Chapter 7 Conditional Probability and Conditional Expectation

Jingrui SUN

Email: sunjr@sustech.edu.cn

Southern University of Science and Technology

September 8, 2024

Outline I

Outline

Part 7.1: The Discrete Case

Suppose X and Y are discrete r.v.s. For any x and y, let us consider the events $\{X = x\}$ and $\{Y = y\}$ with $P\{Y = y\} > 0$.

Define the conditional p.m.f. of X given that $Y = y$ by

$$p_{X|Y}(x \mid y) \triangleq P\{X = x \mid Y = y\} = \frac{P\{X = x, Y = y\}}{P\{Y = y\}} = \frac{p(x, y)}{p_Y(y)}.$$

Then we define the conditional p.d.f. of X given that $Y = y$ by

$$\begin{aligned} F_{X|Y}(x \mid y) &\triangleq P\{X \leq x \mid Y = y\} = \frac{P\{X \leq x, Y = y\}}{P\{Y = y\}} \\ &= \sum_{a \leq x} p_{X|Y}(a \mid y). \end{aligned}$$

Hence both the conditional p.m.f. and the conditional p.d.f of X given that $Y = y$ are functions of x with y being viewed as fixed.

The conditional expectation of X given that $Y = y$ is defined by

$$E(X \mid Y = y) \triangleq \sum_x xP\{X = x \mid Y = y\} = \sum_x x \cdot p_{X|Y}\{x \mid y\}.$$

Special case: If X and Y are independent, then

$$\begin{aligned} p_{X|Y}(x \mid y) &= P\{X = x\} = p_X(x), \\ F_{X|Y}(x \mid y) &= F_X(x), \\ E(X \mid Y = y) &= E(X). \end{aligned}$$

Example 1: Suppose that $p(x, y)$, the joint p.m.f. of X and Y is given by

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

Calculate $p_{X|Y}(x \mid Y = 1)$.

Example 1: Suppose that $p(x, y)$, the joint p.m.f. of X and Y is given by

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

Calculate $p_{X|Y}(x \mid Y = 1)$.

Solution: Note that

$$p_Y(1) = \sum_x p(x,1) = p(1,1) + p(2,1) = 0.6,$$

$$p_{X|Y}(1 \mid 1) = \frac{P\{X=1, Y=1\}}{P\{Y=1\}} = \frac{p(1,1)}{p_Y(1)} = 5/6.$$

Similarly, $p_{X|Y}(2 \mid 1) = 1/6$.

Example 2: If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X + Y = n$, i.e., $E(X \mid X + Y = n)$.

Example 2: If X and Y are independent Poisson random variables with respective means λ_1 and λ_2, calculate the conditional expected value of X given that $X + Y = n$, i.e., $E(X \mid X + Y = n)$.

Solution: Recall that, under the above conditions, the random variable $X + Y$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$. Then

$$
\begin{aligned}
P\{X = k \mid X + Y = n\} &= \frac{P\{X = k, X + Y = n\}}{P\{X + Y = n\}} \\
&= \frac{P\{X = k, Y = n - k\}}{P\{X + Y = n\}} \\
&= \frac{P\{X = k\}P\{Y = n - k\}}{P\{X + Y = n\}} \\
&= \frac{\frac{e^{-\lambda_1}\lambda_1^k}{k!}\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}}{\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} \\
&= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^k \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{n-k}.
\end{aligned}
$$

Hence, Binomial distribution with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Therefore $E(X \mid X+Y=n)$ is just the expectation of a random variable whose distribution is binomial with parameters n and $\frac{\lambda_1}{\lambda_1+\lambda_2}$.

Thus the value is $n\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)$.

Example 3: If X and Y are independent binomial random variables with identical parameters n and p, calculate the conditional p.m.f. of X given that $X + Y = m$.

Example 3: If X and Y are independent binomial random variables with identical parameters n and p, calculate the conditional p.m.f. of X given that $X + Y = m$.

