Conditional Distributions

- X, Y discrete: the **conditional pmf** of X given $Y = y$ is defined to be

$$p_{X|Y}(x|y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p(x, y)}{p_Y(y)}, \quad p_Y(y) > 0.$$

- Given $Y = y$, the randomness of X is described by $p(x, y)$ but $p(x, y)$ is NOT a pmf wrt[a] x since $\sum_{\text{all } x} p(x, y) \neq 1$. We need this normalizing constant $p_Y(y)$ to make it a valid pmf.

[a] wrt = with respect to

- X, Y continuous: the **conditional pdf** of X given $Y = y$ is defined to be

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad f_Y(y) > 0.$$

- Given $Y = y$, $f(x,y)$ is NOT a pdf wrt x, since $\int f(x,y)dx = f_Y(y) \neq 1$. So we need $f_Y(y)$ in the denominator to make it a legit pdf.

  Check the Wolfram Demo.

- If X and Y independent,

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{f_X(x)f_Y(y)}{f_Y(y)} \quad \text{due to independence} \\ &= f_X(x) \end{aligned}$$

$$p_{X|Y}(x|y) = p_X(x)$$

- In general, X and Y are dependent and then $f_{X|Y}(x|y) \neq f_X(x)$: Given the extra information that $Y = y$, the distribution of X is no longer the same as the marginal $f_X(x)$.

Example 1 (2.1.1 on p.74, Revisit)

X / Y	0	1	2	3	$p_X(x)$
0	$1/8$	$1/8$	0	0	$2/8$
1	0	$2/8$	$2/8$	0	$4/8$
2	0	0	$1/8$	$1/8$	$2/8$
$p_Y(y)$	$1/8$	$3/8$	$3/8$	$1/8$	

a)] Find the conditional pmf $p_{Y|X}(y|x)$, conditional expectation $\mathbb{E}(Y|X=x)$ and conditional variance Var$(Y \mid X = x)$.

	0	1	2	3	$\mathbb{E}(Y\|X=x)$	Var$(Y\|X=x)$
$p_{Y\|X}(y\|0)$	0.5	0.5	0	0	0.5	1/4
$p_{Y\|X}(y\|1)$	0	0.5	0.5	0	1.5	1/4
$p_{Y\|X}(y\|2)$	0	0	0.5	0.5	2.5	1/4

Conditional Expectations

$$\mathbb{E}(X|Y=y) \quad = \quad \sum_x x \cdot \mathbb{P}(X=x|Y=y) = \sum_x x \cdot p_{X|Y}(x|y).$$

$$\mathbb{E}(X|Y=y) \quad = \quad \int x \cdot f_{X|Y}(x|y)dx$$

What does the symbol $\mathbb{E}(X \mid Y)$ mean? You can view it as a function of Y, i.e., $\mathbb{E}(X \mid Y) = g(Y)$ with its value at $Y = y$ given by

$$g(y) = \mathbb{E}(X \mid Y = y).$$

Therefore $\mathbb{E}(X \mid Y)$ is a random variable. We can talk about its distribution (HW1, p7) and compute its mean and variance.

Example 1 (2.1.1 on p.74, Revisit)

$\mathbb{E}(Y|X)$ is a r.v., which equals $\mathbb{E}(Y|X=x)$ with probability $p_X(x)$.

That is,

$$\begin{aligned} \mathbb{E}(Y|X) &= 0.5 \quad \text{with prob} \quad p_X(0) = 1/4, \\ \mathbb{E}(Y|X) &= 1.5 \quad \text{with prob} \quad p_X(1) = 1/2, \\ \mathbb{E}(Y|X) &= 2.5 \quad \text{with prob} \quad p_X(2) = 1/4. \end{aligned}$$

What's the expectation of the r.v. $\mathbb{E}(Y|X)$?

$$\mathbb{E}\big[\mathbb{E}(Y|X)\big] = (0.5)\frac{1}{4}+(1.5)\frac{1}{2}+(2.5)\frac{1}{4} = \frac{0.5+(1.5)(2)+2.5}{4} = \frac{6}{4} = 1.5$$

which is the same as

$$\mathbb{E}Y = (0)\frac{1}{8} + (1)\frac{3}{8} + (2)\frac{3}{8} + (3)\frac{1}{8} = \frac{12}{8} = 1.5.$$

This is true for any joint dist, $\mathbb{E}[\mathbb{E}(Y|X)] = \mathbb{E}Y$, due to the iterative rule for expectation.

