty, \qquad (7.14)$$

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

Since

$$\int_0^\infty x^{\alpha-1} (1-F(x))dx<\infty,$$

it follows evidently from (7.13) and (7.14) that

$$\int_0^\infty x^{r-1}(1-F_{k,n}(x))dx<\infty. \qquad (7.15)$$

Similarly, one gets that

$$\int_0^\infty x^{r-1}F_{k,n}(-x)dx<\infty, \qquad (7.16)$$

If ***$k \geq r/\alpha$.*** Finally, (7.12) and (7.13) imply that

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

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

Remark 7.2. All saying in example 7.2 enables us to state additionally that if $X\geq 0$ and $EX^\alpha<\infty$, then

$$0\leq EX_{k.n}^r<\infty$$

for all k and n such that $1\leq k\leq (n- r+1)/\alpha$.

If $X\leq 0$ and $E(-X)^\alpha<\infty$, then

$$0\leq E(-X_{k,n})^r<\infty, \quad r/\alpha \leq k \leq n.$$

Some useful relations for moments come from the evident identity

$$X_{1,n}+\ldots+X_{n,n} = X_1+\ldots+X_n \qquad (7.17)$$

and related equalities. For instance, the simplest corollary of (7.17) is as follows:

$$E(X_{1,n}+\ldots+X_{n,n}) = E(X_1+\ldots+X_n) = nEX, \qquad (7.18)$$

if there exists a population expectation. Say, if $EX= 0$ and $n=2$ in (7.18), then we get that

$$E X_{2,2} = - EX_{1,2.}$$

A natural generalization of (7.17) has the form

$$g(\sum_{k=1}^n h(X_{k,n})) = g(\sum_{k=1}^n h(X_k)), \qquad (7.19)$$

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

Example 7.3. The following identities based on (7.19) can be useful in some situations:

$$E\,(\sum_{k=1}^{n} X_{k,n}^{m})^{r} = E\,(\sum_{k=1}^{n} X_{k}^{m})^{r},\ m=1,2,\dots,\ r=1,2,\dots. \qquad (7.20)$$

The case m=r=1 was considered in (7.18). Similarly we get that

$$\sum_{k=1}^{n} EX_{k,n}^{m} = \sum_{k=1}^{n} EX_{k}^{m} = nEX^{m} \qquad (7.21)$$

for any *m* provided that the corresponding moment EX^m exists. If m=1 and r=2, one gets that

$$\sum_{k=1}^{n} EX_{k,n}^{2} + 2\sum_{k=1}^{n-1}\ \sum_{r=k+1}^{n} E(X_{k,n}X_{r,n)} =$$

$$E(\sum_{k=1}^{n} X_{k})^{2} = nEX^{2} + n(n-1)(EX)^{2}. \qquad (7.22).$$

Due to equality (7.21) (for m=2), identity (7.22) can be simplified:

$$\sum_{k=1}^{n-1}\ \sum_{r=k+1}^{n} E(X_{k,n}X_{r,n}) = \binom{n}{2}(EX)^{2}. \qquad (7.23)$$

Exercise 7.4. Prove that the following identity holds for covariances between order statistics:

$$\sum_{k=1}^{n}\ \sum_{r=1}^{n} Cov(EX_{k,n}, X_{r,n}) = nVarX. \qquad (7.24)$$

Example 7.4. One more similar identity

$$\sum_{k=1}^{n}\ \sum_{r=1}^{n} X_{k,n}^{m}\, X_{r,n}^{s} = \sum_{k=1}^{n}\ \sum_{r=1}^{n} X_{k}^{m}\, X_{r}^{s} \qquad (7.25)$$

implies the relation

$$\sum_{k=1}^{n}\sum_{r=1}^{n} \mathrm{E}(X_{k,n}^{m} X_{r,n}^{s})=nEX^{m+s}+n(n-1)EX^{m}EX^{s}. \qquad (7.26)$$

Exercise 7.5. Prove that the following relation holds for any 1≤k≤n-1 and m=1,2,…:

$$kEX_{k+1,n}^{m}+(n-k)\, EX_{k,n}^{m}=nE\, X_{k,n-1}^{m}. \qquad (7.27)$$

Check your solutions

***Exercise 7.1* (solution).** Suppose that all order statistics have finite moments $E|X_{k,n}|^r<\infty$. It is clear that

$$P\{X_{1,n}\le X_1\le X_{n,n}\}=1.$$

Then,

$$P\{ |X_1|^r\le \max\{|X_{1,n}|^r,|X_{n,n}|^r\}=1$$

and the following inequality, which contradicts to the initial condition, is valid:

$$E|X|^r =E\ |X_1|^r\le E\max\{|X_{1,n}|^r,\ |X_{n,n}|^r\}\le E(|X_{1,n}|^r +|X_{n,n}|^r) < \infty.$$

Hence, there is at least one order statistic such that $E|X_{k,n}|^r= \infty$.

***Exercise 7.2* (hint).** In the first case consider the evident inequalities

$$E|X_{k,n}| \le\ E(|X_{r,n}| +|X_{k+1,n}|),\ r=1,2,\dots.k-1.$$

In the second situation it suffices to use inequalities

$$E|X_{k,n}| \le E(|X_{r,n}| +|X_{k-1,n}|),\ r= k+1,k+2,\dots.n.$$

***Exercise 7.3* (hint).** Prove that the inverse function of F satisfies the following relations:

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

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

Then use equality (7.2) to prove the statement of the exercise.

Exercise 7.4 **(hint)**. Use the evident identity

$$\left(\sum_{k=1}^{n} (X_{k,n}-EX_{k,n})\right)^2 = \left(\sum_{k=1}^{n} (X_k-EX_k)\right)^2.$$

Exercise 7.5 **(solution)**. Recall that

$$E(X_{r,n})^m = \frac{n!}{(r-1)!(n-r)!}\int_{-\infty}^{\infty} x^m (F(x))^{r-1}(1-F(x))^{n-r}dF(x).$$

Then

$$kEX_{k+1,n}^{m} + (n-k)\,EX_{k,n}^{m} =$$

$$\frac{k(n!)}{k!(n-k-1)!}\int_{-\infty}^{\infty} x^m (F(x))^{k}(1-F(x))^{n-k-1}dF(x)+$$

$$\frac{(n-k)(n!)}{(k-1)!(n-k)!}\int_{-\infty}^{\infty} x^m (F(x))^{k-1}(1-F(x))^{n-k}dF(x) =$$

$$\frac{n!}{(k-1)!(n-k-1)!}\int_{-\infty}^{\infty} x^m \{(F(x))^{k}(1-F(x))^{n-k-1} + (F(x))^{k-1}(1-F(x))^{n-k}\}dF(x)=$$

$$\frac{n!}{(k-1)!(n-k-1)!}\int_{-\infty}^{\infty} x^m \{(F(x))^{k-1}(1-F(x))^{n-k-1}dF(x).$$

The latter expression coincides evidently with

$$nEX_{k,n-1}^{m}.$$

Chapter 8. Moments of uniform and exponential order statistics
Моменты равномерных и экспоненциальных порядковых статистик

В предыдущей главе были приведены различные общие формулы и соотношения для произвольных порядковых статистик. Сейчас рассмотрим различные моментные свойства порядковых статистик в случае двух наиболее распространенных исходных распределений- равномерного и экспоненциального.

In chapter 4 we proved some representations for uniform and exponential order statistics, which enable us to express these order statistics via sums or products of independent random variables. By virtue of the corresponding expressions one can easily find single and joint moments of exponential and uniform order statistics.

Uniform order statistics. Indeed, in the case of the standard uniform distribution one can use expression (7.3) to find single moments of order statistics $U_{k,n}$. In fact, in this case for any α>-k we get the following result:

$$E(U_{k,n})^{\alpha} = \frac{n!}{(k-1)!(n-k)!}\int_0^1 x^{\alpha} x^{k-1}(1-x)^{n-k}dx=$$

$$\frac{n!}{(k-1)!(n-k)!}B(\alpha+k, n-k+1) = \frac{n!\Gamma(\alpha+k)\Gamma(n-k+1)}{(k-1)!(n-k)!\Gamma(n+\alpha+1)} =$$

$$\frac{n!\Gamma(\alpha+k)}{(k-1)!\Gamma(n+\alpha+1)}\quad, \qquad (8.1)$$

where B(a, b) and Γ(s) denote the beta function and the gamma function respectively, which are tied by the relation

$$B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b).$$

Note also that

$$\Gamma(n) = (n-1)! \text{ for } n =1,2,\ldots.$$

If α is an integer, then the RHS on (8.1) is simplified. For instance,

$$EU_{k,n} = \frac{n!\Gamma(k+1)}{(k-1)!\Gamma(n+2)} =$$

$$\frac{n!k!}{(k-1)!(n+1)!} = \frac{k}{n+1}, \ 1\le k\le n, \qquad (8.2)$$

and

$$E(1/U_{k,n}) = \frac{n!\Gamma(k-1)}{(k-1)!\Gamma(n)} = \frac{n}{k-1}, \quad 2 \le k\le n. \qquad (8.3)$$

Similarly,

$$E(U_{k,n})^2 = \frac{k(k+1)}{(n+1)(n+2)}, \ 1 \le k\le n, \qquad (8.4)$$

and

$$E(1/(U_{k,n})^2) = \frac{n(n-1)}{(k-1)(k-2)}, \ 3\le k\le n. \qquad (8.5)$$

In general form, for r =1,2,…, we have

$$E(U_{k,n})^r = \frac{k(k+1)...(k+r-1)}{(n+1)(n+2)...(n+r)}, \ 1\le k\le n, \qquad (8.6)$$

and

$$E(1/(U_{k,n})^r) = \frac{n(n-1)...(n-r+1)}{(k-1)(k-2)...(k-r)}, \ r+1\le k\le n. \qquad (8.7)$$

It follows from (8.2) and (8.4) that

$$\mathrm{Var}\,(U_{k,n}) = \frac{k(n-k+1)}{(n+1)^2(n+2)}, \ 1\le k\le n. \qquad (8.8)$$

Exercise 8.1. Find the variance of $1/U_{k,n}$.

Exercise 8.2. Find the third central moments of $U_{k,n}$.

Some of the just-given moments can be obtained by means of representations (4.24) and (4.29).

Example 8.1. We know from (4.29) that

$$EU_{k,n} = E(S_k | S_{n+1}=1),$$

where

$$S_n = \nu_1+\nu_2+\ldots+\nu_n,\ n=1,2,\ldots,$$

and $\nu_1,\nu_2,\ldots$ are independent identically distributed random variables having the standard exponential distribution. Further, the symmetry arguments enable us to see that

$$E(S_k | S_{n+1}=1) = \sum_{r=1}^{k} E(\nu_r | \nu_1+\nu_2+\ldots+\nu_{n+1}=1) = kE(\nu_1 | \nu_1+\nu_2+\ldots+\nu_{n+1}=1) =$$

$$\frac{k}{n+1}(\nu_1+\nu_2+\ldots+\nu_{n+1} | \nu_1+\nu_2+\ldots+\nu_{n+1}=1) = \frac{k}{n+1}.$$

From example 4.6 we know (see (4.24)) that

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

and $\frac{S_k}{S_{n+1}}$ is independent on the sum $S_{n+1}=\nu_1+\nu_2+\ldots+\nu_{n+1}$. Then

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

Due to the independence of $\frac{S_k}{S_{n+1}}$ and S_{n+1} we have the following relation:

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

Thus,

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

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

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

Finally,

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

$$\frac{\Gamma(\alpha+k)\Gamma(n+1)}{\Gamma(k)\Gamma(\alpha+n+1)} = \frac{n!\Gamma(\alpha+k)}{(k-1)!\Gamma(\alpha+n+1)}$$

and the latter expression coincides with (8.1).

