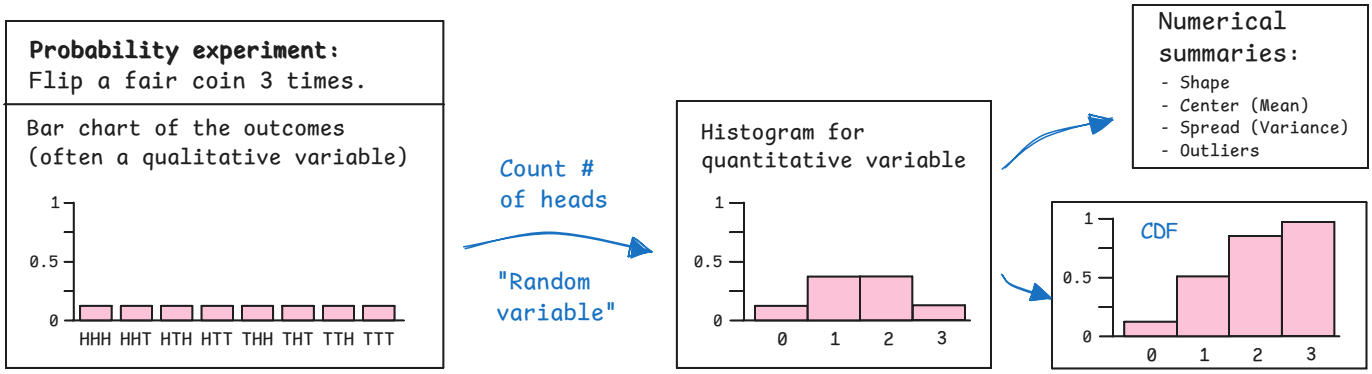
13 - Random variables [P] 3.1-3.3 [PS] 6.1



Definition. A **random variable** (RV) is a function X from a sample space S to the real numbers. Informally, an RV converts qualitative and quantitative outcomes to quantitative outcomes.

Example 1. A biased coin has probability p of heads and probability $q = 1 - p$ of tails. Flip this coin infinitely many times. This probability experiment is called a Bernoulli process. Some RV's associated with this experiment:

(a) Binomial random variable X: number of heads in the first n flips.

Fair coin with n=3: f(0) = P(X=0) = 1/8, f(1) = P(X=1) = 3/8,
f(2) = P(X=2) = 3/8, f(3) = P(X=3) = 1/8

(b) Geometric random variable Y: number of tails before a first head.

Fair coin: g(0) = P(Y=0) = 1/2, g(1) = P(Y=1) = 1/4,
g(2) = P(Y=2) = 1/8, g(3) = P(Y=3) = 1/16, etc.

(c) "Gambler's RV": $ gained by assigning $ values to outcomes in n flips.

Definition. An RV is **discrete** if the set of outcomes is countable, and **continuous** if it takes on values on a continuous scale.

The **probability distribution aka probability mass function (PMF)** of a discrete RV X is often written as:

$$f(x) = P(X = x)$$

The **cumulative distribution function (CDF)** of X is written as

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

In Example 1a, the CDF of X is:

x	0	1	2	3
F(x)	0.125	0.5	0.875	1

Example 2. 3 of 20 servers are defective. Of the 20 servers, 2 are randomly chosen to be inspected. Let X be the random variable for the number of defective servers in the inspected sample. Find the PMF and CDF of X.

(a) $f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$, $f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$,

$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$.

Thus, the probability distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

0.71 0.27 0.02

(b) Its CDF is:

x	0	1	2
F(x)	0.71	0.98	1.00

In practice, the sample space of an RV is not mentioned.

