Lecture 2

Conditional Probabilities and Conditional Expectation

- Discrete conditional expectation
- Continuous conditional expectation
- Computing expectations by conditioning
- Computing probabilities by conditioning

2.1 The Discrete Case

- The *conditional probability mass function* of X given that $Y = y$ is

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

- The *conditional expectation* of X given that $Y = y$ is

$$\begin{aligned} E[X|Y = y] &= \sum_x x \cdot P\{X = x|Y = y\} \\ &= \sum_x x \cdot p_{X|Y}(x|y) \end{aligned}$$

The sum is taken over the possible values $x_1, x_2, \ldots$ of X.

Example 2.1 Suppose that $p(x, y)$ is given by

$X \backslash Y$	1	2
1	0.5	0.1
2	0.1	0.3

Determine the conditional pmf $p_{X|Y}(x|1)$.

$$P(Y=1) = P(1|X=1) + P(1|X=2) = 0.6$$

$$P_{X|Y}(x=1) = \frac{P(X=x, Y=1)}{P(Y=1)} = \begin{cases} \frac{5}{6} \\ \frac{1}{6} \end{cases}$$

Example 2.2 If X_1 and X_2 are independent binomial random variables with respective parameters (n_1, p) and (n_2, p), calculate the conditional pmf of X_1 given that $X_1 + X_2 = m$.

$X_1 \sim B(n_1, p)$ $\quad\quad$ $X_1 + X_2 \sim B(n_1+n_2, p)$.

$X_2 \sim B(n_2, p)$

$$P(X_1 = k \mid X_1 + X_2 = m) = \frac{P(X_1 = k,\ X_1 + X_2 = m)}{P\{X_1 + X_2 = m\}}$$

$$= \frac{P(X_1 = k) \cdot P(X_2 = m-k)}{P(X_1 + X_2 = m)}$$

$$= \frac{\binom{n_1}{k} \cdot p^k \cdot (1-p)^{n_1-k} \cdot \binom{n_2}{m-k} \cdot p^{m-k} \cdot (1-p)^{n_2-(m-k)}}{\binom{n_1+n_2}{m} \cdot p^m \cdot (1-p)^{n_1+n_2-m}} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}$$

2.2 The Continuous Case

- The *conditional probability density function* of X, given that $Y = y$, is defined by

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

- The idea is that

$$\begin{aligned} f_{X|Y}(x|y)dx &= \frac{f(x,y)\,dxdy}{f_Y(y)\,dy} \\ &\approx \frac{P\{x \le X \le x+dx,\ \ y \le Y \le y+dy\}}{P\{y \le Y \le y+dy\}} \\ &= P\{x \le X \le x+dx \mid y \le Y \le y+dy\}. \end{aligned}$$

- The *conditional expectation* of X, given $Y = y$, is defined by

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

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

Example 2.3 Suppose that the joint density of X and Y is

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

Compute $E[X|Y=y]$, for $0 < y < 1$.

$$f_Y(y) = \int_0^1 6xy(2-x-y)\,dx = 6\cdot\int_0^1 2xy - x^2y - xy^2\,dx = 6\cdot\left[x^2y - \tfrac{1}{3}x^3\cdot y - \tfrac{1}{2}y^2\cdot x^2\right] = -3y^2+4y$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f(y)} = \frac{6xy(2-x-y)}{-3y^2+4y} = \frac{6x(2-x-y)}{-3y+4}$$

$$E[X|Y] = \int_{-\infty}^{\infty} x\cdot f_{X|Y}(x|y)\,dx = \int_0^1 6x^2\left(\frac{2-x-y}{-3y+4}\right)dx = \frac{6}{4-3y}\int_0^1 2x^2 - x^3 - x^2y\,dx$$

$$= \frac{6}{4-3y}\cdot\left[\tfrac{2}{3}x^3 - \tfrac{1}{4}x^4 - \tfrac{1}{3}yx^3\right]_0^1$$

$$= \frac{\cancel{6}}{4-3y}\cdot\frac{5-4y}{\cancel{12}\,2} = \frac{5-4y}{8-6y}$$

function of y.

