Chapter 8

Conditional Probability Mass Functions

8.1 Introduction

In Chapter 4 we discussed the concept of conditional probability. We recall that a conditional probability $P[A|B]$ is the probability of an event A, given that we know that some other event B has occurred. Except for the case when the two events are independent of each other, the knowledge that B has occurred will change the probability $P[A]$. In other words, $P[A|B]$ is our new probability in light of the additional knowledge. In many practical situations, two random mechanisms are at work and are described by events A and B. An example of such a compound experiment was given in Example 4.2. To compute probabilities for a compound experiment it is usually convenient to use a conditioning argument to simplify the reasoning. For example, say we choose one of two coins and toss it 4 times. We might inquire as to the probability of observing 2 or more heads. However, this probability will depend upon which coin was chosen, as for example in the situation where one coin is fair and the other coin is weighted. It is therefore convenient to define *conditional probability mass functions*, $p_X[k|\text{coin 1 chosen}]$ and $p_X[k|\text{coin 2 chosen}]$, since once we know which coin is chosen, we can easily specify the PMF. In particular, for this example the conditional PMF is a binomial one whose value of p depends upon which coin is chosen and with k denoting the number of heads (see (5.6)). Once the conditional PMFs are known, we have by the law of total probability (see (4.4)) that the probability of observing k heads for this experiment is given by the PMF

$$\begin{aligned} p_X[k] &= p_X[k|\text{coin 1 chosen}]P[\text{coin 1 chosen}] \\ &\quad + p_X[k|\text{coin 2 chosen}]P[\text{coin 2 chosen}]. \end{aligned}$$

Therefore, the desired probability of observing 2 or more heads is

$$\begin{aligned} P[X \geq 2] &= \sum_{k=2}^{4} p_X[k] \\ &= \sum_{k=2}^{4} (p_X[k|\text{coin 1 chosen}]P[\text{coin 1 chosen}] \\ &\qquad + p_X[k|\text{coin 2 chosen}]P[\text{coin 2 chosen}]). \end{aligned}$$

The PMF that is required depends directly on the conditional PMFs (of which there are two). The use of conditional PMFs greatly simplifies our task in that *given* the event, i.e., the coin chosen, the PMF of the number of heads observed readily follows. Also, in many problems, including this one, it is actually the conditional PMFs that are specified in the description of the experimental procedure. It makes sense, therefore, to define a conditional PMF and study its properties. For the most part, the definitions and properties will mirror those of the conditional probabillity $P[A|B]$, where A and B are events defined on $\mathcal{S}_{X,Y}$.

8.2 Summary

The utility of defining a conditional PMF is illustrated in Section 8.3. It is especially appropriate when the experiment is a compound one, in which the second part of the experiment depends upon the outcome of the first part. The definition of the conditional PMF is given in (8.7). It has the usual properties of a PMF, that of being between 0 and 1 and also summing to one. Its properties and relationships are summarized by Properties 8.1–8.5. The conditional PMF is related to the joint PMF and the marginal PMFs by these properties. They are also depicted in Figure 8.4 for easy reference. If the random variables are independent, then the conditional PMF reduces to the usual marginal PMF as shown in (8.22). For general probability calculations based on the conditional PMF one can use (8.23). In Section 8.5 it is shown how to use conditioning arguments to simplify the derivation of the PMF for $Z = g(X,Y)$. The PMF can be found using (8.24), which makes use of the conditional PMF. In particular, if X and Y are independent, the procedure is especially simplified with examples given in Section 8.5. The mean of the conditional PMF is defined by (8.30). It is computed by the usual procedures but uses the conditional PMF as the "averaging" PMF. It is next shown that the mean of the unconditional PMF can be found by averaging over the means of the conditional PMFs as given by (8.35). This simplifies the computation. Generation of realizations of random vectors (X,Y) can be simplified using conditioning arguments. An illustration and MATLAB code segment is given in Section 8.7. Finally, an application of conditioning to the modeling of human learning is described in Section 8.8. Utilizing the posterior PMF, which is a conditional PMF, one can demonstrate that "learning" takes place as the result of observing the outcomes of repeated experiments. The degree of learning is embodied in the posterior PMF.

8.3 Conditional Probability Mass Function

We continue with the introductory example to illustrate the utility of the conditional probability mass function. Summarizing the introductory problem, we have an experimental procedure in which we first choose a coin, either coin 1 or coin 2. Coin 1 has a probability of heads of p_1, while coin 2 has a probability of heads of p_2. Let X be the discrete random variable describing the outcome of the coin choice so that

$$X = \begin{cases} 1 & \text{if coin 1 is chosen} \\ 2 & \text{if coin 2 is chosen.} \end{cases}$$

Since $\mathcal{S}_X = \{1, 2\}$, we assign a PMF to X of

$$p_X[i] = \begin{cases} \alpha & i = 1 \\ 1 - \alpha & i = 2 \end{cases} \tag{8.1}$$

where $0 < \alpha < 1$. The second part of the experiment consists of tossing the chosen coin 4 times in succession. Call the outcome of the number of heads observed as Y and note that $\mathcal{S}_Y = \{0, 1, 2, 3, 4\}$. Hence, the overall set of outcomes of the compound experiment is $\mathcal{S}_{X,Y} = \mathcal{S}_X \times \mathcal{S}_Y$, which is shown in Figure 8.1. The overall

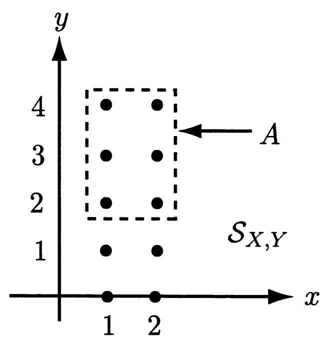


Figure 8.1: Mapping for coin toss example, x denotes the coin chosen while y denotes the number of heads observed.

outcome is described by the random vector (X, Y), where X is the coin chosen and Y is the number of heads observed for the 4 coin tosses. If we wish to determine the probability of 2 or more heads, then this is the probability of the set A shown in Figure 8.1. It is given mathematically as

$$\begin{aligned}P[A] &= \sum_{\{(i,j):(i,j)\in A\}} p_{X,Y}[i,j] \\ &= \sum_{i=1}^{2}\sum_{j=2}^{4} p_{X,Y}[i,j]. \end{aligned} \tag{8.2}$$

Hence, we need only specify the joint PMF to determine the desired probability. To do so we make use of our definition of the joint PMF as well as our earlier concepts from conditional probability (see Chapter 4). Recall from Chapter 7 the definitions of the joint PMF and marginal PMF as

$$\begin{aligned}p_{X,Y}[i,j] &= P[X=i,Y=j] \\ p_X[i] &= P[X=i].\end{aligned}$$

By using the definition of conditional probability for events we have

$$\begin{aligned}p_{X,Y}[i,j] &= P[X=i,Y=j] && \text{(definition of joint PMF)} \\ &= P[Y=j|X=i]P[X=i] && \text{(definition of conditional prob.)} \\ &= P[Y=j|X=i]p_X[i] && \text{(definition of marginal PMF).}\end{aligned} \tag{8.3}$$

From (8.1) we have $p_X[i]$ and from the experimental description we can determine $P[Y=j|X=i]$. When $X=1$, we toss a coin with a probability of heads p_1, and when $X=2$, we toss a coin with a probability of heads p_2. Also, we have previously shown that for a coin with a probability of heads p_i that is tossed 4 times, the number of heads observed has a binomial PMF. Thus, for $i=1,2$

$$P[Y=j|X=i] = \binom{4}{j} p_i^j(1-p_i)^{4-j} \qquad j=0,1,2,3,4. \tag{8.4}$$

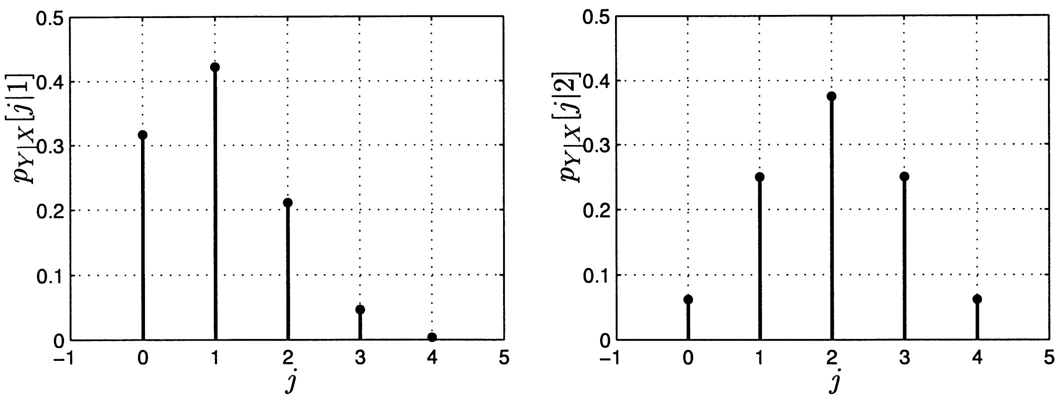
Note that the probability depends on the outcome $X=i$ via p_i. Also, *for a given value* of $X=i$, the probability has all the usual properties of a PMF. These properties are

$$\begin{aligned}0 \le P[Y=j|X=i] \le 1 \\ \sum_{j=0}^{4} P[Y=j|X=i] = 1.\end{aligned}$$

It is therefore appropriate to define $P[Y=j|X=i]$ as a *conditional PMF.* We will denote it by

$$p_{Y|X}[j|i] = P[Y=j|X=i] \qquad j=0,1,2,3,4.$$

Examples are plotted in Figure 8.2 for $p_1 = 1/4$ and $p_2 = 1/2$. Returning to our



(a) $i = 1,\ p_1 = 1/4$ (b) $i = 2,\ p_2 = 1/2$

Figure 8.2: Conditional PMFs given by (8.4).

problem we can now determine the joint PMF. Using (8.3) we have

$$p_{X,Y}[i,j] = p_{Y|X}[j|i]p_X[i] \tag{8.5}$$

and using (8.4) and (8.1) the joint PMF is

$$\begin{aligned} p_{X,Y}[i,j] &= \binom{4}{j} p_1^j (1-p_1)^{4-j}\alpha & i = 1; j = 0,1,2,3,4 \\ &= \binom{4}{j} p_2^j (1-p_2)^{4-j}(1-\alpha) & i = 2; j = 0,1,2,3,4. \end{aligned}$$

Finally the desired probability is from (8.2)

$$\begin{aligned} P[A] &= \sum_{j=2}^{4} p_{X,Y}[1,j] + \sum_{j=2}^{4} p_{X,Y}[2,j] \\ &= \sum_{j=2}^{4} \binom{4}{j} p_1^j (1-p_1)^{4-j}\alpha + \sum_{j=2}^{4} \binom{4}{j} p_2^j (1-p_2)^{4-j}(1-\alpha). \end{aligned}$$

As an example, if $p_1 = 1/4$ and $p_2 = 3/4$, we have for $\alpha = 1/2$, that $P[A] = 0.6055$, but if $\alpha = 1/8$, then $P[A] = 0.8633$. Can you explain this?