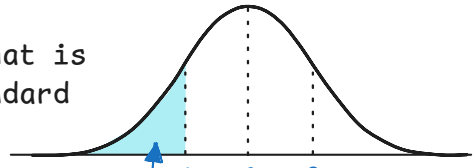
The **probability density function (PDF)** of a continuous RV X is a nonnegative function $f(x)$ defined over the real numbers such that:

$$\int_{-\infty}^{\infty} f(x)\,dx = 1 \qquad \int_{a}^{b} f(x)\,dx = P(a \le X \le b)$$

The **cumulative distribution function (CDF)** of X is

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$$

Example 3. Lifespan of a GPU is modelled as an RV X that is normally distributed with a mean of 6 years and a standard deviation of 2 years. If its CDF is F(x), find F(4).



F(4) = (100-68)/2 = 16%.

Example 4. The PDF of an RV X is given by $f(x) = cx^{-1/2}$ for $0 \le x \le 1$ and $f(x) = 0$ otherwise
(a) Find c.
(b) Find its CDF.
(c) Find the median of X.

(a) $1 = \int_0^1 cx^{-1/2}dx = [2cx^{1/2}]_0^1 = 2c \Rightarrow c = 1/2.$

(b) For $x \ge 0$, $F(x) = \int_0^x (1/2)t^{-1/2}dt = [t^{1/2}]_0^x = \sqrt{x} \Rightarrow F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sqrt{x} & \text{if } 0 \le x \le 1 \\ 1 & \text{if } x > 1 \end{cases}$

(c) $1/2 = F(x) = \sqrt{x} \Rightarrow x = 1/4.$

Answers:
1. (a) 1/17 (b) 1/19
2. (a) 1 (b) 1/2
3. f(0)=2/7, f(1)=4/7, f(2)=1/7
4. (a) 3/16 (b) P(3<X<4)=0.3505
4. (b) $F(x) = \begin{cases} 0, & x < 0 \\ x^{3/2}, & 0 \le x < 4 \end{cases}$
5.(a) $F(x) = \begin{cases} 0, & x < 0, \\ 1 - \exp(-x/1000), & x \ge 0 \end{cases}$
(b) 0.3679; (c) 0.8647
6. (a) 2 (c) sqrt(2)
7. (c) ln(2)

Exercises.
1. Find the value c so that the following functions are PDFs of a discrete RV:
(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2$; (b) $f(x) = c\binom{3}{x}\binom{3}{3-x}$, for $x = 0, 1, 2$;

2. The proportion of people who respond to a mail-order solicitation is a continuous random variable X that has the density function $f(x) = 2(x+2)/5$ for $0 < x < 1$ and $f(x) = 0$ elsewhere.
(a) Show that $P(0 < X < 1) = 1$. (b) Find the probability that more than 1/4 but fewer than 3/4 of the people contacted will respond to this type of solicitation.

3. A shipment of 7 televisions contains 2 defective ones. A hotel makes a random purchase of 3 televisions. If X is the number of defective ones purchased by the hotel, find the probability distribution of X.

4. Consider the density function $f(x) = \begin{cases} k\sqrt{x} & \text{for } 0 \le x < 4 \\ 0 & \text{elsewhere} \end{cases}$.
(a) Evaluate k.
(b) Find F(x) and use it to find P(3<X<4).

5. The time to failure in hours of a PC component has PDF $f(x) = \frac{1}{1000}\exp(-x/1000)$ for $x \ge 1$ and $f(x) = 0$ for $x < 0$.
(a) Find F(x).
(b) Find the probability that the component lasts more than 1000 hours.
(c) Find the probability that the component fails before 2000 hours.

6. Consider the Pareto random variable X whose PDF has the form $f(x) = \begin{cases} c/x^3 & \text{for } x \ge 1 \\ 0 & \text{for } x < 1 \end{cases}$
(a) Find c. (b) Find the CDF of X. (c) Find the median of X.

7. Consider the exponential random variable X whose PDF is given by $f(x) = \begin{cases} e^{-x} & \text{for } x \ge 0 \\ 0 & \text{for } x < 0 \end{cases}$
(a) Verify f is a valid PDF (b) Find the CDF of X.
(c) Find the median of X.