Solution: Note that $X + Y$ is binomial random variable with parameter $2n$ and p. Thus

$$P\{X + Y = m\} = \binom{2n}{m} p^m (1-p)^{2n-m}.$$

Now for $0 \leq k \leq \min\{n, m\}$ (Reason, see below)

$$P\{X = k \mid X + Y = m\} = \frac{P\{X = k\}P\{Y = m - k\}}{P\{X + Y = m\}}$$

$$= \frac{\binom{n}{k} p^k (1-p)^{n-k} \binom{n}{m-k} p^{m-k} (1-p)^{n-(m-k)}}{\binom{2n}{m} p^m (1-p)^{2n-m}}$$

$$= \frac{\binom{n}{k}\binom{n}{m-k}}{\binom{2n}{m}}.$$

Note that k must be less than n and as X and Y are non-negative, k also must be less than m.

This distribution is known as the hypergeometric distribution. It arises as the distribution of the number of black balls that are chosen when a sample of m balls is randomly selected from an urn containing n black and n white balls.

Outline

Part 7.2: 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 that $Y = y$, is defined for all values of y such that $\underline{f_Y(y) > 0}$, by

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

The conditional expectation of X given that $Y = y$ is defined for all values of y such that $\underline{f_Y(y) > 0}$, by

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

Example 1: Suppose the joint p.d.f. of X and Y is given by

$$f(x,y) = \begin{cases} 6xy(2-x-y), & 0<x<1, 0<y<1, \\ 0, & \text{otherwise.} \end{cases}$$

Compute the conditional expectation of X given that $Y=y$, where $0<y<1$.

Example 1: Suppose the joint p.d.f. of X and Y is given by

$$f(x,y) = \begin{cases} 6xy(2-x-y), & 0 < x < 1, 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$, where $0 < y < 1$.

Solution: Compute the conditional p.d.f. first:

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} = \frac{6xy(2-x-y)}{\int_0^1 6xy(2-x-y)dx} = \frac{6x(2-x-y)}{4-3y}.$$

Then

$$E(X \mid Y = y) = \int_0^1 x \frac{6x(2-x-y)}{4-3y} dx = \frac{5-4y}{8-6y}.$$

Example 2: Suppose X and Y have joint p.d.f. as

$$f(x, y) = \begin{cases} \frac{1}{2} y e^{-xy}, & 0 < x < \infty, 0 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $E(e^{\frac{X}{2}} \mid Y = 1)$.

Example 2: Suppose X and Y have joint p.d.f. as

$$f(x,y) = \begin{cases} \frac{1}{2} y e^{-xy}, & 0 < x < \infty, 0 < y < 2, \\ 0, & \text{otherwise.} \end{cases}$$

Compute $E(e^{\frac{X}{2}} \mid Y = 1)$.

Solution:

$$f_{X|Y}(x \mid Y = 1) = \frac{f(x,1)}{f_Y(1)} = \frac{\frac{1}{2}e^{-x}}{\int_0^1 \frac{1}{2}e^{-x}dx} = e^{-x},$$

$$E\left(e^{\frac{X}{2}} \mid Y = 1\right) = \int_0^\infty e^{\frac{x}{2}} f_{X|Y}(x \mid Y = 1)dx = \int_0^\infty e^{\frac{x}{2}} e^{-x} dx = 2.$$

(Think why for the above step and the general case)

Outline

Part 7.3: Computing Expectations by Conditioning

Suppose X and Y are two random (either discrete or continuous) variables. Let us denote by $E[X \mid Y]$, (or $E(X \mid Y)$) the function of the random variable Y whose value at $Y = y$ is $E[X \mid Y = y]$.

Note that $E[X \mid Y]$ is itself a random variable (as a function of random variables Y !!).

An extremely important property of conditional expectation is that for any two random variables X and Y, we have

$$E(X) = E[E(X \mid Y)].$$

Proof: Assume X and Y are continuous r.v.s (for discrete case, the proof is similar), then what we need to show is

$$\int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{\infty} E(X \mid Y = y) f_Y(y) dy. \tag{1}$$

Substituting the following

$$\begin{aligned} E(X \mid Y = y) &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y) dx \\ f_{X|Y}(x \mid y) &= \frac{f(x, y)}{f_Y(y)}, \\ f_X(x) &= \int_{-\infty}^{\infty} f(x, y) dy, \end{aligned}$$

into the left hand side of (1), we obtain

$$\begin{aligned}\int_{-\infty}^{\infty} x f_X(x) dx &= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y) dy \right) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f(x,y) dy dx \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x \frac{f(x,y)}{f_Y(y)} dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} E(X \mid Y = y) f_Y(y) dy.\end{aligned}$$

(1) is thus proved. Note that for discrete case, the important formula $E(X) = E[E(X \mid Y)$ takes the form

$$E(X) = \sum_y E(X \mid Y = y) P(Y = y).$$

Example 1: (The mean of a Geometric distribution) A coin, having probability p of coming up heads, is to be successively flipped until the first head appears. What is the expected number of flips required?