Iterative Rule $\mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \mathbb{E}X$[a].

$$\mathbb{E}\big[\mathbb{E}(X \mid Y)\big] = \int \Big[\mathbb{E}(X \mid Y = y)\Big] f_Y(y) dy$$

$$= \int \left[\int x \cdot f_{X|Y}(x|y) dx\right] f_Y(y) dy = \int \left[\int x \cdot \frac{f(x,y)}{f_Y(y)} dx\right] f_Y(y) dy$$

$$= \iint x \cdot \frac{f(x,y)}{f_Y(y)} f_Y(y) dx dy = \iint x \cdot f(x,y) dx dy$$

$$= \iint x \cdot f(x,y) dy dx = \int x \left[\int f(x,y) dy\right] dx$$

$$= \int x f_X(x) dx = \mathbb{E}X.$$

Similarly we have $\mathbb{E}\big[\mathbb{E}(g(X)|Y)\big] = \mathbb{E}[g(X)]$.

---

[a] What's more useful is $\mathbb{E}_X\left[X\right] = \mathbb{E}_Y\left[\mathbb{E}(X \mid Y)\right]$.

- Sometimes, I'll write the conditional expectation $\mathbb{E}[\cdot \mid Y]$ as $\mathbb{E}_{X|Y}[\cdot]$ especially when $[\cdot]$ has a lengthy expression, where $\mathbb{E}_{X|Y}$ just means that taking expectation of X with respect to the conditional distribution of X given Y[a].

- I also use notations like $\mathbb{E}_Y$ in the slides, to remind you that this expectation is over Y only, wrt the marginal distribution $f_Y(y)$. Similarly, $\mathbb{E}_X$ refers to the expectation over X wrt $f_X(x)$

- Usually the meaning of expectation is clear from the context, e.g., $\mathbb{E}g(X)$ must be $\mathbb{E}_X g(X)$, so you don't need to write subscripts in your homework/exam.

[a]Note that $\mathbb{E}_{X|Y}$ would only average over X but treat Y as a constant.

The general **Iterative Rule**

$$\mathbb{E}g(X,Y) = \mathbb{E}_Y \mathbb{E}_{X|Y}\left[g(X,Y)\right]$$

$$\text{LHS} = \mathbb{E}g(X,Y) = \iint f(x,y)g(x,y)dxdy$$

$$\begin{aligned} \text{RHS} &= \mathbb{E}_Y\left[\mathbb{E}_{X|Y}g(X,Y)\right] \\ &= \int f_Y(y)\left[\int f_{X|Y}(x|y)g(x,y)dx\right]dy \\ &= \iint f_Y(y)\frac{f(x,y)}{f_Y(y)}g(x,y)dxdy \\ &= \iint f(x,y)g(x,y)dxdy \end{aligned}$$

This is essentially the same as the Chain Rule of probability.

Useful Properties

- Linearity

$$\mathbb{E}(aX_1 + bX_2|Y) = a\mathbb{E}(X_1|Y) + b\mathbb{E}(X_2|Y)$$

- Take constants outside an expectation

$$\mathbb{E}[g(Y)|Y] = g(Y), \quad \mathbb{E}[g(Y)X|Y] = g(Y)\mathbb{E}(X|Y)$$

  In particular, $\mathbb{E}[\mathbb{E}(X|Y)|Y] = \mathbb{E}(X|Y)$

- Iterative rule

$$\mathbb{E}_Y\left[\mathbb{E}(X|Y)\right] = \mathbb{E}(X), \quad \mathbb{E}_Y\left[\mathbb{E}(g(X)|Y)\right] = \mathbb{E}[g(X)]$$