Note from (8.5) that the conditional PMF is also expressed as

$$p_{Y|X}[j|i] = \frac{p_{X,Y}[i,j]}{p_X[i]} \tag{8.6}$$

and is only a renaming for the conditional probability of the event that $A_j = \{s : Y(s) = j\}$ given that $B_i = \{s : X(s) = i\}$. To make this connection we have

$$\begin{aligned} p_{Y|X}[j|i] &= P[Y=j|X=i] = \frac{P[X=i, Y=j]}{P[X=i]} \\ &= \frac{P[A_j \cap B_i]}{P[B_i]} \\ &= P[A_j|B_i] \end{aligned}$$

and hence $p_{Y|X}[j|i]$ is a conditional probability for the events A_j and B_i.

8.4 Joint, Conditional, and Marginal PMFs

As evidenced by (8.6), there are relationships between the joint, conditional, and marginal PMFs. In this section we describe these relationships. To do so we rewrite the definition of the conditional PMF in slightly more generality as

$$p_{Y|X}[y_j|x_i] = \frac{p_{X,Y}[x_i, y_j]}{p_X[x_i]} \tag{8.7}$$

for a sample space $\mathcal{S}_{X,Y}$ which may not consist solely of integer two-tuples. *It is always assumed that* $p_X[x_i] \neq 0$. Otherwise, the definition does not make any sense. The conditional PMF, although appearing to be a function of two variables, x_i and y_j, should be viewed as a *family* or set of PMFs. Each PMF in the family is a valid PMF *when x_i is considered to be a constant.* In the example of the previous section, we had $p_{Y|X}[j|1]$ and $p_{Y|X}[j|2]$. The family is therefore $\{p_{Y|X}[j|1], p_{Y|X}[j|2]\}$ and each member is a valid PMF, whose values depend on j. Hence, we would expect that (see Problem 8.4)

$$\begin{aligned} \sum_{j=-\infty}^{\infty} p_{Y|X}[j|1] &= 1 \\ \sum_{j=-\infty}^{\infty} p_{Y|X}[j|2] &= 1 \end{aligned}$$

but *not* $\sum_{i=-\infty}^{\infty} p_{Y|X}[j|i] = 1$ (see also Problem 4.9). Before proceeding to list the relationships between the various PMFs, we give an example of the calculation of the conditional PMF based on (8.7).

Example 8.1 – Two dice toss

Two dice are tossed with all outcomes assumed to be equally likely. The number of dots observed on each die are added together. What is the conditional PMF of the sum if it is known that the sum is even? We begin by letting Y be the sum and define $X = 1$ if the sum is even and $X = 0$ if the sum is odd. Thus, we wish to determine

$p_{Y|X}[j|1]$ and $p_{Y|X}[j|0]$ for all j. The sample space for Y is $\mathcal{S}_Y = \{2, 3, \ldots, 12\}$ as can be seen from Table 8.1, which lists the sum of the two dice outcomes as a function of the outcomes for each die. The boldfaced entries are the ones for which

	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
$i=1$	**2**	3	**4**	5	**6**	7
$i=2$	3	**4**	5	**6**	7	**8**
$i=3$	**4**	5	**6**	7	**8**	9
$i=4$	5	**6**	7	**8**	9	**10**
$i=5$	**6**	7	**8**	9	**10**	11
$i=6$	7	**8**	9	**10**	11	**12**

Table 8.1: The sum of the number of dots observed for two dice – boldface indicates an even sum.

the sum is even and therefore comprise the sample space for $p_{Y|X}[j|1]$. Note that each outcome (i, j) has an assumed probability of occurring of 1/36. Now, using (8.7)

$$p_{Y|X}[j|1] = \frac{p_{X,Y}[1, j]}{p_X[1]} \qquad j = 2, 4, 6, 8, 10, 12 \tag{8.8}$$

where $p_{X,Y}[1, j]$ is the probability of the sum being even and also equaling j. Since we assume in (8.8) that j is even (otherwise $p_{Y|X}[j|1] = 0$), we have that $p_{X,Y}[1, j] = p_Y[j]$ for $j = 2, 4, 6, 8, 10, 12$. Also, there are 18 even outcomes, which results in $p_X[1] = 1/2$. Thus, (8.8) becomes

$$\begin{aligned} p_{Y|X}[j|1] &= \frac{p_Y[j]}{1/2} \\ &= \frac{N_j(1/36)}{1/2} \\ &= \frac{1}{18} N_j \end{aligned}$$

where N_j is the number of outcomes in $\mathcal{S}_{X,Y}$ for which the sum is j. From Table 8.1 we can easily find N_j so that

$$p_{Y|X}[j|1] = \begin{cases} \frac{1}{18} & j = 2 \\ \frac{3}{18} & j = 4 \\ \frac{5}{18} & j = 6 \\ \frac{5}{18} & j = 8 \\ \frac{3}{18} & j = 10 \\ \frac{1}{18} & j = 12. \end{cases} \tag{8.9}$$

Note that as expected $\sum_j p_{Y|X}[j|1] = 1$. The reader is asked to verify by a similar calculation that (see Problem 8.7)

$$p_{Y|X}[j|0] = \begin{cases} \frac{2}{18} & j=3 \\ \frac{4}{18} & j=5 \\ \frac{6}{18} & j=7 \\ \frac{4}{18} & j=9 \\ \frac{2}{18} & j=11. \end{cases} \tag{8.10}$$

These conditional PMFs are shown in Figure 8.3. Also, note that $p_{Y|X}[j|0] \neq$

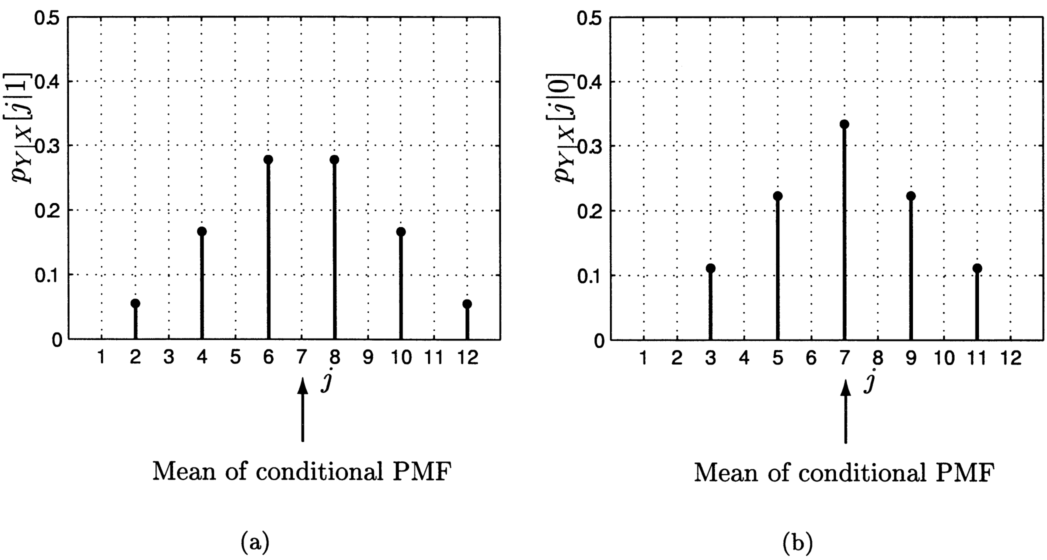


Figure 8.3: Conditional PMFs for Example 8.1.

$1 - p_{Y|X}[j|1]$. Each conditional PMF is generally different.

◇

There are several relationships between the joint, marginal, and conditional PMFs. We now summarize these as properties.

Property 8.1 – Joint PMF yields conditional PMFs.

If the joint PMF $p_{X,Y}[x_i, y_j]$ is known, then the conditional PMFs are found as

$$p_{Y|X}[y_j|x_i] = \frac{p_{X,Y}[x_i,y_j]}{\sum_j p_{X,Y}[x_i,y_j]} \tag{8.11}$$

$$p_{X|Y}[x_i|y_j] = \frac{p_{X,Y}[x_i,y_j]}{\sum_i p_{X,Y}[x_i,y_j]}. \tag{8.12}$$

Proof: Since the marginal PMF $p_X[x_i]$ is found as $\sum_j p_{X,Y}[x_i, y_j]$, the denominator of (8.7) can be replaced by this to yield (8.11). The equation (8.12) is similarly proven.

