$$
\begin{aligned}
P_{X|A}(-2) &= \frac{13}{8} P(X=-2, A) \\
&= \frac{13}{8} P_{XY}(-2,0) = \frac{1}{8}, \\
P_{X|A}(-1) &= \frac{13}{8} P(X=-1, A) \\
&= \frac{13}{8} \big[P_{XY}(-1,0) + P_{XY}(-1,1)\big] = \frac{2}{8} = \frac{1}{4}, \\
P_{X|A}(\ 0\ ) &= \frac{13}{8} P(X=0, A) \\
&= \frac{13}{8} \big[P_{XY}(0,0) + P_{XY}(0,1)\big] = \frac{2}{8} = \frac{1}{4}, \\
P_{X|A}(\ 1\ ) &= \frac{13}{8} P(X=1, A) \\
&= \frac{13}{8} \big[P_{XY}(1,0) + P_{XY}(1,1)\big] = \frac{2}{8} = \frac{1}{4}, \\
P_{X|A}(\ 2\ ) &= \frac{13}{8} P(X=2, A) \\
&= \frac{13}{8} P_{XY}(2,0) = \frac{1}{8}.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
E[X|A] &= \sum_{x_i \in R_X} x_i P_{X|A}(x_i) \\
&= (-2)\frac{1}{8} + (-1)\frac{1}{4} + (0)\frac{1}{4} + (1)\frac{1}{4} + (2)\frac{1}{8} = 0.
\end{aligned}
$$

c. To find $E\big[|X|\,\big|\, -1 < Y < 2\big]$, we use the conditional PMF and LOTUS. We have

$$
\begin{aligned}
E[|X||A] &= \sum_{x_i \in R_X} |x_i| P_{X|A}(x_i) \\
&= |-2| \cdot \frac{1}{8} + |-1| \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} = 1.
\end{aligned}
$$

Conditional expectation has some interesting properties that are used commonly in practice. Thus, we will revisit conditional expectation in Section 5.1.5, where we discuss properties of conditional expectation, conditional variance, and their applications.

Law of Total Probability:

Remember the law of total probability: If $B_1, B_2, B_3, ...$ is a partition of the sample space S, then for any event A we have

$$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A|B_i)P(B_i).$$

If Y is a discrete random variable with range $R_Y = \{y_1, y_2, \ldots\}$, then the events $\{Y = y_1\}$, $\{Y = y_2\}$, $\{Y = y_3\}$, $\cdots$ form a partition of the sample space. Thus, we can use the law of total probability. In fact we have already used the law of total probability to find the marginal PMFs:

$$P_X(x) = \sum_{y_j \in R_Y} P_{XY}(x, y_j) = \sum_{y_j \in R_Y} P_{X|Y}(x|y_j)P_Y(y_j).$$

We can write this more generally as

$$P(X \in A) = \sum_{y_j \in R_Y} P(X \in A|Y = y_j)P_Y(y_j), \quad \text{for any set } A.$$

We can write a similar formula for expectation as well. Indeed, if $B_1, B_2, B_3, \ldots$ is a partition of the sample space S, then

$$EX = \sum_i E[X|B_i]P(B_i).$$

To see this, just write the definition of $E[X|B_i]$ and apply the law of total probability. The above equation is sometimes called the law of total expectation [2].

Law of Total Probability:

$$P(X \in A) = \sum_{y_j \in R_Y} P(X \in A|Y = y_j)P_Y(y_j), \quad \text{for any set } A.$$

Law of Total Expectation:

1. If $B_1, B_2, B_3, \ldots$ is a partition of the sample space S,

$$EX = \sum_i E[X|B_i]P(B_i) \qquad (5.3)$$

2. For a random variable X and a discrete random variable Y,

$$EX = \sum_{y_j \in R_Y} E[X|Y = y_j]P_Y(y_j) \qquad (5.4)$$

Example 5.6

Let $X \sim Geometric(p)$. Find EX by conditioning on the result of the first "coin toss."

Solution

Remember that the random experiment behind $Geometric(p)$ is that we have a coin with $P(H) = p$. We toss the coin repeatedly until we observe the first heads. X is the total number of coin tosses. Now, there are two possible outcomes for the first coin toss: H or T. Thus, we can use the law of total expectation (Equation 5.3):

$$\begin{aligned} EX &= E[X|H]P(H) + E[X|T]P(T) \\ &= pE[X|H] + (1-p)E[X|T] \\ &= p \cdot 1 + (1-p)(EX+1). \end{aligned}$$

In this equation, $E[X|T] = 1 + EX$, because the tosses are independent, so if the first toss is tails, it is like starting over on the second toss. Solving for EX, we obtain

$$EX = \frac{1}{p}.$$

Example 5.7

Suppose that the number of customers visiting a fast food restaurant in a given day is $N \sim Poisson(\lambda)$. Assume that each customer purchases a drink with probability p, independently from other customers and independently from the value of N. Let X be the number of customers who purchase drinks. Find EX.

Solution

By the above information, we conclude that given $N = n$, then X is a sum of n independent $Bernoulli(p)$ random variables. Thus, given $N = n$, X has a binomial distribution with parameters n and p. We write

$$X|N = n \;\; \sim \;\; Binomial(n, p).$$

That is,

$$P_{X|N}(k|n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Thus, we conclude

$$E[X|N=n] = np.$$

Thus, using the law of total probability, we have

$$\begin{aligned} E[X] &= \sum_{n=0}^{\infty} E[X|N=n] P_N(n) \\ &= \sum_{n=0}^{\infty} np P_N(n) \\ &= p \sum_{n=0}^{\infty} n P_N(n) = pE[N] = p\lambda. \end{aligned}$$