1 Discrete Random Variables Part I

1.1 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. A *random variable* is a function that maps the sample space S to the set of real numbers $\mathbb{R}$. That is, X is a random variable if

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

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

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

Definition 3. 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

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

Definition 4. There are two "informal" classifications of random variables we consider in this course.

- We say that a random variable is *discrete* if its range is a discrete subset of $\mathbb{R}$ (i.e., a finite or a countably infinite set).
- A random variable is *continuous* if its range is an interval that is a subset of $\mathbb{R}$ (e.g. $[0,1], (0,\infty), \mathbb{R}$).

There are random variables that are neither discrete or continuous, such as mixed random variables.

Remark 1. A random variable may be discrete even though the underlying sample space might not be (see Problem 1.14).

Definition 5. Random variables X and Y are *independent* if for any events A, B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\,\mathbb{P}(Y \in B).$$

1.1.1 Probability Mass Function (PMF)

Random variables induce a probability on its range $X(S) \subseteq \mathbb{R}$. This is very convenient because we no longer have to consider probabilities on sets, but rather probabilities on the real line.

Definition 6. The *probability (mass) function* of a discrete random variable X is the function

$$f_X(x) = \mathbb{P}(X = x) := \mathbb{P}(\{\omega \in S : X(\omega) = x\}) = \mathbb{P}(X^{-1}(x)) = (\mathbb{P} \circ X^{-1})(x).$$

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

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

1. $$0 \le f_X(x) \le 1 \quad \text{for all } x$$
2. $$\sum_{x \in X(S)} f_X(x) = 1.$$

An important implication of this fact is that we no longer have to specify what the underlying sample space S 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 $X(S)$ and its distribution f_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.

1.1.2 Cumulative Distribution Function (CDF)

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

Definition 7. The *cumulative distribution function* (CDF) of a random variable X is

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

The CDF is not a probability measure, but instead satisfies 4 characterizing properties

1. $0 \le F_X(x) \le 1$
2. $F_X(x) \le F_X(y)$ for $x < y$
3. $\lim_{x \to -\infty} F_X(x) = 0$, and $\lim_{x \to \infty} F_X(x) = 1$
4. F_X is right continuous, i.e., $F_X(x) = F_X(x^+) = \lim_{t \downarrow x} F_X(t)$ for all $x \in \mathbb{R}$

In fact, any function that satisfies these properties defines a CDF, so we can find the random variable associated with this function. As a consequence, if two random variables have the same CDF, they encode the same probability measure on $X(S)$.

Definition 8. 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

$$X \sim Y$$

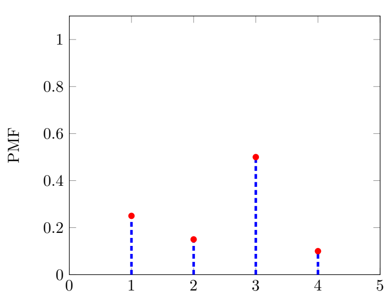
Remark 2. Random variables X and Y being equal in distribution does not mean $X = Y$ (see Problem 1.13). It just means that the probability of X and Y taking any particular value is the same. In fact, X and Y don't even have to be functions defined on the same sample space.

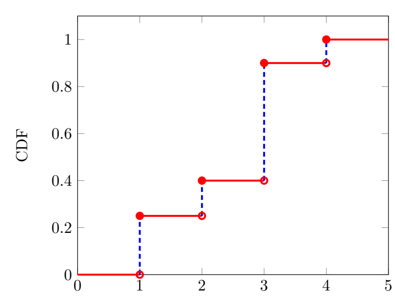
1.1.3 Connection Between the PMF and CDF

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

1. If X is discrete with PMF f_X, then
$$F_X(x) = \sum_{y \le x} f_X(y).$$
Notice that $F_X(x)$ is constant between consecutive points in the support of f_X.
2. If X is discrete with CDF F_X, then
$$f_X(x) = F_X(x) - F_X(x^-) =: F_X(x) - \lim_{t \uparrow x} F_X(t).$$
Notice that $f_X(x)$ 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.

1.2 Important Discrete Distributions (Part I)

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

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

Definition 9. 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

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

- **PMF:**

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

- **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 $A\spadesuit$	$U[1,52]$

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

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

Definition 10. 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:**

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

- **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	$\mathrm{Hyp}(50,7,7)$

1.2.3 Bernoulli Distribution: Bern(p)

The Bernoulli distribution models experiments with two possible outcomes.