Example 1: (The mean of a Geometric distribution) A coin, having probability p of coming up heads, is to be successively flipped until the first head appears. What is the expected number of flips required?

Solution: (Use the method of conditioning)
Let N be the number of flips required, and let

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

Then

$$E(N) = E(N \mid Y = 1)P(Y = 1) + E(N \mid Y = 0)P(Y = 0),$$

$$E(N \mid Y = 1) = 1, \quad E(N \mid Y = 0) = 1 + E(N)$$

(Think why here! Consult Example 3.11 of the Reference book, see p108, 8th ed).

Hence

$$\begin{aligned} E(N) &= E(N \mid Y = 1)p + E(N \mid Y = 0)(1 - p) \\ &= p + (1 - p)(1 + E(N)) \\ &\Rightarrow E(N) = \frac{1}{p}. \end{aligned}$$

Example 2: 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 his mine after five hours. Assuming that the miner is at all times equally to choose any one of the doors, what is the expected length of time until the miner reaches safety?

Solution: Let X denote the time until the miner reaches safety, and let Y denote the door he initially chooses. Then

$$\begin{aligned}
E(X) = E[E(X \mid Y)] &= E(X \mid Y = 1)P(Y = 1) \\
&\quad + E(X \mid Y = 2)P(Y = 2) + E(X \mid Y = 3)P(Y = 3) \\
&= \frac{1}{3}[E(X \mid Y = 1) + E(X \mid Y = 2) + E(X \mid Y = 3)] \\
&= \frac{1}{3}[2 + 3 + E(X) + 5 + E(X)].
\end{aligned}$$

Therefore $E(X) = 10$.

Example 3: Sam will read either one chapter of his probability book or one chapter of his history book. If the number of misprints in a chapter of his probability book is Poisson-distributed with mean 2 and if the number of misprints in his history chapter is Poisson distributed with mean 5, then assuming that Sam is equally likely to choose either book, what is the expected number of misprints that Sam will come across?

Solution: (Again use the method of conditioning.) Letting X be the number of misprints and letting

$$Y = \begin{cases} 1, & \text{if Sam chooses the probability book} \\ 2, & \text{if Sam chooses the history book} \end{cases}$$

then

$$\begin{aligned} E(X) &= E[E(X \mid Y)] \\ &= E(X \mid Y = 1)P(Y = 1) + E(X \mid Y = 2)P(Y = 2) \\ &= \frac{1}{2}(2 + 5) = \frac{7}{2}. \end{aligned}$$

Example 4: (Random Sum Formula) Let $\{X_i; i = 1, 2, 3, \cdots\}$ be i.i.d. random variables. Let $S = \sum_{i=1}^{N} X_i$, where N is a random integer and independent of $\{X_i; i = 1, 2, 3, \cdots\}$. Calculate $E(S)$.

(What is the main difficulty here? Random Sum!)

Solution:

$$
\begin{aligned}
E(S) &= E[E(S \mid N)] \\
&= \sum_{n=0}^{\infty} E(S \mid N = n) P(N = n) \\
&= \sum_{n=0}^{\infty} E\left(\sum_{i=1}^{N} X_i \mid N = n\right) P(N = n) \\
&= \sum_{n=0}^{\infty} E\left(\sum_{i=1}^{n} X_i \mid N = n\right) P(N = n) \\
&= \sum_{n=0}^{\infty} E\left(\sum_{i=1}^{n} X_i\right) P(N = n),
\end{aligned}
$$

(N and $\sum_{i=1}^{n} X_i$ are independent, but N and $\sum_{i=1}^{N} X_i$ are NOT independent! Understand the IDEA and method more now?)

$$\begin{aligned}
E(S) &= \sum_{n=0}^{\infty} nE(X)P(N=n) \quad \text{(i.i.d)} \\
&= E(X)\sum_{n=0}^{\infty} nP(N=n) \\
&= E(X)E(N).
\end{aligned}$$

Example 5: (The variance of a random number of random variables) Compute the variance of S defined in Example 4.

