LECTURE 14
CONDITIONAL DISTRIBUTIONS AND EXPECTATIONS

1 Conditional distributions

Recall that we previously defined the conditional probability of an event A given another event B as

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

when $P(B) > 0$. The conditional probability allows us to improve our prediction about the likeliness of event A given the information on event B. The approach can be extended to random variables. Given two random variables X and Y, we can define

$$P(Y \in A|X \in B) = \frac{P(Y \in A, X \in B)}{P(X \in B)} \tag{1}$$

provided that $P(X \in B) > 0$, where now A and B are subsets of $\mathbb{R}$. The conditional probability $P(Y \in A|X \in B)$ is the probability of Ytaking a value from the set B given that X has taken a value from the set A. It uses the information in the joint distribution of X and Y to give a more accurate prediction about the behavior of Y.

Suppose that X and Y are discrete random variables. Using $A = \{y\}$ and $B = \{x\}$, we immediately obtain from (1) a definition of the *conditional* PMF.

Definition 1. (Conditional PMF) Let X and Y be two discrete random variables with a joint PMF $p_{X,Y}$ and supports $S_X = \{x_1, x_2, \ldots\}$ and $S_Y = \{y_1, y_2, \ldots\}$ respectively. Let $x \in S_X$ be such that $p_X(x) > 0$, where $p_X(x) = \sum_{u \in S_Y} p_{X,Y}(x, u)$ is the marginal PMF of X. The conditional PMF of Y given $X = x$ is defined as

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

for $y \in S_Y$.

Remark. 1. The conditional PMF is a PMF. Note that

$$p_{Y|X}(y|x) \geq 0,$$

and

$$\begin{aligned}\sum_{y\in S_Y} p_{Y|X}(y|x) &= \sum_{y\in S_Y} \frac{p_{X,Y}(x,y)}{p_X(x)} \\ &= \frac{\sum_{y\in S_Y} p_{X,Y}(x,y)}{p_X(x)} \\ &= \frac{p_X(x)}{p_X(x)} \\ &= 1.\end{aligned}$$

2. We should distinguish between $p_{Y|X}(y|x)$ and $p_{Y|X}(y|X)$. The former is $P(Y = y|X = x)$, i.e. the uncertainty about X has been realized, and X has taken a specific value x. The latter is $P(Y = y|X)$ without specifying what value X has taken, i.e. there is still uncertainty about X. In that case, we should think about $p_{Y|X}(y|X)$, $y \in S_Y$, as a family of functions corresponding to different realizations of X from S_X, i.e. $p_{Y|X}(y|X)$ is a *random* PMF, where randomness comes from the uncertainty about X.

Example. Suppose that the joint distribution of X and Y are as in the following table:

Table 1: Joint PMF of earnings (X) and price (Y) with marginal distributions (at the margins)

		Price of a share (Y)			
		\$100	\$250	\$400	
Earnings per share (X)	\$10	$\frac{2}{6}$	$\frac{1}{6}$	0	$\frac{1}{2}$
	\$20	0	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{2}$
		$\frac{2}{6}$	$\frac{3}{6}$	$\frac{1}{6}$	

The conditional distribution of the price conditional on earnings $X = 10$ is given by:

$$\begin{aligned}p_{Y|X}(100|10) &= \frac{2/6}{1/2} = \frac{2}{3}, \\ p_{Y|X}(250|10) &= \frac{1/6}{1/2} = \frac{1}{3}, \\ p_{Y|X}(400|10) &= \frac{0}{1/2} = 0.\end{aligned}$$

Hence, when earnings $X = 10$, the price is equal to 100 with probability $2/3$ and 250 with probability $1/3$. The price cannot be 400 when $X = 10$. (Compare those numbers with the unconditional probabilities.) Similarly, we find:

$$\begin{aligned} p_{Y|X}(100|20) &= 0, \\ p_{Y|X}(250|20) &= \frac{2}{3}, \\ p_{Y|X}(400|20) &= \frac{1}{3}. \end{aligned}$$

Note that the conditional distribution of Y varies as we change the conditioning variable X.