□

Hence, we see that the conditional PMF is just the joint PMF with x_i fixed and then normalized by $\sum_j p_{X,Y}[x_i, y_j]$ so that it sums to one. In Figure 8.3a, the conditional PMF $p_{Y|X}[j|1]$ evaluated at $j = 8$ is just $p_{X,Y}[1, 8] = 5/36$ divided by the sum of the probabilities $p_{X,Y}[1, \cdot] = 18/36$, where "$\cdot$" indicates all possible values of j. This yields $p_{Y|X}[8|1] = 5/18$.

Property 8.2 – Conditional PMFs are related.

$$p_{X|Y}[x_i|y_j] = \frac{p_{Y|X}[y_j|x_i]p_X[x_i]}{p_Y[y_j]} \tag{8.13}$$

Proof: By interchanging X and Y in (8.7) we have

$$p_{X|Y}[x_i|y_j] = \frac{p_{Y,X}[y_j, x_i]}{p_Y[y_j]}$$

but

$$\begin{aligned} p_{Y,X}[y_j, x_i] &= P[Y = y_j, X = x_i] \\ &= P[X = x_i, Y = y_j] \qquad \text{(since } A \cap B = B \cap A\text{)} \\ &= p_{X,Y}[x_i, y_j] \end{aligned}$$

and therefore

$$p_{X|Y}[x_i|y_j] = \frac{p_{X,Y}[x_i, y_j]}{p_Y[y_j]}. \tag{8.14}$$

Using $p_{X,Y}[x_i, y_j] = p_{Y|X}[y_j|x_i]p_X[x_i]$ from (8.7) in (8.14) yields the desired result (8.13).

□

Property 8.3 – Conditional PMF is expressible using Bayes' rule.

$$p_{Y|X}[y_j|x_i] = \frac{p_{X|Y}[x_i|y_j]p_Y[y_j]}{\sum_j p_{X|Y}[x_i|y_j]p_Y[y_j]} \tag{8.15}$$

Proof: From (8.11) we have that

$$p_{Y|X}[y_j|x_i] = \frac{p_{X,Y}[x_i, y_j]}{\sum_j p_{X,Y}[x_i, y_j]} \tag{8.16}$$

and using (8.14) we have

$$p_{X,Y}[x_i, y_j] = p_{X|Y}[x_i|y_j]p_Y[y_j] \tag{8.17}$$

which when substituted into (8.16) yields the desired result.

□

Property 8.4 – Conditional PMF and its corresponding marginal PMF yields the joint PMF.

$$p_{X,Y}[x_i, y_j] = p_{Y|X}[y_j|x_i]p_X[x_i] \tag{8.18}$$
$$p_{X,Y}[x_i, y_j] = p_{X|Y}[x_i|y_j]p_Y[y_j] \tag{8.19}$$

Proof: (8.18) follows from definition of conditional PMF (8.7) and (8.19) is just (8.17).

□

Property 8.5 – Conditional PMF and its corresponding marginal PMF yields the other marginal PMF.

$$p_Y[y_j] = \sum_i p_{Y|X}[y_j|x_i]p_X[x_i] \tag{8.20}$$

Proof: This is just the law of total probability in disguise or equivalently just $p_Y[y_j] = \sum_i p_{X,Y}[x_i, y_j]$ (marginal PMF from joint PMF).

□

These relationships are summarized in Figure 8.4. Notice that the joint PMF can be used to find all the marginals and conditional PMFs (see Figure 8.4a). The conditional PMF and its corresponding marginal PMF can be used to find the joint PMF (see Figure 8.4b). Finally, the conditional PMF and its corresponding marginal PMF can be used to find the other conditional PMF (see Figure 8.4c). As emphasized earlier, we cannot determine the joint PMF from the marginals. This is only possible if X and Y are independent random variables since in this case

$$p_{X,Y}[x_i, y_j] = p_X[x_i]p_Y[y_j]. \tag{8.21}$$

In addition, *for independent random variables*, the use of (8.21) in (8.7) yields

$$p_{Y|X}[y_j|x_i] = \frac{p_X[x_i]p_Y[y_j]}{p_X[x_i]} = p_Y[y_j] \tag{8.22}$$

or *the conditional PMF is the same as the unconditional PMF.* There is no change in the probabilities of Y whether or not X is observed. This is of course consistent with our previous definition of statistical independence.

Finally, for more general conditional probability calculations we sum the appropriate values of the conditional PMF to yield (see Problem 8.14)

$$P[Y \in A|X = x_i] = \sum_{\{j: y_j \in A\}} p_{Y|X}[y_j|x_i]. \tag{8.23}$$

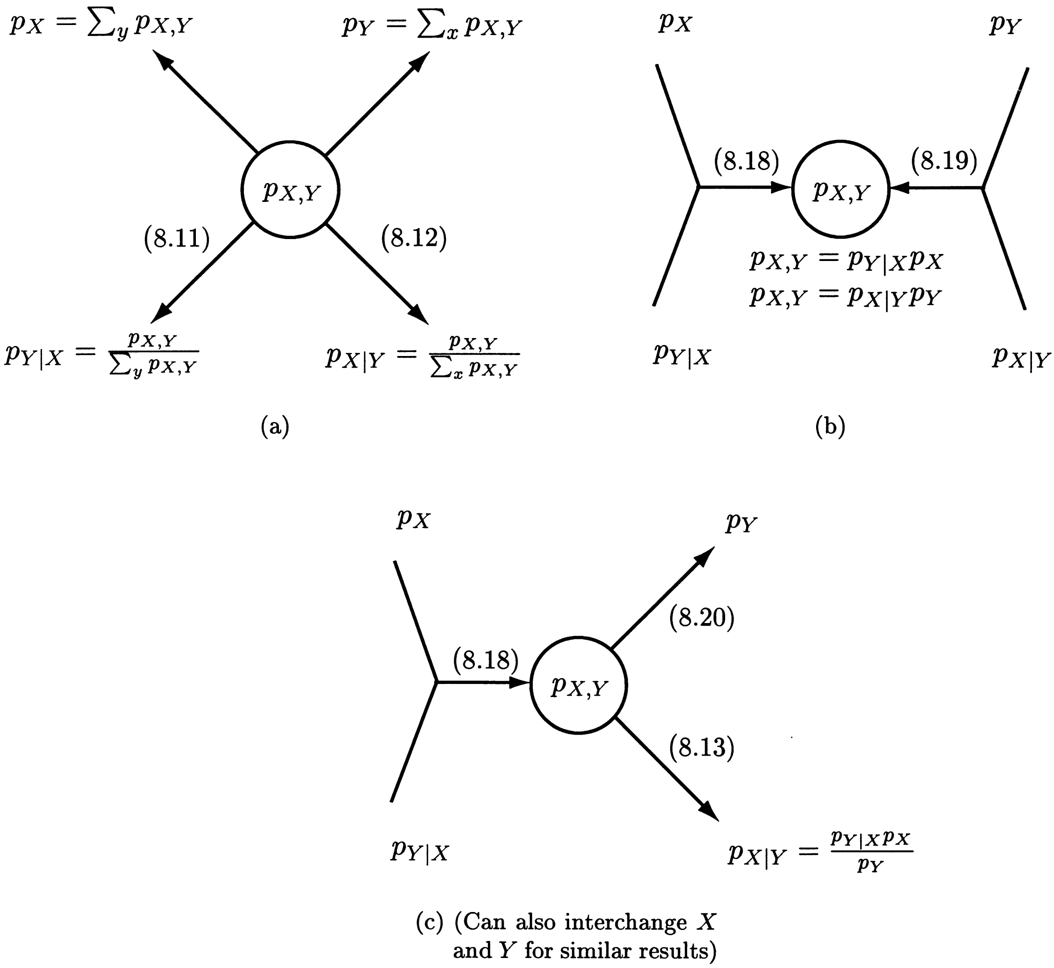


Figure 8.4: Conditional PMF relationships.

8.5 Simplifying Probability Calculations using Conditioning

As alluded to in the introduction, conditional PMFs can be used to simplify probability calculations. To illustrate the use of this approach we once again consider the determination of the PMF for $Z = X + Y$, where X and Y are independent discrete random variables that take on integer values. We have already seen that the solution is $p_Z = p_X \star p_Y$, where $\star$ denotes discrete convolution (see (7.22)). To solve this problem using conditional PMFs, we ask ourselves the question: Could I find the PMF of Z if X were known? If so, then we should be able to use conditioning arguments to first find the conditional PMF of Z given X, and then *uncondition* the result to yield the PMF of Z. Let us say that X is known and that $X = i$. As a result, we have that conditionally $Z = i + Y$, where i is just a constant. This is sometimes denoted by $Z|(X = i)$. But this is a transformation from one discrete random variable Y to another discrete random variable Z. We therefore wish to determine the PMF of a random variable that has been summed with a *constant.* It is not difficult to show that if a discrete random variable U has a PMF $p_U[j]$, then $U + i$ has the PMF $p_U[j - i]$ or the PMF is just shifted to the right by i units. Thus, the *conditional PMF* of Z evaluated at $Z = j$ is $p_{Z|X}[j|i] = p_{Y|X}[j-i|i]$. Now to find the unconditional PMF of Z we use (8.20) with an appropriate change of variables to yield

$$p_Z[j] = \sum_{i=-\infty}^{\infty} p_{Z|X}[j|i]p_X[i]$$

and since $p_{Z|X}[j|i] = p_{Y|X}[j - i|i]$, we have

$$p_Z[j] = \sum_{i=-\infty}^{\infty} p_{Y|X}[j - i|i]p_X[i].$$

But X and Y are independent so that $p_{Y|X} = p_Y$ and therefore we have the final result

$$p_Z[j] = \sum_{i=-\infty}^{\infty} p_Y[j - i]p_X[i]$$

which agrees with our earlier one. Another example follows.

