Unit 3: Discrete Random Variables

Random variables

Definition 3.1.2 Random Variable

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

It is common to denote random variables by capital letters

Example (Random Variable)

Consider experiment where we toss a coin twice, then the Sample space $S = \{HH, HT, TH, TT\}$. We can set these random variables:

- Let X be number of Heads. Then:

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

- Let Y be number of Tails: Then:

$$Y = 2 - X \quad \text{or} \quad Y(s) = 2 - X(s)$$

- Let I be 1 if the first toss is Heads and 0 otherwise, then:

$$I(HH) = 1 \quad , \quad I(HT) = 1 \quad , \quad I(TH) = I(TT) = 0$$

We can encode the sample space as $\{(1,1), (1,0), (0,1), (0,0)\}$ where 1 is Heads and 0 is Tails. From that we can also give formulas to X, Y, and I:

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

1

Conclusion:

Imagine we have a sample space S with a punch of pebbles, each pebble represent the probability of an event. Random variables simply label each pebble with a number (the value of the number depends on how we define our r.v). This ~~helps us~~ provides us the numerical summaries of the experiment without having to explicitly calculate the probabilities.

Discrete Random Variables

Definition 3.2.1 Discrete random variables

A random variable X is discrete if there is a finite list of values $a_1, a_2, \ldots, a_n$ or infinite list $a_1, a_2, \ldots$ such that:

$$P(X = a_j \text{ for some } j) = 1$$

If X is discrete, then the finite or infinite set of values x such that $P(X = x) > 0$ is called the support of X.

The support x of r.v. X can be: ~~integers~~,

- integers if X is discrete r.v.
- real value if X is continuous r.v.
- hybrid of integers and real value

Given a random variable, we would like to describe its behavior using the language of probability. For example, one question is what is the probability of r.v. $X > 0$?

The distribution of the r.v. will provide answers to these questions. The most natural way to see the distribution of a r.v. is probability mass function

2

Definition 3.2.2 (Probability Mass Function)

The probability mass function (PMF) of a discrete r.v. X is the function p_X given by:

$p_X(x) = P(X = x)$ where $\begin{cases} P(X=x) > 0, \text{ if } x \text{ is support of } X \\ P(X=x) = 0 \quad \text{otherwise} \end{cases}$

Example 3.2.3 (PMF)

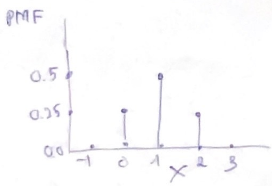
Consider previous example from random variable:

- r.v. X:

$p_X(0) = P(X=0) = 1/4$

$p_X(1) = P(X=1) = 1/2$

$p_X(2) = P(X=2) = 1/4$

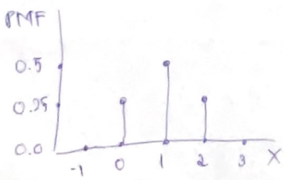


- r.v. Y: $(P\,Y = 2 - X)$

$p_Y(0) = P(Y=0) = 1/4$

$p_Y(1) = P(Y=1) = 1/2$

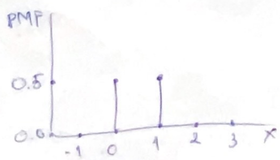
$p_Y(2) = P(Y=2) = 1/4$



- r.v. I:

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

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



Theorem 3.2.5 (Valid PMFs)

The PMF p_X of r.v. X must satisfy the following criterias:

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

Another way to show distribution of a r.v. is Poisson

Example 3.2.6 (Poisson Distribution)

An r.v. X has the Poisson Distribution with parameter $\lambda\,(\lambda > 0)$ if the PMF of X is:

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots$$

We write this as $X \sim \text{Pois}(\lambda)$

Poisson Distribution is widely used in statistics, and a common choice for data that counts the number of occurences of some kind. (More detail in unit 4,5)