In the continuous case, it is convenient to define conditional distributions using joint and marginals PDFs

Definition 2. (Conditional PDF) Let X and Y be continuously distributed with a joint PDF $f_{X,Y}(x,y)$. Let x be such that $f_X(x) > 0$, where $f_X(x) = \int f_{X,Y}(x,y)dy$ is the marginal PDF of X. The conditional PDF of Y given $X = x$ is defined as

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

Remark. 1. Note that the conditional PDF is a PDF:

$$\begin{aligned} f_{Y|X}(y|x) &\geq 0 \text{ for all } y \in \mathbb{R}, \\ \int_{-\infty}^{\infty} f_{Y|X}(y|x)dx &= \int_{-\infty}^{\infty} \left(f_{X,Y}(y,x)/f_X(x)\right) dy \\ &= \int_{-\infty}^{\infty} \left(f_{X,Y}(y,x)\right) dy / f_X(x) \\ &= f_X(x)/f_X(x) \\ &= 1. \end{aligned}$$

2. As in the discrete case, we should distinguish between $f_{Y|X}(y|x)$ and $f_{Y|X}(y|X)$. The former is simply a PDF when viewed as a function of y. The latter is a random PDF: we will have different PDFs depending on the realization of X.

Statistical independence has important consequences for conditional distributions.

Theorem 3. *Suppose that random variables X and Y are independent. Then the conditional distribution of Y given X is the same as the marginal (unconditional) distribution of Y.*

Remark. The result implies that, when X and Y are independent, one cannot improve the description of the behavior of Y by relying on the information in X since the two random variables are fully unrelated.

Proof. The result follows immediately from the definition of the conditional PMF or PDF. In the discrete case, let $p_{Y|X}$, $p_{X,Y}$, p_X and p_Y denote the conditional PMF of X, the joint PMF of Y and X, the marginal PMF of X and the marginal PMF of Y respectively. We have

$$\begin{aligned} p_{Y|X}(y|x) &= \frac{p_{X,Y}(x,y)}{p_X(x)} \\ &= \frac{p_X(x)p_Y(y)}{p_X(x)} \\ &= p_Y(y), \end{aligned}$$

where the second equality follows from independence of X and Y, which implies that the joint PMF is equal to the product of marginal PMFs. In the continuous case, the proof is identical with PMFs replaced by PDFs. □

2 Conditional expectations

Conditional distributions satisfy all the properties of unconditional distributions. In particular, we can define expectations and moments in the same manner as with unconditional distributions.

Definition 4. (Conditional expectation) (a) Let $p_{Y|X}(y|x)$ be the conditional PMF of Y given $X = x$. The conditional expectation of Y given $X = x$ is defined as

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

(b) Let $f_{Y|X}(y|x)$ be the conditional PDF of Y given $X = x$. The conditional expectation of Y given $X = x$ is defined as

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

Remark. 1. The conditional expectation is a property of the conditional distribution: $E(Y|X = x)$ determines the location of the distribution of Y when $X = x$. In general, it is an improvement over the unconditional expectation $E(Y)$, since $E(Y|X = x)$ relies on additional information contained in X and the joint distribution of Y and X.

2. We should distinguish between $E(Y|X = x)$ and $E(Y|X)$. The latter should be viewed as a function of X since the uncertainty about X has not been realized:

$$E(Y|X) = \sum_{y \in S_Y} y p_{Y|X}(y|X)$$

or

$$E(Y|X) = \int_{-\infty}^{\infty} y f_{Y|X}(y|X) dy.$$

After averaging out y (or integrating out), we are left with a function of X. Hence, in general $E(Y|X)$ is a random variable unlike $E(Y)$, which is a constant.

Example. Consider the example in Table 1.

$$\begin{aligned} E(Y|X=10) &= 100\times\frac{2}{3}+250\times\frac{1}{3}+400\times 0=150. \\ E(Y|X=20) &= 100\times 0+250\times\frac{2}{3}+400\times\frac{1}{3}=300. \end{aligned}$$

Hence,

$$E(Y|X)=150\times 1(X=10)+300\times 1(X=20).$$

Note that $E(Y|X)$ varies with X. This illustrates that $E(Y|X)$ is a function of X and therefore a random variable.

Some of the properties of the conditional expectation are given below.

Theorem 5. **(a)** *If X and Y are independent, $E(Y|X)=E(Y)$.*

(b) *$E(Yg(X)|X)=g(X)E(Y|X)$.*

(c) *The law of Iterated Expectation: $EE(Y|X)=EY$.*

(d) *If $E(Y|X)=E(Y)$, then X and Y are uncorrelated.*

Remark. The result in (b) has a very intuitive interpretation: after conditioning on X, we can treat every function of X as known (fixed). Hence, $g(X)$ can be moved outside of the conditional expectation since for the purpose of computing that expectation we treat X as known (fixed).

Proof. The result in (a) follows immediately from the definition of the conditional distribution since, under independence, the conditional distribution of Y is the same as the unconditional (marginal) distribution of Y.

