Conditional Probability and Conditional Expectation

Stochastic Processes - Lecture Notes

Fatih Cavdur

to accompany
Introduction to Probability Models
by Sheldon M. Ross

Fall 2015

Outline I

Conditional Probability and Conditional Expectation

One of the most useful concepts in probability theory is that of conditional probability and conditional expectation since

- in practice, we are often interested in calculating probabilities and expectations when some partial information is available; hence, the desired probabilities and expectations are conditional ones;
- in calculating a desired probability or expectation it is often extremely useful to first "condition" on some appropriate random variable.

Conditional PMF

If X and Y are discrete RVs, the conditional PMF of X given that $Y = y$ is

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

for $P\{Y = y\} > 0$.

Conditional CDF

If X and Y are discrete RVs, the conditional CDF of X given that $Y = y$ is

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

Conditional Expectation

If X and Y are discrete RVs, the conditional expectation of X given that $Y = y$ is

$$\begin{aligned} E(X|Y = y) &= \sum_x xP\{X = x|Y = y\} \\ &= \sum_x x p_{X|Y}(x|y) \end{aligned}$$

Example

If X_1 and X_2 are independent binomial RVs with parameters (n_1, p) and (n_2, p), respectively, what is the PMF of X_1 given that $X_1 + X_2 = m$?

$$\begin{aligned}
P\{X_1 = k | X_1 + X_2 = m\} &= \frac{P\{X_1 = k, X_1 + X_2 = m\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{P\{X_1 = k, X_2 = m - k\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{P\{X_1 = k\} P\{X_2 = m - k\}}{P\{X_1 + X_2 = m\}} \\
&= \frac{\binom{n_1}{k} p^k q^{n_1 - k} \binom{n_2}{m-k} p^{m-k} q^{n_2 - m + k}}{\binom{n_1+n_2}{m} p^m q^{n_1 + n_2 - m}} \\
&= \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}
\end{aligned}$$

Example

If X_1 and X_2 are independent binomial RVs with parameters (n_1, p) and (n_2, p), respectively, what is the PMF of X_1 given that $X_1 + X_2 = m$?

$$P\{X_1 = k | X_1 + X_2 = m\} = \frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}$$

The distribution of X_1 is known as the *hyper-geometric* distribution.

It can be defined as the distribution of the number of blue balls when a sample of m balls is randomly chosen from an urn that contains n_1 blue and n_2 red balls.

Hyper-Geometric Distribution

- X: number of successes
- n: number of draws (without replacement)
- N: total number of items in the population
- K: total number of successes in the population

X follows a hyper-geometric distribution if its PDF is defined as

$$P\{X = k\} = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

In other words, X can be defined as the number of successes when a sample of n items is randomly chosen from population that contains N items of which k are successes.

Example

If X and Y are independent Poisson RVs with respective means λ_1 and λ_2, what is the conditional expectation of X given that $X + Y = n$?

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

Example

If X and Y are independent Poisson RVs with respective means λ_1 and λ_2, what is the conditional expectation of X given that $X + Y = n$?

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

Example

In other words, the conditional distribution of Poisson RV X given that $X + Y = n$, is the binomial distribution with parameters n and $\lambda_1/(\lambda_1 + \lambda_2)$. We can then write

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

Conditional PDF

If X and Y are continuous RVs, the conditional PMF of X given that $Y = y$ is

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

for $f_Y(y) > 0$.

Conditional CDF

If X and Y are continuous RVs, the conditional CDF of X given that $Y = y$ is

$$\begin{aligned} F_{X|Y}(x|y) &= P\{X \leq x | Y = y\} \\ &= \int_{a \leq x} f_{X|Y}(a|y) da \end{aligned}$$

Conditional Expectation

If X and Y are continuous RVs, the conditional expectation of X given that $Y = y$ is

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

Example

Suppose that the joint PDF 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

First compute the conditional density.

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

Example

Now compute the conditional expectation.

$$
\begin{aligned}
E(X|Y=y) &= \int_0^1 \frac{6x^2(2-x-y)dx}{4-3y} \\
&= \frac{2(2-y)}{4-3y} \\
&= \frac{5-4y}{8-6y}
\end{aligned}
$$

Example

Suppose that the joint PDF of X and Y is given by

$$f(x,y) = \begin{cases} 4y(x-y)e^{-(x+y)}, & 0 < x < \infty, 0 < y < x \\ 0, & \text{otherwise} \end{cases}$$

Compute the conditional expectation of X given that $Y = y$.

