CMPT 210: Probability and Computing

Lecture 12

Sharan Vaswani

February 15, 2024

Recap - (Basic) Frievald's Algorithm

- **Q**: For $n \times n$ matrices A, B and D, is $D = AB$?

O(n^2) time using Frievalds algorithm

- Last class, we proved that:

	Yes	No
$D = AB$	1	0
$D \neq AB$	$< \frac{1}{2}$	$\geq \frac{1}{2}$

Ask Sharan to go over the proof

Table 1: Probabilities for Basic Frievalds Algorithm

Probability amplification:
Want to amplify the probability of success.

Text

Frievald's Algorithm

By repeating the *Basic Frievald's Algorithm* m times, we will amplify the probability of success. The resulting complete Frievald's Algorithm is given by:

1 Run the Basic Frievald's Algorithm for m independent runs.

Generate a new vector x.

Generation of the random vector is done independently.

Frievald's Algorithm

By repeating the *Basic Frievald's Algorithm* m times, we will amplify the probability of success. The resulting complete Frievald's Algorithm is given by:

1. Run the Basic Frievald's Algorithm for m independent runs.
2. If *any* run of the Basic Frievald's Algorithm outputs "no", output "no".
3. If *all* runs of the Basic Frievald's Algorithm output "yes", output "yes". Dx = ABx

Every run is testing whether D = AB

Frievald's Algorithm

By repeating the *Basic Frievald's Algorithm* m times, we will amplify the probability of success. The resulting complete Frievald's Algorithm is given by:

1 Run the Basic Frievald's Algorithm for m independent runs.

2 If *any* run of the Basic Frievald's Algorithm outputs "no", output "no".

3 If *all* runs of the Basic Frievald's Algorithm output "yes", output "yes".

If

	Yes	No
$D = AB$	1	0
$D \neq AB$	$< \frac{1}{2^m}$	$\geq 1 - \frac{1}{2^m}$

As m increases, the probability of increasing the right answer increases.

Table 2: Probabilities for Frievald's Algorithm

Frievald's Algorithm

By repeating the *Basic Frievald's Algorithm* m times, we will amplify the probability of success. The resulting complete Frievald's Algorithm is given by:

1 Run the Basic Frievald's Algorithm for m independent runs.
2 If *any* run of the Basic Frievald's Algorithm outputs "no", output "no".
3 If *all* runs of the Basic Frievald's Algorithm output "yes", output "yes".

	Yes	No
$D = AB$	1	0
$D \neq AB$	$< \frac{1}{2^m}$	$\geq 1 - \frac{1}{2^m}$

Table 2: Probabilities for Frievald's Algorithm

If $m = 20$, then Frievald's algorithm will make mistake with probability $1/2^{20} \approx 10^{-6}$.

Computational Complexity: $O(mn^2)$

You run the algorithm m times,

Probability Amplification

Consider a randomized algorithm $\mathcal{A}$ that is supposed to solve a binary decision problem i.e. it is supposed to answer either Yes or No. It has a one-sided error – (i) if the true answer is Yes, then the algorithm $\mathcal{A}$ correctly outputs Yes with probability 1, but (ii) if the true answer is No, the algorithm $\mathcal{A}$ incorrectly outputs Yes with probability $\leq \frac{1}{2}$.

No mistake if D = AB

Probability Amplification

Consider a randomized algorithm $\mathcal{A}$ that is supposed to solve a binary decision problem i.e. it is supposed to answer either Yes or No. It has a one-sided error – (i) if the true answer is Yes, then the algorithm $\mathcal{A}$ correctly outputs Yes with probability 1, but (ii) if the true answer is No, the algorithm $\mathcal{A}$ incorrectly outputs Yes with probability $\leq \frac{1}{2}$.

Let us define a new algorithm $\mathcal{B}$ that runs algorithm $\mathcal{A}$ m times, and if *any* run of $\mathcal{A}$ outputs No, algorithm $\mathcal{B}$ outputs No. If *all* runs of $\mathcal{A}$ output Yes, algorithm $\mathcal{B}$ outputs Yes.

Probability Amplification

Consider a randomized algorithm $\mathcal{A}$ that is supposed to solve a binary decision problem i.e. it is supposed to answer either Yes or No. It has a one-sided error – (i) if the true answer is Yes, then the algorithm $\mathcal{A}$ correctly outputs Yes with probability 1, but (ii) if the true answer is No, the algorithm $\mathcal{A}$ incorrectly outputs Yes with probability $\leq \frac{1}{2}$.

Let us define a new algorithm $\mathcal{B}$ that runs algorithm $\mathcal{A}$ m times, and if *any* run of $\mathcal{A}$ outputs No, algorithm $\mathcal{B}$ outputs No. If *all* runs of $\mathcal{A}$ output Yes, algorithm $\mathcal{B}$ outputs Yes.

Q: What is the probability that algorithm $\mathcal{B}$ correctly outputs Yes if the true answer is Yes, and correctly outputs No if the true answer is No?

