MATH 461 LECTURE NOTE
WEEK 12

DAESUNG KIM

1. Conditional Distribution (Sec 6.4-6)

Suppose X and Y are discrete with the joint pmf $p(x,y)$, that is $\mathbb{P}(X=x,Y=y)=p(x,y)$. Let y satisfy $p_Y(y)=\sum_x p(x,y)>0$. The conditional pmf of X given $Y=y$ is defined by

$$p_{X|Y}(x|y)=\mathbb{P}(X=x|Y=y)=\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}=\frac{p(x,y)}{p_Y(y)}.$$

Note that if X and Y are independent, then $p_{X|Y}(x|y)=p_X(x)$. The conditional cdf of X given $Y=y$ is

$$F_{X|Y}(t|y)=\mathbb{P}(X\le t|Y=y)=\sum_{x\le t}p_{X|Y}(x|y).$$

Example 1. If X and Y are independent Poisson random variables with respective parameters λ_1 and λ_2, calculate the conditional distribution of X given that $X+Y=n$.

Suppose X and Y are jointly continuous with joint density $f(x,y)$. For y with $f_Y(y)>0$, the conditional density of X given $Y=y$ is defined by

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}.$$

If X and Y are independent, then $f_{X|Y}(x|y)=f_X(x)$. Then, the conditional probability and the conditional cdf of X given $Y=y$ can be written as

$$\mathbb{P}(X\in A|Y=y)=\int_A f_{X|Y}(x|y)\,dx$$

$$F_{X|Y}(t|y)=\mathbb{P}(X\le t|Y=y)=\int_{-\infty}^{t} f_{X|Y}(x|y)\,dx.$$

Example 2. Suppose that the joint density of X and Y is given by

$$f(x,y)=\begin{cases}\frac{e^{-\frac{x}{y}}e^{-y}}{y}, & 0<x,y<\infty,\\ 0, & \text{otherwise.}\end{cases}$$

Find $f_{X|Y}(x|y)$ and $\mathbb{P}(X>1|Y=y)$.

Bivariate normal random variable. Jointly continuous random variables X and Y are bivariate normal if their density is given by

$$f(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}\left((\frac{x-\mu_X}{\sigma_X})^2+(\frac{y-\mu_Y}{\sigma_Y})^2-2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)}$$

where $\sigma_X,\sigma_Y>0$, $\rho\in(-1,1)$, and $\mu_X,\mu_Y\in\mathbb{R}$. We denote by

$$\begin{pmatrix}X\\Y\end{pmatrix}\sim N\left(\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix},\begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right).$$

Proposition 3. (i) *$X\sim N(\mu_X,\sigma_X^2)$ and $Y\sim N(\mu_Y,\sigma_Y^2)$. In particular, $\mathbb{E}[X]=\mu_X$, $\mathbb{E}[Y]=\mu_Y$, $\mathrm{Var}(X)=\sigma_X^2$, and $\mathrm{Var}(Y)=\sigma_Y^2$.*
(ii) *The random variable X given $Y=y$ is normal with mean $\mu_X+\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)$ and variance $\sigma_X^2(1-\rho^2)$.*

Proof. Let $\overline{x} = \frac{x-\mu_X}{\sigma_X}$ and $\overline{y} = \frac{y-\mu_Y}{\sigma_Y}$, then

$$\begin{aligned} f(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\overline{x}^2+\overline{y}^2-2\rho\overline{x}\overline{y}\right)} \\ &= \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(\overline{x}-\rho\overline{y})^2} e^{-\frac{1}{2}\overline{y}^2}. \end{aligned}$$

Since

$$\begin{aligned} \int_{\mathbb{R}} e^{-\frac{1}{2(1-\rho^2)}(\overline{x}-\rho\overline{y})^2}\,dx &= \int_{\mathbb{R}} e^{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left(x-\left(\mu_X+\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)\right)^2}\,dx \\ &= \int_{\mathbb{R}} e^{-\frac{1}{2\sigma_X^2(1-\rho^2)}x^2}\,dx \\ &= \sqrt{2\pi\sigma_X^2(1-\rho^2)}, \end{aligned}$$

we have

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{1}{2}\overline{y}^2} = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

and so $Y \sim N(\mu_Y, \sigma_Y^2)$. The same argument for X yields $X \sim N(\mu_X, \sigma_X^2)$. A direct computation leads to

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{1}{\sqrt{2\pi\sigma_X^2(1-\rho^2)}} e^{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left(x-\left(\mu_X+\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right)\right)^2}$$

as desired. □

Remark 4. The parameter ρ represents how X and Y correlated.

Joint distribution of maximum and minimum. Let $X_1, X_2, \cdots, X_n$ be independent jointly continuous random variables with the common cdf $F(t)$. Let $U = \max\{X_1, X_2, \cdots, X_n\}$ and $V = \min\{X_1, X_2, \cdots, X_n\}$.

