

Random variables

- A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.
- For a given sample space S of some experiment, a **random variable** (rv) is any rule that associates a number with each outcome in S .
- To put it more mathematically, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.
- Remark: a random variable is **NOT** a sample space.

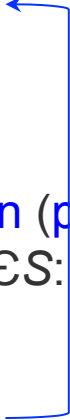
Discrete vs. Continuous

- X is a **discrete random variable** if its possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on (“**countably**” infinite).
- X is a **continuous random variable** if it takes all possible values in an interval of numbers or all numbers in a disjoint union of such intervals. No possible value of the variable has positive probability, that is, $P(X=c) = 0$ for any possible value c .
- X can also be a random variable with a **mixture** distribution of both discrete and continuous components.

PMF

- The probability model for a discrete random variable X , lists its possible values and their probabilities.

Value of X	x_1	x_2	x_k
Probability	p_1	p_2	p_k

- Every probability, p_i , is a number between 0 and 1.
 - $p_1 + p_2 + \dots + p_k = 1$
 - The probability distribution or probability mass function (pmf) of a discrete rv is defined for every number x by $p(x) = P(X=x) = P(\text{all } s \in S: X(s)=x)$.
 - How to check if some function $p(x)$ is a proper PMF?
- 

Bernoulli RV

- The arguably simplest probability model is Bernoulli. Any random variable whose possible values are only 0 and 1 is called a **Bernoulli random variable**.

Ex. Flip a coin. $S=\{H, T\}$. X is a Bernoulli random variable. $X(H)=1$, $X(T)=0$.

$$P(X=1) = 0.5, P(X=0) = 0.5.$$

Ex. Roll a die. $S=\{1, 2, 3, 4, 5, 6\}$. X is a Bernoulli random variable. $X(1)=1$, $X(2)=1$, $X(3)=0$, $X(4)=0$, $X(5)=0$, $X(6)=0$.

$$P(X=1) = 1/3, P(X=0) = 2/3.$$

Example

Ex. Flip three fair coins. (*Binomial*)

$S = \{\text{HHH, HHT, HTH, HTT, THT, THH, TTH, TTT}\}$. Let's define random variable X to be the number of heads in the experiment, i.e., $X(\text{HHH})=3$, $X(\text{THT})=1$, etc.

X

0 TTT

1 TTH THT HTT

2 THH HTH HHT

3 HHH

Value of X	0	1	2	3
Probability	0.125	0.375	0.375	0.125

One can calculate the probability of an event by adding the probabilities p_i of the particular values of x_i that make up the event. For example, if we want to know the probability of getting less than 2 heads, we can use

$$P(X < 2) = P(X=0) + P(X=1) = 0.125 + 0.375 = 0.5$$

$$\text{Note: } P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.875$$

CDF

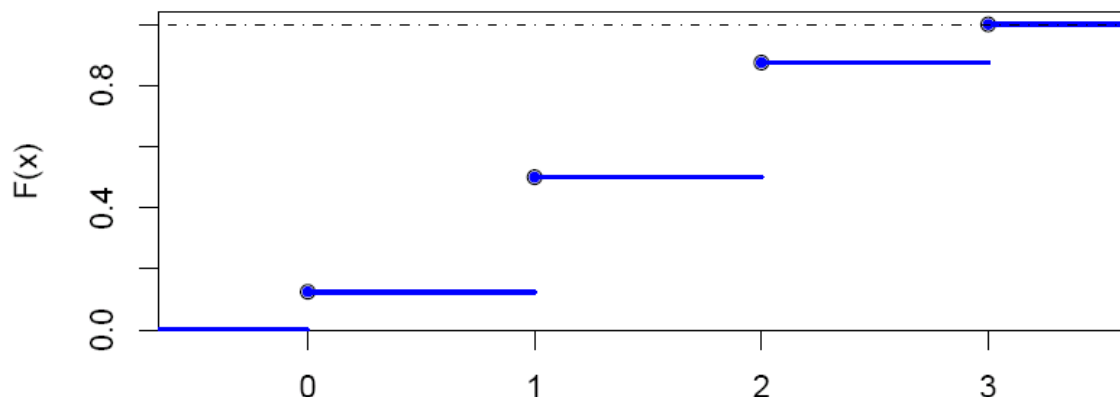
- The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv variable X with pmf $p(x)$ is defined for every number x by

$$F(x) = P(X \leq x) = \sum_{y: y \leq x} p(y).$$

For any number x , $F(x)$ is the probability that the observed value of X will be at most x .

