

# CMPT210

## Probability and Computing \*

Evan F.

November 2024

### Contents

<b>1 Resources &amp; Links</b>	<b>4</b>
<b>2 Functions</b>	<b>4</b>
2.1 Total	4
2.2 Surjective	4
2.3 Injective	4
2.4 Bijective	4
2.5 k-to-1	4
<b>3 Counting</b>	<b>4</b>
3.1 Product Rule	4
3.1.1 Generalization	4
3.2 Division Rule	4
3.3 Combinations (Subsets)	5
3.3.1 With Repetition	5
3.4 Permutations	5
3.4.1 With Repetition	5
3.5 Binomial Theorem	5
3.5.1 Multinomial Theorem	5
3.6 Pigeonhole Principle	5
<b>4 Basics of Probability</b>	<b>5</b>
4.1 Complement Rule	5
4.2 Mutual Exclusivity	5
4.3 Union	5
4.4 Inclusion Exclusion	6
4.5 Conditional Probability	6
4.6 Total Probability	6
4.7 Union Bound	6
<b>5 Bayes Theorem</b>	<b>6</b>
5.1 Probability Trees	6
<b>6 Independence</b>	<b>6</b>
6.1 Relation to Mutual Exclusivity	6
6.2 Pairwise Independence	6
6.3 Mutual Independence	7

---

\*Special thanks to Randy Goffe, Al Gore, Bill Gates, Alex, Al Ghiorso, Steve Jobs, Admiral James D. Watkins, Simon Fraser (the person, not the university), Michael Dell, and Michael Bloomberg. It is 12:19:10 AM. Good morning.

<b>7</b>	<b>Frievald's</b>	<b>7</b>
7.1	Probability Amplification . . . . .	7
<b>8</b>	<b>Random Variables</b>	<b>7</b>
8.1	Indicator Random Variables . . . . .	7
8.2	Pairwise Independence . . . . .	7
8.3	Mutual Independence . . . . .	7
<b>9</b>	<b>PDF</b>	<b>8</b>
<b>10</b>	<b>CDF</b>	<b>8</b>
<b>11</b>	<b>Distribution</b>	<b>8</b>
11.1	Bernoulli . . . . .	8
11.2	Uniform . . . . .	8
11.3	Binomial . . . . .	9
11.4	Geometric . . . . .	10
<b>12</b>	<b>Expectation</b>	<b>11</b>
12.1	Bernoulli . . . . .	11
12.2	Uniform . . . . .	11
12.3	Binomial . . . . .	12
12.4	Geometric . . . . .	12
12.5	Expectation of Transformations . . . . .	12
12.6	Conditional Expectation . . . . .	12
12.7	Law of Total Expectation . . . . .	13
12.8	Linearity of Expectation . . . . .	13
12.9	Expectations of Products . . . . .	13
<b>13</b>	<b>Max Cut via Erdos' Algorithm</b>	<b>13</b>
13.1	Erdos' Algorithm . . . . .	14
13.2	Claim . . . . .	14
13.3	Proof . . . . .	14
<b>14</b>	<b>Randomized Quick Select</b>	<b>15</b>
14.1	Algorithm . . . . .	15
14.2	Claim . . . . .	15
14.3	Proof . . . . .	15
14.3.1	Lemma . . . . .	15
14.3.2	Induction . . . . .	16
<b>15</b>	<b>Joint Distribution</b>	<b>16</b>
15.1	Summing Out (Marginalization) . . . . .	16
<b>16</b>	<b>Variance</b>	<b>16</b>
16.1	Properties of Variance . . . . .	17
16.1.1	Constants . . . . .	17
16.1.2	Linearity of Variance . . . . .	17
16.2	Bernoulli . . . . .	18
16.3	Uniform . . . . .	18
16.4	Geometric . . . . .	19
16.5	Binomial . . . . .	20
<b>17</b>	<b>Standard Deviation</b>	<b>20</b>
<b>18</b>	<b>Covariance</b>	<b>20</b>
18.1	Formulae . . . . .	20

<b>19 Correlation</b>	<b>21</b>
19.1 Correlation's Bounds (Copied)	21
<b>20 Markov's Theorem</b>	<b>21</b>
20.1 With a Lower Bound	22
<b>21 Chebyshevs's Theorem</b>	<b>22</b>
21.1 Offset	23
21.2 Lone Tails	23
21.2.1 Tighter Bound	23
21.3 Pairwise Independent Sampling	23
21.4 Weak Law of Large Numbers	24
<b>22 Chernoff Bound</b>	<b>24</b>
<b>23 Problems</b>	<b>24</b>
23.1 Coupon Collector Problem	24
23.2 Birthday Paradox	25
23.3 Simpson's Paradox	25
23.4 Voter Poll	25
23.5 Randomized Load Balancing	26
23.5.1 Background	26
23.5.2 $\Pr[T_i > 600]$	26
23.5.3 Considering All Servers	26
<b>24 Addendum</b>	<b>27</b>
24.1 Expansion of $(x_1 + x_3 + \dots + x_n)^2$	27
24.2 $x - 1 \leq e^{-x}$	27
<b>25 Formulas &amp; Summary</b>	<b>28</b>
25.1 Functions	28
25.2 Total	28
25.3 Surjective	28
25.4 Injective	28
25.5 Bijective	28
25.6 k-to-1	28
25.7 Probability	28
25.7.1 Independence	28
25.8 Random Variables	28
25.9 Expectation	29
25.10 Joint Distribution	29
25.11 Variance	29
25.12 Standard Deviation	29
25.13 Covariance	30
25.14 Correlation	30
25.15 Markov	30
25.16 Chebyshev	30
25.17 Chernoff	30
25.18 Other	30
<b>26 Wrapping Up</b>	<b>31</b>

# 1 Resources & Links

<https://www.cs.cmu.edu/~odonnell/papers/probability-and-computing-lecture-notes.pdf>  
<https://vaswanis.github.io/210-W24/L25.pdf>  
<https://vaswanis.github.io/210-F24.html>  
<https://aka.amogsus.com/onenote-macm201>

## 2 Functions

Assigns elements in the domain to another in the codomain.

### 2.1 Total

A total function is one that maps every value in the domain to something in the codomain.

### 2.2 Surjective

For all  $b \in B$ , there exists an  $a \in A$  that points to that  $b$

$f : A \rightarrow B; \forall b \in B \exists a \in A \text{ s.t. } f(a) = b$

$|A| \geq |B|$

### 2.3 Injective

All  $a \in A$  map to a unique  $b \in B$ .

$f : A \rightarrow B; \forall a \in A \exists \text{ unique } b \in B \text{ s.t. } f(a) = b$

If  $f(a) = f(b), a = b$

$|A| \leq |B|$

### 2.4 Bijective

If its both surjective and injective.

$|A| = |B|$

### 2.5 k-to-1

Maps exactly  $k$  elements of the domain to every element of the codomain. If  $f : A \rightarrow B$  is a  $k$ -to-1 function,  $|A| = k|B|$ .

## 3 Counting

### 3.1 Product Rule

Where  $M, C, S$  are sets,

$$|M \times C \times S| = |M| \times |C| \times |S|$$

#### 3.1.1 Generalization

Generalized product rule: If  $S$  is the set of length  $k$  sequences such that the first entry can be selected in  $n_1$  ways, after the first entry is chosen, the second one can be chosen in  $n_2$  ways, and so on, then  $|S| = n_1 \times n_2 \times \dots \times n_k$ . If  $n_1 = n_2 = n_3 = \dots = n_k$ , we recover the product rule.

### 3.2 Division Rule

Application of  $k$ -to-1 functions. Map a problem to the  $k$ -to-1 function, such as ears to people. The set  $|\text{Ears}| = 2|\text{People}|$ .

### 3.3 Combinations (Subsets)

The number of subsets of length  $k$  from a set of length  $n$  is  $\binom{n}{k}$ .

#### 3.3.1 With Repetition

The number of combinations (with repetition) of getting  $n$  total items given  $k$  type of items is  $\binom{n+k-1}{k-1}$

### 3.4 Permutations

The number of ways to permute a string of length  $n$ , where all  $n$  items are distinct is  $n!$ .

#### 3.4.1 With Repetition

The number of ways to permute a string of length  $n$ , with  $a$  of the same token,  $b$  of a another token, etc, is

$$\frac{n!}{a!b!c! \dots} = \binom{n}{a, b, c}$$

### 3.5 Binomial Theorem

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}$$

#### 3.5.1 Multinomial Theorem

$$(a + b + c + \dots)^n = \sum_{k_1+k_2+\dots=n} \binom{n}{k_1, k_2, k_3, \dots} a_1^{k_1} b_2^{k_2} c_3^{k_3} \dots$$

### 3.6 Pigeonhole Principle

If there are  $n$  and  $> n$  pigeons in the holes, at least 1 hole has more than 1 pigeon.