Example 5: (The variance of a random number of random variables) Compute the variance of S defined in Example 4.

Solution: Note that $\mathrm{Var}(S) = E\left(S^2\right) - [E(S)]^2$. Now for any $N = n$, we have

$$
\begin{aligned}
E\left(S^2 \mid N = n\right) &= E\left[\left(\sum_{i=1}^{N} X_i\right)^2 \mid N = n\right] = E\left[\left(\sum_{i=1}^{n} X_i\right)^2 \mid N = n\right] \\
&= E\left[\left(\sum_{i=1}^{n} X_i\right)^2\right] \text{(Same Idea as above)} \\
&= \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) + \left[E\left(\sum_{i=1}^{n} X_i\right)\right]^2 .
\end{aligned}
$$

$$
\begin{aligned}
E(S^2 \mid N = n) &= n\mathrm{Var}\,(X) + [nE(X)]^2 \quad \text{(i.i.d.)} \\
&= n\mathrm{Var}\,(X) + n^2[E(X)]^2, \\
&\Rightarrow E(S^2) = E(N)\mathrm{Var}\,(X) + E(N^2)[E(X)]^2.
\end{aligned}
$$

(Getting the above yourself !!)

Noting that $E(S) = E(N)E(X)$, we obtain

$$
\begin{aligned}
\mathrm{Var}\,(S) &= E(N)\mathrm{Var}\,(X) + \left[E(N^2) - [E(N)]^2\right][E(X)]^2 \\
&= E(N)\mathrm{Var}\,(X) + \mathrm{Var}\,(N)[E(X)]^2.
\end{aligned}
$$

Outline

Part 7.4: Computing Probabilities by Conditioning

Let A denote an arbitrary event and define the indicator r.v. X by

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

We then have $E(X) = P(A)$ (Think why here) and

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

for any r.v. Y with $P(Y = y) > 0$. Now

$$\begin{aligned} P(A) = E(X) &= E[E(X \mid Y)] \\ &= \begin{cases} \displaystyle\sum_y P(A \mid Y = y) P(Y = y), & \text{if } Y \text{ is discrete}, \\ \displaystyle\int_{-\infty}^{\infty} P(A \mid Y = y) f_Y(y) dy, & \text{if } Y \text{ is continuous}. \end{cases} \end{aligned}$$

Example 1: Suppose that X and Y are independent continuous r.v.s having densities $f_X(x)$ and $f_Y(y)$ respectively. Compute $P(X < Y)$.

Example 1: Suppose that X and Y are independent continuous r.v.s having densities $f_X(x)$ and $f_Y(y)$ respectively. Compute $P(X < Y)$.

Solution:

$$
\begin{aligned}
P(X < Y) &= \int_{-\infty}^{\infty} P(X < Y \mid Y = y) f_Y(y) dy \\
&\quad \text{(Use the above just obtained result)} \\
&= \int_{-\infty}^{\infty} P(X < y \mid Y = y) f_Y(y) dy \qquad \text{(Why so?)} \\
&= \int_{-\infty}^{\infty} P(X < y) f_Y(y) dy \qquad \text{(Independence!)} \\
&= \int_{-\infty}^{\infty} F_X(y) f_Y(y) dy. \qquad \text{(Definition!)}
\end{aligned}
$$

Example 2: Suppose that X and Y are independent continuous r.v.s. Find the distribution of $X + Y$.

Example 2: Suppose that X and Y are independent continuous r.v.s. Find the distribution of $X+Y$.

Solution: For any $a \in \mathbb{R}$,

$$
\begin{aligned}
P\{X+Y<a\} &= \int_{-\infty}^{\infty} P\{X+Y<a \mid Y=y\} f_Y(y)dy \\
&= \int_{-\infty}^{\infty} P\{X<a-y \mid Y=y\} f_Y(y)dy \\
&= \int_{-\infty}^{\infty} P\{X<a-y\} f_Y(y)dy \qquad \text{(Independence!)} \\
&= \int_{-\infty}^{\infty} F_X(a-y) f_Y(y)dy.
\end{aligned}
$$

(This is called Convolution Law!!)

