7 Order Statistics

- 7.1 Motivation
- 7.2 Multinomial distribution ✓
- 7.3 Distributions related to order statistics ✓

7.1 Motivation

Let $\mathbf{X} = (X_1, \ldots, X_n)$ denote a random sample from a population with a continuous distribution function F_X. Since F_X is assumed to be *continuous*, the probability of any two of these random variables assuming the same value is zero.

After reordering the n values we get

$$X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$$

in which, as mentioned, the $\leq$ sign could also be replaced by $<$.

These values are collectively termed the order statistic of the random sample $\mathbf{X} = (X_1, \ldots, X_n)$. The subject of order statistics generally deals with properties of $X_{(r)}$, $r = 1, 2, \ldots, n$ which is called the r-th order statistic.

For example, $X_{(1)} = \min(X_1, \ldots, X_n)$ is the smallest order statistic and corresponds to the sample minimum, and $X_{(n)} = \max(X_1, \ldots, X_n)$ is the largest order statistic and corresponds to the sample maximum.

Order statistics are particularly useful in nonparametric statistics because of the following:

Theorem 7.27 (Probability-integral transformation)

If the random variable X has a continuous cdf F_X then the random variable $Y = F_X(X)$ has the uniform probability distribution over the interval (0,1).

Further, given a sample $\mathbf{X} = (X_1, \ldots, X_n)$ of n i.i.d. random variables with cdf F_X, the transformation $U_{(r)} = F_X(X_{(r)})$ produces a random variable $U_{(r)}$ which is the r-th order statistic from the uniform population in (0,1), *regardless* of what F_X is, i.e. $U_{(r)}$ is distribution-free.

Proof: at lecture.

Remark 7.19

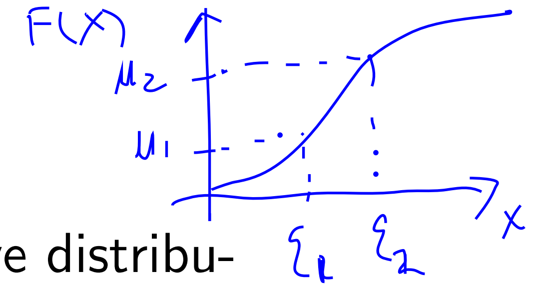
The above theorem has also an extremely important practical application in the generation (computer simulation) of observations from any specific continuous distribution function. There are several well-developed uniform random number generators that implement methods to generate sequences of uniform in $(0,1)$ pseudo-random numbers.

These numbers are pseudo since in fact they are generated by a deterministic algorithm (therefore are not random) but look as random (hence the word pseudo-random) in the sense that they pass usual statistical tests about randomness of the generated sequence.

Every program system (Fortran, SPLUS, C, SAS, etc.) has such uniform random number generators and we will not discuss their specific implementation here. What we would like to discuss is how we could use these uniform random number generators to generate random numbers with arbitrary continuous cumulative distribution function F_X. The answer is:

1) Generate Y as uniformly distributed in $(0,1)$ using the uniform random number generator runif(1)

2) Calculate $\xi = F_X^{-1}(Y)$.



Then ξ is distributed according to $F_X(.)$ since its cumulative distribution function is:

$$P(\xi < x) = P(F_X^{-1}(Y) < x) = P(F_X^{-1}(F_X(X)) < x) = P(X < x) = F_X(x).$$

Some further important applications of order statistics are listed below:

- $X_{(n)}$ is of interest in studying floods, earthquakes and other extreme phenomena, sports records, financial markets etc.
- $X_{(1)}$ is useful, for example, in estimating strength of a chain that would depend on the weakest link.
- the sample median defined as $X_{[(n+1)/2]}$ for n odd and any number between $X_{(n/2)}$ and $X_{(n/2+1)}$ for n even, is a measure of location and an estimate of the population central tendency.
- the sample midrange $(X_{(n)} + X_{(1)})/2$ is also a measure of central tendency, whereas the sample range $X_{(n)} - X_{(1)}$ is a measure of dispersion.