Probability Amplification - Analysis

If A_i denotes run i of Algorithm $\mathcal{A}$, then

$$\Pr[\mathcal{B} \text{ outputs Yes} \mid \text{true answer is Yes}]$$

Each run is independent since each run's output is independent of the other's result.

$$= \Pr[\mathcal{A}_1 \text{ outputs Yes } \cap \mathcal{A}_2 \text{ outputs Yes } \cap \ldots \cap \mathcal{A}_m \text{ outputs Yes} \mid \text{true answer is Yes}]$$

All runs must say Yes.

$$= \prod_{i=1}^{m} \Pr[\mathcal{A}_i \text{ outputs Yes} \mid \text{true answer is Yes}] = 1 \qquad \text{(Independence of runs)}$$

Probability Amplification - Analysis

If A_i denotes run i of Algorithm $\mathcal{A}$, then

$$\Pr[\mathcal{B} \text{ outputs Yes} \mid \text{true answer is Yes}]$$
$$= \Pr[\mathcal{A}_1 \text{ outputs Yes } \cap \mathcal{A}_2 \text{ outputs Yes } \cap \ldots \cap \mathcal{A}_m \text{ outputs Yes} \mid \text{true answer is Yes}]$$
$$= \prod_{i=1}^{m} \Pr[\mathcal{A}_i \text{ outputs Yes} \mid \text{true answer is Yes}] = 1 \qquad \text{(Independence of runs)}$$

Bounded by 1/2

$$\Pr[\mathcal{B} \text{ outputs No} \mid \text{true answer is No}]$$

Negation of the above statement

$$= 1 - \Pr[\mathcal{B} \text{ outputs Yes} \mid \text{true answer is No}]$$
$$= 1 - \Pr[\mathcal{A}_1 \text{ outputs Yes } \cap \mathcal{A}_2 \text{ outputs Yes } \cap \ldots \cap \mathcal{A}_m \text{ outputs Yes} \mid \text{true answer is No}]$$
$$= 1 - \prod_{i=1}^{m} \Pr[\mathcal{A}_i \text{ outputs Yes} \mid \text{true answer is No}] \geq 1 - \frac{1}{2^m}.$$

Probability Amplification - Analysis

If A_i denotes run i of Algorithm $\mathcal{A}$, then

$$\Pr[\mathcal{B} \text{ outputs Yes} \mid \text{true answer is Yes}]$$
$$= \Pr[\mathcal{A}_1 \text{ outputs Yes } \cap \mathcal{A}_2 \text{ outputs Yes } \cap \ldots \cap \mathcal{A}_m \text{ outputs Yes} \mid \text{true answer is Yes}]$$
$$= \prod_{i=1}^{m} \Pr[\mathcal{A}_i \text{ outputs Yes} \mid \text{true answer is Yes}] = 1 \qquad \text{(Independence of runs)}$$
$$\Pr[\mathcal{B} \text{ outputs No} \mid \text{true answer is No}]$$
$$= 1 - \Pr[\mathcal{B} \text{ outputs Yes} \mid \text{true answer is No}]$$
$$= 1 - \Pr[\mathcal{A}_1 \text{ outputs Yes } \cap \mathcal{A}_2 \text{ outputs Yes } \cap \ldots \cap \mathcal{A}_m \text{ outputs Yes} \mid \text{true answer is No}]$$
$$= 1 - \prod_{i=1}^{m} \Pr[\mathcal{A}_i \text{ outputs Yes} \mid \text{true answer is No}] \geq 1 - \frac{1}{2^m}.$$

When the true answer is Yes, both $\mathcal{B}$ and $\mathcal{A}$ correctly output Yes. When the true answer is No, $\mathcal{A}$ incorrectly outputs Yes with probability $< \frac{1}{2}$, but $\mathcal{B}$ incorrectly outputs Yes with probability $< \frac{1}{2^m} << \frac{1}{2}$. By repeating the experiment, we have "amplified" the probability of success.

<< : means very small

Questions?

Concentration inequalities

Random Variables

Definition: A random "variable" R on a probability space is a total function whose domain is the sample space $\mathcal{S}$. The codomain is usually a subset of the real numbers.

F: S -> V
Where V is a subset of R:

Random Variables

Definition: A random "variable" R on a probability space is a total function whose domain is the sample space $\mathcal{S}$. The codomain is usually a subset of the real numbers.

Example: Suppose we toss three independent, unbiased coins. Let C be the number of heads that appear.

Unbiased: Pr(head) = Pr(tails)

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

C is a total function that maps each outcome in $\mathcal{S}$ to a number as follows: $C(HHH) = 3$, $C(HHT) = C(HTH) = C(THH) = 2$, $C(HTT) = C(THT) = C(TTH) = 1$, $C(TTT) = 0$.

C is a random variable that counts the number of heads in 3 tosses of the coin.

Everything up to the 12th lecture will be included

Random Variables

Definition: A random "variable" R on a probability space is a total function whose domain is the sample space $\mathcal{S}$. The codomain is usually a subset of the real numbers.