Bernoulli and Binomial

Definition 3.3.1 (Bernnoulli Distribution)

An r.v. X is said to be Bernoulli Distribution with parameter p if $P(X=1) = p$ and $P(X=0) = 1-p$ $(0 < p < 1)$

This is denoted as $X \sim \text{Bern}(p)$

Definition 3.3.2 (Indicator random variable)

The indicator random variable of event A is the r.v. which equals 1 if A occurs and 0 otherwise, denoted by I_A or $I(A)$. Also, naturally: $I_A \sim \text{Bern}(p)$ with $p = P(A)$

~~Bin~~. Bernoulli trial is an experiment that can result in either a "success" or "failure", the parameter p is often called the success probability of the $\text{Bern}(p)$ distribution.

Binomial distribution is when a sequence ~~a~~ n of independent Bernoulli trials are performed, each with the same success probability p. Let X be the number of successes, the distribution of X is called Binomial Distribution with parameters n and p, denoted as $X \sim Bin(n,p)$

Note: $Bern(p)$ is the same distribution as $Bin(n,p)$, the Bernoulli is a special case of the Binomial

Theorem 3.3.5 (Binomial PMF)

If $X \sim Bin(n,k)$, then the PMF of X is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

for $k = 0, 1, \dots, n$ (and $P(X=k) = 0$ otherwise)

When X is Binomial, then $n-X$ is also Binomial

Theorem 3.3.7

Let $X \sim Bin(n,p)$, and $q = 1-p$. Then $n-X \sim Bin(n,q)$

Hypergeometric

Consider a story with an urn of w white balls and b black balls, then drawing n balls out of the urn without replacement. Let X be the number of white balls in the sample. Then X is said to have Hypergeometric Distribution with paremeters w, b, n. Denoted as $X \sim HGeom(w, b, n)$

Theorem 3.4.3 (Hypergeometric PMF)

If $X \sim HGeom(w, b, n)$, then the PMF of X is:

$$P(X=k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

for integers k satisfy $\begin{cases} 0 \le k \le w \\ 0 \le n-k \le b \end{cases}$, $P(X=k) = 0$ otherwise

What kind of problem required HGeom ?

The essential structure of Hypergeometric story is that items in a population are classified using 2 sets of tags. In the urn story, the 2 sets of tags are:

- The balls are black or white
- The balls are sampled or not sampled

Furthermore, at least 1 set of tag is assigned randomly. In the urn story, the balls are sampled randomly

Differences between Binomial and Hypergeometric

Warning 3.4.5

- Both are discrete distributions taking on integer values between 0 and n
- Both can be interpreted as the number of successes ~~of~~ in n Bernoulli trials
- The difference is that the Bernoulli trials involved in Binomial are independent, while the ones involved in Hypergeometric are dependent.

Discrete Uniform

Story 3.5.1 (Discrete Uniform distribution)

Let C be a finite, nonempty set of numbers. Choose 1 of these numbers uniformly at random, call the chosen number X, then X is said to have the Discrete Uniform distribution with parameter C, denoted as $X \sim \text{DUnif}(C)$. The PMF of $X \sim \text{DUnif}(C)$ is:

$$P(X=x) = \frac{1}{|C|}, \quad \text{for } x \in C \quad (\text{and } 0 \text{ otherwise})$$

For any $A \subseteq C$, we have:

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

This is basically a counting problems

Example 3.5.2

There are 100 slips of paper in a hat, labeled from $1, \dots, 100$. 5 slips are drawn. Consider:

