Probability and Random Process

COSE382

We like numbers than what really is

Grown-ups like numbers. When you tell them about a new friend, they never ask questions about what really matters. They never ask: What does his voice sound like? What games does he like best? Does he collect butterflies? They ask: How old is he? How many brothers does he have? How much does he weigh? How much money does he have? Only then do they think they know him. If you tell grown-ups, I saw a beautiful red brick house, with geraniums at the windows and doves at the roof, they wont be able to imagine such a house. You have to tell them, I saw a house worth a thousand francs. Then they exclaim, What a pretty house!

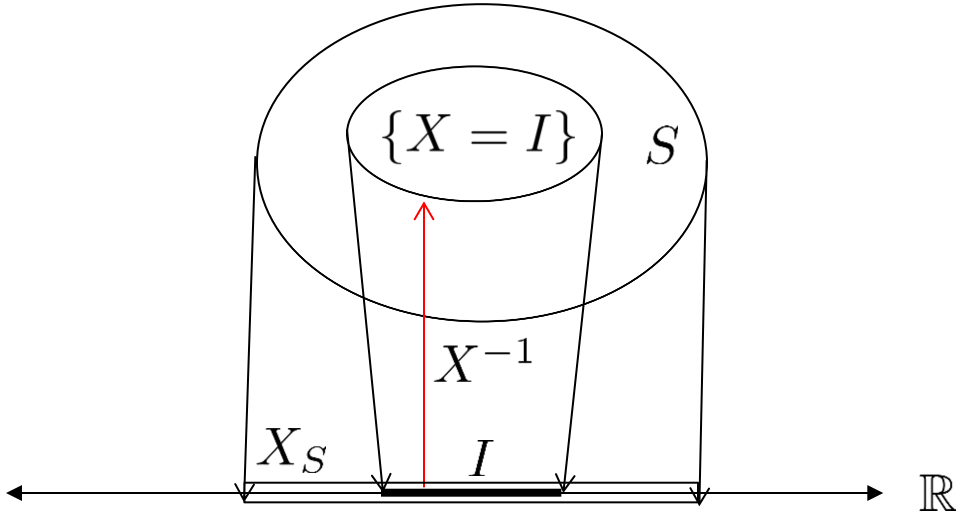
Random Variables

Definition 3.1.1

Given an experiment with sample space S, a *random variable* (r.v.) X is a function from the sample space S to the real numbers $\mathbb{R}$.

$$X : S \to \mathbb{R} \text{ (can be many random variables on a } S\text{)}$$

- X_S denotes the range of X, $\{X(s)|s \in S\} \subset \mathbb{R}$
- $\{X = x\}$ denotes the set $X^{-1}(x) = \{s|X(s) = x\} \subset S$ (an event)
- $\{X \le x\}$ denotes the set $X^{-1}((-\infty, x]) = \{s|X(s) \le x\} \subset S$ (an event)



Example

Example 3.1.2(Coin tosses) Consider an experiment tossing a fair coin twice.
$S = \{HH, HT, TH, TT\}$.

- Let X be the number of Heads. $X_S = \{0, 1, 2\}$

$$X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0.$$

- Let Y be the number of Tails. $Y_S = X_s$, and $Y = 2 - X$, i.e. $Y(s) = 2 - X(s)$ for all s.

- Let I be 1 if first toss lands Heads and 0 otherwise. $I_S = \{0, 1\}$

  We can also encode the sample space as {(1,1),(1,0),(0,1),(0,0)}, where 1 is the code for Heads and 0 is the code for Tails. Then we can give explicit formulas for X, Y, I:

$$\begin{aligned} X(s_1, s_2) &= s_1 + s_2, \\ Y(s_1, s_2) &= 2 - s_1 - s_2, \\ I(s_1, s_2) &= s_1 \end{aligned}$$

  where for simplicity we write $X(s_1, s_2)$ to mean $X((s_1, s_2))$, etc.