Example

First compute the conditional density.

$$\begin{aligned} f_{X|Y}(x|y) &= \frac{f(x,y)}{f_Y(y)} \\ &= \frac{4y(x-y)e^{-(x+y)}}{\int_y^\infty 4y(x-y)e^{-(x+y)}dx}, \quad x > y \\ &= \frac{(x-y)e^{-x}}{\int_y^\infty (x-y)e^{-x}dx}, \quad \text{let } w = x - y \\ &= \frac{(x-y)e^{-x}}{\int_y^\infty we^{-(y+w)}dw}, \quad x > y \\ &= (x-y)e^{-(x-y)}, \quad x > y \end{aligned}$$

Example 3.7

Now compute the conditional expectation.

$$\begin{aligned} E(X|Y=y) &= \int_y^{\infty} x(x-y)e^{-(x-y)}dx \\ &= \int_0^{\infty} (w+y)we^{-w}dw \\ &= E(W^2) + yE(W) \\ &= 2 + y \end{aligned}$$

Computing Expectations by Conditioning

Using

$$E(X) = E[E(X|Y)]$$

we can write

$$E(X) = \sum_y E(X|Y=y)P\{Y=y\}$$

for discrete RVs and

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

for continuous RVs.

Example

Suppose that the expected number of accidents per week is 4, and also suppose that the number of people who are injured in each accident are IID RVs with a mean of 2. If we assume that the number of accidents is independent of the number of injured people, what is the expected number of injuries during a week?

Let N be the number of accidents during a week and X_i the number of injured people due to the ith accident. We can then write the total number of injuries as $\sum_i^N X_i$.

Example

$$
\begin{aligned}
E\left(\sum_i^N X_i\right) &= E\left[E\left(\sum_i^N X_i|N\right)\right] \\
&= E\left[E\left(\sum_{i=1}^N X_i|N=n\right)\right] \\
&= E\left[E\left(\sum_{i=1}^N X_i\right)\right] \\
&= E[NE(X)] \\
&= E(N)E(X)
\end{aligned}
$$

The expected number of accidents is then 8.

Example: The Mean of a Geometric RV

Let N be the number of trials required and Y be defined as

$$Y = \begin{cases} 1, & \text{if first trial is a success} \\ 0, & \text{otherwise} \end{cases}$$

We can then write

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

We then have

$$E(N) = \frac{1}{p}$$

Example

A miner is trapped in a mine containing 3 doors. The first door leads to a tunnel that takes him to safety after 2 hours of travel. The second and third doors lead to tunnels that return him to the mine after 3 hours and 5 hours, respectively. If he randomly chooses a door at all times, what is the expected lenght of time until he reaches safety?

Example

Let X be the time until he reaches safety and Y be the door number he initially chooses.

$$
\begin{aligned}
E(X) &= \sum_{i=1}^{3} E(X|Y=i)P\{Y=i\} \\
&= \frac{\sum_{i=1}^{3} E(X|Y=i)}{3} \\
&= \frac{2+[3+E(X)]+[5+E(X)]}{3} \Rightarrow E(X)=10
\end{aligned}
$$

Computing Variances by Conditioning

Variance of a Geometric RV

To find $\text{Var}(X)$, we first condition on the result of the first trial.

$$
\begin{aligned}
E(X^2) &= E[E(X^2|Y)] \\
&= E\left[\sum_{i=1}^{2} E(X^2|Y=i)P\{Y=i\}\right] \\
&= p + E\left[(1+X)^2\right](1-p) \\
&= 1 + 2(1-p)E(X) + (1-p)E(X^2) \Rightarrow E(X^2) = \frac{2-p}{p^2}
\end{aligned}
$$

The variance is then computed as

$$\text{Var}(X) = \frac{1-p}{p^2}$$

Conditional Variance

The conditional variance of X given that $Y = y$ is

$$\begin{aligned}\text{Var }(X|Y=y) &= E\left\{[X - E(X|Y=y)]^2 \middle| Y=y\right\} \\ &= E\left[X^2 \middle| Y=y\right] - [E(X|Y=y)]^2\end{aligned}$$

