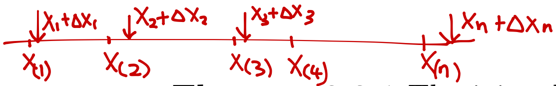
Lecture 12

2.3. 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.



$F(X_{(i)}+\Delta X_{(i)}) - F(X_{(i)})$

Theorem 2.3.1 The joint density $f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ is given by:

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

Indeed, 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:

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

$$[x_{(n)}, x_{(n)} + \Delta x_{(n)}), [x_{(n)} + \Delta x_{(n)}, \infty)$$

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

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

distribution. 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

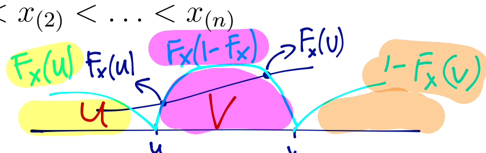
$$\frac{n!}{0!1!0!1!\ldots 0!1!0!}[F_X(x_{(1)} + \Delta x_{(1)}) - F_X(x_{(1)}] \ldots [F_X(x_{(n)} + \Delta x_{(n)}) - F_X(x_{(n)}].$$

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

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

we get

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



Theorem 2.3.2 It holds:

$V - \cancel{\emptyset} + \cancel{\emptyset} - \cancel{\emptyset} + \cancel{\emptyset} - \cancel{\emptyset} + \cancel{\emptyset} - U$

$f_{(X_{(i)}, X_{(j)})}(u,v)$

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

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

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

For $X_1, X_2, \ldots, X_n$

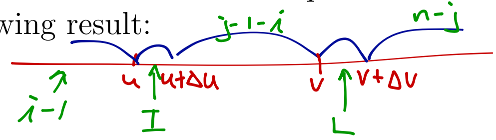
Proof: Since the previous theorem gives us the joint distribution, the marginal distributions formulated in Theorem 2.3.2 can be obtained through integration. This method is straightforward as an idea but the integration is tiresome. A much simpler method can be used which appeals to probability theory instead to pure mathematical integration. It will be illustrated at the lecture.

2.3.3. Note: From Theorem 2.3.2 we realize that for the particular case of the F_X being uniform distribution on (0,1), we get

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

which is the density of the *Beta distribution* with parameters r and $n-r+1$.

Joint densities of couples $(X_{(i)}, X_{(j)})$ can also be derived through integration of the joint density of all the n order statistics. But it is again easier to use some probabilistic arguments instead. The idea is illustrated in the following result:



Theorem 2.3.4 It holds for $1 \le i < j \le n$:

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

$f_{(X_{(i)}, X_{(j)})}(u,v)\,\Delta u \Delta v \approx \frac{n!}{(i-1)!(j-1-i)!(n-j)!}$

$f_X(u)\Delta u f_X(v)\Delta v\, F_X(u)^{i-1}(F(v)-F(u+\Delta u))^{j-1-i}(1-F(v+\Delta v))^{n-j}$

$\Delta u \Delta v$

for $-\infty < u < v < \infty$. Next Page!

$(\Delta u, \Delta v) \to 0$

Proof (idea): We obtain the cdf $F_{X_{(i)},X_{(j)}}(u,v)$ first and then find its partial derivative $f_{X_{(i)},X_{(j)}}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v)$. To obtain $F_{X_{(i)},X_{(j)}}(u,v)$, let U be a random variable that counts the number of $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n})$ that are less or equal to u and V be the random variable that counts the number of $\mathbf{X} = (\mathbf{X_1}, \mathbf{X_2}, \dots, \mathbf{X_n})$ greater than u and less 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, obviously

$X_n - X_{(1)}$

$\frac{X_n + X_{(1)}}{2}$

$f_{(X_{(1)}, X_{(n)})} \to u = X_{(n)} - X_{(1)}$

$V = X_{(n)}$

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

$f_{\binom{u}{v}}(u,v) = f_{\binom{X_{(1)}}{X_{(n)}}}(x_{(1)}, x_{(n)}) \left| \begin{matrix} \frac{\partial x_{(1)}}{\partial u} & \frac{\partial x_{(1)}}{\partial v} \\ \frac{\partial x_{(n)}}{\partial u} & \frac{\partial x_{(n)}}{\partial v} \end{matrix} \right|$

$$\sum_{k=i}^{j-1} \sum_{m=j-k}^{n-k} \frac{n!}{k!m!(n-k-m)!} [F_X(u)]^k [F_X(v) - F_X(u)]^m [1 - F_X((v)]^{n-k-m} + P(U \ge j)$$