Definition 11. 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

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

- **PMF:**
$$f_X(0) = 1 - p, \quad f_X(1) = p$$
- **CDF:**
$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1 - p, & \text{if } 0 \le x < 1, \\ 1, & \text{if } x \ge 1 \end{cases}$$

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

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

1.2.4 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 12. 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

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

- **PMF:**
$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n$$
- **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	Bin($10, \frac{1}{6}$)
Buy 1 tickets from a Lotto where 7 numbers are drawn from 50	# Jackpots	Bin($1, \binom{50}{77}^{-1}$)
Generate each digit of a 5 digit number randomly from 1 to 9	# odd digits	Bin($5, \frac{5}{9}$)

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 1 (*Binomial Approximation of the Hypergeometric Distribution*)

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

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

1.3 Example Problems

1.3.1 PMF and CDF 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 denote sthe gain, what is f_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

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

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

and $f_X(x) = 0$ otherwise.

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

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

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

We see that $x = 7$ maximizes the PMF f_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 $f_X(2)$?

Solution 1.3. We want to compute $f_X(2) = \mathbb{P}(X = 2)$. 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 $f_X(2) = \frac{9}{36} = \frac{1}{4}$.

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

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

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

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

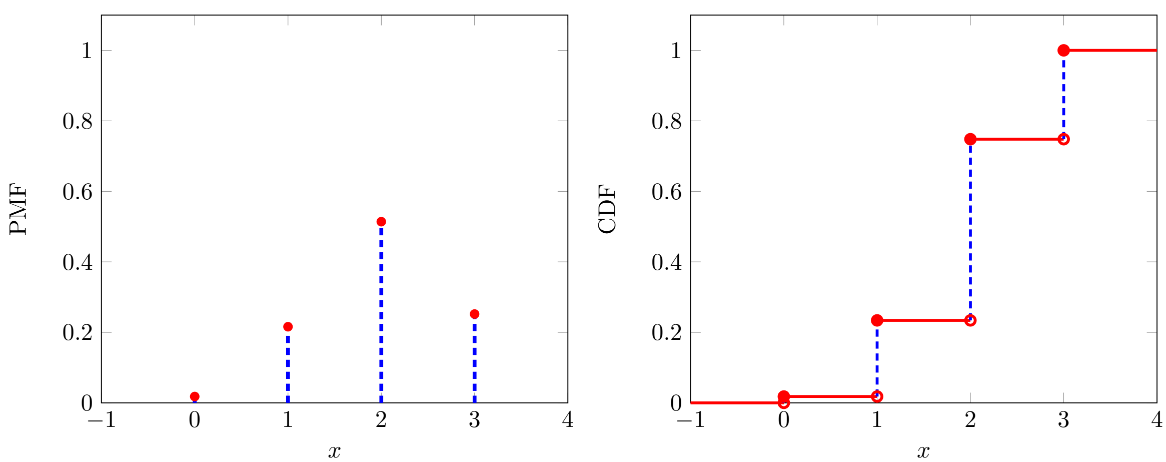
and, since the PMF sums to 1,

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

The CDF is thus

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

The plot of the CDF is below:



Remark 3. 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 function of X?

Solution 1.6. The underlying sample space is $S = \{H, T\}$. 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 $S = \{0, 1, \ldots 9\}$ 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 $\{0, 1, \ldots, 99\}$.

Problem 1.8. 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 1.8. 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 $X \sim \text{Hyp}(N = 52, r = 4, n = 5)$. 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 1.9. 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 1.9. This is modeled using a binomial distribution. Let X denote the number of times the tack lands point up. Then $X \sim \text{Bin}(10, 0.6)$ 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 1.10. 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 1.10. 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

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

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

Alternative Solution: By the assumptions, the sample space $S = [5]^{10}$ (which denotes where each passenger got off) has equally likely outcomes. To find $f_X(x) = \mathbb{P}(X = x)$, 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 S,

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

Remark 4. Clearly, f_X is not uniform, so the number of passengers that got off at the first stop is not uniform over the range $X(S) = \{0, 1, 2, \ldots, 10\}$. 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 1.11. 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 1.11. Since we are trying keys randomly without replacement, the location of the correct key is uniform over the set $\{1, \ldots, n\}$ by symmetry. We can model this with a uniform distribution, so $X \sim \mathrm{U}[1, n]$. Therefore,

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

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

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 $f_X(x) = \mathbb{P}(X = x)$, 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 S,

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

Problem 1.12. 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 1.12. 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.

1.3.2 Derivation of Distributions

Problem 1.13. $(*)$ Find an example of random variables such that $X \sim Y$, but $X \neq Y$.

Solution 1.13. 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 5. 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 $X \sim Y$ 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 1.14. Find an example of random variables that is discrete while its underlying sample space is not.

Solution 1.14. A random variable may be discrete even though the underlying sample space might not be. For example, if $S = [0, 1]$, 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 1.15. (*) Derive the PMF for the hypergeometric function.

Solution 1.15. 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 f_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 $f_X(x)$ 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,

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

Problem 1.16. (*) 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 1.16. 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 $f_X(k)$. 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.

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

Remark 6. 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 1.17. (*) 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 $x \in \mathbb{R}$,

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

Solution 1.17. Recall that for f_X is supported on integers such that

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

which is equal to $0 \leq x \leq n$ 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 f_X with $r = pN$,

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

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

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