Example 8.2 – PMF for $Z = \max(X, Y)$

Let X and Y be discrete random variables that take on integer values. Also, assume independence of the random variables X and Y and that the marginal PMFs of X and Y are known. To find the PMF of Z we use (8.20) or the law of total probability to yield

$$p_Z[k] = \sum_{i=-\infty}^{\infty} p_{Z|X}[k|i]p_X[i]. \tag{8.24}$$

Now p_X is known so that we only need to determine $p_{Z|X}$ for $X = i$. But given that $X = i$, we have that $Z = \max(i, Y)$ for which the PMF is easily found. We have thus reduced the original problem, which is to determine the PMF for the random variable obtained by transforming from (X, Y) to Z, to determining the PMF for a function of *only one random variable.* Letting $g(Y) = \max(i, Y)$ we see that the function appears as shown in Figure 8.5. Hence, using (5.9) for the PMF of a single transformed discrete random variable we have

$$p_{Z|X}[k|i] = \sum_{\{j:g(j)=k\}} p_{Y|X}[j|i].$$

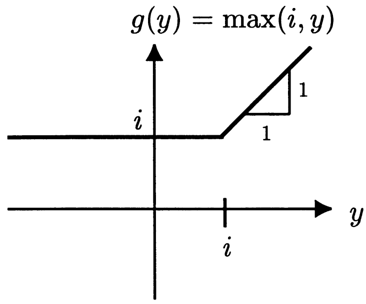


Figure 8.5: Plot of the function $g(y) = \max(i, y)$.

Solving for j in $g(j) = k$ (refer to Figure 8.5) yields no solution for $k < i$, the multiple solutions $j = \ldots, i-1, i$ for $k = i$, and the single solution $j = k$ for $k = i+1, i+2, \ldots$. This produces

$$p_{Z|X}[k|i] = \begin{cases} 0 & k = \ldots, i-2, i-1 \\ \sum_{j=-\infty}^{i} p_{Y|X}[j|i] & k = i \\ p_{Y|X}[k|i] & k = i+1, i+2, \ldots \end{cases} \tag{8.25}$$

Using this in (8.24) produces

$$\begin{aligned} p_Z[k] &= \sum_{i=-\infty}^{k-1} p_{Z|X}[k|i]p_X[i] + p_{Z|X}[k|k]p_X[k] + \sum_{i=k+1}^{\infty} p_{Z|X}[k|i]p_X[i] && \text{(break up sum)} \\ &= \sum_{i=-\infty}^{k-1} p_{Y|X}[k|i]p_X[i] + \sum_{j=-\infty}^{k} p_{Y|X}[j|k]p_X[k] + 0 && \text{(use (8.25))} \\ &= \sum_{i=-\infty}^{k-1} p_Y[k]p_X[i] + \sum_{j=-\infty}^{k} p_Y[j]p_X[k] && \text{(since } X \text{ and } Y \text{ are independent)} \\ &= p_Y[k] \sum_{i=-\infty}^{k-1} p_X[i] + p_X[k] \sum_{j=-\infty}^{k} p_Y[j]. \end{aligned}$$

Note that due to the independence assumption this final result can also be written as

$$p_Z[k] = \sum_{i=-\infty}^{k-1} p_{X,Y}[i,k] + \sum_{j=-\infty}^{k} p_{X,Y}[k,j]$$

so that the PMF of Z is obtained by summing all the points of the joint PMF shown in Figure 8.6 for $k = 2$, as an example. These point comprise the set $\{(x,y) : \max(x,y) = 2 \text{ and } x = i, y = j\}$. It is now clear that we could have solved this problem in a more direct fashion by making this observation. As in most problems, however, the solution is usually trivial once it is known!

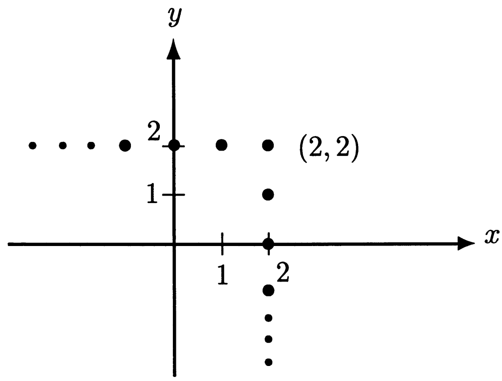


Figure 8.6: Points of joint PMF to be summed to find PMF of $Z = \max(X, Y)$ for $k = 2$.

$\diamond$

As we have seen, a general procedure for determining the PMF for $Z = g(X, Y)$ when X and Y are independent is as follows:

1. Fix $X = x_i$ and let $Z|(X = x_i) = g(x_i, Y)$

2. Find the PMF for $Z|X$ by using the techniques for a transformation of a single random variable Y into another random variable Z. The formula is from (5.9), where the PMFs are first converted to conditional PMFs

$$\begin{aligned} p_{Z|X}[z_k|x_i] &= \sum_{\{j:g(x_i,y_j)=z_k\}} p_{Y|X}[y_j|x_i] \quad \text{for each } x_i \\ &= \sum_{\{j:g(x_i,y_j)=z_k\}} p_Y[y_j] \quad \text{for each } x_i \quad \text{(due to independence).} \end{aligned}$$

3. Uncondition the conditional PMF to yield the desired PMF

$$p_Z[z_k] = \sum_i p_{Z|X}[z_k|x_i]p_X[x_i].$$

In general, to compute probabilities of events it is advantageous to use a conditioning argument, whether or not X and Y are independent. Where previously we have used the formula

$$P[Y \in A] = \sum_{\{j:y_j \in A\}} p_Y[y_j]$$

to compute the probability, a conditioning approach would replace $p_Y[y_j]$ by $\sum_i p_{Y|X}[y_j|x_i]p_X[x_i]$ to yield

$$P[Y \in A] = \sum_{\{j:y_j \in A\}} \sum_i p_{Y|X}[y_j|x_i]p_X[x_i] \tag{8.26}$$

to determine the probability. Equivalently, we have that

$$P[Y \in A] = \underbrace{\sum_i \left[\sum_{\{j: y_j \in A\}} p_{Y|X}[y_j|x_i] \right]}_{\text{conditioning}} \underbrace{p_X[x_i]}_{\text{unconditioning}} . \tag{8.27}$$

In this form we recognize the conditional probability of (8.23), which is

$$P[Y \in A | X = x_i] = \sum_{\{j: y_j \in A\}} p_{Y|X}[y_j|x_i]$$

and the unconditional probability

$$P[Y \in A] = \sum_i P[Y \in A | X = x_i] p_X[x_i] \tag{8.28}$$

with the latter being just a restatement of the law of total probability.

8.6 Mean of the Conditional PMF

Since the conditional PMF is a PMF, it exhibits all the usual properties. In particular, we can determine attributes such as the expected value of a random variable Y, when it is known that $X = x_i$. This expected value is the *mean of the conditional PMF* $p_{Y|X}$. Its definition is the usual one

$$\sum_j y_j p_{Y|X}[y_j|x_i] \tag{8.29}$$

where we have replaced p_Y by $p_{Y|X}$. It should be emphasized that since the conditional PMF depends on x_i, so will its mean. Hence, the mean of the conditional PMF is a constant when we set x_i equal to a fixed value. We adopt the notation for the mean of the conditional PMF as $E_{Y|X}[Y|x_i]$. This notation includes the subscript "$Y|X$" to remind us that the averaging PMF is the conditional PMF $p_{Y|X}$. Also, the use of "$Y|x_i$" as the argument will remind us that the averaging PMF is the conditional PMF that is specified by $X = x_i$ in the family of conditional PMFs. The mean is therefore defined as

$$E_{Y|X}[Y|x_i] = \sum_j y_j p_{Y|X}[y_j|x_i]. \tag{8.30}$$

Although we have previously asserted that the mean is a constant, here *it is to be regarded as a function of* x_i. An example of its calculation follows.

Example 8.3 – Mean of conditional PMF – continuation of Example 8.1

We now compute all the possible values of $E_{Y|X}[Y|x_i]$ for the problem described in Example 8.1. There $x_i = 1$ or $x_i = 0$ and the corresponding conditional PMFs are given by (8.9) and (8.10), respectively. The means of the conditional PMFs are therefore

$$\begin{aligned}
E_{Y|X}[Y|1] &= 2\left(\frac{1}{18}\right) + 4\left(\frac{3}{18}\right) + 6\left(\frac{5}{18}\right) + 8\left(\frac{5}{18}\right) + 10\left(\frac{3}{18}\right) + 12\left(\frac{1}{18}\right) = 7 \\
E_{Y|X}[Y|0] &= 3\left(\frac{2}{18}\right) + 5\left(\frac{4}{18}\right) + 7\left(\frac{6}{18}\right) + 9\left(\frac{4}{18}\right) + 11\left(\frac{2}{18}\right) = 7
\end{aligned}$$

and are shown in Figure 8.3. In this example the means of the conditional PMFs are the same, but will not be in general. We can expect that $g(x_i) = E_{Y|X}[Y|x_i]$ *will vary with* x_i.

◇

We could also compute the variance of the conditional PMFs. This would be

$$\operatorname{var}(Y|x_i) = \sum_j \left(y_j - E_{Y|X}[Y|x_i]\right)^2 p_{Y|X}[y_j|x_i]. \tag{8.31}$$

The reader is asked to do this in Problem 8.22. (See also Problem 8.23 for an alternate expression for $\operatorname{var}(Y|x_i)$.) Note from Figure 8.3 that we do not expect these to be the same.

⚠ What is the "conditional expectation"?

The *function* $g(x_i) = E_{Y|X}[Y|x_i]$ is the mean of the conditional PMF $p_{Y|X}[y_j|x_i]$. Alternatively, it is known as the *conditional mean.* This terminology is widespread and so we will adhere to it, although we should keep in mind that it is meant to denote the usual mean of the *conditional* PMF. It is also of interest to determine the expectation of other quantities besides Y with respect to the conditional PMF. This is called the *conditional expectation* and is symbolized by $E_{Y|X}[g(Y)|x_i]$. The latter is called the conditional expectation of $g(Y)$. For example, if $g(Y) = Y^2$, then it becomes the conditional expectation of Y^2 or equivalently the *conditional second moment.* Lastly, the reader should be aware that the conditional mean is the optimal predictor of a random variable based on observation of a second random variable (see Problem 8.27).

