Joint distributions (part 2)

Lecture 6a (STAT 24400 F24)

Continuous joint distributions

If the pair (X, Y) is continuously distributed, then there is a joint density $f(x, y)$ which is piecewise continuous, nonnegative, and integrates to 1, such that

$$\mathbb{P}((X, Y) \in A) = \iint_A f(x, y)\, \mathrm{d}y\, \mathrm{d}x$$

for any "reasonable" region $A \subseteq \mathbb{R}^2$.

(we may write $f_{X,Y}$ or f)

Continuous joint distributions (CDF and density)

Joint CDF:

$$\begin{aligned} F(x, y) = \mathbb{P}(X \le x, Y \le y) &= \mathbb{P}\{(X, Y) \in (-\infty, x] \times (-\infty, y]\} \\ &= \int_{s=-\infty}^{x} \int_{t=-\infty}^{y} f(s, t)\, \mathrm{d}t\, \mathrm{d}s \end{aligned}$$

Take the derivative with respect to x and y:

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

Generalization: $\frac{\partial^k}{\partial x_1 \cdots \partial x_k} F(x_1, \cdots, x_k) = f(x_1, \cdots, x_k)$

Continuous joint distributions (marginal CDF and density)

Marginal CDF for X:

$$\begin{aligned} F_X(x) = \mathbb{P}(X \le x) &= \mathbb{P}\{(X, Y) \in (-\infty, x] \times (-\infty, \infty)\} \\ &= \int_{s=-\infty}^{x} \int_{y=-\infty}^{\infty} f(s, y)\, \mathrm{d}y\, \mathrm{d}s \end{aligned}$$

Marginal density for X:

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = \int_{y=-\infty}^{\infty} f(x, y)\, \mathrm{d}y$$

Compare to the discrete case:

$$p_X(x) = \sum_y p(x, y)$$

Discrete / continuous / mixed?

Recall: for univariate distributions,

If X has zero mass at any point (i.e., $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$), then X is continuous.

However, for a joint distribution on (X, Y),

It might be the case that

$$\mathbb{P}\big((X, Y) = (x, y)\big) = 0 \text{ for all points } (x, y) \in \mathbb{R}^2,$$

but the distribution is still not continuous;
that is, the joint density does not exist.

Examples (mass on probability zero set)

Examples

1. (X, Y) is a point drawn uniformly at random from the unit circle.
2. Sample a person at random. $X = \#$ of children and $Y =$ height (continuous).

A joint density does not exist for either of the examples.

Why is there no joint density for these examples?

If (X, Y) had density f, then any zero-area region $A \subseteq \mathbb{R}^2$,

$$\mathbb{P}\big((X, Y) \in A\big) = \iint_A f(x, y)\, dy\, dx = 0$$

In fact, each example has a zero-area region $A \subseteq \mathbb{R}^2$ such that

$$\mathbb{P}\big((X, Y) \in A\big) = 1 \qquad \Longrightarrow \qquad \text{Contradiction.}$$

Uniform distribution on a region

Example

Let (X, Y) be sampled uniformly at random from the unit square $[0, 1]^2$.

Then the density of (X, Y) is

$$f(x, y) = \begin{cases} 1, & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, for any region $A \subseteq [0, 1]^2$,

$$\mathbb{P}\big((X, Y) \in A\big) = \iint_A f(x, y)\, dy\, dx = \iint_A 1\, dy\, dx = \text{Area}(A)$$

Example

More generally: if (X, Y) is sampled uniformly from a region $B \subseteq \mathbb{R}^2$

$$f(x, y) = \begin{cases} 1/\text{Area}(B), & (x, y) \in B, \\ 0, & \text{otherwise.} \end{cases}$$

and so then for any $A \subseteq B$, $\mathbb{P}\big((X, Y) \in A\big) = \frac{\text{Area}(A)}{\text{Area}(B)}$.

Bivariate normal distribution

Bivariate normal distribution

Parameters

Means $\mu_1, \mu_2 \in \mathbb{R}$, variances $\sigma_1^2, \sigma_2^2 > 0$, correlation $\rho \in (-1, 1)$

Density function For $(x, y) \in \mathbb{R}^2$,

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right\}$$

Special case: standard bivariate normal with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, $\rho = 0$:

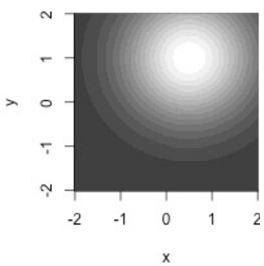
$$f(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}\left(x^2 + y^2\right)\right\}$$

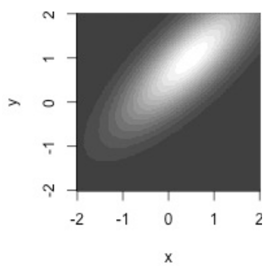
Examples (bivariate normal heat maps)

Examples

Density level-curves with $\mu_1 = 0.5$, $\mu_2 = 1$, $\sigma_1^2 = \sigma_2^2 = 1$:

$\rho = 0$ or $\rho = 0.8$





We may calculate that marginally (by integrating out one variable),

$$X \sim \mathrm{N}(\mu_1, \sigma_1^2), \quad Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$$

Examples (marginal discrete/continuous)

Example (one variable is discrete and the other is continuous)

Suppose that windspeed on cloudy days is distributed as *Exponential*(0.2), and on sunny days is distributed as *Exponential*(0.8).

