Chapter 2

Distributions of Order Statistics

We give some important formulae for distributions of order statistics. For example,

$$F_{k:n}(x) = P\{X_{k,n} \leqslant x\} = I_{F(x)}\ (k,\ n-k+1),$$

where

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

denotes the incomplete Beta function. The probability density function of $X_{k,n}$ is given as follows:

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

where f is a population density function. The joint density function of order statistics $X_{1,n}, X_{2,n}, \dots, X_{n,n}$ has the form

$$f_{1,2,\dots,n:n}(x_1,x_2,\dots,x_n) = \begin{cases} n! \prod\limits_{k=1}^{n} f(x_k), & -\infty < x_1 < x_2 < \cdots < x_n < \infty, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

There are some simple formulae for distribution of maxima and minima.

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

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

Then d.f. of $M(n) = \max\{X_1, X_2, \dots, X_n\}$ has the form

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

Similarly we can get the distribution of $m(n) = \min\{X_1, X_2, \dots, X_n\}$.
One has

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

M. Ahsanullah et al., *An Introduction to Order Statistics*,
Atlantis Studies in Probability and Statistics 3,
DOI: 10.2991/978-94-91216-83-1_2, © Atlantis Press 2013

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.

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

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 $X_1, X_2, \ldots, X_n$ present a sample from a population d.f. F (not necessary continuous):

$$P\{M(n) \leqslant x\} = F^n(x) \tag{2.3}$$

and

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

Exercise 2.4. Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a geometrically distributed random variable X, such that $P\{X = m\} = (1-p)p^m$, $m = 0, 1, 2, \ldots$.
Find

$$P\{Y \geqslant r,\ Z < s\}, \quad r < s,$$

where $Y = \min\{X_1, X_2, \ldots, X_n\}$ and $Z = \max\{X_1, X_2, \ldots, X_n\}$.
There is no difficulty to obtain d.f.'s for single order statistics $X_{k,n}$. Let

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

We 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,

$$\begin{aligned} F_{k:n}(x) &= P\{\text{ at least } k \text{ variables among } X_1, X_2, \ldots, X_n \text{ are less or equal } x\} \\ &= \sum_{m=k}^{n} P\{\text{ exactly } m \text{ variables among } X_1, X_2, \ldots, X_n \text{ are less or equal } x\} \\ &= \sum_{m=k}^{n} \binom{n}{m} (F(x))^m (1 - F(x))^{n-m}, \quad 1 \leqslant k \leqslant n. \end{aligned} \tag{2.5}$$

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 $0 \leqslant x \leqslant 1$, 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 we have tables 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 $X_{1,n}, X_{2,n}, \ldots, X_{n,n}$.

It seems that the joint d.f.

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

promises to be very complicated. Hence, we consider probabilities

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

for any values $-\infty \leqslant y_1 < x_1 \leqslant y_2 < x_2 \leqslant \cdots \leqslant y_n < x_n \leqslant \infty$.

It is evident, that the event

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

is a union of $n!$ disjoint events

$A(\alpha(1),\alpha(2),\ldots,\alpha(n)) = \{y_1 < X_{\alpha(1)} \leqslant x_1, y_2 < X_{\alpha(2)} \leqslant x_2, \ldots, y_n < X_{\alpha(n)} \leqslant x_n\}$, where the vector $(\alpha(1),\alpha(2),\ldots,\alpha(n))$ runs all permutations of numbers $1,2,\ldots,n$. 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

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

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

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

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

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

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

$$\begin{aligned} f_{m:n}(x) &= \int\cdots\int f_{1,2,\dots,n:n}(x_1,\dots,x_{k-1},x,x_{k+1},\dots,x_n)dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_n \\ &= n!f(x)\int\cdots\int\prod_{k=1}^{m-1} f(x_k)\prod_{k=m+1}^{n} f(x_k)dx_1\cdots dx_{m-1}dx_{m+1}\cdots dx_n, \end{aligned} \tag{2.11}$$

where the integration is over the domain

$$-\infty < x_1 < \cdots < x_{m-1} < x < x_{m+1} < \cdots < x_n < \infty.$$

The symmetry of

$$\prod_{k=1}^{m-1} f(x_k)$$

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

$$\prod_{k=m+1}^{n} f(x_k)$$

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

$$\begin{aligned} &\int\cdots\int\prod_{k=1}^{m-1} f(x_k)\prod_{k=m+1}^{n} f(x_k)dx_1\cdots dx_{m-1}dx_{m+1}\cdots dx_n \\ &\quad = \frac{1}{(m-1)!}\prod_{k=1}^{m-1}\int_{-\infty}^{x} f(x_k)dx_k \frac{1}{(n-m)!}\prod_{k=m+1}^{n}\int_{x}^{\infty} f(x_k)dx_k \\ &\quad = \frac{(F(x))^{m-1}(1-F(x))^{n-m}}{(m-1)!(n-m)!}. \end{aligned} \tag{2.12}$$

Combining (2.11) and (2.12), we get 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.

