

Lecture 3: Random Variables

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

October 15, 2024

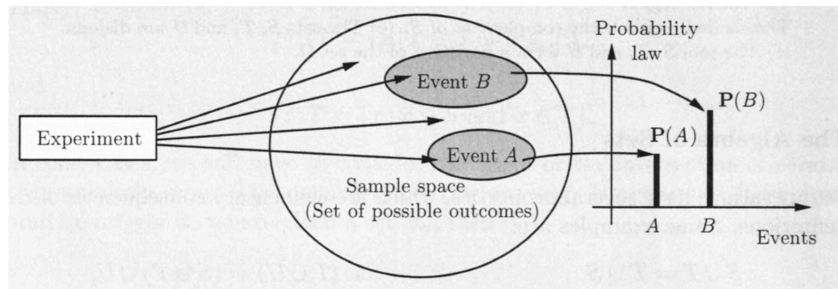
Outline

- 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

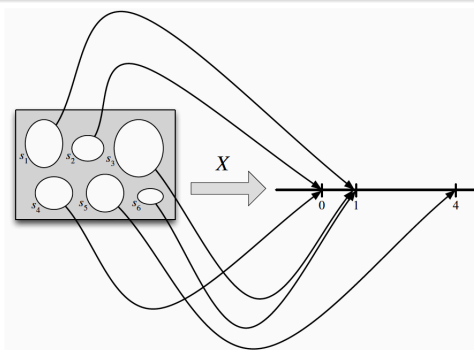
Probabilistic Model



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

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

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

Discrete Random Variable

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 .

Probability Mass Function

Definition

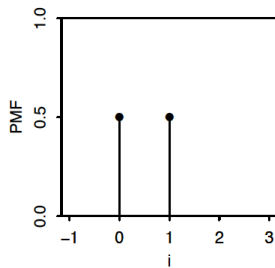
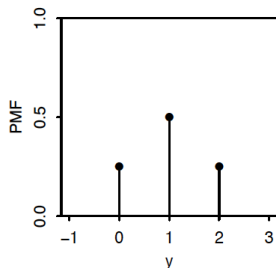
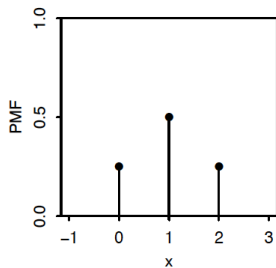
The *probability mass function* (PMF) of a discrete r.v. X is the function p_X given by $p_X(x) = P(X = x)$. Note that this is positive if x is in the support of X , and 0 otherwise.

Example: Coin Tosses

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

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

Example: Coin Tosses



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 $p = P(A)$.

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

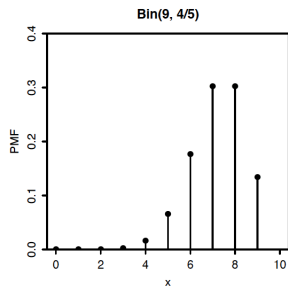
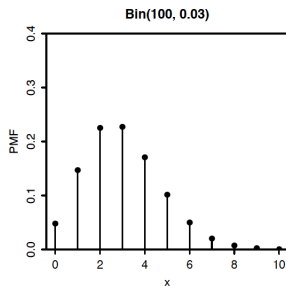
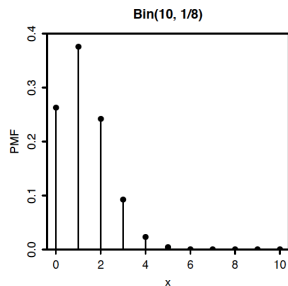
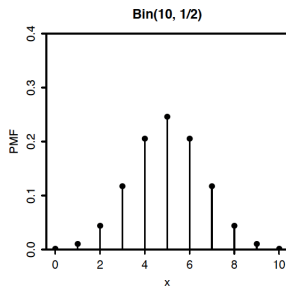
Theorem