60% of days are cloudy and 40% are sunny.

Let

$$X = \text{ windspeed }, \qquad Y = \mathbb{1}_{\text{Sunny}}$$

Remarks

In this example, there is no point-mass at any point (x, y) in the joint distribution for (X, Y).

However, the distribution is not continuous, and there is no joint density.

Examples (marginal discrete/continuous) (cont.)

Question 1: If the windspeed is above 4, what's the chance that it's cloudy?

$$\mathbb{P}(Y = 0 \mid X > 4) = \frac{\mathbb{P}(Y = 0, X > 4)}{\mathbb{P}(X > 4)} = \frac{\mathbb{P}(Y = 0, X > 4)}{\mathbb{P}(Y = 0, X > 4) + \mathbb{P}(Y = 1, X > 4)}$$

$$= \frac{0.6 \cdot e^{-0.2 \times 4}}{0.6 \times e^{-0.2 \times 4} + 0.4 \cdot e^{-0.8 \times 4}}$$

$$= 0.943$$

where we used the conditional probability definition (or the multiplication law)

$$\mathbb{P}(Y = 0, X > 4) = \mathbb{P}(Y = 0)\,\mathbb{P}(X > 4 \mid Y = 0)$$
$$\mathbb{P}(Y = 1, X > 4) = \mathbb{P}(Y = 1)\,\mathbb{P}(X > 4 \mid Y = 1)$$

Examples (marginal discrete/continuous) (cont.)

Question 2: What is the marginal distribution of X?

On the support $x \in (0, \infty)$,

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) \\ &= \textstyle\sum_y \mathbb{P}(X \leq x, Y = y) \\ &= \mathbb{P}(X \leq x, Y = 0) + \mathbb{P}(X \leq x, Y = 1) \\ &= \mathbb{P}(Y = 0)\,\mathbb{P}(X \leq x \mid Y = 0) + \mathbb{P}(Y = 1)\,\mathbb{P}(X \leq x \mid Y = 1) \\ &= 0.6(1 - e^{-0.2x}) + 0.4(1 - e^{-0.8x}) \end{aligned}$$

Taking the derivative:

$$f_X(x) = 0.6(-(-0.2) \cdot e^{-0.2x}) + 0.4(-(-0.8) \cdot e^{-0.8x}) = 0.12e^{-0.2x} + 0.32e^{-0.8x}\ .$$

Independence

Recall: random variables X and Y are **independent** if for all values (x, y),

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

For a continuous joint distribution on (X, Y), it is equivalent to characterize independence based on density:

- Equivalently, it holds for all x, y that
$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$
- Equivalently, it holds for all x, y that
$$f_{X,Y}(x,y) = (\text{some function of } x) \times (\text{some function of } y)$$

Examples (independence in bivariate normal)

Bivariate normal distribution on (X, Y) with parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho$:

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right)\right\}$$

Recall that marginally, $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$.

If we set correlation $\rho = 0$, then X and Y are independent,
because the density $f(x, y)$ factors as (function of x)·(function of y):

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \cdot \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}$$

Examples (independence in uniform distribution)

Uniform distribution on the unit square:

- Density
$$f(x,y) = \mathbb{1}_{0\le x\le 1, 0\le y\le 1}$$
- We can see that $f(x, y)$ factors over x and y:
$$f(x,y) = \mathbb{1}_{0\le x\le 1, 0\le y\le 1} = \mathbb{1}_{0\le x\le 1} \cdot \mathbb{1}_{0\le y\le 1}$$
- Therefore $X \perp\!\!\!\perp Y$
- In fact, X and Y each have marginal distribution $\mathit{Uniform}[0, 1]$.

The same is true for a uniform distribution over any *rectangular* region B.

Examples (non-independence)

However,
the factorization does not work for uniform distribution on other regions.

Example Uniform density on the unit disk:

$$f(x,y) = \frac{1}{\pi} \cdot \mathbb{1}_{x^2+y^2\le 1} \quad \leftarrow \text{cannot be factored}$$

- $(x, y) = (0.1, 0.9)$ is possible, but $(x, y) = (0.9, 0.9)$ is not possible.
- {possible (X, Y) values} $\neq$ {possible X values} $\times$ {possible Y values}
- Factorization is impossible.