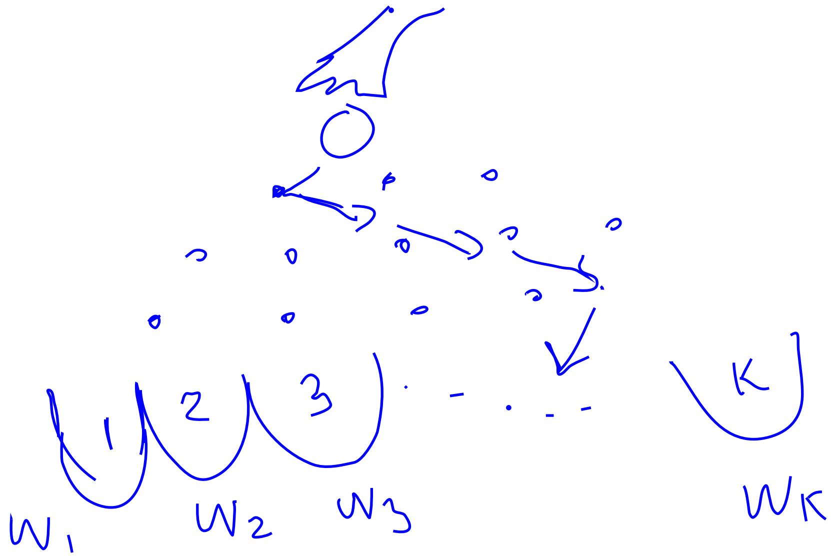
7.2 Multinomial distribution

The multinomial distribution is a discrete distribution that includes the Binomial distribution as a particular case and, as such, is interesting in its own right. However, for us, it is important that it plays a crucial role in the methodology of order statistics and, for this reason, we consider it here first.

$X = (X_1, X_2, \ldots, X_k)$ n trials

X_i

$\sum_{i=1}^{k} X_i = n$ $\sum_{i=1}^{k} w_i = 1$



Suppose a single trial can result in k $(k \geq 2)$ possible outcomes numbered $1, 2, \ldots, k$ and let $w_i = P(\text{a single trial results in outcome } i)$ and $\sum_{i=1}^{k} w_i = 1$.

For n independent trials, let X_i denote the number of trials resulting in outcome i (then $\sum_{i=1}^{k} X_i = n$). Then we say that the distribution of

$$(X_1, X_2, \ldots, X_k) \sim \text{Multinomial}(n; w_1, w_2, \ldots, w_k)$$

and it holds

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1! x_2! \ldots x_k!} w_1^{x_1} w_2^{x_2} \ldots w_k^{x_k},$$

where $0 < w_i < 1$ and $\sum_{i=1}^{k} w_i = 1$.

$n=100$

Example 7.58

Consider rolling a die 100 times and registering the number $X_1, X_2, \ldots, X_6$ of outcomes "uppermost 1", "uppermost 2",...,"uppermost 6" during the 100 tosses.

$K=6$

$w_i = \frac{1}{6}$

$i=1,\ldots,6$

Solution:

If the die is fair, then we have $k = 6$ and the vector

$$(X_1, X_2, \ldots, X_6) \sim \text{Multinomial}(100; \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$$

Remark 7.20

Note that when $k = 2$ this is just the familiar Binomial distribution denoted as $\text{Bin}(n, w_1)$ (we do not need w_2 since it is known that $w_2 = 1 - w_1$ for the binomial. We also that

$$\mathbb{E}(X_i) = nw_i; \quad \text{Var}(X_i) = nw_i(1 - w_i), \quad i = 1, 2, \ldots, k$$

holds by noting that the marginal distribution of the i-th component of the multinomial is in fact binomial $(\text{Bin}(n; w_i))$.

In a slightly more complicated way, one can also show that $\text{Cov}(X_i, X_j) = -nw_iw_j$ holds for $j \neq i$, that is, the k components of the multinomial vector are negatively correlated.

$n=7$ $w_i = \frac{1}{6}$

Example 7.59

A fair die is rolled 7 times. Find the probability that all the 6 possible outcomes will happen at least once.

anything 1,2,3,4,5,6

| 1 | 2 | 3 | 4 | 5 | 6 | ? |

Solution:

Since we have six symmetric cases in which $X_i = 2$, $i = 1, 2, \dots, 6$ and $X_j = 1$ for $i \neq j$ we can include the multiplication by six in our calculation as follows:

$$6 \times P(X_1 = 2, X_2 = 1, X_3 = 1, X_4 = 1, X_5 = 1, X_6 = 1)$$
$$= 6 \times \frac{7!}{2!1!1!1!1!1!1!} \left(\frac{1}{6}\right)^7$$
$$= \frac{35}{648}$$

Example 7.60

The vector X_1, X_2, X_3 has a multinomial $(8; 0.2, 0.3, 0.5)$ distribution.

i) Find $P(X_1 = 2, X_2 = 2, X_3 = 4)$.

ii) Find $\mathbb{E}(X_2)$.

iii) Find $\mathrm{Var}(X_2)$.

iv) Find $\mathrm{Cov}(X_1, X_3)$.