## 4 Basics of Probability

There are a bunch of things related to probability, mostly the most fundamental concepts, that I cannot be bother to add here.

### 4.1 Complement Rule

$$\Pr[E] = 1 - \Pr[E^c]$$

### 4.2 Mutual Exclusivity

For events  $E_1, E_2, E_3, \dots, E_n$ , they are mutually exclusive if for any pair of events  $i \neq j$ ,  $E_i \cap E_j = \emptyset$

### 4.3 Union

If  $E_1, E_2, E_3, \dots, E_n$  are mutually exclusive, then

$$\Pr[E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n] = \Pr[E_1] + \Pr[E_2] + \Pr[E_3] + \dots + \Pr[E_n]$$

## 4.4 Inclusion Exclusion

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

## 4.5 Conditional Probability

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(A|B)\Pr(B) = \Pr(A \cap B) = \Pr(B|A)\Pr(A)$$

## 4.6 Total Probability

$$\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$$

## 4.7 Union Bound

$$\Pr[A_1 \cup A_2 \cup \dots \cup A_n] \leq \Pr[A_1] + \Pr[A_2] + \dots + \Pr[A_n]$$

# 5 Bayes Theorem

$$\Pr(E|A) = \frac{\Pr(E \cap A)}{\Pr(A)} = \frac{\Pr(A|E)\Pr(E)}{\Pr(E)} = \frac{\Pr(A|E)\Pr(A)}{\Pr(A|E)\Pr(E) + \Pr(A|E^c)\Pr(E^c)}$$

## 5.1 Probability Trees

If you see this text, then I've been too lazy to add content here. Read the Carnegie Mellon University notes from [1](#), it's probably in there.

# 6 Independence

One event does not affect another.

$$\Pr(X \cap Y) = \Pr(X)\Pr(Y)$$

$$\Pr(X|Y) = \Pr(X)$$

$$\Pr(Y|X) = \Pr(Y)$$

## 6.1 Relation to Mutual Exclusivity

Independence **DOES NOT** imply mutual exclusivity nor vice versa.

Independence:

$$\Pr(X \cap Y) = \Pr(X)\Pr(Y)$$

Mutual Exclusivity:

$$\Pr(X \cap Y) = 0$$

## 6.2 Pairwise Independence

Events  $(E_1, E_2, E_3, \dots)$  are pairwise independent iff for every pair of events  $E_i$  and  $E_j$  ( $i \neq j$ ),  $\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2)$ ;  $\Pr(E_1|E_2) = \Pr(E_1)$ ;  $\Pr(E_2|E_1) = \Pr(E_2)$ .

### 6.3 Mutual Independence

Events  $(E_1, E_2, E_3, \dots)$  are mutually independent iff for every subset  $S \subseteq \{1, 2, \dots, n\}$ ,  $\Pr[\bigcap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i]$ .

## 7 Frievald's

Verify matrix multiplication, i.e. if  $AB = D$  for given  $n \times n$ -dimension matrices  $A, B, D$ .

1. Generate an  $n$ -bit vector, randomly, with each position independently being a 1 or 0.
2. Computing  $ABx = Dx$  by first computing  $Bx$ , then  $A(Bx)$ , and then comparing to  $Dx$
3. If the resulting  $n$ -bit vector has the same value for every position, the algorithm returns 1, else 0

If  $AB = D$ , the algorithm always returns 1.

If  $AB \neq D$ , the algorithm returns the correct answer with probability of more than  $\frac{1}{2}$

	Pr Algorithm Returns $D = AB$	Pr Algorithm Returns $D \neq AB$
$D = AB$	1	0
$D \neq AB$	$\leq \frac{1}{2}$	$\geq \frac{1}{2}$

Proof of Frievald's is not provided here. I'm too lazy.

### 7.1 Probability Amplification

By repeating the Frievald's algorithm  $i$  times,

$$\begin{aligned} \Pr(\text{Outputs no} \mid \text{True answer is no}) &= 1 - \Pr(\text{Outputs yes} \mid \text{True answer if no}) \\ &= 1 - \Pr(A_1 = \text{Yes} \cap \dots \cap A_i = \text{Yes} \mid \text{True answer is no}) \\ &\geq 1 - \prod_{k=0}^i \frac{1}{2} \end{aligned}$$

## 8 Random Variables

A total function whose domain is the sample space  $S$ . The co-domain is typically  $\subseteq \mathbb{R}$ . It maps outcomes  $(E \in S)$  to the co-domain.

### 8.1 Indicator Random Variables

An indicator random variable  $I_A$  for an event  $A$  is defined as:

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{if } A \text{ does not occur.} \end{cases}$$

### 8.2 Pairwise Independence

Pairwise independent iff for every  $x_1 \in \text{Range}(R_1)$  and  $x_2 \in \text{Range}(R_2)$ ,  $\Pr(R_1 = i \cap R_2 = j) = \Pr(R_1 = i) \Pr(R_2 = j)$ . Yes, this can be very tedious to prove. For an exam, I'd expect a question to be about proving its not pairwise independent, so just find a counter example.

### 8.3 Mutual Independence

Many random variables are mutually independent iff for every  $x_1 \in \text{Range}(R_1), x_2 \in \text{Range}(R_2), x_3 \in \text{Range}(R_3), \dots$ , events  $[R_1 = x_1], [R_2 = x_2], [R_3 = x_3]$  are mutually independent.

## 9 PDF

Probability density function

$$\begin{aligned} \text{PDF}_X(k) &= \Pr(X = k) && \text{if } k \in \text{Range}(X) \\ \text{PDF}_X(k) &= 0 && \text{if } k \notin \text{Range}(X) \end{aligned}$$

$$\sum_{x \in \text{Range}(X)} \Pr[X = x] = 1$$

## 10 CDF

Cumulative distribution function

$$\text{CDF}_X(k) = \Pr[X \leq k]$$

For  $R \rightarrow$  random variable to the value of a roll of a die,

$$\text{CDF}_X(4.5) = \Pr[X \leq 4.5] = \Pr(X = 1) + \Pr(X = 2) + \Pr(X = 3) + \Pr(X = 4) = \frac{2}{3}$$

## 11 Distribution

### 11.1 Bernoulli

$$\begin{aligned} \text{Range}(X) &= \{0, 1\} \\ \text{PDF}_X(1) &= p \\ \text{PDF}_X(0) &= 1 - p \\ \text{CDF}_X(x) &= 0 && \text{for } x < 0 \\ &= 1 - p && \text{for } 0 \leq x \leq 1 \\ &= 1 && \text{for } x \geq 1 \end{aligned}$$

### 11.2 Uniform

$$\begin{aligned} \text{Range}(X) &= \{v_1, v_2, \dots, v_n\} \\ \text{PDF}_X(x) &= \frac{1}{|X|} \\ \text{CDF}_X(x) &= 0 && \text{for } x < v_1 \\ &= \frac{k}{n} && \text{for } v_k \leq x < v_{k+1} \\ &= 1 && \text{for } x \geq v_n \end{aligned}$$



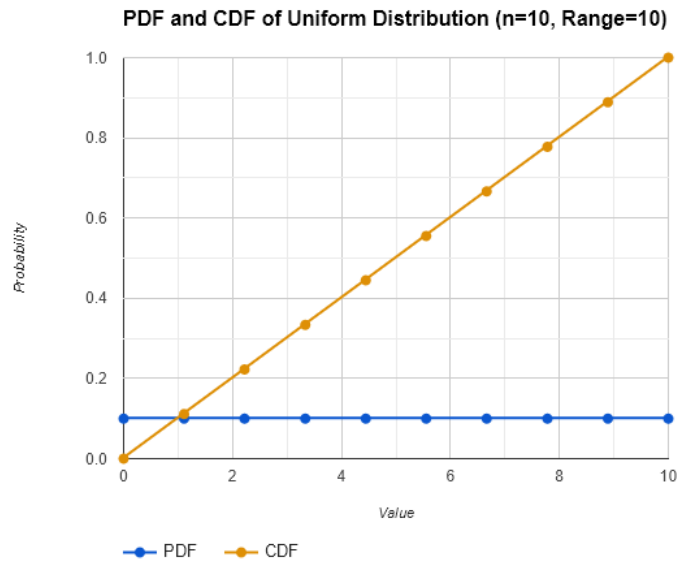


Figure 1: Uniform PDF & CDF

### 11.3 Binomial

$$\text{Range}(X) = \{0, 1, \dots, n\}$$

$$\text{PDF}_X(x) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{Binomial Theorem}$$