Example: Suppose we toss three independent, unbiased coins. Let C be the number of heads that appear.

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

C is a total function that maps each outcome in $\mathcal{S}$ to a number as follows: $C(HHH) = 3$, $C(HHT) = C(HTH) = C(THH) = 2$, $C(HTT) = C(THT) = C(TTH) = 1$, $C(TTT) = 0$.

C is a random variable that counts the number of heads in 3 tosses of the coin.

Example: I toss a coin, and define the random variable R which is equal to 1 when I get a heads, and equal to 0 when I get a tails.

Bernoulli random variables: Random variables with the codomain $\{0, 1\}$ are called Bernoulli random variables. E.g. R is a Bernoulli r.v.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What is the domain, range of R?

the sample space, which is (1, 2, 3, 4, 5,6)^2

Range(R) = {2, … 12}

Range is the values R can take on

Domain are the values that R can have as input.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What is the domain, range of R?

With replacement

Q: Three balls are randomly selected from an urn containing 20 balls numbered 1 through 20. The random variable M is the maximal value on the selected balls. What is the domain, range of M?

You have Dom(M) = {1, 2, 3, 4, 5, ….. 20}^3

Range(M) = {1, 2, 3, … 20}

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What is the domain, range of R?

Text

Q: Three balls are randomly selected from an urn containing 20 balls numbered 1 through 20. The random variable M is the maximal value on the selected balls. What is the domain, range of M?

Q: In the above example, what is $2 \times M((1,4,6))$? Is M an invertible function?

12

No, since you can have different tuples map to the same value.

Random Variables and Events

Indicates whether an event has happened.

Indicator Random Variable: An indicator random variable maps every outcome to either 0 or 1.

Example: Suppose we throw two standard dice, and define M to be the random variable that is 1 iff both throws of the dice produce a prime number, else it is 0.

$M : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \to \{0,1\}$. $M((2,3)) = 1$, $M((3,6)) = 0$.

An indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0.

Random Variables and Events

Indicator Random Variable: An indicator random variable maps every outcome to either 0 or 1.

Example: Suppose we throw two standard dice, and define M to be the random variable that is 1 iff both throws of the dice produce a prime number, else it is 0.

$M : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} \to \{0,1\}$. $M((2,3)) = 1$, $M((3,6)) = 0$.

An indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0.

Example: When throwing two dice, if E is the event that both throws of the dice result in a prime number, then random variable $M = 1$ iff event E happens, else $M = 0$.

Random Variables and Events

Indicator Random Variable: An indicator random variable maps every outcome to either 0 or 1.

Example: Suppose we throw two standard dice, and define M to be the random variable that is 1 iff both throws of the dice produce a prime number, else it is 0.

$M : \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \to \{0, 1\}$. $M((2, 3)) = 1$, $M((3, 6)) = 0$.

An indicator random variable partitions the sample space into those outcomes mapped to 1 and those outcomes mapped to 0.

I_# is an indicator random variable for event E.

Example: When throwing two dice, if E is the event that both throws of the dice result in a prime number, then random variable $M = 1$ iff event E happens, else $M = 0$.

The indicator random variable corresponding to an event E is denoted as $\mathcal{I}_E$, meaning that for $\omega \in E$, $\mathcal{I}_E[\omega] = 1$ and for $\omega \notin E$, $\mathcal{I}_E[\omega] = 0$. In the above example, $M = \mathcal{I}_E$ and since $(2, 4) \notin E$, $M((2, 4)) = 0$ and since $(3, 5) \in E$, $M((3, 5)) = 1$.

Indicates whether an event has happened or not.

Random Variables and Events

In general, a random variable that takes on several values partitions $\mathcal{S}$ into several blocks.

Example: When we toss a coin three times, and define C to be the r.v. that counts the number of heads, C partitions $\mathcal{S}$ as follows: $\mathcal{S} = \{\underbrace{HHH}_{C=3}, \underbrace{HHT, HTH, THH}_{C=2}, \underbrace{HTT, THT, TTH}_{C=1}, \underbrace{TTT}_{C=0}\}$.

Each block is a subset of the sample space and is therefore an event. For example, $[C = 2]$ is the event that the number of heads is two and consists of the outcomes $\{HHT, HTH, THH\}$.

Event

Event that the random variable takes on the value of 2

[C = i] = {w in S I c[W} = i}

Pr(C = 2) = Pr({HHT, HTH, THH}) = 3/8.

Random Variables and Events

In general, a random variable that takes on several values partitions $\mathcal{S}$ into several blocks.

Example: When we toss a coin three times, and define C to be the r.v. that counts the number of heads, C partitions $\mathcal{S}$ as follows: $\mathcal{S} = \{\underbrace{HHH}_{C=3}, \underbrace{HHT, HTH, THH}_{C=2}, \underbrace{HTT, THT, TTH}_{C=1}, \underbrace{TTT}_{C=0}\}$.

Each block is a subset of the sample space and is therefore an event. For example, $[C = 2]$ is the event that the number of heads is two and consists of the outcomes $\{HHT, HTH, THH\}$.

