8.1 Introduction

Recall conditional probability (Ch. 4).

Example: Fair die Event A = \{4\}

Event describing our previous knowledge \( B = \{2, 4, 6\} \)

\[
P(A/B) = \frac{N_{AB}}{N_B} = \frac{N_{AB}/N_S}{N_B/N_S} = \frac{P(AB)}{P(B)}
\]

Two ways.

(i) Toss die 50 times

- Count total # of 2's, 4's, and 6's (reduced sample size)
- Count total # of 4's.

Get \( \frac{N_A}{N_B} = \frac{9}{25} \approx 0.3 \) (relative frequency interpretation of probability)

(ii) \( P(A/B) = \frac{P(AB)}{P(B)} \)

\( AB \cap B = \{4\} \cap \{2, 4, 6\} = \{4\} \)

\[
P(A|B) = \frac{P(AB)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{3}
\]

(using original sample space \( S \))
Ex: 2 (Example 4.2) - A compound experiment

\[
\begin{array}{c|c}
R - p_1 & R - p_2 \\ \hline
B - (1-p_1) & B - (1-p_2) \\
\hline
\text{Urn 1} & \text{Urn 2}
\end{array}
\]

Compound experiment performed:
Urn chosen at random
Then a ball selected from that urn.

\[
P[\text{red ball selected} ~ | ~ A] = P[A/\text{Urn 1}]P[\text{Urn 1}] + P[A/\text{Urn 2}]P[\text{Urn 2}]
\]

Ex: 3 (ch.8) Expt: choose 2 coins; toss it 4 times

\[
P[\{2 or more heads\}] = ?
\]

but can depend on which coin chosen of course.

So more convenient to define

conditional PMF

\[
P_X[k | \text{coin 1 chosen}]
\]

\[
P_X[k | \text{coin 2 chosen}]
\]
since we know which coin was chosen, we have the PMF, which in this case is the binomial. Once we have the PMFs we use the law of total probability:

\[ P(A) = \sum_{i=1}^{N} P(A|B_i)P(B_i) \quad (4.4) \]

to define:

\[ p_X[k] = p_X[k|\text{coin 1 chosen}] P(\text{coin 1 chosen}) \]
\[ + p_X[k|\text{coin 2 chosen}] P(\text{coin 2 chosen}) \]

and finally:

\[ P(X \geq 2) = \sum_{k=2}^{\infty} p_X[k] \]
8.3 CPMF

Introductory example:

(i) Choose coin 1 or coin 2  \( S_X = \{1, 2\} \)

\[ p_X(i) = \begin{cases} \alpha, & i = 1 \\ (1-\alpha), & i = 2 \end{cases} \quad (8.1) \]

(ii) Second part of experiment: Toss coin 4 times.

but # of heads = \( Y \)  \( S_Y = \{0, 1, 2, 3, 4\} \)

so, sample space  \( S_{X,Y} = S_X \times S_Y \)

\[
\begin{aligned}
 & & & Y \\
 & & 4 & 3 & 2 & 1 & 0 \\
 & 4 & & & & & \\
 & 3 & & & & & \\
 & 2 & & & & & \\
 & 1 & & & & & \\
 & 0 & & & & & \\
 & & & & x & & \\
 & & 1 & 2 & & & \\
 \end{aligned}
\]

Clearly,

\[ P[A] = P[\{2\text{ heads or more}\}] \]

\[ = \sum_{(i,j): (i,j) \in A} p_{x,y}[i,j] \]
\[
= \sum_{i=1}^{2} \sum_{j=2}^{4} p_{X,Y}(x_i, y_j) \quad (8.2)
\]

To determine \(p_{X,Y}(x_i, y_j)\),

Recall (Ch.7) \(p_{X,Y}(x_i, y_j) = P[X=x_i, Y=y_j]\)

\(p_X(x_i) = P[X=x_i]\)

Using definition of conditional probability of \(X_{\text{obs}}\):

\(p_{X,Y}(x_i, y_j) = P[X=x_i, Y=y_j]\)

\(= P[\{s : X(s) = x_i, Y(s) = y_j\}]\)

\(= P[\{s : Y(s) = y_j\} / \{s : X(s) = x_i\}] \cdot P[X=x_i]\)

\(= P[Y=y_j / X=x_i] \cdot P[X=x_i]\) \quad (8.3)

Now for a given value of \(X = x_i\),

\(P[Y=y_j / X=x_i] = \binom{4}{j} p_i^j (1-p_i)^{4-j} \quad (8.4)\)

This has the usual properties of a \(PMF\):

\(0 \leq P[Y=y_j / X=x_i] \leq 1 \quad \sum_{j=0}^{4} P[Y=y_j / X=x_i] = 1\)
So we define,

\[ P_{Y/X}[j|i] = P[Y=j/X=i] \]