$$\text{CDF}_X(x) = 0 \quad \text{for } x < 0$$

$$= \sum_{i=0}^k \binom{i}{k} p^i (1-p)^{n-i} \quad \text{for } k \leq x < k+1$$

$$= 1 \quad \text{for } x \geq n$$

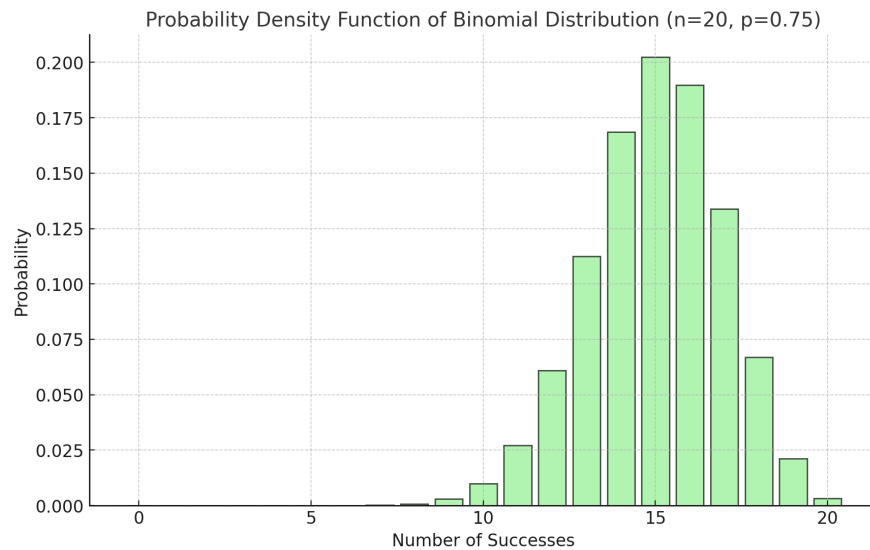


Figure 2: Binomial PDF

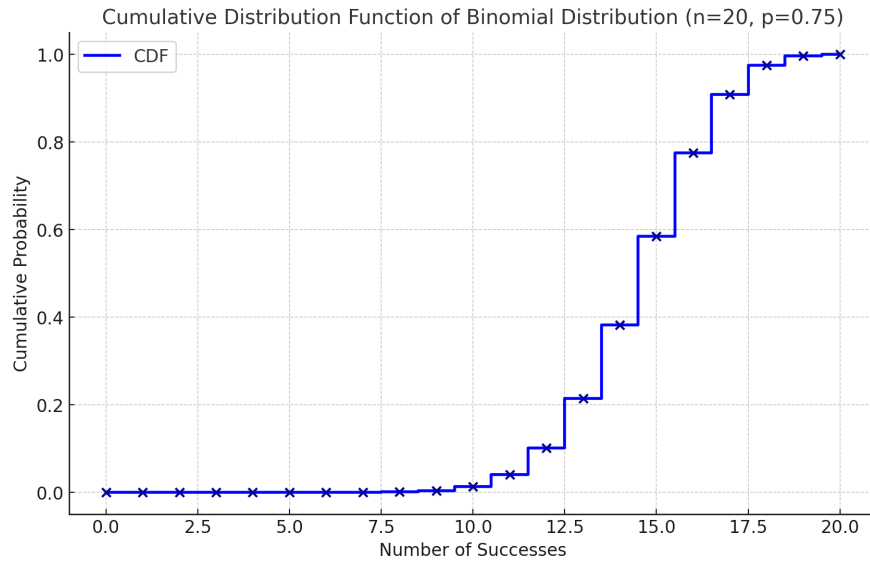


Figure 3: Binomial CDF

## 11.4 Geometric

$$\text{Range}(X) = \{0, 1, \dots, \infty\}$$

$$\text{PDF}_X(x) = (1 - p)^{k-1} p$$

$$\text{CDF}_X(x) = 0$$

for  $x < 0$

$$= \sum_{i=1}^k (1 - p)^{i-1} p$$

for  $k \leq x < k + 1$

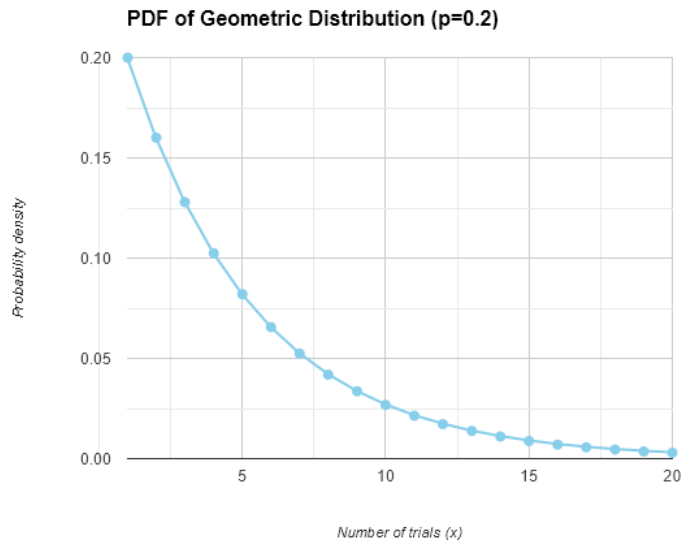


Figure 4: Geometric PDF

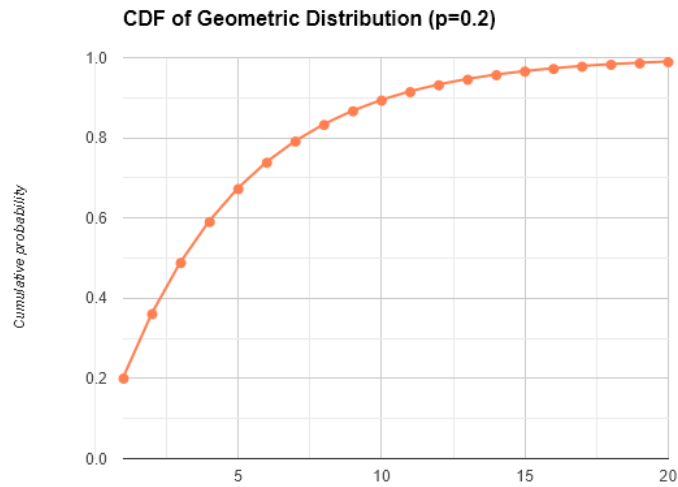


Figure 5: Geometric CDF

## 12 Expectation

The average output value of a random variable.

$$\mathbb{E}(X) = \sum_{x \in \text{Range}(X)} x \Pr(X = x)$$

And also

$$\mathbb{E}(X) = \sum_{w \in S} \Pr[w] R[w]$$

### 12.1 Bernoulli

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x \in \{0,1\}} x \Pr(X = x) \\ &= 0 \cdot \Pr(X = 0) + 1 \cdot \Pr(X = 1) \\ &= \Pr(X = 1) \\ &= p \end{aligned}$$

### 12.2 Uniform

$$\begin{aligned} \mathbb{E}(X) &= \sum_{x \in \{v_1, v_2, \dots, v_n\}} x \Pr(X = x) \\ &= \frac{1}{n} [v_1 + v_2 + \dots + v_n] \end{aligned}$$

### 12.3 Binomial

Let  $X \sim \text{Bin}(n, p)$  be composed of  $n$  indicator variables, each being  $\Pr(X = 1) = p$ .

