Math 425
Introduction to Probability
Lecture 34

Kenneth Harris
kaharri@umich.edu

Department of Mathematics
University of Michigan

April 12, 2009

Definition – Discrete Case

☞ Let X and Y be jointly discrete random variables. We defined the conditional mass of X given that $Y = y$ (provided $\mathbf{P}\{Y = y\} > 0$) by

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

☞ Since $p_{X|Y}(x|y)$ is a probability, we can define the conditional expectation of X given $Y = y$ (provided $\mathbf{P}\{Y = y\} > 0$) by

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

Example

Example. Two tetrahedral die are thrown with equally likely face values $1, 1, 2, 3$. Let Y denote the value on the first die and X denote the sum. The following table provides the relevant probabilities:

		X = 2	3	4	5	6
	1	0.5	0.25	0.25		
Y	2		0.5	0.25	0.25	
	3			0.5	0.25	0.25

☞ Note that $E[X|Y = y]$ is a function of y.

$$\begin{aligned} E[X|Y = 1] &= 2(0.5) + 3(0.25) + 4(0.25) = 2.75 \\ E[X|Y = 2] &= 3(0.5) + 4(0.25) + 5(0.25) = 3.75 \\ E[X|Y = 3] &= 4(0.5) + 5(0.25) + 6(0.25) = 4.75 \end{aligned}$$

Theorem – Conditioning

☞ Expectation is related to conditional expectation in much the same way that probability is related to conditional probability.

Theorem (Conditioning)

Let X and Y be discrete random variables. Then

$$E[X] = \sum_y E[X|Y = y] \cdot p_Y(y).$$

Proof of Theorem

Proof.
We defined conditional expectation by

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

Multiply by $p_Y(y)$ and sum over y:

$$\begin{aligned}\sum_y E[X|Y=y] \cdot p_Y(y) &= \sum_y \sum_x x \cdot p_{X|Y}(x|y) \cdot p_Y(y) \\ &= \sum_y \sum_x x \cdot p_{X,Y}(x,y) \\ &= \sum_x x \cdot \sum_y p_{X,Y}(x,y) \\ &= \sum_x x \cdot p_X(x) \\ &= E[X].\end{aligned}$$

□

Example

Example. Two tetrahedral die are thrown with equally likely face values 1, 1, 2, 3. Let Y denote the value on the first die and X denote the sum.
The following table provides the relevant probabilities:

		X = 2	3	4	5	6
	1	0.5	0.25	0.25		
Y	2		0.5	0.25	0.25	
	3			0.5	0.25	0.25

☞ The expectation of X is

$$\begin{aligned}E[X] &= E[X|Y=1] \cdot \mathbf{P}\{Y=1\} + E[X|Y=2] \cdot \mathbf{P}\{Y=2\} \\ &\quad + E[X|Y=3] \cdot \mathbf{P}\{Y=3\} \\ &= 2.75(0.5) + 3.75(0.25) + 4.75(0.25) \\ &= 3.5\end{aligned}$$

Example – Expectation of Geometric Variable

Example. Let N be a geometric random variable with parameter p. Compute $E[N]$.

☞ Let Y be the event of success on the first trial. Condition on Y:

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

The geometric random variable is memoryless, so

$$E[N|Y=0] = 1 + E[N]$$

☞ Solving for $E[N]$,

$$E[N] = \frac{1}{p}$$

This agrees with the computation in Section 4.8.1.

Example – Variance of Geometric Variable

Example. Let N be a geometric random variable with parameter p. Compute $Var(N)$.

