Weekly Notes for EE2012 2014/15 – Week 5

T. J. Lim

February 6, 2015

Book sections covered this week: 3.4

1 Conditional Distributions

We previously introduced the concept of the conditional probability of an event A given the occurrence of another event B, defined on the same probability space. This is denoted $P(A|B)$ and is defined as $P(A \cap B)/P(B)$. If A is now the event $\{X = x\}$, where X is a random variable, then $P(A|B)$ becomes a conditional probability mass function (pmf).

1.1 Conditional PMF

By definition,

$$p_X(x|B) = P[X = x|B] = \frac{P[\{X = x\} \cap B]}{P[B]}. \tag{1}$$

It can be verified that $p_X(x|B)$ is a valid pmf over S_X, the range of X:

- Note that the events $\{X = x\}$, $x \in S_X$, form a partition of the sample space, and hence
$$\sum_{x \in S_X} P[\{X = x\} \cap B] = P[B].$$
Substituting this into (1) yields $\sum_{x \in S_X} p_X(x|B) = 1$.
- Since conditional probabilities are probability values, it must hold that $0 \leq p_X(x|B) \leq 1$.

Example 1: Toss 5 fair coins, and let X be the number of Heads obtained. Let $B =$ "1st toss is a Head", and find $p_X(k|B)$.

Using the definition (1), we have e.g.

$$p_X(2|B) = \frac{P[\{\text{HHTTT,HTHTT,HTTHT,HTTTH}\}]}{P[B]} = \frac{1}{4}. \tag{2}$$

In fact, we realize that

$$\begin{aligned} P[\{X = k\} \cap B] &= P[k-1 \text{ Heads in tosses 2 to 5 and 1st toss is H}] \\ &= P[k-1 \text{ Heads in tosses 2 to 5}]P[\text{1st toss is H}] \\ &= \binom{4}{k-1}\frac{1}{16} \cdot \frac{1}{2} \end{aligned}$$

where the second line comes from noting that tosses 2 to 5 are independent of toss 1, and so

$$p_X(k|B) = \binom{4}{k-1}\frac{1}{16}, \quad k = 1, 2, 3, 4, 5. \tag{3}$$

This is not the same as $p_X(k)$.

Another way to solve the problem is to observe that

$$p_X(k|B) = P[k-1 \text{ Heads in tosses 2 to 5}]$$

and then immediately arrive at the result above.

The change from $p_X(k)$ to $p_X(k|B)$ comes from the fact that, if we know or hypothesize that B occurred, it would be impossible for $\{X = 0\}$ to have simultaneously occurred. Therefore, we must set $p_X(0|B)$ to 0, and this is exactly what appears in the solution. ■

Some important results related to conditional PMFs are:

1. If the event B is one involving X directly, i.e. the conditioning event can be written as $\{X \in B\}$, then[1]

$$p_X(k|B) = \begin{cases} \frac{p_X(k)}{P[B]}, & k \in B \\ 0 & \text{otherwise} \end{cases}. \tag{4}$$

   In other words, the probability mass at all values outside of the set B are set to zero, and the remaining non-zero values are normalized so they sum to unity. This is logical because the event conditioned upon gives us information only about the *values* of X, and thus should not change the relative likelihoods of seeing the values that have non-zero probability of appearing.

2. If $B_1, \ldots, B_n$ form a partition of the sample space, then the theorem on total probability yields:

$$p_X(k) = \sum_{i=1}^{n} p_X(k|B_i)P[B_i] \tag{5}$$

   This expression is useful because in most cases of practical interest, conditional probabilities are easier to find than non-conditional ones.

Example 2: The height and gender of a group of students are measured, and we define the r.v. X as follows:

$$X(\zeta) = \begin{cases} 0 & \text{if height of } \zeta < 165 \text{ cm} \\ 1 & \text{if height of } \zeta \in [165, 175] \text{ cm} \\ 2 & \text{if height of } \zeta > 175\text{cm} \end{cases} \tag{6}$$

For boys, the PMF of X is $p_X(0) = 0.1$, $p_X(1) = 0.6$ and $p_X(2) = 0.3$; for girls, it is $p_X(0) = 0.5$, $p_X(1) = 0.4$, $p_X(2) = 0.1$. Find the overall PMF of X.

