Probability Theory
A $\sigma$-algebra $\mathcal{F}$ is a collection of subsets of the sample space $\Omega$ such that the following requirements are satisfied:

**S1** The empty set is contained in $\mathcal{F}$.

**S2** If a set $E$ is contained in $\mathcal{F}$, then its complement $E^c$ is contained in $\mathcal{F}$.

**S3** The countable union of sets in $\mathcal{F}$ is contained in $\mathcal{F}$. 
Let $\mathcal{F}$ be a $\sigma$-algebra over the sample space $\Omega$. A **probability measure** on $\mathcal{F}$ is a function $\text{Pr}: \mathcal{F} \to [0, 1]$ satisfying

**P1** The certain event satisfies $\text{Pr}[\Omega] = 1$.

**P2** If the events $E_1, E_2, \ldots$ in $\mathcal{F}$ are mutually disjoint, then

$$\text{Pr}\left[\bigcup_{k=1}^{\infty} E_k\right] = \sum_{k=1}^{\infty} \text{Pr}[E_k].$$
Exercise

The smallest (with respect to inclusion) non-empty events belonging to a $\sigma$-algebra $\mathcal{F}$ are called **atoms**. Show that if $\mathcal{F}$ is a finite $\sigma$-algebra, then each event $A$ in $\mathcal{F}$ is the union of finitely many atoms.
Solution

Seeking a contradiction, suppose that $C$ is an event in $\mathcal{F}$ that is not a union of finitely many atoms.

Let $A$ denote the family of all atoms of $\mathcal{F}$. Let $B = \bigcup A$.

Since $\mathcal{F}$ is finite, the event $C \setminus B$ must contain an atomic event $A$. However, this is impossible, since $B$ is the (finite) union of all atomic events.
Random Variables
Definition of a Random Variable

Definition

Let $\mathcal{F}$ be a $\sigma$-algebra over the sample space $\Omega$. A random variable $X$ is a function $X: \Omega \rightarrow \mathbb{R}$ such that the preimage $X^{-1}(B)$ of each Borel set $B$ in $\mathbb{R}$ is an event in $\mathcal{F}$.

It suffices to show that

$$\{z \in \Omega \mid X(z) \leq x\}$$

is an event contained in $\mathcal{F}$ for all $x \in \mathbb{R}$. 
Let \((\Omega, \mathcal{F})\) be a measurable space.

Let \(A\) be a subset of \(\Omega\). Then the indicator function \(I_A : (\Omega, \mathcal{F}) \to \mathbb{R}\) given by

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

is a random variable if and only if \(A \in \mathcal{F}\). We call \(I_A\) the **indicator random variable** of the event \(A\).
A random variable is called **simple** if and only if it is a linear combination of a finite number of indicator random variables with disjoint support.

In other words, if $X$ is a simple random variable, then there exist pairwise disjoint events $A_1, \ldots, A_n$ and real numbers $s_1, \ldots, s_n$ such that

$$X = \sum_{k=1}^{n} s_k I_{A_k}.$$ 

Any nonnegative random variable can be approximated by a sequence of simple random variables.
A **discrete random variable** is a random variable with countable range, which means that the set \( \{ X(z) \mid z \in \Omega \} \) is countable.

The convenience of a discrete random variable \( X \) is that one can define events in terms of values of \( X \), for instance in the form \( X \in A \) which is short for

\[
\{ z \in \Omega \mid X(z) \in A \}.
\]

If the set \( A \) is a singleton, \( A = \{ x \} \), then we write \( X = x \).
Exercise

Let $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2, 3, 4\}\}$. Is $X(x) = 1 + x$ a random variable with respect to the $\sigma$-algebra $\mathcal{F}$?
Solution

The preimage of \{3\} is

\[ X^{-1}(\{3\}) = \{2\} , \]

but this is not an event in \( \mathcal{F} \). So \( X \) is not a random variable.
Expectation and Variance
Definition

Let $X$ be a discrete random variable over the probability space $(\Omega, \mathcal{F}, \Pr)$. The **expectation value** of $X$ is defined to be

$$E[X] = \sum_{\alpha \in X(\Omega)} \alpha \Pr[X = \alpha],$$

when this sum is unconditionally convergent in $\mathbb{R}$, the extended real numbers.

The expectation value is also called the **mean** of $X$. 

Linearity of Expectation

Proposition

For random variables $X_1, X_2, \ldots, X_n$, we have

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n].$$

For any real number $a$, we have

$$E[aX_k] = aE[X_k].$$
Pigeonhole Principle of Expectation

Proposition

*A random variable cannot always be less than its expected value.*
**Proposition**

A random variable cannot always be less than its expected value.

