1 Multivariate Distributions

We now develop a theory of probability to describe the simultaneous behavior of multiple (possibly dependent) random variables. This is the analogue of multi-variable functions from calculus.

1.1 Bivariate Distributions

We want to build a theory of probability for more than 1 variable. We first consider the bivariate (2 variable) case where X and Y are random variables defined on the same sample space taking values $(x, y) \in \mathbb{R}^2$. The case with n random variables is similar and will be described in Section 1.3. We will see that all definitions are straightforward generalization of the univariate (single variable) case.

Remark 1. We will define everything for discrete and continuous random variables to be precise, but the ideas of the joint, marginal, and conditional distributions are very similar between the two. We just replace the PMF with the PDF and replace sums with integrals just like in the univariate case.

1.1.1 Joint Distributions

The probabilities of objects involving both X and Y are encoded by the joint CDF.

Definition 1 (Joint Cumulative Distribution Function)**.** The *joint CDF* of random variables X and Y is the function $F_{X,Y} : \mathbb{R}^2 \to [0, 1]$ defined by

$$F_{X,Y}(x, y) = \mathbb{P}(X \leq x, Y \leq y) \qquad x, y \in \mathbb{R}.$$

Just like in the univariate case, the CDF allows us to compute the probability of any random variable taking values in any subset of $\mathbb{R}^2$. Just like the PMF and PDF, in the case when the random variables are discrete or continuous, there is an equivalent notion of probability functions.

Definition 2 (Joint Probability Mass Function)**.** The *joint PMF* of X and Y is

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

for $x \in X(\Omega), y \in Y(\Omega)$ and 0 otherwise.

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

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

Definition 3 (Joint Probability Density Function)**.** The *joint probability density function* of X and Y is

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

The joint PDF is not a probability (much like in the univariate case), but it satisfies the normalization property

1. $f_{X,Y} \geq 0$
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dxdy = 1$.

1.1.2 Marginal Distributions

The probabilities of only one random variable are encoded by the marginal distributions. These notions are the same as in the univariate case, and can be recovered by "integrating out" the other random variables we are not interested in.

Definition 4 (Marginal Cumulative Distribution Function)**.** Suppose that X and Y are random variables with joint CDF $F_{X,Y}(x, y)$. The *marginal CDF* of X is

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(X \leq x, Y \leq \infty) = \lim_{y\to\infty} F_{X,Y}(x, y).$$

Similarly, the marginal CDF of Y is

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq \infty, Y \leq y) = \lim_{x\to\infty} F_{X,Y}(x, y).$$

When X and Y are either discrete or continuous, we have the following notions of the probability functions, which behave like the PMF and PDF we covered earlier.

Definition 5 (Marginal Probability Mass Function)**.** Suppose that X and Y are *discrete* random variables with joint PMF $p_{X,Y}(x, y)$. The *marginal PMF* of X is

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(X = x, Y \in Y(\Omega)) = \sum_{y\in Y(\Omega)} p_{X,Y}(x, y).$$

Similarly, the marginal distribution of Y is

$$p_Y(y) = \mathbb{P}(Y = y) = \mathbb{P}(X \in X(\Omega), Y = y) = \sum_{x\in X(\Omega)} p_{X,Y}(x, y).$$

Definition 6 (Marginal Probability Density Function)**.** Suppose that X and Y are *continuous* random variables with joint probability function $f_{X,Y}(x, y)$. The *marginal PDF* of X is

$$f_X(x) = \int f_{X,Y}(x, y)\, dy.$$

Similarly, the marginal distribution of Y is

$$f_Y(y) = \int f_{X,Y}(x, y)\, dx.$$

Remark 2. We can go from the joint distributions to the marginal distributions, but we cannot go the other way around. By only looking at the marginal distributions, we do not know how the random variables behave together without further assumptions.

1.1.3 Conditional Distributions

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

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

This gives the following natural definition for random variables.

Definition 7 (Conditional Probability Mass Function)**.** The *conditional PMF* of X given $Y = y$ is

$$p_{X\mid Y}(x \mid y) = \mathbb{P}(X = x \mid Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p_{X,Y}(x, y)}{p_Y(y)} \text{ provided that } p_Y(y) > 0.$$

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

$$p_{Y\mid X}(y \mid x) = \mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{p_{X,Y}(x, y)}{p_X(x)}, \text{ provided that } p_X(x) > 0.$$

For each fixed y, the function $p_X(x \mid y)$ is the probability mass function of the random variable $X \mid Y = y$ and has the usual properties, such as summing to 1. We can define the conditional PDF in the analogous way even though PDFs are not necessarily probabilities.

Definition 8 (Conditional Probability Density Function)**.** The *conditional PDF* of X given $Y = y$ is

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

Similarly, the *conditional PDF* of Y given $X = x$ is

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

From the conditional probability functions, we get the analogues of the Bayes Rule and the law of total probability.

