1 Multivariate Distributions

1.1 Basic Terminology

1.1.1 Probability Mass Functions

We want to build a theory of probability for more than 1 variable. Suppose that X and Y are discrete random variables defined on the same sample space. The probabilities of objects involving both X and Y are encoded by the joint PMF.

Definition 1. The *joint probability (mass) function* of X and Y is

$$f_{X,Y}(x,y) = \mathbb{P}(\{\omega \in S : X(\omega) = x\} \cap \{\omega \in S : Y(\omega) = y\})$$

for $x \in X(S), y \in Y(S)$ and 0 otherwise. As in the univariate case, a shorthand notation for this is

$$f_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y).$$

The joint PMF is still a probability function in the sense that

1. $0 \leq f_{X,Y}(x,y) \leq 1$
2. $\sum_{x,y} f_{X,Y}(x,y) = 1$.

The probabilities of only one random variable are encoded by the marginal PMF.

Definition 2. Suppose that X and Y are *discrete* random variables with joint probability function $f_{X,Y}(x,y)$. The *marginal probability mass function* of X is

$$f_X(x) = \mathbb{P}(X = x) = \mathbb{P}(X = x, Y \in Y(S)) = \sum_{y \in Y(S)} f(x,y).$$

Similarly, the marginal distribution of Y is

$$f_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(X \in X(S), Y = y) = \sum_{x \in X(S)} f(x,y).$$

Remark 1. The marginal probability mass functions are the same as the PMFs we encountered before.

1.1.2 Independence

Recall we say that events A and B are independent, if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. In the context of random variables, this is the same as

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y) \quad \forall x, y.$$

Definition 3. X and Y are *independent* random variables if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

for all values of (x,y).

1.1.3 Conditional Distributions

Recall that for events A, B with $\mathbb{P}(B) \neq 0$ we defined

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

In the context of random variables, this is the same as

$$\mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} \quad \text{and} \quad \mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)}$$

Definition 4. The *conditional probability mass function* of X given $Y = y$ is

$$f_{X \mid Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)} \text{ provided that } f_Y(y) > 0.$$

Similarly, the *conditional probability mass function* of Y given $X = x$ is

$$f_{Y \mid X}(y \mid x) = \frac{f(x, y)}{f_X(x)}, \text{ provided that } f_X(x) > 0.$$

For each fixed y, the function $f_X(x \mid y)$ is the probability mass function of the random variable $X \mid Y = y$ and has the usual properties, such as summing to 1.

1.1.4 Expected Value

Definition 5. Suppose X and Y are discrete random variables with joint probability function $f_{X,Y}(x, y)$. Then for any function $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}\left[g(X, Y)\right] = \sum_{(x,y)} g(x, y) f_{X,Y}(x, y).$$

Properties:

1. *Linearity of Expectation:* If X and Y are any random variables, then

$$\mathbb{E}[ag_1(X, Y) + bg_2(X, Y)] = a \cdot \mathbb{E}[g_1(X, Y)] + b \cdot \mathbb{E}[g_2(X, Y)].$$

   In particular, if X and Y are any random variables (not necessarily independent), then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

2. *Product of two Independent Random Variables:* If X and Y are independent, then

$$\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\,\mathbb{E}[g_2(Y)].$$

   In particular, if X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

1.2 Random Vectors

All the terminology above can be extended to a collection $X_1, X_2, \ldots, X_n$ of random variables in the obvious way.

Definition 6. For a collection of n discrete random variables, $X_1, ..., X_n$, the joint probability function is defined as

$$f_{X_1,...,X_n}(x_1, x_2, ..., x_n) = \mathbb{P}(X_1 = x_1, X_2 = x_2, ..., X_n = x_n).$$

and we call the vector $(X_1, \ldots, X_n)$ a random vector.

Definition 7. $X_1, X_2, \ldots, X_n$ are *independent* if

$$f_{X_1,...,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

for all values of $(x_1, \ldots, x_n)$.