Distribution and probability mass function

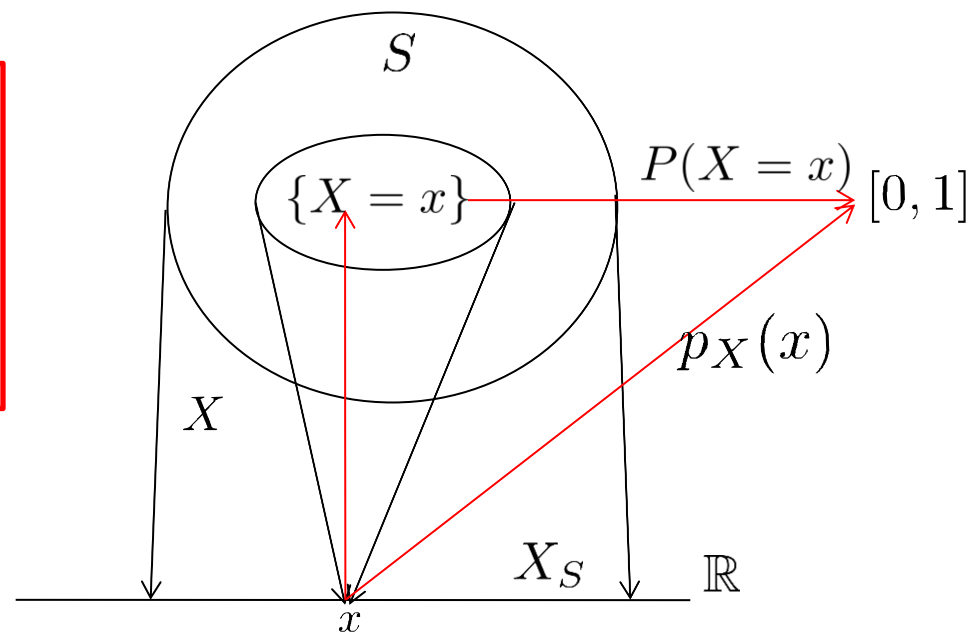
Definition 3.2.1(Discrete random variable).
A random variable X is said to be *discrete* if X_S is discrete (that is countable: X_S is consisted of a finite or infinite list of values $X_S = \{x_1, x_2, ..\}$)

Definition 3.2.2(Probability mass function).
The *probability mass function*(PMF) of a discrete r.v. X is the function $p_X : \mathbb{R} \to [0,1]$ given by

$$p_X(x) = P(X = x) = P(\{X = x\})$$

.

$$\begin{tikzcd} S \arrow[r, "P"] \arrow[d, "X"'] & {[0,1]} \\ \mathbb{R} \arrow[ur, "p_X"'] & \end{tikzcd}$$



- PMF is a function from $\mathbb{R}$ to $[0,1]$ that inherits the characteristics of the X and probability space (S,P).

  "We can forget about the underlying random experiment and its associated probability law and just work with $\mathbb{R}$ and the PMF of X"

$$
\begin{tikzcd}
\text{Experiment specified} & (S,P) \arrow[r, "P"] \arrow[d, "X"'] & {[0,1]} \arrow[d, leftrightarrow, "i.d."] & P : A \mapsto [0,1] \subset \mathbb{R} \\
\text{General} & \mathbb{R} \arrow[r, "p_X"] & {[0,1]} & p_X : \mathbb{R} \to [0,1] \subset \mathbb{R}
\end{tikzcd}
$$

- We often say X is distributed with respect to p_X.
- What if X_S is continuous ?
  - $p_X(x) = P(X = x)$ should be zero for continuous random variables
  - We will define a density function (pdf) instead.

Example 3.2.4(Coin tosses continued).

- X: the number of Heads. The PMF of X is the function p_X given by

$$\begin{aligned} p_X(0) &= P(X=0) = 1/4, \\ p_X(1) &= P(X=1) = 1/2, \\ p_X(2) &= P(X=2) = 1/4, \end{aligned}$$

  and $p_X(x) = 0$ for all other values of $x \in \mathbb{R}$

- Y: the number of Tails.

$$\begin{aligned} P(Y=y) &= P(2-X=y) = P(X=2-y) = p_X(2-y) \\ p_Y(0) &= P(Y=0) = 1/4, \\ p_Y(1) &= P(Y=1) = 1/2, \\ p_Y(2) &= P(Y=2) = 1/4, \end{aligned}$$

  and $p_Y(\mathrm{y}) = 0$ for all other values of y.

- I: the indicator of the first toss landing Heads.

$$p_I(0) = P(I=0) = 1/2,$$
$$p_I(1) = P(I=1) = 1/2,$$

  and $p_I(x) = 0$ for all other values of x.

Properties of PMFs

Theorem 3.2.7(Valid PMFs). Let X be a discrete r.v. with $X_S = \{x_1, x_2, ...\}$. The PMF p_X of X must satisfy the following two criteria:

- Nonnegative: $p_X(x) > 0$ if $x = x_j$ for some j, and $p_X(x) = 0$ otherwise.
- Sums to 1: $\sum_{j=i}^{\infty} p_X(x_j) = 1$.

