$$S_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$$

$$\lim_{n \to \infty} P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right)$$

Lecture 3: Random Variables

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

October 15, 2024

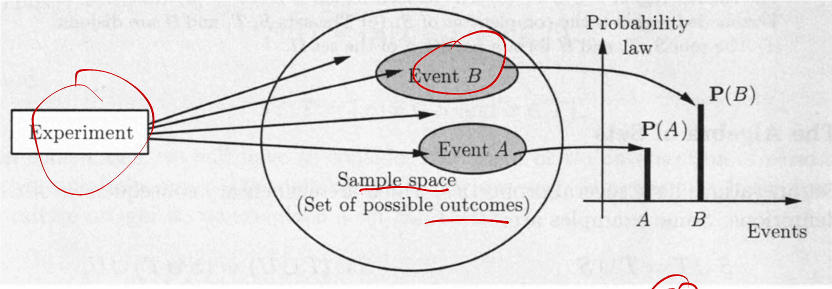
Outline

Outline

Probabilistic Model Motivation

1°. $\cup, \cap, \bar{\ }, \setminus \ldots$

$+, -, \times, \div$



Motivation 2: event $\to \mathbb{R}$

Language $\to$ Numerical value

Definition of Random Variables

$X: S \to R$

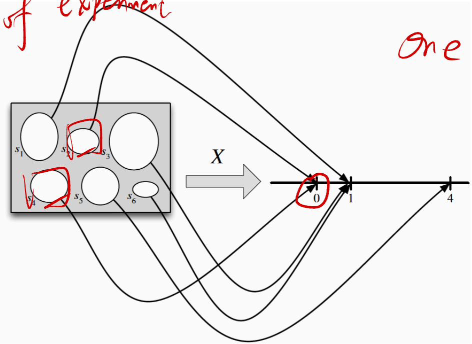
Definition

deterministic

Given an experiment with sample space S, a *random variable* (r.v.) is a function from the sample space S to the real numbers R. It is common, but not required, to denote random variables by capital letters.

X: Summary of experiment

~~One-one mapping~~



Example: Coin Tosses

$S \to \mathbb{R}$

① $X(HH)=2$; $X(HT)=1$; $X(TH)=1$;

$X(TT)=0$; $\{0,1,2\}$

Consider an experiment where we toss a fair coin twice. The sample space consists of four possible outcomes: $S = \{HH, HT, TH, TT\}$. Here are some random variables on this space (for practice, you can think up some of your own). Each r.v. is a numerical summary of some aspect of the experiment

② $Y(s) = 2 - X(s)$ sample point s, $s \in S$,

$Y(HH) = 0$

$Y(HH) = 2 - X(HH) = 2 - 2 = 0$

- X: the number of Heads.
- Y: the number of Tails.
- I: equals 1 if the first toss lands Heads and 0 otherwise.

③ $I(HH)=1$; $I(HT)=1$;

$I(TH)=0$; $I(TT)=0$;

Discrete Random Variable

$X = a_j$ $\triangleq$ $\{s : X(s) = a_j\}$

$P(X = a_j)$ event

$= P(\{s : X(s) = a_j\})$ set of sample points $\subseteq S$

Definition

A random variable X is said to be *discrete* if there is a finite list of values $a_1, a_2, \ldots, a_n$ or an infinite list of values $a_1, a_2, \cdots$ such that $P(X = a_j \text{ for some } j) = 1$. If X is a discrete r.v., then the finite or countably infinite set of values x such that $P(X = x) > 0$ is called the *support* of X.

fair coin tossing $\{H, T\}$ X : # of heads.

Support of X is $\{0, 1\}$

$P(X=1) = P(\text{"H"}) = \frac{1}{2} > 0$ $X = 1$ or 0

$P(X=0) = P(\text{"T"}) = \frac{1}{2} > 0$,

Probability Mass Function

$X = x$

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

$= P(\{s : X(s) = x\})$

Definition

The *probability mass function* (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P(X = x)$. Note that this is positive if x is in the support of X, and 0 otherwise.

$X : S \longrightarrow R_X \subseteq R$

$x \in R_X$

Example: Coin Tosses

$\{X=x\} \triangleq \{s : X(s)=x, s\in S\}$

① $X=0 \Leftrightarrow \{s\}=\{TT\}$; $P_X(0)=P(X=0)=P(\text{"TT"})=\frac{1}{4}$.