Definition 8. If $g : \mathbb{R}^n \to \mathbb{R}$, and $X_1, ..., X_n$ are discrete random variables with joint probability function $f_{X_1,...,X_n}(x_1, ..., x_n)$, then

$$\mathbb{E}\left[g(X_1, ..., X_n)\right] = \sum_{(x_1,...,x_n)} g(x_1, ..., x_n) f_{X_1,...,X_n}(x_1, ..., x_n).$$

1.2.1 Functions of Random Vectors

We have the following formula for the *probability mass function of* $U = g(X_1, X_2, \ldots, X_n)$.

$$f_U(u) = \mathbb{P}(U = u) = \sum_{\substack{(x_1,...x_n) \text{ such that} \\ g(x_1,...,x_n)=u}} f_{X_1,...,X_n}(x_1, \ldots, x_n).$$

We now list some common functions of random variables (many we have already seen).

1. **Sum of Independent Poisson is Poisson:** If $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ are independent, then
$$T = X + Y \sim \text{Poi}(\lambda_1 + \lambda_2).$$
2. **Conditional Poisson is Binomial:** Let $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ be independent. Then, given $X + Y = n$, X follows binomial distribution. That is,
$$X \mid X + Y = n \sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$
Similarly, for Y, we have
$$Y \mid X + Y = n \sim \text{Bin}\left(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right).$$
3. **Sum of Independent Binomials is Binomial:** If $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$ independently, then
$$T = X + Y \sim \text{Bin}(n + m, p).$$
4. **Sum of Independent Bernouilli is Binomial:** Let $X_1, X_2, \ldots, X_n$ be independent $\text{Bern}(p)$ random variables. Then,
$$T = X_1 + X_2 + \ldots + X_n \sim \text{Bin}(n, p).$$
5. **Sum of Independent Geometric is Negative Binommial:** Let $X_1, X_2, \ldots, X_k$ be independent $\text{Geo}(p)$ random variables. Then,
$$T = X_1 + X_2 + \ldots + X_k \sim \text{NegBin}(k, p).$$

Remark 2. Properties 3, 4, and 5 follow directly from the construction of these random variables.

1.3 Important Multivariable Distributions

1.3.1 Mulitinomial Distribution

The multinomial distribution models the number of each outcome in multiple independent experiments with k possible outcomes. The multivariate distribution is a generalization of the binomial distribution.

Definition 9. Consider an experiment in which:

1. Individual trials have k possible outcomes, and the probabilities of each individual outcome are denoted $p_i,\ 1 \leq i \leq k$, so that $p_1 + p_2 + \cdots + p_k = 1$.

2. Trials are independently repeated n times, with X_i denoting the number of times outcome i occurred, so that $X_1 + X_2 + \cdots + X_k = n$.

We say that $X_1, ..., X_k$ has a *Multinomial distribution* with parameters n and $p_1, ..., p_k$, and is denoted by

$$(X_1, ..., X_k) \sim \text{Mult}(n, p_1, ..., p_k).$$

- **Joint PMF:**
$$f_{X_1,...,X_k}(x_1, ..., x_k) = \frac{n!}{x_1!x_2! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$
The terms $\frac{n!}{x_1!x_2!\cdots x_k!} = \binom{n}{x_1,...,x_k}$ are called multinomial coefficcients.
However, since we must have $p_1 + p_2 + \cdots + p_k = 1$ and $X_1 + X_2 + \cdots + X_k = n$, the kth variable is uniquely determined by the first $k - 1$ variables,
$$p_k = 1 - p_1 - p_2 - \ldots - p_{k-1} \qquad \text{and} \qquad x_k = n - x_1 - x_2 - \ldots - x_{k-1}$$
so the PMF is sometimes written as
$$f_{X_1,...,X_{k-1}}(x_1, ..., x_{k-1}) = \frac{n!}{x_1!x_2! \cdots x_{k-1}!(n - \sum_{i=1}^{k-1} x_i)!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} \left(1 - \sum_{i=1}^{k-1} p_i\right)^{n - \sum_{i=1}^{k-1} x_i}$$
Remark 3. Notice that when $k = 2$, then we have the PMF of the Binomial distribution.