Since it is an event, we can compute its probability i.e. $\Pr[C = 2] = \Pr[\{HHT, HTH, THH\}] = \Pr[\{HHT\}] + \Pr[\{HTH\}] + \Pr[\{THH\}]$. Since this is a uniform probability space, $\Pr[\omega] = \frac{1}{8}$ for $\omega \in \mathcal{S}$ and hence $\Pr[C = 2] = \frac{3}{8}$.

Random Variables and Events

In general, a random variable that takes on several values partitions $\mathcal{S}$ into several blocks.

Example: When we toss a coin three times, and define C to be the r.v. that counts the number of heads, C partitions $\mathcal{S}$ as follows: $\mathcal{S} = \{\underbrace{HHH}_{C=3}, \underbrace{HHT, HTH, THH}_{C=2}, \underbrace{HTT, THT, TTH}_{C=1}, \underbrace{TTT}_{C=0}\}$.

Each block is a subset of the sample space and is therefore an event. For example, $[C = 2]$ is the event that the number of heads is two and consists of the outcomes $\{HHT, HTH, THH\}$.

Since it is an event, we can compute its probability i.e. $\Pr[C = 2] = \Pr[\{HHT, HTH, THH\}] = \Pr[\{HHT\}] + \Pr[\{HTH\}] + \Pr[\{THH\}]$. Since this is a uniform probability space, $\Pr[\omega] = \frac{1}{8}$ for $\omega \in \mathcal{S}$ and hence $\Pr[C = 2] = \frac{3}{8}$.

Q: What is $\Pr[C = 0]$, $\Pr[C = 1]$ and $\Pr[C = 3]$?

Q: What is $\sum_{i=0}^{3} \Pr[C = i]$?

Summing over all the partitions is equivalent to summing the probability of every event, which is equivalent to Pr(S)

1/8

3/8

1/8

Random Variables and Events

In general, a random variable that takes on several values partitions $\mathcal{S}$ into several blocks.

Example: When we toss a coin three times, and define C to be the r.v. that counts the number of heads, C partitions $\mathcal{S}$ as follows: $\mathcal{S} = \{\underbrace{HHH}_{C=3}, \underbrace{HHT, HTH, THH}_{C=2}, \underbrace{HTT, THT, TTH}_{C=1}, \underbrace{TTT}_{C=0}\}$.

Each block is a subset of the sample space and is therefore an event. For example, $[C = 2]$ is the event that the number of heads is two and consists of the outcomes $\{HHT, HTH, THH\}$.

Since it is an event, we can compute its probability i.e.
$\Pr[C = 2] = \Pr[\{HHT, HTH, THH\}] = \Pr[\{HHT\}] + \Pr[\{HTH\}] + \Pr[\{THH\}]$. Since this is a uniform probability space, $\Pr[\omega] = \frac{1}{8}$ for $\omega \in \mathcal{S}$ and hence $\Pr[C = 2] = \frac{3}{8}$.

Q: What is $\Pr[C = 0]$, $\Pr[C = 1]$ and $\Pr[C = 3]$?

Q: What is $\sum_{i=0}^{3} \Pr[C = i]$?

Since a random variable R is a total function that maps every outcome in $\mathcal{S}$ to some value in the codomain, $\sum_{i \in \text{Range of R}} \Pr[R = i] = \sum_{i \in \text{Range of R}} \sum_{\omega \text{ s.t. } R(\omega)=i} \Pr[\omega] = \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1$.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What are the outcomes in the event $[R = 2]$?

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What are the outcomes in the event $[R = 2]$?

Q: What is $\Pr[R = 4]$, $\Pr[R = 9]$?

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define R to be the random variable equal to the sum of the dice. What are the outcomes in the event $[R = 2]$?

Q: What is $\Pr[R = 4]$, $\Pr[R = 9]$?

Q: If M is the indicator random variable equal to 1 iff both throws of the dice produces a prime number, what is $\Pr[M = 1]$?

9/36

{2, 3}, {3, 5}, {5, 2}, {2, 5}, {5, 3}, {3, 3}, {2, 2}, {5, 5{

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0, 1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0, 1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

$\sum_{x \in V} \text{PDF}_R[x] = \sum_{x \in \text{Range(R)}} \Pr[R = x] = 1.$

Sum of a pdf is 1.

If there is something that cannot happen, the pdf of that event is 0.

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\mathrm{PDF}_R : V \to [0, 1]$, such that $\mathrm{PDF}_R[x] = \Pr[R = x]$ if $x \in \mathrm{Range(R)}$ and equal to zero if $x \notin \mathrm{Range(R)}$.

$\sum_{x \in V} \mathrm{PDF}_R[x] = \sum_{x \in \mathrm{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): If the codomain is a subset of the real numbers, then the cumulative distribution function is the function $\mathrm{CDF}_R : \mathbb{R} \to [0, 1]$, such that $\mathrm{CDF}_R[x] = \Pr[R \leq x]$ Equality is important

Importantly, neither PDF_R nor CDF_R involves the sample space of an experiment.

