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 underlying probability spaces, which is often tedious to work with in practice. For example, the probability space for the flips of 100 coins is $\{H, T\}^{100}$ but if we only care about the number of heads, then it will be much easier to work with the sample space $\{0, 1, \ldots, 100\}$. Random variables allow us to bypass the underlying probability space and work with the simpler sample space.

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 (Sample Space)**.** The values in $\mathbb{R}$ that a random variable takes is called the *sample space* of the random variable, and is denoted by

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

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 $(S, \mathcal{B}, \mathbb{P}_X)$ defines a new probability space with outcomes given by the sample space, 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 first introduce a way to encode the distribution of X that makes sense for all random variables. The probabilities of the form $\mathbb{P}(X \le x)$ encodes all the information to describe the distribution of X.

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. From the CDF, we can easily compute any probability, since the intervals $(a, b]$ and $(-\infty, a]$ are disjoint,

$$\mathbb{P}(a < X \le b) = \mathbb{P}(X \le b) - \mathbb{P}(X \le a) = F_X(b) - F_X(a), \tag{1}$$

so the CDFs encode the same information as a probability distribution.

Remark 2. It is possible that a CDF does not fall under either of these categories, such as mixed random variables with have CDFs with both jump discontinuities and strictly increasing parts.

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: $F_X(x) \leq F_X(y)$ *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

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

Remark 3. Random variables X and Y being equal in distribution does not mean $X = Y$ (see Problem 1.10). 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.

From the point of view of CDFs, we have the following natural classifications of random variables.

Definition 7 (Discrete vs Continuous Random Variable)**.** If the CDF of X is

1. a *piecewise constant* function, then X is a *discrete* random variable.
2. a *continuous* function, then X is a *continuous* random variable.

Remark 4. It is possible that a CDF does not fall under either of these categories, such as mixed random variables with have CDFs with both jump discontinuities and strictly increasing parts.

1.2 Discrete Random Variables

If the CDF is *piecewise constant* function, then X is a *discrete* random variable. Equivalently, this happens if the sample sapce is a discrete subset of $\mathbb{R}$.

Remark 5. A random variable may be discrete even though the underlying sample space might not be (see Problem 1.11).

1.2.1 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 8. 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. $$0 \le p_X(x) \le 1 \quad \text{for all } x$$

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

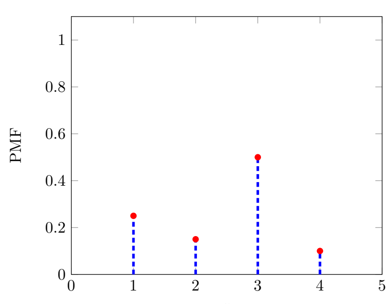
An important implication of this fact is that we no longer have to specify what the underlying sample space Ω with probability measure $\mathbb{P}$ to compute probabilities. If X encodes the quantities we need to assign probabilities to, then we can work directly with the sample space $X(\Omega)$ and 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.

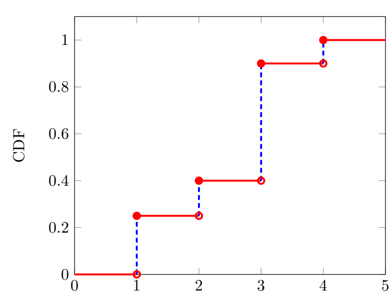
The PMF and CDF encode the same information for discrete random variables.

1. If X is discrete with PMF p_X, then
$$F_X(x) = \sum_{y \le x} p_X(y).$$
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. We can think of the PMF as the discrete rate of change of the CDF.

1.3 Continuous Random Variables

In contrast to discrete random variables, continuous random variables can take uncountably many values so it is impossible to assign a probability to every element in the sample space.

1.4 Probability Density Function (PDF)

We can try to define the PMF for a continuous random variable, but the fact that $\mathbb{P}(X = x) = 0$ for all $x \in \mathbb{R}$ is problematic. We get around this by defining a new way to encode how likely a certain value x is without defining it as a probability.