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

Binomial PMF

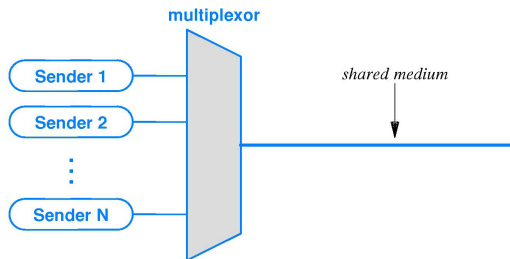


Binomial PMF

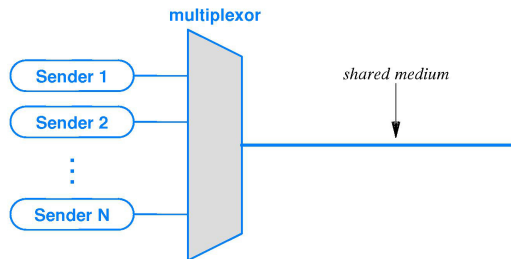
Theorem

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

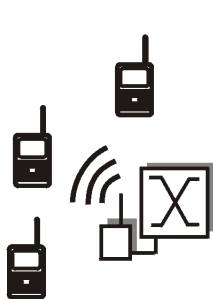
Example: Statistical Multiplexing



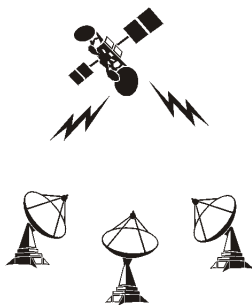
Example: Statistical Multiplexing



Example: Multiple Access (Aloha Protocol)



shared wireless



satellite

Example: Multiple Access (Aloha Protocol)

- 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)?

Example: Multiple Access (Aloha Protocol)

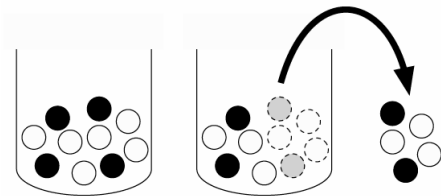
Outline

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



Story: Hypergeometric Distribution

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

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

Theorem

The $\text{HGeom}(w, b, n)$ and $\text{HGeom}(n, w + b - n, w)$ 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.

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

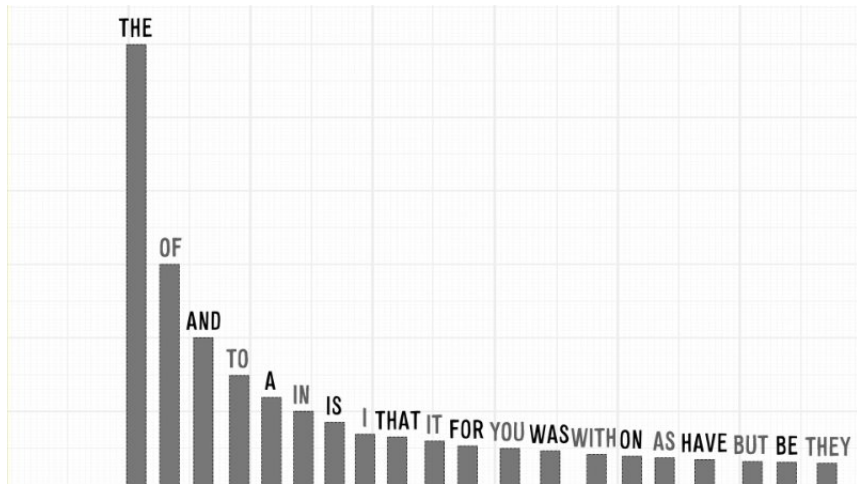
Let C be a finite, nonempty set of numbers. Choose one of these numbers uniformly at random (i.e., all values in C are equally likely). Call the chosen number X . Then X is said to have the *Discrete Uniform distribution* with parameter C ; we denote this by $X \sim \text{DUnif}(C)$.