- For X a discrete rv, the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a **step function**.

The three coin flips example



Parameter and Family

- Suppose $p(x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

Ex. For Bernoulli rv's, the parameter is the probability of being 1 (or 0), that is,

$$p = P(X=1)$$

Expectation and Variance

- Random variables have distributions, so they have centers and spreads.
- The **expected value** (**mean value** or **expectation**) of a random variable describes its **theoretical long-run average value**.
- We typically use μ or $E(X)$ to denote the mean, $\text{Var}(X)$ to denote the variance and σ or $\text{SD}(X)$ to denote the standard deviation of a rv X .

Motivating examples

Ex. How many heads would you expect if you flipped a fair coin twice?

$S = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}.$

$X =$ number of heads.

0 TT

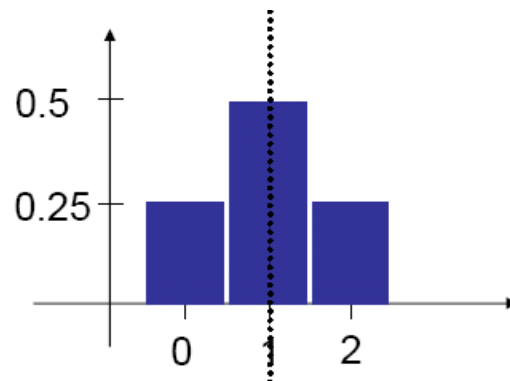
1 HT TH

2 HH

$p(X=0) = 0.25; p(X=1) = 0.5; p(X=2) = 0.25.$

Each outcome is weighted by its probability.

$$\mu = 0 \times 0.25 + 1 \times 0.5 + 2 \times 0.25 = 1$$



Example

Ex. How many heads would you expect if you flipped a coin three times?

$$\mu = 0 \times 0.125 + 1 \times 0.375 + 2 \times 0.375 + 3 \times 0.125 = 1.5$$

This can never occur in a single trial of 3 flips. However, **on average** we would expect to get 1.5 heads if we repeated the experiment many times.

Definition

- Suppose X is a discrete random variable whose probability model is given by

Value of X	x_1	x_2	x_k
Probability	p_1	p_2	p_k

The expected value of X is given by

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x) = x_1 p_1 + x_2 p_2 + \cdots x_k p_k$$

Example

Ex. Expectation of a Bernoulli rv.

$$p(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & x \neq 0,1 \end{cases}$$

$$\mu = 0 \times (1-p) + 1 \times p = p.$$

Example

Ex. The general form for the pmf of X = number of children born up to and including the first boy is,

$$p(x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, 3, \dots \\ 0 & \text{otherwise} \end{cases}$$

1. Verify that this is a proper pmf.
2. Calculate the expected value of X .

The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is X , what is the profit of the bookstore?

The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is X , what is the profit of the bookstore?
- ▶ The profit is $h(X) = 100X - 400$. Is $h(X)$ a Random Variable? Then what is the expectation of $h(X)$?

The Expected Value of a Function

- ▶ A bookstore purchases ten copies of a books at \$60 each to sell at \$120, and any unsold copies after three months can be redeemed for \$20. If the number of copies sold is X , what is the profit of the bookstore?
- ▶ The profit is $h(X) = 100X - 400$. Is $h(X)$ a Random Variable? Then what is the expectation of $h(X)$?
- ▶ If Random Variable X has range D and pmf $p(x)$, then the expected value of function $h(X)$ is given by

$$E(h(X)) = \sum_{x \in D} (h(x) \cdot p(x))$$

The Linear Function Case

- ▶ In the case of linear function, we have a much more convenient formula

$$E(aX + b) = a \cdot E(X) + b$$

The Linear Function Case

- ▶ In the case of linear function, we have a much more convenient formula

$$E(aX + b) = a \cdot E(X) + b$$

- ▶ To prove,

$$E(aX + b) = \sum_{x \in D} (ax + b) \cdot p(x) = a \sum_{x \in D} x \cdot p(x) + b \sum_{x \in D} p(x) = aE(X) + b$$