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

of order statistics $X_{k(1),n}, X_{k(2),n}, \ldots, X_{k(r),n}$, where $1 \leqslant k(1) < k(2) < \cdots < k(r) \leqslant n$.

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

$$\begin{aligned} f_{i,j:n}(x_1, x_2) &= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \\ &\times (F(x_1))^{i-1}(F(x_2)-F(x_1))^{j-i-1}(1-F(x_2))^{n-j}f(x_1)f(x_2), \end{aligned} \tag{2.14}$$

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

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

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

One has

$$F_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \iiint_D \prod_{k=1}^{n} f(u_k)\,du_1 \cdots du_n, \tag{2.15}$$

where

$$D = \{U_1, \ldots, U_n : \ U_1 < U_2 < \cdots < U_n;\ U_1 < x_1,\ U_2 < x_2, \ldots,\ U_n < x_n\}.$$

Note that (2.15) is equivalent to the expression

$$F_{1,2,\ldots,n:n}(x_1, x_2, \ldots, x_n) = n! \iiint_{\widehat{\mathrm{D}}} du_1 \cdots du_n, \tag{2.16}$$

where integration is over

$$\widehat{\mathrm{D}} = \{U_1, \ldots, U_n : \ U_1 < U_2 < \cdots < U_n;\ U_1 < F(x_1),\ U_2 < F(x_2), \ldots,\ U_n < F(x_n)\}.$$

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 $x \geqslant y$, then

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

Otherwise,

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

Exercise 2.2 (answer).

$$P\{Y \leqslant x\} = 1 - H(x,\infty) - H(\infty,x) + H(x,x),$$

where

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

and

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

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

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

If $x < y$, then

$$P\{Y \leqslant x,\ Z \leqslant y\} = P\{X_1 \leqslant x,\ X_2 \leqslant y\} + P\{X_1 \leqslant y,\ X_2 \leqslant x\} - P\{X_1 \leqslant x,\ X_2 \leqslant x\}$$

$$= H(x,y) + H(y,x) - H(x,x).$$

Exercise 2.4 (solution). We see that

$$P\{Y \geqslant r,\ Z < s\} = P\{r \leqslant X_k < s,\ k = 1,2,\dots,n\}$$

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

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

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

because $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ and the gamma function satisfies equality $\Gamma(k) = (k-1)!$ for $k = 1,2,\dots$.

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

$$\frac{n!\,x^{k-1}(1-x)^{n-k}}{(k-1)!(n-k)!}.$$

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

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

If $x_2\leqslant x_1$, then evidently

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

and

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

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 $X_1,X_2,\dots,X_n$ (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 $A=\{X_{r,n}\leqslant x_1,\ X_{s,n}\leqslant x_2\}$ is a union of some disjoint events

$$\begin{aligned}a_{i,j,n-i-j}=\{\,&i \text{ elements of the sample fall into } (-\infty,x_1],\\ &j \text{ elements fall into interval } (x_1,x_2], \text{ and}\\ &(n-i-j) \text{ elements lie to the right of } x_2\,\}.\end{aligned}$$

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 $r\leqslant i\leqslant n$, $j\geqslant 0$ and $s\leqslant i+j\leqslant n$. Hence,

$$\begin{aligned}F_{r,s:n}(x_1,x_2)=P\{A\}&=\sum_{i=r}^{n}\sum_{j=\max\{0,s-i\}}^{n-i}P\{A_{i,j,n-i-j}\}\\ &=\sum_{i=r}^{n}\sum_{j=\max\{0,s-i\}}^{n-i}\frac{n!}{i!j!(n-i-j)!}(F(x_1))^i(F(x_2)-F(x_1))^j(1-F(x_2))^{n-i-j}.\end{aligned}$$

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

$$f_{k(1),k(2),\dots,k(r):n}(x_1,x_2,\dots,x_r)=$$
$$\begin{cases}\dfrac{n!}{\prod_{m=1}^{r+1}(k(m)-k(m-1)-1)!} & \prod_{m=1}^{r+1}(F(x_m)-F(x_{m-1}))^{k(m)-k(m-1)-1}\prod_{m=1}^{r}f(x_m),\\ & \text{if } x_1<x_2<\dots<x_r, \text{ and}\\ 0, & \text{otherwise.}\end{cases}$$

In particular, if $r=2$, $1\leqslant i<j\leqslant n$, and $x_1<x_2$, then

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

http://www.springer.com/978-94-91216-82-4

An Introduction to Order Statistics
Ahsanullah, M.; Nevzorov, V.B.; Shakil, M.
2013, X, 244 p., Hardcover
ISBN: 978-94-91216-82-4
A product of Atlantis Press