⚠

We now give another example of the computation of the conditional mean.

Example 8.4 – Toss one of two dice.

There are two dice having different numbers of dots on their faces. Die 1 is the usual type of die with faces having 1, 2, 3, 4, 5, or 6 dots. Die 2 has been mislabled with its faces having $2,3,2,3,2$, or 3 dots. A die is selected at random and tossed. Each face of the die is equally likely to occur. What is the expected number of dots observed for the tossed die? To solve this problem first observe that the outcomes will depend upon which die has been tossed. As a result, the *conditional expectation* of the number of dots will depend upon which die is initially chosen. We can view this problem as a conditional one by letting

$$X = \begin{cases} 1 & \text{if die 1 is chosen} \\ 2 & \text{if die 2 is chosen} \end{cases}$$

and Y is the number of dots observed. Thus, we wish to determine $E_{Y|X}[Y|1]$ and $E_{Y|X}[Y|2]$. But if die 1 is chosen, the conditional PMF is

$$p_{Y|X}[j|1] = \frac{1}{6} \qquad j = 1,2,3,4,5,6 \tag{8.32}$$

and if die 2 is chosen

$$p_{Y|X}[j|2] = \frac{1}{2} \qquad j = 2,3. \tag{8.33}$$

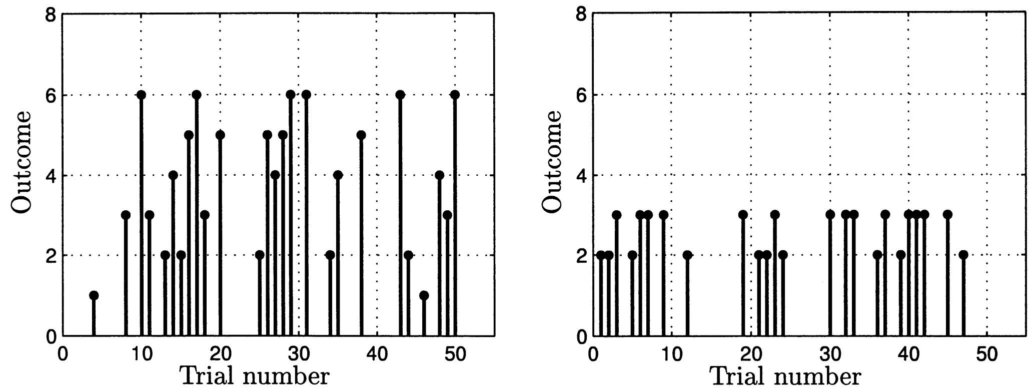
The latter conditional PMF is due to the fact that for die 2 half the sides show 2 dots and the other half of the sides show 3 dots. Using (8.30) with (8.32) and (8.33), we have that

$$\begin{aligned} E_{Y|X}[Y|1] &= \sum_{j=1}^{6} j p_{Y|X}[j|1] = \frac{7}{2} \\ E_{Y|X}[Y|2] &= \sum_{j=2}^{3} j p_{Y|X}[j|2] = \frac{5}{2}. \end{aligned} \tag{8.34}$$

An example of typical outcomes for this experiment is shown in Figure 8.7. For 50 trials of the experiment Figure 8.7a displays the outcomes for which die 1 was chosen and Figure 8.7b displays the outcomes for which die 2 was chosen. It is interesting to note that the estimated mean for Figure 8.7a is 3.88 and for Figure 8.7b it is 2.58. Note that from (8.34) the theoretical conditional means are 3.5 and 2.5, respectively.

$\diamondsuit$

In the previous example, we have determined the conditional means, which are the means of the conditional PMFs. We also might wish to determine the *unconditional mean*, which is the mean of Y. This is the number of dots observed as a result of the overall experiment, without first conditioning on which die was chosen. In essence, we wish to determine $E_Y[Y]$. Intuitively, this is the average number of dots observed if we combined Figures 8.7a and 8.7b together (just overlay Figure 8.7b onto Figure 8.7a) and continued the experiment indefinitely. Hence, we wish to determine $E_Y[Y]$ for the following experiment:



(a) Outcomes when die 1 chosen

(b) Outcomes when die 2 chosen

Figure 8.7: Computer simulated outcomes of randomly selected die toss experiment.

1. Choose die 1 or die 2 with probability of 1/2.

2. Toss the chosen die.

3. Count the number of dots on the face of tossed die and call this the outcome of the random variable Y.

A simple MATLAB program to simulate this experiment is given as

```
for m=1:M
  if rand(1,1)<0.5
    y(m,1)=PMFdata(1,[1 2 3 4 5 6]',[1/6 1/6 1/6 1/6 1/6 1/6]');
  else
     y(m,1)=PMFdata(1,[2 3]',[1/2 1/2]');
  end
end
```

where the subprogram `PMFdata.m` is listed in Appendix 6B. After the code is executed there is an array `y`, which is $M \times 1$, containing M realizations of Y. By taking the sample mean of the elements in the array `y`, we will have estimated $E_Y[Y]$. But we expect about half of the realizations to have used the fair die and the other half to use the mislabled die. As a result, we might suppose that the unconditional mean is just the average of the two conditional means. This would be $(1/2)(7/2) + (1/2)(5/2) = 3$, which turns out to be the true result. This conjecture is also strengthened by the results of Figure 8.7. By overlaying the plots we have 50 outcomes of the experiment for which the sample mean is 3.25. Let's see how to verify this.

To determine the theoretical mean of Y, i.e., the unconditional mean, we will need $p_Y[j]$. But given the conditional PMF and the marginal PMF we know from Figure 8.4c that the joint PMF can be found. Hence, from (8.32) and (8.33) and $p_X[i] = 1/2$ for $i = 1, 2$, we have

$$\begin{aligned} p_{X,Y}[i,j] &= p_{Y|X}[j|i]p_X[i] \\ &= \begin{cases} \frac{1}{12} & i = 1; j = 1,2,3,4,5,6 \\ \frac{1}{4} & i = 2; j = 2,3. \end{cases} \end{aligned}$$

To find $p_Y[j]$ we use

$$\begin{aligned} p_Y[j] &= \sum_{i=1}^{2} p_{X,Y}[i,j] \\ &= \begin{cases} p_{X,Y}[1,j] = \frac{1}{12} & j = 1,4,5,6 \\ p_{X,Y}[1,j] + p_{X,Y}[2,j] = \frac{1}{12} + \frac{1}{4} = \frac{1}{3} & j = 2,3. \end{cases} \end{aligned}$$

Thus, the unconditional mean becomes

$$\begin{aligned} E_Y[Y] &= \sum_{j=1}^{6} j p_Y[j] \\ &= 1\left(\frac{1}{12}\right) + 2\left(\frac{1}{3}\right) + 3\left(\frac{1}{3}\right) + 4\left(\frac{1}{12}\right) + 5\left(\frac{1}{12}\right) + 6\left(\frac{1}{12}\right) \\ &= 3. \end{aligned}$$

This value is sometimes called the *unconditional expectation*. Note that for this example, we have upon using (8.34)

$$E_Y[Y] = E_{Y|X}[Y|1]p_X[1] + E_{Y|X}[Y|2]p_X[2]$$

or the *unconditional mean is the average of the conditional means*. This is true in general and is summarized by the relationship

$$E_Y[Y] = \sum_i E_{Y|X}[Y|x_i]p_X[x_i]. \tag{8.35}$$

To prove this relationship is straightforward. Starting with (8.35) we have

$$\begin{aligned}
& \sum_i E_{Y|X}[Y|x_i] p_X[x_i] \\
= & \sum_i \left(\sum_j y_j p_{Y|X}[y_j|x_i] \right) p_X[x_i] \quad \text{(definition of conditional mean)} \\
= & \sum_i \sum_j y_j \frac{p_{X,Y}[x_i, y_j]}{p_X[x_i]} p_X[x_i] \qquad \text{(definition of conditional PMF)} \\
= & \sum_j y_j \sum_i p_{X,Y}[x_i, y_j] \\
= & \sum_j y_j p_Y[y_j] \qquad \text{(marginal PMF from joint PMF)} \\
= & E_Y[Y].
\end{aligned}$$

In (8.35) we can consider $g(x_i) = E_{Y|X}[Y|x_i]$ as the transformed outcome of the coin choice part of the experiment, where $X = x_i$ is the outcome of the coin choice. Since before we choose the coin to toss, we do not know which one it will be, we can consider $g(X)$ as a transformed *random variable* whose values are $g(x_i)$. By this way of viewing things, we can define a random variable as $g(X) = E_{Y|X}[Y|X]$ and therefore rewrite (8.35) as

$$E_Y[Y] = E_X[g(X)]$$

or explicitly we have that

$$E_Y[Y] = E_X[E_{Y|X}[Y|X]]. \tag{8.36}$$

In effect, we have computed the expectation of a random variable in two steps. Step 1 is to compute a conditional expectation $E_{Y|X}$ while step 2 is to undo the conditioning by averaging the result with respect to the PMF of X. An example is the previous coin tossing experiment. The utility in doing so is that the conditional PMFs were easily found and hence also the means of the conditional PMFs, and finally the averaging with respect to p_X is easily carried out to yield the desired result. We illustrate the use of (8.36) with another example.