2.3 Computing Expectations by Conditioning

- Denote by $E[X|Y]$ the value $E[X|Y=y]$. Note that $E[X|Y]$ is itself a random variable because we let Y vary.

- A very important formula is the *total expectation* formula:

$$E[X] = E[E[X|Y]] \qquad (\text{or } E_X[X] = E_Y[E_{X|Y}[X|Y]]).$$

- If Y is discrete, then the total expectation formula reads:

$$E[X] = \sum_y E[X|Y=y] \cdot P\{Y=y\}.$$

- If Y is continuous, then the total expectation formula reads:

$$E[X] = \int_{-\infty}^{\infty} E[X|Y=y] \cdot f_Y(y)\, dy.$$

- This formula often makes the computation of $E[X]$ much easier.

Example 2.4 A miner is trapped in a mine containing three doors.

- 1st door: takes him to safety after two hours of travel
- 2nd door: takes him to back to the mine after three hours of travel
- 3rd door: takes him to back to the mine after five hours of travel

Suppose that all times the miner chooses one of the doors with equal probability (i.e. he has no memory). What is the expected length of time until he reaches safety?

X: the time until the miner reaches the safety.

Y: the door he initially selected.

(2 hours)

second door → go back to origin → another start

$$E[X] = \sum_{i=1}^{3} E[X|y_i]\cdot P(y_i) = 2\times\frac{1}{3} + (3+E[X])\frac{1}{3} + \frac{1}{3}[5+E(X)]$$

$$E(X) = \frac{2}{3} + 1 + \cancel{\frac{1}{3}E(X)} + \frac{5}{3} + \cancel{\frac{1}{3}E(X)} \Rightarrow \frac{1}{3}E(X) = \frac{10}{3} \Rightarrow E(X) = 10$$

Example 2.5

- Consider a sequence of random variables $X_1, X_2, \ldots, X_n$, whose expected values are all equal to $E[X]$.
- If n is fixed, then $E\left[\sum_{i=1}^{n} X_i\right] = E[X_1] + \cdots + E[X_n] = n \cdot E[X]$.
- However, if N (where $N \neq \infty$) is a random variable independent of X_i, then what is $E\left[\sum_{i=1}^{N} X_i\right]$? It cannot be $N \cdot E[X]$!
- The random variable $\sum_{i=1}^{N} X_i$ is called a *compound random variable*, whose number of terms to sum is also random.
- By the total expectation formula:

a number

$$E\left[\sum_{i=1}^{N} X_i\right] = \underset{N}{E}\left[E\left[\sum_{i=1}^{N} X_i \middle| N\right]\right] = E\left[N E[X]\right] = E[N]E[X].$$

$\square$

Example 2.6 (The Matching Rounds Problem)

(a) At a party n men throw their hats into the center of a room. The hats are mixed up and each man randomly selects one. Find the expected number of men who select their own hats.

X_i. whether ith man selects his own hats.

$$X_i = \begin{cases} 1 & \frac{1}{n}. \\ 0 & \frac{n-1}{n} \end{cases} \qquad E[X_i] = \frac{1}{n}$$

$$E[\sum_{i=1}^{n} E(X_i)] = \sum_{i=1}^{n} E[X_i] = n \cdot \frac{1}{n} = 1$$ has nothing to do with n.

$E(R_n) = n.$

(b) Suppose that those choosing their own hats depart, while others put their selected hats in the center of the room, mix them up, and then reselect. This process continues until everyone has his own hat. Show that $E[R_n] = n$, where R_n is the number of rounds of selection that are necessary when n individuals are initially present.

$E[R_1]=1$. Suppose $E[R_k]=k$, $k=1,2 \cdots n-1$.

