Lecture notes 3

Conditional Probability and Conditional Expectation

1. The Discrete Case

Recall that for any two events E and F, the conditional probability of E given F is defined, as long as $P(F) > 0$, by

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Hence if X and Y are discrete random variables, then it is natural to define the conditional probability mass function of X given $Y = y$, by

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

for all values of y such that $f_Y(y) > 0$, where $f(x,y) = P(X = x, Y = y)$ is the joint pmf of X and Y and $f_Y(y) = P(Y = y)$ is the marginal pmf of Y.

Similarly, the conditional cumulative distribution function of X given $Y = y$ is defined, for all y such that $P(Y = y) > 0$, by

$$F_{X|Y}(x \mid y) = P(X \leq x \mid Y = y) = \sum_{a \leq x} f_{X|Y}(a \mid y).$$

Finally, the conditional expectation of X given $Y = y$ is defined by

$$E(X \mid Y = y) = \sum_x x P(X = x \mid Y = y) = \sum_x x f_{X|Y}(x \mid y).$$

Example:

Let X be the number of jobs taken within 5 years after university graduation.

Let Y be the number of promotions within 5 years after university graduation.

Joint p.m.f. of X and Y

		Y			
		1	2	3	
	0	10/45	15/45	3/45	28/45
X	1	10/45	6/45	0	16/45
	2	1/45	0	0	1/45
		21/45	21/45	3/45	1

The joint p.m.f. of X and Y is given by

$$f(x,y) = \begin{cases} 10/45, & x=0, y=1 \\ 10/45, & x=1, y=1 \\ 1/45, & x=2, y=1 \\ 15/45, & x=0, y=2 \\ 6/45, & x=1, y=2 \\ 3/45, & x=0, y=3 \end{cases}$$

The marginal p.m.f. of X is given by

$$g(x) = P(X=x) = \begin{cases} 28/45, & x=0 \\ 16/45, & x=1 \\ 1/45, & x=2 \end{cases}$$

The marginal p.m.f. of Y is given by

$$h(y) = P(Y=y) = \begin{cases} 21/45, & y=1 \\ 21/45, & y=2 \\ 3/45, & y=3 \end{cases}$$

The conditional p.m.f. of X given $Y=1$ is

$$\begin{aligned} f_{X|Y}(x \mid 1) &= \frac{P(X=x, Y=1)}{P(Y=1)} \\ &= \begin{cases} \frac{(10/45)}{(21/45)} = \frac{10}{21} & x=0; \\ \frac{(10/45)}{(21/45)} = \frac{10}{21} & x=1; \\ \frac{(1/45)}{(21/45)} = \frac{1}{21} & x=2. \end{cases} \end{aligned}$$

The conditional expectation of X given $Y=1$ is

$$\begin{aligned} E(X \mid Y=1) &= 0(\frac{10}{21}) + 1(\frac{10}{21}) + 2(\frac{1}{21}) \\ &= \frac{12}{21} \\ &= \frac{4}{7}. \end{aligned}$$

Note:

(a)

$$E[g(X) \mid Y=y] = \sum_x g(x) f_{X|Y}(x \mid y)$$

(b) If X and Y are independent,

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

Example: Suppose that $f(x, y)$, the joint p.m.f. of X and Y is given by

$$f(1,1) = 0.5,\ \ f(1,2) = 0.1,\ \ f(2,1) = 0.1,\ \ f(2,2) = 0.3.$$

Find (i) the probability mass function of X given $Y = 1$; (ii) the conditional expectation of X given $Y = 1$.

Solution: (i) Note that $P(Y = 1) = f(1,1) + f(2,1) = 0.6$. Hence

$$\begin{aligned} P(X = 1 \mid Y = 1) &= \frac{P(X = 1, Y = 1)}{P(Y = 1)} \\ &= \frac{0.5}{0.6} \\ &= \frac{5}{6} \end{aligned}$$

Similarly,

$$\begin{aligned} P(X = 2 \mid Y = 1) &= \frac{P(X = 2, Y = 1)}{P(Y = 1)} \\ &= \frac{0.1}{0.6} \\ &= \frac{1}{6} \end{aligned}$$

(ii) The conditional expectation of X given $Y = 1$ is given by

$$\begin{aligned} E(X \mid Y = 1) &= \sum_x x P(X = x \mid Y = 1) \\ &= 1(\frac{5}{6}) + 2(\frac{1}{6}) \\ &= \frac{7}{6} \end{aligned}$$

Example: If X and Y are independent Poisson random variables with respectively means λ_1 and λ_2, find the conditional expected value of X given $X + Y = n$.