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:

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

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

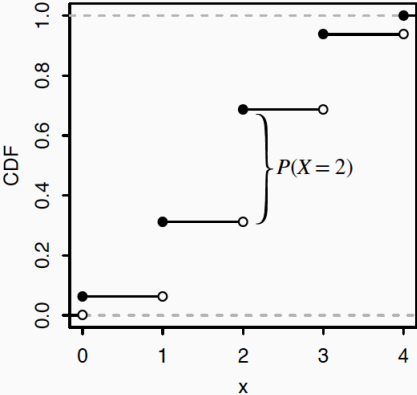
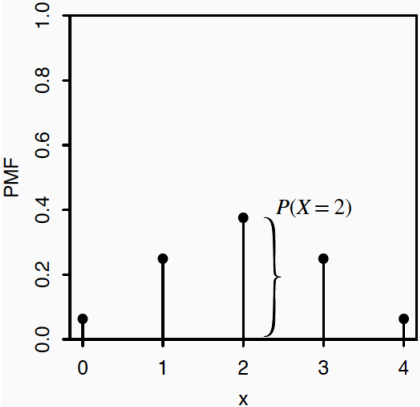
Definition

Theorem

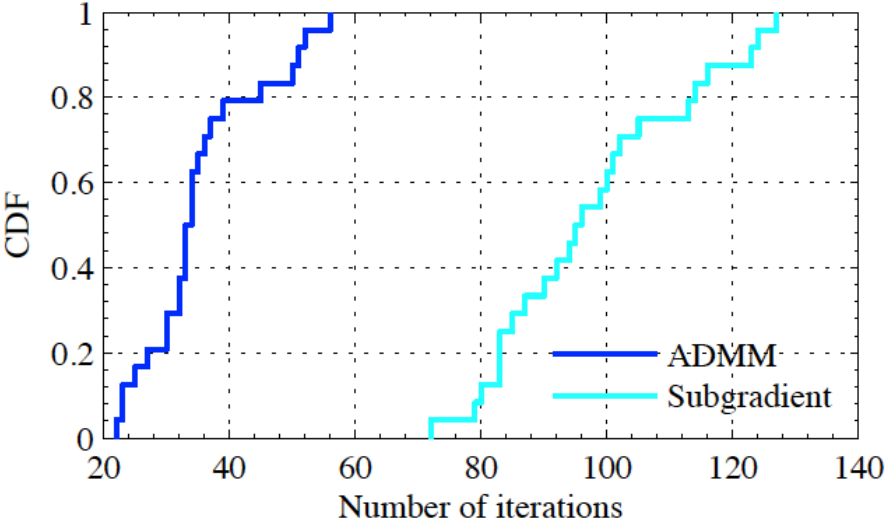
The cumulative distribution function (CDF) of an r.v. X is the function F_X given by $F_X(x) = P(X \leq x)$. When there is no risk of ambiguity, we sometimes drop the subscript and just write F (or some other letter) for a CDF.

Example

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



Example



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

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

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

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

Example: Maximum of Two Die Rolls

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 $\max(X, Y)$.

Example: Sympathetic Magic

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

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

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

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 $X = X_1 + \cdots + X_n$ 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 $X + Y \sim \text{Bin}(n + m, p)$

Proof 1: LOTP

Proof 2: Representation

Proof 3: Story

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

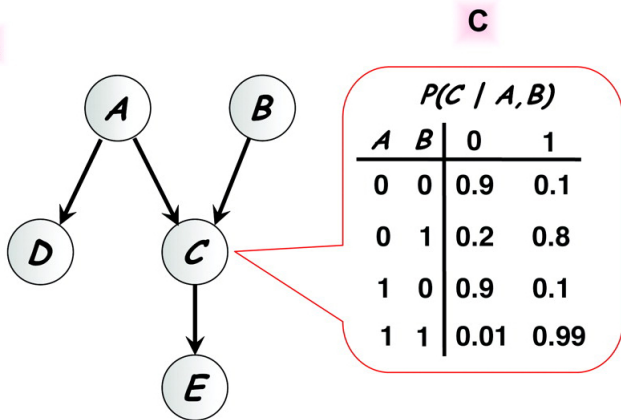
$$P(X = x, Y = y | Z = z) = P(X = x | Z = z) P(Y = y | Z = z).$$