Theorem 1 (*Bayes' Rule*)

1. *For discrete random variables X and Y*

$$p_{Y \mid X}(y \mid x) = \frac{p_{X \mid Y}(x \mid y) p_Y(y)}{p_X(x)}. \qquad \mathbb{P}(Y = y \mid X = x) = \frac{\mathbb{P}(X = x \mid Y = y)\,\mathbb{P}(Y = y)}{\mathbb{P}(X = x)}$$

2. *For continuous random variables X and Y*

$$f_{Y \mid X}(y \mid x) = \frac{f_{X \mid Y}(x \mid y) f_Y(y)}{f_X(x)}$$

Theorem 2 (*Law of Total Probability*)

1. *For discrete random variables X and Y*

$$p_X(x) = \int_{-\infty}^{\infty} p_{X \mid Y}(x \mid y) p_Y(y)\, dy. \qquad \mathbb{P}(X = x) = \sum_{y \in Y(\Omega)} \mathbb{P}(X = x \mid Y = y)\,\mathbb{P}(Y = y).$$

2. *For continuous random variables X and Y*

$$f_X(x) = \int_{-\infty}^{\infty} f_{X \mid Y}(x \mid y) f_Y(y)\, dy.$$

1.1.4 Independence

Recall we say that events A and B are independent, if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$. Since the CDFs encode the behaviors of random variables we get the following definition.

Definition 9 (Independence)**.** X and Y are *independent* random variables if

$$F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y) = \mathbb{P}(X \le x)\,\mathbb{P}(Y \le y) = F_X(x) F_Y(y)$$

for all values of (x, y).

If X and Y have a joint PMF / PDF, then this is equivalent to

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \qquad \text{or} \qquad f_{X,Y}(x,y) = f_X(x)f_Y(y) \quad \forall x,y$$

or for any y such that $f_Y(y), p_Y(y) > 0$,

$$p_{X \mid Y}(x \mid y) = p_X(x) \qquad \text{or} \qquad f_{X \mid Y}(x \mid y) = f_X(x) \quad \forall x$$

There is a converse of this result as well, that says that if the probability functions factorizes, then it must correspond to independent random variables.

Proposition 1 (*Factorization Characterization*)

If the joint PDF of X and Y factorizes as

$$f_{X,Y}(x,y) = g(x)h(y) \qquad \text{for all } x,y \in \mathbb{R}$$

for some non-negative functions g and h, then X and Y are independent. Furthermore, if either g or h is a valid PDF, then the other is one too, and they correspond to the marginal PDFs of X and Y respectively. An analogous statement holds for the PMF.

Remark 3. If X and Y are independent, then it is possible to recover the joint distribution from the marginals. In this case, the joint distribution is simply the product of the marginals.

1.2 Joint Summary Statistics

We now introduce the summary statistics that generalizes the expected value and variances to the multivariate setting.

1.2.1 Expected Value

Definition 10 (Expected Value)**.** Suppose X and Y are discrete/continuous random variables with joint probability functions $p_{X,Y}/f_{X,Y}$. Then for any function $g : \mathbb{R}^2 \to \mathbb{R}$,

$$\mathbb{E}\left[g(X,Y)\right] = \sum_{(x,y)} g(x,y)p_{X,Y}(x,y) \qquad \text{or} \qquad \mathbb{E}\left[g(X,Y)\right] = \iint g(x,y)f_{X,Y}(x,y)\,dxdy$$

depending on if X, Y are jointly discrete or continuous.

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

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

1.2.2 Covariance

The covariance measures the joint variability of two random variables.

Definition 11 (Covariance)**.** For two random variables X and Y, the covariance between X and Y is

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X))(Y-\mathbb{E}(Y))\right]$$

provided the expression exists.

1.2.3 Properties

1. *Relationship with Variance:* $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

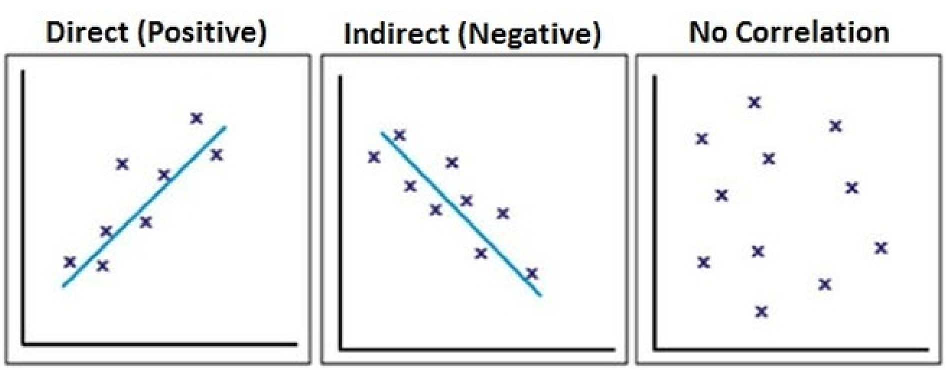
2. *Equivalent formula:*
$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y].$$

3. *Relationship with independence I:* If X and Y are independent,
$$\mathrm{Cov}(X,Y) = 0.$$
The converse of this statement is false!. There are pairs of random variables that have zero covariance, but are dependent (see Problem 1.15).

4. *Relationship with Independence II:* If X and Y have zero covariance, then
$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

5. *Cauchy–Schwarz Inequality:* For any random variables X and Y,
$$|\,\mathbb{E}[XY]| \le \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

6. *The Sign of the Covariance*: Suppose X, Y are positively related (when X large, Y likely large; when X small, Y likely small), then
$$\mathrm{Cov}(X,Y) > 0$$
Conversely, suppose X, Y are negatively related (when X large, Y likely small; when X small, Y likely large), then
$$\mathrm{Cov}(X,Y) < 0.$$



1.2.4 Correlation

The correlation measures how linearly related two random variables are.