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X_1 + X_2 + X_3 + \dots + X_n) \\ &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) + \dots + \mathbb{E}(X_n) \\ &= n \cdot p\end{aligned}$$

### 12.4 Geometric

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x \in \{1, 2, 3, \dots, \infty\}} x \Pr(X = x) \\ &= \mathbb{E}(X|A) \Pr(A) + \mathbb{E}(X|A^c) \Pr(A^c) \\ &= 1 \cdot p + \mathbb{E}(X|A^c)(1 - p) \\ \mathbb{E}(X|A^c) &= \sum_{i=1}^{\infty} i \Pr(X = i|A) \\ &= \sum_{i=0}^{\infty} i \Pr(X = i - 1) \\ &= \sum_{i=0}^{\infty} (i + 1) \Pr(X = i) \\ &= \sum_{i=0}^{\infty} i \Pr(X = i) + \sum_{i=0}^{\infty} \Pr(X = i) \\ &= \mathbb{E}(X) + 1 \\ \mathbb{E}(X) &= 1 \cdot p + (\mathbb{E}(X) + 1)(1 - p) \\ &= p + \mathbb{E}(X)(1 - p) + (1 - p) \\ &= p + \mathbb{E}(X) - \mathbb{E}(X)p + 1 - p \\ &= \mathbb{E}(X) - \mathbb{E}(X)p + 1 \\ &\rightarrow \\ \mathbb{E}(X) - \mathbb{E}(X) + \mathbb{E}(X)p &= 1 \\ \mathbb{E}(X)p &= 1 \\ &\rightarrow \\ \mathbb{E}(X) &= \frac{1}{p}\end{aligned}$$

### 12.5 Expectation of Transformations

$$\mathbb{E}(g(X)) = \sum_{x \in \text{Range}(X)} g(x) \Pr(X = x)$$

### 12.6 Conditional Expectation

$$\mathbb{E}(X|A) = \sum_{x \in \text{Range}(X)} x \Pr(X = x|A)$$

## 12.7 Law of Total Expectation

If  $A_1, A_2, A_3, \dots, A_n$  form partitions of  $S$ , then

$$\begin{aligned}\mathbb{E}[R] &= \sum_{x \in \text{Range}(R)} x \Pr[R = x] \\ &= \sum_{x \in \text{Range}(R)} x \sum_i \Pr[R = x | A_i] \Pr[A_i] \\ &= \sum_i \Pr[A_i] \sum_{x \in \text{Range}(R)} x \Pr[R = x | A_i] \\ &= \sum_i \mathbb{E}[R | A_i] \Pr[A_i]\end{aligned}$$

## 12.8 Linearity of Expectation

$R = R_1 + R_2 + \dots + R_n$

$$\begin{aligned}\mathbb{E}[R] &= \sum_{x \in S} \Pr[x] R[x] \\ &= \sum_{x \in S} [\Pr[x] (R_1[x] + R_2[x] + R_3[x] + \dots + R_n[x])] \\ &= \sum_{x \in S} [\Pr[x] R_1[x] + \Pr[x] R_2[x] + \Pr[x] R_3[x] + \dots + \Pr[x] R_n[x]] \\ &= \sum_{x \in S} \Pr[x] R_1[x] + \sum_{x \in S} \Pr[x] R_2[x] + \sum_{x \in S} \Pr[x] R_3[x] + \dots + \sum_{x \in S} \Pr[x] R_n[x] \\ &= \sum_{i=1}^n \sum_{x \in S} \Pr[x] R_i[x] \\ &= \sum_{i=1}^n \mathbb{E}(R_i)\end{aligned}$$

## 12.9 Expectations of Products

The only case we should be expected to use this directly are if  $A$  and  $B$  are independent. In assignments, if they were not independent, we solved for them instead of calculating.

$$\begin{aligned}\mathbb{E}[AB] &= \sum_{a \in \text{Range}(A)} \sum_{b \in \text{Range}(B)} a \cdot b \cdot \Pr[A = a \cap B = b] \\ &= \sum_{a \in \text{Range}(A)} \sum_{b \in \text{Range}(B)} a \cdot b \cdot \Pr[A = a] \Pr[B = b] && \text{If } A, B \text{ pairwise independent} \\ &= \sum_{a \in \text{Range}(A)} a \cdot \Pr[A = a] \sum_{b \in \text{Range}(B)} b \cdot \Pr[B = b] \\ &= \mathbb{E}[A] \mathbb{E}[B]\end{aligned}$$

$$\mathbb{E}[ABC] = \mathbb{E}[A] \mathbb{E}[B] \mathbb{E}[C] \quad \text{If mutually independent}$$

## 13 Max Cut via Erdos' Algorithm

Goal: Partition a graph  $G = (V, E)$  into two sets of vertices such that the number of vertices between the two sets is maximized. This is an NP-hard problem. We're looking for an approximate solution,

such that if  $OPT$  is the optimal solution with the maximized amount of edges, our solution will have at least half as many ( $\alpha = 0.5$  for  $\alpha OPT$ ).

Let  $\delta(U)$  return the set of edges between the set  $U$  and  $U^c$ , such that  $U \cup U^c = V$ .

### 13.1 Erdos' Algorithm

For each vertex, independently and randomly choose to put it in  $U$  with probability  $\frac{1}{2}$ .

### 13.2 Claim

The expected number of edges will be greater than or equal to half of the optimal solution.

$$|\delta(U)| \geq 0.5 OPT$$

### 13.3 Proof

Let  $X_{u,v}$  be a random variable equal to 1 iff the edge  $(u, v)$  is in  $\delta(U)$ .

$$\begin{aligned} \mathbb{E}(|\delta(U)|) &= \mathbb{E}\left(\sum_{(u,v) \in E} X_{u,v}\right) \\ &= \sum_{(u,v) \in E} \mathbb{E}(X_{u,v}) \\ &= \sum_{(u,v) \in E} \Pr(X_{u,v} = 1) \end{aligned}$$

$$\begin{aligned} \Pr(X_{u,v} = 1) &= \Pr((u, v) \in \delta(U)) \\ &= \Pr((u \in U \cap v \notin U) \cup (u \notin U \cap v \in U)) \\ &= \Pr(u \in U \cap v \notin U) + \Pr(u \notin U \cap v \in U) \\ &= \Pr(u \in U)\Pr(v \notin U) + \Pr(u \notin U)\Pr(v \in U) \\ &= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} \end{aligned}$$

$$\mathbb{E}(|\delta(U)|) = \sum_{(u,v) \in E} \frac{1}{2} = \frac{|E|}{2} \geq \frac{OPT}{2}$$

→

$$\mathbb{E}(|\delta(U)|) \geq 0.5 \cdot OPT$$

## 14 Randomized Quick Select

### 14.1 Algorithm

---

**Algorithm 1** Randomized Quick Select

---

```
1: Input: Array  $A$  of  $n$  distinct numbers, integer  $k \in [1, n]$ 
2: Output: The  $k$ th smallest element in  $A$ 
3: function QUICKSELECT( $A, k$ )
4:   if Length( $A$ ) = 1 then
5:     return  $A[1]$ 
6:   end if
7:   Select  $p \in A$  uniformly at random
8:   Construct sets  $Left := \{x \in A \mid x < p\}$  and  $Right := \{x \in A \mid x > p\}$ 
9:    $r \leftarrow |Left| + 1$  ▷ Element  $p$  is the  $r$ th smallest element in  $A$ 
10:  if  $k = r$  then
11:    return  $p$ 
12:  else if  $k < r$  then
13:    return QUICKSELECT( $Left, k$ )
14:  else
15:    return QUICKSELECT( $Right, k - r$ )
16:  end if
17: end function
```

---

In plain English: Pick a random element  $j$  in the array to be the pivot. Sort the rest of the elements into a set that is smaller and another for those that are bigger. The [size of the set for smaller numbers] + 1 gives the position for that element  $j$ . If it is the  $k$ th element, we are done. Otherwise, inspect either the set for smaller or bigger numbers depending on if the  $k$ th element must be bigger or smaller than  $j$ .

### 14.2 Claim

This algorithm is better than sorting via Quick Sort or another  $\mathcal{O}(n \log(n))$  algorithm. Randomized Quick Select performs, on average, fewer than  $8n$  comparisons for finding the  $k$ th ( $k \in [n]$ ) element from an array with  $n$  elements.

### 14.3 Proof

#### 14.3.1 Lemma

The expected size of the child (the left or right set if  $k \neq r$ ) is  $< \frac{7n}{8}$ . Define a good event  $\varepsilon$  to be if the array is split in about half; if  $r \in (\frac{n}{4}, \frac{3n}{4}]$  or between the first and third quartile. The probability of this is a half. If  $\varepsilon$  happens,  $|Left| < \frac{3n}{4}$  and  $|Right| < \frac{3n}{4}$ , so  $|Child| < \frac{3n}{4}$ . If  $\varepsilon$  doesn't happen, the worst case is  $|Child| < n$ .

$$\begin{aligned}\mathbb{E}[|Child| \mid \varepsilon] &< \frac{3n}{4} \\ \mathbb{E}[|Child| \mid \varepsilon^c] &< n \\ \Pr(\varepsilon) &= \frac{1}{2}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[|Child|] &= \mathbb{E}[|Child| \mid \varepsilon] \Pr(\varepsilon) + \mathbb{E}[|Child| \mid \varepsilon^c] \Pr(\varepsilon^c) \\ &< \frac{3n}{4} \cdot \frac{1}{2} + n \cdot \frac{1}{2} = \frac{7n}{8}\end{aligned}$$

### 14.3.2 Induction

**Base Case:** If  $n = 1$ ,  $0 < 8 \cdot 1$ .

**(Strong) Induction:** Assuming for all  $m < n$ ,  $[E][\text{comparisons for size } m \text{ array}] < 8m$ .

$$\begin{aligned}\mathbb{E}[n \text{ size array}] &= \mathbb{E}[n - 1 + \text{comparisons in child problem}] \\ &= \mathbb{E}[n - 1] + \mathbb{E}[\text{comparisons in child problem}] \\ &< n - 1 + 8 \cdot [E][\text{Child}] \\ &< n - 1 + 8 \cdot \frac{7n}{8} \\ &= n - 1 + 7n = 8n - 1 < 8n\end{aligned}$$

## 15 Joint Distribution

Used for seeing how random variable relate to each other.

$$\text{PDF}_{X,Y}[x, y] = \Pr(X = x \cap Y = y)$$

### 15.1 Summing Out (Marginalization)

$$\text{PDF}_X[x] = \sum_{y \in \text{Range}(Y)} \text{PDF}_{X,Y}[x, y] = \sum_{y \in \text{Range}(Y)} \Pr(X = x \cap Y = y)$$

## 16 Variance

Variance is the expected value of the squared deviation from the mean of a random variable. It is a measure of dispersion, meaning it is a measure of how far a set of numbers is spread out from their average value.