One more representation, (4.20), enables us to get joint (product) moments of the uniform order statistics.

Example 8.2. Consider two uniform order statistics $U_{r,n}$ and $U_{s,n}$, $1 \le r < s \le n$. As it was shown in (4.20),

$$(U_{1,n}, U_{2,n}, \dots, U_{n,n}) \overset{d}{=}$$

$$(W_1 W_2^{1/2} \dots W_{n-1}^{1/(n-1)} W_n^{1/n},\ W_2^{1/2} \dots W_{n-1}^{1/(n-1)} W_n^{1/n}, \dots, W_n^{1/n}),$$

where $W_1, W_2, \dots$ are independent and have the same standard uniform distribution.

Hence,

$$EU_{r,n}U_{s,n} = E(W_r^{1/r} W_{r+1}^{1/(r+1)} \dots W_n^{1/n}\, W_s^{1/s} W_{s+1}^{1/(s+1)} \dots W_n^{1/n}) =$$

$$E\, (W_r^{1/r} W_{r+1}^{1/(r+1)} \dots W_s^{2/s} W_{s+1}^{2/(s+1)} \dots W_n^{2/n}) =$$

$$E(W_r^{1/r}) E(Wr_{+1}^{1/(r+1)}) \dots E(W_s^{2/s}) E(\, W_{s+1})^{2/(s+1)} \dots E(W_n^{2/n}) =$$

$$\prod_{k=r}^{s-1} \frac{1}{(1+1/k)} \prod_{k=s}^{n} \frac{1}{(1+2/k)} = \frac{r(s+1)}{(n+1)(n+2)}. \tag{8.9}$$

From (8.2), (8.8) and (8.9) we derive the following expressions for covariations and correlation coefficients between the uniform order statistics:

$$\mathrm{Cov}\,(U_{r,n}, U_{s,n}) = EU_{r,n}U_{s,n} - EU_{r,n}\, EU_{s,n} = \frac{r(n-s+1)}{(n+1)^2 (n+2)},\ r \le s, \tag{8.10}$$

and

$$\rho(U_{r,n}, U_{s,n}) = \left(\frac{r(n-s+1)}{s(n-r+1)}\right)^{1/2}. \tag{8.11}$$

It is interesting to note (see , for example, Sathe (1988) and Szekely, Mori (1985)) that for any distribution with a finite second moment , except the uniform distributions,

$$\rho(X_{r,n},X_{s,n}) < (\frac{r(n-s+1)}{s(n-r+1)})^{1/2},$$

that is the equality

$$\rho(X_{r,n},X_{s,n}) = (\frac{r(n-s+1)}{s(n-r+1)})^{1/2}$$

characterizes the family of the uniform distributions.

Exercise 8.3. Find product moments

$$E(U_{r,n}^{\alpha} U_{s,n}^{\beta}),\ \alpha\geq 0,\ \beta\geq 0. \qquad (8.12)$$

Exercise 8.4. Let $X_{1,n},\ldots,X_{n,n}$ be order statistics corresponding to the distribution with the density

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

Find $EX_{r,n}$ and $EX_{r,n}X_{s,n}$, $1\leq r<s\leq n$.

Example 8.3. Representation (4.24) is also useful for finding of different moments of the uniform order statistics. Denote

$$\mu_k = \frac{v_k}{v_1+\ldots+v_{n+1}},\ k=1,2,\ldots,n+1, \qquad (8.13)$$

where $\nu_1, \nu_2,\ldots$ are independent exponentially E(1) distributed random variables. It follows from (4.24) that

$$(U_{1,n},U_{2,n},\ldots,U_{n,n}) \overset{d}{=} (\mu_1,\ \mu_1+\mu_2,\ldots,\ \mu_1+\mu_2+\ldots+\mu_n) \qquad (8.14)$$

The symmetrical structure of $\mu_{1,\ldots,}\mu_{n+1}$ and the natural identity

$$\mu_1+\mu_2+\ldots+\mu_{n+1}=1$$

implies that for any k=1,2,…,n+1,

$$1=E(\mu_1+\mu_2+\ldots+\mu_{n+1}) = (n+1)E\mu_k$$

and

$$E\mu_k=1/(n+1).$$

Thus,

$$EU_{k,n}=E(\mu_1+\mu_2+\ldots+\mu_k) = k/(n+1),\ k=1,2,\ldots,n.$$

Similarly we have equality

$$0 = \mathrm{Var}\,(\mu_1+\mu_2+\ldots+\mu_{n+1}) = \sum_{k=1}^{n+1} \mathrm{Var}(\mu_k)+2\sum_{1\le i<j\le n+1} \mathrm{cov}(\mu_i,\ \mu_j). \qquad (8.15)$$

It is clear that all variances Var (μ_k) take on the same value, say σ^2, as well as all covariances $\mathrm{cov}(\mu_i,\ \mu_j)$, $i\neq j$, are identical. Let

$$d= \mathrm{cov}(\mu_i,\ \mu_j),\ i\neq j.$$

Then we derive from (8.15) that

$$(n+1)\sigma^2+n(n+1)d =0.$$

Due to the latter equality we immediately find the expression for correlation coefficients between μ_i and μ_j, $i\neq j$. In fact,

$$\rho(\mu_i,\ \mu_j)= \mathrm{cov}(\mu_i,\ \mu_j)/(\mathrm{Var}(\mu_i)\mathrm{Var}(\mu_j))^{1/2}= d/\sigma^2= -1/n,\ i\neq j. \qquad (8.16)$$

It means, in particular, that

$$\rho(U_{1,n},U_{n,n})= -\rho(U_{1,n},\ 1-U_{n,n})=-\rho(\mu_1,\ \mu_{n+1})=1/n$$

and this is in agreement with (8.11). To develop further the just-given results we must know the value of σ^2. Recalling that

$$\sigma^2=\mathrm{Var}(\mu_k)\ ,\ k=1,2,\ldots,n+1,$$

and in particular,

$$\sigma^2=\mathrm{Var}(\mu_1) =\mathrm{Var}\,(U_{1,n}),$$

we get from (8.8) that

$$\sigma^2 =n/(n+1)^2(n+2). \qquad (8.17)$$

It means that

$$d = \mathrm{cov}(\mu_i,\ \mu_j) = -\sigma^2/n = -1/(n+1)^2(n+2),\ i\neq j. \qquad (8.18)$$

Now we can find covariances between different uniform order statistics.

Let i≤j. Then

$$\text{cov}\,(U_{i,n},U_{j,n})=\text{cov}\,(\mu_1+\ldots+\mu_i,\ \mu_1+\ldots+\mu_j) =$$

$$i\sigma^2+i(j-1)d= i\sigma^2(1-(j-1)/n)=i(n-j+1)/(n+1)^2(n+2). \quad (8.19)$$

In particular,

$$\text{Var}\,(U_{i,n})= \text{cov}\,(U_{i,n},U_{i,n}) =i(n-i+1)/(n+1)^2(n+2),\ i=1,2,\ldots,n. \quad (8.20)$$

Indeed, (8.20) and (8.19) coincide with equalities (8.8) and (8.10) respectively.

Thus, we suggested some alternative ways to get moments of the uniform order statistics. Below you will find some exercises, solutions of which are based on representation (4.24).

Exercise 8.5. Find $E(U_{1,n}/(1-U_{n,n}))^{\alpha}$.

Exercise 8.6. Find E(Y/V), where $Y= U_{r,n}$, $V= U_{s,n}-U_{m,n}$ and $1\le m<r<s\le n$.

Example 8.4. In example 4.7 we introduced differences

$$\delta_1= U_{1,n} - U_{0,n},\quad \delta_2= U_{2,n}-U_{1,n},\ldots,\quad \delta_n= U_{n,n}-U_{n-1,n},\quad \delta_{n+1}=U_{n+1,n}-U_{n,n},$$

where $U_{0,n}=0$, $U_{n+1,n}=1$, and formed the variational series

$$\delta_{1,n+1}\le \delta_{2,n+1}\le\ldots\le \delta_{n+1,n+1}.$$

We found that the following representation (4.32) is valid for elements of this new variational series:

$$\delta_{k,n+1} \overset{d}{=} \left(\frac{\nu_1}{n+1}+\frac{\nu_2}{n}+\ldots+\frac{\nu_k}{n-k+2}\right)/(\nu_1+\ldots+\nu_{n+1}),\ k=1,2,\ldots,n+1.$$

The technique, used above, enables us to derive that

$$E\delta_{k,n+1}=\frac{1}{n+1}\left(\frac{1}{n+1}+\frac{1}{n}+\ldots+\frac{1}{n-k+2}\right). \quad (8.21)$$

In particular,

$$E\delta_{1,n+1}=\frac{1}{n(n+1)} \sim 1/n^2,\quad n\to\infty,$$

and

$$E\delta_{n+1,n+1} = \frac{1}{n+1}(1+\frac{1}{2}+\ldots+\frac{1}{n+1}) \sim \frac{\log n}{n}, \quad n\to\infty.$$

Exercise 8.7. Find variances of order statistics $\delta_{k,n+1}$, k=1,2,…,n+1.

Exponential order statistics. Let $Z_{1,n} \le Z_{2,n} \le \ldots \le Z_{n,n}$, n=1,2,…, be order statistics corresponding to the standard exponential distribution with d.f.

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

To obtain moments

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

one needs to calculate integrals

$$\frac{n!}{(k-1)!(n-k)!}\int_0^{\infty} x^{\alpha}(H(x))^{k-1}(1-H(x))^{n-k}dH(x)=$$

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

$$\frac{n!}{(k-1)!(n-k)!}\sum_{r=0}^{k-1}(-1)^r\binom{k-1}{r}\int_0^{\infty} x^{\alpha}e^{-x(n-k+r+1)}dx.$$

Since

$$\int_0^{\infty} x^{\alpha}e^{-x(n-k+r+1)}dx = (n-k+r+1)^{-(\alpha+1)}\int_0^{\infty} u^{\alpha}e^{-u}du = \Gamma(\alpha+1)/(n-k+r+1)^{(\alpha+1)},$$

we obtain that

$$E(Z_{k,n})^{\alpha} = \frac{n!}{(k-1)!(n-k)!}\sum_{r=0}^{k-1}(-1)^r\binom{k-1}{r}\Gamma(\alpha+1)/(n-k+r+1)^{(\alpha+1)}. \qquad (8.22)$$

For instance, if k=1, then

$$E(Z_{1,n})^{\alpha} = n\Gamma(\alpha+1)/n^{(\alpha+1)} = \Gamma(\alpha+1)/n^{\alpha}, \quad \alpha > -1. \qquad (8.23).$$

For k=2 and α>-1 we have

$$E(Z_{2,n})^{\alpha} = n(n-1)\,\Gamma(\alpha+1)\{(n-1)^{-(\alpha+1)} - n^{-(\alpha+1)}\}. \qquad (8.24)$$

Some simplifications of the general expression (8.22) are due to representation (4.15).