Definition 12. The *correlation* of X and Y, denoted $\text{Corr}(X,Y)$, is defined by

$$\rho = \text{Corr}(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}} = \frac{\text{Cov}(X,Y)}{\text{SD}(X)\text{SD}(Y)}.$$

We say that X and Y are *uncorrelated* if $\text{Cov}(X,Y) = 0$ (or equivalently $\text{Corr}(X,Y) = 0$). We have implicitly assumed that X and Y have non-zero variance in this definition

The correlation satisfies the following properties

1. $\rho = \text{Corr}(X,Y)$ has the same sign as $\text{Cov}(X,Y)$
2. $-1 \leq \rho \leq 1$
3. $|\rho| = 1 \iff X = aY + b$. If $a > 0$, then $\rho = 1$, and if $a < 0$, then $\rho = -1$.
4. X, Y independent $\Rightarrow \text{Corr}(X,Y) = 0$
5. $\text{Corr}(X,Y) = 0 \not\Rightarrow X, Y$ independent in general
6. $\text{Corr}(X,X) = \text{Cov}(X,X)/\text{SD}(X)^2 = \text{Var}(X)/\text{Var}(X) = 1$
7. *Correlation does not imply causation:* Two variables being correlated does not always imply that one variable causes another to behave in certain ways.

1.3 Multivariate Random Variables

We considered the case of bivariate random variables above, but all the terminology above can be extended to a collection $X_1, X_2, \ldots, X_n$ of random variables in the obvious way. We will state the results for discrete random variables, but the obvious modifications also defines the continuous analogue of every definition.

Definition 13 (Multivariatge Joint PMF)**.** For a collection of n discrete random variables, $X_1, ..., X_n$, the joint probability function is defined as

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

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

The notion of independence generalized naturally.

Definition 14 (Multivariate Independence)**.** $X_1, X_2, \ldots, X_n$ are *independent* if

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

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

The definition of the expected value generalizes naturally as well.

Definition 15 (Mutlivariate Expected Value)**.** 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

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

The notion of a variance is a bit trickier since there are many possible covariances to deal with. The covariances between all entries is summarized by a single matrix called the covariance matrix.

Definition 16 (Covariance Matrix)**.** The covariance matrix of $\boldsymbol{X} = (X_1, X_2, \dots, X_n) \in \mathbb{R}^n$ is a $n \times n$ matrix with entries given by $\mathrm{Cov}(X_i, X_j)$, i.e.

$$\mathrm{Cov}(\boldsymbol{X}) = [\mathrm{Cov}(X_i, X_j)]_{i,j \leq n} \in \mathbb{R}^{n \times n}.$$

To compute the covariance matrix of $\boldsymbol{X} \in \mathbb{R}^n$, we have the following familiar formula

$$\mathrm{Cov}(\boldsymbol{X}) = \mathbb{E}[(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])(\boldsymbol{X} - \mathbb{E}[\boldsymbol{X}])^\intercal] = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\intercal] - \mathbb{E}[\boldsymbol{X}]\,\mathbb{E}[\boldsymbol{X}]^\intercal.$$

where the expectation is applied to each entry of the vector or matrix.

Remark 4. This is a generalization of the univariate formula

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

In fact, one can simply recognize that the vector formula is simply the definition of the covariance applied entrywise

$$\mathrm{Cov}(\boldsymbol{X}) = \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^\intercal] - \mathbb{E}[\boldsymbol{X}]\,\mathbb{E}[\boldsymbol{X}]^\intercal = [\mathbb{E}[X_i X_j] - \mathbb{E}[X_i]\,\mathbb{E}[X_j]]_{i,j \leq n}$$

1.3.1 Linear Combinations of Random Variables

We are often interested in the linear combinations of random variables.

Definition 17. A *linear combination* of the random variables $X_1, ..., X_n$ is any random variable $Y \in \mathbb{R}$ of the form

$$Y = \sum_{i=1}^n a_i X_i \qquad \text{where } a_1, ..., a_n \in \mathbb{R}.$$

Example 1. The *sample mean* $\bar{X}$ of $X_1, \dots, X_n$ is obtained by taking $a_i = \frac{1}{n}$ for all i

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Let $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$. We have the following properties about linear combinations of random variables.

1. **Linearity of Expectation:** For any random variables $X_1, \dots, X_n$,

$$\mathbb{E}\left[\sum_{i=1}^n a_i X_i\right] = \sum_{i=1}^n a_i\,\mathbb{E}[X_i].$$

2. **Bi-Linearity of Covariance:** For any random variables $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$,

$$\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^m b_i Y_i\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j).$$

In particular, for random variables X, Y, U, V be random variables, and $a, b, c, d \in \mathbb{R}$. Then,

$$\begin{aligned}&\mathrm{Cov}(aX + bY, cU + dV)\\ &= ac\mathrm{Cov}(X, U) + ad\mathrm{Cov}(X, V) + bc\mathrm{Cov}(Y, U) + bd\mathrm{Cov}(Y, V)\end{aligned}$$

3. **Variance of Linear Combinations:** The following result shows how the variance of a linear combination is "broken down" into pieces:

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j \operatorname{Cov}(X_i, X_j).$$

In particular, for random variables X, Y, and $a, b \in \mathbb{R}$,

$$\operatorname{Var}(aX + bY) = a^2 \operatorname{Var}(X) + b^2 \operatorname{Var}(Y) + 2ab\operatorname{Cov}(X, Y).$$

If the $X_1, \dots, X_n$ are independent, then they are uncorrelated, so in this case

$$\operatorname{Var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{Var}(X_i)$$

1.3.2 Common Distributions of Random Vectors

We now list the distributions of some linear combinations of random vectors. Many of these properties we have already encountered through examples.