We can compute the variance as

$$\text{Var }(X) = E\,[\text{ Var }(X|Y)] + \text{ Var }\,[E(X|Y)]$$

Computing Probabilities by Conditioning

The probability of event E, conditioned on some discrete RV Y is defined as

$$P(E) = \sum_y P(E|Y=y)P\{Y=y\}$$

and on some continuous RV Y as

$$P(E) = \int_{-\infty}^{+\infty} P(E|Y=y)f_Y(y)dy$$

Example

An insurance company assumes that the number of accidents that each policyholder will have in a year is a Poisson RV with a mean dependent on the policyholder. If the mean of the Poisson RV for a randomly chosen policyholder is given by

$$g(\lambda) = \lambda e^{-\lambda}, \ \lambda \geq 0$$

what is the probability that a randomly chosen policyholder will have n accidents next year?

Example

Let X be the number of accidents for a random person and Y be the Poisson mean for the person.

$$\begin{aligned}
P\{X = n\} &= \int_0^{\infty} P\{X = n | Y = \lambda\} g\lambda d\lambda \\
&= \int_0^{\infty} \frac{e^{-\lambda}\lambda^n}{n!} \lambda e^{-\lambda} d\lambda \\
&= \left(\frac{n+1}{n+1}\right)\left(\frac{2^{n+2}}{2^{n+2}}\right)\frac{1}{n!}\int_0^{\infty} \lambda^{n+1} e^{-2\lambda} d\lambda \\
&= \frac{n+1}{2^{n+2}}\left[\frac{2^{n+2}}{(n+1)!}\int_0^{\infty} \lambda^{n+1} e^{-2\lambda} d\lambda\right] \\
&= \frac{n+1}{2^{n+2}}\left[\int_0^{\infty} f(\lambda) d\lambda\right], \quad (f(\lambda) \sim \mathrm{G}\ (n+2, 2)) \\
&= \frac{n+1}{2^{n+2}}
\end{aligned}$$

Example

Let $X_i \dots X_n$ be IID Bernoulli RVs with corresponding parameters p_i. We want to compute the PMF of $X_1 + \dots + X_n$. To do so, we recursively compute the PMF of $X_1 + \dots + X_k$ first for $k = 1$, then, $k = 2$ etc. We thus let

$$p_k(j) = P\{X_1 + \dots X_k = j\}$$

We note that

$$p_k(k) = \prod_{i=1}^{k} p_i \quad \text{and} \quad p_k(0) = \prod_{i=1}^{k} q_i$$

Example

By conditioning on X_k for $0 < j < k$

$$\begin{aligned}
p_k(j) &= P\{X_1 + \dots X_k = j | X_k = 1\} p_k + P\{X_1 + \dots X_k = j | X_k = 0\} q_k \\
&= P\{X_1 + \dots X_{k-1} = j - 1 | X_k = 1\} p_k \\
&+ P\{X_1 + \dots X_{k-1} = j | X_k = 0\} q_k \\
&= P\{X_1 + \dots X_{k-1} = j - 1\} p_k + P\{X_1 + \dots X_{k-1} = j\} q_k \\
&= p_k P_{k-1}(j-1) + q_k P_{k-1}(j)
\end{aligned}$$

Example

Since we have,

$$p_k(j) = p_k P_{k-1}(j-1) + q_k P_{k-1}(j)$$

we start with the followings and solve the above equation recurusively to obtain the functions $P_2(j), \ldots P_n(j)$

$$P_1(1) = p_1 \text{ and } P_0(0) = q_1$$

Some Applications

- A List Model
- A Random Graph
- Uniform Priors
- Mean Time for Patterns
- The k-Record Values of Discrete RVs

A Random Graph

Consider a graph $G = (V, A)$ where

$$V = \{1, \ldots, n\} \text{ and } A = \{(i, x(i)), i = 1, \ldots, n\}$$

where $X(i)$ are independent RVs such that

$$P\{X(i) = j\} = \frac{1}{n}, \ j = 1, \ldots, n$$

Such a graph is commonly referred as a *random graph*.
We are interested in the probability of a random graph is connected.