Example 8.5 – Random number of coin tosses

An experiment is conducted in which a coin with a probability of heads p is tossed M times. However, M is a *random variable* with $M \sim \text{Pois}(\lambda)$. For example, if a realization of M is generated, say $M = 5$, then the coin is tossed 5 times in succession. We wish to determine the average number of heads observed. Conditionally on knowing the value of M, we have a binomial PMF for the number of heads Y. Hence, for $M = i$ we have upon using the binomial PMF (see (5.6))

$$p_{Y|M}[j|i] = \binom{i}{j} p^j (1-p)^{i-j} \qquad j = 0, 1, \ldots, i; i = 0, 1, \ldots.$$

Now using (8.36) and replacing X with M we have

$$E_Y[Y] = E_M[E_{Y|M}[Y|M]]$$

and for a binomial PMF we know that $E_{Y|M}[Y|i] = ip$ so that

$$E_Y[Y] = E_M[Mp] = pE_M[M].$$

But for a Poisson random variable $E_M[M] = \lambda$, which yields the final result

$$E_Y[Y] = \lambda p.$$

It can be shown more generally that $Y \sim \text{Pois}(\lambda p)$ (see Problem 8.26) so that our result for the mean of Y follows directly from knowledge of the mean of a Poisson random variable.

$\diamondsuit$

8.7 Computer Simulation Based on Conditioning

In Section 7.11 we discussed a simple method for generating realizations of jointly distributed discrete random variables (X, Y) using MATLAB. To do so we required the joint PMF. Using conditioning arguments, however, we can frequently simplify the procedure. Since $p_{X,Y}[x_i, y_j] = p_{Y|X}[y_j|x_i]p_X[x_i]$, a realization of (X, Y) can be obtained by first generating a realization of X according to its marginal PMF $p_X[x_i]$. Then, assuming that $X = x_i$ is obtained, we next generate a realization of Y according to the *conditional PMF* $p_{Y|X}[y_j|x_i]$. (Of course, if X and Y are independent, we replace the second step by the generation of Y according to $p_Y[y_j]$ since in this case $p_{Y|X}[y_j|x_i] = p_Y[y_j]$.) This is also advantageous when the problem description is formulated in terms of conditional PMFs, as in a compound experiment. To illustrate this approach with the one described previously we repeat Example 7.15.

Example 8.6 – Generating realizations of jointly distributed random variables – Example 7.15 (continued)

The joint PMF of Example 7.15 is shown in Figure 8.8, where the solid circles represent the sample points and the values of the joint PMF are shown to the right of the sample points. To use a conditioning approach we need to find p_X and $p_{Y|X}$. But from Figure 8.8, if we sum along the columns we obtain

$$p_X[i] = \begin{cases} \frac{1}{4} & i = 0 \\ \frac{3}{4} & i = 1 \end{cases}$$

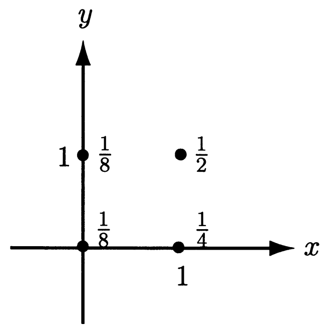


Figure 8.8: Joint PMF for Example 8.6.

and using the definition of the conditional PMF, we have

$$\begin{aligned} p_{Y|X}[j|0] &= \frac{p_{X,Y}[0,j]}{p_X[0]} \\ &= \begin{cases} \frac{1/8}{1/4} = \frac{1}{2} & j=0 \\ \frac{1/8}{1/4} = \frac{1}{2} & j=1 \end{cases} \end{aligned}$$

and

$$\begin{aligned} p_{Y|X}[j|1] &= \frac{p_{X,Y}[1,j]}{p_X[1]} \\ &= \begin{cases} \frac{1/4}{3/4} = \frac{1}{3} & j=0 \\ \frac{1/2}{3/4} = \frac{2}{3} & j=1. \end{cases} \end{aligned}$$

The MATLAB segment of code shown below generates M realizations of (X, Y) using this conditioning approach.

```
for m=1:M
  ux=rand(1,1);
  uy=rand(1,1);
  if ux<=1/4; % Refer to px[i]
    x(m,1)=0;
    if uy<=1/2 % Refer to py|x[j|0]
      y(m,1)=0;
    else
      y(m,1)=1;
    end
  else
    x(m,1)=1;  % Refer to px[i]
```

```
          if uy<=1/3 % Refer to py|x[j|1]
            y(m,1)=0;
          else
            y(m,1)=1;
          end
        end
      end
```

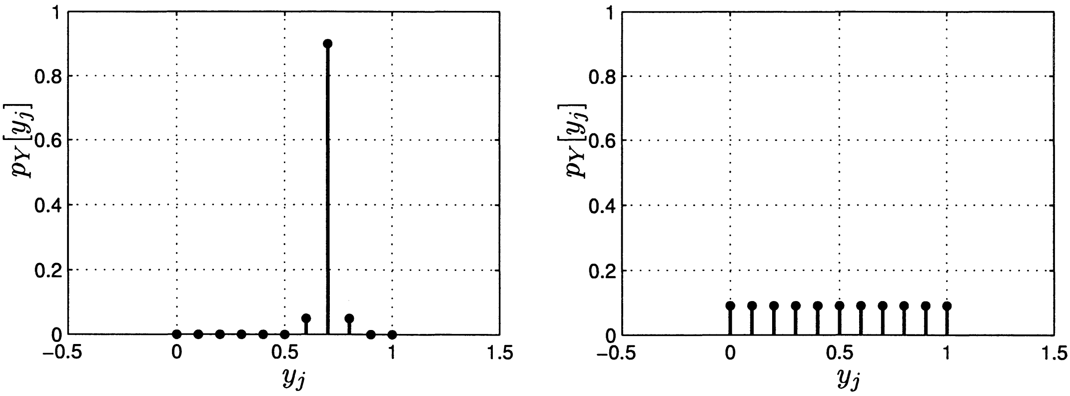
The reader is asked to test this program in Problem 8.29.

◇

8.8 Real-World Example – Modeling Human Learning

A 2 year-old child who has learned to walk can perform tasks that not even the most sophisticated robots can match. For example, a 2 year-old child can easily maneuver her way to a favorite toy, pick it up, and start to play with it. Robots, powered by machine vision and mechanical grippers, have a hard time performing this supposedly simple task. It is not surprisingly, therefore, that one of the holy grails in cognitive science and also machine learning is to figure out how a child does this. If we were able to understand the thought processes that were used to successfully complete this task, then it is conceivable that a machine might be built to do the same thing. Many models of human learning employ a Bayesian framework [Tenenbaum 1999]. This approach appears to be fruitful in that using Bayesian modeling we are able to discriminate with more and more accuracy as we repeatedly perform an experiment and observe the outcome. This is analogous to a child attempting to pick up the toy, dropping it, picking it up again after having learned something about how to pick it up, dropping it, etc., until finally she is successful. Each time the experiment, attempting to pick up the toy, is repeated the child learns something or equivalently narrows down the number of possible strategies. In Bayesian analysis, as we will show next, the width of the PMF decreases as we observe more outcomes. This is in some sense saying that our uncertainty about the outcome of the experiment decreases as it is performed more times. Although not a perfect analogy, it does seem to possess some critical elements of the human learning process. Therefore, we illustrate this modeling with the simple example of coin tossing.

Suppose we wish to "learn" whether a coin is fair ($p = 1/2$) or is weighted ($p \neq 1/2$). One way to do this is to repeatedly toss the coin and count the number of heads observed. We would expect that our certainty about the conclusion, that the coin is fair or not, would increase as the number of trials increases. In the Bayesian model we quantify our knowledge about the value of p by assuming that p is a *random variable*. Our particular coin, however, has a fixed probability of heads. It is just that we do not know what it is and hence our *belief* about the value of p is embodied in an assumed PMF. This is a slightly different interpretation of probability than our previous relative frequency interpretation. To conform to our previous notation we let the probability of heads be denoted by the random variable Y and its values by y_j. Then, we determine its PMF. Our state of knowledge will be high if the PMF is highly concentrated about a particular value, as for example in Figure 8.9a. If, however, the PMF is spread out or "diffuse", our state of knowledge will be low, as for example in Figure 8.9b. Now let's say that we wish to learn the



(a) Y = probability of heads – state of knowledge is high.

(b) Y = probability of heads – state of knowledge is low.

Figure 8.9: PMFs reflecting state of knowledge about coin's probability of heads.

value of the probability of heads. Before we toss the coin we have no idea what it is, and therefore it is reasonable to assume a PMF that is uniform, as, for example, the one shown in Figure 8.9b. Such a PMF is given by

$$p_Y[y_j] = \frac{1}{M+1} \qquad \text{for } y_j = 0, \tfrac{1}{M}, \tfrac{2}{M}, \ldots, \tfrac{M-1}{M}, 1 \tag{8.37}$$

for some large M (in Figure 8.9b $M = 11$). This is also called the *prior PMF* since it summarizes our state of knowledge *before* the experiment is performed. Now we begin to toss the coin and examine our state of knowledge as the number of tosses increases. Let N be the number of coin tosses and X denote the number of heads observed in the N tosses. We know that the PMF of the number of heads is binomially distributed. However, to specify the PMF completely, we require knowledge of the probability of heads. Since this is unknown, we can only specify the PMF of X conditionally or if $Y = y_j$ is the probability of heads, then the conditional PMF of the number of heads for $X = i$ is

$$p_{X|Y}[i|y_j] = \binom{N}{i} y_j^i (1-y_j)^{N-i} \qquad i = 0, 1, \ldots, N. \tag{8.38}$$

Since we are actually interested in the probability of heads or the PMF of Y after observing the outcomes of N coin tosses, we need to determine the conditional PMF $p_{Y|X}[y_j|i]$. The latter is also called the *posterior PMF*, since it is to be determined *after* the experiment is peformed. The reader may wish to compare this terminology with that used in Chapter 4. The posterior PMF contains all the information about the probability of heads that results from our prior knowledge, summarized by p_Y, and our "data" knowledge, summarized by $p_{X|Y}$. The posterior PMF is given by Bayes' rule (8.15) with $x_i = i$ as

$$p_{Y|X}[y_j|i] = \frac{p_{X|Y}[i|y_j]p_Y[y_j]}{\sum_j p_{X|Y}[i|y_j]p_Y[y_j]}.$$