1. **Sum of Independent Binomials is Binomial:** If $X \sim \operatorname{Bin}(n, p)$ and $Y \sim \operatorname{Bin}(m, p)$ independently, then
$$T = X + Y \sim \operatorname{Bin}(n + m, p).$$

2. **Sum of Independent Normal is Normal:** If $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$ and $Y \sim \operatorname{Bin}(\mu_2, \sigma_2^2)$ are independent, then
$$T = X + Y \sim \mathrm{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

3. **Sum of Independent Poisson is Poisson:** If $X \sim \operatorname{Poi}(\lambda_1)$ and $Y \sim \operatorname{Poi}(\lambda_2)$ are independent, then
$$T = X + Y \sim \operatorname{Poi}(\lambda_1 + \lambda_2).$$

4. **Conditional Poisson is Binomial:** Let $X \sim \operatorname{Poi}(\lambda_1)$ and $Y \sim \operatorname{Poi}(\lambda_2)$ be independent. Then, given $X + Y = n$, X follows binomial distribution. That is,
$$X \mid X + Y = n \sim \operatorname{Bin}\left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$
Similarly, for Y, we have
$$Y \mid X + Y = n \sim \operatorname{Bin}\left(n, \frac{\lambda_2}{\lambda_1 + \lambda_2}\right).$$

5. **Sum of Independent Bernouilli is Binomial:** Let $X_1, X_2, \dots, X_n$ be independent $\operatorname{Bern}(p)$ random variables. Then,
$$T = X_1 + X_2 + \ldots + X_n \sim \operatorname{Bin}(n, p).$$

6. **Sum of Independent Geometric is Negative Binommial:** Let $X_1, X_2, \dots, X_k$ be independent $\operatorname{Geo}(p)$ random variables. Then,
$$T = X_1 + X_2 + \ldots + X_k \sim \operatorname{NegBin}(k, p).$$

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

1.4 Example Problems

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

		x		
$p_{X,Y}(x, y)$	1	2	3	$p_Y(y)$
y 1	1/6	1/6	1/6	3/6
2	1/6	1/6	1/6	3/6
$p_X(x)$	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

			x		
$p_{X,Y}(x, y)$	0	1	2	3	$p_Y(y)$
y 0	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2
$p_X(x)$	1/8	3/8	3/8	1/8	1

Part 2: We can see

$$p_{X,Y}(0, 1) = 0 \neq \frac{1}{8} \cdot \frac{1}{2} = p_X(0)p_Y(1)$$

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 $p_{X \mid Y}(x \mid y) = p_{X,Y}(x, y)/p_Y(y)$ we find

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

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

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

1. $0 \leq p_{X,Y}(x,y) \leq 1$
2. $p_{X,Y}(x,y) \leq p_X(x)$
3. $p_{X,Y}(x,y) \leq p_Y(y)$

Solution 1.3.

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

Problem 1.4. Suppose X and Y have joint PMF

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

Find the marginal PMFs p_X and p_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

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

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

Part 2: The Y marginal is

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

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

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

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

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) = e^{-2}\frac{2^x}{x!}e^{-3}\frac{3^y}{y!} = e^{-5}\frac{2^x}{x!}\frac{3^y}{y!}.$$

Problem 1.6. Let X_1 and X_2 be independent exponential random variables with rate $\theta_1 = \frac{1}{\lambda_1}$ and $\theta_2 = \frac{1}{\lambda_2}$ respectively. Let Y denote the index of the smaller of X_1 and X_2, that is

$$Y = \begin{cases} 1 & X_1 < X_2 \\ 2 & X_2 < X_1. \end{cases}$$

In the case that $X_1 = X_2$ we set $Y = 1$. Furthermore, let $X_Y = \min(X_1, X_2)$ to denote the value of the smaller of X_1 and X_2.

1. Find the distribution of X_Y.
2. Find the distribution of Y.
3. Are X_Y and Y independent?

Solution 1.6.

Part 1: We have for $t \geq 0$,

$$F_{X_Y}(t) = \mathbb{P}(\min(X_1, X_2) \leq t) = 1 - \mathbb{P}(\min(X_1, X_2) > t).$$

Notice that by independence,

$$\mathbb{P}(\min(X_1, X_2) > t) = \mathbb{P}(X_1 > t)\,\mathbb{P}(X_2 > t) = e^{-\lambda_1 t}e^{-\lambda_1 t} = e^{-(\lambda_1+\lambda_2)t}.$$

To get the CDF, we differentiate with respect to t,

$$f_{X_Y}(t) = \frac{d}{dt}F_{X_Y}(t) = (\lambda_1+\lambda_2)e^{-(\lambda_1+\lambda_2)t}$$

so X_Y is exponential with parameter $\theta = (\lambda_1+\lambda_2)^{-1}$.

