Joint distributions (part 3)

Lecture 6b (STAT 24400 F24)

Conditional distribution

We often want to ask about probabilities of X based on some (partial) knowledge of Y, e.g. $\mathbb{P}(X \geq 3 \mid Y \geq 7)$.

We may also ask about the distribution of X given exact knowledge of Y, e.g. $\mathbb{P}(X \geq 3 \mid Y = 7)$.

this is the default meaning of "conditional distribution"

Conditional distribution

We often want to ask about probabilities of X based on some (partial) knowledge of Y, e.g. $\mathbb{P}(X \geq 3 \mid Y \geq 7)$.

We may also ask about the distribution of X given exact knowledge of Y, e.g. $\mathbb{P}(X \geq 3 \mid Y = 7)$.

this is the default meaning of "conditional distribution"

For example,

- If we draw a hand of 10 cards, what's the distribution of the number of Kings, given that the hand contains 4 red cards?
- If we sample a random person from the population, what's the distribution of their height, given that the age of the person chosen is 12.5 years?

Conditional distribution: discrete case

If we know that $Y = y$, what is the distribution of X?

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\sum_{x'} p_{X,Y}(x', y)} .$$

Conditional distribution: discrete case

If we know that $Y = y$, what is the distribution of X?

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\sum_{x'} p_{X,Y}(x', y)} .$$

We write this as $p_{X|Y}(x \mid y)$, the conditional PMF of X given $Y = y$.

Conditional distribution: discrete case

If we know that $Y = y$, what is the distribution of X?

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)} = \frac{p_{X,Y}(x, y)}{\sum_{x'} p_{X,Y}(x', y)} .$$

We write this as $p_{X|Y}(x \mid y)$, the conditional PMF of X given $Y = y$.

$p_{X|Y}(\cdot \mid y)$ defines a valid PMF at any fixed y.

For example, it sums to 1:

$$\sum_x p_{X|Y}(x \mid y) = \sum_x \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{\mathbb{P}(Y = y)}{\mathbb{P}(Y = y)} = 1 .$$

Continuous case: problem with point-mass

The continuous case is harder...

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = ????$$

the denominator is zero

Continuous case: problem with point-mass

The continuous case is harder...

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = ????$$

the denominator is zero

We will instead need to work with densities.

Continuous case: univariate density in a tiny interval

First we take another look at density in the univariate (not joint) case:

For small $\epsilon > 0$,

$$\mathbb{P}(x < X < x + \epsilon) = \int_{t=x}^{x+\epsilon} f(t)\, \mathrm{d}t \approx \epsilon \cdot f(x) \; ,$$

if $f(t)$ continuous near x

So, $\mathbb{P}(X = x)$ is zero, but $\mathbb{P}(X \approx x)$ (X falls near x) is validly approximated via the density $f(x)$.

Continuous case: conditional density in a tiny region

For the joint case, we can't ask questions like $\mathbb{P}(X = x \mid Y = y)$,
but $\mathbb{P}(X \approx x \mid Y \approx y)$ should be approximated via the *conditional density*.

Continuous case: conditional density in a tiny region

For the joint case, we can't ask questions like $\mathbb{P}(X = x \mid Y = y)$,
but $\mathbb{P}(X \approx x \mid Y \approx y)$ should be approximated via the *conditional density*.

$$\mathbb{P}(x < X < x + \epsilon \mid y < Y < y + \delta) = \frac{\mathbb{P}(x < X < x + \epsilon, y < Y < y + \delta)}{\mathbb{P}(y < Y < y + \delta)}$$
$$= \frac{\int_{u=x}^{x+\epsilon} \int_{v=y}^{y+\delta} f_{X,Y}(u,v)\, \mathrm{d}v\, \mathrm{d}u}{\int_{v=y}^{y+\delta} f_Y(v)\, \mathrm{d}v} \approx \frac{\epsilon \cdot \delta \cdot f_{X,Y}(x,y)}{\delta \cdot f_Y(y)} = \epsilon \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)} \; .$$