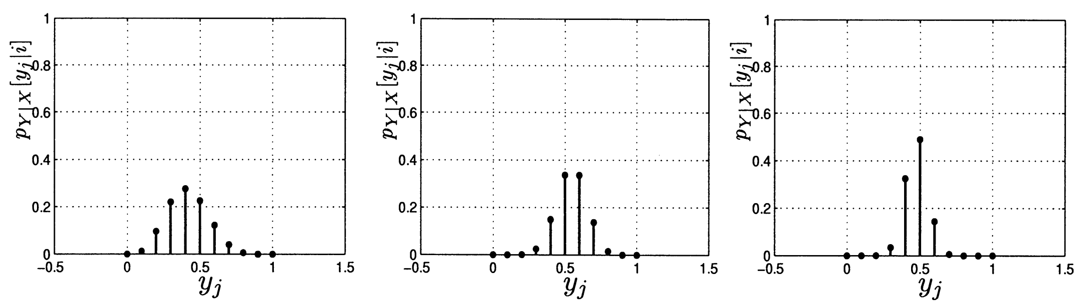
Using (8.37) and (8.38) we have

$$p_{Y|X}[y_j|i] = \frac{\binom{N}{i} y_j^i(1-y_j)^{N-i}\frac{1}{M+1}}{\sum_{j=0}^{M}\binom{N}{i} y_j^i(1-y_j)^{N-i}\frac{1}{M+1}} \qquad y_j = 0, 1/M, \ldots, 1; i = 0, 1, \ldots, N$$

or finally,

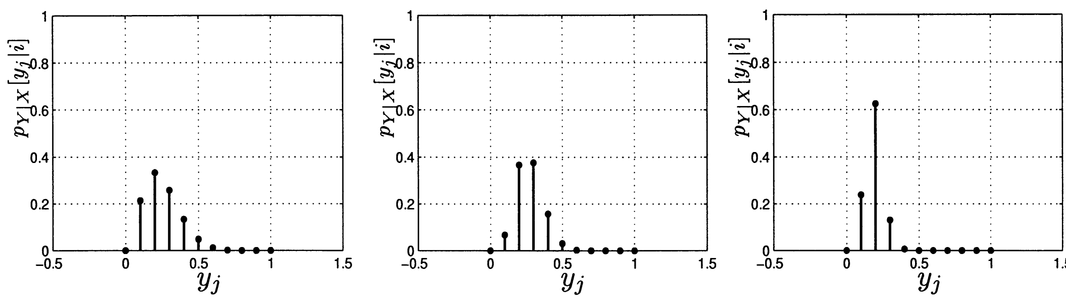
$$p_{Y|X}[y_j|i] = \frac{y_j^i(1-y_j)^{N-i}}{\sum_{j=0}^{M} y_j^i(1-y_j)^{N-i}} \qquad y_j = 0, 1/M, \ldots, 1; i = 0, 1, \ldots, N. \tag{8.39}$$

Note that the posterior PMF depends on the number of heads observed, which is i. To understand what this PMF is saying about our state of knowledge, assume that we toss the coin $N = 10$ times and observe $i = 4$ heads. The posterior PMF is shown in Figure 8.10a. For $N = 20$, $i = 11$ and $N = 40$, $i = 19$, the posterior PMFs are shown in Figures 8.10b and 8.10c, respectively. Note that as the number



(a) $N = 10$, $i = 4$ (b) $N = 20$, $i = 11$ (c) $N = 40$, $i = 19$

Figure 8.10: Posterior PMFs for coin tossing analogy to human learning – coin appears to be fair. The y_j's are possible probability values for a head.



(a) $N = 10$, $i = 2$ (b) $N = 20$, $i = 5$ (c) $N = 40$, $i = 7$

Figure 8.11: Posterior PMFs for coin tossing analogy to human learning – coin appears to be weighted. The y_j's are possible probability values for a head.

of tosses increases the posterior PMF becomes narrower and centered about the value of 0.5. The Bayesian model has "learned" the value of p, with our confidence increasing as the number of trials increases. Note that for no trials (just set $N = 0$ and hence $i = 0$ in (8.39)) we have just the uniform prior PMF of Figure 8.9b. From our experiments we could now conclude with some certainty that the coin is fair. However, if the outcomes were $N = 10$, $i = 2$, and $N = 20$, $i = 5$, and $N = 40$, $i = 7$, then the posterior PMFs would appear as in Figure 8.11. We would then conclude that the coin is weighted and is biased against yielding a head, since the posterior PMF is concentrated about 0.2. See [Kay 1993] for futher descriptions of Bayesian approaches to estimation.

References

Kay, S., *Fundamentals of Statistical Signal Processing; Estimation Theory, Vol. I*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

Tennebaum, J.B. "Bayesian modeling of human learning", in *Advances in Neural Information Processing Systems 11*, MIT Press, Cambridge, MA, 1999.

Problems

8.1 (w) A fair coin is tossed. If it comes up heads, then $X = 1$ and if it comes up tails, then $X = 0$. Next, a point is selected at random from the area A if $X = 1$ and from the area B if $X = 0$ as shown in Figure 8.12. Note that the area of the square is 4 and A and B both have areas of 3/2. If the point selected is in an upper quadrant, we set $Y = 1$ and if it is in a lower quadrant, we set $Y = 0$. Find the conditional PMF $p_{Y|X}[j|i]$ for all values of i and j. Next, compute $P[Y = 0]$.

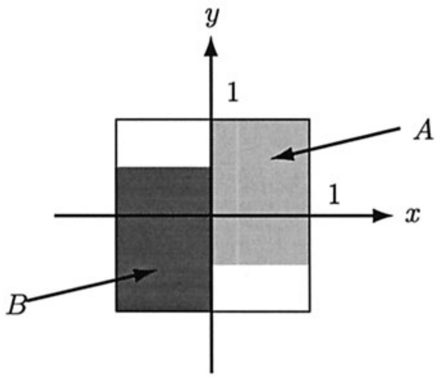


Figure 8.12: Areas for Problem 8.1.

8.2 (⌣) **(w)** A fair coin is tossed with the outcome mapped into $X = 1$ for a head and $X = 0$ for a tail. If it comes up heads, then a fair die is tossed. The outcome of the die is denoted by Y and is set equal to the number of dots observed. If the coin comes up tails, then we set $Y = 0$. Find the conditional PMF $p_{Y|X}[j|i]$ for all values of i and j. Next, compute $P[Y = 1]$.

8.3 (w) A fair coin is tossed 3 times in succession. All the outcomes (i.e., the 3-tuples) are equally likely. The random variables X and Y are defined as

$$\begin{aligned} X &= \begin{cases} 0 & \text{if outcome of first toss is a tail} \\ 1 & \text{if outcome of first toss is a head} \end{cases} \\ Y &= \text{number of heads observed for the three tosses} \end{aligned}$$

Determine the conditional PMF $p_{Y|X}[j|i]$ for all i and j.

8.4 (t) Prove that $\sum_{j=-\infty}^{\infty} p_{Y|X}[y_j|x_i] = 1$ for all x_i.

8.5 (⌣) **(w)** Are the following functions valid conditional PMFs

a. $p_{Y|X}[j|x_i] = (1 - x_i)^j x_i \qquad j = 1, 2, \ldots; x_i = 1/4, 1/2, 3/4$

b. $p_{Y|X}[j|x_i] = \binom{N}{j} x_i^j (1 - x_i)^{N-j} \qquad j = 0, 1, \ldots, N; x_i = -1/2, 1/2$

c. $p_{Y|X}[j|x_i] = c x_i^j \qquad j = 2, 3, \ldots; x_i = 2$ for c some constant?

8.6 (⌣) **(f)** If

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{6} & i = 0, j = 0 \\ \frac{1}{3} & i = 0, j = 1 \\ \frac{1}{3} & i = 1, j = 0 \\ \frac{1}{6} & i = 1, j = 1 \end{cases}$$

find $p_{Y|X}$ and $p_{X|Y}$.

8.7 (f) Verify the conditional PMF given in (8.10).

8.8 (⌣̈) (f) For the sample space shown in Figure 8.1 determine $p_{Y|X}$ and $p_{X|Y}$ if all the outcomes are equally likely. Explain your results.

8.9 (w) Explain the need for the denominator term in (8.11) and (8.12).

8.10 (w) If $p_{Y|X}$ and p_Y are known, can you find $p_{X,Y}$?

8.11 (⌣̈) (w) A box contains three types of replacement light bulbs. There is an equal proportion of each type. The types vary in their quality so that the probability that the light bulb *fails* at the jth use is given by

$$\begin{aligned} p_{Y|X}[j|1] &= (0.99)^{j-1}0.01 \\ p_{Y|X}[j|2] &= (0.9)^{j-1}0.1 \\ p_{Y|X}[j|3] &= (0.8)^{j-1}0.2 \end{aligned}$$

for $j = 1, 2, \ldots$. Note that $p_{Y|X}[j|i]$ is the PMF of the bulb failing at the jth use if it is of type i. If a bulb is selected at random from the box, what is the probability that it will operate satisfactorily for at least 10 uses?

8.12 (f) A joint PMF $p_{X,Y}[i,j]$ has the values shown in Table 8.2. Determine the conditional PMF $p_{Y|X}$. Are the random variables independent?

	$j=1$	$j=2$	$j=3$
$i=1$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{2}{10}$
$i=2$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{10}$
$i=3$	$\frac{3}{10}$	$\frac{1}{20}$	$\frac{1}{20}$

Table 8.2: Joint PMF for Problem 8.12.