X: the number of matches in the first round.

$$E[R_n] = \sum_{i=0}^{n} \underline{E(R_n \mid X=i)} \cdot P\{X=i\}$$

the game starts again with n-i people

$$= \sum_{i=0}^{n} \underline{(1+E(R_{n-i}))} \cdot P\{X=i\}$$

$$= \sum_{i=0}^{n} 1 \cdot P(X=i) + \sum_{i=0}^{n} E(R_{n-i}) P(X=i)$$

$$= 1 + E(R_n) \cdot P(X=0) + \sum_{i=1}^{n} E(R_{n-i}) P(X=i)$$

by math induction

$$E(R_n) = 1 + E(R_n) \cdot P(X=0) + \sum_{i=1}^{n} (n-i) P(X=i)$$

$= 1 + E(R_n) \cdot P(X=0) + n \sum_{i=1}^{n} P(X=i) - \sum_{i=1}^{n} i \, P(X=i)$

Solution: Proceed by mathematical induction. The case $E[R_1] = 1$ is obvious. Suppose $E[R_k] = k$ for $k = 1, 2, \ldots, n-1$. Let X be the number of matches that occur in the first round. Then,

$$\begin{aligned}
E[R_n] &= \sum_{i=0}^{n} E[R_n | X = i] \cdot P\{X = i\} \\
&= \sum_{i=0}^{n} (1 + E[R_{n-i}]) \cdot P\{X = i\} \\
&= \sum_{i=0}^{n} P\{X = i\} + \sum_{i=0}^{n} E[R_{n-i}] \cdot P\{X = i\} \\
&= 1 + E[R_n] \cdot P\{X = 0\} + \sum_{i=1}^{n} E[R_{n-i}] \cdot P\{X = i\}.
\end{aligned}$$

By the induction hypothesis $E[R_{n-i}] = n - i$. Thus,

$$\begin{aligned} E[R_n] &= 1 + E[R_n] \cdot P\{X = 0\} + \sum_{i=1}^{n} (n-i) \cdot P\{X = i\} \\ &= 1 + E[R_n] \cdot P\{X = 0\} + n \sum_{i=1}^{n} P\{X = i\} - \sum_{i=1}^{n} i \cdot P\{X = i\} \\ &= 1 + E[R_n] \cdot P\{X = 0\} + n(1 - P\{X = 0\}) - E[X]. \end{aligned}$$

By part (a), $E[X] = 1$. Thus,

$$E[R_n] = E[R_n] \cdot P\{X = 0\} + n(1 - P\{X = 0\}).$$

Putting the $E[R_n]$ terms on L.H.S. yields

$$E[R_n](1 - P\{X = 0\}) = n(1 - P\{X = 0\}).$$

Hence, $E[R_n] = n$. $\square$

(c) Find the conditional expected number of matches given that the first person did not have a match.

Solution: Let X denote the number of matches, and let X_1 equal 1 if the first person has a match and 0 otherwise. Then,

$$\begin{aligned} E[X] &= E[X|X_1 = 0] \cdot P\{X_1 = 0\} + E[X|X_1 = 1] \cdot P\{X_1 = 1\} \\ &= E[X|X_1 = 0] \cdot \frac{n-1}{n} + E[X|X_1 = 1] \cdot \frac{1}{n}. \end{aligned}$$

By part (a), $E[X] = 1$. If the first person has a match, then the remaining $n-1$ people select among their own $n-1$ hats, which by part (a), has an expected number of matches equal to 1. Thus, $E[X|X_1 = 1] = 1 + 1 = 2$. This gives

match for the first man

$$E[X|X_1 = 0] = \frac{n-2}{n-1}.$$

the expected number for the remaing n-1 person.

□

Some useful formulas:

$$\sum_{i=0}^{\infty} x^i = 1 + x + x^2 + \cdots = \frac{1}{1-x}, \qquad -1 < x < 1$$

$$\sum_{i=1}^{\infty} i x^{i-1} = 1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}.$$

Differentiate the first one to get the second one.

Yet another useful formula:

$$\sum_{i=1}^{\infty} i^2 x^{i-1} = 1^2 + 2^2 x + 3^2 x^2 + \cdots = \frac{1+x}{(1-x)^3}.$$

Multiply the second equation with x and differentiate.

2.4 Computing Probabilities by Conditioning

- Let E be an event and let Y be a random variable.
- By conditioning on Y, we have the *total probability formula*:

$$\begin{aligned} P(E) &= \sum_y P(E|Y=y) \cdot P\{Y=y\}, \quad \text{if } Y \text{ is discrete} \\ &= \int_{-\infty}^{\infty} P(E|Y=y) \cdot f_Y(y)\, dy, \quad \text{if } Y \text{ is continuous.} \end{aligned}$$

Example 2.8 In the Matching Rounds Problem, what is the probability of no matches?