Random sampling with replacement	Random sampling without replacement
a) What is the distribution of how many of drawn slips have a value of at least 80 written on them?	
. The distribution is $\text{Bin}(5, 0.21)$	. The distribution is $\text{HGeom}(21, 79, 5)$
b) What is the distribution of the value of the j-th draw $(1 \leq j \leq 5)$?	
. Let X_j be the value of the jth draw, By symmetry, $X_j \sim \text{DUnif}(1,2,\dots 100)$	. Let Y_j be the value of jth draw. By symmetry, $Y_j \sim \text{DUnif}(1,2,\dots 100)$
c) What is the probability that the number 100 is drawn at least once?	
. $P(X_j = 100 \text{ at least once}_j)$ $= 1 - P(X_1 \neq 100, \dots X_5 \neq 100)$ $= 1 - (99/100)^5 \approx 0.049$	. $P(Y_j = 100 \text{ for some } j)$ $= P(Y_1 = 100) + \dots + P(Y_5 = 100)$ $= 0.05$

7

Cumulative distribution functions

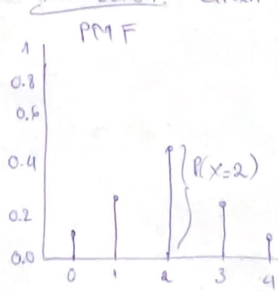
Definition 3.6.1 and Theorem 3.6.4 (Valid CDFs)

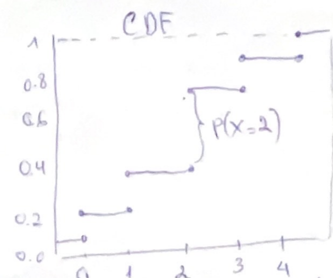
The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$.

Any CDF F has these properties:

- Increasing: if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$
- Right-continuous: CDF continuous except having some jumps, for any a, we have $F(a) = \lim_{x \to a} F(x)$
- Convergence to 0 and 1 in the limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

Visualization: Given $\text{Bin}(4, 1/2)$





. The height of a jump in CDF at x equals the value of PMF at x

8

How many ways can we express the distribution of a r.v. ?

3 ways: PMF, CDF and story.

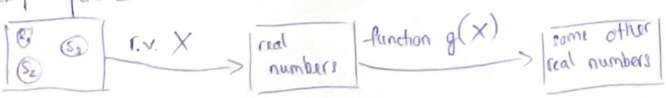
. PMF are generally easier to work with since CDF require summation (compare to CDF).

. Story of Binomial and Hypergeometric are usually used to derive the PMF, since they are more intuitive to proof than go straight to PMF calculation.

Functions of random variables

A function of random variables is just another random variable. If we think about it, a r.v. is mapping elements in a sample space to real numbers, and a function of r.v. then maps those real numbers to another set of real numbers

Sample space S



~~Warning~~ **Definition 3.7.1 (Function of an r.v.)**

For an experiment with sample space S, an r.v. X, and a function $g: \mathbb{R} \to \mathbb{R}$, $g(X)$ is the r.v. that maps s to $g(X(s))$ for all $s \in S$.

Warning 3.7.3 (category errors)

Don't mistaken distribution to random variable. We should think the distribution is a summary or blueprint that describe the random variable. Two different random variables can have the same distribution.

Independence of random variables

Just like notion of independence of events, we have the notion of independence of random variables. Intuitively, if 2 r.v.s are independent, knowing the value of X does not give information about the value of Y, and vice versa.

Definition 3.8.1 (Independence of 2 r.v.s)

Random variables X and Y are said to be independent if:

$$P(X \le x, Y \le y) = P(X \le x) \cdot P(Y \le y),$$

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

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

for all x, y with $\begin{cases} x & \text{supports of } X \\ y & \text{support of } Y \end{cases}$

Also, if X and Y are independent then any functions of X is independent of any function of Y, which is i.i.d

Definition 3.8.3 (i.i.d)

Often, two r.v.s that are independent and have the same distribution is called "independent and identically distributed", or i.i.d for short

Taking the ~~sum~~ sum of i.i.d Bernoulli r.v.s, we can form the story of Binomial distribution in algebraic form.