So, using (8.3)

\[ P_{X,Y}[i,j] = P_{Y/X}[j|i] \cdot P_X[i] \]

\[ = \binom{4}{j} p_1^j (1-p_1)^{4-j} \cdot \alpha \quad i=1, j=0,1,\ldots, 4 \]

\[ = \binom{4}{j} p_2^j (1-p_2)^{4-j} \cdot \alpha \quad i=2, j=0,1,\ldots, 4 \]

Finally,

\[ P[A] = \sum_{j=2}^{4} P_{X,Y}[i,j] + \sum_{j=2}^{4} P_{X,Y}[2,j] \]

We also have, from (8.5),

\[ P_{Y/X}[j,i] = \frac{P_{X,Y}[i,j]}{P_X[i]} \quad (8.6) \]
6.4 Joint, Conditional & Marginal PMFs

More generally,

$$P_{y/x} (y_j|x_i) = \frac{P_{x,y}(x_i,y_j)}{P_x(x_i)} \quad \text{(8.7)}$$

Once again, $P_{y/x}(y_j|x_i)$ valid PMFs for $x_i = \text{constant}$.

For the previous example,

$$\sum_{j=0}^{10} P_{y/x}(y_j|x) = 1$$

$$\sum_{j=-6}^{6} P_{y/x}(y_j|x) = 1$$

Example 8.1

<table>
<thead>
<tr>
<th>$x$</th>
<th>$j=1$</th>
<th>$j=2$</th>
<th>$j=3$</th>
<th>$j=4$</th>
<th>$j=5$</th>
<th>$j=6$</th>
<th>$j=7$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
</tbody>
</table>

Outcomes
want conditional PMF of the sum if known sum $x$.

Let $y$ be the sum.

$X = 1$, sum even;
$X = 0$, sum odd.

want $P_{Y/X}[j | 11]$ and $P_{Y/X}[j | 10]$ $j = 2, 3, \ldots, 12$

$S_Y = \{2, 3, \ldots, 12\}$

we know,

$$P_{Y/X}[j | 11] = \frac{P_{X,Y}[1, j]}{P_X[1]}, j = 2, 4, 6, 8, 10, 12$$

$$P_{Y/X}[j | 11] = \begin{cases} \frac{1}{18}, & j = 2 \\ \frac{3}{18}, & j = 4 \\ \frac{5}{18}, & j = 6 \\ \frac{5}{18}, & j = 8 \\ \frac{3}{18}, & j = 10 \\ \frac{1}{18}, & j = 12 \\ \end{cases} \frac{1/36}{(2/36, 1/2)}$$

$$\sum_j P_{Y/X}[j | 11] = 1$$

(Note: $\sum_j P_{Y/X}[j | 11] = 1$)
Simly:

\[ P_{Y/X} [y_j / x_i] = \begin{cases} 
\frac{2}{15}, & j = 3 \\
\frac{4}{15}, & j = 5 \\
\frac{6}{15}, & j = 7 \\
\frac{4}{15}, & j = 9 \\
\frac{2}{15}, & j = 11 
\end{cases} \]

**Note:** Could also have obtained the conditional PMF based on the restricted sample space.

**Property 8.1**

\[ P_{Y/X} [y_j / x_i] = \frac{P_{X,Y} [x_i, y_j]}{\sum_{j} P_{X,Y} [x_i, y_j]} \]

\[ P_{X/Y} [x_i / y_j] = \ldots \]

**Property 8.2**

\[ P_{X/Y} [x_i, y_j] = \frac{P_{Y/X} [y_j / x_i] \cdot P_x [x_i]}{P_y [y_j]} \]
Property 3.3: Bayes' Rule

\[
P_{Y|x}(y_j|x_i) = \frac{P_{X|Y}(x_i|y_j) \cdot P_Y(y_j)}{\sum_j P_{X|Y}(x_i|y_j) \cdot P_Y(y_j)}
\]

Property 4.1

\[
P_{X,Y}(x_i,y_j) = P_{Y|x}(y_j|x_i) \cdot P_X(x_i)
\]
\[
P_{X,Y}(x_i,y_j) = P_{X|Y}(x_i|y_j) \cdot P_Y(y_j)
\]

Property 4.5

\[
P_Y(y_j) = \sum_i P_{X,Y}(x_i,y_j)
\]
\[
= \sum_i P_{Y|x}(y_j|x_i) \cdot P_X(x_i)
\]
8.5 Probability Calculations using conditionals.

Ex. 8.2 PMF for $Z = \max(X, Y)$

8.6 Mean of conditional PMF.

Recall: Conditional PMF $P_{Y/X}[y_j|x_i]$ is a PMF for fixed $i$.

So we can determine its mean.

$$E_{Y/X}[Y/X_i] = \sum_j y_j P_{Y/X}[y_j|x_i]$$

Conditional mean will of course depend on $x_i$.

Ex. 8.3 (Ref. Ex. 8.1 Two dice tossed.

$Y$-sum $X = 1$ if sum even,

$\sum = 0$ if sum odd.