Example 8.5. Let us recall that the exponential order statistics are expressed in terms of sums of independent random variables :

$$(Z_{1,n}, Z_{2,n},\ldots,Z_{n,n}) \overset{d}{=} \left(\frac{\nu_1}{n}, \frac{\nu_1}{n} + \frac{\nu_2}{n-1}, \ldots, \frac{\nu_1}{n} + \frac{\nu_2}{n-1} + \ldots + \frac{\nu_{n-1}}{2} + \nu_n\right),$$

where $\nu_1,\nu_2,\ldots$ are independent exponential E(1) random variables. Immediately we obtain that

$$EZ_{k,n} = E\left(\frac{\nu_1}{n} + \frac{\nu_2}{n-1} + \ldots + \frac{\nu_k}{n-k+1}\right) = \sum_{r=1}^{k} \frac{1}{n-r+1} \qquad (8.25)$$

and

$$Var(Z_{k,n}) = \sum_{r=1}^{k} Var\left(\frac{\nu_r}{n-r+1}\right) = \sum_{r=1}^{k} \frac{1}{(n-r+1)^2}, \qquad (8.26)$$

so far as $E\nu = Var\nu = 1$, if ν has the standard exponential distribution. It follows from (8.25) and (8.26) that

$$E(Z_{k,n})^2 = \sum_{r=1}^{k} \frac{1}{(n-r+1)^2} + \left(\sum_{r=1}^{k} \frac{1}{n-r+1}\right)^2. \qquad (8.27)$$

Comparing (8.25) and (8.27) with (8.22) (under α=1 and α=2) we derive the following identities:

$$\frac{n!}{(k-1)!(n-k)!} \sum_{r=0}^{k-1} (-1)^r \binom{k-1}{r} / (n-k+r+1)^2 = \sum_{r=1}^{k} \frac{1}{n-r+1} \qquad (8.28)$$

and

$$\frac{2(n!)}{(k-1)!(n-k)!} \sum_{r=0}^{k-1} (-1)^r \binom{k-1}{r} / (n-k+r+1)^3 =$$

$$\sum_{r=1}^{k}\frac{1}{(n-r+1)^2}+(\sum_{r=1}^{k}\frac{1}{n-r+1})^2. \qquad (8.29)$$

Remark 8.1. It is interesting to see that $EZ_{1,n}=1/n$ and Var $Z_{1,n}=1/n^2$ tend to zero, as $n\to\infty$, while

$$E\,Z_{n,n}=\sum_{r=1}^{n}\frac{1}{n-r+1}=\sum_{r=1}^{n}\frac{1}{r}\sim \log n\to\infty,\ n\to\infty, \qquad (8.30)$$

and

$$\mathrm{Var}\,Z_{n,n}=\sum_{r=1}^{n}\frac{1}{r^2}\to \pi^2/6,\ n\to\infty. \qquad (8.31)$$

Exercise 8.8. Find central moments $E(Z_{k,n}-EZ_{k,n})^3$.

Exercise 8.9. Find covariances between order statistics $Z_{r,n}$ and $Z_{s,n}$.

Exercise 8.10. Let $W = aZ_{r,n}+bZ_{s,n}$, where $r<s$. Find the variance of W.

Check your solutions

Exercise 8.1 **(answer).**

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

Exercise 8.2 **(answer).**

$$E(U_{k,n}-EU_{k,n})^3=\frac{2k(n-k+1)(n-2k+1)}{(n+1)^3(n+2)(n+3)},\ 1\le k\le n.$$

Exercise 8.3 **(answer).**

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

Exercise 8.4 **(hint and answer).** Use the relation

$$X_{r,n} \overset{d}{=} (U_{r,n})^{1/a},$$

which expresses $X_{r,n}$ via the uniform order statistics, and the result of exercise 8.3 to get equalities

$$\mathrm{E}X_{r,n} = \frac{n!\Gamma(r+1/a)}{(r-1)!\Gamma(n+1+1/a)}$$

and

$$\mathrm{E}X_{r,n}X_{s,n} = \mathrm{E}(U_{r,n})^{1/a}(U_{s,n})^{1/a} = \frac{n!\Gamma(r+1/a)\Gamma(s+2/a)}{(r-1)!\Gamma(s+1/a)\Gamma(n+1+2/a)}.$$

Exercise 8.5 (solution). Taking into account representation (4.24), we have relation

$$U_{1,n}/(1-U_{n,n}) \overset{d}{=} \nu_1/\nu_{n+1},$$

where ν_1 and ν_{n+1} are independent and have the standard E(1) exponential distribution. Then

$$\mathrm{E}(U_{1,n}/(1-U_{n,n}))^{\alpha} = \mathrm{E}(\nu_1/\nu_{n+1})^{\alpha} = \mathrm{E}\nu^{\alpha}\,\mathrm{E}\nu^{-\alpha},$$

where ν has density function $\exp(-x)$, $x>0$. Note that

$$\mathrm{E}\nu^{\beta} = \int_0^{\infty} x^{\beta}e^{-x}dx$$

is finite only if $\beta>-1$ and $\mathrm{E}\nu^{\beta} = \Gamma(\beta+1)$ for $\beta>-1$. Hence

$$\mathrm{E}(U_{1,n}/(1-U_{n,n}))^{\alpha} = \infty, \text{ if } |\alpha|\geq 1,$$

and

$$\mathrm{E}(U_{1,n}/(1-U_{n,n}))^{\alpha} = \Gamma(\alpha+1)\Gamma(1-\alpha),$$

if $|\alpha|<1$. We can recall some properties of gamma functions, such as

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

and

$$\Gamma(\alpha)\Gamma(1-\alpha) = \pi/\sin\pi\alpha$$

for $\alpha>0$, and then derive that

$$E(U_{1,n}/(1-U_{n,n}))^{\alpha} = \alpha\pi/\sin\pi\alpha$$

for $1>\alpha>0$.

Similarly,

$$E(U_{1,n}/(1-U_{n,n}))^{\alpha} = \Gamma(\alpha+1)\Gamma(1-\alpha) = (-\alpha)\Gamma(\alpha+1)\Gamma(-\alpha) = \alpha\pi/\sin\pi\alpha$$

for $-1<\alpha<0$.

Exercise 8.6 **(solution).** Coming back to representation (4.24) we derive that

$$Y/V \overset{d}{=} (\nu_1+\ldots\nu_r)/(\nu_{m+1}+\ldots\nu_r+\nu_{r+1}+\ldots\nu_s) =$$

$$(\nu_1+\ldots\nu_m)/(\nu_{m+1}+\ldots\nu_s) + (\nu_{m+1}+\ldots\nu_r)/ (\nu_{m+1}+\ldots\nu_s).$$

Hence,

$$E(Y/V) = E(\nu_1+\ldots\nu_m)E(1/(\nu_{m+1}+\ldots\nu_s)) + E(\nu_{m+1}+\ldots\nu_r)/(\nu_{m+1}+\ldots\nu_s).$$

Note that

$$E(\nu_{m+1}+\ldots\nu_r)/(\nu_{m+1}+\ldots\nu_s) = EU_{r-m,s-m-1} = (r-m)/(s-m)$$

and

$$E(\nu_1+\ldots\nu_m) = m.$$

Consider now the sum $\nu_{m+1}+\ldots\nu_s$, which has the gamma distribution with parameter (s-m). Hence (due to the fact that s-m>1) we obtain that

$$E(1/(\nu_m+\ldots\nu_s)) = \frac{1}{(s-m-1)!}\int_0^\infty x^{s-m-2}e^{-x}dx = \Gamma(s-m-1)/\Gamma(s-m) = 1/(s-m-1).$$

Finally we get

$$E(Y/V) = m/(s-m-1) + (r-m)/(s-m).$$

Exercise 8.7 (solution). Comparing relations (4.24), (4.32) and (8.14) we see that

$$\delta_{k,n+1} \overset{d}{=} \frac{\mu_1}{n+1} + \frac{\mu_2}{n} + \ldots + \frac{\mu_k}{n-k+2},\ k=1,2,\ldots, n+1,$$

where $\mu_1, \mu_2, \ldots, \mu_{n+1}$ are defined in (8.13). The variance of $\delta_{k,n+1}$ is expressed via variances of μ's, which are equal to

$$\sigma^2=n/(n+1)^2(n+2)$$

(see (8.17)), and covariances $\mathrm{cov}(\mu_i, \mu_j)$, $i\neq j$, which (as we know from (8.18)) are identical:

$$d= -\sigma^2/n=-1/(n+1)^2(n+2).$$

Really, we have

$$\mathrm{Var}(\delta_{k,n+1}) =\mathrm{Var}(\frac{\mu 1}{n+1}+\frac{\mu 2}{n}+\ldots+\frac{\mu k}{n-k+2}) =$$

$$\sum_{i=1}^{k} \mathrm{Var}(\frac{\mu_i}{n-i+2})+2\sum_{1\le i<j\le k} \mathrm{cov}(\frac{\mu_i}{n-i+2},\frac{\mu_j}{n-j+2})=$$

$$\sigma^2\sum_{i=1}^{k} \frac{1}{(n-i+2)^2}+2d\sum_{1\le i<j\le k} \frac{1}{(n-i+2)(n-j+2)}=$$

$$\frac{n}{(n+1)^2(n+2)}\sum_{i=1}^{k} \frac{1}{(n-i+2)^2}\;-$$

$$\frac{2}{(n+1)^2(n+2)}\sum_{1\le i<j\le k} \frac{1}{(n-i+2)(n-j+2)}, \qquad k=1,2,\ldots,n+1.$$

In particular,

$$\mathrm{Var}(\delta_{1,n+1})=n/(n+1)^4(n+2)$$

and

$$\mathrm{Var}(\delta_{2,n+1})= n/(n+1)^4(n+2)+ 1/n(n+1)^2(n+2)-2/(n+1)^3(n+2).$$

Exercise 8.8 **(solution)**. Recalling (4.15) one gets that

$$E(Z_{k,n}-EZ_{k,n})^3 = E(\frac{\nu_1-1}{n}+\frac{\nu_2-1}{n-1}+...+\frac{\nu_k-1}{n-k+1})^3 =$$

$$\sum_{r=1}^{k} E(\frac{\nu_r-1}{n-r+1})^3 = \sum_{r=1}^{k} \frac{1}{(n-r+1)^3} E(\nu-1)^3,$$

where ν has the standard E(1) exponential distribution. We have also that

$$E(\nu-1)^3 = E\nu^3-3E\nu^2+3E\nu-1=\Gamma(4)-3\Gamma(3)+3-1=2.$$

Hence,

$$E(Z_{k,n}-EZ_{k,n})^3 = 2\sum_{r=1}^{k} \frac{1}{(n-r+1)^3}.$$

Exercise 8.9 **(solution).** Let r≤s. Due to (4.15),

$$\mathrm{Cov}(Z_{r,n},Z_{s,n}) = \mathrm{Cov}(\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_r}{n-r+1}, \frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_s}{n-s+1}) =$$

$$\mathrm{Cov}(\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_r}{n-r+1}, \frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_r}{n-r+1}) +$$

$$\mathrm{Cov}(\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_r}{n-r+1}, \frac{\nu_{r+1}}{n-r}+...+\frac{\nu_s}{n-s+1}).$$

Since sums

$$\frac{\nu_1}{n}+\frac{\nu_2}{n-1}+...+\frac{\nu_r}{n-r+1}$$

and

$$\frac{\nu_{r+1}}{n-r}+...+\frac{\nu_s}{n-s+1}$$

are independent, we get that

$$\mathrm{Cov}(Z_{r,n},Z_{s,n})=$$

$$\mathrm{Cov}\,(\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu r}{n-r+1},\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu r}{n-r+1})=$$

$$\mathrm{Var}\,(\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu r}{n-r+1})=\mathrm{Var}Z_{r,n}=\sum_{k=1}^{r}\frac{1}{(n-k+1)^2}.$$

Exercise 8.10 **(solution).** By virtue of (4.15) we get that

$$W\overset{d}{=}a(\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu r}{n-r+1})+b(\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu s}{n-s+1})=$$

$$(a+b)(\frac{\nu 1}{n}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu r}{n-r+1})+b(\frac{\nu r+1}{n-r}+\frac{\nu 2}{n-1}+\ldots+\frac{\nu s}{n-s+1}).$$

The independence of summands enables us to find the variance of the sum:

$$\mathrm{Var}\,W=(a+b)^2\sum_{k=1}^{r}\frac{1}{(n-k+1)^2}+b^2\sum_{k=r+1}^{s}\frac{1}{(n-k+1)^2}.$$

Chapter 9. Moment relations for order statistics: normal distribution Моментные соотнощения для порядковых статистик: нормальное распределение