Theorem 3.8.4 (Binomial story from Bernoulli i.i.d)

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 $X = X_1 + \dots + X_n$ where X_i are i.i.d. $Bern(p)$

Also, the sum of independent Binomial r.v.s with the same probability is another Binomial r.v

Theorem 3.8.5 (Sum of Binomial r.v.s)

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

Challenge: Prove above theorem using LOTP, Representation, Story.

Just like independence of events, we also can state conditional independence of r.v.s

Definition 3.8.6 (Conditional independence of r.v.s)

Random Variables X and Y are conditionally independent given r.v Z if for all $x, y \in \mathbb{R}$ and all z support Z.

$$P(X \le x, Y \le y \mid Z = z) = P(X \le x \mid Z = z) \cdot P(Y \le y \mid Z = z)$$

For discrete r.v.s, an equivalent definition is.

$$P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)$$

Definition 3.8.7 (Conditional PMF)

For any discrete r.v.s X and Z, the function $P(X = x \mid Z = z)$, considered as function of x for fixed z, is called the conditional PMF of X given $Z = z$.

And just like independence of events, independence of r.v.s does not imply conditional independence, and vice versa

Example 3.8.8 Independent r.v.s doesn't imply conditional independence

Consider a game of matching pennies, have 2 player A and B. If pennies match, A wins, otherwis B wins. Let:

$$\begin{cases} X = 1 & \text{if A lands Heads} \\ X = -1 & \text{if A lands Tails} \\ Y = 1 & \text{if B lands Heads} \\ Y = -1 & \text{if B lands Tails} \\ Z = XY \begin{cases} = 1, & \text{A wins} \\ = -1, & \text{B wins} \end{cases} \end{cases}$$

From this we can see that X and Y are independent but given $Z = 1$, we know that pennies match. So X and Y are conditional ~~in~~ dependent

Example 3.8.9 Conditional independence doesn't imply independence

Suppose you gonna play 2 games of tennis (X and Y) with either one of the twins. Your chance of winning against each twin is $1/2$ and $3/4$. And you can't tell who you are playing against after 2 games. Let Z be the indicator of playing against the twin with $1/2$ chance of winning, so:

$\begin{cases} \text{Conditinon on } Z=1,\ X \text{ and } Y \text{ are i.i.d Bern}(1/2) \\ \text{Condition on } Z=0,\ X \text{ and } Y \text{ are i.i.d Bern}(3/4) \end{cases}$

So X and Y are conditionally independent. But unfortunately, without condition Z, by observing X and Y, we can tell who we are playing against, which means X and Y are not independent.

Practice Problems

Problem 1

Benford's law state that in a very large variety of real-life data sets, the first digit approx follows a particular distribution with about 30% chance of a 1, an 18% chance of a 2, in general

$$P(D=j) = \log_{10}\left(\frac{j+1}{j}\right), \quad \text{for } j \in \{1, 2, \dots, 9\}$$

where D is the first digit of randomly chosen element.

Is this a valid PMF?

~~Check sum of~~ A PMF is $\begin{cases} \text{. non negative} \\ \text{. sum to } 1 \end{cases}$

We can see that $P(D=j)$ is nonnegative.

Now its sum is: $\sum_{j=1}^{9} \log_{10}\left(\frac{j+1}{j}\right) = \sum_{j=1}^{9} \left(\log_{10}(j+1) - \log_{10}(j)\right)$

We can see that all terms cancel each others except $(\log_{10} 10 - \log_{10} 1) = 1$. Therefore, $P(D=j)$ sum to 1.

$\Rightarrow$ This is a valid PMF

Problem 2

In a chess tornament, 10 games are played, independently.

~~Eea~~ Each game either $\begin{cases} \text{win for 1 player with } 0.4 \text{ probability} \\ \text{draw with } 0.6 \text{ probability} \end{cases}$

What is the probability that exactly 5 games end in a draw?