We obtained the conditional PMFs $P_{Y/X}[y_j|1]$,

$P_{Y/X}[y_j|0]$

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

$E_{Y/X}[Y|0] = \ldots \ldots$
Can also redefine the variance of the conditional PMF.

\[
\text{var} \left( Y | x_i \right) = \sum (y_j - E_{Y|x_i}[Y|x_i])^2 \cdot p_{Y|x_i}(y_j|x_i)
\]

**NOTE:** \( E_{Y|x_i}[Y|x_i] \) is the mean of conditional PMF. This also referred to as **conditional mean**.

Can also determine conditional mean of \( g(Y) \).

i.e. \( E_{Y|x_i}[g(Y)|x_i] \) in the usual way.

**Ex. 8.4**

**Two dice**

- Die 1: 1, 2, 3, 4, 5 or 6 dots.
- Die 2: 2, 3, 2, 3, 2 or 3 dots.

\{ Each face equally likely \}

Die selected randomly.

Expected # of dots for the tossed die?

So, want conditional mean.
Set up: \[ X = \begin{cases} 1 & \text{if die 1 chosen} \\ 2 & \text{if die 2 chosen} \end{cases} \]

\[ Y = \# \text{ of dots observed} \]

Want \( E_{Y|x}[Y|1] \) and \( E_{Y|x}[Y|2] \)

For die 1, \( P_{Y|x}[j|1] = \frac{1}{6}, \quad j = 1, 2, \ldots, 6 \)

For die 2, \( P_{Y|x}[j|2] = \frac{1}{2}, \quad j = 2, 3 \)

So: \[ E_{Y|x}[Y|1] = \sum_{j=1}^{6} y_j \cdot \frac{1}{6} = \frac{1}{6} + \frac{2}{6} + \frac{3}{6} + \cdots + \frac{6}{6} = \frac{21}{6} = \frac{7}{2} \]

\[ E_{Y|x}[Y|2] = \sum_{j=2}^{3} y_j \cdot \frac{1}{2} = \frac{2}{2} = 1 \]

Text shows a comparison with actual simulation.

Figs. 8.7(a), 8.7(b)

Also, can simplify \( E[Y] - \text{unconditional mean} \):

\[ E[Y] = \sum_{x} E_{Y|x}[Y|x_i] \cdot P_X[x_i] \quad (8.35) \]
8.8 Real-World Example - Modeling Human Learning

(Using a Bayesian framework).

Example: Coin Tossing.

Wish to determine whether coin fair (p = \frac{1}{2})
or not (p \neq \frac{1}{2}).

Method 1: Toss coin repeatedly. (Relative frequency interpretation)
Obtain \( p \approx \frac{N_{\text{H}}}{N_{\text{Total}}} \)

Method 2: Let \( p \) be a RV. Then we find its
PMF.

\[ \begin{array}{c}
p & 1 - p \\
\uparrow & \downarrow \\
\text{our state of knowledge} & \text{low on the value of } p. \\
\end{array} \]

\[ \begin{array}{c}
\uparrow & \downarrow \\
\text{our state of knowledge higher.} \\
\end{array} \]
1. Assume some PMF for $p$.

$$P(y_j) = \frac{1}{n+1}, \quad y_j = 0, 1, \ldots, \frac{n-1}{m}, 1$$

2. Now do the "learning experiments".

Toss coin $N$ times.

Let $X = \# \text{ of heads}$.

We know $X$ binomially distributed.

But require probability of heads ($y_j$).

Since this not known, we have

$$P(x|y_i|y_j) = \binom{N}{i} (y_j)^i (1-y_j)^{N-i}, \quad i = 0, 1, \ldots, N$$

3. We want to learn "$y_j$" as we toss the coin and count the number of heads.
That is, we are interested in the PMF of \( Y \) (probability of Heads) after observing outcomes of \( N \) tosses. So we want \( P_{Y|x}[y_j|i] \) — called the posterior PMF since determined after experimentation.

Using Bayes rule:

\[
P_{Y|x}[y_j|i] = \frac{P_{X|Y}[i|y_j] \cdot P_Y(y_j)}{\sum_j P_{X|Y}[i|y_j] \cdot P_Y(y_j)}
\]

\[
= \frac{(N_i y_j (1-y_j)^{N-i} \cdot \frac{1}{M+1}}{\sum_{j=0}^{M} \left( \begin{array}{c} N \\ i \end{array} \right) y_j^i (1-y_j)^{N-i} \cdot \frac{1}{M+1}}
\]

\( y_j = \frac{0, 1, \ldots, 1}{M} \)

\( i = 0, 1, \ldots, N \)
and

\[ P_{Y|X}[y_j|i] = \frac{y_j^i (1-y_j)^{(N-i)}}{\sum_{j=0}^{m} y_j^i (1-y_j)^{(N-i)}} \]

\[ y_j = 0, \frac{1}{m}, \ldots, 1 \]

\[ i = 0, 1, \ldots, N \]

3.10: Posterior PMFs for coin tossing analogy to human learning – coin to be fair. The \( y_j \)'s are possible probability values for a head.

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.