EE2012 2014/15 PROBLEM SET 5

Conditional PMF and Expected Values

1. Let $A = \{X \geq 3\}$. Find the conditional PMF $p_X(x|A)$ for the following random variables:

   (a) X is binomial with $n = 5$, $p = 0.2$.
   Ans: We have $p_X(k) = \binom{5}{k} 0.2^k 0.8^{5-k}$, $k = 0, 1, \ldots, 5$. Therefore,

   $$P[A] = \sum_{k=3}^{5} \binom{5}{k} 0.2^k 0.8^{n-k} = 0.05792. \tag{1}$$

   Then using the formula $p_X(k|A) = p_X(k)/P[A]$ for $k \in A$, we have

   $$p_X(k|A) = \begin{cases} 0.8840 & k = 3 \\ 0.1105 & k = 4 \\ 0.0055 & k = 5 \\ 0 & \text{elsewhere} \end{cases} \tag{2}$$

   (b) X is geometric with $p = 0.4$.
   Ans: In this case,

   $$P[A] = \sum_{k=3}^{\infty} 0.6^{k-1} 0.4 = 0.6^2 = 0.36,$$

   and

   $$p_X(k|A) = \frac{1}{0.36} 0.6^{k-1} 0.4 = 1.852(0.6^k), \tag{3}$$

   $k = 3, 4, \ldots$

   (c) X is Poisson with $E[X] = 2$.
   Ans: Now we have

   $$P[A] = 1 - \sum_{k=0}^{2} \frac{2^k}{k!} e^{-2} = 0.6767. \tag{4}$$

   Therefore,

   $$p_X(k|A) = \frac{1}{0.6767} \frac{2^k}{k!} e^{-2} = 0.2 \frac{2^k}{k!}, \tag{5}$$

   $k = 3, 4, \ldots$

2. Using the conditional PMFs of Question 1, find $P[X = 4 \text{ or } 5|A]$ in all three cases.
   Ans: Just compute $p_X(4|A) + p_X(5|A)$ in all cases.

3. Suppose two bus services connect two locations A and B. Bus service 1 takes on average 30 minutes to make the journey from A to B, and service 2 takes on average 20 minutes to make the same journey (because it uses a different route). Assuming that the probability of Service 1 arriving before Service 2 is 2/3, find the travel time from A to B averaged over a large number of bus rides.

   Ans: Let the travel time from A to B be T minutes, and let event A_i represent "Bus service i is used". Then $E[T|A_1] = 30$ and $E[T|A_2] = 20$. We are also told that $P[A_1] = \frac{2}{3}$ and $P[A_2] = \frac{1}{3}$. Therefore,

$$E[T] = E[T|A_1]P[A_1] + E[T|A_2]P[A_2] \tag{6}$$
$$= 26\frac{2}{3}. \tag{7}$$

   This can be interpreted as the travel time averaged over many days of making the same trip, some of which are on service 1 and some on service 2.

4. A store sells only three models of laptop, A, B and C. The lifetime X of the laptop (in years, rounded to the nearest integer) has the following conditional geometric distributions:

$$p_X(k|\text{"Model A"}) = 0.8^{k-1}0.2 \tag{8}$$
$$p_X(k|\text{"Model B"}) = 0.75^{k-1}0.25 \tag{9}$$
$$p_X(k|\text{"Model C"}) = 0.7^{k-1}0.3 \tag{10}$$

   where $k = 1, 2, \ldots$ Laptop A costs \$2,000, laptop B costs \$1,800 and laptop C costs \$1,600.

   (a) Based on their prices, 50% of laptop customers opt for Model C, 30% for Model B, and 20% for Model A. (The statistical lifetimes of these models are not known to the public.) Find the average lifetime of a laptop sold by this store.

   Ans: With the above geometric distributions, we have $E[X|\text{"Model A"}] = 5$, $E[X|\text{"Model B"}] = 4$ and $E[X|\text{"Model C"}] = 3.33$. The probabilities of a customer buying Models A, B and C are 0.2, 0.3 and 0.5, respectively. Therefore, the total probability rule gives us

$$E[X] = 5(0.2) + 4(0.3) + 3.33(0.5) = 3.87. \tag{11}$$

   In other words, the average lifetime of a laptop is 3.87 years.

