ESE 520 Probability and Stochastic Processes

Lecture 7

"Functions of jointly continuous random vectors. Conditional densities."

Let $X = (X_1, ..., X_n) : \Omega \to \mathbb{R}^n$ be a random vector, $n \neq 1$.

Consider a *transformation* $\phi : \mathbb{R}^n \to \mathbb{R}^n$, where $\phi = (\phi_1, ..., \phi_n)$ and $\phi_i : \mathbb{R}^n \to \mathbb{R}$. In other words, ϕ is a *vector function* and all its components ϕ_i are scalar functions of n variables.

We assume that:

i) ϕ is a differentiable transformation;

ii) there is also a *differentiable inverse transformation* $\psi : \mathbb{R}^n \to \mathbb{R}^n$.

In other words,

$$\psi(\phi)(x_1, ..., x_n) = (x_1, ..., x_n)$$

and

$$\phi(\psi)(y_1, ..., y_n) = (y_1, ..., y_n).$$

Also, as it is known from Multidimensional Calculus, the differentiability of ψ means that the *Jacobi matrix* $D\psi(y)$ exists and is non-degenerate. The Jacobi matrix is defined as $n \times n$ matrix the i-th row of which is given by the vector

$$(\frac{\partial \psi_i(y_1, ..., y_n)}{\partial y_1} .... \frac{\partial \psi_i(y_1, ..., y_n)}{\partial y_n}).$$

We formally define a new vector $Y = (Y_1, ..., Y_n)$ as

$$Y := \phi(X).$$

Question: Given we know the joint pdf $f_X(x)$, what is the joint pdf $f_Y(y)$?

To find $f_Y(y)$, we first find the probability measure of Y:

$$P_Y(B) = P(Y = \phi(X) \in B) = P(\phi(X) \in B) =$$

[we use here the existence of inverse transformation]

$$P(X \in \psi(B)) = \int_{\psi(B)} f_X(x)dx =$$

[we make here change of variables $x = \psi(y)$ so that $dx = |\det D\psi(y)|dy$]

$$= \int_B f_X(\psi(y))|\det D\psi(y)|dy.$$

From the definition of a joint pdf, it follows that the function under the integral sign in the last relation must be the pdf we seek. Therefore,we obtain

$$f_Y(y) = f_X(\psi(y))|\det D\psi(y)|. \tag{1}$$

Example 1. Consider a random vector $X = (X_1, X_2)$, where X_1 and aX_2 are independent random variables each having $\mathcal{N}(0,1)$ probability distribution. One can think of (X_1, X_2) as a randomly selected point on $\mathbb{R}^2$. We want to find the joint pdf of that point expressed in polar coordinates.

Solution: Denote the polar coordinates of a randomly selected point by (R, θ) as a random vector and we'll use (r, θ) for its values. In other words, we have a transformation $(X_1, X_2) \to (R, \theta)$, where the direct transformation is defined by

$$r = \phi_1(x_1, x_2) = \sqrt{x_1^2 + x_2^2},\ \theta = \phi_2(x_1, x_2) = \arctan\frac{x_2}{x_1}.$$

The inverse transformation is then defined as

$$x_1 = \psi_1(r, \theta) = r\cos\theta,\ x_2 = \psi_1(r, \theta) = r\sin\theta.$$

The inverse transformation is differentiable and we easily find its Jacobian to be equal to $|\det D\psi(r, \theta)| = r$.

Applying formula (1), we obtain

$$f_{R,\theta}(r, \theta) = \frac{1}{2\pi} r e^{-r^2/2}$$

for all $r > 0$ and $0 \le \theta < 2\pi$ since

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{(x_1^2 + x_2^2)}{2}}.$$

Additionally, we can easily find the marginal pdf's for random variables R and θ. Simple integration of $f_{R,\theta}(r,\theta)$ in θ gives

$$f_R(r) = \begin{cases} re^{-r^2/2}, & r > 0 \\ 0, & r \leq 0. \end{cases}$$

The distribution of R is called *Rayleigh probability distribution.*

Similarly, the integration of the joint density in r yields

$$f_\theta(\theta) = \begin{cases} \frac{1}{2\pi}, & 0 \leq \theta < 2\pi \\ 0, & \theta \notin [0, 2\pi]. \end{cases}$$

In other words, the marginal density of the random variable θ has the uniform distribution on the interval $[0, 2\pi]$.

∘ When working with random vectors it is natural to operate with *conditional probability distributions.*

Assume for simplicity that we have a 2-dimensional jointly continuous random vector (X_1, X_2) (everything said below will also apply to the case of a discrete vector as well).

Definition 1. The *conditional probability density* of the random variable X_1 given $X_2 = x_2$ is defined as

$$f_{X_1|X_2=x_2}(x_1, x_2) := \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_2}(x_2)}, \tag{2}$$

where $f_{X_1,X_2}(x_1, x_2)$ is the joint pdf of the vector (X_1, X_2) and $f_{X_2}(x_2)$ is the marginal pdf of X_2.

