1 Random Variables

We define the concept of a random variable which will allow us to describe probabilities without having to go through the trouble specifying the sample spaces, which is often tedious to work with in practice.

Definition 1 (Random Variable)**.** A *random variable* is a (measurable) function that maps the sample space Ω to the set of real numbers $\mathbb{R}$. That is, X is a random variable if

$$X : \Omega \to \mathbb{R}.$$

Definition 2 (Range)**.** The values in $\mathbb{R}$ that a random variable takes is called the *range* of the random variable, and is denoted by

$$X(\Omega) = \{X(\omega) \in \mathbb{R} : \omega \in \Omega\}.$$

Definition 3 (Pre-Image)**.** The values in the sample space that are mapped to a set A by the random variable X is called the *pre-image* of A under X, and is denoted by

$$X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} = \{X \in A\}.$$

Associated with a random variable is a natural probability measure on $\mathbb{R}$, which encodes the liklihood of the random variables taking any particular set of numbers.

Definition 4 (Distribution)**.** The *distribution* $\mathbb{P}_X$ is a probability measure on $\mathbb{R}$ given by the push-forward of $\mathbb{P}$ by X. That is, for any (measurable) set $A \subset \mathbb{R}$,

$$\mathbb{P}_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)).$$

The set of all events we can assign probabilities is denoted by $\mathcal{B}$ and is called the *Borel σ-algebra*.

Remark 1. The set $\mathcal{B}$ contains almost every set that you can imagine. In more advanced probability courses, we require that in the definition of a random variable that X is a measurable function. This ensures that its distribution of X is well–defined on $\mathcal{B}$.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space then $(X(\Omega), \mathcal{B}, \mathbb{P}_X)$ defines a probability space with sample space given by the range of X, events given by (measurable) subsets of $\mathbb{R}$, and probability measure given by the distribution of X. This is very convenient notion because we no longer have to consider probabilities on sets, but rather probabilities on the real line.

1.1 Cumulative Distribution Function (CDF)

We are often interested in probabilities of the form $\mathbb{P}(X \le x)$. We will see that these probabilities encodes the same information as a PMF.

Definition 5 (CDF)**.** The *cumulative distribution function* (CDF) of a random variable X is

$$F_X(x) = \mathbb{P}(X \le x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) \le x\}), \quad x \in \mathbb{R}.$$

The CDF is not a probability measure, but it encodes all the information of a probability measure.

Theorem 1 (*Characterization of a CDF*)

The CDF F satisfies

(i) Right-continuous: i.e., $F_X(x) = F_X(x^+) = \lim_{t \downarrow x} F_X(t)$ for all $x \in \mathbb{R}$

(ii) Non-decreasing: $F_X(x) \le F_X(y)$ for $x < y$

(iii) Boundary Conditions: satisfies $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$.

Conversely, any function F with these properties (i), (ii) and (iii) is the cdf of some random variable.

As a consequence, if two random variables have the same CDF, they encode the same probability measure on $X(\Omega)$.

Definition 6 (Equal in Distribution)**.** Two random variables X and Y are *equal in distribution* if $F_X(t) = F_Y(t)$ for all $t \in \mathbb{R}$. We denote this by

$$X \stackrel{d}{=} Y.$$

Remark 2. Random variables X and Y being equal in distribution does not mean $X = Y$ (see Problem 4.1). It just means that the probability of X and Y taking any particular value is the same. In fact, X and Y don't even have to be functions defined on the same sample space.

1.2 Independence

The notion of independence for sets translates in the expected way for random variables.

Definition 7 (Independence)**.** Random variables $X_1, X_2, \ldots, X_n$ are *independent* if for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

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

Remark 3. Independence is quite a strong condition, since it must hold for all $x_1, x_2, \ldots, x_n \in \mathbb{R}$. This is the analogue of mutual independence of all the events $\{X_i \leq x_i\}$ for $x_i \in \mathbb{R}$.

Independence is naturally inherited by functions of independent random variables, since applying a non-random function to a random variable does not give more information about the other variables.

Theorem 2 (*Independence of Functions of Random Variables*)

Let $f, g : \mathbb{R} \to \mathbb{R}$ be non-random functions on $\mathbb{R}$. If X and Y are independent random variables, then the random variables $f(X)$ and $g(Y)$ are also independent.

In statistics, we will often sample several independent realizations from the same distribution.

Definition 8 (i.i.d.)**.** A sequence of random variables $X_1, \ldots, X_n$ are called independent and identically distributed (i.i.d.) if they are independent and have the same distribution.

Remark 4. Although we often use independent and identically distributed together throughout the course, the notions of independence and identically distributed are separate concepts.

We will see later that i.i.d. is the default assumption for the basic statement of many results about the limits of several random variables.

1.3 Discrete Random Variables

We now introduce the discrete random variables, which will be used to illustrate several fundamental concepts in probability before generalizing to other types of random variables.

Definition 9 (Discrete Random Variable)**.** We say that a random variable is *discrete* if its range $X(\Omega)$ is a discrete subset of $\mathbb{R}$ (i.e., a finite or a countably infinite set).

Remark 5. A random variable may be discrete even though the underlying sample space might not be (see Problem 4.2).

1.3.1 Probability Mass Function (PMF)

A discrete random variable can only take countably many points, so its distribution is completely encoded by the probability of each individual value.

Definition 10. The *probability (mass) function* of a discrete random variable X is the function

$$p_X(x) = \mathbb{P}(X = x) := \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}) = \mathbb{P}(X^{-1}(x)).$$

The value of $p_X(x)$ is zero when x is outside the range of the random variable X, so we usually only specify p_X on $X(\Omega)$. The values of x such that $p_X(x)$ is nonzero is called the *support* of X.

If $(\Omega, \mathscr{F}, \mathbb{P})$ is our original probability model on the underlying sample space, the PMF induces a probability on the range $X(\Omega)$ through the (push-forward) measure $p_X(x) = \mathbb{P}(X^{-1}(x))$. It follows that the PMF defines a *(discrete) probability distribution* defined on $X(\Omega) \subseteq \mathbb{R}$ instead of Ω:

1. $$0 \leq p_X(x) \leq 1 \quad \text{for all } x$$

2. $$\sum_{x \in X(\Omega)} p_X(x) = 1.$$

An important implication of this fact is that we no longer have to specify what the underlying sample space Ω with probability measure $\mathbb{P}$ to compute probabilities. If X encodes the quantities we need to assign probabilities to, then we can work directly with the sample space $X(\Omega)$ and its distribution p_X. The upside from this point of view is that studying probability has now been connected to studying functions, and we have many mathematical tools to do this, e.g. linear algebra, calculus, etc.

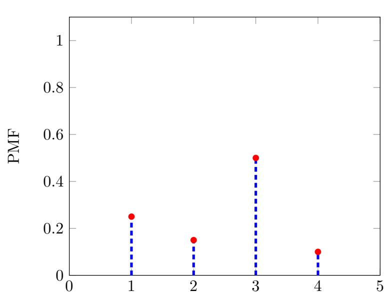
1.3.2 Connection Between the PMF and CDF

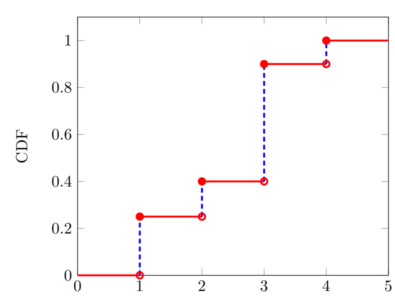
The PMF and CDF encode the same information for discrete random variables.

1. If X is discrete with PMF p_X, then
$$F_X(x) = \sum_{y \leq x} p_X(y).$$
Notice that $F_X(x)$ is constant between consecutive points in the support of p_X.