Abstract from specific examples

PDF and CDF correspond to a specific random variable.

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\mathrm{PDF}_R : V \to [0,1]$, such that $\mathrm{PDF}_R[x] = \Pr[R = x]$ if $x \in \mathrm{Range(R)}$ and equal to zero if $x \notin \mathrm{Range(R)}$.

Remove use of sample space

$\sum_{x \in V} \mathrm{PDF}_R[x] = \sum_{x \in \mathrm{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): If the codomain is a subset of the real numbers, then the cumulative distribution function is the function $\mathrm{CDF}_R : \mathbb{R} \to [0,1]$, such that $\mathrm{CDF}_R[x] = \Pr[R \leq x]$.

Remove use of Codomain

Importantly, neither PDF_R nor CDF_R involves the sample space of an experiment.

Example: If we flip three coins, and C counts the number of heads, then

$\mathrm{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and

Probability that C takes on the value zero

$\mathrm{CDF}_C[2.3] = \Pr[C \leq 2.3] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

variable C takes on values less than or equal to 2.3

Pdf(4) = 0 since we can never see 4 heads in three coin flips.

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0, 1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

$\sum_{x \in V} \text{PDF}_R[x] = \sum_{x \in \text{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): If the codomain is a subset of the real numbers, then the cumulative distribution function is the function $\text{CDF}_R : \mathbb{R} \to [0, 1]$, such that $\text{CDF}_R[x] = \Pr[R \leq x]$.

Importantly, neither PDF_R nor CDF_R involves the sample space of an experiment.

Example: If we flip three coins, and C counts the number of heads, then
$\text{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and
$\text{CDF}_C[2.3] = \Pr[C \leq 2.3] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

Q: What is $\text{CDF}_C[5.8]$? .

1 since you have summed across the range.

Distribution Functions

Probability density function (PDF): Let R be a random variable with codomain V. The probability density function of R is the function $\text{PDF}_R : V \to [0,1]$, such that $\text{PDF}_R[x] = \Pr[R = x]$ if $x \in \text{Range(R)}$ and equal to zero if $x \notin \text{Range(R)}$.

$\sum_{x \in V} \text{PDF}_R[x] = \sum_{x \in \text{Range(R)}} \Pr[R = x] = 1$.

Cumulative distribution function (CDF): If the codomain is a subset of the real numbers, then the cumulative distribution function is the function $\text{CDF}_R : \mathbb{R} \to [0,1]$, such that $\text{CDF}_R[x] = \Pr[R \leq x]$.

Importantly, neither PDF_R nor CDF_R involves the sample space of an experiment.

Example: If we flip three coins, and C counts the number of heads, then
$\text{PDF}_C[0] = \Pr[C = 0] = \frac{1}{8}$, and
$\text{CDF}_C[2.3] = \Pr[C \leq 2.3] = \Pr[C = 0] + \Pr[C = 1] + \Pr[C = 2] = \frac{7}{8}$.

Q: What is $\text{CDF}_C[9.8]$?

Since most things are smaller than infinity, you will approach one.

Since most things are larger than -infinity, you will approach zero as you obtain less events.

For a general random variable R, as $x \to \infty$, $\text{CDF}_R[x] \to 1$ and $x \to -\infty$, $\text{CDF}_R[x] \to 0$.

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define T to be the random variable equal to the sum of the dice. Plot PDF_T and CDF_T

S: (1, 2, 3, 4, 5, 6)^2

T(2, 3) = 5

T: S -> V st T(i, j) = i + j

PDF_T: V -> [0, 1]

PDF(2) = Pr(T = 2) = Pr({1, 1}) = 1/36

PDF(3.2) = 0 since it is not part of the codomain

PDF[3] = Pr(T = 3) = Pr({1, 2}, {2, 1}) = 2/36

CDF: R -> [0, 1]

CDF(2.3) = Pr(c <= 2.3) = Pr(c = 2) = 1/36

CDF(3) = Pr(C <= 3) = Pr(C = 2) + Pr(C = 3) = 3/36

CDF(2.9999) = Pr(C <= 2.999) = Pr(C = 2)

CDF(3.00001) = Pr(C <= 3.00001) = Pr(C = 3) + Pr(C = 2)

Back to throwing dice

Q: Suppose we throw two standard dice one after the other. Let us define T to be the random variable equal to the sum of the dice. Plot PDF_T and CDF_T

Recall that $T : \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \to V$ where $V = \{2, 3, 4, \ldots 12\}$.

$\text{PDF}_T : V \to [0, 1]$ and $\text{CDF}_T : \mathbb{R} \to [0, 1]$.

For example, $\text{PDF}_T[4] = \Pr[T = 4] = \frac{3}{36}$ and $\text{PDF}_T[12] = \Pr[T = 12] = \frac{1}{36}$.