Let G be the number of games end in a draw. ~~P~~ $G \sim Bin(10, 0.6)$

$\Rightarrow P(G=5) = \binom{10}{5} 0.6^5 \, 0.4^{(10-5)} = 0.2007$

Problem 3

There are 2 coins, one with p1 probability of Heads and the other with p2 probability of Heads. One coin is chosen and flip $n \geq 2$ times. Let X be the numbers of Heads.

a) Which is the PMF of X?

There are 2 ~~sce~~ scenarios, depends on which coin is chosen:

$\begin{cases} P(X=k \mid \text{Coin } p_1) = \binom{n}{k} p_1^k (1-p_1)^{n-k} \\ P(X=k \mid \text{Coin } p_2) = \binom{n}{k} p_2^k (1-p_2)^{n-k} \end{cases}$

By LOTP: $P(X=k) = P(X=k \mid \text{Coin } p_1).P(\text{Coin } p_1) + P(X=k \mid \text{Coin } p_2).P(\text{Coin } p_2)$

$= \frac{1}{2}\binom{n}{k} p_1^k (1-p_1)^{n-k} + \frac{1}{2}\binom{n}{k} p_2^k (1-p_2)^{n-k}$

b) Is the distribution of X Binomial if $p_1 = p_2$?

Yes. $P(X=k)$ would reduce to $\binom{n}{k} p_1^k (1-p_1)^{n-k}$ or $\binom{n}{k} p_2^{(k)} (1-p_2)^{n-k}$

c) Is the distribution of X Binomial if $p_1 \neq p_2$?

No. A mixture of Binomials is not Binomial, unless $p_1 = p_2$

Also, the tosses are not independent. If n is large, the earlier tosses will give information about the later tosses.

Problem 4

There are n eggs, each hatch a chick with probability p. Each of these chicks survives with probability r, independently.

Let H be number of eggs that hatch
X be number of eggs that hatch and survive

Find the distribution of H and X.

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

Homework problems

Problem 1

Let X be the number of purchases that a customer will make on the online site for a certain company. PMF of X is:

$$P(X=k) = e^{-\lambda}\lambda^k / k! \quad \text{for } k = 0, 1, 2, \dots$$

a) Find $P(X \geq 1)$ and $P(X \geq 2)$ without summing infinite series

Taking complement:

- $P(X \geq 1) = 1 - P(X=0) = 1 - e^{-\lambda}$
- $P(X \geq 2) = 1 - P(X \leq 1) = 1 - e^{-\lambda} - \lambda e^{-\lambda}$

b) Suppose the company only knows about ppl who have made at least one purchase. What is the PMF (conditional) of X given $X \geq 1$?

$$P(X=k \mid X \geq 1) = \frac{P(X=k)}{P(X \geq 1)} = \frac{e^{-\lambda}\lambda^k}{k!\,(1-e^{-\lambda})} \quad \text{for } k = 1, 2, \dots$$

$$\left\langle P(A \mid B) = \frac{P(A \cap B)}{P(B)} \right\rangle$$

Problem 2

A book has n typos. There are 2 proofreaders, Prue and Frida. Prue catches typo with probability p_1 and misses with $q_1 = 1 - p_1$. Frida catches typo with probability p_2 and misses with $q_2 = 1 - p_2$.

Let X_1 be typos catched by Prue
X_2 be typos catched by Frida
X be typos catched by either or both Prue and Frida

a) Find distribution of X:

$$X \sim \text{Bin}(n, 1 - q_1 q_2)$$

Note: Here we have only 1 probability of "at least typos catched by Prue or Frida (using complement)". Let say if we have 2 ~~probabibi~~ probabilities, then Binomial won't be possible because "mixture of Binomials are not necessarily Binomial".

b) Assume $p_1 = p_2$. Find distribution (conditional) of X_1 given that $X_1 + X_2 = t$