Solution:

$$\begin{aligned} P(X = k \mid X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} \\ &= \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \end{aligned}$$

since X and Y are independent. Recall that $X+Y$ has a Poisson distribution with mean $\lambda_1+\lambda_2$, therefore

$$\begin{aligned} P(X=k \mid X+Y=n) &= \frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}[\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}]^{-1} \\ &= \frac{n!}{k!(n-k)!}\frac{\lambda_1^k\lambda_2^{n-k}}{(\lambda_1+\lambda_2)^n} \\ &= C_k^n(\frac{\lambda_1}{\lambda_1+\lambda_2})^k(\frac{\lambda_2}{\lambda_1+\lambda_2})^{n-k} \end{aligned}$$

which is the p.m.f. of a binomial random variable with parameters n and $\lambda_1/(\lambda_1+\lambda_2)$. Hence,

$$E(X \mid X+Y=n) = n(\frac{\lambda_1}{\lambda_1+\lambda_2})$$

Example: Consider $n+m$ independent trials, each of which results a success with probability p. Compute the expected number of successes in the first n trials given that there are k successes in all.

Solution: Let Y denote the total number of successes and

$$X_i = \begin{cases} 1, & if\ the\ i^{th}\ trial\ is\ a\ success \\ 0, & otherwise. \end{cases}$$

Then the required answer is given by

$$E(\sum_{i=1}^n X_i \mid Y=k) = \sum_{i=1}^n E(X_i \mid Y=k)$$

Now

$$\begin{aligned} E(X_i \mid Y=k) &= 1[P(X_i=1 \mid Y=k)] + 0[P(X_i=0 \mid Y=k)] \\ &= P(X_i=1 \mid Y=k) \\ &= \frac{k}{n+m}. \end{aligned}$$

Therefore

$$E(\sum_{i=1}^n X_i \mid Y=k) = n(\frac{k}{n+m}).$$

2. The Continuous Case
   If X and Y have a joint probability density function $f(x,y)$, then the conditional probability density function of X, given $Y=y$, is defined for all values of y such that $f_Y(y)>0$, by
   $$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f(y)}.$$
   The conditional expectation of X, given $Y=y$, is defined for all values of y such that $f_Y(y)>0$, by
   $$E(X \mid Y=y) = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx.$$
   Note that
   $$E[g(x) \mid Y=y] = \int_{-\infty}^{\infty} g(x) f_{X|Y}(x \mid y) dx.$$

3. Computing Expectations by Conditioning

   Double Expectation Theorem:
   $$E(X) = E_Y[E_{X|Y}(X \mid Y)] = E[E(X \mid Y)]. \tag{1}$$
   If Y is a discrete random variable, then (1) means
   $$E(X) = \sum_y E(X \mid Y=y)P(Y=y). \tag{2}$$
   If Y is a continuous random variable, then (1) means
   $$E(X) = \int_{-\infty}^{\infty} E(X \mid Y=y) f_Y(y) dy. \tag{3}$$
   Proof: (the discrete case)
   $$\begin{aligned}
   \sum_y E(X \mid Y=y)P(Y=y) &= \sum_y \sum_x xP(X=x \mid Y=y)P(Y=y) \\
   &= \sum_y \sum_x x \frac{P(X=x, Y=y)}{P(Y=y)} P(Y=y) \\
   &= \sum_y \sum_x xP(X=x, Y=y) \\
   &= \sum_x x \sum_y P(X=x, Y=y) \\
   &= \sum_x xP(X=x) \\
   &= E(X).
   \end{aligned}$$
   The conditional expectation is often useful in computing the variance of a random variable. In particular, we have
   $$\begin{aligned}
   Var(X) &= E(X^2) - [E(X)]^2 \\
   &= E[E(X^2|Y)] - \{E[E(X|Y)]\}^2
   \end{aligned}$$

Example: A coin having probability p, $0 < p < 1$ of coming up heads, is successively flipped until the first head appears. What is the expected number of flips required?