Part 2: There are only two cases, so

$$\begin{aligned}\mathbb{P}(Y=2) = \mathbb{P}(X_1 > X_2) &= \int_0^\infty \int_{x_2}^\infty \lambda_1\lambda_2 e^{-\lambda_1 x_1}e^{-\lambda_2 x_2}\,dx_1dx_2 \\ &= \int_0^\infty \lambda_2 e^{-\lambda_1 x_2}e^{-\lambda_2 x_2}\,dx_1dx_2 \\ &= \frac{\lambda_2}{\lambda_1+\lambda_2}.\end{aligned}$$

Therefore,

$$\mathbb{P}(Y=1) = 1 - \mathbb{P}(Y=2) = \frac{\lambda_1}{\lambda_1+\lambda_2}.$$

Part 3: We have

$$\begin{aligned}\mathbb{P}(X_Y > t, Y = 1) = \mathbb{E}[\mathbb{1}(X_1 > t, Y = 1)] &= \int_t^\infty \int_{x_1}^\infty \lambda_1\lambda_2 e^{-\lambda_1 x_1} e^{-\lambda_2 x_2}\, dx_2 dx_1 \\ &= \int_t^\infty \lambda_1 e^{-(\lambda_1+\lambda_2)x_1}\, dx_1 \\ &= \frac{\lambda_1}{\lambda_1+\lambda_2} e^{-(\lambda_1+\lambda_2)t} \\ &= \mathbb{P}(X_Y > t)\,\mathbb{P}(Y = 1)\end{aligned}$$

since the indicator on the set $\{X_1 > t, Y = 1\}$ means that we integrate over the region $X_2 > X_1 > t$. An identical computation show that

$$\mathbb{P}(X_Y > t, Y = 2) = \mathbb{P}(X_Y > t)\,\mathbb{P}(Y = 2),$$

so X_Y and Y are independent. This is a somewhat surprising result since X_Y looks like it depends on the minimal index Y.

Remark 6. Notice that in this problem X_1, X_2 and X_Y are continuous random variables while Y is discrete. However, the notations and definitions for the joint distributions naturally extend to this case as well.

Problem 1.7. Let $N \sim \mathrm{Poi}(\lambda)$ be a Poisson random variable with the mean λ. Then, consider N i.i.d. random variables, independent of N, taking values 1 or 2 with probabilities p and $q = 1 - p$ respectively. Let N_j be the number of these random variables taking value j, so that $N_1 + N_2 = N$. Show that N_1 and N_2 are independent Poisson random variables with means λp and λq respectively.

Solution 1.7. We want to compute the joint PMF of N_1 and N_2,

$$p_{N_1,N_2}(n_1, n_2) = \mathbb{P}(N_1 = n_1, N_2 = n_2) = \mathbb{P}(N_1 = n_1, N - N_1 = n_2) = \mathbb{P}(N_1 = n_1, N = n_1 + n_2).$$

By the definition of conditional probability

$$\mathbb{P}(N_1 = n_1, N = n_1 + n_2) = \mathbb{P}(N_1 = n_1 \mid N = n_1 + n_2)\,\mathbb{P}(N = n_1 + N_2).$$

By construction, when we have $N = n_1 + n_2$, then the conditional distribution of $N_1 \mid N = n_1 + n_2$ is binomial with $n_1 + n_2$ trials and probability p of success (see Problem 1.21). Therefore,

$$\begin{aligned}\mathbb{P}(N_1 = n_1 \mid N = n_1 + n_2)\,\mathbb{P}(N = n_1 + N_2) &= \binom{n_1+n_2}{n_1} p^{n_1}(1-p)^{n_2} \cdot e^{-\lambda}\frac{\lambda^{n_1+n_2}}{(n_1+n_2)!} \\ &= \frac{(n_1+n_2)!}{n_1!n_2!} p^{n_1}(1-p)^{n_2} \cdot e^{-\lambda(p+1-p)}\frac{\lambda^{n_1+n_2}}{(n_1+n_2)!} \\ &= e^{\lambda p}\frac{(\lambda p)^{n_1}}{n_1!} p^{n_1} \cdot e^{\lambda(1-p)}\frac{(\lambda(1-p))^{n_2}}{n_2!}.\end{aligned}$$

The right hand side can be recognized as the product of the PMFs of a $\mathrm{Poi}(\lambda p)$ and $\mathrm{Poi}(\lambda(1-p))$ random variable. It immediately follows that

$$p_{N_1,N_2}(n_1, n_2) = e^{\lambda p}\frac{(\lambda p)^{n_1}}{n_1!} p^{n_1} \cdot e^{\lambda(1-p)}\frac{(\lambda(1-p))^{n_2}}{n_2!}$$

so N_1 and N_2 are independent since it is the product of PMFs, and the marginal distribution of N_1 and N_2 are $N_1 \sim \mathrm{Poi}(\lambda p)$ and $N_2 \sim \mathrm{Poi}(\lambda(1-p))$.

Remark 7. This is called the Poisson splitting theorem and it allows you to split a Poisson process into two independent Poisson process by randomly marking each point independently.

Problem 1.8. 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 1.8.

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

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

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

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

Remark 8. We will in the next lesson that we could have used the fact that $(X_1,\dots,X_6)\sim \mathrm{Mult}(n,\frac{1}{6},\dots,\frac{1}{6})$ and used the properties of the multinomial to derive all of the above parts.

Problem 1.9. Let (X,Y) have the following joint probability mass function

$X\backslash Y$	0	1	2
1	0.1	0.2	0.1
2	0.1	0.3	0.2

1. Compute $\mathbb{E}[XY]$.
2. Compute $\mathbb{E}[X^2Y]$.

Solution 1.9.