Example 3: 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 parameter λ, what is the probability that Rebecca does not sell any suits?

Example 3: 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 parameter λ, what is the probability that Rebecca does not sell any suits?

Solution: Let X be the No. of suits that Rebecca sells and N the No. of customers who enter the store.

$$\begin{aligned}
P\{X=0\} &= \sum_{n=0}^{\infty} P\{X=0 \mid N=n\} P\{N=n\} \\
&= \sum_{n=0}^{\infty} P\{X=0 \mid N=n\} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \sum_{n=0}^{\infty} (1-p)^n \frac{e^{-\lambda}\lambda^n}{n!} \text{ (Different customers are independent)} \\
&= e^{-\lambda} e^{\lambda(1-p)} = e^{-\lambda p}.
\end{aligned}$$

Example 4 (Example 3 continued): What is the probability that Rebecca sells k suits?

Example 4 (Example 3 continued): What is the probability that Rebecca sells k suits?

Solution: First note given that $N = n, X$ has binomial distribution with parameter n and p, thus

$$P\{X = k \mid N = n\} = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}, \quad \text{if } k \leq n,$$

and

$$P\{X = k \mid N = n\} = 0, \quad \text{if } k > n.$$

Then

$$
\begin{aligned}
P\{X=k\} &= \sum_{n=0}^{\infty} P\{X=k \mid N=n\} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{n=k}^{\infty} \frac{[\lambda(1-p)]^{n-k}}{(n-k)!} \\
&= \frac{e^{-\lambda}(\lambda p)^k}{k!} \sum_{n=0}^{\infty} \frac{[\lambda(1-p)]^{n}}{n!} \qquad \text{(Think why)} \\
&= \frac{e^{-\lambda}(\lambda p)^k}{k!} e^{\lambda(1-p)} = \frac{e^{-\lambda p}(\lambda p)^k}{k!}.
\end{aligned}
$$

That is, $X \sim \text{Poisson}(\lambda p)$.

Example 5 (The Ballot problem): In an election, candidates A receives n votes and B receives m votes, where $n > m$. Assume that all orderings are equally likely, show that the probability that A is always ahead in the count of votes is $\frac{n-m}{n+m}$.

Example 5 (The Ballot problem): In an election, candidates A receives n votes and B receives m votes, where $n > m$. Assume that all orderings are equally likely, show that the probability that A is always ahead in the count of votes is $\frac{n-m}{n+m}$.

Solution: Let $P_{n,m}$ denote the desired probability. By conditioning on which candidate receives the last vote counted, we have

$$P_{n,m} = P\{A \text{ always ahead}\}$$

$$= P\{A \text{ always ahead} \mid A \text{ receives last vote}\}P\{A \text{ receives the last vote}\}$$

$$+ P\{A \text{ always ahead} \mid B \text{ receives last vote}\}P\{B \text{ receives the last vote}\}$$

$$= \frac{n}{n+m}P\{A \text{ always ahead} \mid A \text{ receives last vote}\}$$

$$+ \frac{m}{n+m}P\{A \text{ always ahead} \mid B \text{ receives last vote}\}$$

$$= \frac{n}{n+m}P_{n-1,m} + \frac{m}{n+m}P_{n,m-1} \quad \text{(Think why here!)}.$$

Using the above relationship, we can prove the conclusion by the method of induction as follows.

- If $n+m=1$, $P_{n,m}=P_{1,0}=1$ (Easy, but why?) which can be written as $\frac{1-0}{1+0}$. Thus TRUE for $n+m=1$.

- Assume that $P_{n,m}=\frac{n-m}{n+m}$ is true when $n+m=k$.
  If $n+m=k+1$, then since $(n-1)+m=k=n+(m-1)$ and by using the above proven relationship and the induction assumption we have

$$P_{n,m}=\frac{n}{n+m}P_{n-1,m}+\frac{m}{n+m}P_{n,m-1}$$
$$=\frac{1}{n+m}\frac{n(n-1-m)+m[n-(m-1)]}{n+m-1}=\frac{n-m}{n+m}.$$

The proof is completed.