- **Marginal PMF:** The number of times the outcome i occurred is
$$X_j \sim \text{Bin}(n, p_j), \quad \text{for } j = 1, 2, \ldots, k\ .$$

- **Sum of Marginals:** The number of times the outcomes i or j occurred is
$$X_i + X_j \sim \text{Bin}(n, p_i + p_j), \quad \text{for } i \neq j.$$

- **Conditional PMF:** The number of times i occured given that i and j occurred t times is
$$X_i \mid X_i + X_j = t \sim \text{Bin}\left(t, \frac{p_i}{p_i + p_j}\right), \quad \text{for } i \neq j.$$

- **Expected Values:** The expected value of the outcomes are given by
$$\mathbb{E}[X_i X_j] = n(n-1)p_i p_j \text{ for } i \neq j \qquad \text{and} \qquad \mathbb{E}[X_i] = np_i \text{ for } i = 1, \ldots, k$$

Example 1. The following experiments can be modeled by a multinomial distribution

Experiment	X	Distribution
Draw 10 cards from a deck with replacement	# of each suit	$\text{Mult}(10, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$
Roll a dice n times	# of each roll	$\text{Mult}(n, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})$

1.4 Example Problems

1.4.1 Applications

Problem 1.1. Let $X \in \{1, 2, 3\}$ and $Y \in \{1, 2\}$, and suppose that every outcome of (X, Y) is equally likely. What is the joint PMF for the vector (X, Y)?

Solution 1.1. We can compute all the probabilities one by one and encode the joint PMF of X and Y in the table

$f_{X,Y}(x,y)$		x = 1	2	3	$f_Y(y)$
y	1	1/6	1/6	1/6	3/6
	2	1/6	1/6	1/6	3/6
$f_X(x)$		2/6	2/6	2/6	1

Problem 1.2. Suppose a fair coin is tossed 3 times. Define the random variables X = "number of Heads", and

$$Y = \begin{cases} 1 & \text{Head occurs on the first toss,} \\ 0 & \text{Tail occurs on the first toss.} \end{cases}$$

1. Find the joint PMF for (X, Y).
2. Are X and Y independent?
3. What is the conditional distribution of X given Y?
4. What is the probability that $X + Y = 2$?

Solution 1.2.

Part 1: We can compute all the probabilities one by one and encode the joint PMF of X and Y in the table

$f_{X,Y}(x,y)$		x = 0	1	2	3	$f_Y(y)$
y	0	1/8	2/8	1/8	0	1/2
	1	0	1/8	2/8	1/8	1/2
$f_X(x)$		1/8	3/8	3/8	1/8	1

Part 2: We can see

$$f_{X,Y}(0,1) = 0 \neq \frac{1}{8} \cdot \frac{1}{2} = f_X(0) f_Y(1)$$

which implies that X and Y are not independent (which makes perfect sense, as the number of heads we have should depend on whether we had heads in the first toss).

Part 3: Using the formula $f_{X \mid Y}(x \mid y) = f_{X,Y}(x,y)/f_Y(y)$ we find

	x = 0	1	2	3
$f_{X \mid Y}(x \mid y = 0)$	2/8	4/8	2/8	0
$f_{X \mid Y}(x \mid y = 1)$	0	2/8	4/8	2/8

Part 4: We have $X+Y=2$ if and only if $X=2, Y=0$ or $X=1, Y=1$. We can sum these terms up in the joint PMF

$$\mathbb{P}(X+Y=2)=f(2,0)+f(1,1)+f(0,2)=\frac{1}{8}+\frac{1}{8}=\frac{1}{4}.$$

Problem 1.3. Let X and Y be any discrete random variables. Show that

1. $0 \leq f_{X,Y}(x,y) \leq 1$
2. $f_{X,Y}(x,y) \leq f_X(x)$
3. $f_{X,Y}(x,y) \leq f_Y(y)$

Solution 1.3.

1. We have $f_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y)$ and all probabilities must be between 0 and 1.
2. We have $f_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y) \leq \mathbb{P}(X=x)=f_X(x)$ since $\{X=x,Y=y\} \subseteq \{X=x\}$.
3. We have $f_{X,Y}(x,y)=\mathbb{P}(X=x,Y=y) \leq \mathbb{P}(Y=x)=f_Y(y)$ since $\{X=x,Y=y\} \subseteq \{Y=y\}$.

Problem 1.4. Suppose X and Y have joint PMF

$$f_{X,Y}(x,y)=\frac{1}{6}\left(\frac{1}{2}\right)^x\left(\frac{2}{3}\right)^y, \quad x,y=0,1,2...$$

Find the marginal PMFs f_X and f_Y of X and Y.

