3 Distribution Theory for Pairs of Random Variables

Many interesting probabilistic statements about a pair X, Y of random variables concern the way X and Y vary together as functions of the domain $\Omega_X \times \Omega_Y$, which is a Cartesian Product. This lecture will introduce joint distribution functions, conditional distributions, joint and conditional expectation and independence of two random variables. Most of results for two random variables can be generalized for n random variables. Also we will see how to find the distributions of functions of random variables by using transformation techniques.

3.1 Joint distribution functions and Statistical independence

3.1.1 Joint distribution functions

Definition 1 *The joint distribution function of X and Y is the function*

$$F_{(X,Y)} : \mathbb{R}^2 \longrightarrow [0,1]$$

given by

$$F_{(X,Y)}(x,y) = \mathbb{P}(X \leq x \textit{ and } Y \leq y).$$

Definition 2 *If X and Y are discrete then their joint mass function $p_{(X,Y)}$: $\mathbb{R}^2 \to [0,1]$ is given by*

$$p_{(X,Y)}(x,y) = \mathbb{P}(X = x, Y = y).$$

If X and Y are continuous then we cannot talk of their point mass function since it is identically zero. Instead we need another density function.

Definition 3 *X and Y are (jointly) continuous with (joint) probability density function*

$$f_{(X,Y)} : \mathbb{R}^2 \longrightarrow [0, \infty)$$

if

$$F_{(X,Y)}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(u, v) du dv$$

for each $x, y \in \mathbb{R}$.

Remark:

1. If $F_{(X,Y)}$ is sufficiently differentiable at the point (x, y), then we normally specify

$$f_{(X,Y)}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x, y).$$

2.Probabilities.

$$\mathbb{P}(a \leq X \leq b, c \leq Y \leq d) = \int_{c}^{d} \int_{a}^{b} f_{(X,Y)}(x, y) dx dy.$$

The properties of the joint distributions and mass functions and density functions are very much the same as those of the corresponding functions of a single variable.

3.1.2 Marginal distributions

If $F_{(X,Y)}(x, y)$ is the joint distribution of X and Y, then the distributions $F_X(x)$ and $F_Y(y)$ are called *marginal distribution functions.* Given the joint distribution function of X and Y, the marginal distribution functions can be found in the following way:

$$\begin{aligned} F_X(x) &= \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, -\infty < Y < \infty) \\ &= \lim_{y \to \infty} F_{(X,Y)}(x, y) \end{aligned}$$

and

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \le y) = \mathbb{P}(-\infty < X < \infty, Y \le y) \\ &= \lim_{x \to \infty} F_{(X,Y)}(x,y). \end{aligned}$$

Applying this idea to discrete and continuous random variables, we have the following results:

1. If X and Y are discrete random variables with *joint mass function* $p_{(X,Y)}(\cdot,\cdot)$, then the *marginal mass functions* $p_X(\cdot)$ and $p_Y(\cdot)$ are

$$\begin{aligned} p_X(x) &= \mathbb{P}(X = x) = \mathbb{P}(\cup_y(\{X = x\} \cap \{Y = y\})) \\ &= \sum_y \mathbb{P}(X = x, Y = y) = \sum_y p_{(X,Y)}(x,y), \end{aligned}$$

and similarly $p_Y(y) = \sum_x p_{(X,Y)}(x,y)$.

2. If X and Y are continuous random variables with joint density function $f_{(X,Y)}(x,y)$, then

$$F_X(x) = \mathbb{P}(X \le x, -\infty < Y < \infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{\infty} f_{(X,Y)}(u,y)dy \right) du.$$

It follows that the *marginal density function* of X is

$$f_X(x) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dy.$$

Similarly the *marginal density function* of Y is

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx.$$

Example 1 *Let X and Y be random variables with joint pdf*

$$f_{(X,Y)}(x,y) = \frac{1}{y} e^{-y-\frac{x}{y}} \qquad \text{for } 0 < x, y < \infty.$$

What is the marginal distribution of Y?

$$f_Y(y) = \int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dx = \int_0^{\infty} \frac{1}{y} e^{-y-\frac{x}{y}} dx = e^{-y}, \quad y > 0,$$

and hence Y is exponentially distributed.

3.1.3 Statistical Independence

Definition 4 ***Independence.*** *X and Y are independent if $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events for all $x, y \in \mathbb{R}$. In other words, X and Y are independent if*

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

More generally, X and Y are independent if $F_{X,Y}(x, y)$ can be factorised as the product $s(x)t(y)$ of a function of x alone and a function of y alone, for all $x, y \in \mathbb{R}$.