Back to throwing dice

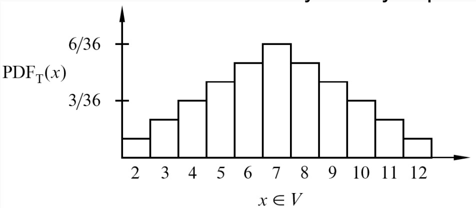
Q: Suppose we throw two standard dice one after the other. Let us define T to be the random variable equal to the sum of the dice. Plot PDF_T and CDF_T

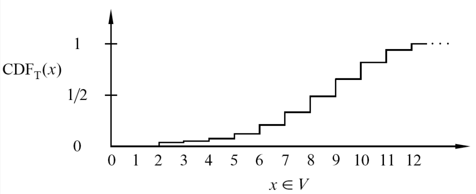
Recall that $T : \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\} \to V$ where $V = \{2, 3, 4, \ldots 12\}$.

$\mathrm{PDF}_T : V \to [0, 1]$ and $\mathrm{CDF}_T : \mathbb{R} \to [0, 1]$.

For example, $\mathrm{PDF}_T[4] = \Pr[T = 4] = \frac{3}{36}$ and $\mathrm{PDF}_T[12] = \Pr[T = 12] = \frac{1}{36}$.

May have you plot out CDF or PDR in exam





CDF for discrete distributions is like a step function

Questions?

Distributions

Many random variables turn out to have the same PDF and CDF. In other words, even though R and T might be different random variables on different probability spaces, it is often the case that $\mathrm{PDF}_R = \mathrm{PDF}_T$. Hence, by studying the properties of such PDFs, we can study different random variables and experiments.

Distributions

Many random variables turn out to have the same PDF and CDF. In other words, even though R and T might be different random variables on different probability spaces, it is often the case that $\text{PDF}_R = \text{PDF}_T$. Hence, by studying the properties of such PDFs, we can study different random variables and experiments.

Distribution over a random variable can be fully specified using the cumulative distribution function (CDF) (usually denoted by F). The corresponding probability density function (PDF) is denoted by f.

Distribution is interchangable with telling you the CDF or the PDF

Distributions

Many random variables turn out to have the same PDF and CDF. In other words, even though R and T might be different random variables on different probability spaces, it is often the case that $\text{PDF}_R = \text{PDF}_T$. Hence, by studying the properties of such PDFs, we can study different random variables and experiments.

Distribution over a random variable can be fully specified using the cumulative distribution function (CDF) (usually denoted by F). The corresponding probability density function (PDF) is denoted by f.

Common Discrete Distributions in Computer Science:

- Bernoulli Distribution

Next lecture will talk about the PDF and CDF of each of these distributions

- Uniform Distribution
- Binomial Distribution
- Geometric Distribution

Bernoulli Distribution

Canonical Example: We toss a biased coin such that the probability of getting a heads is p. Let R be the random variable such that $R = 1$ when the coin comes up heads and $R = 0$ if the coin comes up tails. R follows the Bernoulli distribution.

Bernoulli Distribution

Canonical Example: We toss a biased coin such that the probability of getting a heads is p. Let R be the random variable such that $R = 1$ when the coin comes up heads and $R = 0$ if the coin comes up tails. R follows the Bernoulli distribution.

PDF$_R$ for Bernoulli distribution: $f\colon \{0,1\} \to [0,1]$ meaning that Bernoulli random variables take values in $\{0,1\}$. It can be fully specified by the "probability of success" (of an experiment) p (probability of getting a heads in the example). Formally, PDF$_R$ is given by:

$$f(1) = p \quad ; \quad f(0) = q := 1 - p.$$

In the example, $\Pr[R = 1] = f(1) = p = \Pr[\text{event that we get a heads}]$.

Bernoulli Distribution

Canonical Example: We toss a biased coin such that the probability of getting a heads is p. Let R be the random variable such that $R = 1$ when the coin comes up heads and $R = 0$ if the coin comes up tails. R follows the Bernoulli distribution.

PDF$_R$ for Bernoulli distribution: $f\colon \{0,1\} \to [0,1]$ meaning that Bernoulli random variables take values in $\{0,1\}$. It can be fully specified by the "probability of success" (of an experiment) p (probability of getting a heads in the example). Formally, PDF$_R$ is given by:

$$f(1) = p \quad ; \quad f(0) = q := 1 - p.$$

In the example, $\Pr[R = 1] = f(1) = p = \Pr[\text{event that we get a heads}]$.

CDF$_R$ for Bernoulli distribution: $F : \mathbb{R} \to [0,1]$:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < 0\text{)} \\ &= 1 - p && \text{(for } 0 \leq x < 1\text{)} \\ &= 1 && \text{(for } x \geq 1\text{)} \end{aligned}$$

Uniform Distribution

Canonical Example: We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

Uniform Distribution

Canonical Example: We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

A random variable R that takes on each possible value in its codomain V with the same probability is said to be uniform.

Uniform Distribution

Canonical Example: We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

A random variable R that takes on each possible value in its codomain V with the same probability is said to be uniform.