Solution 1.4. Recall the identity

$$\sum_{k=0}^{\infty} q^k=\frac{1}{1-q}, \quad 0<q<1.$$

Part 1: The X marginal is

$$\begin{aligned} f_X(x) &= \sum_{y=0}^{\infty}\frac{1}{6}\left(\frac{1}{2}\right)^x\left(\frac{2}{3}\right)^y \\ &= \frac{1}{6}\left(\frac{1}{2}\right)^x\sum_{y=0}^{\infty}\left(\frac{2}{3}\right)^y=\frac{1}{6}\left(\frac{1}{2}\right)^x\frac{1}{1-\frac{2}{3}} \\ &= \frac{1}{2}\left(\frac{1}{2}\right)^x, \quad x=0,1,\ldots \end{aligned}$$

from which we conclude that $X \sim \text{Geo}(1/2)$.

Part 2: The Y marginal is

$$\begin{aligned} f_Y(x) &= \sum_{x=0}^{\infty}\frac{1}{6}\left(\frac{1}{2}\right)^x\left(\frac{2}{3}\right)^y \\ &= \frac{1}{6}\left(\frac{2}{3}\right)^y\sum_{x=0}^{\infty}\left(\frac{1}{2}\right)^x=\frac{1}{6}\left(\frac{2}{3}\right)^y\frac{1}{1-\frac{1}{2}} \\ &= \frac{1}{3}\left(\frac{2}{3}\right)^y, \quad y=0,1,\ldots \end{aligned}$$

from which we conclude that $Y \sim \text{Geo}(1/3)$.

Problem 1.5. Suppose $X \sim \text{Poi}(2)$, $Y \sim \text{Poi}(3)$, and that X and Y are independent. What is the joint probability function of X and Y?

Solution 1.5. By independence, we that for all integer valued $x, y \geq 0$,

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = e^{-2}\frac{2^x}{x!}e^{-3}\frac{3^y}{y!} = e^{-5}\frac{2^x}{x!}\frac{3^y}{y!}.$$

Problem 1.6. If we roll a die n times, let's denote by $X_1, \dots, X_6$ the number of times we rolled a 1, 2,. . . , 6.

1. What is the distribution (or marginal probability function) of X_j for $j = 1, \dots, 6$?
2. Are $X_1, X_2, \dots, X_6$ independent?
3. What is the joint probability function of $(X_1, \dots, X_6)$?
4. Let's denote by $T = X_1 + X_2$ the number of times we had a 1 or two. What's the distribution of $T = X_1 + X_2$?

Solution 1.6.

Part 1: By definition, if X_j denotes the number of times we roll a j in n rolls, then

$$X_j \sim \text{Bin}(n, \frac{1}{6}).$$

Part 2: Intuitively, these are not independent because we must have $X_1 + \cdots + X_6 = n$ so X_6 is totally determined by X_1 to X_5. For example, if we consider the case

$$\mathbb{P}(X_1 = n, X_2 = n, \dots, X_6 = n) = 0$$

but

$$\mathbb{P}(X_1 = n) \cdots \mathbb{P}(X_6 = n) = \left(\frac{1}{6}^n\right)^6 > 0$$

so they are not independent.

Part 3: Let $x_1, \dots, x_6 \in \{1, \dots, n\}$. As noted earlier, if $x_1 + x_2 + \cdots + x_6 \neq n$, then $\mathbb{P}(X_1 = x_1, \dots, X_6 = x_6) = 0$. Thus, let $x_1 + x_2 + \cdots + x_6 = n$. We can arrange the x_1 rolls of 1, x_2 rolls of 2,. . . , x_6 of rolls of 6, among the n trials in

$$\frac{n!}{x_1!x_2!\dots x_6!}$$

many ways, using the formula for the arrangements with repeated objects: the 1 is repeated x_1 times, the 2 is repeated x_2 times, etc. Each of these arrangements has probability

$$\left(\frac{1}{6}\right)^{x_1} \cdot \left(\frac{1}{6}\right)^{x_2} \cdot \dots \cdot \left(\frac{1}{6}\right)^{x_6} = \left(\frac{1}{6}\right)^{x_1+\cdots+x_6} = \left(\frac{1}{6}\right)^n$$