   (b) Find the probability that a laptop bought from this store lasts longer than 2.5 years.

   Ans: The event "lifetime longer than 2.5 years" is equivalent to $\{X \geq 3\}$ (remember the rounding to the nearest year). From the geometric conditional

PMFs, we have

$$P[X \geq 3|\text{"Model A"}] = 0.8^2 = 0.64 \quad (12)$$
$$P[X \geq 3|\text{"Model B"}] = 0.75^2 = 0.5625 \quad (13)$$
$$P[X \geq 3|\text{"Model C"}] = 0.7^2 = 0.49. \quad (14)$$

By total probability, we have

$$P[X \geq 3] = 0.64(0.2) + 0.5625(0.3) + 0.49(0.5) = 0.54.$$

(c) If a laptop lasts longer than 2.5 years, find the probability that it was Model A.

Ans: We use Bayes Rule to obtain

$$P[\text{"Model A"}|X \geq 3] = \frac{0.64 \times 0.2}{0.54} \quad (15)$$
$$= 0.2363. \quad (16)$$

Notice that this value (0.2363) is larger than the original probability of choosing model A (0.2). This is due to the new information that the lifetime of the laptop was observed to be longer than 2.5 years, combined with the fact that Model A has a longer lifetime (on average) than the other two models. If instead we observed that the lifetime was longer than 4.5 years, how would the answer change?

5. A student has only 10 minutes left to complete two questions in an exam, questions 9 and 10. Question 9 is worth 5 marks, and Question 10 is worth 8 marks. If Q9 is answered first (denoted event A), the probability of getting Q9 correct and Q10 wrong (denoted C_9W_{10}) is 0.6, i.e. $P[C_9W_{10}|A] = 0.6$. Conditioned on A the probability of all the other outcomes are

$$P[W_9W_{10}|A] = 0.1 \qquad P[C_9C_{10}|A] = 0.1$$
$$P[W_9C_{10}|A] = 0.2.$$

Similarly, conditioned on A^c (i.e. he answers Q10 first), we have

$$P[W_9C_{10}|A^c] = 0.5 \qquad P[C_9C_{10}|A^c] = 0.05$$
$$P[W_9W_{10}|A^c] = 0.25 \qquad P[C_9W_{10}|A^c] = 0.2.$$

Let X be the total marks scored in Q9 and Q10.

(a) Find $E[X|A]$ and $E[X|A^c]$. Which question should he attempt first, in order to maximize his expected marks from Q9 and Q10?

Ans: We have $S_X = \{0, 5, 8, 13\}$, with conditional PMFs

$$p_X(0|A) = 0.1,\ p_X(5|A) = 0.6, \qquad p_X(8|A) = 0.2,\ p_X(13|A) = 0.1$$
$$p_X(0|A^c) = 0.25,\ p_X(5|A^c) = 0.2, \qquad p_X(8|A^c) = 0.5,\ p_X(13|A^c) = 0.05.$$

Therefore, $E[X|A] = 5(0.6) + 8(0.2) + 13(0.1) = 5.9$, and $E[X|A^c] = 5(0.2) + 8(0.5) + 13(0.05) = 5.65$, and he should choose strategy A, i.e. answer Q9 first, in order to maximize his expected score.

(b) Find $P[X \geq 8|A]$ and $P[X \geq 8|A^c]$. To maximize the probability of scoring 8 or more marks, which question should he do first?

Ans: From the conditional PMFs in part (a), we can compute

$$\begin{aligned} P[X \geq 8|A] &= 0.3 && (17) \\ P[X \geq 8|A^c] &= 0.55. && (18) \end{aligned}$$

Therefore, in order to maximize the probability of scoring 8 or more marks, he should answer Q10 first.

(c) The student has no time to perform such complicated calculations, and chooses strategy A with probability p, and strategy A^c with probability $1-p$, where $p \notin \{0,1\}$. Show that A and C_9 are not independent.