Similarly, the conditional probability density of the random variable X_2 given $X_1 = x_1$ is defined as

$$f_{X_2|X_1=x_1}(x_1, x_2) := \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)}. \tag{3}$$

Both functions defined in (2) and (3) are indeed 1-dimensional probability density functions since they are non-negative functions and it holds that

$$\int f_{X_1|X_2=x_2}(x_1, x_2)dx_1 = 1$$

and

$$\int f_{X_2|X_1=x_1}(x_1, x_2)dx_2 = 1.$$

It follows then that

$$P(X_1 \in B_1|X_2 = x_2) = \int_B f_{X_1|X_2=x_2}(x_1, x_2)dx_1$$

and

$$P(X_1 \in B_2|X_1 = x_1) = \int_B f_{X_2|X_1=x_1}(x_1, x_2)dx_2$$

for all Borel sets B_1 and B_2.

Also, using the conditional densities one can find the relation between the expected value of one variable given a fixed value of the other. In other words, we can calculate the functions

$$g_1(x_2) := E(X_1|X_2 = x_2) = \int x_1 f_{X_1|X_2=x_2}(x_1, x_2)dx_1$$

and

$$g_2(x_1) := E(X_2|X_1 = x_1) = \int x_2 f_{X_2|X_1=x_1}(x_1, x_2)dx_2.$$

The graph of the function $y = g_1(x_2)$ is then called the *curve of regression* of the variable X_1 on the variable X_2. Accordingly, the function $y = g_2(x_1)$ is then called the curve of regression of the variable X_2 on the variable X_1.

Example 2. (Example 1 continued) Find the conditional density of the variable X_1 given $R = r$, where $r > 0$ is a fixed value.

Solution: By definition,

$$f_{X_1|R=r}(x_1, r) = \frac{f_{X_1,R}(x_1, r)}{f_R(r)}.$$

Therefore, we have only to find the joint pdf of the vector (X_1, R) since the marginal density $f_R(r)$ was found in Example 1 (Rayleigh distribution).

To find the density of the vector (X_1, R), we consider the transformation

$$\phi : (x_1, x_2) \to (x_1, r).$$

However, the above transformation does not have the unique inverse. In fact, we have two inverse transformations

$$\psi_{\pm}(x_1, r) := (x_1, \pm\sqrt{r^2 - x_1^2}). \tag{4}$$

As the result, we cannot apply directly formula (1) established earlier.

However, we still can follow the general procedure in calculation of the joint density of the vector $Y = (X_1, R)$ when $Y = \phi(X)$:

$$P_Y(B) = P(Y = \phi(X) \in B) = P(X \in \psi_+(B) \cup \psi_-(B)) =$$

[we apply the addition formula next]

$$P(X \in \psi_+(B)) + P(X \in \psi_-(B)) - P(X \in \psi_+(B) \cap X \in \psi_-(B)).$$

Now, the sets $\psi_+(B)$ and $\psi_-(B)$ would overlap only when

$$\{x_1 = \pm r, x_2 = 0\}$$

yielding

$$P(X \in \psi_+(B) \cap X \in \psi_-(B)) \leq P(X_2 = 0) = 0$$

since X_2 is a continuous random variable.

Using this fact and the symmetry, we obtain that

$$P_Y(B) = 2P(X \in \psi_+(B)).$$

Both transformations ψ_+ and ψ_- in (4) are $1:1$ and we can find the corresponding density of Y using the transformation ψ_+ and the formula established earlier.

First, we can calculate following (4) the Jacobian of transformation:

$$|D\psi_+(x_1, r)| = \frac{r}{\sqrt{r^2 - x_1^2}}.$$

The density corresponding to the transformation ψ_+ is then the function

$$\frac{1}{2\pi} e^{-r^2/2} \frac{r}{\sqrt{r^2 - x_1^2}}$$

since

$$f_{X_1,X_2}(\psi_+(x_1, r)) = \frac{1}{2\pi} e^{-r^2/2}.$$

As the result, the joint density $f_{X_1,R}(x_1, r)$ is the given as

$$f_{X_1,R}(x_1, r) = \frac{1}{\pi} e^{-r^2/2} \frac{r}{\sqrt{r^2 - x_1^2}}$$

so that

$$f_{X_1|R=r}(x_1, r) = \frac{1}{\pi} \frac{1}{\sqrt{r^2 - x_1^2}}$$

for all $|x_1| < r$ (and 0 otherwise).

Example 3. Let X, Y be a jointly continuous vector where X and Y are independent random variables with the pdf's $f_X(x)$ and $f_Y(y)$, respectively. We define $Z := X + Y$. What is the pdf of Z?

Solution: We calculate the cdf of Z as

$$F_Z(z) := P(X + Y \le z) = \int\int_{\{x+y\le z\}} f(x, y)dxdy =$$

$$\int\int_{\{x+y\le z\}} f_X(x)f_Y(y)dxdy = [x = z - y] =$$

$$\int_{-\infty}^{\infty} (\int_{-\infty}^{z-y} f_X(x)dx)f_Y(y)dy =$$

$$\int_{-\infty}^{\infty} F_X(z - y)f_Y(y)dy.$$

We can differentiate now both sides in z to obtain:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y)f_Y(y)dy =: f_X \star f_Y(z).$$

The integral in the last relation is said to be the *convolution* of functions f_X and f_Y. Thus, we have shown that the density of the sum of two independent random variables is the convolution of their marginal densities. The result can be extended to the sum of more than two variables.