☞ Compute $E[N^2]$ by conditionalizing on Y, using

$$\begin{aligned}E[N^2|Y=1] &= 1 \\ E[N^2|Y=0] &= E\left[(1+N)^2\right]\end{aligned}$$

(The geometric distribution is memoryless for the second equality.)

$$\begin{aligned}E[N^2] &= E[N^2|Y=1] \cdot p_Y(1) + E[N^2|Y=0] \cdot p_Y(0) \\ &= 1 \cdot p + E\left[(1+N)^2\right] \cdot (1-p) \\ &= p + E[1 + 2N + N^2] \cdot (1-p) \\ &= p + (1-p) + 2(1-p) \cdot E[N] + (1-p) \cdot E[N^2] \\ &= 1 + \frac{2(1-p)}{p} + (1-p) \cdot E[N^2]\end{aligned}$$

Example – continued

$$E[N^2] = 1 + \frac{2(1-p)}{p} + (1-p)\cdot E[N^2]$$
$$= \frac{2-p}{p} + (1-p)\cdot E[N^2]$$

☞ Solving for $E[N^2]$,

$$E[N^2] = \frac{2-p}{p^2}$$

☞ Therefore,

$$Var(N) = E[N^2] - E[N]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}.$$

Example

Example

Consider the following dice game. A pair of dice are rolled. If the sum is 7, then the game ends and you win 0. If the sum is not 7, then you have the option of stopping the game and receiving any amount equal to the that sum or starting over again.

Consider the strategy of choosing a target value i and stopping whenever the roll is at least as big as i.

☞ What value of $i = 2, 3, \dots, 12$ leads to the largest expected winnings with this strategy?

Example – continued

☞ Let X_i denote the return when you use the critical value i.
Let Y be the value of the first roll. Condition on Y:

$$E[X_i] = E[X_i|Y=7]\cdot p_Y(7) + \sum_{j\geq i, j\neq 7} E[X_i|Y=j]\cdot p_Y(j) + \sum_{j<i, j\neq 7} E[X_i|Y=j]\cdot p_Y(j)$$
$$= 0\cdot p_Y(7) + \sum_{j\geq i, j\neq 7} j\cdot p_Y(j) + \sum_{j<i, j\neq 7} E[X_i]\cdot p_Y(j)$$

If $j < i, j \neq 7$, then we roll again, with the same expected outcome.
f $j \geq i, j \neq 7$ we win j, so

$$E[X_i|Y=j] = E[X_i] \quad \text{if } j < i, j \neq 7$$
$$E[X_i|Y=j] = j \quad \text{if } j \geq i, j \neq 7$$

Example – continued

$$E[X_i] = \sum_{j\geq i, j\neq 7} j\cdot p_Y(j) + E[X_i]\cdot \sum_{j<i, j\neq 7} p_Y(j)$$

☞ Solving for $E[X_i]$:

$$E[X_i] = \Big(\sum_{j\geq i, j\neq 7} j\cdot p_Y(j)\Big)\cdot\Big(1 - \sum_{j<i, j\neq 7} p_Y(j)\Big)^{-1}$$

Example – continued

$$E[X_i] \;=\; \Big(\sum_{j\geq i, j\neq 7} j\cdot p_Y(j)\Big)\cdot\Big(1-\sum_{j<i, j\neq 7} p_Y(j)\Big)^{-1}$$

☞ I have computed the expected values for all possible strategies X_i:

$E[X_2] \approx 5.83$ $E[X_3] \approx 5.94$ $E[X_4] \approx 6.12$
$E[X_5] \approx 6.33$ $E[X_6] \approx 6.54$ $E[X_7] \approx 0$
$E[X_8] \approx 6.67$ $E[X_9] \approx 6.25$ $E[X_{10}] \approx 5.33$
$E[X_{11}] \approx 3.4$ $E[X_{12}] \approx 1.5$

☞ The best strategy is to continue rolling until you roll 8 or higher.

Conditional Expectation and Independence

☞ If the random variables X and Y are independent, then for any y

$$E[X|Y=y] \;=\; E[X]$$

Reason.

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

Theorem

☞ The following is a special case of Wald's equation for computing the mean of a sum of a random number of random variables.