2. If X is discrete with CDF F_X, then
$$p_X(x) = F_X(x) - F_X(x^-) =: F_X(x) - \lim_{t \uparrow x} F_X(t).$$
Notice that $p_X(x)$ is zero except for points of discontinuity of F_X.

Example 1. The PMF and CDF of a random variable is visualized below:





The discontinuous jumps of the CDF are exactly the same size as the non-zero values of the PMF.

1.4 Example Problems

Problem 1.1. Consider again the following game: You roll a fair die and win \$2 if the die shows a number between 1 and 4 (inclusive), and otherwise you loose \$5. If X denotes the gain, what is p_X.

Solution 1.1. The underlying sample space is [6], and the underlying probability is uniform on this sample space. Clearly X takes values in $\{-5, 2\}$. We have

$$p_X(2) = \mathbb{P}(X = 2) = \mathbb{P}(X^{-1}(2)) = \mathbb{P}(\omega \in \{1, 2, 3, 4\}) = \frac{2}{3}$$

$$p_X(-5) = \mathbb{P}(X = -5) = \mathbb{P}(X^{-1}(-5)) = \mathbb{P}(\omega \in \{5, 6\}) = \frac{1}{3}$$

and $p_X(x) = 0$ otherwise.

Problem 1.2. Suppose you roll two fair six-sided dice and denote by X the sum. Which x maximizes the PMF $p_X(x)$?

Solution 1.2. We tabulate the PMF p_X of X, which represents the sum of the two dice:

x	2	3	4	5	6	7	8	9	10	11	12
$p_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

We see that $x = 7$ maximizes the PMF p_X.

Problem 1.3. Consider rolling two fair six sided die, and let the random variable X be the minimum of the die rolls. What is $p_X(2)$?

Solution 1.3. We want to compute $p_X(2) = \mathbb{P}(X = 2)$. This happens when we roll

$$(2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 2), (4, 2), (5, 2), (6, 2)$$

There are 9 possibilities, so $p_X(2) = \frac{9}{36} = \frac{1}{4}$.

Problem 1.4. Find the value k which makes the function f given by

$$f(0) = 0.1, \quad f(1) = k, \quad f(2) = 3k, \quad f(3) = 0.3$$

and 0 elsewhere a valid probability function.

Solution 1.4. A PMF function has to be non-negative and sum to 1. We find k such that

$$f(0) + f(1) + f(2) + f(3) = 1 \implies 4k + 0.4 = 1 \implies k = 0.15.$$

Furthermore, one can check that this choice of k ensures that all $f(i) \in [0, 1]$.

Problem 1.5. Suppose students A,B and C each independently answer a question on a test. The probability of getting the correct answer is 0.9 for A, 0.7 for B and 0.4 for C. Let X denote the number of people who get the answer correct.

1. Compute the PMF of X.

2. Draw the CDF of X.

Solution 1.5. Denote by A, B, C the events that students A, B, C get the answer correct, then $\mathbb{P}(A) = 0.9$, $\mathbb{P}(B) = 0.7$ and $\mathbb{P}(C) = 0.4$ and we also know $\mathbb{P}(A^c) = 0.1$, $\mathbb{P}(B^c) = 0.3$ and $\mathbb{P}(C^c) = 0.6$.

We find can explicitly compute all cases

$$p_X(0) = \mathbb{P}(X = 0) = \mathbb{P}(A^c \cap B^c \cap C^c) \overset{\text{indep.}}{=} \mathbb{P}(A^c)\,\mathbb{P}(B^c)\,\mathbb{P}(C^c) = \frac{18}{1000}$$

$$p_X(3) = \mathbb{P}(X = 3) = \mathbb{P}(A \cap B \cap C) \overset{\text{indep.}}{=} \mathbb{P}(A)\,\mathbb{P}(B)\,\mathbb{P}(C) = \frac{252}{1000}$$

$$\begin{aligned} p_X(1) = \mathbb{P}(X = 1) &= \mathbb{P}(A \cap B^c \cap C^c) + \mathbb{P}(A^c \cap B \cap C^c) + \mathbb{P}(A^c \cap B^c \cap C) \\ &= \frac{9 \cdot 3 \cdot 6}{1000} + \frac{7 \cdot 1 \cdot 6}{1000} + \frac{4 \cdot 1 \cdot 3}{1000} = \frac{216}{1000} \end{aligned}$$

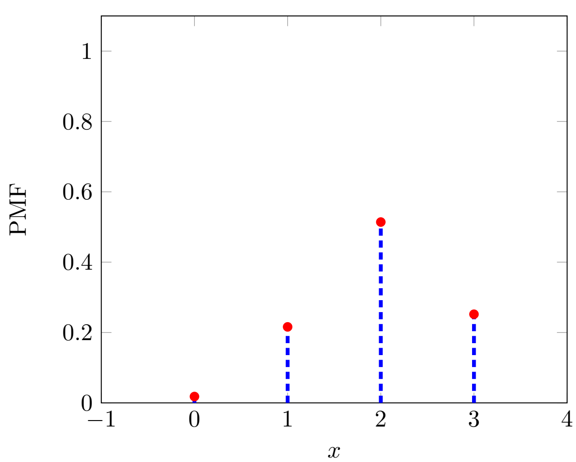
and, since the PMF sums to 1,

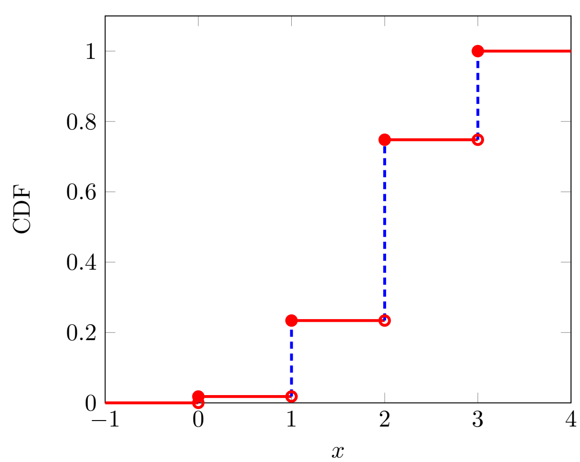
$$p_X(2) = 1 - p_X(0) - p_X(1) - p_X(3) = \frac{514}{1000}.$$

The CDF is thus

$$F_X(x) = \mathbb{P}(X \le x) = \begin{cases} 0, & \text{if } x < 0 \\ p_X(0), & \text{if } 0 \le x < 1 \\ p_X(0) + p_X(1), & \text{if } 1 \le x < 2 \\ p_X(0) + p_X(1) + p_X(2), & \text{if } 2 \le x < 3 \\ p_X(0) + p_X(1) + p_X(2) + p_X(3), & \text{if } x \ge 3 \end{cases} = \begin{cases} 0, & \text{if } x < 0 \\ \frac{18}{1000}, & \text{if } 0 \le x < 1 \\ \frac{234}{1000}, & \text{if } 1 \le x < 2 \\ \frac{748}{1000}, & \text{if } 2 \le x < 3 \\ 1, & \text{if } 3 \le x \end{cases}$$

The plot of the CDF is below:





Remark 6. The end points of the intervals in the CDF are the same as the non-zero values of the PMF. Furthermore, the $\le$ inequality is always on the left of the x and the $<$ inequality is always to the right of the x. This implies the CDF is right continuous. Furthermore, the value of the CDF on each interval is equal to the value of the CDF at the left endpoint (which is true even for the first interval since $F_X(-\infty) = 0$).

Problem 1.6. Consider flipping a fair coin. Let $X = 1$ if the coin is heads, and $X = 3$ if the coin is tails. Let $Y = X^2 + X$. What is the probability mass function of X?