$X=1 \Leftrightarrow s=HT, TH$; $P_X(1)=P(X=1)=P(\text{"}\{HT,TH\}\text{"})=\frac{1}{2}$;

$X=2 \Leftrightarrow s=HH$; $P_X(2)=P(X=2)=P(\text{"HH"})=\frac{1}{4}$

Consider an experiment where we toss a fair coin twice. The sample space consists of four possible outcomes: $S = \{HH, HT, TH, TT\}$. Here are some random variables on this space (for practice, you can think up some of your own). Each r.v. is a numerical summary of some aspect of the experiment

PMF. $R_X=\{0,1,2\}$ $x\in R_X=\{0,1,2\}$

$P_X(0)=\frac{1}{4}$, $P_X(1)=\frac{1}{2}$, $P_X(2)=\frac{1}{4}$

Support of X = R_X.

- $X(s)$ X: the number of Heads.
- $Y(s)$ Y: the number of Tails.
- $I(s)$ I: equals 1 if the first toss lands Heads and 0 otherwise.

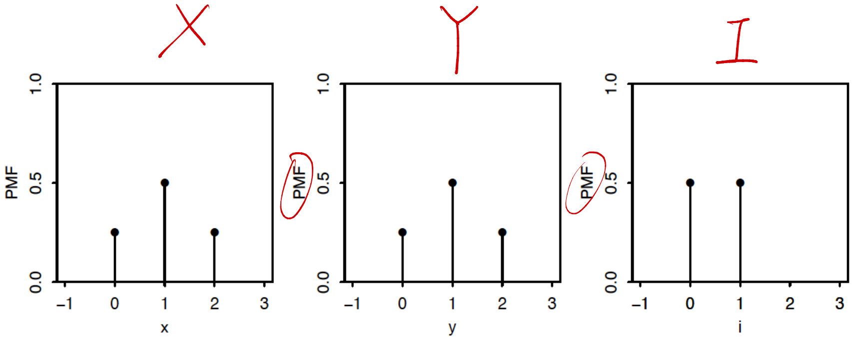
② $P_Y(x)=P_X(x)$, $\forall x\in\{0,1,2\}$

③ $P_I(0)=P(I=0)=P(\{TH,TT\})=\frac{1}{2}$

$P_I(1)=P(I=1)=P(\{HH,HT\})=\frac{1}{2}$

PMF $\Rightarrow$ $P_I(x)=\frac{1}{2}$, $x=0$ or 1.

Example: Coin Tosses



distribution of X

Valid PMFs

Theorem

Let X be a discrete r.v. with support $x_1, x_2, \ldots$ (assume these values are distinct and, for notational simplicity, that the support is countably infinite; the analogous results hold if the support is finite). 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:* $\sum_{j=1}^{\infty} p_X(x_j) = 1$.

Outline

Math: $S \xrightarrow{r.v.} R$. r.v. mapping

Modeling: random Variable.

Bernoulli Distribution

PMF

$P(X=x) = p^{x}(1-p)^{1-x}$, $x = 0$ or 1

Family of Bernoulli distribution

$P(X;\theta) = \theta^{x}(1-\theta)^{1-x}$

$0<\theta<1$

Definition

An r.v. X is said to have the *Bernoulli distribution* with parameter p if $P(X=1)=p$ and $P(X=0)=1-p$, where $0<p<1$. We write this as $X \sim \mathrm{Bern}(p)$. The symbol $\sim$ is read "is distributed as".

Indicator Random Variable

$P(A) = p$

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs.} \\ 0 & \text{otherwise.} \end{cases}$$

Definition

The *indicator random variable* of an event A is the r.v. which equals 1 if A occurs and 0 otherwise. We will denote the indicator r.v. of A by I_A or $I(A)$. Note that $I_A \sim \text{Bern}(p)$ with $p = P(A)$.

indicator function $\neq$ r.v.

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A. \\ 0 & \text{otherwise.} \end{cases}$$

Story: Bernoulli Trial

An experiment that can result in either a "success" or a "failure" (but not both) is called a *Bernoulli trial*. A Bernoulli random variable can be thought of as the *indicator of success* in a Bernoulli trial: it equals 1 if success occurs and 0 if failure occurs in the trial.

Story: Binomial Distribution

$0, 1, \ldots, n$

$X \in [0, n]$

Suppose that n *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 the *Binomial distribution* with parameters n and p. We write $X \sim \mathrm{Bin}(n, p)$ to mean that X has the Binomial distribution with parameters n and p, where n is a positive integer and $0 < p < 1$.