[1]Note that B is now taken as a subset of S_X, the range of X, rather than a subset of S, the underlying sample space. In dealing with random variables, we prefer this more convenient notion of S_X as an effective sample space, so that we can say that event B occurs if and only $X \in B$, as opposed to if and only if $\zeta \in B'$ where B' is a mapping from the set of X values given in B to the corresponding set of ζ values within S.

Ans: What we are given are in fact conditional PMFs. Let A be the event that "selected student is a girl". Then the above information can be represented as

$$p_X(0|A) = 0.5, \quad p_X(1|A) = 0.4, \quad p_X(2|A) = 0.1 \tag{7}$$
$$p_X(0|A^c) = 0.1, \quad p_X(1|A^c) = 0.6, \quad p_X(2|A^c) = 0.3. \tag{8}$$

The total probability theorem, assuming $P[A] = P[A^c] = 0.5$, then yields

$$p_X(0) = p_X(0|A)P[A] + p_X(0|A^c)P[A^c] = 0.3 \tag{9}$$
$$p_X(1) = p_X(1|A)P[A] + p_X(1|A^c)P[A^c] = 0.5 \tag{10}$$
$$p_X(2) = p_X(2|A)P[A] + p_X(2|A^c)P[A^c] = 0.2. \tag{11}$$

In other words, 30% of all students are below 165 cm in height, while 50% of all girls and 10% of all boys are below 165 cm in height. ■

Example 3: Consider a geometric random variable X with parameter p, i.e.

$$p_X(k) = (1-p)^{k-1}p, \quad k = 1, 2, \ldots$$

Find the conditional PMF of X conditioned on $B = \{X \leq N\}$, where N is a positive constant.

Ans: From first principles, we have

$$p_X(k|B) = \frac{P[\{X = k\} \cap \{X \leq N\}]}{P[\{X \leq N\}]} \tag{12}$$
$$= \begin{cases} \frac{P[X=k]}{P[X \leq N]} & \text{if } k \leq N \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

because if $k \leq N$, then $\{X = k\} \cap \{X \leq N\} = \{X = k\}$, otherwise there is no overlap between the two ranges and hence their intersection is empty.

Finally, since

$$P[X \leq N] = \sum_{k=1}^{N} (1-p)^{k-1}p = 1 - q^N,$$

where $q = 1 - p$, we have

$$p_X(k|B) = \begin{cases} \dfrac{q^{k-1}p}{1-q^N} & \text{if } k = 1, 2, \ldots, N \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

Note that (13) could have been derived from the formula (4), but it is nearly as easy to derive it from first principles. ■

1.2 Conditional Mean (or Expected Values)

From the conditional PMF, we can define the conditional mean as

$$E[X|B] = \sum_{x \in S_X} x p_X(x|B) \tag{15}$$

where B is any non-zero probability event. We often also use the notation $\mu_{X|B} = E[X|B]$. It is important to note that X is a random variable, while B is an event even though both are denoted by upper-case Roman letters.

Since $p_X(x|B)$ is just another PMF for X, it will generally have a variance as well as a mean. We define the conditional variance as follows:

$$\sigma^2_{X|B} = E\left[(X - \mu_{X|B})^2|B\right] \tag{16}$$

with the expectation taken with respect to the *conditional* PMF $p_X(x|B)$, and not the original PMF $p_X(x)$, i.e.

$$\sigma^2_{X|B} = \sum_{x \in S_X} (x - \mu_{X|B})^2 p_X(x|B). \tag{17}$$

A careful expansion of terms on the RHS above yields

$$\begin{aligned}
\sigma^2_{X|B} &= \sum_{x \in S_X} (x^2 - 2x\mu_{X|B} + \mu^2_{X|B}) p_X(x|B) && (18)\\
&= \sum_x x^2 p_X(x|B) - 2\mu_{X|B} \sum_x x p_X(x|B) + \mu^2_{X|B} && (19)\\
&= E[X^2|B] - 2\mu^2_{X|B} + \mu^2_{X|B} && (20)\\
&= E[X^2|B] - \mu^2_{X|B}. && (21)
\end{aligned}$$

This is identical to the expression for $\mathrm{var}(X)$ except that all expectations are now conditioned on B.