Solution 1.6. The underlying sample space is $\Omega = \{H, T\}$. Y takes values in $\{3, 12\}$, so f_Y is supported on $\{3, 12\}$. Notice that $Y^{-1}(12) = X^{-1}(3) = \{T\}$ and $Y^{-1}(3) = X^{-1}(1) = \{H\}$

$$f_Y(y) = \mathbb{P}(Y = y) = \begin{cases} \frac{1}{2} & \text{if } y = 3 \\ \frac{1}{2} & \text{if } y = 12 \end{cases}$$

Problem 1.7. Suppose that a bowl contains 10 balls, each uniquely numbered 0 through 9. Two balls are drawn with replacement and let X_1 be the number of the first ball and X_2 be the number of the second ball. Find the PMF of $X = X_1 + 10X_2$.

Solution 1.7. We have X_1 and X_2 are independent and $X_1 \sim X_2$. X_1 is uniformly distributed over the set $\Omega = \{0, 1, \dots 9\}$ and so is X_2. We have

$$\begin{aligned}
\mathbb{P}(X = 0) &= \mathbb{P}(X_1 = 0)\,\mathbb{P}(X_2 = 0) = 0.1^2 = 0.01 \\
\mathbb{P}(X = 1) &= \mathbb{P}(X_1 = 1)\,\mathbb{P}(X_2 = 0) = 0.1^2 = 0.01 \\
&\vdots \\
\mathbb{P}(X = 98) &= \mathbb{P}(X_1 = 8)\,\mathbb{P}(X_2 = 9) = 0.1^2 = 0.01 \\
\mathbb{P}(X = 99) &= \mathbb{P}(X_1 = 9)\,\mathbb{P}(X_2 = 9) = 0.1^2 = 0.01
\end{aligned}$$

We have that X is uniformly distributed on the set of $\{0, 1, \dots, 99\}$.

2 Important Discrete Distributions

We now introduce several named discrete distributions. We use the symbol $\sim$ to mean *distributed as*.

2.1 Summary of Named Discrete Distributions

2.1.1 (Discrete) Uniform Distribution: $\mathrm{DUnif}[a, b]$

The (discrete) uniform distribution models variables with equally likely outcomes on an interval.

Definition 11. Suppose the range of the random variable X is $\{a, a+1, \ldots, b\}$, where $a, b \in \mathbb{Z}$, and suppose all values are equally likely. Then we say that X has a *discrete uniform distribution* on $\{a, a+1, \ldots, b\}$, and is denoted by

$$X \sim \mathrm{DUnif}[a, b].$$

- **PMF:**
$$p_X(x) = \frac{1}{b-a+1}, \quad \text{for } x \in \{a, a+1, \ldots, b\}.$$

- **CDF:**
$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{\lfloor x \rfloor - a + 1}{b-a+1}, & \text{if } x \in \{a, a+1, \ldots, b\}, \\ 1, & \text{if } x \geq b, \end{cases}$$

  where $\lfloor x \rfloor = \max\{z \in \mathbb{Z} : z \leq x\}$ is the rounding-down function ("floor").

Example 2. The following experiments can be modeled by a uniform distribution:

Experiment	X	Distribution
Roll a 6 sided die	# showing on die	$U[1, 6]$
Draw a number between 1 and 50	# Drawn	$U[1, 50]$
Shuffle a deck of cards	position of $A\spadesuit$	$U[1, 52]$

2.1.2 Hypergeometric Distribution: $\mathrm{Hyp}(N, r, n)$

The hypergeometric distribution counts the number of successes in a sample **without** replacement.

Definition 12. Consider a population that consists of N objects that can be divided into a group of r indistinguishable "successes" and a group of $N - r$ indistinguishable "failures". If X is the number of successes in a random subset of size n drawn from the population **without** replacement, then we say X follows a *hypergeometric distribution* with parameters (N, r, n), and is denoted by

$$X \sim \mathrm{Hyp}(N, r, n).$$

- **PMF:**
$$p_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} \quad \text{for } \max\{0, n-(N-r)\} \leq x \leq \min\{r, n\}$$

- **CDF:** There is no closed form in terms of elementary functions.

Example 3. The following experiments can be modeled by a hypergeometric distribution

Experiment	X	Distribution
Drawing 5 cards from a deck of cards	# of Ace's	$\mathrm{Hyp}(52, 4, 5)$
Lotto where 7 numbers are drawn from 50	# Matches	$\mathrm{Hyp}(50, 7, 7)$

2.1.3 Bernoulli Distribution: Bern(p)

The Bernoulli distribution models experiments with two possible outcomes.

Definition 13. Suppose an experiment (called a *Bernoulli trial*) has a probability of success p. If X denotes the number of successes in a **single** Bernoilli trial, then we say X follows the *Bernoulli distribution*, and is denoted by

$$X \sim \mathrm{Bern}(p).$$

- **PMF:**
$$p_X(0) = 1 - p, \quad p_X(1) = p$$

- **CDF:**
$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - p, & \text{if } 0 \leq x < 1, \\ 1, & \text{if } x \geq 1 \end{cases}$$

Example 4. The following experiments can be modeled by a Bernoulli distribution

Experiment	X	Distribution
Roll a 6 sided die	# of 1's	$\mathrm{Bern}(\frac{1}{6})$
Lotto where 7 numbers are drawn from 50	# Jackpots	$\mathrm{Bern}(\binom{50}{7}^{-1})$

A Bernoulli random variable can be encoded by the occurrence of an event A.

Definition 14 (Indicator Random Variable)**.** Let A be an event. The indicator random variable is given by

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A. \end{cases}$$

Remark 7. Notice that $\mathbb{1}_A \sim \mathrm{Bern}(\mathbb{P}(A))$.

2.1.4 Binomial Distribution: Bin(n, p)

The binomial distribution counts how many trials are successful after **multiple** independent experiments. Equivalently, it also models the number of successes in samples **with replacement**.

Definition 15. Suppose a Bernoulli trial has a probability of success p. If X is the number of successes in n independent Bernoilli trials, then we say X follows the *Binomial distribution*, and is denoted by

$$X \sim \mathrm{Bin}(n, p).$$

- **PMF:**
$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

- **CDF:** There is no closed form in terms of elementary functions.

Example 5. The following experiments can be modeled by a Binomial distribution

Experiment	X	Distribution
Roll a 10 6 sided die	# of 1's	$\mathrm{Bin}(10, \frac{1}{6})$
Buy 1 tickets from a Lotto where 7 numbers are drawn from 50	# Jackpots	$\mathrm{Bin}(1, \binom{50}{77}^{-1})$
Generate each digit of a 5 digit number randomly from 1 to 9	# odd digits	$\mathrm{Bin}(5, \frac{5}{9})$

Relationship with the Bernoulli Distribution: Since a Bernoulli random variable takes values 0 or 1, the number of successes in n Bernoulli trials is simply the sum. Therefore, for i.i.d. Bernoulli random variables $X_1, X_2, \dots, X_n \sim \mathrm{Bern}(p)$,

$$X = X_1 + X_2 + \cdots + X_n \sim \mathrm{Bin}(n, p).$$

Relationship with the Hypergeometric Distribution: Intuitively, when the population is large then sampling with or without replacement should not make much of a difference provided that the sample size is small with respect to the population size. The Binomial distribution arises as a limit of Hypergeometric distribution when the number of successes r is a fixed proportion of the population size,

$$\frac{r}{N} = p \text{ and } N \to \infty$$

Theorem 3 (*Binomial Approximation of the Hypergeometric Distribution*)

Let $p \in (0, 1)$ and let $X \sim \mathrm{Hyp}(N, pN, n)$ and $Y \sim \mathrm{Bin}(n, p)$. Then for all $k \in \mathbb{R}$,

$$\lim_{N\to\infty} \mathbb{P}(X \leq k) = \mathbb{P}(Y \leq k).$$

2.1.5 Geometric Distribution: Geo(p)

The geometric distributions models the number of fails until the **first** success.

