

## 4.2a Conditional distributions

Given a bivariate random vector  $(X, Y) \in \mathbb{R}^2$

we are interested in, say

$$P(Y \in B \mid X \in A), \text{ for } A, B \subset \mathbb{R}.$$

Examples

①  $(X, Y) = (\text{height}, \text{weight})$  of a (random) person

$$P(Y > 200 \text{ lb.} \mid X = 6') ?$$

②  $X =$  number people out of 100 sampled in a  
Iowa poll who said they'd vote for  
John Smith

$Y =$  John Smith wins Iowa on the election day

$$P(Y = \text{yes} \mid X = 53) ?$$

Let  $(X, Y)$  be discrete.

$$P(Y \in B | X \in A) = \frac{P(X \in A \cap Y \in B)}{P(X \in A)}$$

$$P(X \in A \cap Y \in B) = \sum_{x \in A} \sum_{y \in B} f_{XY}(x, y)$$

$$\begin{aligned} P(X \in A) &= \sum_{x \in A} \sum_y f_{XY}(x, y) \\ &= \sum_{x \in A} f_X(x) \end{aligned}$$

So

$$\begin{aligned} P(Y \in B | X \in A) &= \frac{\sum_{x \in A} \sum_{y \in B} f_{XY}(x, y)}{\sum_{x \in A} f_X(x)} \\ &= \sum_{y \in B} \frac{\sum_{x \in A} f_{XY}(x, y)}{\sum_{x \in A} f_X(x)} \end{aligned}$$

In particular, let  $A = \{x\}$ ,  $B = \{y\}$ , then

$$P(Y=y | X=x) = \frac{f_{XY}(x, y)}{f_X(x)} \rightarrow \text{conditional pmf}$$

**Def.** Let  $(X, Y)$  be a discrete Random vector with joint pmf  $f_{XY}(x, y)$

- Then, for any  $x$  such that  $f_X(x) = P(X=x) > 0$  the **conditional pmf** of  $Y$  given that  $X=x$  is a function of  $y$  denoted by  $f(y|x)$

and

$$f(y|x) = P(Y=y | X=x)$$

$$= \frac{f_{XY}(x, y)}{f_X(x)}$$

$$= \frac{f_{XY}(x, y)}{\sum_y f_{XY}(x, y)} .$$

Remark:

- $f(y|x) \geq 0 \quad \forall y$

- $\sum_y f(y|x) = \sum_y \frac{f_{XY}(x, y)}{\sum_y f_{XY}(x, y)}$

$$= \frac{\sum_y f_{XY}(x, y)}{\sum_y f_{XY}(x, y)} = 1.$$

- Sometimes, we use notation  $f_{Y|X}(y|x)$
- From definition, both the marginal pmf  $f_X, f_Y$  and the conditional pmf  $f_{Y|X}, f_{X|Y}$  are completely determined by the joint pmf.  $f_{XY}$

Moreover  $f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$

implies  $f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x)$

Similarly,  $f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y)$

whenever  $f_X(x) > 0$  and  $f_Y(y) > 0$

So the Marginal pmf and Conditional pmf are combined to completely determine the joint pmf

Let's consider how a continuous bivariate Random vector  $(X, Y) \in \mathbb{R}^2$  we'll develop a similar notion of Conditional pdf for the continuous case by mimicking the conditional pmf for the discrete case

Heuristic but incorrect Argument

$$P(Y \in B \mid X=x) = \frac{P(Y \in B, X=x)}{P(X=x)}$$

$$P(Y \in B, X=x) = \int_B f_{XY}(x, y) dy$$

$$P(X=x) = \int_{\mathbb{R}} f_{XY}(x, y) dy$$

Hence  $P(Y \in B \mid X=x) = \frac{\int_B f_{XY}(x, y) dy}{f_X(x)}$

$$= \int_B \underbrace{\frac{f_{XY}(x, y)}{f_X(x)}}_{\text{define this integrand to be the conditional p.d.f.}} dy \quad (*)$$

define this integrand to be  
the conditional p.d.f.

the "conditional" p.d.f.

The problem with above argument is that both  
 $P(X=x) = 0$  and  
 $P(Y \in B, X=x) \leq P(X=x) = 0.$

The correct treatment is to use limits:  
define

$$P(Y \in B | X=x) := \lim_{\varepsilon \rightarrow 0} P(Y \in B | X \in (x-\varepsilon, x+\varepsilon))$$

provided that such a limit exists and is unique

To assess the RHS and its limit, we need to use the joint pdf:

$$\begin{aligned} P(Y \in B | X \in (x-\varepsilon, x+\varepsilon)) &= \frac{P(Y \in B \cap X \in (x-\varepsilon, x+\varepsilon))}{P(X \in (x-\varepsilon, x+\varepsilon))} \\ &= \frac{\iint_{B \cap x-\varepsilon \leq x \leq x+\varepsilon} f_{XY}(x,y) dx dy}{\int_{x-\varepsilon}^{x+\varepsilon} f_X(x) dx} \end{aligned}$$