A Random Graph

Define N equal the first k such that

$$X^k(1) \in \{1, X(1), \ldots, X^{k-1}(1)\}$$

By letting $p(C)$ be the probability that the graph is connected, we can write

$$p(C) = \sum_{k=1}^{n} P\{C|N = kP\{N = k\}$$

A Random Graph

Lemma

Given a random graph with nodes $0, \ldots, r$ and r arcs as $(i, Y(i))$, $i = 1, \ldots, r$, where

$$Y_i = \begin{cases} j, & P\{Y_i = j\} = 1/(r+k),\ j = 1, \ldots, r \\ 0, & P\{Y_i = 0\} = k/(r+k) \end{cases}$$

then

$$p(C) = \frac{k}{r+k}$$

which means that there are $r+1$ nodes (r ordinary nodes and 1 super node), and a chosen arc from out of each ordinary node goes to the super node and an ordinary node with respective probabilities. See the proof in the text!

A Random Graph

We can write

$$p(c) = \frac{E(N)}{n}$$

where we can show that

$$E(N) = \sum_{i=1}^{\infty} P\{N \geq i\}$$

and then

$$p(c) = \frac{(n-1)!}{n^n} \sum_{j=0}^{n-1} \frac{n^j}{j!}$$

A Random Graph

For simple approximation, assume that X is Poisson with mean n,

$$P\{X < n\} = e^{-n}\left(\sum_{j=0}^{n-1}\frac{n^j}{j!}\right)$$

Since a Poisson RV with mean n can be regarded as the sum of n independent Poisson with mean 1, it follows from the CLT that, for n large, such an RV is approximately normal, and thus,

$$p(C) \approx \frac{e^n(n-1)!}{2n^n}$$

by applying the Stirling approximation,

$$p(C) \approx \sqrt{\frac{\pi}{2(n-1)}}e\left(\frac{n-1}{n}\right)^n \Rightarrow \lim_{n\to\infty} p(C) = \sqrt{\frac{\pi}{2(n-1)}}$$

An Identity for Compound RVs

Let $X_1, X_2, \ldots$ be a sequence of IID RVs, and let S_n be the sum of first n of them. We know that if N is a non-negative integer valued RV that is independent of the sequence,

$$S_N = \sum_{i=1}^{N} X_i$$

is said to be a *compound RV* and the distribution of N is said to be the *compounding distribution*.

An Identity for Compound RVs

Let M be an RV that is independent of the sequence and such that

$$P\{M = n\} = \frac{nP\{N = n\}}{E(N)}, \quad n = 1, 2, \ldots$$

Proposition

We have, for any function h,

$$E[S_N h(S_N)] = E(N)E[X_1 h(S_M)]$$

An Identity for Compound RVs

Corollary

$$P\{S_N = 0\} = P\{N = 0\}$$

$$P\{S_N = k\} = \frac{1}{k}E(N)\sum_{j=1}^{k} j\alpha_j P\{S_{M-1} = k - j\}, \quad k > 0$$

Poisson Compounding Distribution

Let N be a Poisson RV with mean λ.

$$\begin{aligned} P\{M-1=n\} &= P\{M=n+1\} \\ &= \frac{(n+1)P\{N=n+1\}}{E(N)} \\ &= \frac{1}{\lambda}(n+1)e^{-\lambda}\frac{\lambda^{n+1}}{(n+1)!} \\ &= e^{-\lambda}\frac{\lambda^n}{n!} \end{aligned}$$

Since $M-1$ is also Poisson with mean λ, with $P\{S_N=n\}$,

$$P_0 = e^{-\lambda} \quad \text{and} \quad P_k = \frac{k}{\lambda}\sum_{j=1}^{k} j\alpha_j P_{k-j}, \;\; k>0$$

Binomial Compounding Distribution

If N is a binomial RV with parameters r and p,

$$\begin{aligned} P\{M-1=n\} &= \frac{(n+1)P\{N=n+1\}}{E(N)} \\ &= \frac{n+1}{rp}\binom{r}{n+1}p^{n+1}(1-p)^{r-n-1} \\ &= \frac{n+1}{rp}\frac{r!}{(r-1-n)!(n+1)!}p^{n+1}(1-p)^{r-n-1} \\ &= \frac{(r-1)!}{(r-1-n)!n!}p^n(1-p)^{r-n-1} \sim \mathrm{B}\,(r-1,p) \end{aligned}$$

Let $N(r)$ be binomial with r and p and $P_r(k) = P\{S_{N(r)}=k\}$

$$P_r(0) = (1-p)^r \quad \text{and} \quad P_r(k) = \frac{rp}{k}\sum_{j=1}^{k} j\alpha_j P_{r-1}(k-j), \;\; k>0$$

Thanks. Questions?