Definition 16. Suppose a Bernoulli trial has a probability of success p. The independent trials are repeated until a success has been observed. If X denotes the number of failures that we observed before the first success, then we say X follows the *geometric distribution*, and is denoted by

$$X \sim \mathrm{Geo}(p)$$

- **PMF:**

$$p_X(x) = p(1-p)^x \quad \text{for } x = 0, 1, 2, \dots.$$

- **CDF:**

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - (1-p)^{\lfloor x \rfloor + 1} & \text{if } \geq 0 \end{cases}$$

Example 6. The following experiments can be modeled by a uniform distribution:

Experiment	X	Distribution
Repeated flips of a coin	# tails until the first head	Geo(0.5)
Repeated flips of a coin	# flips until the first head	Geo(0.5) + 1

Memoryless Property: Intuitively, having a long string of failures should not mean that a success is more likely when conducting independent experiments.

Theorem 4 (*Memoryless Property*)

Let $X \sim \mathrm{Geo}(p)$ and s, t be non-negative integers. Then, the following holds

$$\mathbb{P}(X \geq s + t \mid X \geq s) = \mathbb{P}(X \geq t).$$

In fact, the geometric distribution is the only discrete distribution with this property.

2.1.6 Negative Binomial Distribution: $\text{NegBin}(k, p)$

The negative binomial distributions models the number of fails until a certain amount successes.

Definition 17. Suppose a Bernoulli trial has a probability of success p. The independent trials are repeated until k successes have been observed. If X denotes the number of failures that we observed before k successes, then we say X follows the *negative Binomial distribution*, and is denoted by

$$X \sim \text{NegBin}(k, p)$$

- **PMF:**
$$p_X(x) = \binom{x+k-1}{x} p^k (1-p)^x \quad \text{for } x = 0, 1, 2, \ldots.$$
- **CDF:** There is no closed form in terms of elementary functions.

Example 7. The following experiments can be modeled by a negative binomial distribution:

Experiment	X	Distribution
Repeated flips of a coin	# tails until 3 heads	$\text{NegBin}(3, 0.5)$
Repeated flips of a coin	# flips until 3 heads	$\text{NegBin}(3, 0.5) + 3$

Comparison with Binomial Distribution: In the *negative binomial distribution*, you know the number of successes, but you don't know the number of trials (since # fails = # trials - # successes). In the *binomial distribution*, you know the number of trials, but you don't know the number of successes. One can also interpret the coefficient in the PMF of the negative binomial as a negative binomial coefficient (see Problem 4.10).

2.1.7 Poisson Distribution: $\text{Poi}(\lambda)$

The Poisson distribution models the number of occurrences of an event in a given period of time (or space) when the events happen one after another and the occurrence of one event does not influence another.

Definition 18. Let λ encode the mean rate of events, then we say X follows the *Poisson distribution*, and is denoted by

$$X \sim \text{Poi}(\lambda).$$

- **PMF:**
$$p_X(x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots.$$
- **CDF:** There is no closed form in terms of elementary functions.

Example 8. The following experiments can be modeled by a negative binomial distribution:

Experiment	X	Distribution
Incoming calls at a call center (at rate 2 per hour)	# of calls per hour	$\text{Poi}(2)$
iPhone manufacturing with failure rate 5%	# of faulty iPhones	$\text{Poi}(0.05)$

Poisson Approximation of the Binomial: The Poisson distribution is an approximation of the Binomial distribution when $p_n \approx \frac{\lambda}{n}$.

Theorem 5 (*Poisson Approximation of the Binomial*)

Given $\lambda > 0$, if $p = p_n \to 0$ in such a way such that $np_n \to \lambda$. Let $X \sim \text{Bin}(n, p_n)$ and $Y \sim \text{Poi}(\lambda)$. Then for all $k \geq 0$,

$$\lim_{n \to \infty} \mathbb{P}(X = k) = \mathbb{P}(Y = k).$$

3 Example Problems

Problem 3.1. Consider drawing a 5 card hand at random from a standard 52 card deck. What is the probability that the hand contains at least 3 Kings?

Solution 3.1. This is modeled using a hypergeometric distribution. We have $N = 52$ cards, out of which $r = 4$ are kings ("successes"), and we are sampling $n = 5$ cards without replacement from the deck. The random number of kings, X, then satisfies $X \sim \mathrm{Hyp}(N = 52, r = 4, n = 5)$. Thus, using the PMF from earlier, we find

$$P(X \geq 3) = P(X = 3) + P(X = 4) = \frac{\binom{4}{3}\binom{48}{2}}{\binom{52}{5}} + \frac{\binom{4}{4}\binom{48}{1}}{\binom{52}{5}} \approx 0.00175$$

Problem 3.2. Suppose a tack when flipped has probability 0.6 of landing point up. If the tack is flipped 10 times, what is the probability it lands point up more than twice?

Solution 3.2. This is modeled using a binomial distribution. Let X denote the number of times the tack lands point up. Then $X \sim \mathrm{Bin}(10, 0.6)$ and

$$\begin{aligned}
\mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \leq 2) \\
&= 1 - [\mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2)] \\
&= 1 - \left[\binom{10}{0}0.6^0 0.4^{10} + \binom{10}{1}0.6^1 0.4^9 + \binom{10}{2}0.6^2 0.4^8\right] \\
&\approx 0.9877
\end{aligned}$$

Problem 3.3. There are 5 stops on a bus line and 10 passengers on the bus. At every stop, there is a machine that records how many passengers got off at that stop. Assume the passengers are each equally likely to get off at any stop. Let X denote the number of passengers recorded by the machine at the first stop. Find the PMF of X.

Solution 3.3. By our assumptions, each passenger chooses a bus stop independently, and there is a $\frac{1}{5}$ chance of the passenger getting off at the first stop. We can model this with a binomial distribution, which counts a success if the passenger gets off at the first stop. Therefore, $X \sim B(10, 0.2)$, so

$$p_X(x) = \binom{10}{x}0.2^x 0.8^{10-x},$$

for $x \in \{0, 1, \ldots, 10\}$.

Alternative Solution: By the assumptions, the sample space $\Omega = [5]^{10}$ (which denotes where each passenger got off) has equally likely outcomes. To find $p_X(x) = \mathbb{P}(X = x)$, we want to count all the possible events A such that A has exactly x 1's appearing. There are $\binom{10}{x}4^{10-x}$ possibilities since there are $\binom{10}{x}$ ways to choose which passengers got off at stop 1 (the number of cases), and 4^{10-x} possible choices for the remaining passengers (the number of possibilities in each case). Since the probability is uniform on Ω,

$$p_X(x) = \frac{\binom{10}{x}4^{10-x}}{5^{10}} = \binom{10}{x}\frac{1}{5^x}\frac{4^{10-x}}{5^{10-x}} = \binom{10}{x}0.2^x 0.8^{10-x}.$$

Remark 8. Clearly, p_X is not uniform, so the number of passengers that got off at the first stop is not uniform over the range $X(\Omega) = \{0, 1, 2, \dots, 10\}$. This means that a sample space that encodes the number of people that got off at a particular stop does not have equally likely outcomes.

Problem 3.4. You have n identical looking keys on a chain, and one opens your office door. Suppose you try the keys in random order. Let X denote the number of keys you try until the door opens. Find the PMF of X.