(by def)

Under suitably mild conditions (see "Tools")

$$\lim_{\varepsilon \rightarrow 0} \int_B \frac{\left( \int_{x-\varepsilon}^{x+\varepsilon} f_{xy}(x,y) dx \right)}{\left( \int_{x-\varepsilon}^{x+\varepsilon} f_x(x) dx \right)} dy$$

(interchanging limit and integral)

$$= \int_B \left( \lim_{\varepsilon \rightarrow 0} \frac{\int_{x-\varepsilon}^{x+\varepsilon} f_{xy}(x,y) dx}{\int_{x-\varepsilon}^{x+\varepsilon} f_x(x) dx} \right) dy$$

$$= \boxed{ \int_B \frac{f_{xy}(x,y)}{f_x(x)} dy }.$$

(if  $f_{xy}(x,y)$  continuous at  $(x, B)$ )

which is ( $\neq$ ) than we aimed for.

We'll take the integrand as the definition for the conditional pdf for the continuous setting.

Deg.

Let  $(X, Y)$  be a bivariate Random vector with joint pdf  $f_{XY}(x, y)$  (and marginal pdfs  $f_X(x)$  and  $f_Y(y)$ )

Then, for any  $x$  such that  $f_X(x) > 0$ , the conditional pdf of  $Y$  given  $X=x$  is

$$f_{Y|X}(y|x) := \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y) dy}$$

Remark.

- $f_{Y|X}(y|x) \geq 0 \quad \forall y, \forall x \quad (\text{where } f_X(x) > 0)$

- $\int f_{Y|X}(y|x) dy = 1$

- The following identities hold

$$f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x) \leftarrow \text{when } f_X(x) > 0$$

$$= f_Y(y) f_{(X|Y)}(x|y) \leftarrow \text{when } f_Y(y) > 0$$

## Conditional distributions

- Given a bivariate Random vector  $(X, Y)$  endowed with a joint pdf / pmf  $f_{XY}(x, y)$ , we have defined the conditional pdf / pmf  $f_{Y|X}(y|x)$  which defines a conditional distribution for  $Y$  given that  $X=x$ , for each  $x \in \mathbb{R}$ .
- By varying  $x$ , we actually have a collection of distributions for Random variable  $Y$  described by the collection of Conditional pdf / pmf on  $y$ .
$$\left\{ f_{Y|X}(y|x) \right\}_{x \in \mathbb{R}}$$
- This is what we mean when we speak of the Conditional distribution of a Random variable  $Y$  given a Random variable  $X$ .

## Conditional expectation

Given  $(X, Y) \sim f_{XY}$

$g : \mathbb{R} \rightarrow \mathbb{R}$  a function

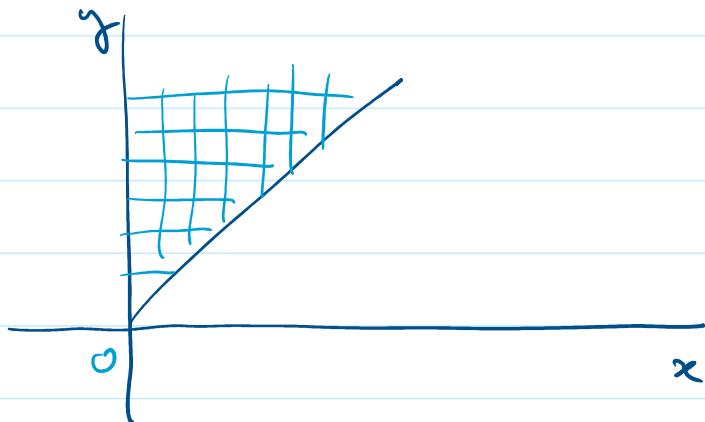
$$\mathbb{E}[g(Y) | X] := \mathbb{E}[g(Y) | X=x]$$

$$:= \int g(y) f_{Y|X}(y|x) dy$$

↑  
 $\sum$  if  $Y$  discrete

Example!

$$\text{Let } (X, Y) \sim f_{XY}(x, y) = \begin{cases} e^{-y} & , 0 < x < y < \infty \\ 0 & \text{otherwise} \end{cases}$$



- $f_{xy}(x,y) \geq 0 \quad \forall x,y \in \mathbb{R}^2$

- $\int_{\mathbb{R}^2} f_{xy}(x,y) dy dx$

$$\begin{aligned}
 &= \iint_{0 \times 0}^{\infty} e^{-y} dx dy = \int_0^{\infty} ye^{-y} dy \\
 &= \int_0^{\infty} (-y) de^{-y} \\
 &\stackrel{\text{ibp}}{=} \int_0^{\infty} e^{-y} dy = 1.
 \end{aligned}$$

- So this is a valid joint pdf.