8.13 (⌣̈) (w) A random vector (X,Y) has a sample space shown in Figure 8.13 with the sample points depicted as solid circles. The four points are equally probable. Note that the points in Figure 8.13b are the corners of the square shown in Figure 8.13a after rotation by $+45°$. For both cases compute $p_{Y|X}$ and p_Y to determine if the random variables are independent.

8.14 (t) Use the properties of conditional probability and the definition of the conditional PMF to prove (8.23). Hint: Let $A = \cup_j\{s : Y(s) = y_j\}$ and note that the events $\{s : Y(s) = y_j\}$ are mutually exclusive.

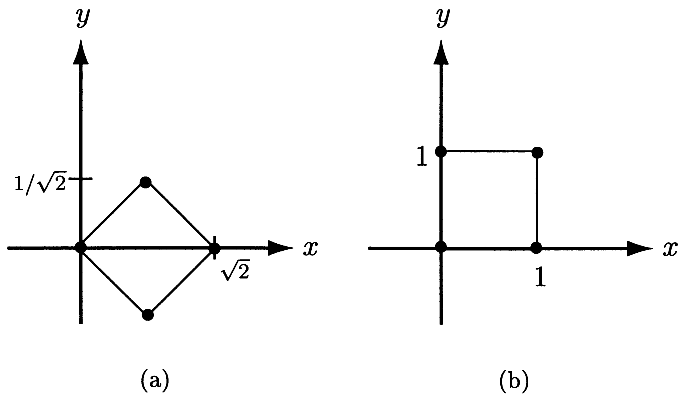


Figure 8.13: Joint PMFs – each point is equally probable.

8.15 (w) If X and Y are independent random variables, find the PMF of $Z = |X - Y|$. Assume that $\mathcal{S}_X = \{0, 1, \ldots\}$ and $\mathcal{S}_Y = \{0, 1, \ldots\}$. Hint: The answer is

$$p_Z[k] = \begin{cases} \sum_{i=0}^{\infty} p_X[i] p_Y[i] & k = 0 \\ \sum_{i=0}^{\infty} (p_Y[i] p_X[i+k] + p_X[i] p_Y[i+k]) & k = 1, 2, \ldots \end{cases}.$$

As an intermediate step show that

$$p_{Z|X}[k|i] = \begin{cases} p_Y[i] & k = 0 \\ p_Y[i+k] + p_Y[i-k] & k \neq 0. \end{cases}$$

8.16 (w) Two people agree to meet at a specified time. Person A will be late by i minutes with a probability $p_X[i] = (1/2)^{i+1}$ for $i = 0, 1, \ldots$, while person B will be late by j minutes with a probability of $p_Y[j] = (1/2)^{j+1}$ for $j = 0, 1, \ldots$. The persons arrive independently of each other. The first person to arrive will wait a maximum of 2 minutes for the second person to arrive. If the second person is more than 2 minutes late, the first person will leave. What is the probability that the two people will meet? Hint: Use the results of Problem 8.15.

8.17 (⌣̈) **(w)** If X and Y are independent random variables, both of whose PMFs take on values $\{0, 1, \ldots\}$, find the PMF of $Z = \min(X, Y)$.

8.18 (w) If X and Y have the joint PMF

$$p_{X,Y}[i, j] = p_1 p_2 (1 - p_1)^i (1 - p_2)^j \qquad i = 0, 1, \ldots; j = 0, 1, \ldots$$

where $0 < p_1 < 1$, $0 < p_2 < 1$, find $P[Y > X]$ using a conditioning argument. In particular, make use of (8.23) and $P[Y > X | X = i] = P[Y > i | X = i]$.

8.19 (f) If X and Y have the joint PMF given in Problem 8.6, find $E_{Y|X}[Y|x_i]$.

8.20 (f) If X and Y have the joint PMF

$$p_{X,Y}[i,j] = \left(\frac{1}{2}\right)^{i+1} \exp(-\lambda)\frac{\lambda^j}{j!} \qquad i = 0,1,\ldots; j = 0,1,\ldots$$

find $E_{Y|X}[Y|i]$ for all i.

8.21 (⌣) (f) Find the conditional mean of Y given X if the joint PMF is uniformly distributed over the points $\mathcal{S}_{X,Y} = \{(0,0),(1,0),(1,1),(2,0),(2,1),(2,2)\}$.

8.22 (⌣) (f) For the joint PMF given in Problem 8.21 determine $\operatorname{var}(Y|x_i)$ for all x_i. Explain why your results appear to be reasonable.

8.23 (t) Prove that $\operatorname{var}(Y|x_i) = E_{Y|X}[Y^2|x_i] - E^2_{Y|X}[Y|x_i]$ by using (8.31).

8.24 (f) Find $E_Y[Y]$ for the joint PMF given in Problem 8.21. Do this by using the definition of the expected value and also by using (8.36).

8.25 (t) Prove the extension of (8.36) which is

$$E_Y[g(Y)] = E_X\left[E_{Y|X}[g(Y)|X]\right]$$

where $h(X) = E_{Y|X}[g(Y)|X]$ is a function of the random variable X which takes on values

$$h(x_i) = E_{Y|X}[g(Y)|x_i] = \sum_j g(y_j) p_{Y|X}[y_j|x_i].$$

This says that $E_Y[g(Y)]$ can be computed using the formula

$$E_Y[g(Y)] = \sum_i \left[\sum_j g(y_j) p_{Y|X}[y_j|x_i]\right] p_X[x_i].$$

8.26 (t) In this problem we prove that if $M \sim \operatorname{Pois}(\lambda)$ and Y conditioned on M is a binomial PMF with parameter p, then the unconditional PMF of Y is $\operatorname{Pois}(\lambda p)$. This means that if

$$p_M[m] = \exp(-\lambda)\frac{\lambda^m}{m!} \qquad m = 0,1,\ldots$$

and

$$p_{Y|M}[j|m] = \binom{m}{j} p^j (1-p)^{m-j} \qquad j = 0,1,\ldots,m$$

then

$$p_Y[j] = \exp(-\lambda p)\frac{(\lambda p)^j}{j!} \qquad j = 0,1,\ldots.$$

To prove this you will need to derive the characteristic function of Y and show that it corresponds to a Pois(λp) random variable. Proceed as follows, making use of the results of Problem 8.25

$$\begin{aligned}\phi_Y(\omega) &= E_Y[\exp(j\omega Y)] \\ &= E_M\left[E_{Y|M}[\exp(j\omega Y)|M]\right] \\ &= E_M\left[[p\exp(j\omega) + (1-p)]^M\right]\end{aligned}$$

and complete the derivation.

8.27 (t) In Chapter 7 the optimal *linear* predictor of Y based on $X = x_i$ was found. The criterion of optimality was the minimum mean square error, where the mean square error was defined as $E_{X,Y}[(Y - (aX + b))^2]$. In this problem we prove that the best predictor, now allowing for *nonlinear predictors* as well, is given by the conditional mean $E_{Y|X}[Y|x_i]$. To prove this we let the predictor be $\hat{Y} = g(X)$ and minimize

$$\begin{aligned}E_{X,Y}[(Y - g(X))^2] &= \sum_i \sum_j (y_j - g(x_i))^2 p_{X,Y}[x_i, y_j] \\ &= \sum_i \left[\sum_j (y_j - g(x_i))^2 p_{Y|X}[y_j|x_i]\right] p_X[x_i].\end{aligned}$$

But since $p_X[x_i]$ is nonnegative and we can choose a different value of $g(x_i)$ for each x_i, we can equivalently minimize

$$\left[\sum_j (y_j - g(x_i))^2 p_{Y|X}[y_j|x_i]\right]$$

where we consider $g(x_i) = c$ as a constant. Prove that this is minimized for $g(x_i) = E_{Y|X}[Y|x_i]$. Hint: You may wish to review Section 6.6.

8.28 (⌣̈) (f) For random variables X and Y with the joint PMF

$$p_{X,Y}[i,j] = \begin{cases} \frac{1}{4} & (i,j) = (-1,0) \\ \frac{1}{8} & (i,j) = (0,-1) \\ \frac{3}{8} & (i,j) = (0,1) \\ \frac{1}{4} & (i,j) = (1,0) \end{cases}$$

we wish to predict Y based on our knowledge of the outcome of X. Find the optimal predictor using the results of Problem 8.27. Also, find the optimal *linear* predictor for this problem (see Section 7.9) and compare your results. Draw a picture of the sample space using solid circles to indicate the sample points in a plane and then plot the prediction for each outcome of $X = i$ for $i = -1, 0, 1$. Explain your results.

8.29 (c) Test out the MATLAB program given in Section 8.7 to generate realizations of the vector random variable (X, Y) whose joint PMF is given in Figure 8.8. Do so by estimating the joint PMF or $p_{X,Y}[i, j]$. You may wish to review Section 7.11.

8.30 (⌣̈) (w,c) For the joint PMF given in Figure 8.8 determine the conditional mean $E_{Y|X}[j|i]$ and then verify your results using a computer simulation. Note that you will have to separate the realizations (x_m, y_m) into two sets, one in which $x_m = 0$ and one in which $x_m = 1$, and then use the sample average of each set as your estimator.

8.31 (w,c) For the joint PMF given in Figure 8.8 determine $E_Y[Y]$. Then, verify (8.36) by using your results from Problem 8.30, and computing

$$\widehat{E_Y[Y]} = \widehat{E_{Y|X}[Y|0]}\hat{p}_X[0] + \widehat{E_{Y|X}[Y|1]}\hat{p}_X[1]$$

where $\widehat{E_{Y|X}[Y|0]}$ and $\widehat{E_{Y|X}[Y|1]}$ are the values obtained in Problem 8.30. Also, the PMF of X,which needs to be estimated, can be done so as described in Section 5.9.

8.32 (w,c) For the posterior PMF given by (8.39) plot the PMF for $i = N/2$, $M = 11$ and increasing N, say $N = 10, 30, 50, 70$. What happens as N becomes large? Explain your results. Hint: You will need a computer to evaluate and plot the posterior PMF.