Hence, the joint PMF of $(X_1, \dots, X_6)$ is

$$f_{X_1,\dots,X_6}(x_1, \dots, x_6) = \begin{cases} \frac{n!}{x_1!x_2!\dots x_6!}\left(\frac{1}{6}\right)^n, & \text{if } x_1 + x_2 + \cdots + x_6 = n, \\ 0 & \text{otherwise.} \end{cases}$$

Part 4: T counts the number of 1's and 2's after n rolls. The probability of rolling a 1 or 2 is $\frac{1}{3}$, so

$$T \sim \text{Bin}(n, \frac{1}{3}).$$

Remark 4. We could have used the fact that $(X_1, \ldots, X_6) \sim \text{Mult}(n, \frac{1}{6}, \ldots, \frac{1}{6})$ and used the properties of the multinomial to derive all of the above parts.

Problem 1.7. Consider drawing 5 cards from a standard 52 card deck of playing cards (4 suits, 13 kinds) **with replacement**. What is the probability that 2 of the drawn cards are hearts, 2 are spades, and 1 is a diamond?

Solution 1.7. Denote by H, S, D, C the number of Hearts, Spades, Diamonds, and Clubs. Then

$$(H, S, D, C) \sim \text{Mult}(5, 0.25, 0.25, 0.25, 0.25)$$

and

$$\mathbb{P}(H = 2, S = 2, D = 1, C = 0) = \frac{5!}{2!2!1!0!}\left(\frac{1}{4}\right)^4$$

Problem 1.8. In the game of Roulette, a small ball is spun around a wheel in such a way so that the probability it lands in a black or red box is 18/38 each, and the probability it lands in a green box is 2/38. Suppose 10 games are played, and let B, R and G denote the number of times the ball landed on black, red, and green, respectively.

- Write down the probability function of (B, R, G) along with all its constraints.
- Given that $B = 5$, calculate the probability that $R = 5$.

Solution 1.8.

Part 1: We know $(B, R, G) \sim \text{Mult}(10, 18/38, 18/38, 2/38)$ so

$$\mathbb{P}(B = b, R = r, G = g) = \frac{10!}{b!r!g!}\left(\frac{18}{38}\right)^{b+r}\left(\frac{2}{38}\right)^g,$$

when $b, r, g \in \{0, 1, \ldots, 10\}$ with $b + r + g = 10$ and 0 otherwise.

Part 2: By definition of conditional probability, and using that marginally $B \sim Bin(10, 18/38)$, we find

$$\begin{aligned}\mathbb{P}(R = 5 \mid B = 5) &= \frac{\mathbb{P}(R = 5, B = 5)}{\mathbb{P}(B = 5)} = \frac{\mathbb{P}(R = 5, B = 5, G = 0)}{\mathbb{P}(B = 5)} \\ &= \frac{\frac{10!}{5!5!}\left(\frac{18}{38}\right)^{10}}{\frac{10!}{5!5!}\left(\frac{18}{38}\right)^5\left(\frac{20}{38}\right)^5} = \left(\frac{18}{20}\right)^5 \approx 0.59049\end{aligned}$$

Problem 1.9. We can model n rounds of fair, independent rock-paper-scissors game using multinomial distribution:

$$(R, P, C) \sim \text{Mult}(n, 1/3, 1/3, 1/3).$$

Suppose that I play 5 games of R-P-S. Given that the sum of Rocks and Papers is 4, what would be the distribution of the number of Rocks I played?

Solution 1.9. Using the conditional probability formula for the multinomial with with $n = 5$, $p_j = 1/3$ for $j = 1, 2, 3$ and $t = 4$, we find

$$R \mid R + P = 4 \sim \text{Bin}\left(4, \frac{1/3}{1/3 + 1/3}\right) = \text{Bin}\left(4, \frac{1}{2}\right)$$

1.4.2 Proofs and Derivations

Problem 1.10. If X and Y are any random variables, show that

$$\mathbb{E}[ag_1(X, Y) + bg_2(X, Y)] = a \cdot \mathbb{E}[g_1(X, Y)] + b \cdot \mathbb{E}[g_2(X, Y)].$$