NB: we stress that the factorization in the definition must hold for all x and y in order that X and Y be independent.

The following general result holds for the independence of functions of random variables. We will state it without proof since it is beyond the scope of this course.

Theorem 1 *Let X and Y be independent random variables (of any type) and let $g, h : \mathbb{R} \longrightarrow \mathbb{R}$ be functions. Then $g(X)$ and $h(Y)$ are also independent.*

From the definition of independence, we have particular results to check independence for discrete and continuous random variables.

Lemma 1 *Let X and Y be discrete random variables with joint probability mass function $p_{(X,Y)}(\cdot, \cdot)$. Then X and Y are independent if and only if*

$$p_{(X,Y)}(x, y) = p_X(x)p_Y(y) \quad for\ all\ x\ and\ y.$$

Lemma 2 *Let X and Y be continuous random variables with joint probability density function $f_{X,Y}(\cdot, \cdot)$. Then X and Y are independent if and only if*

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) \quad for\ all\ x\ and\ y.$$

Example 2 *The number of births during a day at a hospital is assumed to be Poisson with parameter λ. Each birth is a boy with probability p or a girl with probability*

$q = 1 - p$, independently of other births. Let B be the number of boys, G the number of girls, born during the day. Prove that

$$\mathbb{P}(B = x, G = y) = \frac{e^{-\lambda p}(\lambda p)^x}{x!}\,\frac{e^{-\lambda q}(\lambda q)^y}{y!}$$

and deduce that

1. *B has the Poisson distribution with parameter λp;*
2. *G has the Poisson distribution with parameter λq;*
3. *B and G are independent.*

Let N be the number of births, where N has a Poisson distribution with parameter λ.

$$\begin{aligned}\mathbb{P}(B = x, G = y) &= \mathbb{P}(B = x, G = y, N = x + y)\\ &= \mathbb{P}(B = x, G = y | N = x + y)\mathbb{P}(N = x + y)\\ &= \binom{x+y}{x} p^x(1-p)^y \frac{\lambda^{x+y}e^{-\lambda}}{(x+y)!} = \frac{e^{-\lambda p}(\lambda p)^x}{x!}\,\frac{e^{-\lambda q}(\lambda q)^y}{y!}.\end{aligned}$$

$$\begin{aligned}f_B(x) &= \sum_{y=0}^{\infty} \mathbb{P}(B = x, G = y)\\ &= \sum_{y=0}^{\infty} \frac{e^{-\lambda p}(\lambda p)^x}{x!}\,\frac{e^{-\lambda q}(\lambda q)^y}{y!}\\ &= \frac{e^{-\lambda p}(\lambda p)^x}{x!} e^{-\lambda q} \sum_{y=0}^{\infty} \frac{(\lambda q)^y}{y!}\\ &= \frac{e^{-\lambda p}(\lambda p)^x}{x!} e^{-\lambda q} e^{\lambda q} = \frac{e^{-\lambda p}(\lambda p)^x}{x!}.\end{aligned}$$

A similar result holds for G, and so

$$\mathbb{P}(B = x, G = y) = \mathbb{P}(B = x)\mathbb{P}(G = y).$$

3.1.4 Covariance and Correlation coefficient

Lemma 3 ***Law of the unconscious statistician - 2 variables***

If $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ is a sufficiently nice function then

$$\mathbb{E}[g(X,Y)] = \sum_{(x,y)} g(x,y) p_{(X,Y)}(x,y)$$

when X and Y are discrete and

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{(X,Y)}(x,y) dx dy$$

when X and Y are continuous.

Definition 5 ***Covariance.***

The covariance of the random variables X and Y is

$$\mathrm{cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])].$$

Remark: Note that the concept of covariance generalizes that of variance in that $\mathrm{cov}(X,X) = \mathrm{var}(X)$. Expanding the covariance gives

$$\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

Theorem 2 ***Cauchy-Schwarz inequality.***

Let X and Y be random variables, then

$$(\mathbb{E}[XY])^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

with equality if and only if $\mathbb{P}(Y = t_0 X) = 1$, where t_0 is some constant.

Proof: for any real number t, let $u(t) = \mathbb{E}[(tX - Y)^2]$. It is easy to see that $u(t) = t^2\mathbb{E}[X^2] - 2t\mathbb{E}[XY] + \mathbb{E}[Y^2] \geq 0$ for $t \in \mathbb{R}$, so, viewed as a quadratic in t, the discriminant must be non-positive, i.e.,

$$(\mathbb{E}[XY])^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2] \leq 0.$$

Furthermore, $u(t) = 0$ for some $t = t_0$, say, if and only if

$$(\mathbb{E}[XY])^2 - \mathbb{E}[X^2]\mathbb{E}[Y^2] = 0.$$

It is also the case that $\mathbb{E}[(t_0X - Y)^2] = 0$, provided that $Y = t_0X$ almost everywhere.