Solution 3.4. Since we are trying keys randomly without replacement, the location of the correct key is uniform over the set $\{1, \dots, n\}$ by symmetry. We can model this with a uniform distribution, so $X \sim DUnif[1, n]$. Therefore,

$$p_X(x) = \frac{1}{n}$$

for $x \in \{, 1, \dots, n\}$.

Alternative Solution: By our assumptions, the sample space is the permutations of the set $[n]$, which denotes the order of keys we try. Without loss of generality, we may assume that the key labeled 1 is the right key. To find $p_X(x) = \mathbb{P}(X = x)$, we want to count all the possible events A such that 1 appears in the xth position. There are $(n-1)!$ ways that this can happen, since there are $(n-1)$ positions left to assign without replacement after fixing the correct key in the xth position. Since the probability is uniform on Ω,

$$p_X(x) = \frac{(n-1)!}{n!} = \frac{1}{n}.$$

Problem 3.5. Suppose you have a bag with 100 beads, 15 of which are red and the remaining ones are blue. You take 5 beads out of the bag without replacement. Suppose we want to compute the probability that 2 of the 5 sampled beads are red. The best model is the hypergeometric. Call the resulting probability p^{hyper}. We approximate this probability by using an appropriate Binomial distribution. Denote the probability (under the binomial model) that we have 2 red beads by $p^{binomial}$. Compute p^{hyper} and $p^{binomial}$.

Solution 3.5. Under the true model $X \sim \text{Hyp}(N = 100, r = 15, n = 5)$, so

$$p^{hyper} = \mathbb{P}(X = 2) = \frac{\binom{15}{2}\binom{100-15}{5-2}}{\binom{100}{5}} \approx 0.13775$$

Let $p = r/N = 0.15$. If we assumed a binomial distribution $Y \sim \text{Bin}(5, 0.15)$, we'd get

$$p^{binomial} = \mathbb{P}(Y = 2) = \binom{5}{2} 0.15^2 0.85^3 = 0.13818$$

which is quite close to the true probability.

Problem 3.6. A website is counting visitors to their website. Suppose that visitors visit the website at random at a rate of 10 visitors per minute on average, and they visit the site independently and individually from each other. Let X denote the number of visitors after 10 minutes. What is the distribution of X?

Solution 3.6. Let t be measured in minutes. The rate is 10 visitors per minute, so X is a Poisson random variable with rate $\lambda = 10$. Therefore, number of visitors in 10 minutes is

$$X \sim \mathrm{Poi}(10 \cdot 10) = \mathrm{Poi}(100).$$

Problem 3.7. In the manufacturing process of commercial carpet, small faults occur at random in the carpet according to a Poisson process at an average rate of 0.95 per 20 m^2. One of the rooms of a new office block has an area of 90 m^2 and has been carpeted using the same commercial carpet described above. What is the probability that the carpet in that room contains at least 4 faults?

Solution 3.7. Let X denotes the faults in the room. We are given that the rate of faults is 0.95 per 20 m^2 or equivalently a rate of $\frac{0.95}{20}$ faults per m^2, so X is Poisson with rate $\lambda = 90 \cdot \frac{0.95}{20}$,

$$X \sim \mathrm{Poi}\Big(90 \cdot \frac{0.95}{20}\Big) = \mathrm{Poi}(4.275).$$

The complement of the event of at least 4 faults is equal to 3 or less faults, so

$$\begin{aligned}\mathbb{P}(X \geq 4) &= 1 - \mathbb{P}(X \leq 3) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) - \mathbb{P}(X = 3) \\ &= 1 - e^{-4.275}\Big(1 + 4.275 + \frac{4.275^2}{2} + \frac{4.275^3}{6}\Big) = 0.618.\end{aligned}$$

Problem 3.8. Website hits for a given website occur according to a Poisson process with a rate of 100 hits per minute. We say a second is a "break" if there are no hits in that second.

1. What is the probability p of a break in any given second?
2. Compute the probability of observing exactly 10 breaks in 60 consecutive non-overlapping seconds.
3. Compute the probability that one must wait for 30 seconds to get 2 breaks.

Solution 3.8. Let t be measured in seconds. The rate is 100 hits per minute, or equivalently $\frac{100}{60}$ hits per second.

1. If X is the number of hits in one sec, then X is a Poisson random variable rate $\lambda = \frac{100}{60}$,

   $$X \sim \mathrm{Poi}(100/60) = \mathrm{Poi}(5/3).$$

   A break means zero hits in one sec, so

   $$p = \mathbb{P}(X = 0) = e^{-\frac{5}{3}} \frac{\left(\frac{5}{3}\right)^0}{0!} \approx 0.189.$$

2. Take 60 one-sec intervals. Each interval has a probability of p of having a break. Let Y be the number of one-sec intervals (from 60 one-sec intervals) with a break. Then $Y \sim Bin(60, p)$, and

   $$\mathbb{P}(Y = 10) = \binom{60}{10} p^{10}(1-p)^{50} \approx 0.124$$

3. Let Z be the number of one-sec intervals one needs to wait until observing two breaks. Then, $Z \sim \mathrm{NegBin}(2, p)$ and

   $$\mathbb{P}(Z = 30) = \binom{30 + 2 - 1}{30} p^2(1-p)^{30} \approx 0.002.$$

Problem 3.9. At a super busy coffee chain, customers arrive according to a Poisson Process at a rate of $\lambda = 5$ customers per minute.

1. Find the probability that there are more than 2 customers in one minute.
2. Suppose you record the number of customers in 5 consecutive one-minute intervals. What is the probability that in at least 3 of them there were more than 2 customers?
3. Find the probability that a minute with more than 2 customers actually had 6 customers
4. Suppose you are waiting until finally, there is one minute with more than 2 customers. Denote by X the the number of minutes you need to wait. Find the PMF of X.
5. Suppose in 3 minutes, there were n customers. Find the probability that x of these came in the first two minutes.

Solution 3.9. Let t be measured in minutes, and the rate is $\lambda = 5$ customers per minute.

1. If X is the number of customers in one minute, then $X \sim Poi(5 \cdot 1)$. Thus,

$$\begin{aligned} p = \mathbb{P}(X > 2) &= 1 - \mathbb{P}(X \le 2) \\ &= 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) - \mathbb{P}(X = 2) \\ &= 1 - e^{-5}\Big(1 + 5 + \frac{5^2}{2}\Big) \approx 0.875 \end{aligned}$$

2. Let Y be the number of one-minute intervals with more than two customers, then $Y \sim \mathrm{Bin}(5, p)$ with p from earlier. Thus,

$$\begin{aligned} \mathbb{P}(Y \ge 3) &= \mathbb{P}(Y = 3) + \mathbb{P}(Y = 4) + \mathbb{P}(Y = 5) \\ &= \binom{5}{3} p^3 (1-p)^2 + \binom{5}{4} p^4 (1-p) + \binom{5}{5} p^5 (1-p)^0 \approx 0.984 \end{aligned}$$

3. Let X be the number of customers in one minute. Thus,

$$\mathbb{P}(X = 6 \mid X > 2) = \frac{\mathbb{P}(X = 6 \text{ and } X > 2)}{\mathbb{P}(X > 2)} = \frac{\mathbb{P}(X = 6)}{\mathbb{P}(X > 2)} = \frac{e^{-5}\frac{5^6}{6!}}{0.875} \approx 0.167$$

4. Let Z be the number of minutes until first minute with more than 2 customers, then $Z \sim Geo(p)$ with p from earlier. Thus,

$$f_Z(x) = \mathbb{P}(Z = x) = (1-p)^x p, \quad x = 0, 1, 2, \dots$$

5. We want to fine

$$\mathbb{P}(x \text{ in first 2min} \mid n \text{ in 3min}) = \frac{\mathbb{P}(x \text{ in first 2min and n in 3min})}{\mathbb{P}(n \text{ in 3min})}.$$