В предыдущей главе рассматривались моментные характеристики семейств равномерных и экспоненциальных распределений. Эти распределения часто встречаются в математической статистике и различных приложениях (теория надежности, теория массового обслуживания) теории вероятностей. Имеется еще одно семейство распределений, которое играет весьма важную роль в теории вероятностей и математической статистике. Вспомните хотя бы ***ЦЕНТРАЛЬНУЮ ПРЕДЕЛЬНУЮ ТЕОРЕМУ,*** *играющую существеннейшую роль в асимптотических задачах теории вероятностей. Речь идет о семействе нормальных вероятностных распределений. Можно также добавить, что таблицы стандартного (N(0,1)) нормального распределения можно встретить практически во всех математических справочниках и во многих учебниках по теории вероятностей. В данной главе представлены соотношения для моментов порядковых статистик, соответствующих исходным нормальным распределениям.*

Let X_1, X_2,… be independent random variables having the standard normal distribution and $X_{1,n} \le \ldots \le X_{n,n}$ be the corresponding normal order statistics. It is known that the normal distribution is the most popular in the mathematical statistics. Statistical methods related to normal samples are deeply investigated and have a very long history. However, there are some problems when we want to calculate moments of normal order statistics. Indeed, one can write immediately (remembering the general form for moments of order statistics) that

$$EX^{r}_{k,n} = \frac{n!}{(k-1)!(n-k)!} \int_{-\infty}^{\infty} x^r \Phi^{k-1}(x)(1-\Phi(x))^{n-k} \varphi(x)dx, \qquad (9.1)$$

where

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

and

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

There are effective numerical methods to compute integrals (9.1). Unlike the cases of the uniform and exponential order statistics, moments (9.1) have the explicit expressions only in some special situations for small sample sizes.

Example 9.1. Consider the case n=2. We get that

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

$$2\int_{-\infty}^{\infty} \varphi^2(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty} \exp(-x^2)dx= \frac{1}{\sqrt{\pi}}.$$

From the identity

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

we obtain now that

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

Remark 9.1. If we have two samples $X_1, X_2,\dots, X_n$ (from the standard N(0,1) normal distribution) and $Y_1,Y_2,\dots,Y_n$ (from the normal $N(a,\sigma^2)$ distribution with expectation a and variance σ^2, $\sigma>0$), then evidently

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

Remark 9.2. Any normal $N(a,\sigma^2)$ distribution is symmetric with respect to its expectation a. Hence, it is easy to see that for order statistics $Y_{1,n} \le\dots\le Y_{n,n}$ the following relations are true:

$$EY_{k,n}= 2a - EY_{n-k+1,n},\ k=1,2,\dots,n; \tag{9.2}$$

$$E(Y_{k,n} -a)^m = (-1)^m E(Y_{n-k+1,n} -a)^m,\ k=1,2,\dots,n;\ m=1,2,\dots. \tag{9.3}$$

It follows from (9.3) that if $Y_{k+1,2k+1}$ is a sample median from the normal $N(a,\sigma^2)$ distribution, then

$$E(Y_{k+1,2k+1}- a)^{2r-1}= 0,\ r= 1,2,\dots, \tag{9.4}$$

and, in particular,

$$EY_{k+1,2k+1}= a. \tag{9.5}$$

Exercise 9.1. Let $X_{1,3} \le X_{2,3} \le X_{3,3}$ be the order statistics corresponding to the standard normal distribution. Find $EX_{1,3}$, $EX_{2,3}$ and $EX_{3,3}$.

Exercise 9.2. Let $X_{1,3} \le X_{2,3} \le X_{3,3}$ be the order statistics corresponding to the standard normal distribution. Find $E(X_{k,3})^2$ and $Var(X_{k,3})$, k=1,2,3.

The just-obtained results show that 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.

Example 9.2. A lot of statistical procedures for the normal distribution are based on the independence property of vector

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

and the sample mean

$$\overline{X} = (X_1 + X_2 + \ldots + X_n)/n.$$

What is more important for us, this yields the independence of vector

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

And the sample mean $\overline{X}$.

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

$$E(X_{k,n} - \overline{X})\overline{X} = E(X_{k,n} - \overline{X})E\overline{X} = 0 \qquad (9.6)$$

and we obtain the following results:

$$EX_{k,n}\overline{X} = E(\overline{X}^2) = Var\,\overline{X} = 1/n,\ k=1,2,\ldots,n, \qquad (9.7)$$

and hence

$$\sum_{m=1}^{n} EX_{k,n} X_m = 1, \qquad (9.8)$$

as well as

$$\sum_{m=1}^{n} EX_{k,n} X_{m,n} = 1. \qquad (9.9)$$

As corollaries of (9.8) one gets that

$$E(X_{k,n} X_m)=1/n \qquad (9.10)$$

and

$$cov(X_{k,n} ,X_m)= E(X_{k,n} X_m)- EX_{k,n} EX_m=1/n \qquad (9.11)$$

for any k=1,2,…,n, m=1,2,…,n and n=1,2,….

Similarly to (9.6), we can get also that

$$cov(X_{k,n}- \overline{X} , \overline{X}) = 0$$

and hence

$$cov(X_{k,n}, \overline{X}) = cov(\overline{X} , \overline{X})= Var(\overline{X})=1/n. \qquad (9.12)$$

Since at the same time

$$n \overline{X} = (X_1+X_2+…+X_n)$$

and

$$n \overline{X} = (X_{1,n}+X_{2,n}+…+X_{n,n}), \qquad (9.13)$$

one obtains from (9.12) that

$$\sum_{m=1}^{n} cov(X_{k,n} ,X_{m,n}) =1, k=1,2,…,n, \qquad (9.14)$$

and

$$cov(X_{k,n},X_m)=1/n, 1\le k\le n, 1\le m\le n.$$

The symmetry of the normal distribution with respect to its expectation also gives some simplifications.

Example 9.3. Let again $X_{1,n}$, $X_{2,n}$,…, $X_{n,n}$ be order statistics from the standard normal distribution. The symmetry of the normal distribution implies that

$$(X_{1,n},X_{2,n},…,X_{n,n}) \overset{d}{=} (-X_{n,n},X_{n-1,n},…,X_{1,n}). \qquad (9.15)$$

Hence

$$EX_{k,n}=-EX_{n-k+1,n}, k=1,2,…,n, \qquad (9.16)$$

$$E(X_{k,n})^2 =E(X_{n-k+1,n})^2, k=1,2,…,n, \qquad (9.17)$$

$$Var(X_{k,n})=Var(X_{n-k+1,n}),\ k=1,2,\dots,n. \qquad (9.18)$$

It follows also from (9.15) that

$$E(X_{r,n}\ X_{s,n})=E(X_{n-r+1,n}\ X_{n-s+1,n}) \qquad (9.19)$$

and

$$cov(X_{r,n},\ X_{s,n})=cov\ (X_{n-r+1,n},\ X_{n-s+1,n}) \qquad (9.20)$$

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

Exercise 9.3. Let $X_{1,3} \le X_{2,3} \le X_{3,3}$ be the order statistics corresponding to the standard normal distribution. Find covariances $cov(X_{r,3}\ ,\ X_{s,3})$ for $1\le r<s\le 3$.

In some special procedures we need to find moments of maximal or minimal values of dependent normal random variables.

Example 9.4. Let the initial random variables X_1, X_2 have jointly the bivariate normal

$$N(a_1,a_2,\ \sigma_1^2,\sigma_2^2,\rho)$$

distribution. This means that

$$EX_1 = a_1,\ \ EX_2 = a_2,\ \ VarX_1 = \sigma_1^2,\ \ VarX_2 = \sigma_2^2$$

and the correlation coefficient between X and Y equals ρ. Let us find $EX_{1,2}$ and $EX_{2,2}$.

We have evidently the following equalities:

$$X_{2,2} = \max\{X_1,X_2\}= X_1+\max\{0,X_2-X_1\}$$

and

$$EX_{2,2} = EX_1+E\max\{0,Y\} = a_1+E\max\{0,Z\},$$

where Z denotes the difference X_2-X_1, which also has the normal distribution with expectation

$$b= EZ =E(X_2-X_1)= a_2-\ a_1$$

and variance

$$\sigma^2= VarZ=VarX_1+VarX_2-2cov\{X_1,X_2\} = \sigma_1^2 +\sigma_2^2 -2\rho\sigma_1\sigma_2.$$

If $\sigma^2= 0$, i.e. $Z=X-Y$ has the degenerate distribution, then

$$EX_{2,2} = a_1+\max\{0,b\}=\max\{a_1,a_2\}$$

and

$$EX_{1,2}=E(X_1+X_2)-EX_{2,2}= a_1+a_2 - \max\{ a_1,a_2\}= \min\{ a_1,a_2\}.$$

Consider now the case $\sigma^2>0$. In this situation

$$EX_{2,2} = a_1+ E\max\{0,b+\sigma Y\},$$

Where Y has the standard N(0,1) normal distribution with the density function

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

and distribution function

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

It is not difficult to see that

$$E\max\{0, b+\sigma Y\} =b+E \max\{-b, \sigma Y\}=b+\sigma\, E\max\{c, Y\},$$

where $c= -b/\sigma$.

Further,

$$E\max\{c, Y\} = cP\{ Y\leq c\}+ \int_{c}^{\infty} x\varphi(x)dx= c\Phi(c) - \int_{c}^{\infty} d\varphi(x) =c\Phi(c)+ \varphi(c).$$

We have thus, that

$$EX_{2,2}= a_1+ E\max\{0,b+\sigma Y\}= a_1+ b+\sigma\, E\max\{c, Y\}=$$

$$a_1+b+\sigma(c\Phi(c)+ \varphi(c)) = a_2+\sigma(c\Phi(c)+ \varphi(c))$$

and

$$EX_{1,2} = E(X_1+X_2)- EX_{2,2} = a_1-\sigma(c\Phi(c)+ \varphi(c)).$$

In the partial case, when $a_1= a_2 = a$, we have the equalities

$$EX_{2,2} = a+\sigma\, \varphi(0) = a+\frac{\sigma}{\sqrt{2\pi}}$$

and

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

We considered some situations when it is possible to get the exact expressions for moments of order statistics from the normal populations. Note that there are different tables (see, for example, Teichroew. (1956) or Tietjen , Kahaner and Beckman (1977)) which give expected values and some other moments of order statistics for samples of large sizes from the normal distribution.