Since $(X_1, X_2, X_3) \sim \text{Multinomial}(8, 0.2, 0.3, 0.5)$ we have

(i)

$$P(X_1 = 2, X_2 = 2, X_3 = 4) = \frac{8!}{2!2!4!} 0.2^2\, 0.3^2\, 0.5^4 = 0.0945$$

(ii) The marginal distributions are Binomial which implies that $X_2 \sim \text{Bin}(8, 0.3)$ and therefore

$$\mathbb{E}(X_2) = nw_2 = 8 \times 0.3 = 2.4$$

(iii)

$$\text{Var}(X_2) = nw_2(1 - w_2) = 8 \times 0.3 \times 0.7 = 1.68$$

(iv)

$$\text{Cov}(X_2, X_3) = -nw_1 w_2 = -8 \times 0.2 \times 0.5 = -0.8$$

Example 7.61

The vector X_1, X_2, X_3 has a multinomial $(6; 0.5, 0.2, 0.3)$ distribution.

i) Find $P(X_1 = 3, X_2 = 1, X_3 = 2)$.

ii) Find $P(X_1 + X_2 = 2)$. $= P(X_3 = 4)$ $\quad X_3 \sim Bin(6, 0.3)$

$X_1 + X_2 + X_3 = 6$

$X_3 = 6 - (X_1 + X_2)$

$= 6 - 2$

$= 4$

Since $(X_1, X_2, X_3) \sim$ Multinomial$(6, 0.5, 0.2, 0.3)$ we have

(i)

$$P(X_1 = 3, X_2 = 1, X_3 = 2) = \frac{6!}{3!1!2!} 0.5^3 \, 0.2^1 \, 0.3^2 = 0.135$$

(ii) First notice that $P(X_1 + X_2 = 2) = P(X_3 = 4)$ and the marginal distribution is Binomial which implies that $X_3 \sim$ Bin$(6, 0.3)$ and therefore

$$P(X_1 + X_2 = 2) = \frac{6!}{2!4!} 0.3^2 \, 0.7^4 = 0.059525$$

7.3 Distributions related to order statistics

Let X be a random variable with a density $f_X(x)$ and a cumulative distribution function $F_X(x)$ and let there be n independent copies $X_1, X_2, \ldots, X_n$ of X.

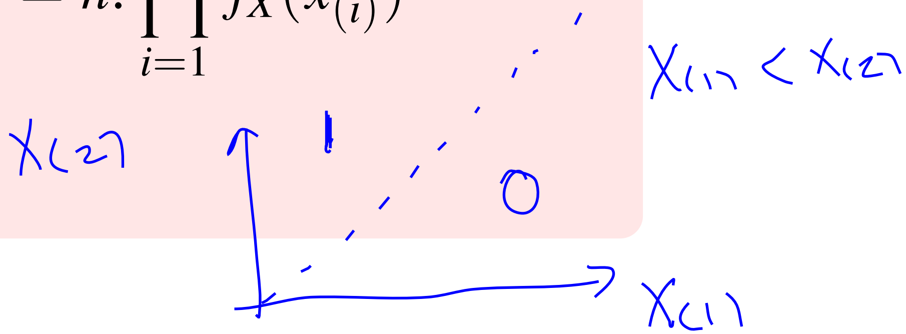
Theorem 7.28

The joint density $f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is given by:

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)}) = n! \prod_{i=1}^{n} f_X(x_{(i)})$$

for $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$.

$n = 2$



Proof

We consider a trial consisting of $k = 2n+1$ possible outcomes that is repeated independently n times. Each of the outcomes is a realization in one of the $2n+1$ intervals:

$$(-\infty, x_{(1)}), [x_{(1)}, x_{(1)} + \Delta x_{(1)}), [x_{(1)} + \Delta x_{(1)}, x_{(2)}), \dots,$$
$$[x_{(n-1)} + \Delta x_{(n-1)}, x_{(n)}), [x_{(n)}, x_{(n)} + \Delta x_{(n)}), [x_{(n)} + \Delta x_{(n)}, \infty)$$

where $\Delta x_{(i)}, i = 1, 2, \dots, n$ are chosen sufficiently small so that no overlap of the intervals occurs. The trial's outcome can be interpreted as a realization of multinomial distribution:

$$\text{Multin}(n; F_X(x_{(1)}), F_X(x_{(1)} + \Delta x_{(1)}) - F_X(x_{(1)}),$$
$$F_X(x_{(2)}) - F_X(x_{(1)} + \Delta x_{(1)}), \dots, (1 - F_X(x_{(n)} + \Delta x_{(n)}))$$

We are looking at the probability for one very particular outcome of the trial, namely realization in the second, fourth, $2n$th interval for this multinomial distribution.