Ans: We can calculate

$$\begin{aligned} P[C_9|A] &= P[C_9W_{10}|A] + P[C_9C_{10}|A] = 0.7 && (19) \\ P[C_9|A^c] &= P[C_9W_{10}|A^c] + P[C_9C_{10}|A^c] = 0.25. && (20) \end{aligned}$$

For independence between A and C_9, where both have non-zero probabilities, we need $P[C_9] = P[C_9|A] = P[C_9|A^c]$. Since $P[C_9|A] \neq P[C_9|A^c]$, A and C_9 cannot be independent.

Important Discrete Random Variables

1. Let N be a geometric random variable with $E[N] = 1/p$.

(a) Find $P[N = k|N \leq m]$.

Ans: From the definition of conditional probability,

$$P[N = k|N \leq m] = \frac{P[N = k, N \leq m]}{P[N \leq m]}.$$

We can compute $P[N \leq m] = \sum_{k=1}^{m} q^{k-1}p = 1 - q^m$. Since $\{N = k\} \cap \{N \leq m\} = \{N = k\}$ when $k \leq m$, and $\emptyset$ otherwise, we have

$$P[N = k|N \leq m] = \frac{q^{k-1}p}{1-q^m}, \quad k = 1, 2, \ldots, m. \qquad (21)$$

(b) Find the probability that N is even.

Ans: We just sum the PMF over the even values of k to get

$$\begin{aligned} P[\text{“}X \text{ is even”}] &= \sum_{k=1}^{\infty} q^{2k-1} p && (22) \\ &= \frac{p}{q} \frac{q^2}{1-q^2} && (23) \\ &= \frac{pq}{1-q^2} = \frac{q}{1+q}. && (24) \end{aligned}$$

2. The number of page requests that arrive at a web server is a Poisson random variable with an average of 3,000 requests per minute.

   (a) Find the probability that there are no requests in a 100 ms period.

   Ans: If 3,000 requests arrive per minute on average, then in 100 ms, the average number of requests would be $3000/600 = 5$. The PMF of X, the number of requests in 100 ms, is thus

   $$p_X(k) = \frac{5^k}{k!} e^{-5}, \quad k = 0, 1, 2 \ldots$$

   Therefore, $p_X(0) = e^{-5} = 0.0067$.

   (b) Find the probability that there are between 5 and 10 requests in a 100 ms period.

   Ans: We calculate the sum $\sum_{k=5}^{10} p_X(k)$ using a calculator or computer to arrive at the value 0.5458.

3. An LCD display has 2000×1000 pixels. A display is accepted if it has 15 or fewer faulty pixels. The probability that a pixel is faulty coming out of the production line is 5×10^{-6}. Find the proportion of displays that are accepted.

   Ans: The number of faulty pixels follows a binomial distribution, with $n = 2 \times 10^6$, and $p = 5 \times 10^{-6}$. We can use a Poisson distribution to approximate the binomial in this case, since $np = 10$ is a small value and n is large. Therefore we have

   $$\begin{aligned} P[\text{“display accepted”}] &= \sum_{k=0}^{15} \frac{10^k}{k!} e^{-10} && (25) \\ &= 0.9513. && (26) \end{aligned}$$

4. A random variable X is uniformly distributed in $\{-3, -2, \ldots, 3, 4\}$.

   (a) Find the mean and variance of X.

   Ans: The mean of X is the mid-point of the distribution, i.e. $E[X] = (-3 + 4)/2 = 0.5$. The variance can be computed as (see the notes)

   $$\sigma_X^2 = \frac{L^2 - 1}{12} = \frac{63}{12} = 5.25 \qquad (27)$$

   where $L = 8$ is the number of consecutive integers inside the range of X.

(b) Find the mean and variance of $Y = -2X^2 + 3$.

Ans: Using the linearity property of the expectation operator,

$$E[Y] = -2E[X^2] + 3. \tag{28}$$

But $E[X^2] = \sigma_X^2 + \mu_X^2 = 5.25 + 0.25 = 5.5$, and thus

$$E[Y] = -2(5.5) + 3 = -8.$$

To find the variance of Y we need

$$\begin{aligned} E[Y^2] &= E[4X^4 - 12X^2 + 9] && (29)\\ &= 4E[X^4] - 12E[X^2] + 9. && (30) \end{aligned}$$

By definition,

$$E[X^4] = \frac{1}{8}\sum_{k=-3}^{4} k^4 = 56.5, \tag{31}$$

and therefore

$$E[Y^2] = 4(56.5) - 12(5.5) + 9 = 169, \tag{32}$$

so that $\sigma_Y^2 = 169 - (-8)^2 = 105$.