Continuous case: conditional density in a tiny region

For the joint case, we can't ask questions like $\mathbb{P}(X = x \mid Y = y)$,
but $\mathbb{P}(X \approx x \mid Y \approx y)$ should be approximated via the *conditional density*.

$$\mathbb{P}(x < X < x + \epsilon \mid y < Y < y + \delta) = \frac{\mathbb{P}(x < X < x + \epsilon, y < Y < y + \delta)}{\mathbb{P}(y < Y < y + \delta)}$$
$$= \frac{\int_{u=x}^{x+\epsilon} \int_{v=y}^{y+\delta} f_{X,Y}(u,v)\, \mathrm{d}v\, \mathrm{d}u}{\int_{v=y}^{y+\delta} f_Y(v)\, \mathrm{d}v} \approx \frac{\epsilon \cdot \delta \cdot f_{X,Y}(x,y)}{\delta \cdot f_Y(y)} = \epsilon \cdot \frac{f_{X,Y}(x,y)}{f_Y(y)} \; .$$

So by analogy, we should define the **conditional density** as

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

Independence (in PMF and PDF)

If (X, Y) is discrete,

$X \perp\!\!\!\perp Y$ is equivalent to: $p_{X|Y}(x \mid y) = p_X(x)$ for all (x, y)

(and, same for reversing X and Y).

If (X, Y) is continuous,

$X \perp\!\!\!\perp Y$ is equivalent to: $f_{X|Y}(x \mid y) = f_X(x)$ for all (x, y)

(and, same for reversing X and Y).

Law of total probability (in terms of conditional PMF & PDF)

For probabilities, given a partition $B_1, B_2, \dots$, we showed

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)$$

Law of total probability (in terms of conditional PMF & PDF)

For probabilities, given a partition $B_1, B_2, \dots$, we showed

$$\mathbb{P}(A) = \sum_i \mathbb{P}(A \cap B_i) = \sum_i \mathbb{P}(A \mid B_i)\,\mathbb{P}(B_i)$$

Similar rules hold for conditional distributions:

- For discrete (X, Y),

$$p_Y(y) = \sum_x p_{X,Y}(x, y) = \sum_x p_{Y|X}(y \mid x) p_X(x)$$

- For a continuous (X, Y),

$$f_Y(y) = \int_{x=-\infty}^{\infty} f_{X,Y}(x, y)\, \mathrm{d}x = \int_{x=-\infty}^{\infty} f_{Y|X}(y \mid x) f_X(x)\, \mathrm{d}x$$

Example (uniform on unit disk)

(X, Y) is chosen uniformly at random from the unit disk, $\{x^2 + y^2 \le 1\}$.

Joint density:

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example (uniform on unit disk)

(X, Y) is chosen uniformly at random from the unit disk, $\{x^2 + y^2 \le 1\}$.

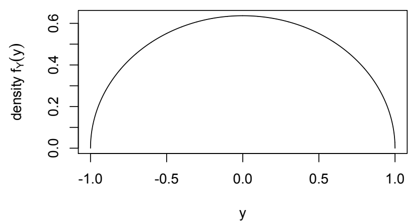
Joint density:

$$f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

Marginal density:

$$f_Y(y) = \int_{x=-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \int_{x=-\sqrt{1-y^2}}^{\sqrt{1-y^2}} \frac{1}{\pi}\,\mathrm{d}x = \frac{2\sqrt{1-y^2}}{\pi} \quad \text{for } y \in [-1,1]$$

(& analogous for $f_X(x)$)



Example (cont.)

Conditional density:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}} = \frac{1}{2\sqrt{1-y^2}}$$

Example (cont.)

Conditional density:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}} = \frac{1}{2\sqrt{1-y^2}}$$

If $Y = y$ for $y \in [-1,1]$, possible values for X lie in $[-\sqrt{1-y^2}, \sqrt{1-y^2}]$.