Proposition 5. *The joint density of U and V is*

$$f_{U,V}(u,v) = n(n-1)(F(u)-F(v))^{n-2} f(u)f(v).$$

2. Joint Distribution of Functions of RVs (Sec 6.7)

Suppose X_1 and X_2 are jointly continuous random variables with joint probability density $f_X(x_1,x_2)$. Let $g = (g_1, g_2)$ and $Y_1 = g_1(X_1,X_2)$, $Y_2 = g_2(X_1,X_2)$ with

$$|J_g(x_1,x_2)| = \begin{vmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{vmatrix} = \left| \frac{\partial g_1}{\partial x_1}\frac{\partial g_2}{\partial x_2} - \frac{\partial g_1}{\partial x_2}\frac{\partial g_2}{\partial x_1} \right| \neq 0$$

where

$$J_g(x_1,x_2) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} \end{pmatrix}.$$

Suppose that there exists a map $h = (h_1, h_2)$ such that the equations $y_1 = g_1(x_1,x_2)$ and $y_2 = g_2(x_1,x_2)$ can be uniquely solved for x_1 and x_2 so that $x_1 = h_1(y_1,y_2)$ and $x_2 = h_2(y_1,y_2)$. Here, J_g is called the Jacobian of g, and h is the inverse of g. Note that if g is continuously differentiable (that is, g_1 and g_2 are differentiable and their derivatives are continuous), then $J_h = J_g^{-1}$ (due to the inverse function theorem).

Let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$, then the change of variable formula provides

$$\begin{aligned}\iint_B f_Y(y_1, y_2)\, dy_1 dy_2 &= \mathbb{P}((Y_1, Y_2) \in B) \\ &= \mathbb{P}((X_1, X_2) \in h(B)) \\ &= \iint_{h(B)} f_X(x_1, x_2)\, dx_1 dx_2 \\ &= \iint_B f_X(h_1(y_1, y_2), h_2(y_1, y_2)) |J_h(y_1, y_2)|\, dy_1 dy_2 \\ &= \iint_B f_X(x_1, x_2) |J_g(x_1, x_2)|^{-1}\, dy_1 dy_2.\end{aligned}$$

Thus, the joint density of Y_1 and Y_2 is

$$\begin{aligned}f_Y(y_1, y_2) &= f_X(h_1(y_1, y_2), h_2(y_1, y_2)) |J_h(y_1, y_2)| \\ &= f_X(h_1(y_1, y_2), h_2(y_1, y_2)) |J_g(h_1(y_1, y_2), h_2(y_1, y_2))|^{-1}.\end{aligned}$$

Example 6. Let X_1 and X_2 be jointly continuous random variables with probability density function $f(X_1, X_2)$. Let $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$. Find the joint density function of Y_1 and Y_2 in terms of $f(X_1, X_2)$.

Example 7. Let (X, Y) denote a random point in the plane, and assume that the rectangular coordinates X and Y are independent standard normal random variables. What is the joint distribution of R, Θ, the polar coordinate representation of (X, Y).

3. Expectation of functions of RVs (Sec 7.2, 3)

Let $g(x, y)$ be a function on $\mathbb{R}^2$. If X, Y are discrete with joint pmf $p(x, y)$, then

$$\mathbb{E}[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y).$$

If X, Y are jointly continuous with joint density $f(x, y)$, then

$$\mathbb{E}[g(X, Y)] = \iint g(x, y) f(x, y)\, dx dy.$$

Example 8. Let X and Y be independent uniform random variables on $(0, 1)$. Find $\mathbb{E}[|X - Y|^k]$.

Expectations of sum and product

(i) $\mathbb{E}[\sum_{i=1}^n X_i] = \sum_{i=1}^n \mathbb{E}[X_i]$

(ii) If X and Y are independent, then

$$\mathbb{E}[g(X) h(Y)] = \mathbb{E}[g(X)] \mathbb{E}[h(Y)].$$

Example 9 (A random walk in the plane)**.** Consider a particle initially located at the origin in the plane $\mathbb{R}^2$. Suppose that the new position after each step is one unit of distance from the previous position and at an angle of orientation from the previous position that is uniformly distributed over $(0, 2\pi)$. Compute the expected square of the distance from the origin after n steps.

The number of events that occur. For events $A_1, A_2, \cdots, A_n$, we consider the corresponding indicator random variables $X_i = I_{A_i}$, defined by

$$I_{A_i} = \begin{cases} 1, & \text{if } A_i \text{ occurs,} \\ 0, & \text{otherwise.} \end{cases}$$

Note that $\mathbb{E}[I_{A_i}] = \mathbb{P}(A_i)$ and $\mathrm{Var}(I_{A_i}) = \mathbb{P}(A_i)(1-\mathbb{P}(A_i))$. The number of the events that occur can be written as the sum of X_i. Let $X = \sum_i X_i$. Then,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{n} \mathbb{E}[X_i] = \sum_{i=1}^{n} \mathbb{P}(A_i),\\
\mathbb{E}[X^2] &= \sum_{i,j=1}^{n} \mathbb{E}[X_iX_j] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2\sum_{i<j}^{n} \mathbb{E}[X_iX_j] = \sum_{i=1}^{n} \mathbb{P}(A_i) + 2\sum_{i<j}^{n} \mathbb{P}(A_i \cap A_j),\\
\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2\\
&= \sum_{i=1}^{n} \mathbb{P}(A_i)(1-\mathbb{P}(A_i)) + 2\sum_{i<j}^{n} \left(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\right),\\
&= \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j}^{n} \left(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\right).
\end{aligned}$$

Example 10. Let X be a binomial random variable with parameters n and p. Compute $\mathbb{E}[X]$, $E[X^2]$ and $\mathrm{Var}(X)$.

Example 11. A group of 20 people consisting of 10 men and 10 women is randomly arranged into 10 pairs of 2 each. Compute the expectation and variance of the number of pairs that consist of a man and a woman.

References

[SR] Sheldon Ross, *A First Course in Probability*, 9th Edition, Pearson

Department of Mathematics, University of Illinois at Urbana-Champaign
E-mail address: daesungk@illinois.edu