Theorem

Suppose that $X_1, X_2, \ldots$ have the same mean. Also suppose that N is a nonnegative, integer-valued random variable that is independent of the X_i's. Then

$$E\Big[\sum_{i=1}^{N} X_i\Big] = E[X_1]\cdot E[N].$$

Theorem

☞ Prove by conditioning on N.

$$\begin{aligned} E\Big[\sum_{i=1}^{N} X_i\Big] &= \sum_{n=1}^{\infty} E\Big[\sum_{i=1}^{N} X_i \Big| N=n\Big]\cdot p_N(n) \\ &= \sum_{n=1}^{\infty} E\Big[\sum_{i=1}^{n} X_i \Big| N=n\Big]\cdot p_N(n) \\ &= \sum_{n=1}^{\infty} E\Big[\sum_{i=1}^{n} X_i\Big]\cdot p_N(n) \qquad X_i\text{'s, N independent!!} \\ &= \sum_{n=1}^{\infty} nE[X_1]\cdot p_N(n) \\ &= E[X_1]\sum_{n=1}^{\infty} n\cdot p_N(n) \\ &= E[X_1]\cdot E[N]. \end{aligned}$$

Example

Example. Suppose the number of times N we roll a die is Poisson distributed with mean 10. What is the expected sum of the rolls?

☞ Let X_i denote the value of the ith roll. Since the X_i's are independent of N, we can use the previous theorem to compute the expected sum:

$$\begin{aligned} E[\sum_{i=1}^{N} X_i] &= E[X_1] \cdot E[N] \\ &= 3.5(10) \\ &= 35. \end{aligned}$$

See Example 7.5d in Ross for another example.

Example

Example

The coat check at the Bistro Acmé is thoroughly incompetent: he gives out the coats in his keeping at random.

A conference of n mathematicians has booked the Bistro for their annual banquet, and have now come to reclaim their coats. The coat check distributes the coats at random. Those who received their coat can leave, the others return the coat the received, and then the process begins again.

☞ What is the expected number of attempts before all n mathematicians have retrieved their own coat?

Example – continued

☞ Let X_i for $i = 1, 2, \ldots$ count the number of mathematicians who retrieve their own coat in attempt i. Let N count the number of attempts required before all n mathematicians have received their coat.

☞ Note that in this problem the X_i's are not independent of each other: the number of people remaining in attempt i is $n - \sum_{j<i} X_j$

☞ Nor are the X_i's independent for N:
If $i < N$, then not everyone attempting to retrieve their coat in attempt i can get their coat; if $i = N$, then everyone remaining must have received their coat; and if $i > N$ then $X_i = 0$.

Example – continued

☞ However, the expected value of X_i depends ONLY on their being at least one person who needs their coat:

$$\begin{aligned} E[X_i] &= E[X_i | N = n] \\ &= \begin{cases} 1 & \text{if } i \le n \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Reason. Recall the gift exchange example in Lecture 31.

Example – continued

☞ By definition of X_i and N: $n = \sum_{i=1}^{N} X_i$.

$$\begin{aligned} n &= E[\sum_{i=1}^{N} X_i] = \sum_{n=1}^{\infty} E[\sum_{i=1}^{N} X_i | N = n] \cdot p_N(n) \\ &= \sum_{n=1}^{\infty} E[\sum_{i=1}^{n} X_i | N = n] \cdot p_N(n) \\ &= \sum_{n=1}^{\infty} (\sum_{i=1}^{n} E[X_i | N = n]) \cdot p_N(n) \\ &= \sum_{n=1}^{\infty} (\sum_{i=1}^{n} E[X_i]) \cdot p_N(n) \\ &= \sum_{n=1}^{\infty} n \cdot p_N(n) = E[N]. \end{aligned}$$

So, the expected number of attempts is n.

Computing Probabilities by Conditioning

☞ Let A be some event, and let I_A be the indicator variable

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}$$

Recall that

$$E[I_A] = \mathbf{P}\{I_A = 1\} = \mathbf{P}\{A\}$$

☞ We can compute $\mathbf{P}\{A\}$ by conditioning on Y

$$\begin{aligned} \mathbf{P}\{A\} &= E[I_A] \\ &= \sum_y E[I_A | Y = y] \cdot p_Y(y) \\ &= \sum_y \mathbf{P}(A \mid Y = y) \cdot p_Y(y) \end{aligned}$$

Example

☞ The game show Choose!! has the following interesting payoff to the winner on a show:

> A contestant is offered a sum of money and can accept or reject it. If this sum is rejected, another offer is made. A total of n offers is made.
>
> The contestant does not know the maximum value to be offered, and the order of the offers are randomly determined before the show.