On one hand, this probability is

$$\frac{n!}{0!1!0!1!\ldots 0!1!0!}[F_X(x_{(1)}+\Delta x_{(1)})-F_X(x_{(1)})]\ldots$$
$$[F_X(x_{(n)}+\Delta x_{(n)})-F_X(x_{(n)})].$$

On the other hand, it is just

$$P(x_{(i)} \leq X_{(i)} < x_{(i)}+\Delta x_{(i)}, i=1,2,\ldots,n).$$

Having in mind the definition

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)},x_{(2)},\ldots,x_{(n)}) = \lim_{\Delta x_{(i)}\to 0} \frac{P(x_{(i)} \le X_{(i)} < x_{(i)} + \Delta x_{(i)}, i = 1,2,\ldots,n)}{\prod_{i=1}^{n} \Delta x_{(i)}}$$

we get

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)},x_{(2)},\ldots,x_{(n)}) = n! \prod_{i=1}^{n} f_X(x_{(i)}),$$

for $x_{(1)} < x_{(2)} < \cdots < x_{(n)}$. □

Theorem 7.29

It holds:

$$f_{X_{(n)}}(y_n) = n[F_X(y_n)]^{n-1} f_X(y_n)$$

$$f_{X_{(1)}}(y_1) = n[1 - F_X(y_1)]^{n-1} f_X(y_1)$$

$$f_{X_{(r)}}(y_r) = \frac{n!}{(r-1)!(n-r)!}[F_X(y_r)]^{r-1}[1 - F_X(y_r)]^{n-r} f_X(y_r)$$

Note that the density of the maximum and of the minimum has already been used by us earlier. The result for $f_{X_{(r)}}(y_r)$ is the general result from which the maximum and the minimum can be obtained for $r = n$ and $r = 1$, respectively.

Since the previous theorem gives us the joint distribution, the marginal distributions formulated in Theorem 7.29 can be obtained through integration. This method is straightforward but the integration is very tiresome. A much simpler method can be used which appeals to probability theory instead of to pure mathematical integration.

Proof.

If we introduce the discrete random variable $Y =$ number of realizations $X_1, X_2, \ldots, X_n$ that happen to be $\leq x$. We see immediately that $Y \sim Bin(x, F_X(x))$. We then derive the cdf $F_{X_{(r)}}(x)$ by noticing that

$$F_{X_{(r)}}(x) = P(X_{(r)} \leq x) = P(Y \geq r)$$

and substituting the latter probability from the binomial distribution formula.

Finally, we differentiate $F_{X_{(r)}}(x)$ to get the density $f_{X_{(r)}}(x)$. To accomplish this, we need to apply the rule for a differentiation of a product of two functions several times. A lot of cancellations happen during the differentiation and we end up with

$$f_{X_{(r)}}(y_r) = \frac{n!}{(r-1)!(n-r)!}[F_X(y_r)]^{r-1}[1 - F_X(y_r)]^{n-r} f_X(y_r). \quad \square$$

$$f_X(x) = 1$$

$$F_X(x) = x$$

Example 7.62

Derive the density of the rth order statistic from a sample of size n from the uniform distribution. $[0,1]$

$$f_{X(r)}(y) = \frac{n!}{(r-1)!(n-r)!} y^{r-1} (1-y)^{n-r} \cdot 1$$

$$= \frac{1}{B(r, n-r+1)} y^{r-1} (1-y)^{n-r}$$

$$X_{(r)} \sim \text{Beta}(r, n-r+1)$$

Solution:

From Theorem 7.29 we realise that, for the particular case of F_X being uniform distribution on $(0, 1)$, we get

$$f_{X_{(r)}}(y) = \frac{n!}{(r-1)!(n-r)!} y^{r-1}(1-y)^{n-r}, \quad y \in (0, 1)$$

(and zero elsewhere), which is the density of the Beta distribution with parameters r and $n - r + 1$.