Check your solutions

Exercise 9.1 **(solution).** It follows from remark 9.2 that E $X_{2,3}$= 0 and

$EX_{1,3}$ = -$EX_{3,3}$. Thus, we need to find $EX_{3,3}$ only. We see 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

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

and

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

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

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

Exercise 9.2 **(solution).** Due to symmetry of the standard normal distribution,

$$E(X_{1,3})^2 = E(X_{3,3})^2 \quad \text{and} \quad VarX_{1,3} = Var\ X_{3,3}.$$

We have also that

$$E(X_{3,3})^2=3\int_{-\infty}^{\infty} x^2\Phi^2(x)\varphi(x)dx=-3\int_{-\infty}^{\infty} x\Phi^2(x)d\varphi(x)=3\int_{-\infty}^{\infty} \varphi(x)d(x\Phi^2(x))=$$

$$3\int_{-\infty}^{\infty} \varphi(x)\Phi^2(x)dx+6\int_{-\infty}^{\infty} x\varphi^2(x)\Phi(x)dx$$

$$=\int_{-\infty}^{\infty} d(\Phi^3(x))+\frac{3}{\pi}\int_{-\infty}^{\infty} x\exp(-x^2)\Phi(x)dx=$$

$$1-\frac{3}{2\pi}\int_{-\infty}^{\infty} \Phi(x)d(\exp(-x^2))= 1+\frac{3}{2\pi}\int_{-\infty}^{\infty} \exp(-x^2)\varphi(x)dx=$$

$$1+\frac{3}{(2\pi)^{3/2}}\int_{-\infty}^{\infty} \exp(-3x^2/2)dx=1+\frac{\sqrt{3}}{2\pi}.$$

Taking into account that

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

(see exercise 9.1) , one obtains that

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

Further,

$$E(X_{2,3})^2 = 6\int_{-\infty}^{\infty} x^2\Phi(x)(1-\Phi(x))\varphi(x)dx =$$

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

We obtain that

$$\int_{-\infty}^{\infty} x^2\Phi(x)\varphi(x)dx \ = \ -\int_{-\infty}^{\infty} x\Phi(x)d\varphi(x) =$$

$$\int_{-\infty}^{\infty} \varphi(x)d(x\Phi(x)) = \int_{-\infty}^{\infty} \varphi(x)\Phi(x)dx+\int_{-\infty}^{\infty} x\varphi^2(x)dx =$$

$$\frac{1}{2}\int_{-\infty}^{\infty} d(\Phi^2(x))+\frac{1}{2\pi}\int_{-\infty}^{\infty} x\exp(-x^2)dx = \frac{1}{2}.$$

Hence,

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

and

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

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

Thus,

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

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

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

and

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

Let us note that

$$E(X_{1,3})^2 + E(X_{2,3})^2 + E(X_{3,3})^2 = E(X_1)^2+ (EX_2)^2+ (EX_3)^2 =3.$$

Exercise 9.3 **(solution).** It follows from (9.20) that

$$\mathrm{cov}(X_{1,3}, X_{2,3}) = \mathrm{cov}\,(X_{2,3}, X_{3,3}).$$

From (9.14) we have also that

$$\mathrm{cov}(X_{1,3}, X_{2,3})+\mathrm{cov}(X_{2,3}, X_{2,3})+\mathrm{cov}(X_{2,3}, X_{3,3}) =1.$$

On combining these relations we get equalities

$$\mathrm{cov}(X_{1,3}, X_{2,3}) = \mathrm{cov}(X_{2,3},X_{3,3}) = (1-\mathrm{cov}(X_{2,3},X_{2,3})) = (1-\mathrm{Var}(X_{2,3}))/2.$$

It was found in exercise 9.2 that

$$\mathrm{Var}(X_{2,3}) = 1-\frac{\sqrt{3}}{\pi}.$$

Finally,

$$\mathrm{cov}(X_{1,3}, X_{2,3}) = \mathrm{cov}(X_{2,3},X_{3,3}) = \frac{\sqrt{3}}{2\pi}.$$

Now we again use (9.14):

$$\mathrm{cov}(X_{1,3}, X_{1,3})+\mathrm{cov}(X_{1,3}, X_{2,3})+\mathrm{cov}(X_{1,3}, X_{3,3}) =1$$

and have the equality

$$\mathrm{cov}(X_{1,3}, X_{3,3}) =1-\mathrm{cov}(X_{1,3}, X_{2,3}) -\mathrm{Var}(X_{1,3}).$$

It was found in exercise 9.2 that

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

Thus,

$$\mathrm{cov}(X_{1,3}, X_{3,3}) = 1-\frac{\sqrt{3}}{2\pi}-(1+\frac{\sqrt{3}}{2\pi}-\frac{9}{4\pi}) = \frac{9}{4\pi}-\frac{\sqrt{3}}{\pi} =\frac{9-4\sqrt{3}}{4\pi}.$$

Учитывая важность различных моментных характеристик порядковых статистик, соответствующих семейству нормальных распределений, приводим ряд ссылок на работы, которые позволят более подробно познакомиться с такого рода моментами.

1. Bose, R.C. and Gupts, S.S. (1959). Moments of order statistics from a normal distribution. Biometrika, **46,** 433-440.

2. Davis, C. S. and Stephens, M. A. (1977). The covariance matrix of normal order statistics, *Commun. Statist., **B6***, 75-81.

3. Govindarajulu, Z. (1963). On moments of order statistics and quasi-ranges from normal populations. *Ann. Math. Statist*., ***34,*** 633-651.

4. Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika ,* **48,** 151-165. Correction **48,** 476.

5. Joshi, P. C. and Balakrishnan, N. (1981). An identity for the moments of normal order statistics with applications, *Scand. Actuar. J.,* 203-213.

6. Ruben, H. (1954). On the moments of order statistics in samples from normal populations. *Biometrika*, **41,** 200-227.

7. Teichroew, D. (1956). Tables of expected values of order statistics and products of order statistics for samples of size twenty and less from the normal distribution. *Ann. Math*. *Statist.,* **27,** 410-426.

8. Tietjen, G. L., Kahaner, D. K., and Beckman, R. J. (1977). Variances and covariances of the normal order statistics for sample sizes 2 to 50. *Selected Tables in Mathematical Statistics,* **5**, 1-73.

Chapter 10. Asymptotic behavior of the middle and intermediate order statistics
Асимптотическое поведение средних и промежуточных порядковых статистик

В предыдущих главах были получены выражения для функций распределения, для плотностей порядковых статистик. При больших значениях n эти выражения представляются весьма громоздкими. С ними трудно работать. Существенно эти формулы упрощаются, если вместо них рассматривать соответствующие асимптотические соотношения. В данной главе исследуется асимптотика средних членов вариационного ряда и промежуточных порядковых статистик.

middle order statistics = средние порядковые статистики (средние члены вариационного ряда)

intermediate order statistics = промежуточные порядковые статистики

extreme order statistics = экстремальные порядковые статистики

asymptotic distributions = асимптотические (предельные) распределения

Lyapunov ratio = дробь Ляпунова

It turns out that one, who wants to study asymptotic distributions of suitably normalized and centered order statistics $X_{k,n}$, must distinguish three different options. Since we consider the case, when n (the size of the sample) tends to infinity, it is natural that k =k(n) is a function of n. The order statistics $X_{k(n),n}$ are said to be ***extreme*** if k=k(n) or n-k(n)+1 is fixed, as n→∞. If

$$0<\liminf_{n\to\infty} k(n)/n \le \limsup_{n\to\infty} k(n)/n<1,$$

then order statistics $X_{k(n),n}$ are said to be ***middle***. At last, the case when

k(n) →∞, k(n)/n→0 or n-k(n) →∞, k(n)/n→1, corresponds to the so-called ***intermediate*** order statistics.

Exercise 10.1. Show that sample quantiles of order p, 0<p<1, present middle order statistics.

The opportunity to express the uniform and exponential order statistics via sums of independent terms enables us to use limit theorems for sums of independent random variables. Really we need to know the following Lyapunov theorem.

Theorem 10.1. Let X_1, X_2,… be independent random variables with expectations $a_k = EX_k$, variances σ_k^2 and finite moments $\gamma_k = E|X_k-a_k|^3$, k=1,2,…. Denote

$$S_n = \sum_{k=1}^{n} (X_k-a_k), \qquad B_n^2 = \mathrm{Var}(S_n) = \sum_{k=1}^{n} \sigma_k^2$$

and

$$L_n = \sum_{k=1}^{n} \gamma_k / B_n^3 . \tag{10.1}$$

If Lyapunov ratio (10.1) converges to zero, as n→∞, then

$$\sup_x |P\{S_n/B_n<x\}- \Phi(x)| \to 0, \tag{10.2}$$

where

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

is the distribution function of the standard normal distribution.

Example 10.1. Consider a sequence of exponential order statistics $Z_{k(n),n}$, n=1,2,…, where k(n) →∞ and limsup k(n)/n<1, as n→∞. As we know from lecture 4, the following relation is valid for the exponential order statistics:

$$Z_{k,n} \overset{d}{=} \frac{\nu_1}{n} + \frac{\nu_2}{n-1} + \ldots + \frac{\nu_k}{n-k+1}, \ k=1,2,\ldots,n,$$

where ν_1, ν_2,… is a sequence of independent identically distributed random variables, having the standard exponential E(1) distribution. Let us check if the Lyapunov theorem can be used for independent terms

$$X_k = \frac{\nu_k}{n-k+1}.$$

We obtain easily that

$$a_k = EX_k = 1/(n-k+1), \sigma_k^2 = \mathrm{Var}(X_k) = 1/(n-k+1)^2$$

and

$$\gamma_k = \gamma/(n\text{-}k+1)^3,$$

where

$$\gamma = E|\nu\text{-}1|^3 = \int_0^1 (1\text{-}x)^3 e^{-x}dx + \int_1^\infty (x\text{-}1)^3 e^{-x}dx = 6\text{-}10/e. \qquad (10.3)$$

In our case

$$B_{k,n}^2 = \mathrm{Var}(Z_{k,n}) = \sum_{r=1}^{k} \sigma_r^2 = \sum_{r=1}^{k} \frac{1}{(n-r+1)^2} = \sum_{r=n-k+1}^{n} \frac{1}{r^2} \qquad (10.4)$$

and

$$\Gamma_{k,n} = \sum_{r=1}^{k} E|X_r\text{-}a_r|^3 = \gamma \sum_{r=1}^{k} \frac{1}{(n-r+1)^3} = \gamma \sum_{r=n-k+1\text{`}}^{n} \frac{1}{r^3}. \qquad (10.5)$$

The restriction limsup k(n)/n <1, as n→∞, means that there exists such *p*, 0<p<1, that k ≤ pn, for all sufficiently large n. For such values *n* we evidently have the following inequalities:

$$B_{k,n}^2 \geq k/n^2 \ , \ \Gamma_{k,n} \leq k\gamma/(n(1\text{-}p))^3 \ ,$$

$$L_{k,n} = \Gamma_{k,n}/\, B_{k,n}^3 \leq \gamma(1\text{-}p)^3/k^{1/2}. \qquad (10.6)$$

Hence, if limsup k(n)/n<1 and k(n) →∞, as n→∞, then the Lyapunov ratio $L_{k(n),n}$ tends to zero and this provides the asymptotic normality of

$$(Z_{k(n),n} - A_{k,n})/B_{k,n},$$

where

$$A_{k,n} = \sum_{r=1}^{k} 1/(n\text{-}r+1).$$

Remark 10.1. It is not difficult to see that that example 10.1 covers the case of sample quantiles $Z_{[pn]+1,n}$, 0<p<1. Moreover if k(n)~pn, 0<p<1, as n→∞, then

$$A_{k,n}=\sum_{r=1}^{k} 1/(n-r+1) = \sum_{r=n-k+1}^{n} 1/r \sim -\log(1-p),\ n\to\infty,$$

and

$$B_{k,n}^{2} = \sum_{r=n-k+1}^{n} 1/r^{2} \sim \int_{n-k+1}^{n} x^{-2}dx \sim p/(1-p)n\ ,$$

as n→∞.

All the saying with some additional arguments enable us to show that for any $0<p<1$, as n→∞,

$$\sup_{x} |P\{(Z_{[pn]+1,n} +\log(1-p))(n(1-p)/p)^{1/2} <x\}- \Phi(x)| \to 0. \qquad (10.7)$$

In particular, the following relation is valid as n→∞ for the exponential sample median:

$$\sup_{x} |P\{n^{1/2} (Z_{[n/2]+1,n} -\log 2) <x\}- \Phi(x)| \to 0. \qquad (10.8)$$

Remark 10.2. More complicated and detailed arguments allow to prove that suitably centered and normalized exponential order statistics $Z_{k(n),n}$ have the asymptotically normal distribution if $\min\{k(n), n-k(n)+1\}\to \infty$, as n→∞.