If the strategy is to maximize the likelihood of obtaining the highest sum offered, how well can a contestant expect to do?

Compare to Example 7.5.j in Ross, pp. 377-8.

Example – continued

☞ Fix k with $0 \leq k < n$ and consider the strategy

- Strategy k: reject the first k offer and then accept the first offer better than each of these first k.

Let A_k be the event that the highest sum offered is chosen under Strategy k.

☞ Let X be the position of the highest offer, so X is uniformly distributed over $1, \ldots, n$. Condition on $X = i$:

$$\begin{aligned} \mathbf{P}\{A_k\} &= \sum_{i=1}^{n} \mathbf{P}(A_K \mid X = i) \cdot \mathbf{P}\{X = i\} \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbf{P}(A_K \mid X = i) \\ &= \frac{1}{n} \sum_{i=k+1}^{n} \mathbf{P}(A_K \mid X = i) \end{aligned}$$

Reason. If the highest offer is in the first k, it will be rejected.

Example – continued

$$\mathbf{P}\{A_k\} = \frac{1}{n}\sum_{i=k+1}^{n} \mathbf{P}(A_K \mid X = i)$$

☞ We will only select the best offer at position $X = i$ if the best offer among the first $i-1$ occurs in one of positions $1, \ldots, k$.

The best offer in the first $i-1$ positions is equally likely to be at any position, and so when $i > k$

$$\mathbf{P}(A_K \mid X = i) = \mathbf{P}(\text{best of first } i-1 \text{ in first } k \mid X = i)$$

$$= \frac{k}{i-1}$$

Thus,

$$\mathbf{P}\{A_k\} = \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1}$$

Example – continued

☞ We approximate the sum we computed for $\mathbf{P}\{A_k\}$:

$$\begin{aligned}\mathbf{P}\{A_k\} &= \frac{1}{n}\sum_{i=k+1}^{n} \frac{k}{i-1} \\ &\approx \frac{k}{n}\int_{k+1}^{n} \frac{1}{x-1}\,dx \\ &= \frac{k}{n}\Big(\int_{n}^{\infty} \frac{1}{x-1}\,dx - \int_{k+1}^{\infty} \frac{1}{x-1}\,dx\Big) \\ &= \frac{k}{n}\log\big(\frac{n-1}{k}\big) \\ &\approx \frac{k}{n}\log\big(\frac{n}{k}\big) = -\frac{k}{n}\log\big(\frac{k}{n}\big)\end{aligned}$$

☞ So, $\mathbf{P}\{A_k\} \approx -\frac{k}{n}\log\big(\frac{k}{n}\big)$.

Example – continued

$$\mathbf{P}\{A_k\} \approx -\frac{k}{n}\log\big(\frac{k}{n}\big)$$

☞ Consider the function $g(x)$ replacing k with x:

$$\begin{aligned}g(x) &= -\frac{x}{n}\log\big(\frac{x}{n}\big) \\ g'(x) &= -\frac{1}{n}\log\big(\frac{x}{n}\big) - \frac{1}{n}\end{aligned}$$

The best we can hope to do is about where $g'(x) = 0$, or equivalently

$$x = \frac{n}{e}.$$

What bearing does this excursion have on our problem?

Example – continued

$$\mathbf{P}\{A_k\} \approx -\frac{k}{n}\log\big(\frac{k}{n}\big)$$

☞ When $k \approx \frac{n}{e}$ we have

$$\mathbf{P}\{A_k\} \approx \frac{1}{e} \approx 0.36788$$

So, there is about a 37% chance of getting the best prize if we watch $\frac{n}{e}$ prizes go by, then accepting the first prize larger than these.

Plots

☞ Plots of

$$g(x) = -\frac{x}{n}\log\left(\frac{x}{n}\right) \qquad 0 \le x \le n$$

for $n = 10, 30, 50, 70$ with maximum at

$$\frac{n}{e} \approx 3.7,\ 11.04,\ 18.4,\ 25.75$$