- Marginal computation:  $f_x(x) = 0 \quad \forall x \leq 0$ .

for  $x > 0$

$$\begin{aligned}
 f_x(x) &= \int_{-\infty}^{\infty} f_{xy}(x,y) dy \\
 &= \int_x^{\infty} f_{xy}(x,y) dy \\
 &= \int_x^{\infty} e^{-y} dy \\
 &= -e^{-y} \Big|_x^{\infty} = e^{-x}.
 \end{aligned}$$

for  $y > 0$ :

$$\begin{aligned}
 f_Y(y) &= \int_{-\infty}^{\infty} f_{XY}(x,y) dx \\
 &= \int_0^y f_{XY}(x,y) dx \\
 &= \int_0^y e^{-y} dx = ye^{-y}.
 \end{aligned}$$

$$f_Y(y) = 0 \text{ if } y \leq 0.$$

. Conditional pdf: if  $y > x > 0$  then

$$\begin{aligned}
 f_{Y|X}(y|x) &= \frac{f_{XY}(x,y)}{f_X(x)} = \frac{e^{-y}}{e^{-x}} = e^{x-y}. \\
 &= 0 \text{ if } y < x.
 \end{aligned}$$

if  $x \in (0, y)$ :

$$\begin{aligned}
 f_{X|Y}(x|y) &= \frac{f_{XY}(x,y)}{f_Y(y)} \\
 &= \frac{e^{-y}}{ye^{-y}} = \frac{1}{y}.
 \end{aligned}$$

$$f_{X|Y}(x|y) = 0 \text{ otherwise.}$$

## • Conditional expectations

$$\begin{aligned} E[Y|X=x] &= \int y f_{Y|X}(y|x) dy \\ &= \int_x^{\infty} y e^{x-y} dy \\ &= e^x \int_x^{\infty} y e^{-y} dy \\ &= x+1. \end{aligned} \quad (\text{check!})$$

$$\begin{aligned} E[Y^2|X=x] &= \int_x^{\infty} y^2 e^{x-y} dy \\ &= e^x \int_x^{\infty} y^2 e^{-y} dy \\ &= x^2 + 2x + 2 \end{aligned} \quad (\text{check!})$$

$$\begin{aligned} \text{Hence } \text{var}(Y|X=x) &= x^2 + 2x + 2 - (x+1)^2 \\ &= 1. \end{aligned}$$

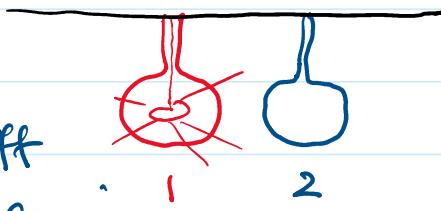
## Example 2

Given two light bulbs with lifetimes  $X, Z \geq 0$

$$X, Z \stackrel{iid}{\sim} \text{Exp}(\lambda)$$

Light bulb 1 is on, 2 is off

Light bulb 2 is turned on as soon as 1 burns out.



Let  $Y = X + Z$  be the time bulb 2 burns out

$$\begin{aligned} P(Y \leq y \mid X = x) &= P(Z \leq y - x \mid X = x) \\ &= P(Z \leq y - x) \quad \text{since } Z \perp\!\!\!\perp X \\ &= F_Z(y - x) = 1 - e^{-\lambda(y-x)} \end{aligned}$$

if  
 $y > x$

Hence

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{d}{dy} (1 - e^{-\lambda(y-x)}) \\ &= \lambda e^{-\lambda(y-x)} \end{aligned}$$

In particular, if  $\lambda = 1$  then this gives the same conditional pdf as the previous example

Example

with

$$f_X(x) = \lambda e^{-\lambda x} \mathbf{1}(x > 0)$$

and

$$f_{Y|X}(y|x) = \begin{cases} \lambda e^{-\lambda(y-x)} & , y > x \\ 0 & \text{otherwise} \end{cases}$$

The joint pdf for  $(X, Y)$  is completely determined via

$$\begin{aligned} f_{XY}(x,y) &= f_X(x) f_{Y|X}(y|x) \\ &= \lambda^2 e^{-\lambda y} \quad \text{if } y > x > 0 \end{aligned}$$

$$\text{and } f_{XY}(x,y) = 0 \quad \text{otherwise}$$

**Remark:**

There are two equivalent ways to define a joint distribution for a bivariate Random vector  $(X, Y)$ :

① By specifying joint pdf  $f_{XY}(x,y)$

② By specifying pdf for  $X$ ,  $f_X(x)$  and the conditional pdf of  $Y$  given  $X$ ,  $f_{Y|X}(y|x)$  and set

$$f_{XY}(x,y) := \begin{cases} f_X(x) f_{Y|X}(y|x) & \text{if } f_X(x) > 0 \\ 0 & \text{o.w.} \end{cases}$$

the pdf  $f_X$  then becomes the marginal pdf with respect to the joint  $f_{X,Y}$ , i.e. the identity

$$\int_y f_{XY}(x,y) dy = f_X(x)$$

holds.

② is usually "easier" in applications