Exercise 10.2. Let

$$S_n = \sum_{k=1}^{n} d_k Z_{k,n}$$

and

$$f_k(n)= \sum_{m=k}^{n} d_m/(n-m+1),\ k=1,2,\ldots n.$$

Let also

$$\sum_{k=1}^{n} |f_k(n)|^3/(\sum_{k=1}^{n} f_k^2(n))^3 \to 0, \qquad (10.9)$$

as n→∞. Show that then random variables

$$(S_n - \sum_{k=1}^{n} f_k(n)) / (\sum_{k=1}^{n} f_k^2(n))^{1/2}$$

converge in distribution to the standard normal distribution.

Exercise 10.3. Let $T_{1,n} = Z_{1,n}$ and $T_{k,n} = Z_{k,n} - Z_{k-1,n}$, $k=2,\ldots,n$, $n=1,2,\ldots$,

be the exponential spacings and

$$R_n = \sum_{k=1}^{n} b_k T_{k,n}.$$

Reformulate the result of exercise 10.2 for sums R_n.

Example 10.2. Consider now the uniform order statistics

$$0 = U_{0.n} \leq U_{1,n} \leq \ldots \leq U_{n,n} \leq U_{n+1,n} = 1,\ n=1,2,\ldots.$$

From chapter 4 we know that $U_{k,n}$ has the same distribution as the ratio S_k/S_{n+1}, where

$$S_m = \sum_{k=1}^{m} \nu_k,\ m=1,2,\ldots,$$

are the sums of independent E(1) random variables $\nu_1, \nu_2, \ldots$. We know also that $EU_{k,n}=k/(n+1)$. Now we can see that

$$P\{b_n(U_{k,n} - \frac{k}{n+1}) < x\} = P\{b_n(\frac{S_k}{S_{n+1}} - \frac{k}{n+1}) < x\} =$$

$$P\{\frac{b_n}{S_{n+1}}((n+1-k)S_k - k(S_{n+1}-S_k)) < x\}. \qquad (10.10)$$

Due to the law of large numbers

$$S_{n+1}/(n+1) \to 1$$

in probability. Hence if

$$\frac{b_n}{n+1}((n+1-k)S_k - k(S_{n+1}-S_k)) = \frac{b_n}{n+1}((n+1-k)(S_k-k) - k(S_{n+1}-S_k-(n-k+1)))$$

has some asymptotic distribution under the suitable choice of norming constants b_n, then this distribution is also asymptotic for

$$\frac{b_n}{S_{n+1}}((n+1-k)S_k-k(S_{n+1}-S_k)).$$

Since random variables (S_k-k) and $(S_{n+1}-S_k-(n-k+1))$ are independent and

$(S_k-k)/k^{1/2}$ and $(S_{n+1}-S_k-(n-k+1))/(n-k+1)^{1/2}$ asymptotically have the standard normal

distribution if $\min\{k,n-k+1\}\to\infty$, as $n\to\infty$, we find that

$$((n+1-k)S_k - k(S_{n+1}-S_k))/(n+1)^{1/2}$$

also converges to the corresponding normal law distribution. Moreover, if

$$\min\{k,n-k+1\}\to\infty, \text{ as } n\to\infty,$$

and

$$b_n = n^{3/2} /k(n-k+1)\}^{1/2}$$

then it provides the convergence of random variables

$$n^{3/2} (U_{k,n}-\frac{k}{n+1})/k^{1/2}(n-k+1)^{1/2}$$

to the standard normal law.

Note that there are some estimates of the rate of convergence to the normal law of suitable centered and normalized uniform order statistics $U_{k,n}$. For example, the following inequality holds for any n=1,2,… and $1\le k\le n$:

$$\sup_x |P\{U_{k,n} -\frac{k}{n+1}< xk^{1/2}(n-k+1)^{1/2}/n^{3/2}\}- \Phi(x)| \le C(k^{-1/2} +(n-k+1)^{-1/2}),$$

where C is some absolute constant.

Exercise 10.4. Consider the uniform sample quantiles $U_{qn,n}$, $0<q<1$, and prove that

$$(U_{qn,n}- q)n^{1/2}/q^{1/2}(1-q)^{1/2}$$

converges in distribution to the standard normal distribution.

Exercise 10.5. Let

$$T_n = \sum_{k=1}^{n} c_kU_{k,n}.$$

Denote $b_{n+1}=0$ and

$$b_m = \sum_{k=m}^{n} c_k,\ m=1,2,\ldots,n.$$

Let

$$b(n)= \sum_{k=1}^{n+1} b_k/(n+1)$$

and

$$D^2(n)= \frac{1}{(n+1)^2} \sum_{k=1}^{n+1} (b_k-b(n))^2.$$

Show that if

$$L_n = \sum_{k=1}^{n+1} |b_k-b(n)|^3/(\sum_{k=1}^{n+1} (b_k-b(n))^2)^{3/2} \to 0,$$

as n→∞, then random variables $(T_n-b(n))/D(n)$ have asymptotically the standard normal distribution.

Example 10.3. Consider sample quantiles $X_{qn,n}$, $0 <q <1$, corresponding to the cdf $F(x)$ and density $f(x)$. Suppose that $f(x) >0$ and $f(x)$ is continuous in some neighborhood of the point $G(q)$, where $G(x)$ is the inverse of F. From exercise 10.4 we know for the uniform sample quantiles $U_{qn,n}$ that

$$(U_{qn,n}- q)n^{1/2}/q^{1/2}(1-q)^{1/2}$$

converges to the standard normal distribution. We can try express sample quantile $X_{qn,n}$ via the corresponding uniform sample quantile as follows:

$$X_{qn,n} = G(U_{qn,n}) = G(q) + (U_{qn,n} –q)G'(q) + R_n, \qquad (10.11)$$

where $R_n= 0(|U_{qn,n}-q|)$ and

$$P\{|R_n|>\varepsilon\}\leq E(U_{qn,n}-q)^2/\varepsilon^2 \leq q(1-q)/n\varepsilon^2. \qquad (10.12)$$

We can see from (10.11) and (10.12) that $X_{qn,n}$ has the same asymptotics as

$$G(q)+(U_{qn,n} –q)G'(q).$$

Hence,

$$(X_{qn,n}-G(q))\sqrt{\frac{n}{q(1-g)}}\,/\,G'(q)$$

has the same limit distribution as

$$(U_{qn,n}- q)n^{1/2}/q^{1/2}(1-q)^{1/2},$$

that is

$$(X_{qn,n}-G(q))\sqrt{\frac{n}{q(1-g)}}/G'(q)$$

converges to the standard normal distribution. We can also find that

$$1/G'(g)=f(G(g))$$

and then write that

$$(X_{qn,n}-G(q))f(G(q))\sqrt{\frac{n}{q(1-g)}}$$

asymptotically has the standard normal distribution. We can express this result as

$$X_{qn,n} \sim N(G(g), q(1-g)/nf^2(G(q))), \qquad (10.13)$$

which means that a sequence $X_{qn,n}$ behaves asymptotically as a sequence of normal random variables with expectation $G(q)$ and variance $q(1-q)/nf^2(G(q))$.

Exercise 10.6*.* Consider sample medians $X_{k+1,2k+1}$ for the normal distribution with density function

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

and for the Laplace distribution with density function

$$f(x)=\frac{1}{2}\exp(-|x|)$$

and investigate the corresponding limit distributions.

Check your solutions

***Exercise 10.1* (solution)**. As we know, sample quantiles of order *p* are defined as $X_{[pn]+1,n}$. It suffices to see that

$$0<\lim_{n\to\infty}([pn]+1)/n = p <1$$

for any 0<p<1.

***Exercise 10.2* (hint).** As in example 10.1 use the representation for the exponential order statistics to express S_n as the sum

$$\sum_{k=1}^{n} f_k(n)\nu_k ,$$

where ν_1, ν_2,… is a sequence of independent exponential E(1) random variables. Now it suffices to consider the corresponding Lyapunov ratio L_n and to show that condition (10.9) provides the convergence L_n to zero.

***Exercise 10.3* (hint).** Express R_n as

$$\sum_{k=1}^{n} d_k Z_{k,n},$$

where $d_k= b_k-b_{k+1}$, k=1,2,…,n, and $d_n=b_n$.

***Exercise 10.4* (hint).** Use example 10.2 for k=qn.

***Exercise 10.5* (solution).** Applying the representation of the uniform order statistics $U_{k,n}$ as the ratio S_k/S_{n+1}, where

$$S_m = \sum_{k=1}^{m} \nu_k,$$

one gets that

$$T_n - ET_n = T_n - \sum_{k=1}^{n} c_k E\,\frac{S_k}{S_{n+1}} = T_n - \sum_{k=1}^{n} \frac{kc_k}{n+1} = T_n\text{-}b(n) =$$

$$\sum_{k=1}^{n} c_k U_{k,n} - b(n) \overset{d}{=} \frac{1}{S_{n+1}} \sum_{k=1}^{n+1} (b_k\text{-}b(n))\,\nu_k.$$

It is easy to see that $S_{n+1}/(n+1)$ converges to one in distribution. Hence to show that $(T_n-b(n))/D(n)$ asymptotically has the normal distribution it suffices to prove the asymptotical normality of the sums

$$W_n=\frac{1}{(n+1)D(n)}\sum_{k=1}^{n+1}(b_k-b(n))\,\nu_k$$

of independent random variables. We see that

$$EW_n=0,\ \ \mathrm{Var}\,W_n=\sum_{k=1}^{n+1}(b_k-b(n))^2/(n+1)^2D^2(n)=1,$$

and the Lyapunov ratio for the sum W_n coincides with γL_n, where

$$\gamma=E|\nu-1|^3=6-10/e$$

is the third central moment for the exponential distribution and

$$L_n=\sum_{k=1}^{n+1}|b_k-b(n)|^3/(\sum_{k=1}^{n+1}(b_k-b(n))^2)^{3/2}.$$

Hence, if L_n converges to zero as *n* tends to infinity, then $(T_n-b(n))/D(n)$ converges to the standard normal distribution.

Exercise 10.6 **(answers).** In the both given situations $q=1/2$, $Q(q)=Q(1/2)=0$. We have

$$f(Q(q))=\frac{1}{\sqrt{2\pi}}$$

for the normal distribution and

$$f(Q(q))=1/2$$

for the Laplace distribution. Using notation (10.13) we have

$$X_{k+1,2k+1}\sim N(0,\pi/2(2k+1))$$

for the normal distribution and

$$X_{k+1,2k+1}\sim N(0,1/(2k+1))$$

for the Laplace distribution.

Chapter 11. Asymptotic behavior of the extreme order statistics Асимптотическое поведение экстремальных порядковых статистик

В предыдущей главе исследовалось предельное распределение средних и промежуточных членов вариационного ряда. С точки зрения многих областей человеческой деятельности (теория надежности, страхование, спортивные рекорды) необходимо знать асимптотическое поведение крайних членов вариационного ряда (экстремальных порядковых статистик). Экстремумы всегда вызывали и вызывают естественный интерес ученых и практиков, о чем, например, свидетельствует набор (далеко не включающий все возможные публикации по данной тематике) работ по данной тематике, приведенный в конце данной главы.

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 $X_{1,n}$. These situations correspond to $X_{n-k(n)+1}$ and $X_{k(n),n}$ for the case k=k(n)=1.

Exercise 11.1. Check that if $X \overset{}{=} -Y$, then the following relation holds for extremes $X_{n,n}=\max\{X_1,X_2,\ldots,X_n\}$ and $Y_{1,n} = \min\{Y_1,Y_2,\ldots,Y_n\}$:

$$X_{n,n} \overset{d}{=} -Y_{1,n}.$$

Remark 11.1. Indeed, the assertion given in exercise 11.1 can be generalized as follows:

$$\text{if } X \overset{d}{=} -Y, \text{ then } X_{n-k+1,n} \overset{d}{=} -Y_{k,n}, \text{ for any } k=1,2,\ldots,n.$$

Exercise 11.2. Let F(x) and f(x) be the distribution function and density function of X correspondingly. Find the joint distribution function $F_n(x,y)$ and the joint density $f_n(x,y)$ of $X_{1,n}$ and $X_{n,n}$, n=2,3,….