Definition 6 ***Correlation coefficient.***

The correlation coefficient of X *and* Y *is*

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}}.$$

It is easy to see that $-1 \leq \rho \leq 1$ by applying the Cauchy-Schwarz inequality to $\widetilde{X} = X - E[X]$ and $\widetilde{Y} = Y - E[Y]$.

Remark: Both the covariance and the correlation coefficient of random variables X and Y are measures of a *linear relationship* between X and Y in the following sense: $\text{cov}(X,Y)$ will be positive when $X-\mathbb{E}[X]$ and $Y-\mathbb{E}[Y]$ tend to have the same sign with high probability, and $\text{cov}(X,Y)$ will be negative when $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ tend to have opposite signs with high probability. $\text{cov}(X,Y)$ tends to measure the linear relationship of X and Y; however, its actual magnitude does not have much meaning since it depends on the variability of X and Y. The correlation coefficient removes, in a sense, the individual variability of each X and Y by dividing the covariance by the product of the standard deviations, and thus the correlation coefficient is a better measure of the linear relationship of X and Y than the covariance. Also the correlation coefficient is unitless.

Definition 7 ***Uncorrelated.***

The random variables X *and* Y *are uncorrelated if* $\text{cov}(X,Y) = 0$.

$\text{Cov}(X,Y) = E[XY] - E[X]E[Y]$

Lemma 4 *If* X *and* Y *are uncorrelated then* $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Lemma 5 *If* X *and* Y *are independent then they are uncorrelated.*

We have previously discussed the relationship between independence and correlation. We have the result that independent random variables are indeed uncorrelated but that, in general, the converse is not true. Below we look at an example where we have two random variables where no correlation does imply independence (and obviously independence implies no correlation).

Example 3 ***Bivariate normal distribution.***

Let $f_{(X,Y)} : \mathbb{R}^2 \longrightarrow \mathbb{R}$ *be given by*

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right) \tag{1}$$

where $\rho \in (-1,1)$.

Check that $f_{(X,Y)}(\cdot,\cdot)$ *is a density function by verifying that*

$$f_{(X,Y)}(x,y) \geq 0 \ \text{ and } \ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{(X,Y)}(x,y)dxdy = 1.$$

$f_{(X,Y)}(\cdot,\cdot)$ *is called the standard bivariate normal density function of some pair* X *and* Y*. Calculation of its marginals shows that* X *and* Y *are* $N(0,1)$ *variables. Furthermore, the covariance*

$$\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

is given by

$$\text{cov}(X,Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf_{(X,Y)}(x,y)dxdy = \rho.$$

If X *and* Y *are uncorrelated, which means* $\rho = 0$*, then*

$$f_{(X,Y)}(x,y) = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}\right)\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}y^2}\right) = f_X(x)f_Y(y)$$

for all x *and* y*, and so* X *and* Y *are independent.*

Example 4 ***Uncorrelated but not independent.***

Suppose that X takes values $-2, -1, 1, 2$, each with equal probability of $1/4$.

$$E[X] = (-2 \times 1/4) + (-1 \times 1/4) + (1 \times 1/4) + (2 \times 1/4) = 0.$$

Let $Y = |X|$; Y takes values $1, 2$ each with equal probability $1/2$. Hence

$$E[Y] = (1 \times 1/2) + (2 \times 1/2) = 1\frac{1}{2}.$$

X	*Y*	*XY*
-2	*2*	*-4*
-1	*1*	*-1*
1	*1*	*1*
2	*2*	*4*

Here, the values taken by XY are each taken with probability $1/4$. So

$$E[XY] = (-4 \times 1/4) + (-1 \times 1/4) + (1 \times 1/4) + (4 \times 1/4) = 0,$$

and

$$\text{cov}(X, Y) = 0 - 0 \times 1\frac{1}{2} = 0.$$

3.2 Conditional Distributions and Conditional expectations

3.2.1 Conditional distributions

In Chapter 1, we discussed the conditional probability $\mathbb{P}(B|A)$. Similarly, we can also define the conditional distribution of one variable Y given the value of another variable X.