~~X_1~~ Let $p = p_1 = p_2$ and $T = X_1 + X_2 \sim \text{Bin}(2n, p)$. Then:

$$P(X_1 = k \mid T = t) = \frac{P(T = t \mid X_1 = k).P(X_1 = k)}{P(T = t)}$$

$$= \frac{\binom{n}{t-k} p^{t-k} q^{n-t+k} \cdot \binom{n}{k} p^k q^{n-k}}{\binom{2n}{t} p^t q^{2n-t}}$$

$$= \frac{\binom{n}{t-k}\binom{n}{k}}{\binom{2n}{t}} \quad \text{for } k \in \{0, 1, \dots, t\}$$

$\Rightarrow \text{HGeom}(n, n, t)$

Problem 3

People arriving at the party one at a time. Let X be the number of people needed to obtain a birthday match. For example, $X = 10$ would mean that the first 9 people arrive all have different birthdays. Find $P(X=3 \cup X=4)$

. The support of X is $\{2, 3, \dots, 366\}$

Now: $P(X=2) = \frac{1}{365}$ (since there is 1 match)

$P(X=3) = \frac{364}{365} \cdot \frac{2}{365}$ (364 no match, 2 match)

$P(X=4) = \frac{364}{365} \cdot \frac{363}{365} \cdot \frac{3}{365}$ (364 and 363 no match, 3 match)

Lets generalize this.

$$P(X=k) = P(X > k-1 \text{ and } X=k)$$

$$= \frac{365 . 364 \dots (365-k+2)}{365^{k-1}} \cdot \frac{k-1}{365}$$

$$= \frac{(k-1) . 364 . 363 \dots (365-k+2)}{365^{k-1}}$$

Therefore:

$$P(X=3 \text{ or } P=4) = P(X=3) + P(X=4) \approx 0.0136$$

Problem 4

Let X be the number of Heads in 10 fair coin tosses.

a) Find the conditional PMF of X, given the first 2 tosses land Heads

Let $\begin{cases} X_2 & \text{be the number of Heads in the first 2 tosses} \\ X_8 & \text{be the number of Heads in the first 8 tosses} \end{cases}$

$$P(X=k \mid X_2=2) = P(X_2 + X_8 = k \mid X_2 = 2)$$

$$= P(X_8 = k - X_2 \mid X_2 = 2)$$

$$= P(X_8 = k-2)$$

$$= \binom{8}{k-2} \cdot \left(\frac{1}{2}\right)^{k-2} \cdot \left(\frac{1}{2}\right)^{8-k+2}$$

$$= \frac{1}{256}\binom{8}{k-2} \quad \text{for } k = 2, 3, \dots, 10$$

b) Find the conditional PMF of X, given at least 2 tosses land Heads

$$P(X=k \mid X \geq 2) = \frac{P(X=k, X \geq 2)}{P(X \geq 2)}$$

$$= \frac{P(X=k)}{1 - P(X=0) - P(X=1)}$$

$$= \frac{\binom{10}{k}\left(\frac{1}{2}\right)^{10}}{1 - \left(\frac{1}{2}\right)^{10} - 10\left(\frac{1}{2}\right)^{10}}$$

$$= \frac{1}{1013}\binom{10}{k} \quad \text{for } k = 2, 3, \dots, 10$$

Note:

When to model data as Binomial and when to model data as Poisson ?

	Binomial	Poisson
Number of trials	. Fix number of independant trials with binary outcomes	. Number of trials is large and indefinite . The concept of "success" depends on occurences in a fixed interval.
Probability of Success	. Remains constant from trial to trial	. No concept of "success", it models the occurrences ~~of~~ in an ~~intera~~ interval.
Rare events	. If . probability of success (p) is small . number of trials (n) is large $\Rightarrow$ Bin (n, p) approaches Pois (λ) with $\lambda = n \cdot p$ Therefore, when n is large and p is small, Poisson is a more appropriate model and vice versa.	
Nature of data	. The model is discrete and more suitable for counting the number of "successes" in a fixed trials.	. The model is also discrete, more suitable to count number of "rare events" (small p) over fixed interval.