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

Example 11.1. Due to the mentioned above relationship between maximal and minimal order statistics we can study one type of them, say, maximal ones. Consider a sequence of order statistics $M(n) =X_{n,n}$, n=1,2,…. 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 in probability to infinity. In the case, when $\beta<\infty$, we need to distinguish two situations.

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

$$P\{M(n)= \beta\}=1-P\{M(n)< \beta\}=1 -P^n(X<\beta)= 1-(1-p)^n \qquad (11.1)$$

and

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

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 to 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 (11.1) gives completed information on M(n) and in this case any centering and norming can not improve our knowledge about the asymptotic behavior of M(n). It is clear also that if $M(n) \to\beta < \infty$, then we can choose such norming constants $b(n) \to\infty$, that M(n)/b(n) converges to zero. The similar situation is also valid for the case $\beta=\infty$. In fact, if $\sup\{x: F(x) <1\}=\infty$, we can take such sequence d_n that

$$P\{X>d_n\}=1-F(d_n)<1/n^2.$$

Indeed, $d_n\to\infty$ as $n \to\infty$. Then for any $\varepsilon>0$ and for sufficiently large n, one gets that $\varepsilon d_n >1$, and

$$P\{M(n)/(d_n)^2 >\varepsilon\}\leq P\{M(n)>d_n\}=1-P\{M(n)\leq d_n\}=$$

$$1-F^n(d_n) \leq 1-(1-\frac{1}{n^2})^n\to 0,$$

as $n \to\infty$. It means that in this case there exists also a sequence b(n), say $b(n)= d_n^2$, such that $M(n)/b(n) \to 0$ in probability (and hence, in distribution). Thus, we see that the appropriate constants can provide the convergence of the normalized maxima M(n)/b(n) to zero. Now the question arises, if it is possible to find centering constants a(n), for which the sequence (M(n)-a(n)) converges to some degenerate distribution. Indeed, if $\beta<\infty$, then, as we know, $M(n) \to\beta$, and hence any centering of the type M(n)-a(n), where $a(n) \to a$ (and, in particular a(n)=0, n=1,2,…) gives the degenerate limit distribution for the sequence M(n)-a(n). Let us consider now the case, when $\beta=\infty$. We can show that in this case no centering (without norming) can provide the convergence M(n)-a(n) to some finite constant. To illustrate this assertion one can take the standard exponential distribution with cdf

F(x)=1-exp(-x), x>0. In this situation

$$P\{M(n)-a_n<x\}=F^n(x+a_n)=(1-\exp(-x-a(n)))^n.$$

Note that if a_n=logn, n→∞, then

$$P\{M(n)-\log n<x\}=(1-\frac{1}{n}\exp(-x))^n \to \exp(-\exp(-x)) \qquad (11.2)$$

and the RHS of (11.2) presents non-degenerate distribution. Let a random variable η have the cdf $H_0(x)=\exp(-\exp(-x))$. Then we can rewrite (11.2) in the form

$$M(n)-\log n \xrightarrow{d} \eta. \qquad (11.3)$$

If we choose other centering constants a_n, then

$$M(n)-a_n=M(n)-\log n-(a_n-\log n)$$

and we see that ($M(n)-a_n$) has a limit distribution only if the sequence (a_n-logn) converges to some finite constant *a*. Then

$$M(n)-a_n \xrightarrow{d} \eta-a. \qquad (11.4)$$

This means that limit distribution functions of centering maxima of independent exponentially distributed random variables ($M(n)-a_n$) must have the form

$$H_a(x)=\exp(-\exp(-x-a))$$

and these distributions are nondegenerate for any *a*.

A new problem arises: 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 ?

Exercise 11.3. Consider the exponential distribution from example 11.2. Let it be known that

$$(M(n)-a_n)/b_n,\ b_n>0,$$

Converges to some nondegenerate distribution with some d.f. G(x). Show that then

$$b_n\to b,\ a_n-\log n \to a,$$

as n →∞, where *b*>0 and *a* are some finite constants, and $G(x)=H_0(a+xb)$ with

$$H_0(x)=\exp(-\exp(-x)).$$

Exercise 11.4. Let X have the uniform U([-1,0]) distribution. Find the limit distribution of nM(n).

Exercise 11.5. Consider X 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 form

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

and

$$H_{1,\alpha}(x)=1,\ x>0. \qquad (11.5)$$

Exercise 11.6. Let X have Pareto distribution with d.f. $F(x)=1-x^{-\alpha}$, $x>1$. Prove that the asymptotic d.f. of $M(n)/n^{1/\alpha}$ is of the form:

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

and

$$H_{2,\alpha}(x)=\exp\{-x^{-\alpha}\},\ x\geq 0. \qquad (11.6)$$

Remark 11.2. Changing suitably the normalized constants a_n and b_n for maximal values considered in exercises 11.5, one gets that any d.f. of the form $H_{1,\alpha}(a+bx)$, where $b>0$ and a are arbitrary constants, can serve as the limit distribution for $(M(n)-a_n)/b_n$. Analogously, making some changing in exercise 11.6 we can prove that any d.f. of the form $H_{2,\alpha}(a+bx)$ also belongs to the set of the limit distributions for the suitably normalized maximal values.

These facts are based on the following useful result, which is due to Khinchine.

Lemma 11.1. Let a sequence of d.f.'s F_n converge weakly to a nondegenerate d.f. G, i.e.

$$F_n(x)\rightarrow G(x),\ n\rightarrow\infty,$$

for any continuity point of G. Then the sequence of d.f.'s

$$H_n(x)=F_n(b_nx+a_n)$$

converges, for some constants a_n and $b_n>0$, to a non-degenerate d.f. *H* if and only if

$$a_n\rightarrow a,\ b_n\rightarrow b,\ n\rightarrow\infty,$$

where $b>0$ and a are some constants, and $H(x)=G(bx+a)$.

Lemma 11.1 can be rewritten in the following equivalent form.

Lemma 11.2. Let F_n be a sequence of d.f.'s and let there exist sequences of constants $b_n>0$ and a_n such that

$$H_n(x)=F_n(a_n+b_nx)$$

weakly converge to and a non-degenerate d.f. G(x). Then for some sequences of constants $\beta_n>0$ and α_n, d.f.'s $R_n(x)= F_n(\alpha_n+\beta_n x)$ weakly converge to a non-degenerate d.f. G*(x) if and only if for some b>0 and a

$$\lim_{n\to\infty} \beta_n/b_n = b$$

and

$$\lim_{n\to\infty} (\alpha_n-a_n)/b_n= a.$$

Moreover, in this situation G*(x)=G(bx+a).

As corollary, we obtain from lemma 11.2 that if a sequence of maximal values $(M(n)-a_n)/b_n$ has some non-degenerate limit distribution then a sequence of $(M(n)- a_n)/ \beta_n$ has the same distribution if and only if $\beta_n \sim b_n$ and $(\alpha_n-a_n)/b_n \to 0$, as $n\to\infty$.

Considering two d.f.'s, G(d+cx) and G(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 G(x), having the most simplest (or convenient) form, is chosen to represent all the distributions of the given type, which we call then *G*-type. As basic for their own types, we suggested above the following distributions:

$$H_0(x)=\exp(-\exp(-x));$$

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

$$H_{2,\alpha}(x)=0,\ x<0, \text{ and } H_{2,\alpha}(x)=\exp\{-x^{-\alpha}\},\ x\ge 0,$$

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 *Gumbe*l, *Frechet* and *Weibull* types of 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 $H_{2,\beta}$, $\alpha\neq\beta$.

It is surprising that we can not obtain some new 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 11.3. We mentioned above that the set of all possible limit distributions for maximal values includes d.f.'s H_0, $H_{1,\alpha}$ and $H_{2,\alpha}$ only. Hence it is important for a statistician to be able to determine 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, 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 $F \in D(H_{1,1})$. 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 their form is rather cumbersome.

As an example of the corresponding results one can find below the following theorem which is due to Gnedenko (Gnedenko, 1943).

Theorem 11.1. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with distribution function F(x) and let b = sup {x: F(x) < 1}. Then *F* belongs $D(H_0)$, if and only if there exists a positive function g(t) such that

$$\frac{(1-F(t+xg(t)))}{1-F(t)} = \exp(-x),$$

for all real x.

The analogous conditions were obtained for *F* which belong $D(H_{1,\alpha})$ and $D(H_{2,\alpha})$.

Hence, simple sufficient conditions are more interesting for us. We present below some of them.

Theorem 11.2. Let d.f. F have positive derivative F′ for all $x>x_0$. If the following relation is valid for some $\alpha>0$:

$$xF'(x)/(1-F(x)) \to \alpha, \qquad (11.7)$$

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

$$a_n=0 \quad \text{and} \quad F(b_n) = 1-1/n. \qquad (11.8)$$

Theorem 11.3. 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

$$(x-x_0)F'(x)/(1-F(x)) \to \alpha, \; x \to x_0, \qquad (11.9)$$

ehen $F \in D(H_{1,\alpha})$. The centering, a_n, and normalizing, b_n, constants can be taken to satisfy relations

$$a_n=x_0 \quad \text{and} \quad F(x_0 - b_n)=1-1/n.$$

Theorem 11.4. 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

$$F''(x)(1-F(x))/(F'(x))^2 = -1,$$

then $F\in D(H_0)$. The centering, a_n, and normalizing, b_n, constants can be taken to satisfy relations

$$F(a_n)=1-1/n \text{ and } b_n=h(a_n),$$

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

Exercise 11.7. Let

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

(the Cauchy distribution). Prove that $F\in D(H_{2,\alpha})$. What normalizing constants, a_n and b_n, can be taken in this case?

Exercise 11.8. Let

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

Show that $F\in D(H_0)$ and the normalizing constants can be taken as follows:

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

Exercise 11.9. Use theorem 11.3 to find the limit distribution and the corresponding normalizing constants for gamma distribution with p.d.f.

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

Example 11.2. Above we considered the most popular distributions (exponential, uniform, normal, Cauchy, gamma, Pareto) and found the correspondence between these distributions and limit distributions for maximal values. It is interesting to mention that there are distributions, which are not in the domain of attraction of any limit law, H_0, $H_{1,\alpha}$ or $H_{2,\alpha}$. In fact, let $X_1,X_2,\ldots$ have geometric distribution with probabilities

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

Then

$$F(x)=P\{X\le x\}=1-p^{[x+1]},\ x\ge 0,$$

Where $[x]$ denotes the entire part of x. Suppose that $H(x)$ (one of functions H_0, $H_{1,\alpha}$, $H_{2,\alpha}$) is a limit d.f. for $(M(n)-a_n)/b_n$ under the suitable choice of normalizing constants a_n and b_n. It means that

$$F^n(a_n+xb_n)\to H(x),\ n\to\infty,$$

or

$$n\log(1-(1-F(a_n+xb_n)) \to \log H(x). \qquad (11.10)$$

Due to the relation $\log(1-x)\sim -x, x\to 0$, we get from (11.10) that

$$n(1-F(a_n+xb_n)) \to -\log H(x),$$

and hence

$$\log n+\log(1-F(a_n+xb_n)) \to h(x)=\log(-\log H(x)), \qquad (11.11)$$

as $n\to\infty$, for all x such that $0<H(x)<1$. In our case

$$\log(1- F(a_n+xb_n))=[1+ a_n+xb_n]\log p$$

and (11.11) can be rewritten as

$$\log n +[1+ a_n+xb_n]\log p\to h(x).$$

One can express $[1+a_n+xb_n]$ as $a_n+\gamma_n + xb_n$, where $0\le \gamma_n<1$. Then

$$(\log n +(a_n+ \gamma_n)\log p)+xb_n\log p\to h(x). \qquad (11.12)$$

It follows from (11.12) that $b_n\to b$, where b is a positive constant. We get from lemma 11.2 that if

$$F^n(a_n+xb_n) \to H(x),$$

then

$$F^n(a_n+xb) \to H(x). \qquad (11.13)$$

Fix some *x* and take $x_1 =x-1/3b$ and $x_2=x+1/3b$. The difference between a_n+x_2b and

a_n+x_1b is less than one. It means that for any *n*, at least two of three points $[a_n+x_1b]$, $[a_n+xb]$ and $[a_n+x_2b]$ must coincide. Then sequences

$$F^n(a_n+x_1b)=(1-p^{[a_n+bx_1]})^n,$$

$$F^n(a_n+x_1b)= (1-p^{[a_n+bx]})^n,$$

and

$$F^n(a_n+x_1b)= (1-p^{[a_n+bx_2]})^n$$

can not have three different limits $H(x_1) < H(x)< H(x_2)$. This contradicts to proposal that $(M(n)-a_n)/b_n$ has a non-degenerate limit distribution.

