Lisa Yan
CS109

Lecture Notes #14
October 14, 2020

# Conditional Expectation

Based on a chapter by Chris Piech

**Pre-recorded lecture**: Sections 1 and 2 (up to 2.2).
**In-lecture**: Sections 2.3 and 2.4.
**Not covered**: Section 2.5.

## 1 Conditional Distributions

Before we looked at conditional probabilities for events. Here we formally go over conditional probabilities for random variables. The equations for the discrete case is an intuitive extension of our understanding of conditional probability:

### *1.1 Discrete*

The conditional probability mass function (PMF) for the discrete case:

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

The conditional cumulative density function (CDF) for the discrete case:

$$F_{X|Y}(a|y) = P(X \leq a|Y = y) = \frac{\sum_{x \leq a} p_{X,Y}(x, y)}{p_Y(y)} = \sum_{x \leq a} p_{X|Y}(x|y)$$

### *1.2 Example: Web Server Requests, Redux*

Let's say we have two independent random Poisson variables for requests received at a web server in a day: $X$ = # requests from humans/day, $X \sim \text{Poi}(\lambda_1)$ and $Y$ = # requests from bots/day, $Y \sim \text{Poi}(\lambda_2)$. Since the convolution of Poisson random variables is also a Poisson we know that the total number of requests $(X + Y)$ is also a Poisson $(X + Y) \sim \text{Poi}(\lambda_1 + \lambda_2)$. What is the probability of having $k$ human requests on a particular day given that there were $n$ total requests?

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

# 2 Conditional Expectation

We have gotten to know a kind and gentle soul, conditional probability. And we know another funky fool, expectation. Let's get those two crazy kids to play together.

Let $X$ and $Y$ be jointly discrete random variables. We define the conditional expectation of $X$ given $Y = y$ to be:

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

## *2.1 Properties of Conditional Expectation*

Here are some helpful, intuitive properties of conditional expectation:

$$E[g(X)|Y = y] = \sum_x g(x) p_{X|Y}(x|y) \qquad \text{if X and Y are discrete}$$

$$E[\sum_{i=1}^{n} X_i|Y = y] = \sum_{i=1}^{n} E[X_i|Y = y]$$

## *2.2 Law of Total Expectation*

The law of total expectation states that: $E[E[X|Y]] = E[X]$.

What?! How is that a thing? Check out this proof:

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

## 2.3 Example: Conditional Dice Sums

You roll two 6-sided dice $D_1$ and $D_2$. Let $X = D_1 + D_2$ and let $Y$ = the value of $D_2$.

- What is $E[X|Y = 6]$?

$$E[X|Y = 6] = \sum_x xP(X = x|Y = 6)$$
$$= \left(\frac{1}{6}\right)(7 + 8 + 9 + 10 + 11 + 12) = \frac{57}{6} = 9.5,$$

  which makes intuitive sense since $6 + E[\text{value of } D_1] = 6 + 3.5$.

- What is $E[X|Y = y]$, where $y = 1, \ldots, 6$?

  Let $W$ = the value of $D_1$. Then $X = Y + W$, and $Y$ and $W$ are independent.

$$\begin{aligned}
E[X|Y = y] &= E[W + Y|Y = y] = E[W + y|Y = y] \\
&= y + E[W|Y = y] && (y \text{ is a constant with respect to} W) \\
&= y + \sum_w wP(W = w|Y = y) \\
&= y + \sum_w wP(W = w) && (W, Y \text{ are independent}) \\
&= y + 3.5
\end{aligned}$$

  Note that $E[X|Y = y]$ depends on the value $y$. In other words, $E[X|Y]$ is a function of the random variable $Y$.

## 2.4 Example: Recursive Code

Consider the following code with random numbers:

```
int Recurse() {
   int x = randomInt(1, 3);  // Equally likely values
   if (x == 1) return 3;
   else if (x == 2) return (5 + Recurse());
   else return (7 + Recurse());
}
```

Let $Y$ = value returned by “Recurse". What is $E[Y]$. In other words, what is the expected return value. Note that this is the exact same approach as calculating the expected run time.