Definition 8 *If X and Y are discrete, the conditional distribution function of Y given $X = x$, written $F_{Y|X}(\cdot|x)$, is defined by*

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x)$$

for any x such that $\mathbb{P}(X = x) > 0$. The conditional (probability) mass function of Y given $X = x$, written $p_{Y|X}(\cdot|x)$, is defined by

$$p_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$$

for any x such that $\mathbb{P}(X = x) > 0$.

NB: Conditional distributions and mass functions are undefined at values of x for which $\mathbb{P}(X = x) = 0$.

In the discrete case we had that the probability

$$\mathbb{P}(Y \leq y | X = x) = \frac{\mathbb{P}(Y \leq y \text{ and } X = x)}{\mathbb{P}(X = x)}.$$

For this expression to make sense we must require that $\mathbb{P}(X = x) > 0$. Hence if we wanted to define conditional densities for continuous random variables in a similar way we would not be able to do so because $\mathbb{P}(X = x) = 0$ if X is a continuous random variable. One way to get around this problem is to think about a very 'small' rectangle placed at x and instead of looking at $\mathbb{P}(Y \leq y | X = x)$ we look at $\mathbb{P}(Y \leq y | x \leq X \leq x + \Delta x)$ where Δx is the base of the rectangle and we have that $\mathbb{P}(x \leq X \leq x + \Delta x) \neq 0$. Now we may proceed heuristically and define the conditional probability as

$$\begin{aligned}
\mathbb{P}(Y \leq y | x \leq X \leq x + \Delta x) &= \frac{\mathbb{P}\left(Y \leq y \text{ and } x \leq X \leq x + \Delta x\right)}{\mathbb{P}(x \leq X \leq x + \Delta x)} \\
&= \frac{\int_{-\infty}^{y} \int_{x}^{x+\Delta x} f_{(X,Y)}(u, v) du dv}{\int_{x}^{x+\Delta x} f_X(u) du} \\
&\simeq \frac{\int_{v=-\infty}^{y} f_{(X,Y)}(x, v) \Delta x dv}{f_X(x) \Delta x} \\
&= \int_{v=-\infty}^{y} \frac{f_{(X,Y)}(x, v) dv}{f_X(x)}.
\end{aligned}$$

Note that we start using an equality sign $=$ which we then later replace by an approximation $\simeq$ for Δx 'small'. But what we need is to make Δx *arbitrarily* small so we let $\Delta x \longrightarrow 0$. Note that this is possible since we have a ratio where the term Δx appears in both the numerator and the denominator, hence the limit exists as $\Delta x \to 0$.

Although the above argument is heuristic we have motivated the following definition.

Definition 9 ***Conditional distribution function.***

Let X and Y be continuous random variables. The conditional distribution function of the random variable Y given $X = x$ is the function $F_{Y|X}(\cdot|x)$ given by

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} \frac{f_{(X,Y)}(x,v)}{f_X(x)} dv \tag{2}$$

for any x such that $f_X(x) > 0$.

And we can also derive the conditional density function from the conditional distribution function.

Lemma 6 ***Conditional density function.***

The conditional density function of Y given X is the function $f_{Y|X}$ given by

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} \tag{3}$$

for any x such that $f_X(x) > 0$.

Example 5 *Let X and Y have joint density function*

$$f_{(X,Y)}(x,y) = \frac{1}{x} \qquad \text{for } 0 < y \le x \le 1.$$

Show that $f_X(x) = 1$ if $x \in (0,1]$ and that $f_{Y|X}(y|x) = 1/x$ if $0 < y \le x \le 1$.

When $0 < x \le 1$, we have

$$f_X(x) = \int_0^x f_{(X,Y)}(x,y)dy = \int_0^x \frac{1}{x} dy = 1.$$

When $0 < y \le x \le 1$, we have

$$f_{Y|X}(y|x) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} = \frac{1}{x}.$$

So X is 'uniformly' distributed on $(0,1]$ and, conditional on the event $\{X = x\}$, Y is 'uniform' on $(0,x]$.

3.2.2 Conditional Expectation

We can also extend the concept of conditionality to expected values. We may be interested in the expected value of a random variable knowing some information. For example, what is the life expectancy of men and women that live in Africa or South America? Or what is the expected growth of the UK economy knowing that the US economy is slowing down.

We consider discrete random variables first. Suppose we are told that $X = x$. Conditional upon this, the new distribution of Y has mass function $p_{Y|X}(y|x)$, which we think of as a function of y. The expected value of this distribution, $\sum_y y p_{Y|X}(y|x)$, is called the *conditional expectation* of Y given $X = x$ and is written as $\psi(x) = \mathbb{E}[Y|X = x]$. Now, we observe that the conditional expectation depends on the value x taken by X, and can be thought of as a function $\psi(X)$ of X itself.