Definition 9 (Absolutely Continuous Random Variable)**.** We say that a continuous random variable X with distribution function F_X is *absolutely continuous* if it is the antiderivative of some function f_X,

$$F_X(x) = \mathbb{P}(X \le x) = \int_{-\infty}^{x} f_X(t)\,dt.$$

Remark 6. It might be the case that $F_X(x)$ is not differentiable at isolated points. However, if $f_X(x)$ is continuous at x, then

$$F_X'(x) = \frac{d}{dx}F_X(x) = f_X(x).$$

Likewise, jump discontinuities of f_X correspond to non-differentiable points of F_X.

Definition 10 (Probability Density Function)**.** The function f_X is called the *probability density function* (PDF). The function $f_X(x)$ and it satisfies the following properties

1. $f_X(x) \ge 0$ for all $x \in \mathbb{R}$;
2. $\int_{-\infty}^{\infty} f_X(x)dx = 1$;

The *support* of f_X (or X) is the closure of the set of non-zero values of the PDF,

$$\mathrm{supp}(f_X) = \mathrm{cl}(\{x \in \mathbb{R} : f_X(x) \ne 0\}).$$

Remark 7. By convention we take the closure of the set, which means that we always include the endpoints of intervals in the support. This is not a big issue since $f_X(x)$ can be arbitrarily defined at the endpoints since the area under the PDF does not change if we redefine the endpoints.

Thus the PDF plays the same role the analogous role of the PMF for discrete random variables, in the sense that it encodes the instantaneous rate probabilities accumulate. More precisely, the PDF $f_X(x)$ is proportional to the probability that X lies in a small interval around x in the sense that for ϵ small

$$\mathbb{P}\left(x - \frac{\epsilon}{2} \le X \le x + \frac{\epsilon}{2}\right) = \int_{x-\frac{\epsilon}{2}}^{x+\frac{\epsilon}{2}} f_X(x)\,dt \approx \epsilon f_X(x).$$

So $f_X(x)$ isn't the probability that $X = x$, but it encodes the likelihood of x compared to other values. In fact, $f_X(x)$ cannot be a probability since it is possible for it to take values larger than 1.

Remark 8. In this course, unless otherwise stated, all continuous random variables will be absolutely continuous so we often drop the prefix "absolutely".

1.5 Independence

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

Definition 11 (Independence)**.** Random variables $X_1, X_2, \dots, X_n$ are *independent* if for all $x_1, x_2, \dots, x_n \in \mathbb{R}$,

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

Remark 9. Independence is quite a strong condition, since it must hold for all $x_1, x_2, \dots, x_n \in \mathbb{R}$. This is the analogue of mutual independence of all the events $\{X_i \le x_i\}$ 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 12 (i.i.d.)**.** A sequence of random variables $X_1, \dots, X_n$ are called independent and identically distributed (i.i.d.) if they are independent and have the same distribution.

Remark 10. 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.

1.6 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

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

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

$$\begin{aligned}p_X(1)=\mathbb{P}(X=1)&=\mathbb{P}(A\cap B^c\cap C^c)+\mathbb{P}(A^c\cap B\cap C^c)+\mathbb{P}(A^c\cap B^c\cap C)\\&=\frac{9\cdot3\cdot6}{1000}+\frac{7\cdot1\cdot6}{1000}+\frac{4\cdot1\cdot3}{1000}=\frac{216}{1000}\end{aligned}$$