   We know that

   - Denominator: The number of customers in 3 minutes follows a $\mathrm{Poi}(5 \cdot 3)$ distribution, so

$$\mathbb{P}(n \text{ in 3 min}) = e^{-15}\frac{15^n}{n!}, \quad n = 0, 1, 2, \dots$$

- Numerator: Since non-overlapping intervals are independent,

$$\begin{aligned}\mathbb{P}(x \text{ in first 2min and } n \text{ in 3min}) &= \mathbb{P}(x \text{ in first 2min and } n-x \text{ in last min})\\ &= \mathbb{P}(x \text{ in first 2min}) \cdot \mathbb{P}(n-x \text{ in last min})\\ &= e^{-10}\frac{10^x}{x!} \cdot e^{-5}\frac{5^{n-x}}{(n-x)!}\end{aligned}$$

Combining and simplifying gives

$$\mathbb{P}(x \text{ in first 2min} \mid n \text{ in 3min}) = \binom{n}{x}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{n-x}, \quad x = 0, 1, \ldots, n,$$

which is the PMF of $\text{Bin}(n, 2/3)$.

Problem 3.10. Shiny versions of Pokemon are possible to encounter and catch starting in Generation 2 (Pokemon Gold/Silver). Normal encounters with Pokemon while running in grass occur according to a Poisson process with rate 1 per minute on average. 1 in every 8192 encounters will be a Shiny Pokemon, on average.

1. Ash runs around in grass for 15 hours, what is the probability he will encounter at least one Shiny pokemon?
2. How long would Ash have to run around in grass so that he has better than 50 percent chance of encountering at least one Shiny pokemon?

Solution 3.10. Let t be measured in hours. The rate of normal encounters is 60 per hour, and the rate of shinies are is $\frac{1}{8192 \text{ encounters}}\frac{60 \text{ enounters}}{\text{hour}} = \frac{60}{8192}$ shinies per hour. Let X be number of pokemon encountered after 1 hour, Y be the number of shiny pokemon encountered after one hour. Then

$$X \sim X_1 \sim \text{Poi}(60) \qquad \text{and} \qquad Y \sim Y_1 \sim \text{Poi}\Big(\frac{60}{8192}\Big)$$

1. Let Z be the number of shiny encountered after 15 hours, then $Z \sim Y_{15} = \text{Poi}(\frac{60}{8192} \cdot 15) = \text{Poi}(0.1099)$, and
$$\mathbb{P}(Z \geq 1) = 1 - \mathbb{P}(Z = 0) = 1 - e^{-0.1099} \approx 0.104.$$
2. If Z is the number of shiny encountered after t hours, then $Z \sim \text{Poi}(\frac{60}{8192} \cdot t)$. Then
$$\begin{aligned}\mathbb{P}(Z \geq 1) = 1 - \mathbb{P}(Z = 0) \geq 0.5 &\Leftrightarrow \mathbb{P}(Z = 0) \leq 0.5\\ &\Leftrightarrow e^{-\frac{60}{8192}\cdot t} \leq 0.5\\ &\Leftrightarrow -\frac{60}{8192} \cdot t \leq -\log(2)\\ &\Leftrightarrow t \geq \log(2) \cdot \frac{8192}{60} \approx 94.6\end{aligned}$$
That mean Ash will have to run for at least 95 hours!

Problem 3.11. A bit error occurs for a given data transmission method independently in one out of every 1000 bits transferred. Suppose a 64 bit message is sent using the transmission system. Let p_{true} be the probability that there are exactly 2 bit errors and p_{approx} be the approximated probability that there are exactly 2 bit errors obtained through the Poisson approximation. Find p_{true} and p_{approx}.

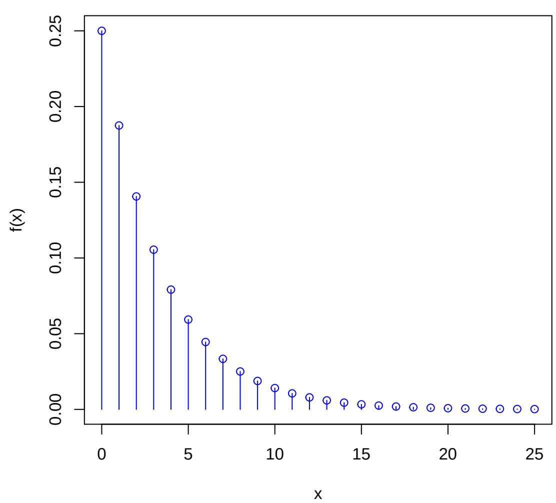
Solution 3.11.

1. Let X be the number of errors, then $X \sim \text{Bin}(64, 1/1000)$. We find
$$\mathbb{P}(X = 2) = \binom{64}{2} \left(\frac{1}{1000}\right)^2 \left(\frac{999}{1000}\right)^{64-2} \approx 0.00189.$$

2. n is large and p is small, so X follows approximately a Poisson with $\lambda = np = 64/1000$, so
$$\mathbb{P}(X = 2) = e^{-\frac{64}{1000}} \frac{\left(\frac{64}{1000}\right)^2}{2!} \approx 0.00192.$$

As expected by the Poisson approximation of the Binomial, both values are close.

Problem 3.12. Which PMF is the figure most likely showing?

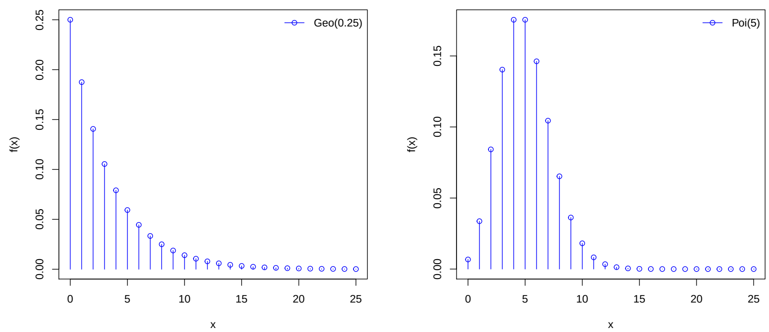


1. Poi(1)
2. Geo(0.25)
3. NegBin(5, 0.75)
4. Bin(25, 0.25)

Solution 3.12. The PDF is strictly decreasing so it cannot be negative binomial (for $k \geq 2$) or binomial since the binomial coefficient means the PMFs have a bump. The Poi(μ) distribution is strictly decreasing for $\mu < 1$ and has a bump for $\mu > 1$. When $\mu = 0$ then it turns out that $\mathbb{P}(X = 0) = \mathbb{P}(X = 1)$. This means that the Geo(0.25) is the only possibility.

Alternative Solution: We can also look at $p_X(0)$ which is 0.25 to determine if it is Poisson or geometric random variable. If $X \sim \text{Geo}(0.25)$ then $p_X(0) = 0.25$. While if $X \sim \text{Poi}(1)$ then $p_X(0) = e^{-1} \approx 0.36$, so the PMF is the one of the Geometric random variable.