Part 1: From the definition of the expected value,

$$\mathbb{E}[XY]=\sum_{x=1}^{2}\sum_{y=0}^{2}xyp_{X,Y}(x,y)=0+0.2+0.2+0+0.6+0.8=1.8$$

Part 2: From the definition of the expected value,

$$\mathbb{E}[X^2Y]=\sum_{x=1}^{2}\sum_{y=0}^{2}x^2yp_{X,Y}(x,y)=0+0.2+0.2+0+1.2+1.6=3.2.$$

Problem 1.10. Let (X,Y) have the joint probability density function

$$f_{X,Y}(x,y)=\begin{cases}6xy, & 0\le x\le 1,\ 0\le y\le 1-x,\\ 0, & \text{otherwise.}\end{cases}$$

1. Compute $\mathbb{E}[XY]$.
2. Compute $\mathbb{E}[X^2Y]$.

Solution 1.10.

Part 1: From the definition of the expected value,

$$\begin{aligned}\mathbb{E}[XY]=\int_0^1\int_0^{1-x}xy\,f_{X,Y}(x,y)\,dy\,dx&=\int_0^1\int_0^{1-x}6x^2y^2\,dy\,dx\\&=\int_0^1 6x^2\cdot\frac{(1-x)^3}{3}\,dx\\&=2\int_0^1(x^2-3x^3+3x^4-x^5)\,dx\\&=2\left(\frac{1}{3}-\frac{3}{4}+\frac{3}{5}-\frac{1}{6}\right)=\frac{1}{30}\end{aligned}$$

Part 2: From the definition of the expected value,

$$\begin{aligned}\mathbb{E}[X^2Y]=\int_0^1\int_0^{1-x}x^2y\,f_{X,Y}(x,y)\,dy\,dx&=\int_0^1\int_0^{1-x}6x^3y^2\,dy\,dx\\&=2\int_0^1(x^3-3x^4+3x^5-x^6)\,dx\\&=2\left(\frac{1}{4}-\frac{3}{5}+\frac{1}{2}-\frac{1}{7}\right)=\frac{1}{70}.\end{aligned}$$

Problem 1.11. Let X, Y be independent random variables with $\text{Var}(X) = \text{Var}(Y) = 1$. What is $\text{Var}(X - Y)$?

Solution 1.11. By independence, $\text{Cov}(X, Y) = 0$, so the variance for linear combinations formula implies

$$\text{Var}(X - Y) = \text{Var}(X) + (-1)^2 \text{Var}(Y) - 2\text{Cov}(X, Y) = 1 + 1 = 2.$$

1.5 Proofs of Key Results

Problem 1.12. 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 1.12. 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 9. We have by the definition of the marginal PMF

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

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

Problem 1.13. 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 1.13. 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

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

A similar proof holds for continuous random variables. The integration is used instead of summation and we apply Fubini's theorem to split the sum into two parts.

Problem 1.14. Prove Proposition 1.

Solution 1.14. Independence follows immediately if $g(x)$ and $h(y)$ integrated to 1, since in that case we can $g(x) = f_X(x)$ and $h(y) = f_Y(y)$. However, if we only know that

$$f_{X,Y}(x,y) = g(x)h(y)$$

then we don't know that g and h integrate to 1. However, we can normalize the functions so that they always integrate to 1 without loss of generality. We define $c = \int_{-\infty}^{\infty} h(y)\,dy$ and consider the functions $\tilde{g}(x) = cg(x)$ and $\tilde{h}(y) = \frac{h(y)}{c}$

$$f_{X,Y}(x,y) = g(x)h(y) = cg(x)\frac{h(y)}{c} = \tilde{g}(x)\tilde{h}(y).$$

Notice that by definition of c,

$$\int_{-\infty}^{\infty} \tilde{h}(y)\,dy = \frac{1}{c}\int_{-\infty}^{\infty} h(y)\,dy = 1.$$

Likewise, by Fubini's theorem and the fact that $f_{X,Y}$ is a joint PDF,

$$\int_{-\infty}^{\infty} \tilde{g}(x)\,dx = \underbrace{\left(\int_{-\infty}^{\infty} \tilde{h}(y)\,dy\right)}_{=1}\int_{-\infty}^{\infty} \tilde{g}(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underbrace{\tilde{g}(x)\tilde{h}(y)}_{=f_{X,Y}(x,y)}\,dxdy = 1.$$

This means that X and Y are independent with marginal PDFs $f_X = \tilde{g}$ and $f_Y = \tilde{h}$. The proof for discrete random variables is identical.

Problem 1.15. Suppose that X and Y are independent. Show that

$$\mathrm{Cov}(X,Y) = 0.$$

Show that the converse is false by providing a counterexample.

Solution 1.15. Suppose that X and Y are independent. We know that $\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$. Therefore, using the equivalent formula,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[X]\,\mathbb{E}[Y] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0.$$

Counterexample: Let $X \sim U(-1,1)$, and let $Y = X^2$. X and Y are not independent because

$$0 = \mathbb{P}\left(X > \frac{1}{2}, Y < \frac{1}{4}\right) \neq \mathbb{P}\left(X > \frac{1}{2}\right)\mathbb{P}\left(Y < \frac{1}{4}\right) > 0$$

since $X > \frac{1}{2} \implies X^2 > \frac{1}{4}$ so it is impossible that $Y = X^2 < \frac{1}{4}$ as well. However, we can compute the covariance,

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\,\mathbb{E}[X^2] = 0$$

since the PDF of X is symmetric, so $\mathbb{E}[X^3] = 0$ and $\mathbb{E}[X] = 0$.

Problem 1.16. Prove the Cauchy–Schwarz inequality,

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

Equality holds if and only if $Y = aX$ for some constant a.

Solution 1.16. Notice that the statement is trivial if either $X = 0$ or $Y = 0$, so we consider the non-trivial cases.