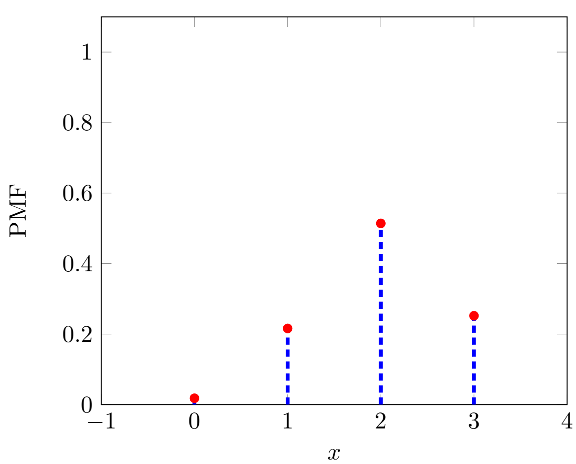
and, since the PMF sums to 1,

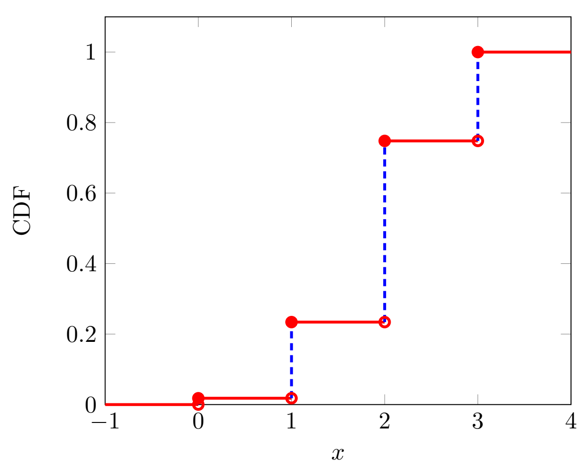
$$p_X(2)=1-p_X(0)-p_X(1)-p_X(3)=\frac{514}{1000}.$$

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 11. The end points of the intervals in the CDF are the same as the non-zero values of the PMF. Furthermore, the $\leq$ 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

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

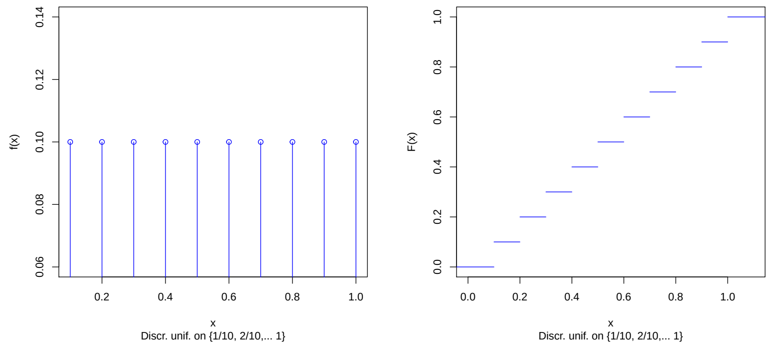
We have that X is uniformly distributed on the set of $\{0, 1, \dots, 99\}$.

Problem 1.8. Suppose we are cutting a stick of length 1 randomly and denote by X the cutting point. The random variable X is continuous with range $[0, 1]$.

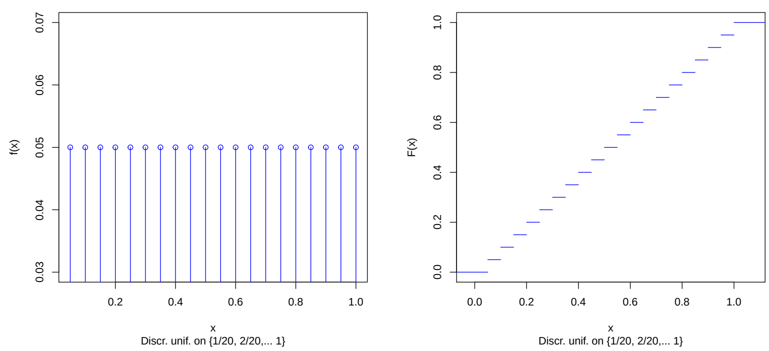
1. Approximate X with a discrete uniform distribution on $\{\frac{1}{10}, \frac{2}{10} \dots, 1\}$. Draw the PMF and CDF.
2. Approximate X with a discrete uniform distribution on $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$. Draw the PMF and CDF.
3. Approximate X with a discrete uniform distribution on $\{\frac{1}{100}, \frac{2}{100} \dots, 1\}$. Draw the PMF and CDF.
4. Approximate X with a discrete uniform distribution on $\{\frac{1}{10000}, \frac{2}{10000} \dots, 1\}$. Draw the PMF and CDF.