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}((X - \mathbb{E}(X))^2) \\ &= \sum_{x \in \text{Range}(X)} (i - \mu)^2 \Pr(X = i) \\ &= \sum_{x \in \text{Range}(X)} (i^2 - 2\mu i + \mu^2) \Pr(X = i) \\ &= \sum_{x \in \text{Range}(X)} i^2 \Pr(X = i) - 2 \sum_{x \in \text{Range}(X)} \mu i \Pr(X = i) + \sum_{x \in \text{Range}(X)} \mu^2 \Pr(X = i) \\ &= \mathbb{E}(X^2) - 2\mu \sum_{x \in \text{Range}(X)} i \Pr(X = i) + \mu^2 \sum_{x \in \text{Range}(X)} \Pr(X = i) \\ &= \mathbb{E}(X^2) - 2\mu \mathbb{E}(X) + \mu^2 \cdot 1 \\ &= \mathbb{E}(X^2) - 2\mu^2 + \mu^2 \\ &= \mathbb{E}(X^2) - \mu^2 \\ &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$



## 16.1 Properties of Variance

### 16.1.1 Constants

$$\begin{aligned}
\text{Var}[aX + b] &= \mathbb{E}[(aR + b)^2] - \mathbb{E}[aR + b]^2 \\
&= (\mathbb{E}[a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2]) - \mathbb{E}[aR] + \mathbb{E}[b]^2 \\
&= (a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2) - (a\mathbb{E}[R] + \mathbb{E}[b])^2 \\
&= (a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2) - (a^2\mathbb{E}[R]^2 + 2ab\mathbb{E}[R] + b^2) \\
&= a^2\mathbb{E}[R^2] + 2ab\mathbb{E}[R] + b^2 - a^2\mathbb{E}[R]^2 - 2ab\mathbb{E}[R] - b^2 \\
&= a^2\mathbb{E}[R^2] - a^2\mathbb{E}[R]^2 \\
&= a^2[\mathbb{E}[R^2] - (\mathbb{E}[R])^2] \\
&= a^2\text{Var}[X]
\end{aligned}$$

### 16.1.2 Linearity of Variance

If  $R_1, R_2, R_3, \dots, R_n$  are pairwise independent,

$$\begin{aligned}
R &= R_1 + R_2 + \dots + R_n \\
\text{Var}[R] &= \mathbb{E}[R^2] - \mathbb{E}[R]^2 \\
&= \mathbb{E}[(R_1 + R_2 + \dots + R_n)^2] - \mathbb{E}[R_1 + R_2 + \dots + R_n]^2 \\
&= \mathbb{E}[(R_1 + R_2 + \dots + R_n)^2] - (\mathbb{E}[R_1] + \mathbb{E}[R_2] + \dots + \mathbb{E}[R_n])^2 \\
&= \mathbb{E}\left[\sum_{i=1}^n R_i^2 + 2 \sum_{1 \leq i < j \leq n} R_i R_j\right] - (\mathbb{E}[R_1] + \mathbb{E}[R_2] + \dots + \mathbb{E}[R_n])^2 \\
&= \left[ \sum_{i=1}^n \mathbb{E}[R_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i R_j] \right] - (\mathbb{E}[R_1] + \mathbb{E}[R_2] + \dots + \mathbb{E}[R_n])^2 \\
&= \left[ \sum_{i=1}^n \mathbb{E}[R_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i R_j] \right] - \left[ \sum_{i=1}^n \mathbb{E}[R_i]^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i] \mathbb{E}[R_j] \right] \\
&= \sum_{i=1}^n \mathbb{E}[R_i^2] + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i R_j] - \sum_{i=1}^n \mathbb{E}[R_i]^2 - 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i] \mathbb{E}[R_j] \\
&= \sum_{i=1}^n \mathbb{E}[R_i^2] - \sum_{i=1}^n \mathbb{E}[R_i]^2 + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i R_j] - 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[R_i] \mathbb{E}[R_j] \\
&= \sum_{i=1}^n [\mathbb{E}[R_i^2] - \mathbb{E}[R_i]^2] + 2 \sum_{1 \leq i < j \leq n} [\mathbb{E}[R_i R_j] - \mathbb{E}[R_i] \mathbb{E}[R_j]] \\
&= \sum_{i=1}^n \text{Var}[R_i] + 2 \sum_{1 \leq i < j \leq n} \text{Cov}[R_i, R_j] \\
&= \sum_{i=1}^n \text{Var}[R_i] + 2 \sum_{1 \leq i < j \leq n} 0 \\
&= \sum_{i=1}^n \text{Var}[R_i]
\end{aligned}$$

## 16.2 Bernoulli

$$\begin{aligned}\mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \mathbb{E}(X^2) - p^2 \\ \mathbb{E}(X^2) &= 0^2 \cdot \Pr(X = 0) + 1^2 \cdot \Pr(X = 1) \\ &= \Pr(X = 1) \\ &= p \\ \mathbb{E}(X^2) - \mathbb{E}(X)^2 &= p - p^2 = p(1 - p) \\ \text{Var}(X) &= p(1 - p)\end{aligned}$$

## 16.3 Uniform

If  $x \sim \text{Uniform}(\{v_1, v_2, \dots, v_n\})$ ,

$$\begin{aligned}\mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ \mathbb{E}(X^2) &= \sum_{i=0}^n X_i^2 \Pr(X = i) \\ &= \sum_{i=0}^n X_i^2 \frac{1}{n} \\ &= \frac{1}{n} \sum_{i=0}^n X_i^2 \\ &= \frac{1}{n} [X_1^2 + X_2^2 + \dots + X_n^2] \\ &= \frac{1}{n} [X_1^2 + X_2^2 + \dots + X_n^2] \\ \mathbb{E}(X) &= \frac{1}{n} [X_1 + X_2 + \dots + X_n] \\ \text{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= \left( \frac{1}{n} [X_1^2 + X_2^2 + \dots + X_n^2] \right) - \left( \frac{1}{n} [X_1 + X_2 + \dots + X_n] \right)^2\end{aligned}$$

## 16.4 Geometric

$$\begin{aligned}\mathbb{E}(X^2) &= \mathbb{E}(X^2|A)\Pr(A) + \mathbb{E}(X^2|A^c)\Pr(A) \\ &= 1 \cdot p + \mathbb{E}(X^2|A^c)(1-p)\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X^2|A^c) &= \sum_{i=0}^{\infty} i^2 \Pr(X=i|A) \\ &= \sum_{i=0}^{\infty} i^2 \Pr(X=i-1) \\ &= \sum_{i=0}^{\infty} (i+1)^2 \Pr(X=i) \\ &= \sum_{i=0}^{\infty} (i^2 + 2i + 1) \Pr(X=i) \\ &= \sum_{i=0}^{\infty} i^2 \Pr(X=i) + 2 \sum_{i=0}^{\infty} i \Pr(X=i) + \sum_{i=0}^{\infty} \Pr(X=i) \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(X) + 1\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X^2) &= 1 \cdot p + (\mathbb{E}(X^2) + 2\mathbb{E}(X) + 1)(1-p) \\ &= 1 \cdot p + \mathbb{E}(X^2)(1-p) + 2\mathbb{E}(X)(1-p) + (1-p) \quad \rightarrow \\ \mathbb{E}(X^2) - \mathbb{E}(X^2)(1-p) &= 1 \cdot p + 2\mathbb{E}(X)(1-p) + (1-p) \\ &\rightarrow \\ \mathbb{E}(X^2)[1 - (1-p)] &= 1 \cdot p + 2\mathbb{E}(X)(1-p) + (1-p) \\ &\rightarrow \\ \mathbb{E}(X^2)p &= 1 \cdot p + 2\frac{1}{p}(1-p) + (1-p) \\ &\rightarrow \\ \mathbb{E}(X^2) &= 1 + \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{p}{p} \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p}\end{aligned}$$

$$\begin{aligned}\mathbb{E}(X^2) - \mathbb{E}(X)^2 &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 \\ &= \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} \\ &= \frac{2(1-p)}{p^2} + \frac{p}{p^2} - \frac{1}{p^2} \\ &= \frac{2(1-p)}{p^2} + \frac{p-1}{p^2} \\ &= \frac{2(1-p)}{p^2} - \frac{1-p}{p^2} \\ &= \frac{(1-p)}{p^2}\end{aligned}$$