Conditional PMF

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



$$P(A, B, C, D, E) = P(A)P(B)P(C | A, B)P(D | A)P(E | C)$$

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

Connection

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 $\text{HGeom}(n, m, r)$.

Connection

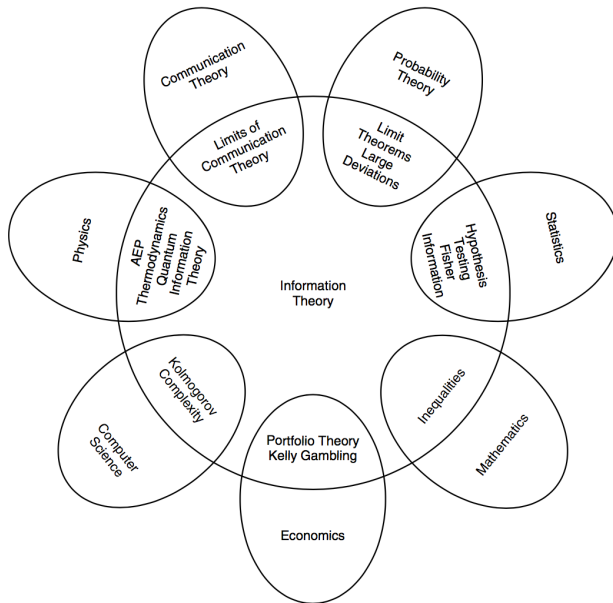
Theorem

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

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

Entropy of Discrete Uniform Distribution

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

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

Balance Puzzle

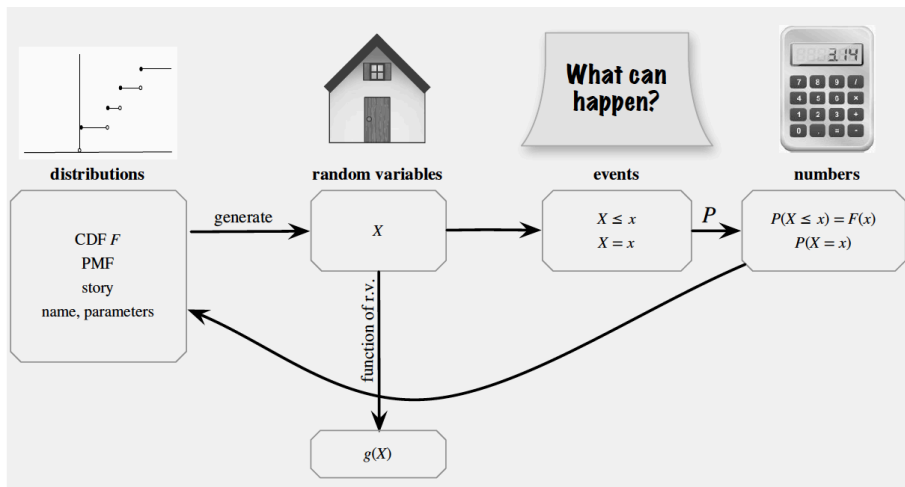
You have 13 apparently identical gold coins. One of them is false but is virtually indistinguishable from the others. You also have a balance with two pans, but without weights. Accordingly, any measurement will tell you if the loaded pans weight the same or, if not, which weighs more. How many measurements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

Solution

Entropy Bounds in General

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	$\frac{3^n - 1}{2} - 1$	$\lceil \log_3(2c + 3) \rceil$

Summary 1



References

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