Solution 1.8.

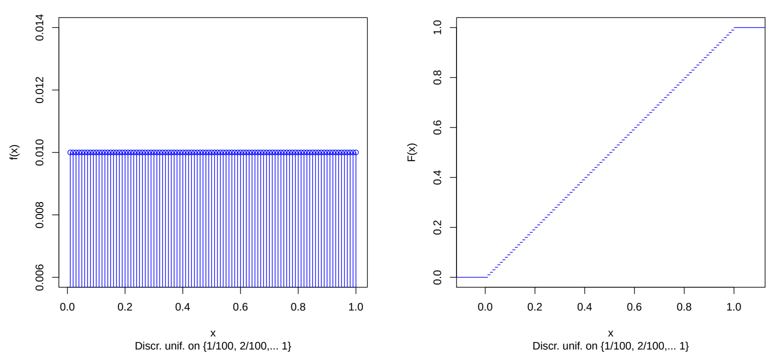
1. We have $f_X(x) = \frac{1}{10}$ for all $x \in \{\frac{1}{10}, \frac{2}{10} \dots, 1\}$.



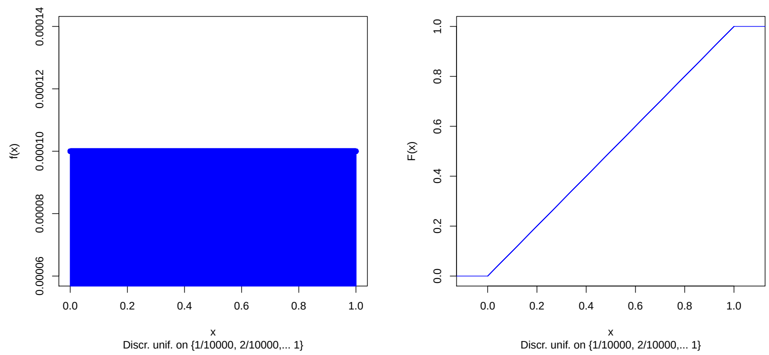
2. We have $f_X(x) = \frac{1}{20}$ for all $\{\frac{1}{20}, \frac{2}{20} \dots, 1\}$.



3. We have $f_X(x) = \frac{1}{100}$ for all $x \in \{\frac{1}{100}, \frac{2}{100} \dots, 1\}$



4. We have $f_X(x) = \frac{1}{10000}$ for all $x \in \{\frac{1}{10000}, \frac{2}{10000} \dots, 1\}$.



Remark 12. Notice that the PMF gradually tends to 0 and the CDF gets closer to CDF of the continuous uniform distribution on $[0, 1]$.

Problem 1.9. Suppose that X is a continuous random variable with PDF

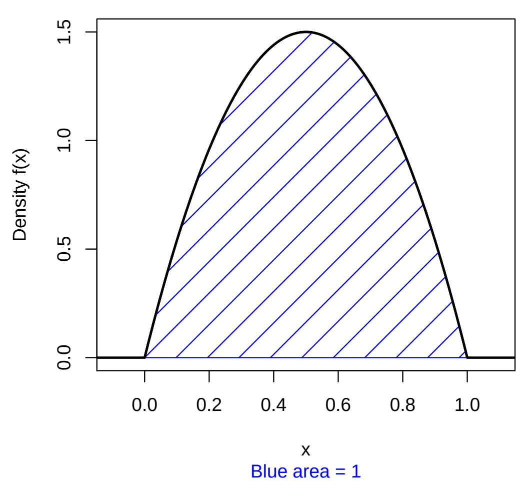
$$f_X(x) = \begin{cases} cx(1-x) & \text{if } 0 \leq x \leq 1, \\ 0 & \text{otherwise} \end{cases}$$

1. Compute c so that this is a valid pdf.
2. Compute the cdf $F_X(x)$.
3. Compute $\mathbb{P}(1/4 \leq X \leq 3/4)$

Solution 1.9.