## 16.5 Binomial

$$\begin{aligned}X &= X_1 + X_2 + \dots \\ \text{Var}(X_i) &= p(1-p) \\ \text{Var}(X) &= \text{Var}(X_1 + X_2 + \dots) \\ &= \text{Var}(X_1) + \text{Var}(X_2) + \dots \\ &= p(1-p) + p(1-p) + \dots \\ &= np(1-p)\end{aligned}$$

## 17 Standard Deviation

Describes how much, on average, values deviate from the mean (expected value) of the random variable. This value is in the same units as the random variable.

$$\sigma_X = \sqrt{\text{Var}(X)}$$

## 18 Covariance

Intuition:

$$\text{Cov}[R, S] = \mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])]$$

Why does this make sense for measuring how two variables are related?

$$\begin{aligned}(R - \mathbb{E}[R]) \\ (S - \mathbb{E}[S])\end{aligned}$$

These each find how far they are from the mean of the random variable. Unlike variance, these can be negative.

$$\begin{aligned}\mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])] &= \sum_{i \in \text{Range}(R)} \sum_{j \in \text{Range}(S)} (r_i - \mathbb{E}[R])(s_j - \mathbb{E}[S])\mathbb{P}(R = r_i \cap S = s_j) \\ &= \sum_{i \in \text{Range}(R)} \sum_{j \in \text{Range}(S)} (r_i - \mu_r)(s_j - \mu_s)\mathbb{P}(R = r_i \cap S = s_j)\end{aligned}$$

If  $(r_i - \mu_r)$  is negative and  $(s_j - \mu_s)$  is positive, they are negatively correlated. Additionally, for some additional info, correlation is only defined for pairs of pair (a tuple) of the R and S r.vs.

### 18.1 Formulae

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

$$\begin{aligned}\text{Cov}[R, S] &= \mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])] \\ &= \mathbb{E}[RS - \mathbb{E}[R]S - \mathbb{E}[S]R + \mathbb{E}[R]\mathbb{E}[S]] \\ &= \mathbb{E}[RS] - \mathbb{E}[\mathbb{E}[R]S] - \mathbb{E}[\mathbb{E}[S]R] + \mathbb{E}[\mathbb{E}[R]\mathbb{E}[S]] \\ &= \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] - \mathbb{E}[S]\mathbb{E}[R] + \mathbb{E}[R]\mathbb{E}[S] \\ &= \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S] - \mathbb{E}[S]\mathbb{E}[R] + \mathbb{E}[R]\mathbb{E}[S] \\ &= \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S]\end{aligned}$$

## 19 Correlation

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \\ -1 \leq \text{Corr}(X, Y) \leq 1$$

-1 means negative correlation; as one random variable goes up, the other goes down. 1 means positive correlation; as one random variable goes up, the other goes up too. The closer to -1 or 1 it is, the stronger the relationship. 0 means no correlation.

### 19.1 Correlation's Bounds (Copied)

The correlation coefficient is defined as:

$$r = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y},$$

where

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)].$$

By the Cauchy-Schwarz inequality:

$$\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]^2 \leq \mathbb{E}[(X - \mu_X)^2] \cdot \mathbb{E}[(Y - \mu_Y)^2].$$

This implies:

$$\text{Cov}(X, Y)^2 \leq \sigma_X^2 \sigma_Y^2.$$

Dividing through by  $\sigma_X^2 \sigma_Y^2$ :

$$\left( \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \right)^2 \leq 1,$$

so:

$$-1 \leq \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \leq 1.$$

Thus:

$$-1 \leq r \leq 1.$$

## 20 Markov's Theorem

For a non-negative random variable and  $x \geq 0$ .

$$x\mathcal{I}\{X \geq x\} \\ \rightarrow \\ x\mathcal{I}\{X \geq x\} \leq X$$

If  $X \geq x$ , then it follows  $x \cdot 1 \leq X$ . If  $X < x$ , then  $0 \leq X$

$$\mathbb{E}[x\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \\ x\mathbb{E}[\mathcal{I}\{X \geq x\}] \leq \mathbb{E}[X] \\ x\text{Pr}[X \geq x] \leq \mathbb{E}[X] \\ \text{Pr}[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$$

## 20.1 With a Lower Bound

If  $b$  is a lower bound on  $R$ , then Markov's Theorem can be tightened to:

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R] - b}{x - b}$$

Proof: Let  $Y = R - b$ . If  $R \geq b$  then  $Y \geq 0$ .

$$\begin{aligned} \Pr[Y \geq y] &\leq \frac{\mathbb{E}[Y]}{y} = \frac{\mathbb{E}[R] - b}{y} \\ &\rightarrow \\ \Pr[R - b \geq y] &= \Pr[R \geq y + b] \leq \frac{\mathbb{E}[R] - b}{y} \\ &\rightarrow \\ &\text{Let } y = x - b \\ \Pr[R \geq x] &\leq \frac{\mathbb{E}[R] - b}{x - b} \end{aligned}$$

## 21 Chebyshev's Theorem

Does not require the random variable to be positive. Requires variance. For a  $x > 0$ .

$$\begin{aligned} \Pr[X \geq x] &\leq \frac{\mathbb{E}[X]}{x} \\ \text{Choose } X &\text{ to be } |X - \mathbb{E}(X)|^2 \\ \text{Choose } x &\text{ to be } y^2 \\ \Pr[|X - \mathbb{E}(X)|^2 \geq y^2] &\leq \frac{\mathbb{E}[|X - \mathbb{E}(X)|^2]}{y^2} \\ \Pr[|X - \mathbb{E}(X)|^2 \geq y^2] &\leq \frac{\mathbb{E}[|X - \mathbb{E}(X)|^2]}{y^2} \\ \Pr[|X - \mathbb{E}(X)|^2 \geq y^2] &\leq \frac{\mathbb{E}[(X - \mathbb{E}(X))^2]}{y^2} \\ \sqrt{|X - \mathbb{E}(X)|^2} &\geq \sqrt{y^2} \\ \Pr(|X - \mathbb{E}(X)| \geq y) &\leq \frac{\text{Var}[X]}{y^2} \\ 1 - \Pr(|X - \mathbb{E}(X)| \geq y) &\geq 1 - \frac{\text{Var}[X]}{y^2} \\ \Pr(|X - \mathbb{E}(X)| < y) &\geq 1 - \frac{\text{Var}[X]}{y^2} \end{aligned}$$

There is also the version where you get the  $\Pr(X \geq \mathbb{E}(X) - c \cdot \sigma \cup X \leq \mathbb{E}(X) + c \cdot \sigma)$

## 21.1 Offset

$$\begin{aligned}\Pr[X \geq x] \\ \Pr[X - \mathbb{E}(X) \geq x - \mathbb{E}(X)] \\ \Pr[X - \mathbb{E}(X) \geq y] &\leq \Pr[|X - \mathbb{E}(X)| \geq y] \\ \Pr(|X - \mathbb{E}(X)| \geq y) &\leq \frac{\sigma^2}{y^2}\end{aligned}$$

For the other side,

$$\begin{aligned}1 - \Pr(|X - \mathbb{E}(X)| \geq y) &\geq 1 - \frac{\sigma^2}{y^2} \\ \Pr(|X - \mathbb{E}(X)| < y) &\geq 1 - \frac{\sigma^2}{y^2}\end{aligned}$$

## 21.2 Lone Tails

$$\begin{aligned}\Pr(|X - \mathbb{E}(X)| \geq y) &\leq \frac{\text{Var}[X]}{y^2} \\ \Pr(X - \mathbb{E}(X) \leq -y) + \Pr(X - \mathbb{E}(X) \geq y) &\leq \frac{\text{Var}[X]}{y^2}\end{aligned}$$

It follows that

$$\Pr(X - \mathbb{E}(X) \geq y) \leq \frac{\text{Var}[X]}{y^2}$$

and

$$\Pr(X - \mathbb{E}(X) \leq -y) \leq \frac{\text{Var}[X]}{y^2}$$

### 21.2.1 Tighter Bound

Proven in `Final-Practice-sols.pdf`, a tighter bound for the one-sided Chebyshev's Theorem is

$$\Pr[R - \mathbb{E}[R] \leq x] \geq \frac{\text{Var}[R]}{x^2 + \text{Var}[R]}$$

## 21.3 Pairwise Independent Sampling

For a  $S_n = G_1 + G_2 + G_3 + \dots + G_n$  where  $G_i$  are pairwise independent and have a) the same mean ( $\mu$ ) and b) the same standard deviation  $\sigma$ .