Binomial PMF

1 : success p

0 : failure $1-p$

$\underbrace{1\ 0\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1 \cdots}_{\text{length } n}$ binary sequence

Theorem

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

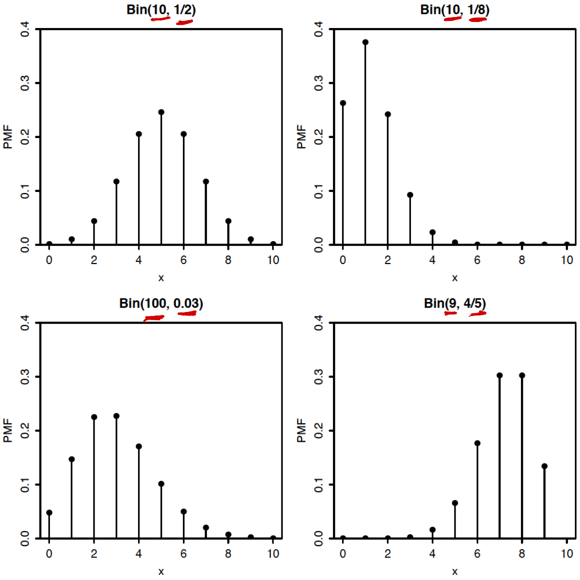
k "1"

$n-k$ "0"

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

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

Binomial PMF



Binomial PMF

Theorem

Let $X \sim \mathrm{Bin}(n, p)$, and $q = 1 - p$ (we often use q to denote the failure probability of a Bernoulli trial). Then $n - X \sim \mathrm{Bin}(n, q)$.

Example: Statistical Multiplexing

X: # of active users at the same time.

user/sender in or out Bernoulli trial.

$X \sim Bin(N, p)$

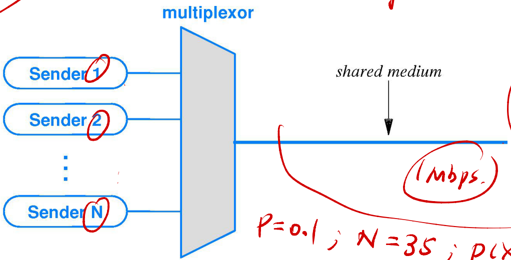
each sender (user) request 100kbps. active prob. p.

No statistical multiplexing

$P(X=k)$

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

$k \in \{0, 1, \dots, N\}$



1Mbps

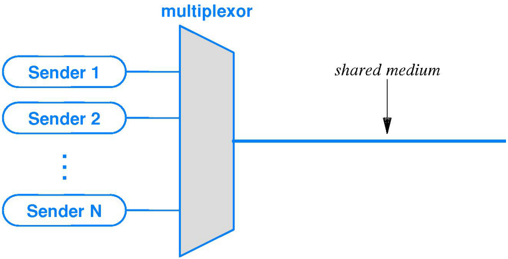
at most 10 users

$p = 0.1$; $N = 35$; $P(X > 10) < 0.0004$

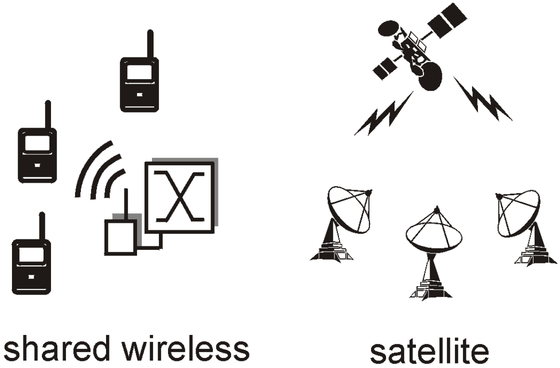
System overload: $P(X > 10) = 1 - P(X \le 10)$

$$= 1 - \sum_{k=0}^{10} \binom{N}{k} p^k (1-p)^{n-k} = 1 - \sum_{k=0}^{10} P(X=k)$$

Example: Statistical Multiplexing



Example: Multiple Access (Aloha Protocol)



Example: Multiple Access (Aloha Protocol)

unit time   time-slot

time-slotted Aloha

X : # of devices transmit simultaneously

$X \sim Bin(N, p)$

- N smart devices sharing a WiFi access point (e.g., in starbucks)
- ≥ 2 devices transmit simultaneously lead to collision
- Aloha Protocol: proposed by Norman Abramson in the later 1960s
- Each device transmits with probability p independently
- What is the transmission rate (the number of successful transmissions per unit time)?

$$f(p) = \frac{P(X=1)\cdot 1}{1} = P(X=1)$$

$$= \binom{N}{1} p^1 (1-p)^{N-1} = N p (1-p)^{N-1}$$

Maximize $f(p)$, $0 \le p \le 1$.

$f'(p) = 0$; $f''(p) < 0$ $\Rightarrow p^* = \frac{1}{N}$

$$\Rightarrow f^*(p) = (1 - \tfrac{1}{N})^{N-1} \xrightarrow{N \to \infty} e^{-1} \approx 0.36$$

Example: Multiple Access (Aloha Protocol)

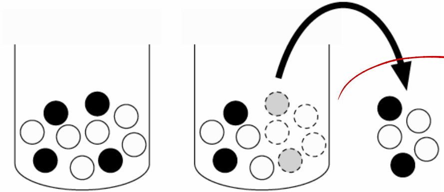
Outline

Urn Model

$\text{Bern}\left(\frac{w}{w+b}\right)$

An urn is filled with w white and b black balls, then drawing n balls out of the urn

- with replacement: $Bin(n, w/(w+b))$ distribution for the number of white balls obtained
- without replacement: Hypergeometric distribution



Story: Hypergeometric Distribution

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

$$\begin{cases} 0 \le k \le w \\ 0 \le n-k \le b \end{cases}$$

$n \le w+b$

Consider an urn with w white balls and b black balls. We draw n balls out of the urn at random without replacement, such that all $\binom{w+b}{n}$ samples are equally likely. Let X be the number of white balls in the sample. Then X is said to have the *Hypergeometric distribution* with parameters w, b, and n; we denote this by $X \sim \mathrm{HGeom}(w, b, n)$.

Hypergeometric PMF

Theorem

If $X \sim \mathrm{HGeom}(w, b, n)$*, then the PMF of* X *is*

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

for integers k *satisfying* $0 \le k \le w$ *and* $0 \le n - k \le b$*, and* $P(X = k) = 0$ *otherwise.*

Identical Distribution

$X \sim$ HGeom(w, b, n)

X: # of white balls in sampled balls.

first sampled tag;

second color tag;

Color tag: white/black

Sampled tag: Yes or No

Theorem

The $\mathrm{HGeom}(w, b, n)$ *and* $\mathrm{HGeom}(n, w + b - n, w)$ *distributions are identical. That is, if* $X \sim \mathrm{HGeom}(w, b, n)$ *and* $Y \sim \mathrm{HGeom}(n, w + b - n, w)$, *then* X *and* Y *have the same distribution.*

Y: first color tag.

second sampled tag;

Outline

Story: Discrete Uniform Distribution

Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X. Then X is said to have the *Discrete Uniform distribution* with parameter C; we denote this by $X \sim \mathrm{DUnif}(C)$.

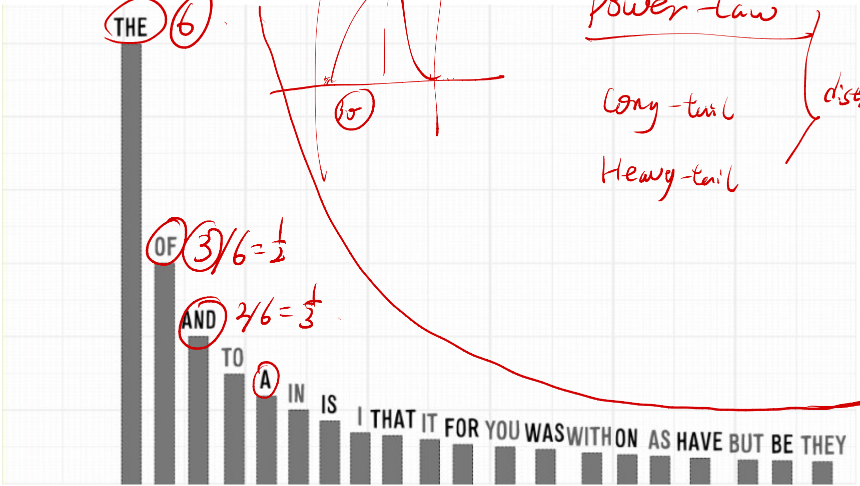
$$P(X = x) = \frac{1}{|C|} \quad (x \in C)$$

Zipf Distribution

- Zipf's Law & Zipf distribution: American linguist George Kingsley Zipf (1902-1950)
- Popularity distribution: popularity of the i^{th} most popular term is proportional to $1/i$.
- If $X \sim Zipf(\alpha > 0)$, then PMF of X is:

$$P(X = k) = \frac{\frac{1}{k^{\alpha+1}}}{\sum_{j=1}^{\infty} \left(\frac{1}{j}\right)^{\alpha+1}}, k = 1, 2, \ldots$$

Example of Zipf Distribution: Word Frequency



6 5 4 3 2 1

THE 6

OF 3/6 = $\frac{1}{2}$

AND 2/6 = $\frac{1}{3}$

Power-law

Long-tail

Heavy-tail

distribution

Examples of Zipf Distribution

- The world population lives in several large cities, a greater number of medium-sized cities, and a vast number of small towns.
- There are a few websites that get lots of hits, a greater number of websites that get a moderate number of hits, and a vast number of websites that hardly get any hits at all.
- A library has a few books that everyone wants to borrow (best sellers), a greater number of books that get borrowed occasionally (classics), and a vast number of books that hardly ever get borrowed.

Outline

Definition

CDF

$P(X \le x)$

discrete r.v. PMF $P(X=x)$

continuous r.v. PDF $f_X(x)$

Theorem

The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \le x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

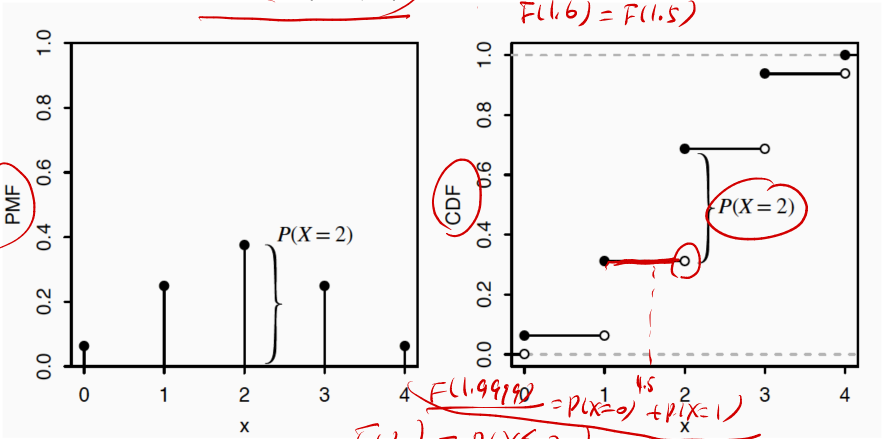
Example

$F(1.5) = P(X \le 1.5)$

$= P(X=0) + P(X=1)$

Let $X \sim Bin(4, 1/2)$, the PMF and CDF of X:

$F(1.6) = F(1.5)$



$F(1.9999) = P(X=0) + P(X=1)$

$F(2) = P(X \le 2)$

$= P(X=0) + P(X=1) + P(X=2)$

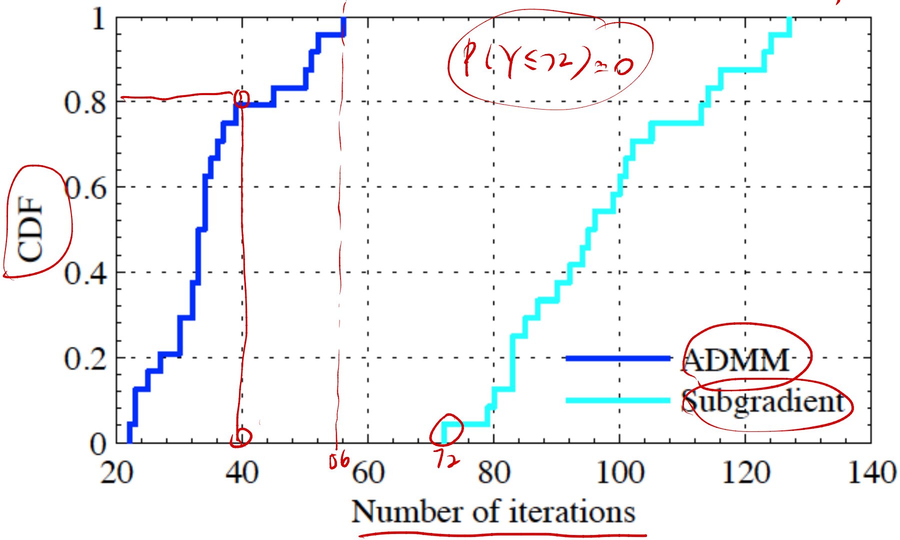
Example

X = # of iterations of ADMM

Y = ... Subgradient

$P(X \le 40) = 0.8$

$P(X \le 56) = 1$



Valid CDFs

Any CDF F has the following properties.

- Increasing: If $x_1 \leq x_2$, then $F(x_1) \leq F(x_2)$.
- Right-continuous: the CDF is continuous except possibly for having some jumps. Wherever there is a jump, the CDF is continuous from the right. That is, for any a, we have

$$F(a) = \lim_{x \to a^+} F(x).$$

- Convergence to 0 and 1 in the limits:

$F(x) = P(X \leq x)$

$$\lim_{x \to -\infty} F(x) = 0 \text{ and } \lim_{x \to \infty} F(x) = 1$$

Outline

Definition

$$\begin{array}{ccccc} S & \xrightarrow{X} & R & \xrightarrow{g} & R \\ s & \longrightarrow & X(s) & \longrightarrow & g(X(s)) \end{array}$$

Theorem

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 $s \in S$.

PMF of $g(X)$

$g(x) = x^2$

Theorem

Let X be a discrete r.v. and $g : \mathbb{R} \to \mathbb{R}$. Then the support of $g(X)$ is the set of all y such that $g(x) = y$ for at least one x in the support of X, and the PMF of $g(X)$ is

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

for all y in the support of $g(X)$.

Example: Maximum of Two Die Rolls

$\max\{X,Y\}$ r.v $\in \{1,2,3,4,5,6\}$

$1^\circ.\ P[\max(X,Y)=1] = P(X=1, Y=1) = P(X=1)P(Y=1)$

$= \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$

We roll two fair 6-sided dice. Let X be the number on the first die and Y the number on the second die. What is the PMF of $\max(X, Y)$.

$2^\circ.\ P[\max(X,Y)=2] = P(X=1, Y=2) + P(X=2, Y=1)$

$+ P(X=2, Y=2) = \frac{3}{36} = \frac{1}{12};$

$3^\circ.\ P[\max(X,Y)=c] = \begin{cases} \frac{5}{36} & c=3 \\ \frac{4}{} & c=5 \\ \frac{2}{36} & c=4 \end{cases}$ $\frac{11}{36}$ $c=6$

Example: Sympathetic Magic

① $Y = 2X$ $\quad P(Y=y) \neq 2P(X=y)$

$$P(Y=y) = P(2X=y) = P(X=\tfrac{y}{2})$$

- Given an r.v. X, trying to get the PMF of $2X$ by multiplying the PMF of X by 2.
- Claiming that because X and Y have the same distribution, X must always equal Y, i.e., $P(X = Y) = 1$.

$P(X=1) = P(X=0) = \frac{1}{2}$

$P(Y=1) = P(Y=0) = \frac{1}{2}$

② Toss coin (H, T) $\quad X$: indicator of Head event.

fair coin $\quad Y$: Tail.

$X, Y \sim \text{Bern}(1/2)$. But $X \neq Y$.

$X+Y=1 \Rightarrow$ if $X=Y \Rightarrow \cancel{X=Y=1/2}$

Outline

Independence of Two R.V.s

Definition

Joint CDF   individual CDF × ... × ...

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

Joint PMF   individual PMF × ...

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

for all x,y with x in the support of X and y in the support of Y.

Independence of Many R.V.s

events A, B, C independent

$$\begin{cases} P(A \cap B) = P(A)P(B) \\ P(B \cap C) = P(B)P(C) \\ P(C \cap A) = P(C)P(A) \end{cases}$$

$P(A \cap B \cap C) = P(A)P(B)P(C)$

Definition

Random variables $X_1, \ldots, X_n$ are *independent* if

Joint CDF $\quad \prod_{i=1}^{n}$ individual CDF$_i$

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

for all $x_1, \cdots, x_n \in \mathbb{R}$. For infinitely many r.v.s, we say that they are independent if every finite subset of the r.v.s is independent.

I.I.D.

We will often work with random variables that are independent and have the same distribution. We call such r.v.s independent and identically distributed, or i.i.d. for short.

- Independent & Identically Distributed I.I.D.
- Independent & NOT Identically Distributed
- Dependent & Identically Distributed
- Dependent & NOT Identically Distributed

Binomial Distribution

Theorem

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

$X_i = I_{\{\text{trial } i \text{ is successful}\}}$

Binomial Distribution

Theorem

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

Proof 1: LOTP

$X \sim Bin(n,p)$; $Y \sim Bin(m,p)$.

$\underline{X \perp\!\!\!\perp Y}$ $\quad 0 \le k \le n+m$

$P(\underline{X+Y=k})$

$$\overset{LOTP}{=} \sum_{j=0}^{k} P(\underline{X+Y=k \mid X=j}) \cdot P(\underline{X=j})$$

$$= \underbrace{\sum_{j=0}^{k} P(Y=k-j \mid X=j) \cdot P(X=j)}_{X \perp\!\!\!\perp Y.}$$

$$= \sum_{j=0}^{k} P(Y=k-j) \cdot P(X=j)$$

$$= \sum_{j=0}^{k} \binom{m}{k-j} \cdot p^{k-j} (1-p)^{m-k+j} \cdot \binom{n}{j} p^{j} (1-p)^{n-j}$$

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

$$= \left(\sum_{j=0}^{k} \binom{m}{k-j} \cdot \binom{n}{j}\right) p^{k} \cdot (1-p)^{n+m-k} = \binom{n+m}{k} \cdot p^{k} (1-p)^{n+m-k}$$

$\binom{n+m}{k}$

$= \sum_{j=0}^{k} \binom{m}{k-j}\binom{n}{j}$

Story proof.

Proof 2: Representation

$X \sim Bin(n,p)$. $X = X_1 + \cdots + X_n$, $X_i \sim$ i.i.d. $Bern(p)$.

$Y \sim Bin(m,p)$, $Y = Y_1 + \cdots + Y_m$, $Y_j \sim$ i.i.d. $Bern(p)$.

$$X + Y = \underbrace{(X_1 + \cdots + X_n) + (Y_1 + \cdots + Y_m)}_{n+m \text{ i.i.d. } Bern(p)}$$

$$\Rightarrow X + Y \sim Bin(n+m, p)$$

Proof 3: Story

each trial. success with prob. p

X : # of success in n independent Bernoulli trials.

Y : m - - - - - -

$X+Y$: # of total success in $n+m$ - - - - - -

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

Conditional Independence of R.V.s

$\tilde{P}(\cdot) = P(\cdot | Z = z)$

Definition

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

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z) P(Y \leq y | Z = z).$$

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

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z) P(Y = y | Z = z).$$

Conditional PMF

PMF

$\{P(X=x), x \in \text{Support}\}$

$\{P(X=x|Z=z), x \in \text{Support}.\}$ Condition PMF

Definition

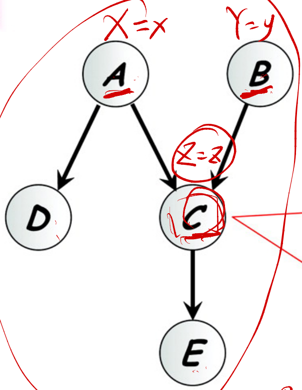
For any discrete r.v.s X and Z, the function $P(X = x|Z = z)$, when considered as a function of x for fixed z, is called the conditional PMF of X given $Z = z$.

Example: Bayesian Network

$X = x$

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

C



$P(C \mid A,B)$

A	B	0	1
0	0	0.9	0.1
0	1	0.2	0.8
1	0	0.9	0.1
1	1	0.01	0.99

$P(A)\,P(B \mid A)\,P(C \mid A,B)\,P(D \mid A,B,C)\,P(E \mid A,B,C,D)$

$$P(A,B,C,D,E) = P(A)P(B)P(C \mid A,B)P(D \mid A)P(E \mid C)$$

Example: Bayesian Network

- A probabilistic graphical model proposed by Judea Pearl in 1985
- Represents a set of random variables and their conditional dependencies
- Node: random variables
- Edge: conditional dependency
- Topology: a directed acyclic graph (DAG)
- Each node has a conditional probability table (CPT) with input from its parent nodes.
- Popular models for inference and leaning

Outline

Connection

- Binomial $\Longrightarrow$ Hypergeometric: **conditioning**
- Hypergeometric $\Longrightarrow$ Binomial: **taking a limit**

Connection

① $X+Y \sim \mathrm{Bin}(n+m, p)$

② $P(X=x \mid X+Y=r) = \dfrac{P(X=x, X+Y=r)}{P(X+Y=r)}$

$= \dfrac{P(X=x, Y=r-x)}{P(X+Y=r)} \overset{X \perp\!\!\!\perp Y}{=} \dfrac{P(X=x)P(Y=r-x)}{P(X+Y=r)}$

Theorem

If $X \sim \mathrm{Bin}(n,p)$, $Y \sim \mathrm{Bin}(m,p)$, and X is independent of Y, then the conditional distribution of X given $X+Y=r$ is $\mathrm{HGeom}(n,m,r)$.

$= \dfrac{\binom{n}{x}\cancel{p^x(1-p)^{n-x}} \cdot \binom{m}{r-x}\cancel{p^{r-x}(1-p)^{m-r+x}}}{\binom{n+m}{r}\cancel{p^r(1-p)^{n+m-r}}}$

$= \dfrac{\binom{n}{x}\binom{m}{r-x}}{\binom{n+m}{r}}$

Connection

$N = w+b \to \infty$.

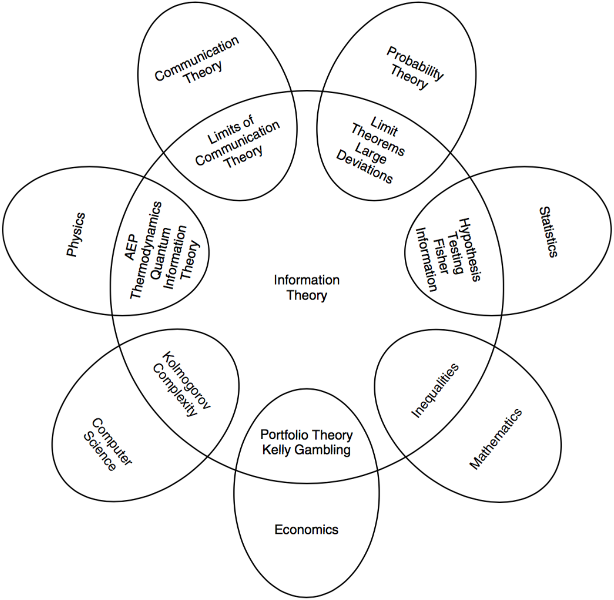
Sampling with/without replacement

Theorem

If $X \sim \mathrm{HGeom}(w, b, n)$ and $N = w + b \to \infty$ such that $p = w/(w + b)$ remains fixed, then the PMF of X converges to the $\mathrm{Bin}(n, p)$ PMF.