To show (b), suppose that X and Y are discrete.

$$\begin{aligned} E(Yg(X)|X) &= \sum_{y\in S_Y} yg(X)p_{Y|X}(y|X) \\ &= g(X)\sum_{y\in S_Y} yp_{Y|X}(y|X) \\ &= g(X)E(Y|X). \end{aligned}$$

For (c), we will show the result for the continuous case. The proof in the discrete case is identical.

$$
\begin{aligned}
E(Y|X) &= \int y f_{Y|X}(y|X)dy \equiv h(X), \\
E(h(X)) &= \int h(x) f_X(x)dx.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
EE(Y|X) &= \int \left(\int y f_{Y|X}(y|x)dy \right) f_X(x)dx \\
&= \int \int y f_{Y|X}(y|x) f_X(x)dydx \\
&= \int \int y f_{X,Y}(y,x)dydx \quad \text{(since } f_{Y|X} f_X = f_{X,Y}) \\
&= \int y \left(\int f_{X,Y}(x,y)dx \right) dy \\
&= \int y f_Y(y)dy \quad \text{(since } \int f_{X,Y}(x,y)dx = f_Y(y)) \\
&= EY.
\end{aligned}
$$

Part (d) follows from the results in (b) and (c). Consider first $E(XY)$. By the law of iterated expectation in (c),

$$
\begin{aligned}
E(XY) &= EE(XY|X) \\
&= E\left(XE(Y|X)\right) \\
&= E(XE(Y)) \\
&= (EX)(EY).
\end{aligned}
$$

Hence,

$$
Cov(X,Y) = E(XY) - (EX)(EY) = 0.
$$

□

The property in part (d) of Theorem 5 is called mean independence.

Definition 6. (Mean independence) A random variable Y is said to be mean independent of X if

$$
E(Y|X) = constant, \tag{2}
$$

i.e. not a function of X.

Remark. 1. By Theorem 5(c) (the law of iterated expectation), $EE(Y|X) = E(Y)$. Hence, the constant in (2) must be equal to $E(Y)$. Thus, mean independence can be alternatively stated as

$$E(Y|X) = E(Y).$$

2. Mean independence implies that the location of the distribution of Y cannot be explained using X. This does not necessarily mean that X and Y are fully (statistically) independent: it is possible that X explains some other features of the distribution of Y. Note that under statistical independence X cannot explain any features of the distribution of Y.

3. We have three levels of "unrelatedness" of random variables. The strongest is statistical independence (full independence) and the weakest is uncorrelatedness (absence of a linear relationship). Mean independence is the intermediate level.

Statistical Independence		Mean Independence		Uncorrelatedness $Cov(Y, X) = 0$
$p_{Y\|X} = p_Y$ or $f_{Y\|X} = f_Y$	$\Rightarrow$	$E(Y\|X) = E(Y)$	$\Rightarrow$	Best linear fit is a line with zero slope

Conditional expectation is related to the concept of regression discussed in Lecture 13.

Theorem 7. *Suppose that the conditional expectation of Y given X is a linear function: $E(Y|X) = \alpha + \beta X$ for some constants α and β. Then,*

$$\begin{aligned} \beta &= \frac{Cov(X, Y)}{Var(X)}, \\ \alpha &= (EY) - \beta(EX). \end{aligned}$$

Remark. The theorem argues that when the conditional expectation of Y given X is linear, it coincides with the regression of Y against X.

Proof. Define

$$\begin{aligned} U &= Y - E(Y|X) \\ &= Y - \alpha - \beta X. \end{aligned}$$

Note that

$$\begin{aligned} E(U|X) &= E(Y - E(Y|X)|X) \\ &= E(Y|X) - E(Y|X) \\ &= 0. \end{aligned}$$

Hence, by (5)(d), X and U are uncorrelated. Furthermore, by Theorem (5)(c),

$$EU = EE(U|X) = 0.$$

Therefore, $Cov(X,U) = EXU$. We have

$$\begin{aligned} 0 &= EXU = EX(Y - \alpha - \beta X), \\ 0 &= EU = E(Y - \alpha - \beta X). \end{aligned}$$

However, these are the same equations that define the regression coefficients in Lecture 13, Theorem 9. $\square$

We can use conditional distributions to define other properties as in the unconditional case: conditional variances, conditional quantiles and etc. For example, the conditional variance of Y given X is defined as

$$Var(Y|X) = E\left((Y - E(Y|X))^2|X\right) = E(Y^2|X) - (E(Y|X))^2.$$

Under mean independence, it is possible that $E(Y|X) = constant$ while $Var(Y|X) = h(X)$ (a function that varies with X). Under statistical independence, however, we would have $E(Y|X) = constant$ and $Var(Y|X) = constant = Var(Y)$.