Theorem Let $A \subset \mathbb{R}$, then the probability of the event $X^{-1}(A)$ is written as $P(X \in A)$ and given as

$$P(X \in A) = \sum_{x \in A} p_X(x)$$

Bernoulli distribution

Definition 3.3.1(Bernoulli distribution).

An r.v. X is said to have the *Bernoulli distribution* with parameter p

- $X_S = \{0, 1\}$ with $P(X = 1) = p$ and $P(X = 0) = 1 - p$ $(0 < p < 1)$.
- We write this as $X \sim \text{Bern}(p)$. ("$\sim$" is read "is distributed as")
- Any r.v. that has two possible values (say 0 and 1) is a Bern(p) distribution.
- "1" is reffered as "success", "0" is reffered as "failure".

Definition 3.3.2 (Indicator random variable).

The *indicator random variable* of an event A, denoted by I_A, is the r.v. which equals 1 if A occurs ("success") and 0 otherwise ("failure").

$$I_A(s) = \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}$$

- Note that $I_A \sim \text{Bern}(p)$ with $p = P(A)$.

Binomial distribution

- Bernoulli trial: An experiment that has two results ("success" and "failure")
- Binomial distribution: Let X be the number of successes when n independent Bernoulli trials (with $P(\text{"success"}) = p$) are performed. The distribution of X is called the *Binomial distribution* and denoted by

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

Theorem 3.3.5(Binomial PMF).

If $X \sim \mathrm{Bin}(n, p)$, then the PMF of X is

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ for } k = 0, 1, ..., n \ (P(X = k) = 0, \text{ otherwise})$$

Theorem 3.3.7 Let $X \sim \mathrm{Bin}(n, p)$, and $q = 1 - p$. Then $n - X \ \sim \mathrm{Bin}(n, q)$.

Discrete Uniform

- Discrete Uniform Distribution: Let C be a finite, nonempty set of numbers. Let X be a number in C chosen by uniformly at random (i.e., all values in C are equally likely). X is said to have the *Discrete Uniform distribution* with parameter C, denoted by

$$X \sim \mathrm{DUnif}(C)$$

- The PMF of X is

$$P(X = x) = \frac{1}{|C|}$$

- For a $A \subset C$

$$P(X \in A) = \frac{|A|}{|C|}$$

Example 3.5.2(Random slips of paper). There are 100 slips of paper, numbered from 1 to 100, in a hat and five of the slips are drawn, one at a time.

First consider random sampling with *replacement (with equal probabilities).*

(a) What is the distribution of how many of the drawn slips have a number ≥ 80 ?

$$\mathrm{Bin}(5, 21/100)$$

(b) What is the distribution of the value of the jth draw (for $1 \leq j \leq 5$)?

$$X_j \sim \mathrm{DUnif}(\{1, \cdots, 100\})$$

(c) What is the probability that the number 100 is drawn at least once?

$$P(X_j = 100 \text{ for at least one } j) = 1 - P(X_1 \neq 100, ..., X_5 \neq 100) = 1 - (99/100)^5 \approx 0.049$$

Example 3.5.2(Random slips of paper). There are 100 slips of paper, numbered from 1 to 100, in a hat and five of the slips are drawn, one at a time.

Second consider random sampling without *replacement (with equal probabilities).*

(d) What is the distribution of the value of the jth draw (for $1 \leq j \leq 5$)?

$$Y_j \sim \text{DUnif}(\{1, \cdots, 100\})$$

(e) What is the probability that the number 100 is drawn at least once?

The events $Y_1 = 100, ..., Y_5 = 100$ are disjoint since we are now sampling without replacement, so