- $\mathbb{E}\Big[\big(X - \mathbb{E}(X|Y)\big)g(Y)\Big] = 0$

$$\begin{aligned}\mathbb{E}_Y\Big[\mathbb{E}_{X|Y}\big(X - \mathbb{E}(X|Y)\big)g(Y)\Big] &= \mathbb{E}_Y g(Y)\Big[\mathbb{E}_{X|Y}\big(X - \mathbb{E}(X|Y)\big)\Big] \\ &= \mathbb{E}_Y g(Y)\Big[\mathbb{E}(X|Y) - \mathbb{E}(X|Y)\Big]\end{aligned}$$

Similarly, we can define conditional variance $\text{Var}(X|Y)$ that is the variance of r.v. $g(Y) = \mathbb{E}(X \mid Y)$ (check the calculation for Example 1).

$$
\begin{aligned}
\text{Var}(X|Y) &= \mathbb{E}\Big[\big(X - \mathbb{E}(X|Y)\big)^2 |Y\Big] \\
&= \mathbb{E}\Big[X^2 - 2X \cdot \mathbb{E}(X|Y) + \mathbb{E}(X|Y)^2 |Y\Big] \\
&= \mathbb{E}\big[X^2|Y\big] - 2\mathbb{E}\big[X \cdot \mathbb{E}(X|Y)|Y\big] + \mathbb{E}\big[\mathbb{E}(X|Y)^2|Y\big] \\
&= \mathbb{E}\big[X^2|Y\big] - 2\mathbb{E}(X|Y)\mathbb{E}(X|Y) + \mathbb{E}(X|Y)^2 \\
&= \mathbb{E}(X^2|Y) - [\mathbb{E}(X|Y)]^2 \\
&= \text{(Conditional 2nd Moment)} - \text{(Conditional Mean)}^2.
\end{aligned}
$$

Note that (shown on the next slide)

$$\text{Var}(X) = \mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y))$$

$$
\begin{aligned}
\mathsf{Var}(X) &= \mathbb{E}(X-\mu_X)^2 = \mathbb{E}(X-\mathbb{E}(X|Y)+\mathbb{E}(X|Y)-\mu_X)^2 \\
&= \mathbb{E}\big[X-\mathbb{E}(X|Y)\big]^2 + \mathbb{E}\big[\mathbb{E}(X|Y)-\mu_X\big]^2 \\
&\quad +2\mathbb{E}\big[(X-\mathbb{E}(X|Y)(\mathbb{E}(X|Y)-\mu_X)\big] \\
&= \mathbb{E}_Y\left\{\mathbb{E}_{X|Y}\left[X-\mathbb{E}(X|Y)\right]^2\right\} + \mathbb{E}_Y\mathbb{E}_{X|Y}\left[\mathbb{E}(X|Y)-\mu_X\right]^2 \\
&\quad +2\mathbb{E}_Y\left\{\mathbb{E}_{X|Y}\left[(X-\mathbb{E}(X|Y)(\mathbb{E}(X|Y)-\mu_X)\right]\right\} \\
&= \mathbb{E}(\mathsf{Var}(X|Y)) + \mathsf{Var}(\mathbb{E}(X|Y))
\end{aligned}
$$

You may get confused with the expression $\mathbb{E}(X|Y)$ on the previous slide. Let's go through the proof again with notation $g(Y) = \mathbb{E}(X|Y)$ and note that $\mathbb{E}g(Y) = \mathbb{E}X = \mu_X$.

$$
\begin{aligned}
\text{Var}(X) &= \mathbb{E}(X - \mu_X)^2 = \mathbb{E}(X - g(Y) + g(Y) - \mu_X)^2 \\
&= \mathbb{E}_{X,Y}\left[X - g(Y)\right]^2 + \mathbb{E}_Y\left[g(Y) - \mu_X\right]^2 \\
&\quad + 2\mathbb{E}_{X,Y}\left[(X - g(Y))(g(Y) - \mu_X)\right] \\
&= \mathbb{E}_Y\left\{\mathbb{E}_{X|Y}\left[X - g(Y)\right]^2\right\} + \mathbb{E}\,\left[g(Y) - \mu_X\right]^2 \\
&\quad + 2\mathbb{E}_Y\left\{\mathbb{E}_{X|Y}\left[(X - g(Y))(g(Y) - \mu_X)\right]\right\} \\
&= \mathbb{E}(\text{Var}(X|Y)) + \text{Var}(\mathbb{E}(X|Y))
\end{aligned}
$$