Looking at this expression, we realize that when calculating $\frac{\partial^2}{\partial u \partial v} F_{X_{(i)},X_{(j)}}(u,v)$, the probability $P(U \ge j)$ is irrelevant since this term only depends on u and *not* on v. Carefully calculating the mixed partial derivative of the other term, 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 even more exhausting.

2.3.5 Note: The above 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

$f_{\binom{u}{v}}(u,v) = 1 * n(n-1)(F_x(v) - F_x(v-u))^{n-2} f_x(v-u) f_x(v)\, dv$

$U = Y - X$
$V = Y$

$f_R(u) = \int_{-\infty}^{\infty} n(n-1)(F_x(v) - F_x(v-u))^{n-2} f_x(v-u) f_x(v)\, dv$

$\begin{cases} y = v \\ x = v - u \end{cases}$

$\left| \left| \begin{matrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{matrix} \right| \right| = \left| \left| \begin{matrix} -1 & 1 \\ 0 & 1 \end{matrix} \right| \right|$

$F_{X_{(r)}}(x) = ?$

Introduce $Y =$ # of realisations from $(X_1, X_2, \dots, X_n)$ therefore $\leq x$

$$F_{X_{(r)}}(x) = P(Y \geq r) = P(X_{(r)} \leq x) = \sum_{k=r}^{n} P(Y = k) =$$

$$= \sum_{k=r}^{n} \binom{n}{k} (F_x(x))^k (1 - F_x(x))^{n-k}$$

$$\frac{\partial}{\partial x}\left[\sum_{k=r}^{n} \binom{n}{k} F_x(x)^k (1 - F_x(x))^{n-k}\right]$$

$$= \left[\sum_{k=r}^{n} \binom{n}{k} k f_x(x) F_x(x)^{k-1} (1 - F_x(x))^{n-k} - \binom{n}{k}(n-k) f_x(x) F_x(x)^k (1 - F_x(x))^{n-k-1}\right.$$

for $k = r$,

$$\binom{n}{r} r f_x(x) F_x(x)^{r-1} (1 - F_x(x))^{n-r} = \frac{n!}{(r-1)!(n-r)!} f_x(x) F_x(x)^{r-1} (1 - F_x(x))^{n-r}$$

$$\binom{n}{r} r = \frac{n!\, \cancel{r}}{\cancel{r!}(n-r)!}$$ $(r-1)!$

$\Rightarrow$ Beta $(r, n-r+1)$

$$E(X_{(r)}) = \frac{r}{n+1}$$

standard formulae for density of transformed random variables. In the case of the range $\boxed{R = X_{(n)} - X_{(1)}}$, for example, we know from Theorem 2.3.4 that

Spread of density

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

and if we make the transformation $u = y - x, v = y$ (with absolute value of Jacobian equal to one), we obtain for the range: for all $u > 0$:

$$f_R(u) = \int_{-\infty}^{\infty} n(n-1)[F_X(v) - F_X(v-u)]^{n-2} f_X(v-u) f_X(v) dv$$

For the uniform distribution, the integrand is non-zero for the intersection of the regions $0 < v - u < 1$ and $0 < v < 1$ which is just the region $0 < u < v < 1$. Therefore after integration, we get in this case $f_R(u) = n(n-1)u^{n-2}(1-u)$ for $0 < u < 1$.

$f_X(x) = 1$, $\quad F_X(x) = \begin{cases} x & 0<x<1 \\ 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases}$

$0 < x < y < 1$

$0 < v - u < v < 1$

$\boxed{0 < u < v < 1}$

$$f_R(u) = \int_u^1 n(n-1)(v-v+u)^{n-2} dv$$
$$= n(n-1) \int_u^1 u^{k-2} dv$$