Problem 3.13. Consider the following PMF



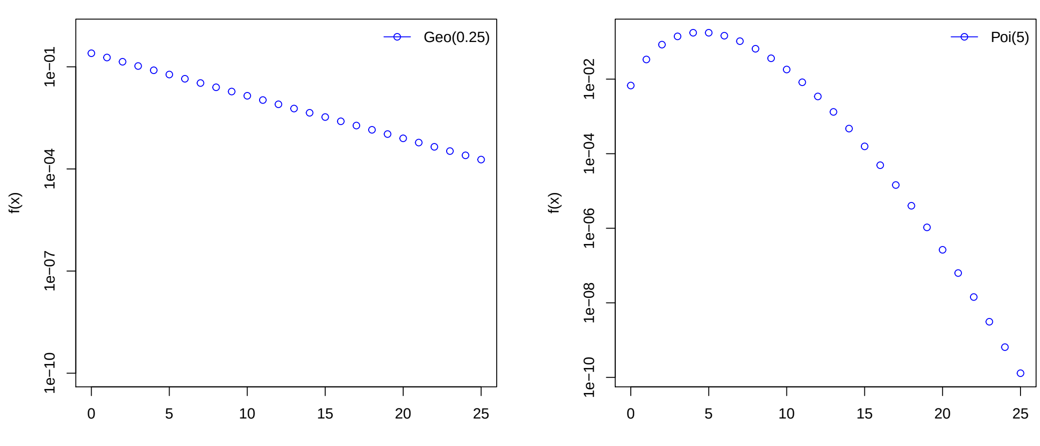
Which is TRUE?

1. $\mathbb{P}(X = 25)$ is much larger for the distribution on the left than for the distribution on the right.
2. $\mathbb{P}(X = 25)$ is much larger for the distribution on the right than for the distribution on the left.
3. $\mathbb{P}(X = 25)$ is about the same for the distributions on the left and on the right.

Solution 3.13. It is hard to see from the picture, but we can explicitly compute the probabilities. Let $X \sim \text{Geo}(0.25)$ and $Y \sim \text{Poi}(5)$.

$$p_X(0.25) = 0.25(1 - 0.25)^{25} \approx 0.0002 \qquad f_Y(25) = e^{-5}\frac{5^{25}}{25!} \approx 1.3 \times 10^{-10}.$$

The PMFs on the log scale is shown below



Probability function of Geo(0.25) with log−scale on y axis

Probability function of Poi(5) with log−scale on y axis

4 Proofs of Key Results

Problem 4.1. Find an example of random variables such that $X \overset{d}{=} Y$, but $X \neq Y$.

Solution 4.1. We define $X = \text{Bern}(0.5)$ and $Y = 1 - X$. Clearly, $X \neq Y$ since $X = 1 \implies Y = 0$ and $X = 0 \implies Y = 1$. However,

$$f_Y(1) = \mathbb{P}(Y = 1) = \mathbb{P}(1 - X = 1) = \mathbb{P}(X = 0) = \frac{1}{2}$$

and

$$f_Y(0) = \mathbb{P}(Y = 0) = \mathbb{P}(1 - X = 0) = \mathbb{P}(X = 1) = \frac{1}{2},$$

so Y has the same PMF as a Bern(0.5) random variable, and in F_X and F_Y are identical since the CDF is in direct correspondence with the PMF.

Remark 9. This is example is equivalent to the following. You flip a single coin. Let X denote the number of heads, and let Y denote the number of tails. It is clear that $X \neq Y$, but $X \overset{d}{=} Y$ since,

$$\mathbb{P}(T) = \mathbb{P}(Y = 1) = \mathbb{P}(X = 1) = \mathbb{P}(H) = \frac{1}{2}$$

and

$$\mathbb{P}(H) = \mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = \mathbb{P}(T) = \frac{1}{2}.$$

Problem 4.2. Find an example of random variables that is discrete while its underlying sample space is not.

Solution 4.2. A random variable may be discrete even though the underlying sample space might not be. For example, if $\Omega = [0, 1]$, the random variable

$$X(\omega) = \mathbb{1}(\omega \leq 0.5) = \begin{cases} 1, & \text{if } \omega \leq 0.5 \\ 0, & \text{otherwise} \end{cases}.$$

Problem 4.3. Derive the PMF for the hypergeometric function.

Solution 4.3. Recall that N denotes the size of the population, and there are r successes and $N - r$ failures. We consider samples of size n from the population without replacement, and let X denote the number of successes in the draw. We first find the support of p_X.

- We cannot have more successes x than the total successes $r \Rightarrow x \leq r$.
- We cannot have more successes x than the total trials $n \Rightarrow x \leq n$.
- We cannot have less than 0 successes $\Rightarrow x \geq 0$.
- When there are more trials than failures $n > (N - r)$ we will for sure have at least $n - (N - r)$ successes $\Rightarrow x \geq n - (N - r)$.
- Altogether, $\max\{0, n - (N - r)\} \leq x \leq \min\{r, n\}$.

We now find the $p_X(x)$ for $\max\{0, n - (N - r)\} \leq x \leq \min\{r, n\}$. There are $\binom{N}{n}$ ways to draw n items from a population of size N without replacement, which is our sample space. We now count the number of ways to get exactly x successes in this sample. There are $\binom{r}{x}$ ways to pick x successes out of the possible r successes, and there are $\binom{N-r}{n-x}$ ways to pick the remaining failures, and so the product encodes the total number of ways to get exactly x successes in this sample. Since the probability is uniform on the sample space of draws,

$$p_X(x) = \mathbb{P}(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}.$$

Problem 4.4. Let $X_1, \dots, X_n$ are independent Bern(p) random variables. Show that the sum $S_n = X_1 + \cdots + X_n$ has a Bin(n,p) distribution. In other words, show that $S_n \sim \text{Bin}(n,p)$.

Solution 4.4. We see that S_n can take values in $\{0, 1, \dots, n\}$ since they are the sum of random variables that takes values in $\{0,1\}$. We want to find $p_X(k)$. Let $(x_1, \dots, x_n) \in \{0,1\}^n$ be such that $\sum_i x_i = k$, which is equivalent to saying that exactly k coordinates are 1. We have

$$\mathbb{P}((X_1, \dots, X_k) = (x_1, \dots, x_n)) = \mathbb{P}(X_1 = x_1) \dots \mathbb{P}(X_n = x_n) = p^k (1-p)^{1-k}$$

by independence. There are $\binom{n}{k}$ ways to choose the coordinates of $(x_1, \dots, x_n)$ such that exactly k are 1 , so by symmetry.

$$p_X(k) = \mathbb{P}(S_n = k) = \sum_{\substack{(x_1, \dots, x_n) \\ \sum x_i = k}} \mathbb{P}((X_1, \dots, X_k) = (x_1, \dots, x_n)) = \binom{n}{k} p^k (1-p)^{1-k}.$$

Remark 10. We can think of the X_i as denoting whether the ith independent draw was a success. The total number of successes in n draws with replacement was a success (since we need the probability of success to be the same for all X_i) is encoded by S_n, which has Binomial distribution.