Solution: Let N be the number of flips required.

Method 1:
Note that N is a Geometric random variable with p.m.f. $f(n) = (1-p)^{n-1}p$, $n = 1, 2, 3, \ldots$. Thus

$$\begin{aligned}
E(N) &= \sum_{n=1}^{\infty} n f(n) \\
&= \sum_{n=1}^{\infty} n(1-p)^{n-1}p \\
&= \sum_{n=1}^{\infty} p\frac{d}{dp}[-(1-p)^n] \\
&= -p\frac{d}{dp}\sum_{n=1}^{\infty}(1-p)^n \\
&= -p\frac{d}{dp}\frac{1-p}{p} \\
&= -p\frac{p(-1)-(1-p)(1)}{p^2} \\
&= \frac{1}{p}
\end{aligned}$$

Method 2:
Let

$$Y = \begin{cases} 1, & \textit{if the first flip results in a head} \\ 0, & \textit{otherwise} \end{cases}$$

By double expectation theorem,

$$\begin{aligned}
E(N) &= E[E(N \mid Y)] \\
&= E(N \mid Y=1)P(Y=1) + E(N \mid Y=0)P(Y=0) \\
&= E(N \mid Y=1)p + E(N \mid Y=0)(1-p)
\end{aligned}$$

Since $E(N \mid Y=1) = 1$ and $E(N \mid Y=0) = 1 + E(N)$, we have

$$E(N) = p + [1 + E(N)](1-p)$$

which implies $E(N) = 1/p$.

Example:
A miner is trapped in a mine containing three doors.
The first door leads to a tunnel which takes him to safety after two hour's travel.
The second door leads to a tunnel which returns him to the mine after three hour's travel.
The third door leads to a tunnel which returns him to the mine after five hour's travel.
Assuming that the miner is at all times equally likely to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Let X be the time until the miner reaches safety, and let Y be the door he initially chooses. By double expectation theorem,

$$\begin{aligned} E(X) &= E[E(X \mid Y)] \\ &= E(X \mid Y=1)P(Y=1) + E(X \mid Y=2)P(Y=2) + E(X \mid Y=3)P(Y=3) \\ &= \frac{1}{3}[E(X \mid Y=1) + E(X \mid Y=2) + E(X \mid Y=3)] \end{aligned}$$

Since $E(X \mid Y=1) = 2$, $E(X \mid Y=2) = 3 + E(X)$, and $E(X \mid Y=3) = 5 + E(X)$, we have

$$E(X) = \frac{1}{3}\{2 + [3 + E(X)] + [5 + E(X)]\}$$

which implies $E(X) = 10$.

4. Computing Probabilities by Conditioning

Let E denote an arbitrary event and define the indicator random variable X by,

$$X = \begin{cases} 1, & if\ E\ occurs \\ 0, & if\ E\ does\ not\ occur \end{cases}$$

It follows from the definition of X that

$$E(X) = P(E)$$

$$E(X \mid Y = y) = P(E \mid Y = y),$$

for any random variable Y.

Therefore

$$P(E) = \begin{cases} \sum_y P(E \mid Y=y)P(Y=y) & if\ Y\ is\ discrete \\ \int_{-\infty}^{\infty} P(E \mid Y=y) f_Y(y)dy & if\ Y\ is\ continuous. \end{cases}$$

Example: Each customer who enters Rebecca's clothing store will purchase a suit with probability p. If the number of customers entering the store is Poisson distributed with mean λ, what is the probability that Rebecca does not sell any suits?