In particular, if $g_1 = x$ and $g_2 = y$ then

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Solution 1.10. We have by the definition,

$$\begin{aligned}\mathbb{E}[ag_1(X, Y) + bg_2(X, Y)] &= \sum_{(x,y)} [ag_1(x, y) + bg_2(x, y)]f_{X,Y}(x, y) \\ &= a \sum_{(x,y)} g_1(x, y)f_{X,Y}(x, y) + b \sum_{(x,y)} g_2(x, y)f_{X,Y}(x, y) \\ &= a \cdot \mathbb{E}[g_1(X, Y)] + b \cdot \mathbb{E}[g_2(X, Y)].\end{aligned}$$

By taking $g_1(x, y) = x$ and $g_2(x, y) = y$ we immediately arrive at the fact that

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

Remark 5. We have by the definition of the marginal PMF

$$\mathbb{E}[X] = \sum_{(x,y)} xf_{X,Y}(x, y) = \sum_x \sum_y xf_{X,Y}(x, y) = \sum_x x \sum_y f_{X,Y}(x, y) = \sum_x xf_X(x)$$

so $\mathbb{E}[X]$ coincides with the expected value for single random variables we saw before.

Problem 1.11. If X and Y are independent random variables, show that

$$\mathbb{E}[g_1(X)g_2(Y)] = \mathbb{E}[g_1(X)]\,\mathbb{E}[g_2(Y)].$$

In particular, if $g_1 = x$ and $g_2 = y$ then

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

Solution 1.11. Since $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ by independence, we have by the definition of the expected value,

$$\begin{aligned}\mathbb{E}[g_1(X)g_2(Y)] &= \sum_{(x,y)} (g_1(x)g_2(y))f_{X,Y}(x, y) \\ \text{independence} &= \sum_{(x,y)} g_1(x)g_2(y)f_X(x)f_Y(y) \\ &= \left(\sum_x g_1(x)f_X(x)\right)\left(\sum_y g_2(y)f_Y(y)\right) = \mathbb{E}[X]\,\mathbb{E}[Y].\end{aligned}$$

By taking $g_1(x) = x$ and $g_2(y) = y$ we immediately arrive at the fact that

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

Problem 1.12. If $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ are independent, show that

$$T = X + Y \sim \text{Poi}(\lambda_1 + \lambda_2).$$

Solution 1.12. We have $X + Y = n$ if and only if $X = m$ and $Y = n - m$ for $m = 0, 1, \dots, n$. Therefore,

$$\begin{aligned}
f_T(n) = \mathbb{P}(X + Y = n) &= \sum_{(x,y): x+y=n} \mathbb{P}(X = x, Y = y) \\
&= \sum_{m=0}^{n} \mathbb{P}(X = m, Y = n - m) \\
\text{independence} \quad &= \sum_{m=0}^{n} \mathbb{P}(X = m)\,\mathbb{P}(Y = n - m) \\
&= \sum_{m=0}^{n} e^{-\lambda_1} \frac{\lambda_1^m}{m!} e^{-\lambda_2} \frac{\lambda_2^{n-m}}{(n-m)!} \\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{m=0}^{n} \frac{n!}{m!(n-m)!} \lambda_1^m \lambda_2^{n-m} \\
\text{Binomial thm} \quad &= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n.
\end{aligned}$$

Problem 1.13. Let $X \sim \text{Poi}(\lambda_1)$ and $Y \sim \text{Poi}(\lambda_2)$ be independent. Show that

$$X \mid X + Y = n \sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$

Similarly, for Y, we have

$$Y \mid X + Y = n \sim \text{Bin}\left(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right)$$

Solution 1.13. Since $X + Y \sim \text{Poi}(\lambda_1 + \lambda_2)$, we have

$$\begin{aligned}
f_{X \mid X+Y} = \frac{f_{X,X+Y}(x,n)}{f_{X+Y}(n)} &= \frac{\mathbb{P}(X = x, X + Y = n)}{\mathbb{P}(X + Y = n)} \\
\text{independence} \quad &= \frac{\mathbb{P}(X = x)\,\mathbb{P}(Y = n - x)}{\mathbb{P}(X + Y = n)} \\
&= \frac{e^{-\lambda_1} \frac{\lambda_1^x}{x!} e^{-\lambda 2} \frac{\lambda_2^{n-x}}{(n-x)!}}{e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}} \\
&= \frac{n!}{x!(n-x)!} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-x}
\end{aligned}$$

which we recognize as the PMF of a $\text{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$ random variable. The proof for the Y given $X + Y = n$ is identical.