Solution: Let E denote the event that no matches occur, and to make explicit the dependence on n, write $P_n = P(E)$. We start by conditioning on whether or not the first man selects his own hat–call these events M and M^c. Then

$$\begin{aligned} P_n = P(E) &= P(E|M)P(M) + P(E|M^c)P(M^c) \\ &= 0 \cdot \frac{1}{n} + P(E|M^c) \cdot \frac{n-1}{n} = P(E|M^c) \cdot \frac{n-1}{n}. \end{aligned}$$

Now, $P(E|M^c)$ is the probability of no matches when $n-1$ men select from a set of $n-1$ hats that does not contain the hat of one of these men. There are two mutually exclusive cases where this can happen.

E

First case, there are no matches and the extra man does not select the extra hat (first man's hat). By regarding the extra hat as belonging to the extra man, it is seen that the probability of this event is P_{n-1}.

Second case, there are no matches and the extra man selects the extra hat. That is, the extra man selects the first man's hat and the remaining $n-2$ men have no matches in selecting their $n-2$ hats. The probability of this event is $\frac{1}{n-1} \cdot P_{n-2}$. This yields that

$$P(E|M^c) = P_{n-1} + \frac{1}{n-1} \cdot P_{n-2}$$

and thus

$$P_n = \frac{n-1}{n} \cdot P(E|M^c) = \frac{n-1}{n} \cdot P_{n-1} + \frac{1}{n} \cdot P_{n-2},$$

or equivalently,

$$P_n - P_{n-1} = -\frac{1}{n}(P_{n-1} - P_{n-2}). = -\frac{1}{n}\left(-\frac{1}{n-1}\right) \cdot (P_{n-2} - P_{n-3})$$

By recursion, we have

$$P_n - P_{n-1} = \frac{-1}{n} \cdots \frac{(-1)^n}{3}(P_2 - P_1) = \frac{2(-1)^n}{n!}\left(\frac{1}{2} - 0\right) = \frac{(-1)^n}{n!}.$$

Finally, by telescoping, we have

$$P_n = \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} - \cdots + \frac{(-1)^n}{n!} \quad n \geq 2. \quad \square$$