Returning to results of exercise 11.1 one can find the possible types of limit distributions for minimal values $m(n)=\min\{X_1,\dots,X_n\}$. 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:

$$L_0(x) = 1-\exp(-\exp(x));$$

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

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

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 k-th extremes - order statistics $X_{n-k+1,n}$ and $X_{k,n}$, when k =2,3, … is some fixed number and n tends to infinity. The following results are valid in these situations.

Theorem 11.5. Let random variables X_1, X_2,… be independent and have a common d.f. *F* and $X_{n-k+1,n}$, n=k,k+1,… , be the (n-k+1)-th order statistics. If for some normalizing constants a_n and b_n,

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

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

$$P\{(X_{n-k+1,n} - a_n)/b_n<x\}\to T(x)\sum_{j=0}^{k-1} (-\log T(x))^j/j! \tag{11.14}$$

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

Theorem 11.6. Let random variables X_1,X_2,… be independent and have a common d.f. *F* and $X_{k,n}$, n=k,k+1,… , be the *k*-th order statistics. If for some normalizing constants a_n and b_n,

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

in distribution, then the limit relation

$$P\{(X_{k,n} - a_n)/b_n<x\}\to H(x)\sum_{j=0}^{k-1} (-\log H(x))^j/j! \tag{11.15}$$

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

Exercise 11.10. Let F(x) =1-exp(-x), x>0. Find the corresponding limit (as n $\to\infty$) distributions, when k=2,3,… is fixed, for sequences $Y_n= (X_{n-k+1,n} - \log n)$.

Check your solutions

Exercise 11.1 (solution). Let F(x) =P{X≤x}. We must compare

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

and

$$P\{-Y_{1,n}\le x\}=P\{Y_{1,n}\ge -x\}=P^n\{Y\ge -x\} = P^n\{-Y\le x\}=P^n\{X\le x\}=F^n(x).$$

Thus, we see that

$$X_{n,n} \overset{d}{=} -Y_{1,n}.$$

Exercise 11.2 (solution). It is clear that $F_n(x,y)=F^n(y)$, if $x\ge y$. Let $x<y$. Then

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

$$F^n(y) - P^n\{x<X\le y\} = F^n(y)-(F(y)-F(x))^n.$$

Now it is evident, that $f_n(x,y) = 0$, if $x\ge y$, and $f_n(x,y) = n(n-1)(F(y)-F(x))^{n-2}f(y)f(x)$, if $x<y$.

Exercise 11.3 (solution). From (11.3) we know that for any x,

$$P\{M(n)-\log n\le x\} \to H_0(x), \quad n \to\infty. \tag{11.16}$$

Moreover, since $H_0(x)$ is a continuous d.f., it follows from (11.16) that

$$\sup_x |P\{M(n)-\log n\le x\} - H_0(x)| \to 0. \tag{11.17}$$

From condition of exercise 11.3 we have also that

$$P\{(M(n)-a_n)/b_n \le x \} \to G(x), \qquad n \to\infty, \tag{11.18}$$

where $G(x)$ is some non-degenerate d.f. One sees that

$$P\{(M(n)-a_n)/b_n \le x \}-G(x)= (P\{(M(n)-\log n) \le xb_n+(a_n-\log n)\}-H_0(xb_n+(a_n-\log n)))+$$

$$(G(x)- H_0(xb_n+(a_n-\log n))) = I_1(x)+I_2(x) \to 0. \tag{11.19}$$

Due to (11.17), $I_1(x) \to 0$, and hence (11.19) implies that

$$H_0(xb_n+(a_n-\log n))) \to G(x). \tag{11.20}$$

In particular,

$$H_0(a_n-\log n)) \to G(0),$$

as $n \to\infty$. It means that a sequence $(a_n-\log n)$ has a limit, say, *a*.

Suppose that G(0)=0. Then a = - ∞. Since G is a non-degenerate d.f., there is a finite positive c, such that 0<G(c)<1. It means that

$$cb_n+(a_n-\log n)) \to \gamma=H_0^{-1}\ (G(c)), \qquad (11.21)$$

where

$$H_0^{-1}(x)= -\log(-\log x)$$

is the inverse function of H_0. Relation (11.21) implies that b(n) tends to infinity, as n→∞, and

$$xb_n+(a_n-\log n))= (x-c)b_n +(cb_n+a_n-\log n) \to\infty$$

for any x>c. In this case G(x)=1 for x>c and hence G(c)=1. This contradicts to our statement that G(c)<1. Hence G(0)>0. Analogously we can prove that G(0)<1. Thus,

$$\lim_{n\to\infty} (a_n-\log n))= a\ , -\infty< a <\infty.$$

Now we get from (11.21) that the sequence b(n) also has the finite limit b=(γ- a)/c.

Taking into account relation (11.20) and increase of G(x), we obtain that b>0. It is evident also that

$$G(x)= \lim_{n\to\infty} H_0(xb(n)+(a_n-\log n)))=H_0(a+xb).$$

Exercise 11.4 (solution). In this case, according to example 11.1, M(n) itself has the degenerate distribution, concentrated at zero, while normalized maxima has non-degenerate distribution. In fact, we have

$$F(x)=1+x,\ -1<x<0,$$

and for any x≤0,

$$P\{nM(n)\le x\}=(F(x/n))^n=(1+x/n)^n\to\exp(x),$$

as n→∞. Indeed, nM(n) ≤0 and hence, P{nM(n)≤x} =1 for any n and x>0.

Exercise 11.5 (solution). In this situation we simply generalize the case given in exercise 11.4:

$$P\{\ n^{1/\alpha}M(n)\le x\}=(F(x/\ n^{1/\alpha}))^n=(1-(-x)^\alpha/n)^n\to\exp((-x)^\alpha),\ -\infty<x\le 0.$$

Exercise 11.6 (solution). We have

$$P\{\ n^{-1/\alpha}M(n)\le x\} = (F(xn^{1/\alpha}))^n = (1-x^{-\alpha}/n)^n\to\exp(-x^{-\alpha}),\ x\ge 0.$$

Exercise 11.7 (hint and answer). Use relations

$$(1-F(x))\sim \frac{1}{\pi x}, \quad F'(x) \sim \frac{1}{\pi x^2}, \; x\to\infty,$$

and the statement of theorem 11.1 to show that it is possible to take $a_n=0$ and $b_n=n$.

Exercise 11.8 (hint). Show that

$$1-F(x)\sim F'(x)/x=\frac{1}{x\sqrt{2\pi}}\exp(-x^2/2),$$

$$F''(x)\sim -xF'(x), \; x\to\infty$$

and then use the statement of theorem 11.3. To find constants a_n you need to solve the equation

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

or simply to check that the sequence

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

satisfies this equation. Theorem 11.3 recommends to take $b_n = h(a_n)$, where

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

but we can see that

$$h(x)\sim 1/x, \; x\to\infty,$$

and hence

$$h(a_n)\sim 1/a_n, \; n\to\infty.$$

From lemma 11.2 we know that the sequence of constants b_n can be changed by a sequence of equivalent constants and this changing saves the limit distribution for maximal values. Moreover, going further we can take the following more simple sequence of equivalent norming constants, namely

$$b_n= (2\log n)^{1/2}.$$

Exercise 11.9 (answer). In this case $F\in D(H_0)$, $a_n= \log n$ and $b_n=1$.

Exercise 11.10 (hint and answer). Use relation (11.2) and theorem 11.5 to find that

$$P\{Y_n<x\}\to \exp(-e^{-x})\sum_{j=0}^{k-1} e^{-jx}\exp(-e^{-x})/j!, \text{ as } n\to\infty.$$

Учитывая важность результатов и разнообразие методов для экстремальных порядковых статистик, приводим здесь ряд соответствующих ссылок, которые могут помочь подробнее разобраться в теории экстремумов.

1). Ahsanullah, M. and Kirmani, S.N.U.A. (2008). Topics in extreme values. New York: Nova Science Publishers Inc.

2). Ahsanullah, M. and Nevzorov, V.B. (2001). Ordered Random Variables. New York: Nova Science Publishers Inc.

3). Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (1993). A first course in order statistics, Wiley, New York.

4). Balakrishnan, N. and Nevzorov, V.B. (2003). A primer on statistical distributions, Wiley, New York.

5). Castillo, E. (1988). Extreme value theory in engineering, Academic, Boston.

6). David, F.N., Nagaraja,H.N (2003). Order statistics , (3rd ed.), Wiley, New York

7). Galambos, J. (1987). The Asymptotic Theory of Extreme Order Statistics. Malabar, Florida: Robert E. Krieger Publishing Co.

8). Gnedenko, B. (1943). Sur la distribution limite du terme maximum d'une serie aletoise. Ann. Math., 44, 423-453.

9). Gumbel, E. .J. (1958). Statistics of Extremes. New York: Columbia Univ. Press.

10). Lindgren, G. (1971). Extreme values of stationary normal processes. Z. Wahrsch. verw.Geb., 17, 39-47.

11). Mardia, K. V. (1964). Asymptotic independence of bivariate extremes. Calcutta Stat. Assoc. Bull., 13, 172-178.

12). Mejzler, D.G . (1978). Limit distributions for the extreme order statistics. Canad. Math.Bull., 21, 447-459.

13). Mises, R. von (1936). La distribution de la plus grande de n valeurs. Revue Mathematique de l’Union Interbalkanique (Athens), 1, 141-160.

14). Renyi, A. (1962). On outstanding values of a sequence of observations. In: Selected paper of A Renyi, 3, 50-65. Budapest: Akademiai Kiado.

15). Sethuramann, J. (1965). On a characterization of the three limiting types of extremes.

Sankhya, A 27, 357-364.

REFERENCES
Библиография

Имеется множество работ, в которых главными действующими лицами являются порядковые статистики. Мы приводим здесь лишь ряд монографий, в которых они подробно изучаются. В этих книгах также можно найти множество ссылок на статьи, связанные с порядковыми статистиками.

1) M. Ahsanullah, V. Nevzorov (2001). Ordered random variables. Nova Science Publishers, NY ,412 p.

2) M. Ahsanullah, V. Nevzorov (2005). Order statistics. Examples and exercises. Nova Science Publishers, NY, 236 p.

3) M.Ahsanullah, V.B. Nevzorov, M.Shakil (2013). An introduction to order statistics. Atlantis Press, Amsterdam, 244 p.

4) B.C.Arnold, N.Balakrishnan, H.N. Nagaraja (1993). A first course in order statistics. John Wiley & Sons, New York, 279 p.

5) H.A.David, H.N.Nagaraja (2003). Order statistics (third edition). John Wiley & Sons, 458 p.

6) В. Б. Невзоров (2000).Рекорды. Математическая теория. Фазис, Москва, 244 стр.