Problem 1.14. If $(X_1, \dots, X_n) \sim \text{Mult}(n, p_1, \dots, p_n)$, show that

1. $$X_j \sim \text{Bin}(n, p_j), \quad \text{for } j = 1, 2, \ldots, k \ .$$

2. $$X_i + X_j \sim \text{Bin}(n, p_i + p_j), \quad \text{for } i \neq j.$$

3. $$X_i \mid X_i + X_j = t \sim \text{Bin}\left(t, \frac{p_i}{p_i + p_j}\right), \quad \text{for } i \neq j.$$

4. $$\mathbb{E}[X_i X_j] = n(n-1)p_i p_j \quad \text{for } i \neq j.$$

Solution 1.14.

Part 1: By definition, X_j denotes the number of occurrences of outcome j in n trials and each occurence has probability p_j of happening so

$$X_j \sim \text{Bin}(n, p_j), \quad \text{for } j = 1, 2, \ldots, k \ .$$

Part 2: By definition, $X_i + X_j$ denotes the number of occurrences of outcome i or j in n trials and the probability of either i or j happening is $p_i + p_j$ so

$$X_i + X_j \sim \text{Bin}(n, p_i + p_j), \quad \text{for } i \neq j.$$

Part 3: Notice that if $X_i + X_j = t$, then X_i takes values in $\{0, 1, \ldots, t\}$. Therefore, for $x \in \{0, 1, \ldots, t\}$ we have

$$f_{X_i | X_i + X_j} = \frac{\mathbb{P}(X_i = x)}{\mathbb{P}(X_i + X_j = t)} = \frac{\mathbb{P}(X_i = x, X_j = t - x)}{\mathbb{P}(X_i + X_j = t)} = \frac{\mathbb{P}(X_i = x, X_j = t - x, \sum_{k \neq i,j} X_k = n - t)}{\mathbb{P}(X_i + X_j = t)}$$

since the total of all outcomes must be n. From the second part, we know that $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$

$$f_{X_i | X_i + X_j} = \frac{\frac{n!}{x!(t-x)!(n-t)!} p_i^x p_j^{t-x} (1 - p_i - p_j)^{n-t}}{\frac{n!}{t!(n-t)!} (p_i + p_j)^t (1 - p_i - p_j)^{n-t}} = \frac{t!}{x!(t-x)!} \left(\frac{p_i}{p_i + p_j}\right)^x \left(\frac{p_j}{p_i + p_j}\right)^{t-x}$$

which we recognize as the PMF of a $\text{Bin}\left(t, \frac{p_i}{p_i + p_j}\right)$ random variable.

Remark 6. This result is intuitive. Since we are given that $X_j + X_j = t$ we know that we have t total occurrences of X_i and X_j. We have that

$$\mathbb{P}(i \text{ happens} \mid i \text{ or } j \text{ happens}) = \frac{\mathbb{P}(i \text{ happens})}{\mathbb{P}(i \text{ or } j \text{ happens})} = \frac{p_i}{p_i + p_j}.$$

Therefore, the number of times i happens given that i or j happens at total of t times is $\text{Bin}\left(t, \frac{p_i}{p_i + p_j}\right)$

Part 4: We need to compute (noting that $x_i + x_j \leq n$ needs to hold):

$$\begin{aligned}
\mathbb{E}[X_i X_j] &= \sum_{\substack{x_i \geq 0, x_j \geq 0 \\ x_i + x_j \leq n}} x_i \cdot x_j \cdot \frac{n!}{x_i! x_j! (n - x_i - x_j)!} p_i^{x_i} p_j^{x_j} (1 - p_i - p_j)^{n - x_i - x_j} \\
&= \sum_{\substack{x_i \geq 1, x_j \geq 1 \\ x_i + x_j \leq n}} x_i \cdot x_j \cdot \frac{n!}{x_i! x_j! (n - x_i - x_j)!} p_i^{x_i} p_j^{x_j} (1 - p_i - p_j)^{n - x_i - x_j} \\
&= \sum_{\substack{x_i \geq 1, x_j \geq 1 \\ x_i + x_j \leq n}} \frac{n!}{(x_i - 1)!(x_j - 1)!(n - x_i - x_j)!} p_i^{x_i} p_j^{x_j} (1 - p_i - p_j)^{n - x_i - x_j}
\end{aligned}$$