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n} \left(\frac{\sigma}{\epsilon}\right)^2$$

$$\begin{aligned}\mathbb{E}\left(\frac{S_n}{n}\right) &= \frac{1}{n} \cdot \mathbb{E}(S_n) = \frac{1}{n} \cdot \mathbb{E}\left(\sum_{i=1}^n G_i\right) = \frac{1}{n} \cdot \sum_{i=1}^n \mathbb{E}(G_i) = \frac{1}{n} \cdot n \cdot \mu = \mu \\ \text{Var}\left(\frac{S_n}{n}\right) &= \frac{1}{n^2} \text{Var}(S_n) = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n G_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[G_i] = \frac{1}{n^2} \sum_{i=1}^n \sigma^2 = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

$$\begin{aligned}\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] &= \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right] \\ &\leq \frac{\sigma^2}{n\epsilon^2}\end{aligned}$$

## 21.4 Weak Law of Large Numbers

Result from above.

$$\begin{aligned}\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] &= \Pr\left[\left|\frac{S_n}{n} - \mathbb{E}\left(\frac{S_n}{n}\right)\right| \geq \epsilon\right] \\ &\leq \frac{\sigma^2}{n\epsilon^2}\end{aligned}$$

As  $n \rightarrow \infty$ ,

$$\begin{aligned}\Pr\left[\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right] &\geq 1 - \frac{\sigma^2}{n\epsilon^2} \\ \lim_{n \rightarrow \infty} \Pr\left[\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right] &= 1 - 0 = 1\end{aligned}$$

## 22 Chernoff Bound

Can be used when the random variable of interest is a sum of other random variables. If,

1.  $T_1, T_2, \dots, T_n$  are **mutually independent** random variables
2.  $0 \leq T_i \leq 1$  is satisfied for all  $i$
3.  $T = \sum_{i=1}^n T_i$

then, for all  $c \geq 1$ ,

$$\begin{aligned}\beta(c) &= c \ln(c) - c + 1 \\ \Pr[T \geq c\mathbb{E}[T]] &\leq e^{-\beta(c)\mathbb{E}[T]}\end{aligned}$$

No proof is provided. The  $T_i$  don't have to be the same distribution, they just need to satisfy the requirements above. It describes how far the random variable is from its expectations with a multiplicative factor.

## 23 Problems

### 23.1 Coupon Collector Problem

L16

A café has a program where if you collect all  $n$  colors of coupons (where you get 1 at random with each order) you get a free coffee. How many orders do you need to make, on average, to get that free coffee? Let  $X$  be the amount of orders needed, and  $X_i$  the amount of order you need to go from having  $i - 1$  coupons to getting the  $i$ th coupon.



$$\begin{aligned}
X &= X_1 + X_2 + \dots + X_n \\
\mathbb{E}[X] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \dots + \mathbb{E}[X_n] \\
\Pr[\text{new coupon}] &= \frac{n - (i - 1)}{n} \\
\mathbb{E}[X_i] &= \frac{n}{n - i + 1} & X_i &\sim \text{Geo}\left(\frac{n - (i - 1)}{n}\right) \\
\sum_{i=1}^n \frac{n}{n - i + 1} &= n \cdot \sum_{i=1}^n \frac{1}{n - i + 1} \\
&\leq 1 + \int_1^n \frac{1}{x} dx \\
&= n[1 + \ln(n)]
\end{aligned}$$

## 23.2 Birthday Paradox

L6, L7

What is the probability there are two students in the class with the same birthday?

$$\begin{aligned}
\Pr[\text{Two students, same birthday}] &= 1 - \Pr[\text{No students with the same birthday}] \\
\Pr[\text{No students with the same birthday}] &= \frac{d(d-1)(d-2)\dots(d-(n-1))}{d^n} \\
\Pr[\text{Two students, same birthday}] &= 1 - \left(1 - \frac{0}{d}\right) \left(1 - \frac{1}{d}\right) \left(1 - \frac{2}{d}\right) \dots
\end{aligned}$$

Apply  $1 - x \leq e^{-x}$  for  $x > 0$  to obtain a closed form bound.

## 23.3 Simpson's Paradox

L10

## 23.4 Voter Poll

L22, L23

We want to find how many voters we need to survey to get a high confidence in the results, say within 0.05 of the fraction  $p$  with a confidence of 95%. By surveying, we are summing the results (Harris vs Trump) and dividing by the number of voters surveyed to get the estimate probability  $p$ . This can be represented with

$$S_n = V_1 + V_2 + \dots + V_n$$

We are asking

$$\Pr\left[\left|\frac{S_n}{n} - p\right| \leq 0.05\right] \geq 0.95$$

Calculating the variance requires knowing  $p$ , unless you just bound the maximum variance by maximizing  $p(1 - p)$ . Each  $V_i$  are mutually independent, enough for linearity of variance. Then, just use this approximate variance on Chebyshev's theorem.

## 23.5 Randomized Load Balancing

L24

### 23.5.1 Background

Units of work (seconds it takes to process a post) is between 0 and 1. How many servers are needed so no server is overloaded? Work is randomly split between servers. 24000 posts per 600 seconds, average of 0.25 seconds per post. Let  $T$  be the random variable for the amount of work assigned to server 1. If  $T > 600$ , it is overloaded. Let  $m$  be the number of servers.

### 23.5.2 $\Pr[T_i > 600]$

The total amount of work the server is assigned can be represented with  $T$ , and composed by  $T = T_1 + T_2 + T_3 + \dots + T_{24000}$ . Calculate the probability the first server is overloaded, using Chernoff bound.

1.  $T$  is a sum of  $T_i$
2. All units of work are mutually independent
3. All  $T_i$  are between 0 and 1

Solve for  $\mathbb{E}(T)$ . Let  $A$  be the event post  $i$  is assigned to this first server.

$$\begin{aligned}\mathbb{E}(T_i) &= \mathbb{E}(T_i|A)\Pr[A] + \mathbb{E}(T_i|A^c)\Pr[A^c] \\ &= \frac{1}{4} \frac{1}{m} + 0(1 - \frac{1}{m}) \\ &= \frac{1}{4m}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[T] &= \sum_{i=1}^{24000} \frac{1}{4m} \\ &= \frac{6000}{m}\end{aligned}$$

Solve  $\Pr[T_i > 600]$  using Chernoff.

$$\begin{aligned}\Pr[T_i > 600] &\leq \Pr[T_i \geq 600] \\ &= \Pr[T_i \geq \frac{m}{10} \mathbb{E}[T_i]] \\ &\leq e^{\beta(\frac{m}{10}) \cdot \mathbb{E}[T_i]}\end{aligned}$$

### 23.5.3 Considering All Servers

$$\begin{aligned}\Pr[\text{A server overloaded}] &= \Pr[T_1 > 600 \cup T_2 > 600 \cup \dots] \\ &\leq \sum_{i=0}^m \Pr[T_i > 600] \\ &= m \cdot \Pr[T_i > 600]\end{aligned}$$

$$\begin{aligned}\Pr[\text{No server overloaded}] &= 1 - \Pr[\text{A server overloaded}] \\ &\geq 1 - m \cdot \Pr[T_i > 600]\end{aligned}$$

## 24 Addendum

### 24.1 Expansion of $(x_1 + x_2 + \dots + x_n)^2$

$$\begin{aligned}(x_1 + x_2 + \dots + x_n)^2 &= \sum_{i=1}^n x_i^2 + 2 \sum_{1 \leq i < j \leq n} x_i x_j \\ &= \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i x_j\end{aligned}$$

### 24.2 $x - 1 \leq e^{-x}$

For  $x > 0$ .

## 25 Formulas & Summary

### 25.1 Functions

Assigns elements in the domain to another in the codomain.

### 25.2 Total

A total function is one that maps every value in the domain to something in the codomain.