Note that the conditional variance is an important special case of the conditional mean of a function of X, i.e.

$$E[g(X)|B] = \sum_{x \in S_X} g(x) p_X(x|B). \tag{22}$$

Using Total Probability to Find the Mean Value From (22), and the Total Probability result $p_X(x) = \sum_i p_X(x|B_i)P(B_i)$ where $\cup_i B_i = S$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$, we have

$$\begin{aligned}
\sum_i E[g(X)|B_i]P(B_i) &= \sum_i \sum_x g(x) p_X(x|B_i) P(B_i) && (23)\\
&= \sum_x g(x) \sum_i \sum_i p_X(x|B_i) P(B_i) && (24)\\
&= \sum_x g(x) p_X(x) = E[g(X)]. && (25)
\end{aligned}$$

This is an important result as it can be used whenever we can partition the sample space in such a way that the conditional expectations are known, as shown in the following examples.

Example 4: There are exactly three bus routes that connect point A and point B. Bus 1 takes an average of 30 minutes to go from A to B; Bus 2 takes an average of 25 minutes; and Bus 3 takes an average of 35 minutes. Assume that with probability

0.2, 0.3 and 0.5, a commuter uses Bus 1, Bus 2 and Bus 3 respectively. Find the average travel time from A to B.

Ans: Let X be the travel time from A to B in minutes, and B_i be the event that Bus i is used. Then the given information can be written as $E[X|B_1] = 30$, $E[X|B_2] = 25$ and $E[X|B_3] = 35$; also $P(B_1) = 0.2$, $P(B_2) = 0.3$ and $P(B_3) = 0.5$.

From (25), we have

$$\begin{aligned} E[X] &= E[X|B_1]P(B_1) + E[X|B_2]P(B_2) + E[X|B_3]P(B_3) \\ &= 30(0.2) + 25(0.3) + 35(0.5) \\ &= 31. \end{aligned}$$

So an average of 31 minutes is needed to travel from A to B by bus. ■

Example 5: Show using (21) that $\text{var}(X) \neq \sum_i \sigma^2_{X|B_i} P(B_i)$. What is the right way to find σ^2_X from $\mu_{X|B_i}$ and $\sigma^2_{X|B_i}$, $i = 1, \ldots, N$?

Ans: The problem here is that $\sigma^2_X = E[(X - \mu_X)^2]$, whereas $\sigma^2_{X|B_i} = E[(X - \mu^2_{X|B_i})|B_i]$. In general, $\mu_{X|B_i} \neq \mu_X$, and therefore $\sigma^2_X \neq \sum_i \sigma^2_{X|B_i} P(B_i)$.

Instead, if we know the conditional mean and conditional variance for a certain partition of the sample space, then we should compute

$$\mu_X = \sum_i \mu_{X|B_i} P(B_i) \tag{26}$$

$$E[X^2|B_i] = \sigma^2_{X|B_i} + \mu^2_{X|B_i} \tag{27}$$

$$E[X^2] = \sum_i E[X^2|B_i] P(B_i) \tag{28}$$

$$\Rightarrow \quad \sigma^2_X = E[X^2] - \mu^2_X. \tag{29}$$

This example illustrates a common misconception – that a random variable with a small variance in every possible situation (partition) must have a small overall variance. As we show in the next example, this cannot be true because for a discrete r.v. X, we can always design partitions with zero conditional variance. ■

Example 6: Let $p_X(k) = 0.5$, $k = \pm 1$. Partition the sample space into $B_1 = \{X \leq 0\}$ and $B_2 = \{X > 0\}$. Then it should be clear that $p_X(k|B_1) = \delta(k+1)$ and $p_X(k|B_2) = \delta(k-1)$, where $\delta(k)$ is the Kronecker delta function[2]. We then have $\sigma_{X|B_1} = \sigma_{X|B_2} = 0$. But clearly $\sigma^2_X = 1 \neq 0 = \sigma^2_{X|B_1} P(B_1) + \sigma^2_{X|B_2} P(B_2)$. This illustrates the result given in Example 5. ■

2 Diagnostic Questions

1. If X has the PMF $p_X(k) = \frac{1}{10}$, $k = 0, 1, \ldots, 9$, find $p_X(k|X > 3)$ and $p_X(k|X < 5)$.

2. What is the smallest possible value of $E[X|X \geq 3]$? Give an example of a PMF which attains this minimal conditional mean.

[2] $\delta(0) = 1$, $\delta(n) = 0$ for all $n \neq 0$.

3. An unfair coin is tossed, with Heads appearing with probability p. If the coin flip is Heads, X is a geometric random variable with probability of success 0.25; otherwise, X is a binomial random variable with $n = 10$ and $p = 0.2$. Find $E[X]$ and $\mathrm{var}(X)$ using conditional means and variances.