1. We need $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$, so

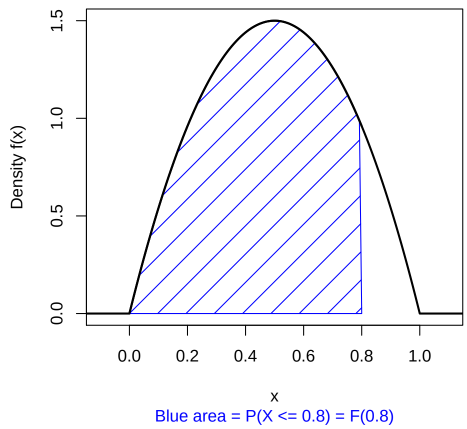
$$\int_0^1 cx(1-x)\, dx = \frac{c}{6} = 1 \implies c = 6.$$

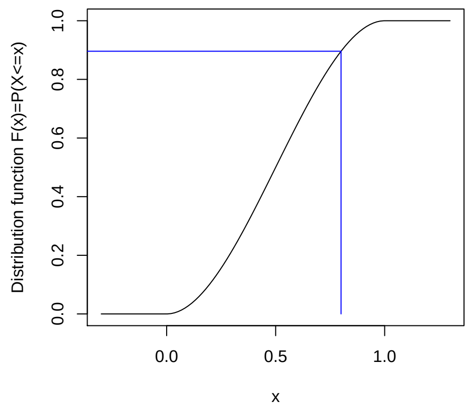


2. The CDF $F_X(x)$ is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt = \int_0^x 6t(1-t)\,dt = 3x^2 - 2x^3, \quad x \in [0,1]$$

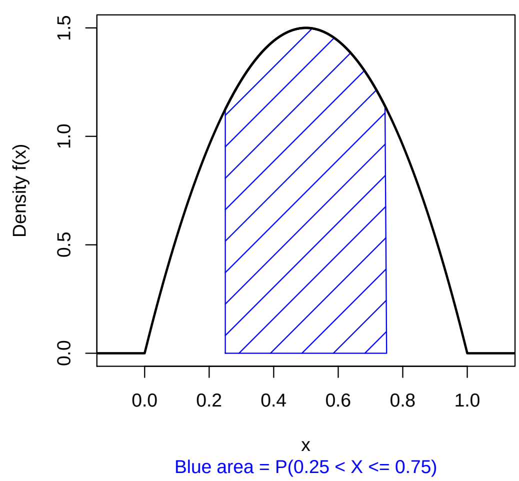
and $F_X(x) = 0$ for $x \leq 0$ and $F_X(x) = 1$ for $x \geq 1$.





3. We compute the probability using the CDF:

$$\mathbb{P}\left(\frac{1}{4} \leq X \leq \frac{3}{4}\right) \overset{\text{cont.}}{=} \mathbb{P}\left(\frac{1}{4} < X \leq \frac{3}{4}\right) = \mathbb{P}(X \leq 3/4) - P(X \leq 1/4)$$
$$= F_X(3/4) - F_X(1/4) = 11/16.$$



Alternative Solution: By integrating the PDF

$$\mathbb{P}(1/4 \leq X \leq 3/4) = \int_{1/4}^{3/4} f_X(t)\,dt = 11/16.$$

1.7 Proofs of Key Results

Problem 1.10. Find an example of random variables such that $X \overset{d}{=} Y$, but $X \neq Y$.

Solution 1.10. 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 13. 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 1.11. Find an example of random variables that is discrete while its underlying sample space is not.

Solution 1.11. A random variable may be discrete even though the underlying sample space might not be. For example, if $\Omega = [0, 1]$, the random variable

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