Definition 10 ***Conditional expectation.*** *Suppose X and Y are discrete random variables. Let $\psi(x) = \mathbb{E}[Y|X = x]$. Then $\psi(X)$ is called the conditional expectation of Y given X, written as $\mathbb{E}[Y|X]$.*

NB: Although 'conditional expectation' sounds like a number, it is actually a random variable.

Theorem 3 *The conditional expectation $\psi(X) = \mathbb{E}[Y|X]$ satisfies*

$$\mathbb{E}[\psi(X)] = \mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y].$$

Proof:

$$\begin{aligned}
\mathbb{E}[\psi(X)] &= \sum_x \psi(x) p_X(x) \\
&= \sum_{x,y} y p_{Y|X}(y|x) p_X(x) \\
&= \sum_{x,y} y p_{(X,Y)}(x, y) \\
&= \sum_y y p_Y(y).
\end{aligned}$$

Now we consider the continuous case. Let X and Y be continuous random variables with conditional density function $f_{Y|X}(y|x)$. Given the event $\{X = x\}$, Y has a new density function, i.e, $f_{Y|X}(y|x)$, hence has a new expectation $\mathbb{E}[Y|X = x]$ calculated by

$$\psi(x) = \mathbb{E}[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy.$$

Since $\psi(x)$ is a function of x, we write $\psi(X) = \mathbb{E}[Y|X]$ which is called *conditional expectation*. So conditional expectation is a function of the random variable X, hence a random variable itself whose value is known once the value of the random variable X is known. Conditional expectation in the continuous case has the same property as in the discrete case, that is, the conditional expectation $\psi(X) = \mathbb{E}[Y|X]$ satisfies

$$\mathbb{E}[\psi(X)] = \mathbb{E}[Y].$$

The proof of this result is as follows. Let R_X be the set of real numbers for which $f_X(x) > 0$ for $x \in R_X$.

$$\begin{aligned}
\mathbb{E}[\psi(X)] &= \int_{R_X} \psi(x) f_X(x)dx \\
&= \int_{R_X} \mathbb{E}[Y|X = x] f_X(x)dx \\
&= \int_{R_X} \left(\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy \right) f_X(x)dx \\
&= \int_{-\infty}^{\infty} y \left(\int_{R_X} f_{Y|X}(y|x) f_X(x)dx \right) dy \\
&= \int_{-\infty}^{\infty} y \left(\int_{R_X} f_{(X,Y)}(x, y)dx \right) dy = \int_{-\infty}^{\infty} y f_Y(y)dy = \mathbb{E}[Y].
\end{aligned}$$

This result is usually written as $\mathbb{E}[\mathbb{E}[Y|X]] = \mathbb{E}[Y]$ and is used extensively, providing a useful method for calculating $\mathbb{E}[Y]$ as

$$\mathbb{E}[Y] = \int_{R_X} \mathbb{E}[Y|X = x] f_X(x)\, dx.$$

Example 6 *Suppose X and Y have the joint density $f_{(X,Y)}(x, y) = cx(y - x)e^{-y}$ for $0 < x \le y < \infty$.*

1. *Find the constant c.*

2. Find the marginals of X *and* Y

3. Show that $\mathbb{E}[X|Y] = Y/2$ *and* $\mathbb{E}[Y|X] = X + 2$.

1. $\int_0^\infty \int_0^y f_{(X,Y)}(x,y)dxdy = 1 \Longrightarrow c = 1$.

2.

$$\begin{aligned} f_X(x) &= \int_x^\infty f_{(X,Y)}(x,y)dy = xe^{-x}, \quad x > 0; \\ f_Y(y) &= \int_0^y f_{(X,Y)}(x,y)dx = \frac{1}{6}e^{-y}y^3, \quad y > 0. \end{aligned}$$

3. It is easy to see that

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = 6x(y-x)y^{-3}; \\ f_{Y|X}(y|x) &= \frac{f_{(X,Y)}(x,y)}{f_X(x)} = (y-x)e^{x-y} \qquad \textit{for } 0 < x \le y < \infty; \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X|Y=y] &= \int_0^y x f_{X|Y}(x|y)dx = \frac{y}{2} \Longrightarrow \mathbb{E}[X|Y] = \frac{Y}{2}; \\ \mathbb{E}[Y|X=x] &= \int_x^\infty y f_{Y|X}(y|x)dy = x + 2 \Longrightarrow \mathbb{E}[Y|X] = X + 2. \end{aligned}$$