Conditional distributions

Recall that $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $A, B \in \mathcal{F}$ be two events. such that $\mathbb{P}(B) > 0$. Then we defined

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Conditional distributions for bivariate discrete random vectors

Let (X, Y) be a discrete random vector. ~~Then~~ This means that (X, Y) are defined on the same probability space. Consider $(y_j : j \in \mathbb{N})$ to be the support of the marginal distribution function of Y that is,

$$\mathbb{P}(Y = y_j) > 0 \text{ and } \sum_{j=1}^{\infty} \mathbb{P}(Y = y_j) = 1.$$

Then we can define the conditional probability mass function as

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)} = \frac{p_{i,j}}{p_Y(j)}.$$

where the numerator is given by the joint probability mass function and the denominator is given by the marginal probability mass function of Y. Similarly one can define the conditional probability mass function of X given the random variable Y.

We can now define the conditional distribution function of X given $\{Y = y_j\}$. Note that

$$F_{X|Y=y_j}(x) = \sum_{i \geq 1 : x_i \leq x}^{\infty} \mathbb{P}(X = x_i \mid Y = y_j)$$

$$= \frac{1}{\mathbb{P}(Y = y_j)} \sum_{i \geq 1 : x_i \leq x}^{\infty} \mathbb{P}(X = x_i ; Y = y_j)$$

$$= \frac{1}{\mathbb{P}(Y = y_j)} \mathbb{P}(X \leq x ; Y = y_j).$$

for all $x \in \mathbb{R}$. Similarly, we can define conditional distribution function of the random variable Y given $\{X = x_i\}$.

<u>Conditional distributions and probability density function of multivariate random vector of continuous type.</u>

In general, conditional distribution can be thought of as the <u>random distribution function</u>.

$F_{X|Y}(x) = \mathbb{P}(X \le x | Y) : (\mathbb{R} \times \Omega, \mathcal{B}(\mathbb{R} \times \mathcal{F}), \mathbb{P})$
$\longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

is a random variable where $(X, Y) : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ is a bivariate discrete random variable. We can use the definition of conditional ~~conditional~~ probability to compute the values

$F_{X|Y=y_j}(x) = \mathbb{P}(X \le x | Y = y_j)$ for $j \in \mathbb{N}$.

If $E = (y_j : j \in \mathbb{N})$ is the support of the distribution of Y, then we can say that

$F_{X|Y}(x) = \mathbb{P}(X \le x | Y) : (\mathbb{R} \times E, \mathcal{B}(\mathbb{R} \times E), \mathbb{P})$
$\longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

is a random variable which takes the value $F_{X|Y=y_j}(x)$ with probability $\mathbb{P}(Y = y_j) = p_Y(j)$.

Furthermore, $x \mapsto F_{X|Y}(x)$ is a random function which satisfies the properties of a distribution function with probability 1. Sometimes, we can say this as $x \mapsto F_{X|Y}(x)$ is a random distribution function. We say that the random function

$F_{X|Y}$ equals $F_{X|Y=y_j}$ with probability $\mathbb{P}(Y=y_j)$ for every $j=1,2,\ldots$

In a similar way, we can define conditional mass function. If (X,Y) is a bivariate random vector of discrete type, then for every $x \in \mathbb{R}$, we can define conditional probability mass function $p_{X|Y}(x) : (E, 2^E, \mathbb{P}) \longrightarrow \left([0,\infty), \mathcal{B}([0,\infty))\right)$ is a random variable. Note that

$p_{X|Y}(x)$ takes the value $p_{X|Y=y_j}(x) = \dfrac{\mathbb{P}(X=x, Y=y_j)}{\mathbb{P}(Y=y_j)}$

with probability $\mathbb{P}(Y=y_j)$ for $j=1,2,\ldots$

Note that $x \longmapsto p_{X|Y}(x)$ is a random nonnegative function which vanishes everywhere except the points $(x_j : j \geq 1)$ (support of the marginal distribution of the random variable X).

and $\displaystyle\sum_{i=1}^{\infty} p_{X|Y}(x_i) = 1$ with probability 1.

That is,

$$\mathbb{P}\left(\sum_{i=1}^{\infty} p_{X|Y}(x_i) = 1\right) = 1.$$

The random function $x \mapsto p_{X|Y}(x)$ takes the function $p_{X|Y=y_j}(x) = \dfrac{\mathbb{P}(X=x,\ Y=y_j)}{\mathbb{P}(Y=y_j)}$ for every $x \in \mathbb{R}$ with probability $\mathbb{P}(Y=y_j)$ ~~and~~ for every $j \geq 1$.

Joint, conditional and marginal probability mass function.

$$\sum_{j=1}^{\infty} p_{X|Y=y_j}(x_i)\ p_Y(j) = \sum_{j=1}^{\infty} p_{X|Y=y_j}(x_i)\ \mathbb{P}(Y=y_j)$$

$$= \sum_{j=1}^{\infty} \frac{\mathbb{P}(X=x_i,\ Y=y_j)}{\mathbb{P}(Y=y_j)}\ \mathbb{P}(Y=y_j)$$

$$= \sum_{j=1}^{\infty} \mathbb{P}(X=x_i,\ Y=y_j)$$

$$= \mathbb{P}(X=x_i) = p_X(i) \longrightarrow *_1 \quad \text{for every } i \in \mathbb{N}.$$

and hence we can conclude that

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} p_{X|Y=y_j}(x_i)\ p_Y(j) = 1 \longrightarrow *_2.$$

Recall ~~that~~ the definitionation of expectation of a discrete random variable. From the above discussion, it follows that

$$\mathbb{E}\left(p_{X|Y}^{(x)}\right) = \sum_{j\geq 1}^{\infty} p_{X|Y=y_j}^{(x)}\, \mathbb{P}(Y=y_j), \text{ for every } x\in\mathbb{R}.$$

From the linearity property of expectation, it follows that

$$\mathbb{E}\left(\sum_{i\geq 1}^{\infty} p_{X|Y}(x_i)\right) = \sum_{i\geq 1}^{\infty} \cancel{p_{X|Y}}\ \mathbb{E}\left(p_{X|Y}^{(x_i)}\right)$$

$$\left(\text{Note that } \mathbb{P}\left(\sum_{i\geq 1}^{\infty} p_{X|Y}^{(x_i)} \leq 1\right) = 1\right).$$

$$= \sum_{i\geq 1}^{\infty}\sum_{j\geq 1}^{\infty} p_{X|Y=y_j}^{(x_i)}\, \mathbb{P}(Y=y_j) \longrightarrow (*_3).$$

Combining $(*_1)$, $(*_2)$ and $(*_3)$, it follows that

$$\mathbb{E}\left(\sum_{i\geq 1}^{\infty} p_{X|Y}(x_i)\right) = \sum_{i\geq 1}^{\infty}\sum_{j\geq 1}^{\infty} p_{X|Y=y_j}^{(x_i)}\, \mathbb{P}(Y=y_j) = 1.$$

$$\longrightarrow (*_F).$$

Conditional probability mass function can be defined through $(*_F)$. By this, we mean that a ~~function $p_{X|Y}(x)$:~~ (random non-negative function $p_{X|Y}$ is a conditional probability mass function if $p_{X|Y}(x) = 0$ if $x \notin \{x_i : i\geq 1\}$ and it satisfies the equation $(*_F)$ and $(*_1)$.

This definition may seem complicated but can be proved to be equivalent to the previous definition of conditional probability mass function. This definition is more preferred as it can be generalized for random variables of continuous type which is less straight forward.

Random vectors of continuous type.

Problem

The conditional probability density function or conditional distribution function can not be defined in a similar way as we did for discrete bivariate random vector.

Note that (X, Y) be a random vector of continuous type implies that Y is a random variable of continuous type. For every $y \in \mathbb{R}$, we have $\mathbb{P}(X \in B \mid Y = y)$ undefined for any $B \in \mathcal{B}(\mathbb{R})$ according to the definition of conditional probability as $\mathbb{P}(Y = y) = 0$ for every $y \in \mathbb{R}$.

We can show that for every $x \in \mathbb{R}$, there is a ~~rand~~ random variable $F_{X|Y}(x) : (\Omega, \mathcal{F}, \mathbb{P})$

$\to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ of continuous type.

Generalizing the formulae obtained for multivariate / bivariate discrete random vector.

Let $(X,Y) : (\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ be a ~~bivarie~~ bivariate continuous random vector that is there exists a ~~fun~~ continuous function $f_{X,Y} : \mathbb{R}^2 \to [0,\infty)^2$ such that

$$F_{X,Y}(x,y) = \mathbb{P}(X \le x ; Y \le y) = \mathbb{P}(\{X \le x\} \cap \{Y \le y\})$$

$$= \int_{-\infty}^{x} du \int_{-\infty}^{y} dv\, f_{X,Y}(u,v) \quad \text{for all } (x,y) \in \mathbb{R}^2.$$

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

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

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

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

Note that

$f_{X|Y}(x|y)$, $f_{Y|X}(y|x)$ are called <u>conditional</u> <u>probability density function</u>.

$F_{X|Y}(x|y)$, $F_{Y|X}(y|x)$ are called <u>conditional</u> <u>probability distribution function</u>.

We can further check that

$$\int_{-\infty}^{\infty} dv \; f_{X|Y}(x|v)\, f_Y(v) = f_X(x) = \frac{\partial}{\partial x}\left[\int_{-\infty}^{x} du \int_{-\infty}^{\infty} dv \; f_{X,Y}(u,v)\right].$$

Note that many books use the fact that $f_{X|Y}(x)$ is a conditional probability ~~distribution~~ density <u>function</u>. But we interprete (as we did in the discrete case) $f_{X|Y}(x) : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ as a ~~random variable~~ ~~conditional~~ continuous random variable which takes the values $(f_{X|Y}(x|y) : y \in \mathbb{R})$ with probability density functions given by $(f_Y(y) : y \in \mathbb{R})$.

Note that $(f_{X|Y}(x) : x \in \mathbb{R})$ and $(F_{X|Y}(x) : x \in \mathbb{R})$ are (uncountable) collections/families of random variables defined on the same probability space where the bivariate random vector (X, Y) lives.

Expectation and moments.

Let $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be a continuous function. Then

$$\mathbb{E}(g(X,Y)) = \begin{cases} \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \; g(x,y) \, f_{X,Y}(x,y) \\ \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} g(x_i, y_j) \, p_{X,Y}(x_i, y_j) \end{cases}$$

Note that we can take $g(X,Y) = X^k$ for $k \geq 1$.

Then we get the moments of the random variable X.

Similarly, we can obtain the moments of Y.

A special attention is given to the function

$$g(X,Y) = XY$$

We then call

$$\mathbb{E}(g(X,Y)) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

the covariance of X and Y. We shall use

$$\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right]$$