Theory of Probability Nov 18, 2020

Order Statistics

Let $X_1, \dots, X_n$ be iid continuous random variables.

Let $X_{(1)}$ = smallest $X_1, \dots, X_n$

$X_{(2)}$ = next smallest,

$\vdots$

$X_{(n)}$ = largest of $X_1, \dots, X_n$.

$$\Rightarrow \quad \underbrace{X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}}_{\text{The order statistics of } X_1, \dots, X_n}$$

What is the pdf?

$X_{(1)} \le X_{(2)} \le \dots \le X_{(n)}$ take on the values $x_1 \le x_2 \le \dots \le x_n$ if and only if

$$X_1 = x_{i_1}$$
$$X_2 = x_{i_2}$$
$$\vdots$$
$$X_n = x_{i_n}$$

for some permutation $(i_1, i_2, \dots, i_n)$ of $(1, 2, \dots, n)$.

So in terms of $X_1, .., X_n$

$$P[x_{i_1} - \tfrac{\varepsilon}{2} \le X_1 \le x_{i_1} + \tfrac{\varepsilon}{2}, \ \ldots, \ x_{i_n} - \tfrac{\varepsilon}{2} \le X_n \le x_{i_n} + \tfrac{\varepsilon}{2}]$$

$$\approx \varepsilon^n f_{X_1 X_2 \cdots X_n}(x_{i_1}, \ldots, x_{i_n})$$

$$= \varepsilon^n f_{X_1}(x_{i_1}) \cdots f_{X_n}(x_{i_n})$$

Now since there are $n!$ permutations of $(1, 2, .., n)$, we have that

$$P[x_1 - \tfrac{\varepsilon}{2} \le X_{(1)} \le x_1 + \tfrac{\varepsilon}{2}, \ \ldots, \ x_n - \tfrac{\varepsilon}{2} \le X_{(n)} \le x_n + \tfrac{\varepsilon}{2}]$$

$\approx n! \, \varepsilon^n f(x_1) \cdots f(x_n)$. since it doesn't matter which $X_i = x_1$, etc.

$$\Rightarrow f_{X_{(1)} \cdots X_{(n)}}(x_1, .., x_n) = n! \, f(x_1) \cdots f(x_n) \quad \text{for}$$

$$x_1 \le x_2 \le \ldots \le x_n .$$

Joint distribution of functions of several random variables

Goal: Given joint pdf $f = f(x_1, x_2)$ for X_1, X_2, and if

$$Y_1 = g_1(X_1, X_2)$$

$$Y_2 = g_2(X_1, X_2)$$

what is the pdf of Y_1, Y_2?

We need two assumptions:

① The mapping $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longrightarrow \begin{pmatrix} g_1(x_1, x_2) \\ g_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$

is uniquely invertible, with

$$x_1 = h_1(y_1, y_2)$$
$$x_2 = h_2(y_1, y_2)$$

② g_1, g_2 are continuously differentiable,

and that

$$J(x_1, x_2) = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix}$$

$$= \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1}$$

$$\neq 0$$

Under these two assumptions we have that

$$f_{Y_1 Y_2}(y_1, y_2) = f_{X_1 X_2}(x_1, x_2) \frac{1}{|J(x_1, x_2)|}$$

$$= f_{X_1 X_2}\big(h_1(y_1, y_2), h_2(y_1, y_2)\big) \frac{1}{\big|J\big(h_1(y_1, y_2), h_2(y_1, y_2)\big)\big|}$$

Idea:

$$P[Y_1 \le y_1, Y_2 \le y_2]$$

$$= P[g_1(X_1, X_2) \le y_1, g_2(X_1, X_2) \le y_2]$$

$$= \iint_{\substack{g_1(x_1,x_2) \le y_1 \\ g_2(x_1,x_2) \le y_2}} f_{X_1 X_2}(x_1, x_2) \, dx_1 \, dx_2$$

Make the following change of variables:

$$x_1 = h_1(y_1, y_2)$$

$$x_2 = h_2(y_1, y_2)$$

$$\Rightarrow f \rightarrow f(h_1, h_2)$$

$$dx_1 dx_2 = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} dy_1 dy_2$$

$\Rightarrow$ Inserting back into the integral, we obtain that the integrand is:

$$f_{X_1 X_2}(h_1, h_2) \frac{1}{J(h_1, h_2)} = f_{Y_1 Y_2}(y_1, y_2).$$