Chapter 3.5 Conditional Distributions

1 General Remarks

While this setting deals with multivariate r.v.s we will consider only simpler cases, usually involving only two random variables. From the context we will see how it works more generally, and will deal with this only a little.

There are three settings for bivariate r.v. (i) bivariate discrete, (ii) bivariate continuous and (iii) some special cases of bivariate with one component discrete and one continuous.

Conditional distributions are (i) distributions and hence with all these corresponding properties (eg sum or integrate to 1) and (ii) the distributions are random, that is they depend on other random variable upon which we are conditioning.

2 Discrete Case

As is usual the bivariate discrete case is the easiest to consider, as we can proceed directly using the axioms probability and properties of distributions or pmf's.

Consider bivariate (X, Y). Suppose y is such that $P(Y = y) > 0$. Then since $\{X = x\}$ is an event we have from the definition of conditional probability that

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} .$$

Now consider, for a fixed value y.

$$q(x) = P(X = x|Y = y) .$$

This function, treating y as fixed and x as the argument, obeys all the properties of a probability mass function (pmf), but of course in general it is a different function for different y.

Notation : Let

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

This is called the conditional pmf of X given the event $Y = y$, or more simply the conditional pmf of X given $Y = y$. This notation is helpful as it reminds us that x is the argument of this function. Other notations are often used, for example as in the Rice text $p_{X|Y}(x|y)$, but then one needs to be careful as to the role of x and y, as y is not an argument of this function. The particular value of y specifies which of these particular conditional distributions one is using.

As with any pmf we can now calculate

$$P(X \in A|Y = y) = \sum_{x \in A} p_{X|Y=y}(x) \ .$$

The observation that $p_{X|Y=y}(x)$ is a pmf will be very helpful when we define the notion of conditional expectation.

Formal definition : The conditional probability distribution of X given $Y = y$ is given by $p_{X|Y=y}$ where

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

if $P(Y = y) = p_Y(y) > 0$ and

$$p_{X|Y=y} \text{ is undefined if } p_Y(y) = 0 \ .$$

The undefined case is not important, in the sense that in practice if we wish to calculate $P(X \in A)$ then

$$\begin{aligned} P(X \in A) &= \sum_{x \in A} p_X(x) \\ &= \sum_{x \in A} \sum_{y} p_{X,Y}(x, y) \\ &= \sum_{x \in A} \sum_{y} p_{X|Y=y}(x) p_Y(y) \ . \end{aligned}$$

Notice in this last line for those y such that $p_Y(y) = 0$, it does not matter what value is assigned to $p_{X|Y=y}(x)$, one still has the same answer for $P(X \in A)$, as 0 times any number still gives 0.

3 Continuous Case

In the bivariate continuous case we cannot proceed directly from the Axioms of probability. There are some other technical issues that we cannot deal with in this course, so we proceed more informally by analogy with the discrete case.

Define the conditional pdf of X given $Y = y$ as

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

Again this notation is helpful to remind us that the argument of this function is x and that y plays a different role. One may also use an alternate notation as in the Rice text, such as $f_{X|Y}(x|y)$. The same comments pertaining to the role of y in the case of conditional pmfs also applies to conditional pmfs. The particular value of y specifies which conditional pdf one uses, and hence the corresponding support etc.

The conditional pdf has all the properties of any pdf, that is it non-negative and integrates to 1 and for intervals or other appropriate sets A

$$P(X \in A|Y = y) = \int_A f_{X|Y=y}(x)dx \ .$$

This property also is used when we discuss conditional expectation.

Using conditional pdf's one can calculate

$$\begin{aligned} P(X \in A) &= \int_A f_X(x)dx \\ &= \int_A \int_{-\infty}^{\infty} f_{X,Y}(x,y)dydx \\ &= \int_A \int_{-\infty}^{\infty} f_{X|Y=y}(x) f_Y(y)dydx \ . \end{aligned}$$

As in the discrete case notice that for y in the complement of the support of f_Y (so that $f_Y(y) = 0$) these values only contribute 0 to the integral. Thus we have a more formal definition.

Formal definition : The conditional probability distribution of X given $Y = y$ is given by $f_{X|Y=y}$ where

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

if $f_Y(y) > 0$ and

$$f_{X|Y=y} \text{ is undefined if } f_Y(y) = 0 \ . \tag{1}$$

In principle the value that we assign in the *undefined* case (1) does not change the value of the integral in the calculation of $P(X \in A)$ above. The so called *undefined* case is never used in such the probability calculation.

4 A Special Case

A useful special case is when X is continuous and Y is discrete, usually integer valued. If $P(Y = y) > 0$ then we have to define the conditional pdf of X given $Y = y$ as $f_{X|Y=y}(x)$ implicitly so that it satisfies

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

Usually in the context the marginal pmf of Y and the conditional pdf $f_{X|Y=y}(\cdot)$ are given so one may calculate conditional probabilities. This is used in particular to describe mixture models and some types of loss models.

Example : In an insurance model there are two types of customers, low and high accident prone customers. In a financial model there are two periods of finance, low and high volatility periods. Let Y take values 1 and 2 to represent these. Let $F_{X|Y=1}$ and $F_{X|Y=2}$ represent the cdf's (and correspondingly their pdf's with similar notation) in these two cases. For example they might be exponential distributions with different means, or two normal distributions with different variances. With this information we are now able to calculate the *marginal* distribution of X.

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P(X \leq x, Y = 1) + P(X \leq x, Y = 2) \\
&= P(X \leq x | Y = 1)P(Y = 1) + P(X \leq x | Y = 2)P(Y = 2) \\
&= F_{X|Y=1}(x)P(Y = 1) + F_{X|Y=2}(x)P(Y = 2)
\end{aligned}
$$

From this and differentiation we then obtain the marginal pdf of X as

$$f_X(x) = f_{X|Y=1}(x)P(Y = 1) + f_{X|Y=2}(x)P(Y = 2) \ .$$

In particular since Y only takes two values let $\alpha = P(Y = 1)$. Then the marginal pdf of X is

$$f_X(x) = f_{X|Y=1}(x)\alpha + f_{X|Y=2}(x)(1 - \alpha) \ .$$

This is an example of a so called finite mixture distribution.

5 Additional Comments

Suppose X, Y have pdf or pmf f or $f_{X,Y}$ (using whichever notation will be more convenient in the context). Since

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

we then have in the case that X, Y are independent that

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

Thus

- if X and Y then the conditional pdf $f_{X|Y=y}$ is equal to the marginal distribution f_X no matter what the value of y happens to be

- if the conditional pdf $f_{X|Y=y}$ is not the same function for all y (ie is different for some distinct values of y) then X and Y must be dependent
- X and Y are independent iff and only if

  $$f_{X,Y} = f_X \times f_Y$$

  Notice this means the two functions are equal, that is they must be equal for all *relevant* arguments. In the case of pmf's this means for all arguments. In the case of pdf's this means for all arguments except possibly on the boundaries of the supports.

To check for dependence or independence we may use either of these ideas, depending on what calculations we have already made or on what information we are given.

If we are given the relevant marginal and conditional pmf or pdf then we can calculate the joint pmf or pdf by

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

Again one should notice these must be the same functions on both sides of the = sign, so they must be equal for all *relevant* (x,y).