**Proof.**

Seeking a contradiction, suppose that $X$ is a discrete random variable that has values always less than $\mu = E[X]$. Then

$$E[X] = \sum_{\alpha \in \mathcal{X}(\Omega)} \alpha \Pr[X = \alpha] < \sum_{\alpha \in \mathcal{X}(\Omega)} \mu \Pr[X = \alpha] = E[X],$$

contradiction.
Proposition

A random variable cannot always be less than its expected value.

Proof.

Seeking a contradiction, suppose that $X$ is a discrete random variable that has values always less than $\mu = \mathbb{E}[X]$. Then

$$\mathbb{E}[X] = \sum_{\alpha \in X(\Omega)} \alpha \Pr[X = \alpha] < \sum_{\alpha \in X(\Omega)} \mu \Pr[X = \alpha] = \mathbb{E}[X],$$

contradiction.

Similarly, a random variable cannot always be larger than its expected value.
Exercise

Consider the complete graph $K_n$ on $n$ vertices. Show that there exists a tournament on $K_n$ that has at least $n!/2^{n-1}$ Hamiltonian paths.

A **tournament** $T_n$ is a directed graph that is obtained from $K_n$ by orienting each edge. This is a round robin tournament with no draws, where an edge $(u, v)$ in the graph $T_n$ means that player $u$ was beating player $v$.

A **Hamiltonian path** is a path of $n-1$ edges that visits each vertex of $T_n$ precisely once, $v_1 \rightarrow v_2 \rightarrow v_3 \rightarrow \cdots \rightarrow v_n$. 
The exercise asserts that some combinatorial structure exists that has a certain property. It asserts that there exists a tournament on $n$ points that has many (namely $n!/2^{n-1}$) Hamiltonian paths.

For $n = 10$, the exercise asserts that there exists a tournament with

$$\frac{n!}{2^{n-1}} = \frac{10!}{2^9} > 7000$$

Hamiltonian paths. Of course, not all tournaments on $n$ points will have that many Hamiltonian paths.
Construct a tournament on $K_n$ by randomly orienting each edge in $K_n$ with probability $1/2$. Consider a random permutation $\pi$ on $n$ points. The vertices $(v_{\pi 1}, v_{\pi 2}, \ldots, v_{\pi n})$ form a Hamiltonian path if and only if $v_{\pi k}$ beats $v_{\pi (k+1)}$ for all $k$ in the range $1 \leq k \leq n-1$. Let $X_\pi$ denote the indicator random variable for the event that $\pi$ yields a Hamiltonian path. Then

$$E[X_\pi] = \Pr[X_\pi = 1] = 1/2^{n-1}.$$ 

Let $X = \sum X_\pi$ be the random variable counting Hamiltonian paths. Since there are $n!$ permutations, the expected number of Hamiltonian paths is

$$E[X] = \sum_{\pi \in S_n} E[X_\pi] = n!/2^{n-1}.$$ 

By the pigeonhole principle of expectation, it follows that some tournament must have at least $n!/2^{n-1}$ Hamiltonian paths.
Concentration Inequalities
Theorem (Markov's Inequality)

If $X$ is a nonnegative random variable and $t$ a positive real number, then

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

Corollary (Markov's Inequality)

If $X$ is a nonnegative random variable and $t$ a positive real number, then

$$\Pr[X \geq t\mathbb{E}[X]] \leq \frac{1}{t}.$$
Chebychev’s Inequality

Theorem (Chebychev’s inequality)

If $X$ is a random variable, then

$$\Pr[|X - E[X]| \geq t] = \Pr[(X - E[X])^2 \geq t^2] \leq \frac{E[(X - E[X])^2]}{t^2} = \frac{\text{Var}[X]}{t^2}.$$
Theorem (Chernoff Bounds)

Let $X$ be the sum of $n$ independent indicator random variables $X_1, X_2, \ldots, X_n$, where $E[X_k] = p_k$. Let $\mu = E[X] = \sum_{k=1}^{n} E[X_k]$. Then

\[
\Pr[X > (1 + \delta)\mu] \leq e^{-\delta^2 \mu / 3},
\]

\[
\Pr[X < (1 - \delta)\mu] \leq e^{-\delta^2 \mu / 2}.
\]
Exercise

Who first proved Markov’s, Chebychev’s, and Chernoff’s inequality?
Exercise

Who first proved Markov’s, Chebychev’s, and Chernoff’s inequality?

Solution

1. Markov’s inequality was first proved by Chebychev.
Exercise

Who first proved Markov’s, Chebychev’s, and Chernoff’s inequality?

Solution

1. Markov’s inequality was first proved by Chebychev.
2. Chebychev’s inequality was first proved by Bienaymé.
Exercise

Who first proved Markov’s, Chebychev’s, and Chernoff’s inequality?

Solution

1. Markov’s inequality was first proved by Chebychev.
2. Chebychev’s inequality was first proved by Bienaymé.
3. