

Lecture 3: Random Variables

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

March 19, 2024

Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Outline

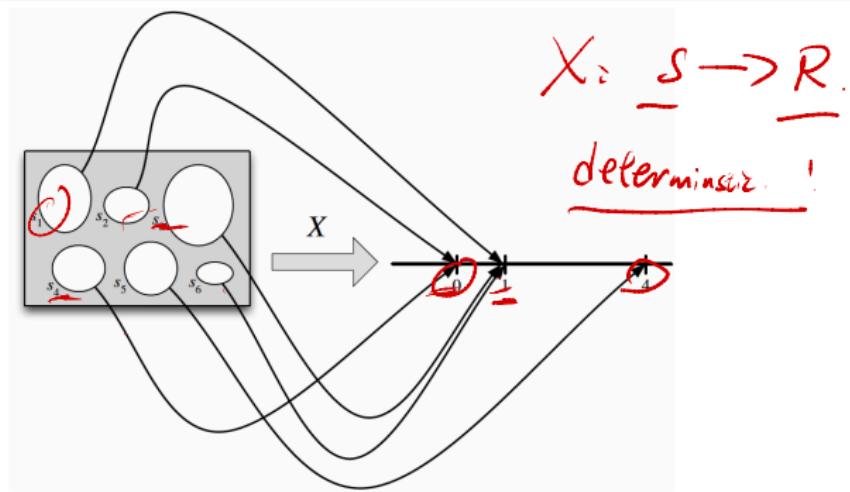
- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Definition of Random Variables



Definition

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.



Example: Coin Tosses

① $X(\underline{HH}) = \underline{2} ; X(\underline{HT}) = \underline{1} ; X(\underline{TH}) = \underline{1} ;$
 $X(\underline{TT}) = \underline{0} ;$

Consider an experiment where we toss a fair coin twice. The sample space consists of four possible outcomes: $S = \{\underline{HH}, \underline{HT}, \underline{TH}, \underline{TT}\}$. 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 $Y(s) = 2 - X(s), \quad \forall s \in S.$

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

$$Y(\underline{HH}) = 2 - X(\underline{HH}) = 0 ;$$
$$\dots$$

③ $I(\underline{HH}) = \underline{1} ; I(\underline{HT}) = \underline{1} ; I(\underline{TH}) = \underline{0} ; I(\underline{TT}) = \underline{0}$

Discrete Random Variable $\underline{X(s)}$

event $\underline{\{X = x\}} \triangleq \left\{ s : \frac{X(s) = x}{s \in S} \right\}$

Subset of sample space S

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

Definition

A random variable X is said to be discrete if there is a finite list of values a_1, a_2, \dots, a_n or an infinite list of values a_1, a_2, \dots 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 .

Coin tossing:

$$\underline{\{H, T\}}$$

$\{0, 1\}$: support of X .

X : # of Heads.

$$\underline{X = 1 \text{ or } 0}$$

$$\{X=1\} \triangleq \underline{\{H\}}, \quad \{X=0\} \triangleq \underline{\{T\}}$$

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

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

so

Probability Mass Function

$$\underline{p_X(x)} = \underline{P(X=x)} = \underline{P(\{s : X(s)=x, s \in S\})}$$

Definition

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

Example: Coin Tosses

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

$$\textcircled{1} \quad X=0 \iff s=TT ; \quad P_{X(0)} = P(X=0) = P(\{TT\}) = \frac{1}{4}.$$

$$X=1 \iff s=HT, TH ; \quad P_{X(1)} = P(X=1) = P(\{HT, TH\}) = \frac{1}{2} ;$$

$$X=2 \iff s=HH ; \quad P_{X(2)} = P(X=2) = P(\{HH\}) = \frac{1}{4}.$$

Consider an experiment where we toss a fair coin twice. The sample space consists of four possible outcomes: $S = \{\underline{HH}, \underline{HT}, \underline{TH}, \underline{TT}\}$.

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

X

• X: the number of Heads.

PMF

$$\{0, 1, 2\}$$

$$\begin{cases} P_{X(0)} = \frac{1}{4} \\ P_{X(1)} = \frac{1}{2} \\ P_{X(2)} = \frac{1}{4} \end{cases} \quad \text{PMF}$$

Y

• Y: the number of Tails.

PMF

(Tax)

I

• I: equals 1 if the first toss lands Heads and 0 otherwise. PMF

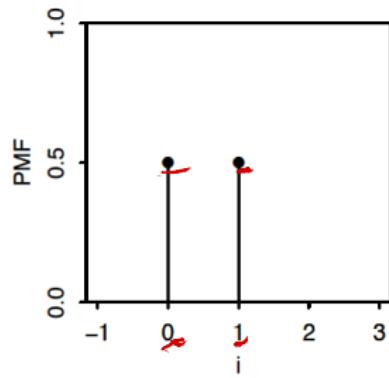
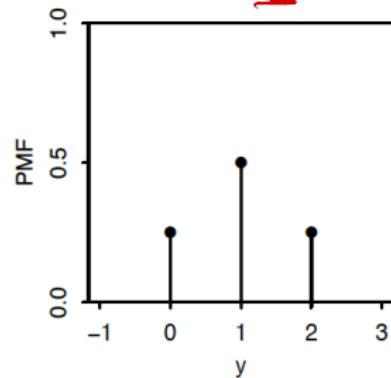
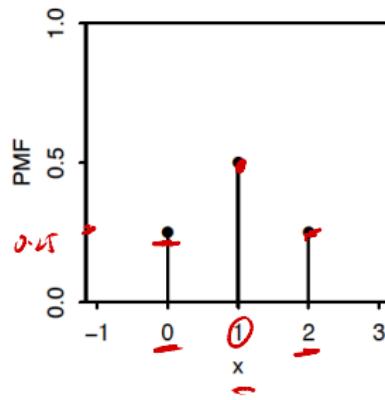
$$\textcircled{3} \quad P_{I(0)} = P(I=0) = P(\{TH, TT\}) = \frac{1}{2} \quad \begin{cases} P_{I(0)} = \frac{1}{2} \\ P_{I(1)} = \frac{1}{2} \end{cases}$$

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

Example: Coin Tosses

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

~~$$P(X=2) = \frac{1}{4}$$~~



X : distribution

Valid PMFs

Theorem

Let X be a discrete r.v. with support x_1, x_2, \dots (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

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Bernoulli Distribution

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 $X \sim \text{Bern}(p)$. The symbol \sim is read “is distributed as”.

Indicator Random Variable

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 $p = \underline{P(A)}$.

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.

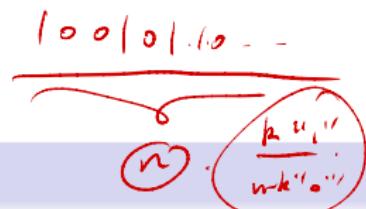
Story: Binomial Distribution

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

0: failure



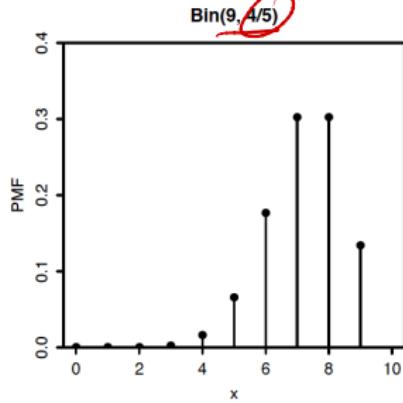
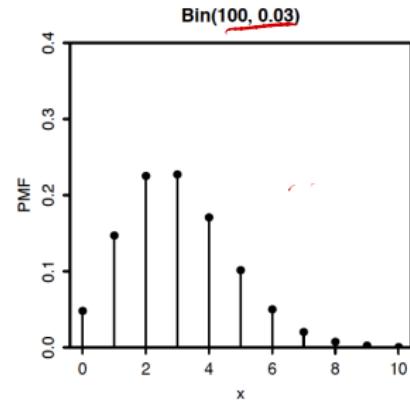
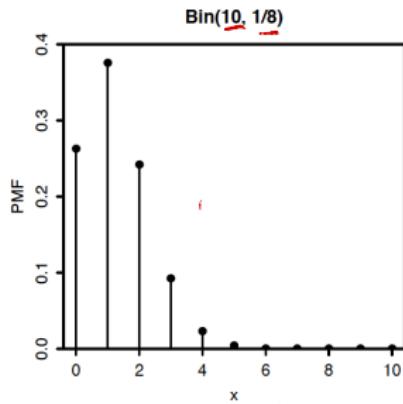
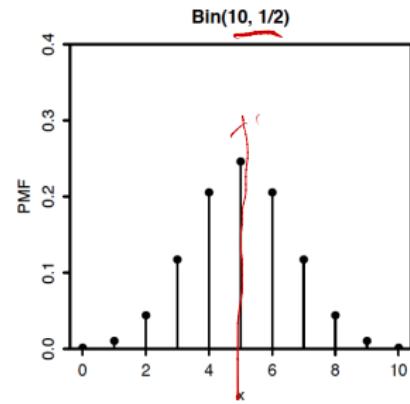
Theorem

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

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

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

Binomial PMF



Binomial PMF

Theorem

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

Example: Statistical Multiplexing

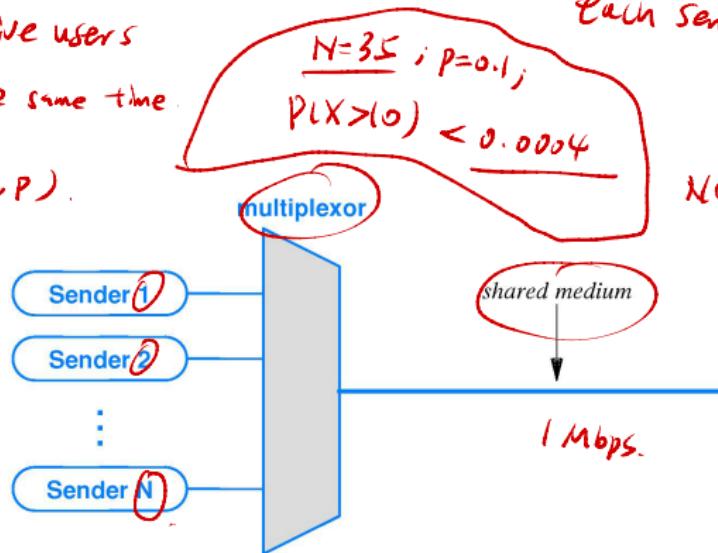
X : # of active users
at the same time.

$$X \sim \text{Bin}(N, p).$$

$$P(X=k)$$

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

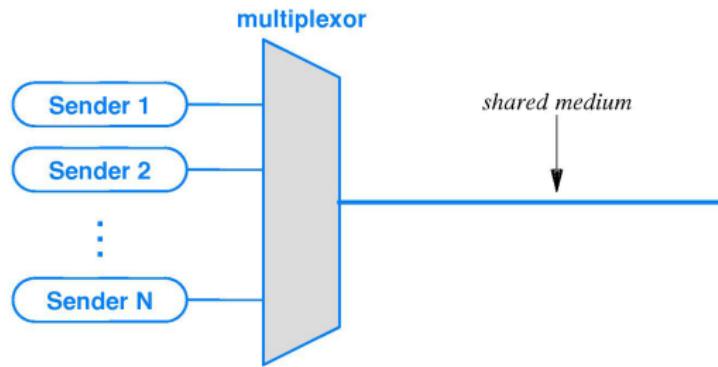
$$k \in \{0, 1, \dots, N\}.$$



$$\text{System overload: } P(X > 10) = 1 - P(X \leq 10) = 1 - \sum_{k=0}^{10} P(X=k)$$

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

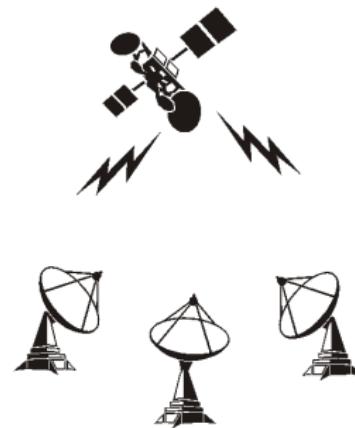
Example: Statistical Multiplexing



Example: Multiple Access (Aloha Protocol)



shared wireless



satellite

Example: Multiple Access (~~Aloha~~ Protocol)

unit time: time-slotted system



X : # of devices transmit simultaneously

$$X \sim \text{Bin}(N, p)$$

- 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)?

$$f'(p) = 0; f''(p) < 0 \quad \begin{matrix} \text{maximize } f(p), \\ 0 \leq p \leq 1 \end{matrix}$$

$$\Rightarrow p^* = \frac{1}{N} \quad \Rightarrow f^*(p) = \left(1 - \frac{1}{N}\right)^{N-1}$$

$$\begin{aligned} P(X=1) \cdot 1 &= P(X=1) = \binom{N}{1} p^* (1-p)^{N-1} \\ &= N \cdot p^* (1-p)^{N-1} \\ &= f(p) \end{aligned}$$

$\xrightarrow{N \rightarrow \infty} e^{-1} \approx 0.36$

Example: Multiple Access (Aloha Protocol)

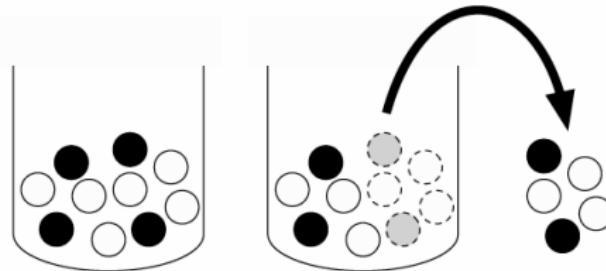
Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Urn Model

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

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



Story: Hypergeometric Distribution

if $X=k$;

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

$0 \leq k \leq w$
 $0 \leq k \leq n$.

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 $X \sim \text{HGeom}(w, b, n)$.

Hypergeometric PMF

$$\frac{\binom{n}{k} \binom{w}{n-k}}{\binom{w+b}{n}}$$

Theorem

If $X \sim \text{HGeom}(w, b, n)$, 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 $0 \leq k \leq w$ and $0 \leq n - k \leq b$, and $P(X = k) = 0$ otherwise.

Identical Distribution

X : # of white balls in sampled balls

color tag first.

sampled tag second.

two types of tags:

① color tag : white/black.

② sampled / unsampled.

Theorem

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

Y : # of sampled balls among the white balls.

sampled tag first

color tag second.

Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Story: Discrete Uniform Distribution

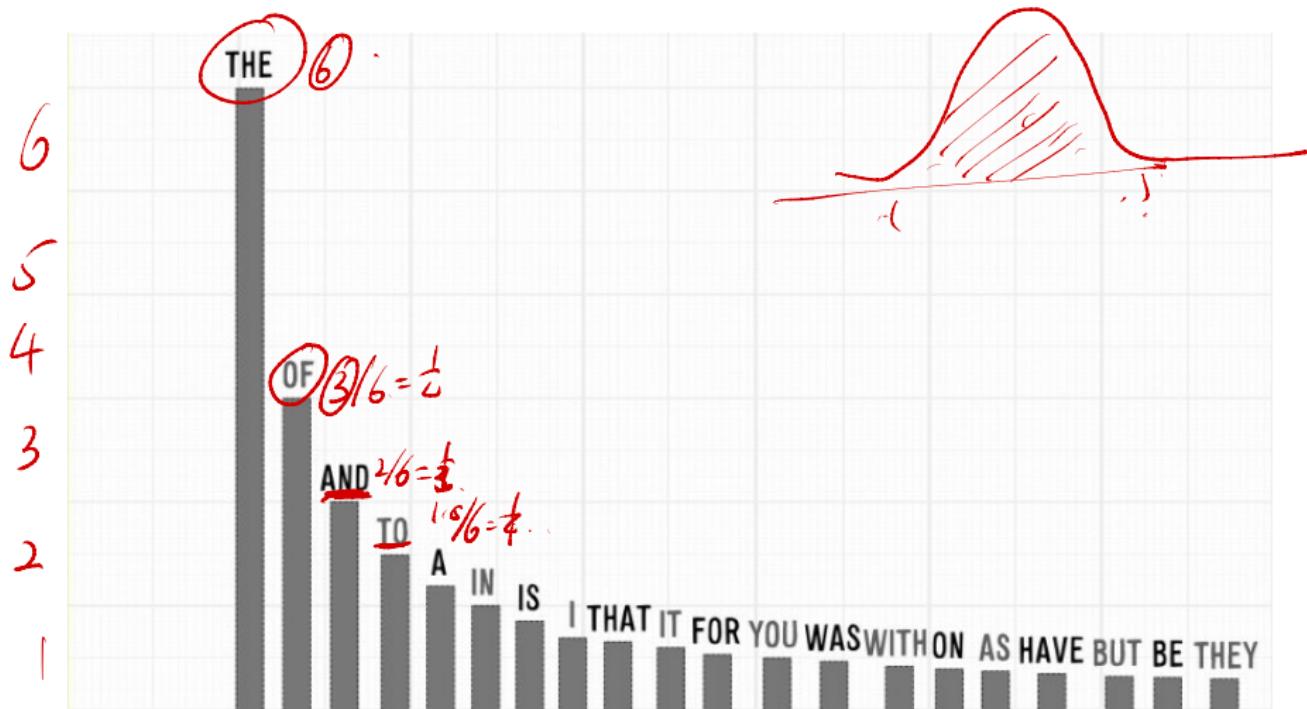
Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X . Then X is said to have the *Discrete Uniform distribution* with parameter C ; we denote this by $X \sim \text{DUnif}(C)$.

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 $X \sim \text{Zipf}(\alpha > 0)$, then PMF of X is:

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

Example of Zipf Distribution: Word Frequency



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

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Definition

$$\left. \begin{array}{l} \text{CDF} \\ p(X \leq x) \end{array} \right\} \begin{array}{l} \text{discrete r.v. : PMF} \\ \frac{p(X=x)}{} \end{array}$$
$$\begin{array}{l} \text{continuous r.v. : PDF} \\ f_X(x) \end{array}$$

Theorem

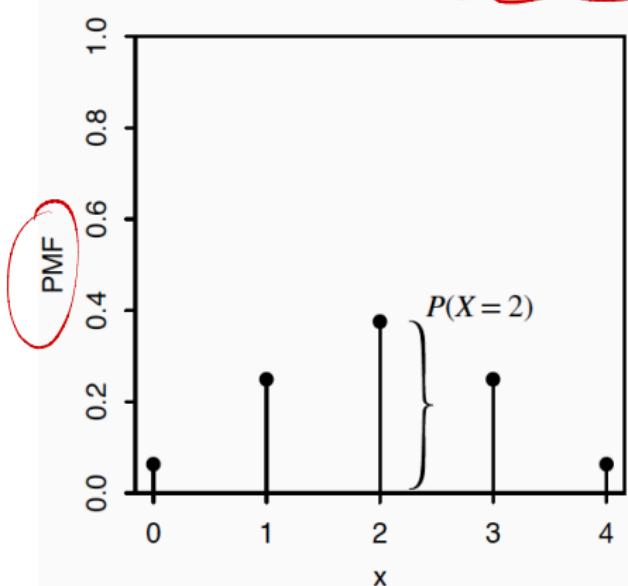
The cumulative distribution function (CDF) of an r.v. X is the function F_x given by $\underline{F_x(x) = P(X \leq x)}$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

$$\underline{0 \leq F(x) \leq 1}$$

Example

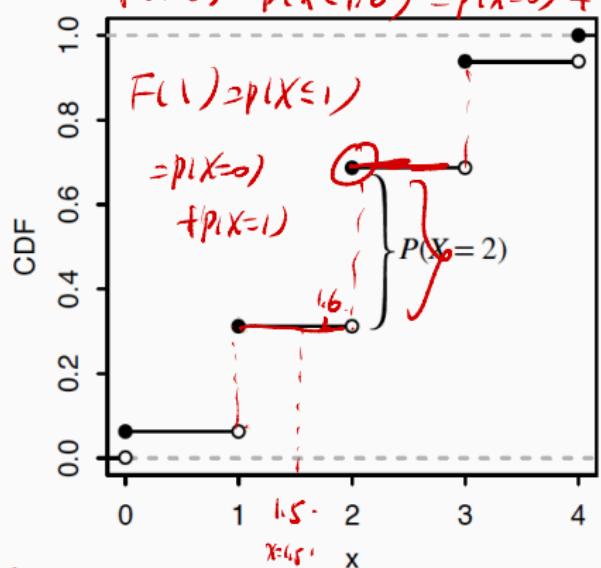
$$P(X=k) = \binom{4}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{4-k} = \frac{1}{16} \binom{4}{k}$$

$k=0,1,2,3,4$, Let $X \sim \text{Bin}(4, 1/2)$, the PMF and CDF of X :



$$\begin{aligned} F(1.5) &= P(X \leq 1.5) \\ &= P(X=0) + P(X=1) \end{aligned}$$

$$F(1.6) = P(X \leq 1.6) = P(X=0) + P(X=1)$$



$$F(2) = P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$$

Example

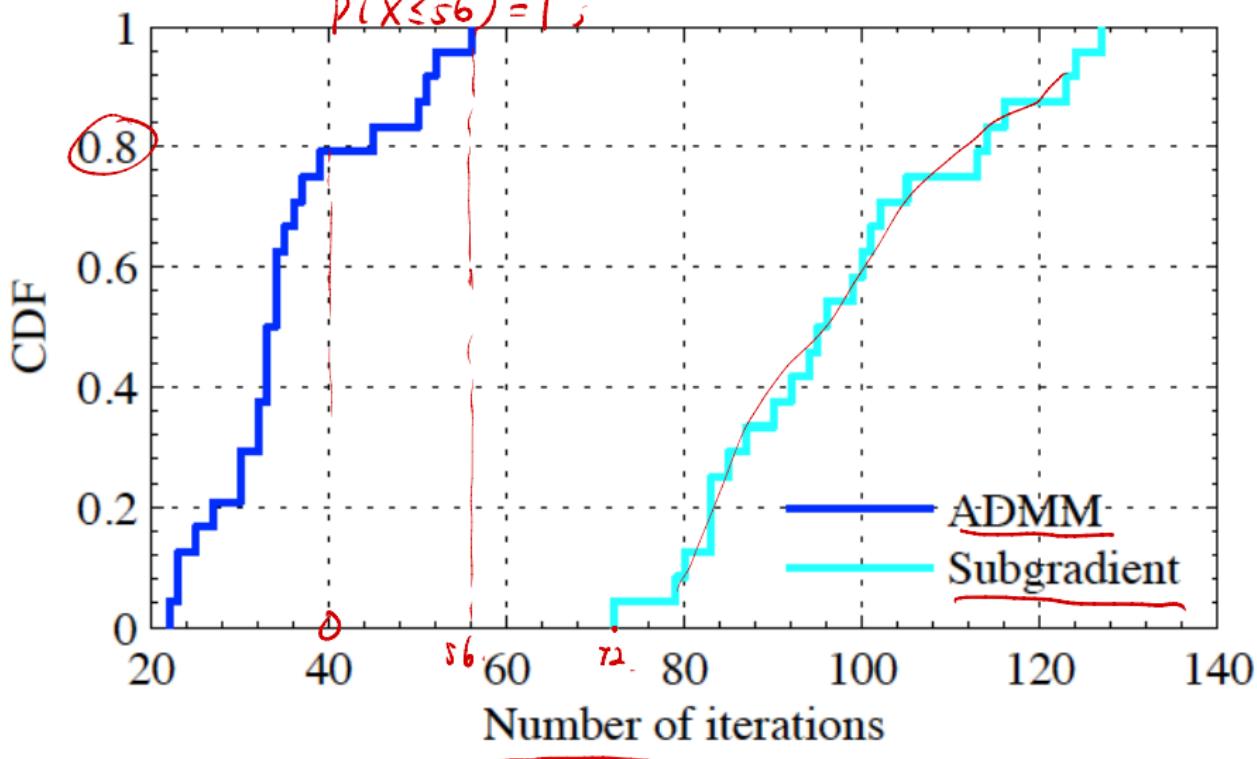
X : # of iterations of ADMM.

$$P(X \leq 40) = 0.8;$$

$$P(X \leq 56) = 1;$$

Y : # of iterations of Subgrad.

$$P(Y \leq 72) = 0$$



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

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

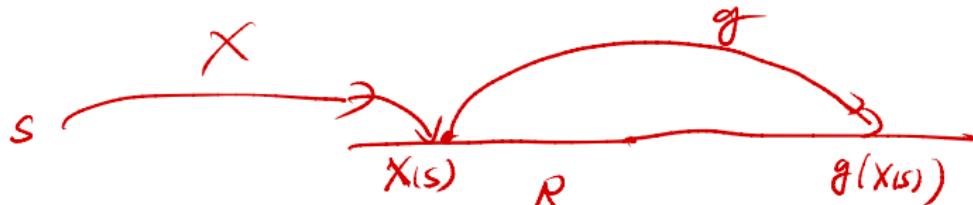
- Convergence to 0 and 1 in the limits:

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

Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Definition



Theorem

$s \xrightarrow{X} X(s) \xrightarrow{g} g(X(s))$

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

PMF of $g(X)$

Theorem

Let X be a discrete r.v. and $g : \mathbb{R} \rightarrow \mathbb{R}$. 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

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

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

Example: Maximum of Two Die Rolls

$$\underline{\max\{X, Y\}} \in \{1, 2, 3, 4, 5, 6\}$$

$$1^{\circ}. P[\max(X, Y) = 1] = P(\underline{X=1}, \underline{Y=1}) = P(X=1)P(Y=1) \\ = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}.$$

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

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 $\max(X, Y)$.

$$= \frac{3}{36} = \frac{1}{12}.$$

$$3^{\circ}. P[\max(X, Y) = 3] = \frac{5}{36}; \quad P[\max(X, Y) = 5] = \frac{1}{4}$$

$$P[\max(X, Y) = 4] = \frac{7}{36}; \quad P[\max(X, Y) = 6] = \frac{11}{36}$$

Example: Sympathetic Magic

$$\textcircled{1} \quad Y = 2X \quad ; \quad P(Y=y) \neq 2P(X=y)$$

$$P(Y=y) = P(2X=y) = P\left(X=\frac{y}{2}\right)$$

- 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., $P(X=Y)=1$.

\textcircled{2} Toss coin $X = \begin{cases} \text{indicator of Head} \\ \text{fair coin} \end{cases}$ $P(X=1) = \frac{1}{2}$
 $P(X=0) = \frac{1}{2}$

$Y = \begin{cases} \text{tail} \end{cases}$

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

$$X \sim Y \sim \text{Bern}(\frac{1}{2})$$

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

$$\cancel{X \neq Y} \quad [\quad X+Y=1 \Rightarrow Y=1-X, \text{ if } X=Y=1 \Rightarrow X=Y=\frac{1}{2} \quad]$$

Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Independence of Two R.V.s

Definition

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 with x in the support of X and y in the support of Y .

Independence of Many R.V.s

A, B, C independent

$$\left. \begin{aligned} P(A, B, C) &= P(A)P(B)P(C) \\ P(A, B) &= P(A)P(B) \\ P(B, C) &= P(B)P(C) \\ P(A, C) &= P(A)P(C) \end{aligned} \right\}$$

Definition

Random variables X_1, \dots, X_n are *independent* if

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

for all $x_1, \dots, x_n \in \mathbb{R}$. 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
- Independent & NOT Identically Distributed
- Dependent & Identically Distributed
- Dependent & NOT Identically Distributed

IID



Binomial Distribution

Theorem

If $X \sim \text{Bin}(n, p)$, 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 the X_i are i.i.d. $\text{Bern}(p)$.

Binomial Distribution

Theorem

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

Proof 1: LOTP

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

$0 \leq k \leq n+m$; X and Y are independent.

$$P(X+Y=k)$$

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

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

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

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

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

$$= \binom{m+n}{k} p^k (1-p)^{n+m-k} \sim \text{Bin}(m+n, p)$$

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

Proof 2: Representation

$X \sim \text{Bin}(n, p)$; $X = X_1 + \dots + X_n$, $X_i \sim \text{i.i.d. Bern}(p)$.

$Y \sim \text{Bin}(m, p)$; $Y = Y_1 + \dots + Y_m$, $Y_j \sim \text{i.i.d. Bern}(p)$.

$$X + Y = (X_1 + \dots + X_n) + (Y_1 + \dots + Y_m)$$



$n+m$ i.i.d. $\text{Bern}(p)$

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

Proof 3: Story

each successful prob. \underline{P} .

X : # of success in n independent Bernoulli trials.

Y : - - - - - m - - - - - - - - -

$X+Y$: # of total success in $n+m$ independent Bernoulli trials.

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

Conditional Independence of R.V.s

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

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

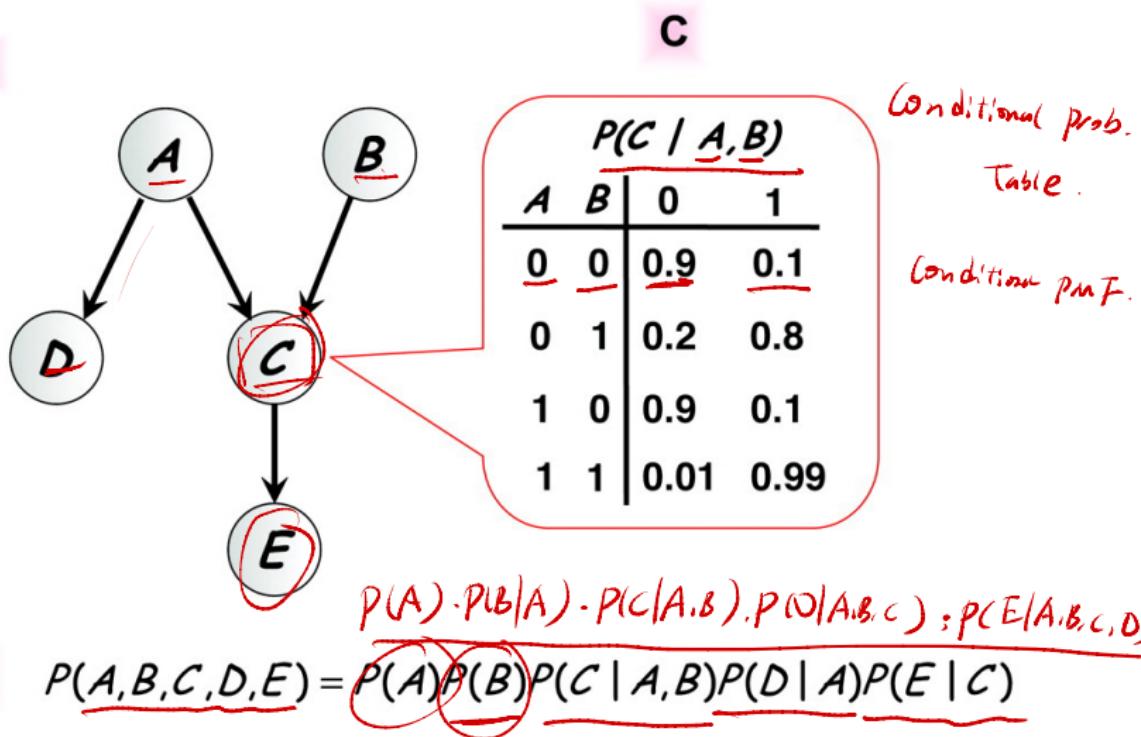
Conditional PMF

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

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



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 learning

Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Connection

- Binomial \Rightarrow Hypergeometric: **conditioning**
- Hypergeometric \Rightarrow Binomial: **taking a limit**

Connection ① $X+Y \sim \text{Bin}(n+m, p)$

$$\begin{aligned} \textcircled{2} \quad P(X=x | X+Y=r) &= \frac{P(X=x, X+Y=r)}{P(X+Y=r)} \\ &= \frac{P(X+Y=r | X=x) P(X=x)}{P(X+Y=r)} = \frac{P(Y=r-x | X=x) P(X=x)}{P(X+Y=r)} \\ &= \frac{P(Y=r-x) \cdot P(X=x)}{P(X+Y=r)} \end{aligned}$$

Theorem

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

$$\begin{aligned} &= \frac{\binom{m}{r-x} p^{r-x} (1-p)^{m-r+x} \cdot \binom{n}{x} p^x (1-p)^{n-x}}{\binom{n+m}{r} p^r (1-p)^{n+m-r}} \\ &= \frac{\binom{m}{r-x} \binom{n}{x}}{\binom{n+m}{r}} \end{aligned}$$

Connection

$$N = w + b \rightarrow \infty,$$

Sampling with/without

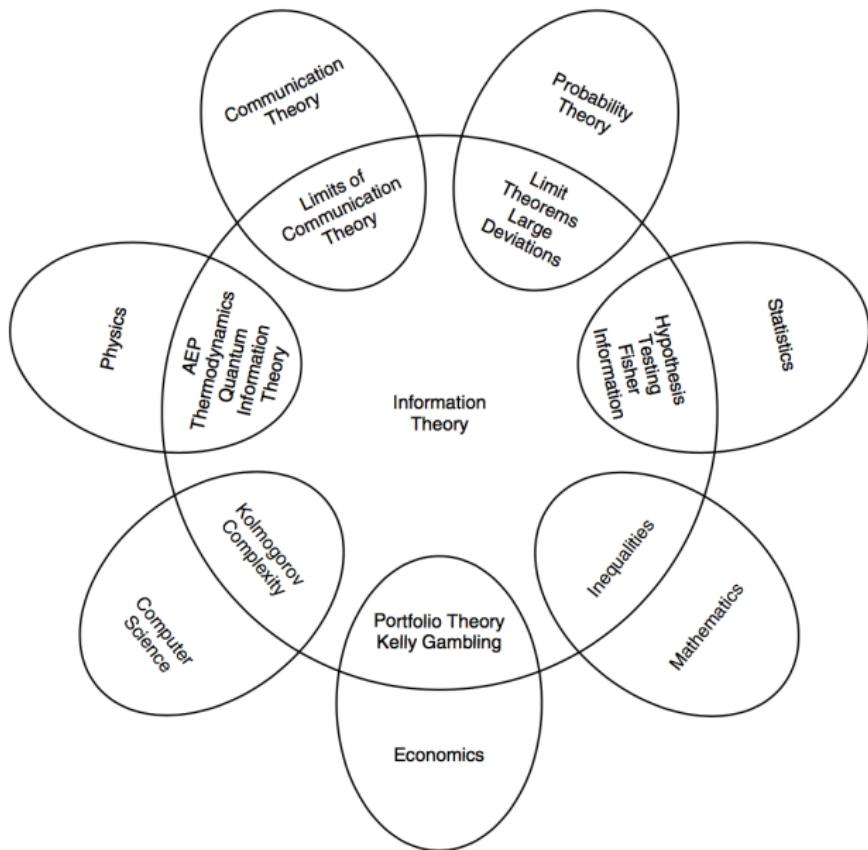
Theorem

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

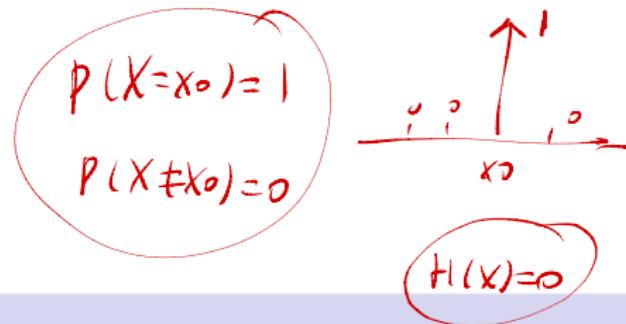
Outline

- 1 Random Variables
- 2 Bernoulli and Binomial
- 3 Hypergeometric
- 4 Discrete Uniform & Zipf Distribution
- 5 Cumulative Distribution Functions
- 6 Functions of Random Variables: Random Variables
- 7 Independence of R.V.s
- 8 Binomial & Hypergeometric
- 9 Information Theory & Entropy

Information Theory & Other Fields



Entropy



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

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

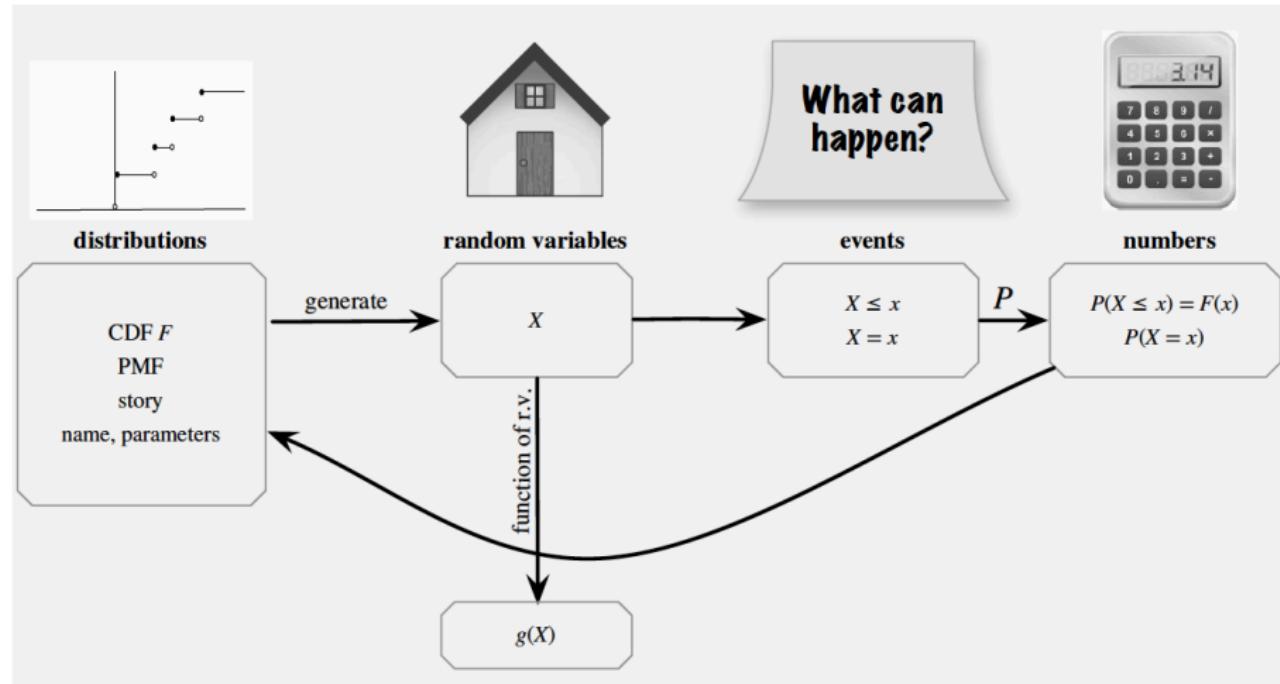
$$= \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{1}{p(x)}$$

Entropy of Discrete Uniform Distribution

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

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

Summary 1



References

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