Solution:
Let X be the number of suits that Rebecca sells.
Let N denote the number of customers who enter the store. By conditioning on N, we have

$$\begin{aligned}
P(X=0) &= \sum_{n=0}^{\infty} P(X=0 \mid N=n)P(N=n) \\
&= \sum_{n=0}^{\infty} P(X=0 \mid N=n)\frac{e^{-\lambda}\lambda^n}{n!}
\end{aligned}$$

Since $P(X=0 \mid N=n) = (1-p)^n$, we have

$$\begin{aligned}
P(X=0) &= \sum_{n=0}^{\infty}(1-p)^n \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty} \frac{[\lambda(1-p)]^n e^{-\lambda}}{n!} \\
&= e^{-\lambda}\sum_{n=0}^{\infty} \frac{[\lambda(1-p)]^n}{n!} \\
&= e^{-\lambda}e^{\lambda(1-p)} \\
&= e^{-\lambda p}
\end{aligned}$$

Theorem:

$$E(X) = E[E(X \mid Y)]$$
$$Var(X) = E[Var(X \mid Y)] + Var[E(X \mid Y)]$$

Proof: (When X is a discrete random variable)

$$\begin{aligned}
E[E(X \mid Y)] &= \sum_y E(X \mid Y=y)P(Y=y) \\
&= \sum_y \sum_x xP(X=x \mid Y=y)P(Y=y) \\
&= \sum_y \sum_x x\frac{P(X=x, Y=y)}{P(Y=y)}P(Y=y) \\
&= \sum_x x \sum_y P(X=x, Y=y) \\
&= \sum_x xP(X=x) \\
&= E(X).
\end{aligned}$$

The proof for continuous random variable X is omitted.

For the second conclusion of the theorem,

$$\begin{aligned}
Var(X|Y) &= E\{X^2 - 2XE(X|Y) + [E(X|Y)]^2 \mid Y\} \\
&= E(X^2|Y) - 2E(X|Y)E(X|Y) + [E(X|Y)]^2 \\
&= E(X^2|Y) - [E(X|Y)]^2 \\
E[Var(X|Y)] &= E(X^2) - E\{[E(X|Y)]^2\} \\
Var[E(X|Y)] &= E\{[E(X|Y) - E(X)]^2\} \\
&= E\{[E(X|Y)]^2 - 2E(X)E(X|Y) + [E(X)]^2\} \\
&= E\{[E(X|Y)]^2\} - 2E(X)E(X) + [E(X)]^2 \\
&= E\{[E(X|Y)]^2\} - [E(X)]^2
\end{aligned}$$

Therefore $E[Var(X|Y)] + Var[E(X|Y)] = E(X^2) - [E(X)]^2 = VAR(X)$.

Example 3 (Expectation of a random sum):
Suppose that the expected number of accidents per week at an industrial plant is 4 and that the numbers of workers injured in each accident are random variables with a common mean of 2. Assume also that the number of workers injured in each accident is independent of the number of accidents that occur. What is the expected number of workers injured during a week?

Solution: Let N denote the number of accidents per week and X_i be the number of workers injured in the i^{th} accident, $i = 1, 2, \ldots$; Then the total number of workers injured can be expressed as $\sum_{i=1}^{N} X_i$ with the convention that the sum equals zero when $N = 0$. We have

$$E[\sum_{i=1}^{N} X_i] = E[E(\sum_{i=1}^{N} X_i \mid N)] = E[NE(X_1)] = E(N)E(X_1) = 4 \times 2 = 8$$

since

$$E(\sum_{i=1}^{N} X_i \mid N = n) = E(\sum_{i=1}^{n} X_i \mid N = n) = E[\sum_{i=1}^{n} X_i] = nE(X_1).$$

Furthermore, if we assume that all X_i, where $i = 1, 2, \ldots,$ are independent and have a common variance, then

$$\begin{aligned} Var(\sum_{i=1}^{N} X_i \mid N = n) &= Var(\sum_{i=1}^{n} X_i \mid N = n) \\ &= Var(\sum_{i=1}^{n} X_i) \\ &= nVar(X_1) \end{aligned}$$

$$E[Var(\sum_{i=1}^{N} X_i | N)] = E[Var(X_1)N] = Var(X_1)E(N).$$

$$Var[E(\sum_{i=1}^{N} X_i | N)] = Var[NE(X_1)] = [E(X_1)]^2 Var(N).$$

and hence

$$Var(\sum_{i=1}^{N} X_i) = Var(X_1)E(N) + [E(X_1)]^2 Var(N).$$