For any $t \in \mathbb{R}$, we have

$$0 \leq \mathbb{E}[(tX - Y)^2] = at^2 - 2bt + c$$

where $a = \mathbb{E}[X^2]$, $b = \mathbb{E}[XY]$ and $c = \mathbb{E}[Y^2]$. A quadratic polynomial $at^2 - 2bt + c$ is non-negative if and only if it has at most one root, which happens if the discriminant satisfies

$$D = 4b^2 - 4ac \leq 0 \implies b^2 \leq ac \implies |b| \leq \sqrt{ac}$$

so $|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$. This proves the first part of the statement.

We now consider the equality case. Suppose now that we have equality $|\mathbb{E}[XY]| = \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$, so $|b| = \sqrt{ac}$. This implies that $D = 0$, so the quadratic polynomial has exactly one real root. Let $\lambda = \frac{b}{a}$ denote the value of this root, so

$$\mathbb{E}[(\lambda X - Y)^2] = a\lambda^2 - 2b\lambda + c = 0.$$

We have that $\mathbb{E}[(\lambda X - Y)^2] = 0$ if and only if $\lambda X - Y = 0$ with probability one, so $Y = \lambda X = \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}X$ with probability 1. Therefore, if $X \neq aY$ for any a, then $Y \neq \frac{\mathbb{E}[XY]}{\mathbb{E}[X^2]}X$ so $|\mathbb{E}[XY]| \neq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$.

To prove the converse, suppose that $X = aY$. We have

$$|\mathbb{E}[XY]| = |a||\mathbb{E}[Y^2]| = \sqrt{\mathbb{E}((aY)^2)}\sqrt{\mathbb{E}(Y^2)} = |a|\,\mathbb{E}[Y^2]$$

so equality holds.

Problem 1.17. Show that $\rho = \mathrm{Corr}(X, Y)$ satisfies $|\rho| \leq 1$ and $|\rho| = 1$ if and only $Y = aX + b$ for some constants a and b.

Solution 1.17. By the Cauchy–Schwarz inequality, applied to $X - \mathbb{E}[X]$ and $Y - \mathbb{E}[Y]$ we have

$$|\mathrm{Cov}(X, Y)| = |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \leq \sqrt{\mathbb{E}[(X - \mathbb{E}[X])^2]}\sqrt{\mathbb{E}[(Y - \mathbb{E}[Y])^2]} = \sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}.$$

Rearranging terms implies that

$$|\mathrm{Corr}(X, Y)| = \frac{|\mathrm{Cov}(X, Y)|}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}} \leq 1.$$

Next, we have that equality happens if and only if $Y - \mathbb{E}[Y] = a(X - \mathbb{E}[X])$ for some constant a. This means that there must be a linear relation between Y and X if equality were to hold. To see that any linear relation achieves equality, suppose that $Y = aX + b$ for some constants a and b, so by bilinearity

$$|\mathrm{Cov}(X, Y)| = |\mathrm{Cov}(X, aX + b)| = |a\mathrm{Cov}(X, X) + b\mathrm{Cov}(X, 1))| = |a|\,\mathrm{Var}(X)$$

and

$$\sqrt{\mathrm{Var}(Y)} = \sqrt{\mathrm{Var}(aX + b)} = |a|\sqrt{\mathrm{Var}(X)},$$

so

$$|\mathrm{Corr}(X, Y)| = 1.$$

Remark 10. We can repeat the second computation without the absolute values to conclude that $\mathrm{Corr}(X,Y) = 1$ implies that $Y = aX + b$ for some constant $a > 0$ and $\mathrm{Corr}(X,Y) = -1$ implies that $Y = aX + b$ for some constant $a < 0$

Problem 1.18. Prove the binlinearity property of covariances

$$\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^n b_i Y_i\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j).$$

Solution 1.18. This is a direct consequence of linearity of expectation and the distributive property of numbers

$$\sum_{i=1}^n a_i \times \sum_{j=1}^m b_j = \sum_{i=1}^n \sum_{j=1}^m a_i b_j.$$

By the definition of the covariance,

$$\begin{aligned}
\mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^n b_i Y_i\right] &= \mathbb{E}\left[\left(\sum_{i=1}^n a_i X_i - \mathbb{E}\left[\sum_{i=1}^n a_i X_i\right]\right)\left(\sum_{i=1}^m b_i Y_i - \mathbb{E}\left[\sum_{i=1}^m b_i Y_i\right]\right)\right] \\
\text{linearity of expectation } &= \mathbb{E}\left[\left(\sum_{i=1}^n a_i (X_i - \mathbb{E}[X_i])\right)\left(\sum_{i=1}^m b_i (Y_i - \mathbb{E}[Y_i])\right)\right] \\
\text{distributive property } &= \mathbb{E}\left[\sum_{i=1}^n \sum_{i=1}^m a_i b_i (X_i - \mathbb{E}[X_i])(Y_i - \mathbb{E}[Y_i])\right] \\
\text{linearity of expectation } &= \sum_{i=1}^n \sum_{i=1}^m a_i b_i\, \mathbb{E}[(X_i - \mathbb{E}[X_i])(Y_i - \mathbb{E}[Y_i])] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j).
\end{aligned}$$

Problem 1.19. Prove the formula for the variance of linear combinations of random variables,

$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \,\mathrm{Var}(X_i) + 2 \sum_{1 \le i < j \le n} a_i a_j \mathrm{Cov}(X_i, X_j).$$