$$E[Y] = E[Y|X = 1]P(X = 1) + E[Y|X = 2]P(X = 2) + E[Y|X = 3]P(X = 3)$$

First lets calculate each of the conditional expectations:

$$E[Y|X=1] = 3$$
$$E[Y|X=2] = E[5+Y] = 5+E[Y]$$
$$E[Y|X=3] = E[7+Y] = 7+E[Y]$$

Now we can plug those values into the equation. Note that the probability of X taking on 1, 2, or 3 is 1/3:

$$\begin{aligned} E[Y] &= E[Y|X=1]P(X=1) + E[Y|X=2]P(X=2) + E[Y|X=3]P(X=3) \\ &= 3(1/3) + (5+E[Y])(1/3) + (7+E[Y])(1/3) \\ &= 15 \end{aligned}$$

## *2.5 Example: Random Number of Random Variables*

Each of 100 people are equally likely to visit (or not visit) the website `BestJokesEver.com`. Each person who visits the website will spend a certain number of minutes on the website per day, distributed as $\text{Poi}(8)$. The number of people and the time that each person spends on the website are independent. Let $W$ be the time spent by all visitors on a given day. What is $E[W]$?

**Solution** : Let $X$ be the number of people of 100 who visit the website. $X \sim \text{Bin}(100, 0.5)$. Let $Y_i$ be the number of minutes spent per day by visitor $i$, from $i = 1, \ldots, X$, where $Y_i \sim \text{Poi}(8)$. Finally, define $W = \sum_{i=1}^{X} Y_i$.

$$E[W] = E\left[\sum_{i=1}^{X} Y_i\right] = E\left[E\left[\sum_{i=1}^{X} Y_i | X\right]\right]$$

For a given $X = x$, we know that for all $i$, $P(Y_i = y|X = x) = P(Y_i = y)$, because $Y_i$ and $X$ are independent. Therefore

$$E\left[\sum_{i=1}^{X} Y_i | X = x\right] = \sum_{i=1}^{x} E[Y_i|X = x] = \sum_{i=1}^{x} E[Y_i] = xE[Y_1]$$

and thus $E[\sum_{i=1}^{X} Y_i|X] = XE[Y_1]$, where we note that $E[Y_i] = E[Y_1]$ for all $i$. Again, since $Y_1$ and $X$ are independent,

$$E[W] = E[XE[Y_1]] = E[Y_1]E[X] = 8 \cdot 50 = 400.$$

## 2.6 Example: Hiring Software Engineers

You are interviewing $n$ software engineer candidates and will hire only 1 candidate. All orderings of candidates are equally likely. Right after each interview you must decide to hire or not hire. You can not go back on a decision. At any point in time you can know the relative ranking of the candidates you have already interviewed.

The strategy that we propose is that we interview the first $k$ candidates and reject them all. Then you hire the next candidate that is better than all of the first $k$ candidates. What is the probability that the best of all the $n$ candidates is hired for a particular choice of $k$? Let's denote that result $P_k(Best)$. Let $X$ be the position in the ordering of the best candidate:

$$\begin{aligned}
P_k(Best) &= \sum_{i=1}^{n} P_k(Best|X=i)P(X=i) \\
&= \frac{1}{n}\sum_{i=1}^{n} P_k(Best|X=i) && \text{since each position is equally likely}
\end{aligned}$$

What is $P_k(Best|X=i)$? if $i \leq k$ then the probability is 0 because the best candidate will be rejected without consideration. Sad times. Otherwise we will chose the best candidate, who is in position $i$, only if the best of the first $i-1$ candidates is among the first $k$ interviewed. If the best among the first $i-1$ is not among the first $k$, that candidate will be chosen over the true best. Since all orderings are equally likely the probability that the best among the $i-1$ candidates is in the first $k$ is:

$$\frac{k}{i-1} \qquad \text{if } i > k$$

Now we can plug this back into our original equation:

$$\begin{aligned}
P_k(Best) &= \frac{1}{n}\sum_{i=1}^{n} P_k(Best|X=i) \\
&= \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1} && \text{since we know } P_k(Best|X=i) \\
&\approx \frac{1}{n}\int_{i=k+1}^{n} \frac{k}{i-1} di && \text{By Riemann Sum approximation} \\
&= \frac{k}{n}\ln(i=1)\Big|_{k+1}^{n} = \frac{k}{n}\ln\frac{n-1}{k} \approx \frac{k}{n}\ln\frac{n}{k}
\end{aligned}$$

If we think of $P_k(Best) = \frac{k}{n}\ln\frac{n}{k}$ as a function of $k$ we can take find the value of $k$ that optimizes it by taking its derivative and setting it equal to 0. The optimal value of $k$ is $n/e$. Where $e$ is Euler's number.