In particular, using the properties of the beta distribution, we can now show that for the rth order statistic $X_{(r)}$ of the beta distribution we have

$$\mathbb{E}(X_{(r)}) = \frac{r}{n+1} \qquad \text{and} \qquad \mathrm{Var}(X_{(r)}) = \frac{r(n+1-r)}{(n+1)^2(n+2)}.$$

Exercise 7.36 (at lecture)

Electrical components of a certain type have lifetime Y with probability density given by

$$f(y) = \begin{cases} \frac{1}{100} e^{-\frac{y}{100}}, & y > 0, \\ 0 & \text{elsewhere} \end{cases}$$

i) Suppose that two such components operate independently and in series in a certain system (that is, the system fails when either component fails). Find the density function for X, the lifetime of the system.

$Y_{(1)}$

ii) Now suppose that the components operate in parallel (that is, the system does not fail until both components fail). Find the density function for X, the lifetime of the system.

$Y_{(2)}$

Joint densities of couples $(X_{(i)}, X_{(j)})$ can also be derived through integration of the joint but it is again easier to use some probabilistic arguments instead.

Theorem 7.30

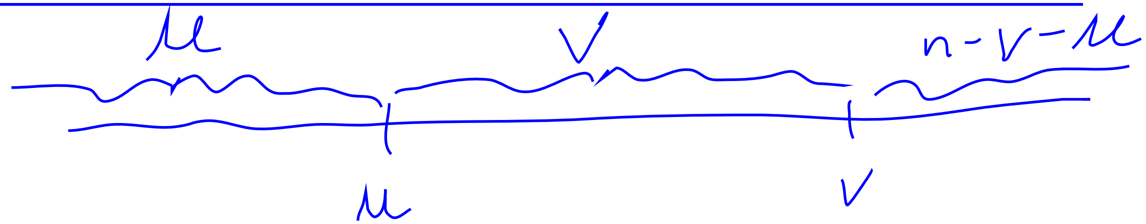
It holds for $1 \le i < j \le n$:

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{n!}{(i-1)!(j-1-i)!(n-j)!} \times f_X(u) f_X(v) [F_X(u)]^{i-1} [F_X(v) - F_X(u)]^{j-1-i} [1 - F_X(v)]^{n-j}$$

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

Proof (sketch)

(skip)



We first obtain the cdf $F_{X_{(i)},X_{(j)}}(u,v)$ first and then find its mixed partial derivative to calculate the density:

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v)$$

To obtain $F_{X_{(i)},X_{(j)}}(u,v)$, we introduce two discrete random variables. Let U count the number of observations that are less than or equal to u and let V count the number of observations greater than u and less than or equal to v. Then

$$(U,V,n-U-V) \sim \text{Multinomial}(n; F_X(u), F_X(v) - F_X(u), 1 - F_X(v))$$

Then, we observe that

$$F_{X_{(i)},X_{(j)}}(u,v) = P(U \geq i, U+V \geq j)$$
$$= \sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} P(U=k, V=m) + P(U \geq j)$$

Looking at this expression, we realise that when calculating

$$\frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v),$$

the probability $P(U \geq j)$ is irrelevant since this term only depends on u and not on v.

Hence

$$f_{X_{(i)},X_{(j)}}(u,v) = \frac{\partial^2}{\partial u \partial v} \sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} \frac{n!}{k!m!(n-k-m)!} \times$$
$$[F_X(u)]^k [F_X(v) - F_X(u)]^m [1 - F_X((v)]^{n-k-m}$$

By carefully calculating the mixed partial derivative using the product rule for differentiation, again we will notice huge cancellations occur and this yields the formula given in the Theorem.

Of course, joint densities of three or more order statistics could also be derived using similar arguments like above but the calculations will be more exhausting

The previous derivations have very important applications. For instance, the density of the sample median or of the sample range can now easily be derived by using standard formulae for density of transformed random variables that we discussed in Chapter 1. Here is an illustration of the approach.

Example 7.63

Derive the density of the sample range $R = X_{(n)} - X_{(1)}$ of n i.i.d. uniform $[0, 1]$ random variables.