Solution 1.19. Since $\mathrm{Var}(X) = \mathrm{Cov}(X,X)$, the proof follows directly from the bilinearity of covariance. We have

$$\begin{aligned}
\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) &= \mathrm{Cov}\left[\sum_{i=1}^n a_i X_i, \sum_{i=1}^n a_i X_i\right] \\
\text{bilinearity } &= \sum_{i,j=1}^n a_i a_j \mathrm{Cov}(X_i, X_j) \\
\text{split into diagonal and offdiagonal } &= \sum_i^n a_i^2 \mathrm{Cov}(X_i, X_i) + \sum_{i \ne j} a_i a_j \mathrm{Cov}(X_i, X_j) \\
\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X), \mathrm{Var}(X) = \mathrm{Cov}(X,X) &= \sum_i^n a_i^2 \,\mathrm{Var}(X_i) + 2 \sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j).
\end{aligned}$$

Problem 1.20. If $X \sim \mathrm{Poi}(\lambda_1)$ and $Y \sim \mathrm{Poi}(\lambda_2)$ are independent, show that

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

Solution 1.20. We have $X+Y=n$ if and only if $X=m$ and $Y=n-m$ for $m=0,1,\dots,n$. Therefore,

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

Problem 1.21. Let $X\sim \mathrm{Poi}(\lambda_1)$ and $Y\sim\mathrm{Poi}(\lambda_2)$ be independent. Show that

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

Similarly, for Y, we have

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

Solution 1.21. Since $X+Y\sim\mathrm{Poi}(\lambda_1+\lambda_2)$, we have

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

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

Problem 1.22. Let $X_i\sim \mathrm{N}(\mu_i,\sigma_i^2)$, $i=1,2,\dots,n$ be independent then,

$$\sum_{i=1}^n (a_iX_i+b_i)\sim \mathrm{N}\left(\sum_{i=1}^n a_i\mu_i+b_i, \sum_{i=1}^n a_i^2\sigma_i^2\right).$$

Solution 1.22. The easiest proof of this fact uses moment generating functions, which will be explained a later week. In the meantime, we will present a geometric proof of this result using a multidimensional change of variables. For simplicity, we consider the case that $n = 2$ and the means of the random variables are 0 and $a_i = 1$ and $b_i = 1$. The general case can be recovered by an induction argument and a standardization argument.

Let $X_1 \sim \mathrm{N}(0, \sigma_1)$ and $X_2 \sim \mathrm{N}(0, \sigma_2)$ be independent. We need to show that

$$X_1 + X_2 \sim \mathrm{N}(0, \sigma_1^2 + \sigma_2^2).$$

By standardization, $X_1 = \sigma_1 Z_1$ and $X_2 = \sigma_2 Z_2$. We want to find the PDF of

$$F_{X_1+X_2}(t) = \mathbb{P}(\sigma_1 Z_1 + \sigma_2 Z_2 \leq t) = \frac{1}{2\pi} \iint_{\sigma_1 z_1 + \sigma_2 z_2 \leq t} e^{-\frac{z_1^2+z_2^2}{2}} dz_1 dz_2.$$

Using the change of variables corresponding to the rotation of the half plane to one perpendicular to the z_2 axis,

$$w_1 = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}} z_1 + \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} z_2 \qquad w_2 = \frac{\sigma_1}{\sqrt{\sigma_1^2+\sigma_2^2}} z_1 - \frac{\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} z_2$$

we have

$$\{\sigma_1 z_1 + \sigma_2 z_2 \leq t\} = \left\{ w_1 \leq \frac{t}{\sqrt{\sigma_1^2+\sigma_2^2}} \right\}$$

so

$$\frac{1}{2\pi} \iint_{\sigma_1 z_1 + \sigma_2 z_2 \leq t} e^{-\frac{z_1^2+z_2^2}{2}} dz_1 dz_2 = \frac{1}{2\pi} \iint_{w_1 \leq \frac{t}{\sqrt{\sigma_1^2+\sigma_2^2}}} e^{-\frac{w_1^2+w_2^2}{2}} dw_1 dw_2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t}{\sqrt{\sigma_1+\sigma_2}}} e^{-\frac{w_1^2}{2}} dw_1.$$

We conclude that

$$F_{X_1+X_2}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{t}{\sqrt{\sigma_1+\sigma_2}}} e^{-\frac{w_1^2}{2}} dw_1 = \mathbb{P}\left(Z \leq \frac{t}{\sqrt{\sigma_1+\sigma_2}} \right) = \mathbb{P}(\sqrt{\sigma_1+\sigma_2} Z \leq t)$$

which is the CDF of a $\mathrm{N}(0, \sigma_1^2 + \sigma_2^2)$ random variable.

Remark 11. Essentially we have exploited the rotational invariance of the standard normal. Notice that the PDF is rotational symmetric since

$$f_{Z_1,Z_2}(z_1, z_2) = \frac{1}{2\pi} e^{-\frac{z_1^2+z_2^2}{2}} = \frac{1}{2\pi} e^{-\frac{r^2}{2}} = f_{Z_1,Z_2}(r)$$

is a function of $r = \sqrt{z_1^2 + z_2^2}$, which means that the density only depends on the distance from points to the origin. The special change of variables rotated the half plane $\{\sigma_1 z_1 + \sigma_2 z_2 \leq t\}$ to be perpendicular to the z_2 axis to reduce the problem to computing the probability $\sqrt{\sigma_1^2 + \sigma_2^2} Z \leq t$.