Lecture 3

Some Applications

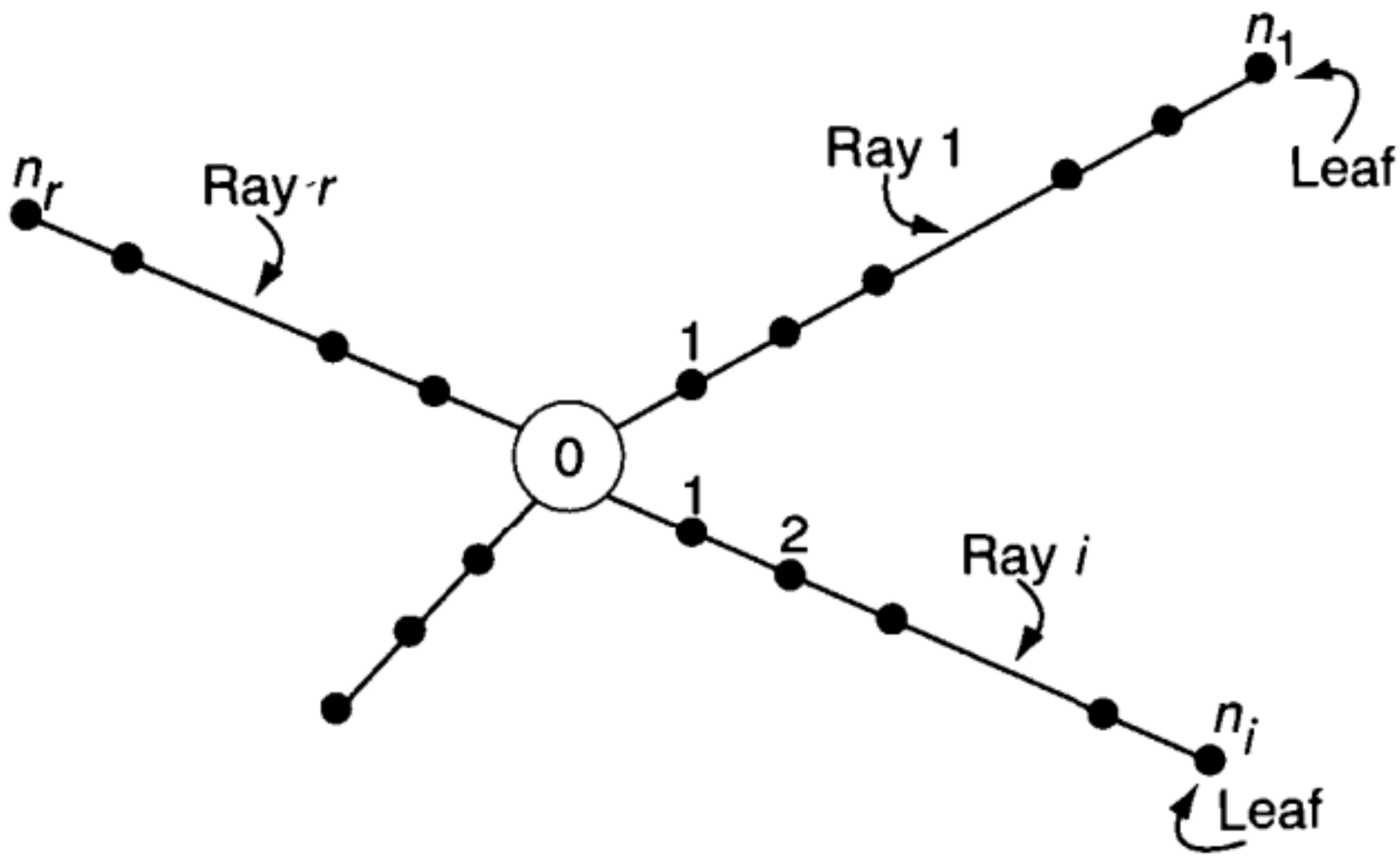
The following applications will be discussed

1. A list model
2. A random graph
3. Mean time for patterns
4. A packing problem
5. Star graph

3.5 Star Graph

- A graph consisting of a central vertex, labeled 0, and rays emanating from that vertex is called a *star graph*.
- Let r be the number of rays and let ray i consist of n_i vertices (not counting the central vertex), for $i = 1, 2, \dots, r$.
- Suppose that a particle moves along the vertices of the graph so that it is equally likely to move from the vertex it is presently at to any of the neighbors of that vertex, where two vertices are said to be neighbors if they are joined by an edge.
- The vertices at the far ends of the rays are called *leafs*.
- Question: What is the probability that, starting at node 0, the first leaf visited is the one on ray i, $i = 1, 2, \dots, r$?

A star graph



- Let L denote the first leaf visited. Let R be the first ray (i.e., the first step) visited.

- Conditioning on R yields

$$P\{L = i\} = \sum_{j=1}^{r} \underbrace{\frac{1}{r}}_{\nearrow P\{R=j\}.} \cdot P\{L = i | R = j\}.$$

- If the particle moves from node 0 to ray i, then the probability that it will reach the leaf before leaving the ray is $1/n_i$ (up $n_i - 1$ before down 1).

- Consider whether the particle will reach a leaf or start over:

$$\begin{aligned} P\{L = i | R = i\} &= \frac{1}{n_i} + \left(1 - \frac{1}{n_i}\right) P\{L = i\} \\ P\{L = i | R = j\} &= \left(1 - \frac{1}{n_j}\right) P\{L = i\}, \text{ for } j \neq i. \end{aligned}$$

- Finally,

$$\begin{aligned} r \cdot P\{L = i\} &= \sum_{j=1}^{r} P\{L = i | R = j\} \\ &= \frac{1}{n_i} + \sum_{j=1}^{r} \left(1 - \frac{1}{n_j}\right) P\{L = i\} \end{aligned}$$

or

$$P\{L = i\} = \frac{1/n_i}{\sum_{j=1}^{r} 1/n_j}.$$