PDF$_R$ for Uniform distribution: $f : V \to [0,1]$ such that for all $v \in V$, $f(v) = 1/|V|$. In the example, $f(1) = f(2) = \ldots = f(6) = \frac{1}{6}$.

Uniform Distribution

Canonical Example: We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

A random variable R that takes on each possible value in its codomain V with the same probability is said to be uniform.

PDF$_R$ for Uniform distribution: $f : V \to [0,1]$ such that for all $v \in V$, $f(v) = {}^{1}/_{|V|}$. In the example, $f(1) = f(2) = \ldots = f(6) = \frac{1}{6}$.

CDF$_R$ for Uniform distribution: For n elements in V arranged in increasing order – $(v_1, v_2, \ldots, v_n)$, the CDF is:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < v_1\text{)} \\ &= {}^{k}/_{n} && \text{(for } v_k \le x < v_{k+1}\text{)} \\ &= 1 && \text{(for } x \ge v_n\text{)} \end{aligned}$$

Uniform Distribution

Canonical Example: We roll a standard die. Let R be the random variable equal to the number that shows up on the die. R follows the uniform distribution.

A random variable R that takes on each possible value in its codomain V with the same probability is said to be uniform.

PDF$_R$ for Uniform distribution: $f : V \to [0,1]$ such that for all $v \in V$, $f(v) = 1/|V|$. In the example, $f(1) = f(2) = \ldots = f(6) = \frac{1}{6}$.

CDF$_R$ for Uniform distribution: For n elements in V arranged in increasing order – $(v_1, v_2, \ldots, v_n)$, the CDF is:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < v_1\text{)} \\ &= k/n && \text{(for } v_k \leq x < v_{k+1}\text{)} \\ &= 1 && \text{(for } x \geq v_n\text{)} \end{aligned}$$

Q: If X has a Bernoulli distribution, when is X also uniform?

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1 - p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\Pr[E_k] = \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots$$

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\begin{aligned} \Pr[E_k] &= \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots \\ &= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)} \end{aligned}$$

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\begin{aligned}
\Pr[E_k] &= \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots \\
&= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)} \\
&= p^k (1-p)^{n-k} + p^k (1-p)^{n-k} + \ldots
\end{aligned}$$

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\begin{aligned}
\Pr[E_k] &= \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots \\
&= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)} \\
&= p^k (1-p)^{n-k} + p^k (1-p)^{n-k} + \ldots
\end{aligned}$$

$$\implies \Pr[E_k] = \binom{n}{k} p^k (1-p)^{n-k}$$

(Number of terms = number of ways to choose the k tosses that result in heads = $\binom{n}{k}$)

Binomial Distribution

Canonical Example: We toss n biased coins independently. The probability of getting a heads for each coin is p. Let R be the random variable equal to the number of heads in the n coin tosses. R follows the Binomial distribution.

PDF$_R$ for Binomial distribution: $f : \{0, 1, 2, \ldots, n\} \to [0, 1]$. For $k \in \{0, 1, \ldots, n\}$, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Proof: Let E_k be the event we get k heads. Let A_i be the event we get a heads in toss i.

$$E_k = (A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup (A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap A_{k+2}^c \cap \ldots \cap A_n^c) \cup \ldots$$

$$\begin{aligned}
\Pr[E_k] &= \Pr[(A_1 \cap A_2 \ldots A_k \cap A_{k+1}^c \cap A_{k+2}^c \cap \ldots \cap A_n^c)] + \Pr[A_1^c \cap A_2 \ldots A_k \cap A_{k+1} \cap \ldots \cap] + \ldots \\
&= \Pr[A_1] \Pr[A_2] \Pr[A_k] \Pr[A_{k+1}^c] \Pr[A_{k+2}^c] \ldots \Pr[A_n^c] + \ldots \qquad \text{(Independence of tosses)} \\
&= p^k (1-p)^{n-k} + p^k (1-p)^{n-k} + \ldots
\end{aligned}$$

$$\implies \Pr[E_k] = \binom{n}{k} p^k (1-p)^{n-k}$$

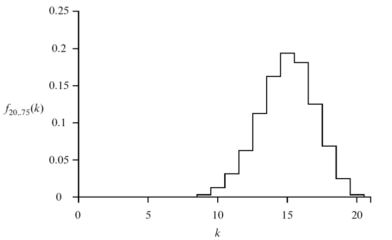
(Number of terms = number of ways to choose the k tosses that result in heads = $\binom{n}{k}$)

Binomial Distribution

For the Binomial distribution, $\mathrm{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.

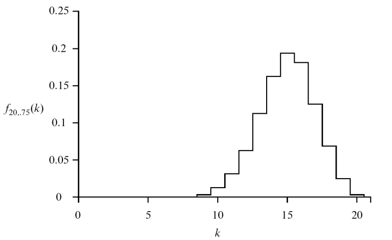
Binomial Distribution

For the Binomial distribution, $\text{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Binomial Distribution

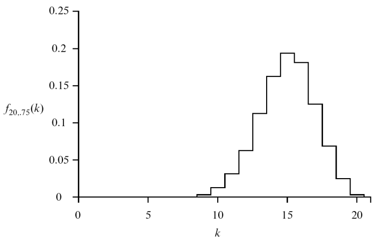
For the Binomial distribution, $\mathrm{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Q: Prove that $\sum_{k \in \mathrm{Range(R)}} \mathrm{PDF}_R[k] = 1$.