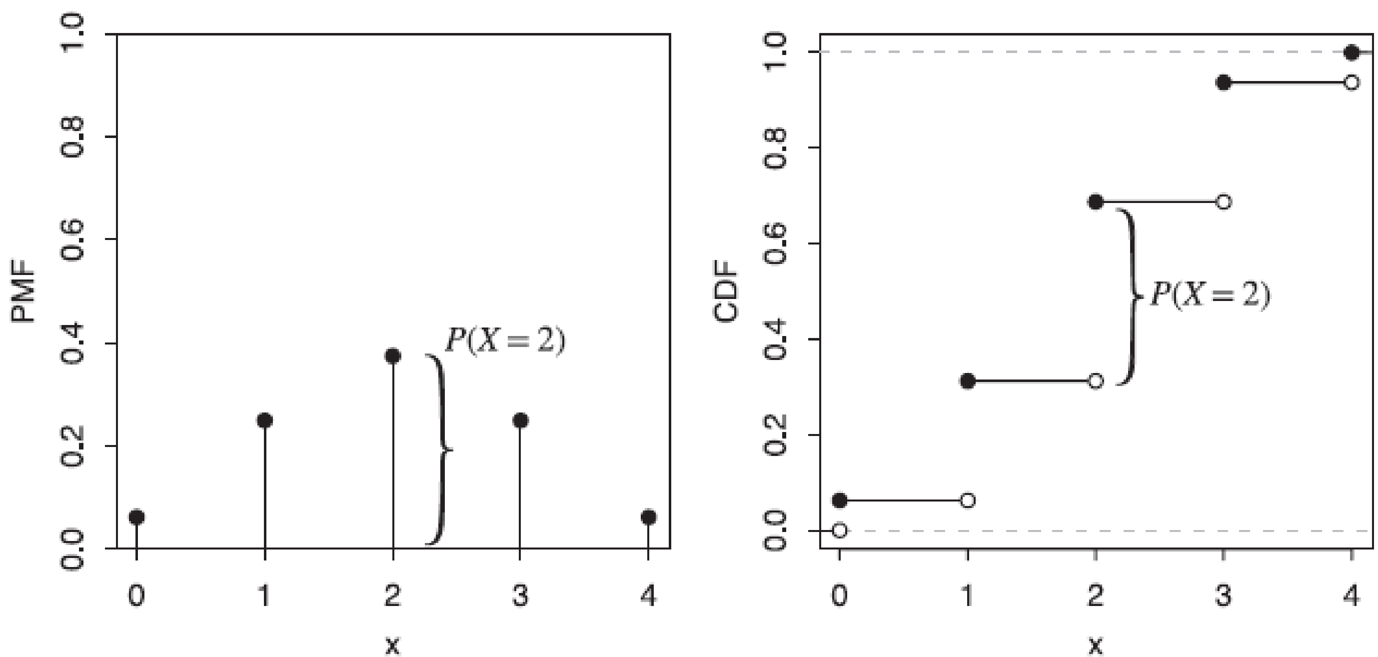
$$P\left(\dot{\bigcup}_{j=1}^{5}(Y_j = 100)\right) = \sum_{j=1}^{5} P(Y_j = 100) = 0.05.$$

Cumulative distribution functions

Definition 3.6.1. The *cumulative distribution function* (CDF) of an r.v. X is the function F_X given by

$$F_X(x) = P(X \le x).$$

Example 3.6.2. Let $X \sim \text{Bin}(4,1/2)$. The PMF and CDF of X are



Functions of random variables

Definition 3.7.1 (Function of an r.v.). For an experiment with sample space S, an r.v. X, and a function $g : \mathbb{R} \to \mathbb{R}$,

$$Y := g(X)$$

is the r.v. that maps s to $g(X(s))$ for all $s \in S$.

Example 3.7.2 (Random walk). A particle moves n steps on a number line, starting from 0. At each step it moves $+1$ or -1 with equal probabilities. Let Y be the position after n steps. Find the PMF of Y.

Solution: Each step is a Bernoulli trial. The number of steps the particle takes "+1", say X, is a Bin(n,1/2) random variable. If $X = j$, then the particle's position is $j - (n - j) = 2j - n$. Let Y be the position of the particle at the step n. Then $Y = 2X - n$ The PMF of Y is

$$P(Y = k) = P(2X - n = k) = P(X = (n + k)/2) = \binom{n}{\frac{n+k}{2}} \left(\frac{1}{2}\right)^n$$

if k is an integer between $-n$ and n (inclusive) such that $n + k$ is an even number (otherwise $P(Y = k) = 0$).

Theorem 3.7.3 (PMF of $g(X)$). Let X be a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$. The PMF of $Y = g(X)$ is

$$P(Y = y) = P(g(X) = y) = \sum_{x:g(x)=y} P(X = x),$$

If g is invertible, $P(Y = y) = P(X = g^{-1}(y))$

Example 3.7.4 In the random walk, let D be the particle's distance from the origin after n steps. Assume that n is even. Find the PMF of D.