Problem 4.5. Let $p \in (0,1)$ and let $X \sim \text{Hyp}(N, pN, n)$ and $Y \sim \text{Bin}(n,p)$. Show that for all $x \in \mathbb{R}$,

$$\lim_{N \to \infty} \mathbb{P}(X \le x) = \mathbb{P}(Y \le x).$$

Solution 4.5. Recall that for p_X is supported on integers such that

$$\max\{0, n - (N - Np)\} \le x \le \min\{Np, n\}$$

which is equal to $0 \le x \le n$ for N sufficiently large since $p \in (0,1)$. Therefore, both PMF functions are discrete and supported on $\{0, 1, \dots, n\}$ so it suffices to compute its PMF functions, from which one can trivially compute the CDF. We first rewrite the PMF of p_X with $r = pN$,

$$\begin{aligned}
p_X(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} &= \frac{r!}{x!(r-x)!} \cdot \frac{(N-r)!}{(n-x)!(N-r-(n-x))!} \cdot \frac{n!(N-n)!}{N!} \\
&= \frac{n!}{x! \cdot (n-x)!} \cdot \frac{r!}{(r-x)!} \cdot \frac{(N-r)!}{(N-r-(n-x))!} \\
&= \binom{n}{x} \cdot \prod_{i=1}^{x} (r - x + i) \cdot \prod_{j=1}^{n-x} (N - r - (n-x) + j) \prod_{k=1}^{n} \frac{1}{(N-n+k)} \\
&= \binom{n}{x} \cdot \prod_{i=1}^{x} \frac{r-x+i}{N-x+i} \cdot \prod_{j=1}^{n-x} \frac{(N-r-(n-x)+j)}{N-n+m}
\end{aligned}$$

The result follows from the fact that $\frac{r}{N} \to p$, so for any fixed i and j,

$$\lim_{N \to \infty} \frac{r-x+i}{N-x+i} = p \qquad \text{and} \qquad \lim_{N \to \infty} \frac{(N-r-(n-x)+j)}{N-n+m} = 1-p$$

which implies that

$$\lim_{N \to \infty} p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Problem 4.6. If $p \in (0,1)$, show that

$$p_X(x) = p(1-p)^x \quad \text{for } x = 0,1,2,\dots.$$

is a valid PMF.

Solution 4.6. Clearly, $p_X(x) \geq 0$ on its support. We need to check if the sum is over the support 1. Using the formula for the sum of a geometric series i.e. if $|q| < 1$, then $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$, we see that

$$\sum_{x \geq 0} p(1-p)^x = \frac{p}{1-(1-p)} = 1$$

as required.

Problem 4.7. Find the CDF of a Geo(p) random variable.

Solution 4.7. For $x = 0,1,2,\dots$, we have by the sum of a geometric series i.e. if $|q| < 1$, then $\sum_{k=0}^{n-1} q^k = \frac{1-q^n}{1-q}$ that

$$F_X(x) = \sum_{k \leq x} p(1-p)^k = p\frac{1-(1-p)^{x+1}}{1-(1-p)} = 1-(1-p)^{x+1}.$$

Clearly, $F_X(x) = 0$ for $x < 0$ and by linearly interpolating between the discontinuities, we see that $F_X(x) = 1-(1-p)^{\lfloor x \rfloor + 1}$ for $x \geq 0$.

Problem 4.8. Let $X \sim \text{Geo}(p)$ and s,t be non-negative integers. Show that

$$\mathbb{P}(X \geq s+t \mid X \geq s) = \mathbb{P}(X \geq t).$$

Solution 4.8. Notice that for any integer r,

$$\mathbb{P}(X \geq r) = 1 - \mathbb{P}(X \leq r-1) = 1 - F_X(r-1) = (1-p)^r.$$

Therefore,

$$\mathbb{P}(X \geq s+t \mid X \geq s) = \frac{\mathbb{P}(X \geq s+t, X \geq s)}{\mathbb{P}(X \geq s)} = \frac{\mathbb{P}(X \geq s+t)}{\mathbb{P}(X \geq s)} = \frac{(1-p)^{s+t}}{(1-p)^s} = \mathbb{P}(X \geq t).$$

Problem 4.9. Suppose that X is discrete and for all s,t be non-negative integers

$$\mathbb{P}(X \geq s+t \mid X \geq s) = \mathbb{P}(X \geq t).$$

Show that X must have a Geometric distribution.

Solution 4.9. By assumption, for every $n \geq 1$

$$\mathbb{P}(X \geq n+1 \mid X \geq 1) = \frac{\mathbb{P}(X \geq n+1)}{\mathbb{P}(X \geq 1)} = \mathbb{P}(X \geq n).$$

Rearranging this implies that

$$\mathbb{P}(X \geq n+1) = \mathbb{P}(X \geq 1)\,\mathbb{P}(X \geq n) = (1 - \mathbb{P}(X = 0))\,\mathbb{P}(X \geq n) = (1-p)\,\mathbb{P}(X \geq n)$$

where we set $p = \mathbb{P}(X = 0)$ (which is consistent with the meaning in the Geometric distribution). Since this holds for all n, we can continue inductively to see that

$$\mathbb{P}(X \geq n+1) = (1-p)^{n+1}.$$

Therefore,

$$\begin{aligned}\mathbb{P}(X = n) = \mathbb{P}(X \geq n) - \mathbb{P}(X \geq n+1) &= (1-p)^n - (1-p)^{n+1} \\ &= (1-(1-p))(1-p)^n = p(1-p)^n.\end{aligned}$$

Problem 4.10. Show that

$$\binom{x+k-1}{x} = (-1)^x \binom{-k}{x}.$$

Solution 4.10. By definition,

$$\binom{x+k-1}{x} = \frac{(x+k-1)(x+k-2)\cdots(k+1)k}{x!}.$$

We can factor out (-1) from each term in the numerator (there are a total of x of them) to conclude that

$$\begin{aligned}\frac{(x+k-1)(x+k-2)\cdots(k+1)k}{x!} &= (-1)^x \frac{(-k)(-k-1)\cdots(-k-x+2)(-k-x+1)}{x!} \\ &= (-1)^x \binom{-k}{x}\end{aligned}$$

if we extend the definition of the binomial coefficient to negative integers.

Problem 4.11. Show that

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots.$$

is a valid PMF.

Solution 4.11. Clearly, $p_X(x)$ is non-negative on its support. We need to check if the sum is over the support 1. Using the exponential series $e^t = \sum_{k=0}^\infty \frac{t^k}{k!}$, we have

$$\sum_{x \geq 0} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x \geq 0}\frac{\lambda^x}{x!} = e^{-\lambda+\lambda} = 1.$$

Problem 4.12. Let $\lambda > 0$, and suppose that $p = p_n \to 0$ in such a way such that $np_n \to \lambda$. Let $X \sim \mathrm{Bin}(n, p_n)$ and $Y \sim \mathrm{Poi}(\lambda)$. Show that for $x \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}(X = k) = \mathbb{P}(Y = k).$$

Solution 4.12. By assumption, $p = \frac{\lambda}{n}$. For every fixed integer $x \geq 0$,

$$
\begin{aligned}
p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} &= \frac{n!}{x!(n-x)!} \Big(\frac{\lambda}{n}\Big)^x \Big(1 - \frac{\lambda}{n}\Big)^{n-x} \\
&= \frac{\lambda^x}{x!} \Big(1 - \frac{\lambda}{n}\Big)^n \Big(1 - \frac{\lambda}{n}\Big)^{-x} \prod_{k=0}^{x-1} \frac{n-k}{n}.
\end{aligned}
$$

For each fixed x, we have

$$
\lim_{n\to\infty} \Big(1 - \frac{\lambda}{n}\Big)^{-x} = 1 \qquad \text{and} \qquad \lim_{n\to\infty} \prod_{k=0}^{x-1} \frac{n-k}{n} = 1.
$$

Furthermore, $e^{-\lambda} = \lim_{n\to\infty}(1 - \frac{\lambda}{n})^n$ by definition, so

$$
\lim_{n\to\infty} \mathbb{P}(X = x) = \lim_{n\to\infty} \frac{\lambda^x}{x!} \underbrace{\Big(1 - \frac{\lambda}{n}\Big)^n}_{\to e^{-\lambda}} \underbrace{\Big(1 - \frac{\lambda}{n}\Big)^{-x}}_{\to 1} \underbrace{\prod_{k=0}^{x-1} \frac{n-k}{n}}_{\to 1} = e^{-\lambda} \frac{\lambda^x}{x!} = \mathbb{P}(Y = x).
$$

Furthermore, $\mathbb{P}(X = x) = \mathbb{P}(Y = x) = 0$ for all other values.