Solution:

The sample range is a transformation of the two-dimensional random vector $(X_{(1)}, X_{(n)})$. The density of this vector is known due to Theorem 7.30:

$$f_{X_{(1)},X_{(n)}}(x, y) = n(n-1)[F_X(y) - F_X(x)]^{n-2} f_X(x) f_X(y)$$

where $F_X(.)$ and $f_X(.)$ are the uniform in $[0, 1]$ cdf and the density, respectively with $f_X(x) = 1$ and $F_X(x) = x$ for $0 < x < 1$ which gives

$$f_{X_{(1)},X_{(n)}}(x, y) = n(n-1)(y - x)^{n-2}$$

Set $X = X_{(1)}$ and $Y = X_{(n)}$. Let us introduce the variable of interest by making the transformation

$f_{uv} \to f_u$

$\begin{pmatrix} X_{(1)} \\ X_{(n)} \end{pmatrix} \to \begin{pmatrix} u \\ v \end{pmatrix}$

$X_{(n)} - X_{(1)}$ $\qquad$ $X_{(n)}$

$$U = Y - X \quad \text{and} \quad V = Y.$$

The value of the Jacobian of this transformation is equal to -1 since:

$X = Y - u$

$$X = V - U \quad \text{and} \quad Y = V.$$

Hence

$$J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} -1 & 1 \\ 0 & 1 \end{vmatrix} = -1 \cdot 1 - 0 \cdot 1 = -1$$

Therefore the absolute value of the Jacobian is equal to one.

Using the density transformation formula we have:

$$
\begin{aligned}
f_{U,V}(u,v) &= f_{X_{(1)},X_{(n)}}(x(u,v),y(u,v))|J(u,v)| \\
&= f_{X_{(1)},X_{(n)}}(v-u,v)\cdot|-1| \\
&= n(n-1)(F_X(v)-F_X(v-u))^{n-2}f_X(v-u)f_X(v)\cdot 1 \\
&= n(n-1)(\cancel{v}-(\cancel{v}-u))^{n-2} \\
&= n(n-1)u^{n-2}
\end{aligned}
$$

The density of the range $U = X_{(n)} - X_{(1)}$ is obtained by integrating out the unwanted variable V from the joint density:

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\mathrm{d}v$$

We need to be careful when doing this: since

$$0 < X_{(1)} < X_{(n)} < 1$$

$0 < V - U$

$U < V$

holds for the original variables, we get

$$0 < \underline{V - U} < \underline{V < 1}$$

for the transformed variables. This implies:

$$0 < U < V < 1$$

For a fixed u we have $u < v < 1$.

Hence, for a fixed u, the value of v ranges in $(u, 1)$. Therefore:

$$
\begin{aligned}
f_U(u) &= \int_u^1 f_{(U,V)}(u,v)dv \\
&= \cancel{\int_u^1 n(n-1)(v - v + u)^{n-2} \times 1 \mathrm{d}v} \\
&= \int_u^1 n(n-1)u^{n-2}\mathrm{d}v \qquad = n(n-1)u^{n-2}(1-u)
\end{aligned}
$$

for $0 < u < 1$ (and zero else). v $\quad n(n-1)u^{n-2} \cdot v \Big]_u^1$

The "trick" in the above derivation was to introduce, on top of the transformation $u = y - x$ of interest, one more transformation $(v = y)$ so that we could transform the original vector $(X_{(1)}, X_{(n)})$ in a new random vector (U, V) whose first component is the statistic of interest.

This allows us, by using the density transformation formula for random vectors, to get the joint density of (U, V) first. Then we integrated out the unwanted variable V to obtain the marginal density of the main component of interest (the component U).

Such a trick is used very often when working with order statistics. Clearly, the variable V plays an intermediate role here and in its choice, we are mainly guided by the convenience of the calculations.

For example, we could have chosen $v = x$ instead as a component of our transformation. You are advised to go along similar lines by using this new transform instead, derive the marginal of U again and convince yourself that you get the same result for the density of the range.

Exercise 7.37 (at lecture)

A continuous random variable X has a standard exponential distribution

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0 & \text{elsewhere} \end{cases}$$

$n = 3$

For a random sample of size 3, let $X_{(1)}, X_{(2)}, X_{(3)}$ denote the ordered sample.

i) Find the joint distribution of $X_{(1)}$ and $X_{(3)}$.

ii) Obtain the distributions of $X_{(1)}$ and $X_{(3)}$.

iii) Evaluate $\mathbb{E}(X_{(1)})$.

iv) Evaluate $\mathbb{E}(X_{(3)})$.

v) Find the sampling distribution of the range $R = X_{(3)} - X_{(1)}$.