### 25.3 Surjective

For all  $b \in B$ , there exists an  $a \in A$  that points to that  $b$

$$f : A \rightarrow B; \forall b \in B \exists a \in A \text{ s.t. } f(a) = b$$

$$|A| \geq |B|$$

### 25.4 Injective

All  $a \in A$  map to a unique  $b \in B$ .

$$f : A \rightarrow B; \forall a \in A \exists \text{ unique } b \in B \text{ s.t. } f(a) = b$$

If  $f(a) = f(b)$ ,  $a = b$

$$|A| \leq |B|$$

### 25.5 Bijective

If its both surjective and injective.

$$|A| = |B|$$

### 25.6 k-to-1

Maps exactly k elements of the domain to every element of the codomain. If  $f : A \rightarrow B$  is a k-to-1 function,  $|A| = k|B|$ .

### 25.7 Probability

$$\Pr[E] = 1 - \Pr[E^c]$$

$$\Pr[E_1 \cup E_2 \cup E_3 \cup \dots \cup E_n] = \Pr[E_1] + \Pr[E_2] + \Pr[E_3] + \dots + \Pr[E_n] \quad \text{Mutually exclusive}$$

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

$$\Pr(A|B)\Pr(B) = \Pr(A \cap B) = \Pr(B|A)\Pr(A)$$

$$\Pr(A) = \Pr(A|B)\Pr(B) + \Pr(A|B^c)\Pr(B^c)$$

$$\Pr[A_1 \cup A_2 \cup \dots \cup A_n] \leq \Pr[A_1] + \Pr[A_2] + \dots + \Pr[A_n]$$

$$\Pr(E|A) = \frac{\Pr(E \cap A)}{\Pr(A)} = \frac{\Pr(A|E)\Pr(E)}{\Pr(A)} = \frac{\Pr(A|E)\Pr(E)}{\Pr(A|E)\Pr(E) + \Pr(A|E^c)\Pr(E^c)}$$

#### 25.7.1 Independence

Events  $(E_1, E_2, E_3, \dots)$  are pairwise independent iff for every pair of events  $E_i$  and  $E_j$  ( $i \neq j$ ),  $\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2)$ ;  $\Pr(E_1|E_2) = \Pr(E_1)$ ;  $\Pr(E_2|E_1) = \Pr(E_2)$ .

Events  $(E_1, E_2, E_3, \dots)$  are mutually independent iff for every subset  $S \subseteq \{1, 2, \dots, n\}$ ,  $\Pr[\cap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i]$ .

### 25.8 Random Variables

Pairwise independent iff for every  $x_1 \in \text{Range}(R_1)$  and  $x_2 \in \text{Range}(R_2)$ ,  $\Pr(R_1 = x_1 \cap R_2 = x_2) = \Pr(R_1 = x_1)\Pr(R_2 = x_2)$ . Many random variables are mutually independent iff for every  $x_1 \in \text{Range}(R_1), x_2 \in \text{Range}(R_2), x_3 \in \text{Range}(R_3), \dots$ , events  $[R_1 = x_1], [R_2 = x_2], [R_3 = x_3]$  are mutually independent.

## 25.9 Expectation

$$X \sim \text{Ber}(p) \rightarrow \mathbb{E}(X) = p$$

$$X \sim \text{Uniform}(\{v_1, v_2, \dots, v_n\}) \rightarrow \mathbb{E}(X) = \frac{1}{n}[X_1 + X_2 + \dots + X_n]$$

$$X \sim \text{Geo}(p) \rightarrow \mathbb{E}(X) = \frac{1}{p}$$

$$X \sim \text{Bin}(n, p) \rightarrow \mathbb{E}(X) = np$$

$$\mathbb{E}(g(X)) = \sum_{x \in \text{Range}(X)} g(x) \Pr(X = x)$$

$$\mathbb{E}(X|A) = \sum_{x \in \text{Range}(X)} x \Pr(X = x|A)$$

$$\mathbb{E}[R] = \sum_i \mathbb{E}[R|A_i] \Pr[A_i]$$

If  $A_1, A_2, A_3, \dots, A_n$  form partitions of S

$$\mathbb{E}[R] = \sum_{i=1}^n \mathbb{E}(R_i)$$

$$\mathbb{E}[AB] = \sum_{a \in \text{Range}(A)} \sum_{b \in \text{Range}(B)} a \cdot b \cdot \Pr[A = a \cap B = b]$$

$$\mathbb{E}[AB] = \mathbb{E}[A]\mathbb{E}[B]$$

If pairwise independent

$$\mathbb{E}[ABC] = \mathbb{E}[A]\mathbb{E}[B]\mathbb{E}[C]$$

If mutually independent

## 25.10 Joint Distribution

$$\text{PDF}_{X,Y}[x, y] = \Pr(X = x \cap Y = y)$$

$$\text{PDF}_X[x] = \sum_{y \in \text{Range}(Y)} \text{PDF}_{X,Y}[x, y] = \sum_{y \in \text{Range}(Y)} \Pr(X = x \cap Y = y)$$

## 25.11 Variance

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

$$\text{Var}[aX + b] = a^2 \text{Var}[X]$$

$$\text{Var}[R] = \sum_{i=1}^n \text{Var}[R_i] + 2 \sum_{1 \leq i < j \leq n} \text{Cov}[R_i, R_j]$$

$$\text{Var}[R] = \sum_{i=1}^n \text{Var}[R_i]$$

If pairwise independent

$$X \sim \text{Ber}(p) \rightarrow \text{Var}(X) = p(1 - p)$$

$$X \sim \text{Uniform}(\{v_1, v_2, \dots, v_n\}) \rightarrow \text{Var}(X) =$$

$$\left( \frac{1}{n}[X_1^2 + X_2^2 + \dots + X_n^2] \right) - \left( \frac{1}{n}[X_1 + X_2 + \dots + X_n] \right)^2$$

$$X \sim \text{Geo}(p) \rightarrow \text{Var}(X) = \frac{(1 - p)}{p^2}$$

$$X \sim \text{Bin}(n, p) \rightarrow \text{Var}(X) = np(1 - p)$$

## 25.12 Standard Deviation

$$\sigma_X = \sqrt{\text{Var}(X)}$$

### 25.13 Covariance

$$\text{Cov}[R, S] = \mathbb{E}[(R - \mathbb{E}[R])(S - \mathbb{E}[S])] = \mathbb{E}[RS] - \mathbb{E}[R]\mathbb{E}[S]$$

$$\text{Cov}[R, S] = 0$$

If R, S independent

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}[X, Y]$$

### 25.14 Correlation

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

### 25.15 Markov

For a non-negative random variable and  $x > 0$ :

$$\Pr[X \geq x] \leq \frac{\mathbb{E}[X]}{x}$$

$$\Pr[R \geq x] \leq \frac{\mathbb{E}[R] - b}{x - b}$$

If R has a lower bound of  $b$

### 25.16 Chebyshev

For a random variable  $X$  and all  $x > 0$ :

$$\Pr(|X - \mathbb{E}(X)| \geq y) \leq \frac{\text{Var}[X]}{y^2}$$

$$\Pr(|X - \mathbb{E}(X)| < y) \geq 1 - \frac{\text{Var}[X]}{y^2}$$

$$\Pr[R - \mathbb{E}[R] \leq x] \geq \frac{\text{Var}[R]}{x^2 + \text{Var}[R]}$$

One-sided

$$\Pr\left[\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right] \leq \frac{1}{n} \left(\frac{\sigma}{\epsilon}\right)^2$$

For  $S_i$  all pairwise independent, same  $\mu$ , same  $\sigma$

$$\lim_{n \rightarrow \infty} \Pr\left[\left|\frac{S_n}{n} - \mu\right| \leq \epsilon\right] = 1$$

For  $S_i$  all pairwise independent, same  $\mu$ , same  $\sigma$

### 25.17 Chernoff

1.  $T_1, T_2, \dots, T_n$  are **mutually independent** random variables
2.  $0 \leq T_i \leq 1$  is satisfied for all  $i$
3.  $T = \sum_{i=1}^n T_i$
4.  $c \geq 1$

$$\beta(c) = c \ln(c) - c + 1$$

$$\Pr[T \geq c\mathbb{E}[T]] \leq e^{-\beta(c)\mathbb{E}[T]}$$

### 25.18 Other

$$(x_1 + x_2 + \dots + x_n)^2 = \sum_{i=1}^n x_i^2 + 2 \sum_{1 \leq i < j \leq n} x_i x_j = \sum_{i=1}^n x_i^2 + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n x_i x_j$$

## 26 Wrapping Up

In total, this document has taken somewhere above 15 hours, probably 20, to construct and write. Was it worth it? Probably not. Forever and ever, may your days be merry and your life be meaningful.