Example (cont.)

Conditional density:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}} = \frac{1}{2\sqrt{1-y^2}} \quad \text{for } x \in [-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

If $Y = y$ for $y \in [-1,1]$, possible values for X lie in $[-\sqrt{1-y^2}, \sqrt{1-y^2}]$.

Example (cont.)

Conditional density:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}} = \frac{1}{2\sqrt{1-y^2}} \quad \text{for } x \in [-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

If $Y = y$ for $y \in [-1, 1]$, possible values for X lie in $[-\sqrt{1-y^2}, \sqrt{1-y^2}]$.

Equivalently:

$$X \mid Y = y \sim \text{Uniform}[-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

(& analogous for $Y \mid X$)

Example (cont.)

Conditional density:

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{\frac{1}{\pi}}{\frac{2\sqrt{1-y^2}}{\pi}} = \frac{1}{2\sqrt{1-y^2}} \quad \text{for } x \in [-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

If $Y = y$ for $y \in [-1, 1]$, possible values for X lie in $[-\sqrt{1-y^2}, \sqrt{1-y^2}]$.

Equivalently:

$$X \mid Y = y \sim \text{Uniform}[-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

(& analogous for $Y \mid X$)

Remarks : The *marginal* distribution of X (or of Y) is not uniform,
but the *conditional* distribution of $X|Y$ (or of $Y|X$) is uniform.

Functions of jointly distributed random variables

Expectations

If $Y = g(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ have a joint PMF p, then

$$\mathbb{E}(Y) = \sum_{(x_1,\ldots,x_n)} g(x_1, \ldots, x_n) \cdot p(x_1, \ldots, x_n) \, .$$

Functions of jointly distributed random variables

Expectations

If $Y = g(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ have a joint PMF p, then

$$\mathbb{E}(Y) = \sum_{(x_1,\ldots,x_n)} g(x_1, \ldots, x_n) \cdot p(x_1, \ldots, x_n) \, .$$

If instead $X_1, \ldots, X_n$ have a joint density f, then

$$\mathbb{E}(Y) = \int_{x_1} \ldots \int_{x_n} g(x_1, \ldots, x_n) \cdot f(x_1, \ldots, x_n) \, \mathrm{d}x_n \, \ldots \, \mathrm{d}x_1 \, .$$

Functions of jointly distributed random variables

Expectations

If $Y = g(X_1, \ldots, X_n)$ where $X_1, \ldots, X_n$ have a joint PMF p, then

$$\mathbb{E}(Y) = \sum_{(x_1,\ldots,x_n)} g(x_1, \ldots, x_n) \cdot p(x_1, \ldots, x_n) \ .$$

If instead $X_1, \ldots, X_n$ have a joint density f, then

$$\mathbb{E}(Y) = \int_{x_1} \ldots \int_{x_n} g(x_1, \ldots, x_n) \cdot f(x_1, \ldots, x_n) \, \mathrm{d}x_n \ \ldots \ \mathrm{d}x_1 \ .$$

(Alternatively, can compute the distribution of Y via the joint distribution of the X_i's, and then compute $\mathbb{E}(Y)$, but this is generally less efficient.)

Expectation of product function under independence

If $X \perp\!\!\!\perp Y$, then $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$.

Expectation of product function under independence

If $X \perp\!\!\!\perp Y$, then $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proof: Let's prove for the continuous case:

the density factors since $X \perp\!\!\!\perp Y$

$$\mathbb{E}(X \cdot Y) = \int_x \int_y x \cdot y \cdot f_{X,Y}(x,y) \, \mathrm{d}y \, \mathrm{d}x = \int_x \int_y x \cdot y \cdot f_X(x) f_Y(y) \, \mathrm{d}y \, \mathrm{d}x$$
$$= \left(\int_x x \cdot f_X(x) \, \mathrm{d}x \right) \cdot \left(\int_y y \cdot f_Y(y) \, \mathrm{d}y \right) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

