Math 40 Probability and Statistical Inference

Winter 2021

Yoonsang Lee (yoonsang.lee@dartmouth.edu)

Lecture 8: Multivariate random variables

Joint cdf and density (3.1)

Now we have more than one (scalar-valued) random variables, say X and Y.
The **joint** cdf of X, Y is the probability

$$F(x, y) = Pr(X \leq x, Y \leq y)$$

1. $0 \leq F(x, y) \leq 1$.
2. $F(-\infty, y) = F(x, -\infty) = 0$ and $F(\infty, \infty) = 1$.
3. If $x_1 \leq x_2$, then $F(x_1, y) \leq F(x_2, y)$ for all y. Also for $y_1 \leq y_2$, $F(x, y_1) \leq F(x, y_2)$ for all x.
4. If $x_1 \leq x_2$ and $y_1 \leq y_2$, then $F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \geq 0$.

The **marginal** cdf of X is the limit of the joint cdf when $y \to \infty$

$$F(x) = \lim_{y \to \infty} F(x, y)$$

Joint cdf and density (3.1)

The **joint** density (or pdf) of X and Y is the mixed partial derivative of the cdf

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

1. $f(x, y) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x, y) dxdy = 1$.

Joint cdf and density (3.1)

Expectation $E[g(X,Y)] = \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy = 1$
For any a and b, we have

$$E[aX + bY] = aE[X] + bE[Y]$$

Joint cdf and density (3.1)

Exercise 3.1.2 (a) Prove that Property 4 of the bivariate cdf implies the property 3.

Joint cdf and density (3.1)

Exercise 3.1.4 Let $H(x, y) = max(x, y)/(x + y)$ for $x > 0$ and $y > 0$ and 0 elsewhere. Can H be a cdf? Hint: first prove that if F is a cdf, then $\lim_{x\to\infty} F(x, x) = 1$.

Joint cdf and density (3.1)

Exercise 3.1.10 Suppose that $f(x)$ is a density. Are (a) $f(x)f(y)$, (b) $0.5(f(x)+f(y))$, (c) $\lambda f(x)+(1-\lambda)f(y)$ where $0 \leq \lambda \leq 1$, bivariate densities?

Independence (3.2)

We call two events A and B are independent if and only if

$$Pr(A \cap B) = Pr(A)Pr(B)$$

For two random variables, they are independent if and only if

$$F(x, y) = F_X(x)F_Y(y)$$

This is equivalent to

$$f(x, y) = f_X(x)f_Y(y)$$

Independence (3.2)

Example 2.7 Mary and John talking. Mary and John participate in a zoom conference call set up at 10 am. The bivariate density of the times they join the group is given by $e^{-(x+y)}$, where x stands for Mary and y stands for John (minutes after 10 am).
(a) Prove that Mary and John join the meeting independently.

Independence (3.2)

Example 2.7 Mary and John talking. Mary and John participate in a zoom conference call set up at 10 am. The bivariate density of the times they join the group is given by $e^{-(x+y)}$, where x stands for Mary and y stands for John (minutes after 10 am).
(b) Calculate the probability that John joins the group two minutes after Mary and check the answer via simulations.

Independence (3.2)

Example 2.7 Mary and John talking. Mary and John participate in a zoom conference call set up at 10 am. The bivariate density of the times they join the group is given by $e^{-(x+y)}$, where x stands for Mary and y stands for John (minutes after 10 am).
(c) Estimate the size of sample (nExp) to achieve standard error < 0.001.

Convolution (3.2.1)

If we know the joint density of X and Y, say $f(x, y)$, what is the density of $Z = X + Y$?
The cdf is given by

$$F_Z(z) = Pr(Z \leq z) = Pr(X + Y \leq z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-x} f(x, y) dy dx$$

From this, the pdf is given by

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} f(x, z - x) dx$$

If X and Y are independent, $f(x, y) = f_X(x) f_Y(y)$, that is

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

This integral is called the **convolution** of f_X and f_Y.

Convolution (3.2.1)

Example 3.10 Sum of two uniformly distributed random variables. Find the cdf and pdf of the sum of two independent uniformly distributed random variables on $(0, 1)$.

Independence (3.2)

Some facts

- A convolution of two independent normal distributions is a normal distribution.
- If X and Y are independent random variables then any functions of these variables are independent as well.
- If X and Y are independent random varaibles,

$$E(XY) = E(X)E(Y)$$

Independence (3.2)

If we assume independence between X and Y, many things become convenient. However, showing independence is not a trivial task except a few cases.

Independence (3.2)

If we assume independence between X and Y, many things become convenient. However, showing independence is not a trivial task except a few cases.

Example 3.17 Uniform fairs. U and V are independent random variables uniformly distributed on $(0, 1)$. Prove that $X = \min(U, V)$ and $Y = \max(U, V)$ are not independent.

Exericses

3.2.4 Prove that the mean of the random variable with density

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

is equal to $E(X) + E(Y)$.

Exericses

3.2.10 Let X and Y are independent uniformly distributed on $(0, 1)$. Approximate the cdf of $X + Y$ using the central limit theorem with $n = 2$.

- From CLT, $\frac{X-1/2+Y-1/2}{\sqrt{2/12}} \sim \mathcal{N}(0, 1)$.
- Here $1/2$ on the numerator is the mean of X (and Y)). $\frac{1}{12}$ is the variance of X (and Y).
- After a bit of algebra, we have $X + Y = \frac{Z}{\sqrt{6}} + 1$, where $Z \sim \mathcal{N}(0, 1)$.
- That is, $X + Y \sim \mathcal{N}(1, \frac{1}{6})$.