Outline

Information Theory & Other Fields



Entropy

$X \sim p(x) = \begin{cases} 1 & x = x_0 \\ 0 & \text{otherwise.} \end{cases}$ $\mathcal{X} = \{x_0\}$

$X = x_0$ a.s.

$H(X) = p(x_0)\log_2 \frac{1}{p(x_0)}$

$= 1 \cdot \log_2 \frac{1}{1} = \log_2 1 = 0$

Definition

Given a random variable X with a probability mass function $p(x)$ and a support $\mathcal{X}$. The entropy of X is defined by

bit

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) log_2 p(x)$$

$$= \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)}$$

Entropy of Discrete Uniform Distribution

X discrete r.v

- X has a uniform distribution over k outcomes.
- $p(x) = 1/k$
- Then the entropy of X is

$H(X) \leq \log_2 k$

$$H(X) = -\sum_{x=1}^{k} p(x) log_2 p(x) = -\sum_{x=1}^{k} \frac{1}{k} log_2 \frac{1}{k} = log_2 k$$

$\log_2 2 = 1$ (bit)

Balance Puzzle

$=$ $<$ $>$

① one weight $\log_2 3$ (bits)

k. n weights $n\log_2 3$ (bits)

$k \le \frac{3^n-1}{2}$.

$k=13$

$\Rightarrow 3^n \ge 27$

$\Rightarrow n \ge 3$

You have 13 apparently identical gold coins. One of them is false but is virtually indistinguishable form the others. You also have a balance with two pans, but without weights. Accordingly, any measurement will tell you if the loaded pans weight the same or, if not, which weighs more. How many measurements are needed to find the false coin?

② k coins, one of them is false.

$\log_2 k + \log_2 2 = \log_2 2k$ (bits)

③ $n\log_2 3 \ge \log_2 2k \Rightarrow 3^n \ge 2k \Rightarrow 3^n \ge 2k+1$

Solution

Entropy Bounds in General

$n \log_2 3 \geq \log_2 k \Rightarrow 3^n \geq k$

$n \geq \log_3 k$

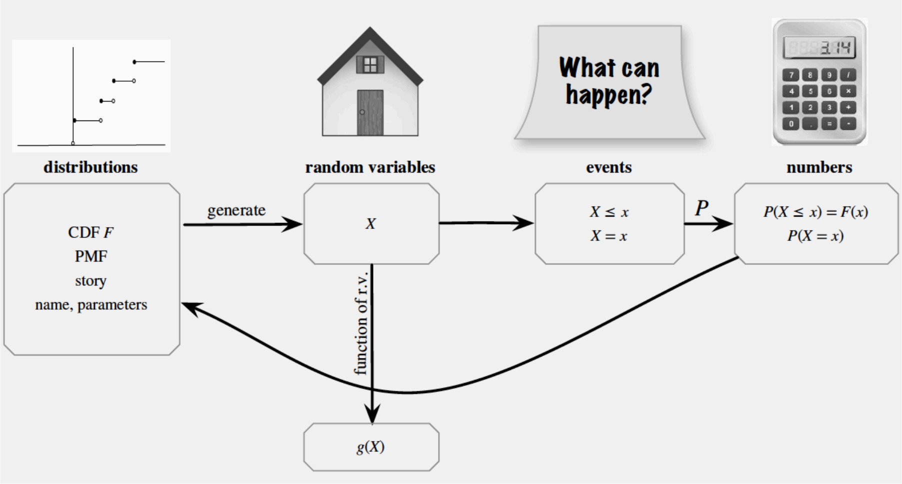
Known	Goal	Maximum Coins for n weighings	Number of Weighings for c coins
Whether target coin is lighter or heavier than others	Identify coin	3^n	$\lceil \log_3(c) \rceil$
Target coin is different from others	Identify coin	$\frac{3^n-1}{2}$ [1]	$\lceil \log_3(2c+1) \rceil$
Target coin is different from others, or all coins are the same	Identify if unique coin exists, and whether it is lighter or heavier	$\frac{3^n-1}{2}-1$	$\lceil \log_3(2c+3) \rceil$

$k \leq \frac{3^n-3}{2}$

$n \log_2 3 \geq \log_2 (k+1) + \log_2 2 = \log_2 2(k+1)$

$\Rightarrow 3^n \geq 2k+2 \quad \Rightarrow 3^n \geq 2k+3$

Summary 1



What can happen?
distributions
random variables
events
numbers
CDF F
PMF
story
name, parameters
generate
X
$X \le x$
$X = x$
P
$P(X \le x) = F(x)$
$P(X = x)$
function of r.v.
$g(X)$

References

- Chapter 3 of **BH**
- Chapter 2 of **BT**