Solution: We can write $D = |Y|$; this is a function of Y, but it isn't one-to-one. The event $D = 0$ is the same as the event $Y = 0$. For $k = 2, 4, ..., n$, the event $D = k$ is the same as the event $\{Y = k\} \cup \{Y = -k\}$. So the PMF of D is

$$\begin{aligned} P(D = 0) &= \binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n \\ P(D = k) &= P(Y = k) + P(Y = -k) = 2\binom{n}{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n \text{ for } k = 2, 4, ..., n. \end{aligned}$$

Functions of two random variables

Definition 3.7.5 (Function of two r.v.s). Given an experiment with sample space S, if X and Y are r.v.s that map $s \in S$ to $X(s)$ and $Y(s)$ respectively, then for a function $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $Z := g(X, Y)$ is the r.v. that maps s to $g(X(s), Y(s))$.

Example 3.7.6 (Maximum of two dice rolls). We roll two fair 6-sided dice. Let X be the number on the first dice and Y the number on the second dice. Let $Z = \max(X, Y)$

s	X	Y	$\max(X, Y)$
(1,2)	1	2	2
(1,6)	1	6	6
(2,5)	2	5	5
(3,1)	3	1	3
(4,3)	4	3	4
(5,4)	5	4	5
(6,6)	6	6	6

$$\begin{aligned} P(\max(X, Y) = 5) &= P(X = 5, Y \leq 4) + P(X \leq 4, Y = 5) + P(X = 5, Y = 5) \\ &= 2P(X = 5, Y \leq 4) + 1/36 = 2(4/36) + 1/36 = 9/36. \end{aligned}$$

Independence of random variables

Definition 3.8.1 (Independence of two r.v.s). Random variables X and Y are said to be *independent* if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y),$$

for all $x, y \in \mathbb{R}$. In the discrete case, this is equivalent to the condition

$$P(X = x, Y = y) = P(X = x)P(Y = y),$$

for all $x, y \in \mathbb{R}$.

Definition 3.8.2 (Independence of multiple r.v.s). Random variables $X_1, ..., X_n$ are *independent* if

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1) \cdots P(X_n \leq x_n),$$

for all $x_1, ..., x_n \in \mathbb{R}$.

For infinitely many r.v.x, we say that they are independent if every finite subset of the r.v.s is independent.

Example 3.8.4. In a roll of two fair dice, if X is the number on the first dice and Y is the number on the second dice, then

- X and Y are independent

$$P(X = i, Y = j) = P(X = i)P(Y = j) = \frac{1}{36}$$

- $X + Y$ is not independent of $X - Y$.

$$0 = P(X + Y = 12, X - Y = 1) \neq P(X + Y = 12)P(X - Y = 1) = \frac{1}{36} \cdot \frac{5}{36},$$

i.i.d.: Independent and Identically distributed

Definition 3.8.5 (i.i.d.). Random variables that are independent and have the same distribution are called *independent and identically distributed*, or *i.i.d.* for short.

- When $X_1, X_2, X_3, \cdots, X_n$ are said to be i.i.d., all X_i's have the same distribution and independent.
- Example 1: Let X be the result of a dice roll, and let Y be the result of a second, independent dice roll. Then, X and Y are i.i.d.
- Example 2: Let $\{X_i\}$ be an i.i.d. sequence with $X_i \sim \text{Bern}(p)$

Theorem 3.8.7.

If X $\sim$Bin(n, p), viewed as the number of successes in n independent Bernoulli trials with success probability p, then we can write

$$X = X_1 + \cdots + X_n$$

where the X_i are i.i.d. Bern(p).

Theorem 3.8.8.

If $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$, and X is independent of Y, then

$$X + Y \sim \text{Bin}(n + m, p)$$

Conditional PMF

Definition 3.8.10 (Conditional PMF). For any discrete r.v.s X and Z, the function

$$P(X = x|Z = z),$$

when considered as a function of x for fixed z, is called the *conditional PMF of X given $Z = z$*.

Example. If $X \sim \text{Bin}(n, p), Y \sim \text{Bin}(m, p)$, and X and Y are independent, find the conditional PMF of X given $X + Y = r$

$$\begin{aligned} P(X = x|X + Y = r) &= \frac{P(X + Y = r \mid X = x)P(X = x)}{P(X + Y = r)} \\ &= \frac{P(Y = r - x)P(X = x)}{P(X + Y = r)} \\ &= \frac{\binom{m}{r-x}p^{r-x}(1-p)^{m-r+x}\binom{n}{x}p^{x}(1-p)^{n-x}}{\binom{n+m}{r}p^{r}(1-p)^{n+m-r}} \\ &= \frac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}} \end{aligned}$$

where $P(X + Y = r \mid X = x) = P(x + Y = r|X = x) = P(x + Y = r)$ by the independence of X and Y.

The conditional PMF above is called Hyper-geometric, $\text{HGeom}(n, m, r)$.