Cumulative Distribution Function

1. The CDF of X is given by

$$F_X(x) = \begin{cases} 0 & x < -2 \\ 0.5 & -2 \le x \le 0 \\ (2+x)/4 & 0 \le x \le 2 \\ 1 & x > 2 \end{cases} \tag{33}$$

(a) Sketch the CDF and identify the type of random variable X is.

Ans: X is a mixed random variable due to the step discontinuity in the CDF at $x = -2$.

(b) Use the CDF to find $P[-1 < X \le 1]$ and $P[X > 0]$.

Ans: $P[-1 < X \le 1] = P[X \le 1] - P[X \le -1] = F_X(1) - F_X(-1) = 0.75 - 0.5 = 0.25$.

$P[X > 0] = 1 - P[X \le 0] = 1 - F_X(0) = 0.5$.

2. Let ζ be a value selected at random from the unit interval $[0, 1)$, and define $X(\zeta) = (1-\zeta)^{-1/2}$.

(a) Find the range of X.

Ans: Since $\zeta \ge 0$, $1 - \zeta \le 1$, and therefore $(1-\zeta)^{-1/2} \ge 1$. Since the largest value of ζ is 1, which makes $1 - \zeta = 0$, the largest value of X is ∞. Therefore, $S_X = [1, \infty)$.

(b) Find and sketch the CDF of X.

Ans: The event $\{X \leq x\}$ is equivalent to

$$(1-\zeta)^{-1/2} \leq x \Rightarrow \zeta \leq 1 - \frac{1}{x^2}.$$

But the probability of $\zeta \leq z$ is z, for $z \in [0,1)$. Therefore, for $x \in [1,\infty)$, we have

$$F_X(x) = P[X \leq x] = 1 - \frac{1}{x^2}.$$

When $x < 1$, then $F_X(x) = 0$ since in this case $\{X \leq x\} = \emptyset$.

(c) Find the probability of the events $\{X > 1\}$, $\{5 < X < 7\}$ and $\{X \leq 20\}$.

Ans: $P[X > 1] = 1 - F_X(1) = 1$. $P[5 < X < 7] = F_X(7) - F_X(5) - P[X = 7] = \frac{1}{25} - \frac{1}{49}$ (note that $P[X = 7] = 0$ since the CDF is continuous at $x = 7$). $P[X \leq 20] = F_X(20) = \frac{399}{400}$.

3. A point is selected at random inside a square defined by $\{(x,y) : 0 \leq x \leq b, 0 \leq y \leq b\}$. Assume the point is equally likely to fall anywhere in the square. Let the random variable Z be defined as the minimum of the two coordinates of the point where the dart lands.

(a) Find the range of Z.

Ans: $S_Z = [0,b]$ because when $(x,y) = (0,0)$, then $Z = 0$, and when $(x,y) = (b,b)$, then $Z = b$. All other values in between are also possible.

(b) Find and sketch the CDF of Z.

Ans: The event $\{Z \leq z\}$, $0 \leq z \leq b$, is equivalent to $\{0 \leq x \leq b, 0 \leq y \leq z\} \cup \{0 \leq x \leq z, 0 \leq y \leq b\}$. The area of this region is $A = 2bz - z^2$, and since the area of the whole square region is b^2, then the probability of (x,y) falling in this region is

$$F_Z(z) = \frac{A}{b^2} = \frac{2bz - z^2}{b^2}.$$

For $z < 0$, $F_Z(z) = 0$, and for $z > b$, $F_Z(z) = 1$. The CDF of Z is therefore

$$F_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{2bz - z^2}{b^2} & 0 \leq z \leq b \\ 1 & z > b \end{cases}.$$

(c) Use the CDF to find the probabilities of $\{Z > 0\}$, $\{Z \leq b/2\}$ and $\{Z > b/4\}$.

Ans: $P[Z > 0] = 1 - F_Z(0) = 1$. $P[Z \leq b/2] = F_Z(b/2) = 3/4$ and $P[Z > b/4] = 1 - F_Z(b/4) = 7/16$.