Expectation of product function under independence

If $X \perp\!\!\!\perp Y$, then $\mathbb{E}(X \cdot Y) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proof: Let's prove for the continuous case:

the density factors since $X \perp\!\!\!\perp Y$

$$\mathbb{E}(X \cdot Y) = \int_x \int_y x \cdot y \cdot f_{X,Y}(x,y) \, \mathrm{d}y \, \mathrm{d}x = \int_x \int_y x \cdot y \cdot f_X(x) f_Y(y) \, \mathrm{d}y \, \mathrm{d}x$$
$$= \left(\int_x x \cdot f_X(x) \, \mathrm{d}x \right) \cdot \left(\int_y y \cdot f_Y(y) \, \mathrm{d}y \right) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

More generally:

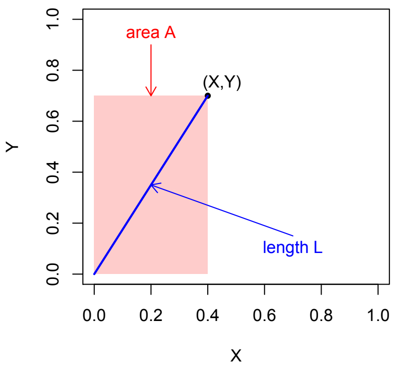
If $X \perp\!\!\!\perp Y$, then for any functions g, h, $\quad g(X) \perp\!\!\!\perp h(Y)$ and so

$$\mathbb{E}(g(X) \cdot h(Y)) = \mathbb{E}(g(X)) \cdot \mathbb{E}(h(Y))$$

Example (functions of (x, y) uniform on unit square)

The point (X, Y) is drawn uniformly from the unit square $[0, 1] \times [0, 1]$.

What is the expected value of A, the area of the rectangle?
What is the expected value of L, the length of the segment?



Example: (unit square cont.)

Joint density of (X, Y):

$$f(x, y) = \begin{cases} 1, & (x, y) \in [0, 1] \times [0, 1], \\ 0, & (x, y) \notin [0, 1] \times [0, 1]. \end{cases}$$

Example: (unit square cont.)

Joint density of (X, Y):

$$f(x, y) = \begin{cases} 1, & (x, y) \in [0, 1] \times [0, 1], \\ 0, & (x, y) \notin [0, 1] \times [0, 1]. \end{cases}$$

Equivalent:

$$f(x, y) = \mathbb{1}_{0 \le x \le 1} \cdot \mathbb{1}_{0 \le y \le 1}$$

The joint density factors $\Rightarrow$ $X \perp\!\!\!\perp Y$

Example: (unit square cont.)

Joint density of (X, Y):

$$f(x, y) = \begin{cases} 1, & (x, y) \in [0, 1] \times [0, 1], \\ 0, & (x, y) \notin [0, 1] \times [0, 1]. \end{cases}$$

Equivalent:

$$f(x, y) = \mathbb{1}_{0 \le x \le 1} \cdot \mathbb{1}_{0 \le y \le 1}$$

The joint density factors $\Rightarrow$ $X \perp\!\!\!\perp Y$

We can verify that marginally, X and Y are each Uniform$[0, 1]$

Example: (unit square cont.)

The area is $A = X \cdot Y$, so by independence,

$$\mathbb{E}(A) = \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

Example: (unit square cont.)

The area is $A = X \cdot Y$, so by independence,

$$\mathbb{E}(A) = \mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}.$$

The length is $L = \sqrt{X^2 + Y^2}$, so:

$$\mathbb{E}(L) = \iint \sqrt{x^2 + y^2} \cdot f(x, y)\, \mathrm{d}y\, \mathrm{d}x = \int_{x=0}^{1} \int_{y=0}^{1} \sqrt{x^2 + y^2}\, \mathrm{d}y\, \mathrm{d}x.$$