How to understand

$$\mathsf{Var}(X) = \mathbb{E}(\mathsf{Var}(X|Y)) + \mathsf{Var}(\mathbb{E}(X|Y))$$

Let X denote the height of a randomly chosen student from stat410.

Suppose students can be divided into several sub-populations (r.v. Y).

The height (r.v. X) variation comes from two sources:

- Variation within each sub-population (variation of X given Y)
- Variation among the mean height for each sub-population (variation of $\mathbb{E}(X \mid Y)$)

The total variation is the sum of these two.

Example 2: The joint pdf is

$$f(x, y) = 60x^2 y, \quad 0 \leq x, y \leq 1, \; x + y \leq 1, \qquad \text{zero, elsewhere.}$$

(`JointDistributions.pdf`, `ConditionalDistributions.pdf`)

We have computed the marginal pdf

$$\begin{aligned} f_X(x) &= 30x^2(1-x)^2, \quad 0 < x < 1, \qquad \mathbb{E}(X) = \frac{1}{2}, \\ f_Y(y) &= 20y(1-y)^3, \quad 0 < y < 1, \qquad \mathbb{E}(Y) = \frac{1}{3}. \end{aligned}$$

a) Find the conditional pdf $f_{X|Y}(x|y)$ of X given $Y = y$, $0 < y < 1$.

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{3x^2}{(1-y)^3}, \quad 0 < x < 1 - y.$$

Check $f_{X|Y}(x|y)$ is a valid pdf wrt x: apparently $f_{X|Y}(x|y) \geq 0$,

$$\int f_{X|Y}(x|y) dx = \int_0^{1-y} 3x^2/(1-y)^3 dx = 1.$$

b) Find $\mathbb{P}(X > \frac{1}{2} | Y = \frac{1}{3})$.

Calculate this conditional probability using conditional pdf:

$$f_{X|Y}(x \mid 1/3) = \frac{3x^2}{(1-1/3)^3} = \frac{3x^2}{(2/3)^3}, \quad 0 < x < 2/3.$$

$$\begin{aligned} \mathbb{P}\left(X > \frac{1}{2} \middle| Y = \frac{1}{3}\right) &= \int_{1/2}^{2/3} \frac{3x^2}{(2/3)^3} dx = \frac{x^3}{(2/3)^3}\Big|_{1/2}^{2/3} \\ &= 1 - \frac{27}{64} = \frac{37}{64}. \end{aligned}$$

How to use conditional pmf/pdf to evaluate $\mathbb{P}(a < X < b \mid Y = y)$?

- For discrete rvs, you can either use conditional pmf

$$\sum_{a<x<b} p_{X|Y}(x \mid y) = \sum_{a<x<b} \frac{p(x,y)}{p_Y(y)} = \frac{\sum_{a<x<b} p(x,y)}{p_Y(y)},$$

or just follow the definition of conditional probability

$$\frac{\mathbb{P}(a < X < b, Y = y)}{\mathbb{P}(Y = y)} = \frac{\sum_{a<x<b} p(x,y)}{\sum_{\text{all } x} p(x,y)}.$$

- For **continuous** rvs, we CANNOT evaluate this probability via $\mathbb{P}(a < X < b, Y = y)/\mathbb{P}(Y = y)$ as in the discrete case, since $\mathbb{P}(Y = y) = 0$, instead we need to use conditional pdf

$$\mathbb{P}(a < X < b \mid Y = y) = \int_a^b f_{X|Y}(x|y)dx.$$

Go through other examples from `ConditionalDistributions.pdf`