Like in the computation of the expected value of a binomial, we factor out terms to make the summation look like the sum of a PMF,

$$
\begin{aligned}
&= n(n-1)p_ip_j \sum_{\substack{x_i-1\geq 0, x_j-1\geq 0\\ x_i-1+x_j-1\leq n-2}} \frac{(n-2)!\times p_i^{x_i-1}p_j^{x_j-1}(1-p_i-p_j)^{n-2-(x_i-1)-(x_j-1)}}{(x_i-1)!(x_j-1)!(n-2-(x_i-1)-(x_j-1))!}\\
&= n(n-1)p_ip_j \underbrace{\sum_{\substack{y_i\geq 0, y_j\geq 0\\ y_i+y_j\leq n-2}} \frac{(n-2)!}{(y_i)!(y_j)!(n-2-y_i-y_j)!}p_i^{y_i}p_j^{y_j}(1-p_i-p_j)^{n-2-y_i-y_j}}_{=1 \text{ Sum of PMF of Mult}(n-2,p_i,p_j,1-p_i-p_j)}\\
&= n(n-1)p_ip_j
\end{aligned}
$$

where we used the change of variables $y_i = x_i - 1$, $y_j = x_j - 1$.

Alternative Proof: We can compute the expected value using linearity of expectation. We can write $X_i = \sum_{k=1}^n \mathbb{1}_{A_k}$ where A_k is the event that outcome i occured on the kth trial, and

$$\mathbb{1}_{A_k} = \begin{cases} 1 & A_k \text{ happens} \\ 0 & A_k \text{ does not happens.} \end{cases}$$

Similarly, $X_j = \sum_{\ell=1}^n \mathbb{1}_{B_\ell}$ where B_ℓ is the event that outcome j occured on the ℓth trial, and

$$\mathbb{1}_{B_\ell} = \begin{cases} 1 & B_\ell \text{ happens} \\ 0 & B_\ell \text{ does not happens.} \end{cases}$$

Therefore,

$$\mathbb{E}[X_iX_j] = \mathbb{E}\left[\sum_{k=1}^n \mathbb{1}_{A_k} \sum_{\ell=1}^n \mathbb{1}_{B_\ell}\right] = \sum_{k,\ell=1}^n \mathbb{E}\left[\mathbb{1}_{A_k}\mathbb{1}_{B_\ell}\right].$$

We have two cases

1. $k = \ell$: Suppose that $k = \ell$. Since $\mathbb{1}(A_k)\mathbb{1}(B_k) = 1$ if an only if A_k and B_k happen, we have

   $$\mathbb{E}[\mathbb{1}_{A_k}\mathbb{1}_{B_\ell}] = \mathbb{E}[\mathbb{1}_{A_k}\mathbb{1}_{B_k}] = \mathbb{P}(A_k \cap B_k) = 0$$

   since both outcome i and j can't happen at the same time.

2. $k \neq \ell$: Suppose that $k \neq \ell$. Since $\mathbb{1}_{A_k}\mathbb{1}_{B_\ell} = 1$ if an only if A_k and B_ℓ happens

   $$\mathbb{E}[\mathbb{1}_{A_k}\mathbb{1}_{B_\ell}] = \mathbb{P}(A_k \cap B_\ell) = \mathbb{P}(A_k)\,\mathbb{P}(B_\ell) = p_ip_j$$

   since the trials are independent, so the outcomes A_k and B_ℓ are independent (they refer to different trials).

Since there are $n(n-1)$ ways to pick indices $k \neq \ell$, we have

$$\mathbb{E}[X_iX_j] = \sum_{k,\ell=1}^n \mathbb{E}\left[\mathbb{1}_{A_k}\mathbb{1}_{B_\ell}\right] = n(n-1)\,\mathbb{E}\left[\mathbb{1}_{A_1}\mathbb{1}_{B_2}\right] = n(n-1)p_ip_j.$$