Binomial Distribution

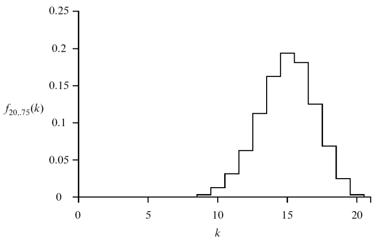
For the Binomial distribution, $\text{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Q: Prove that $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = 1$.
By the Binomial Theorem, $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$.

Binomial Distribution

For the Binomial distribution, $\text{PDF}_R(k) = \binom{n}{k} p^k (1-p)^{n-k}$.



Q: Prove that $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = 1$.
By the Binomial Theorem, $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p+1-p)^n = 1$.

CDF$_R$ for Binomial distribution: $F : \mathbb{R} \to [0,1]$:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < 0\text{)} \\ &= \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} && \text{(for } k \le x < k+1\text{)} \\ &= 1. && \text{(for } x \ge n\text{)} \end{aligned}$$

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1 - p)^{k-1}\, p$.

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1 - p)^{k-1} p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1-p)^{k-1}\, p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$E_k = A_1^c \cap A_2^c \cap \ldots \cap A_k$$

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1 - p)^{k-1}\, p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c]\Pr[A_2^c] \ldots \Pr[A_k] \quad \text{(Independence of tosses)} \end{aligned}$$

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \dots\} \to [0, 1]$. For $k \in \{1, 2, \dots, \infty\}$, $f(k) = (1 - p)^{k-1}\, p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \dots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \dots \cap A_k] = \Pr[A_1^c] \Pr[A_2^c] \dots \Pr[A_k] \quad \text{(Independence of tosses)} \\ \implies \Pr[E_k] &= (1 - p)^{k-1} p \end{aligned}$$

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1-p)^{k-1}\, p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c] \Pr[A_2^c] \ldots \Pr[A_k] \quad \text{(Independence of tosses)} \\ \implies \Pr[E_k] &= (1-p)^{k-1} p \end{aligned}$$

Q: Prove that $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = 1$.

Geometric Distribution

Canonical Example: We toss a biased coin independently multiple times. The probability of getting a heads is p. Let R be the random variable equal to the number of tosses needed to get the first heads. R follows the geometric distribution.

PDF$_R$ for Geometric distribution: $f : \{1, 2, \ldots\} \to [0, 1]$. For $k \in \{1, 2, \ldots, \infty\}$, $f(k) = (1 - p)^{k-1}\, p$.

Proof: Let E_k be the event that we need k tosses to get the first heads. Let A_i be the event that we get a heads in toss i.

$$\begin{aligned} E_k &= A_1^c \cap A_2^c \cap \ldots \cap A_k \\ \Pr[E_k] &= \Pr[A_1^c \cap A_2^c \cap \ldots \cap A_k] = \Pr[A_1^c]\Pr[A_2^c]\ldots\Pr[A_k] \quad \text{(Independence of tosses)} \\ \implies \Pr[E_k] &= (1-p)^{k-1} p \end{aligned}$$

Q: Prove that $\sum_{k \in \text{Range(R)}} \text{PDF}_R[k] = 1$.

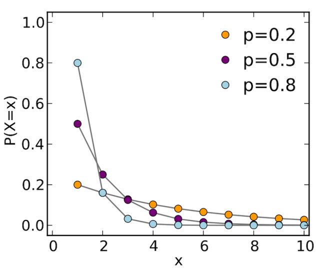
By the sum of geometric series, $\sum_{k \in \text{Range}(R)} \text{PDF}_R[k] = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \frac{p}{1-(1-p)} = 1$.

Geometric Distribution

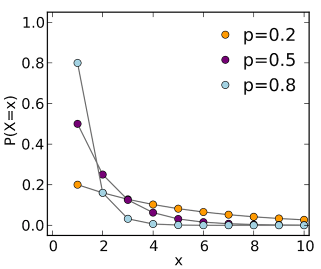
For the Geometric distribution, $\mathrm{PDF}_R(k) = (1-p)^{k-1}p$.

Geometric Distribution

For the Geometric distribution, $\mathrm{PDF}_R(k) = (1-p)^{k-1}p$.



Geometric Distribution



For the Geometric distribution, $\text{PDF}_R(k) = (1-p)^{k-1}p$.

CDF$_R$ for Geometric distribution: $F : \mathbb{R} \to [0,1]$:

$$\begin{aligned} F(x) &= 0 && \text{(for } x < 1\text{)} \\ &= \sum_{i=1}^{k} (1-p)^{i-1}p && \text{(for } k \le x < k+1\text{)} \end{aligned}$$

Questions?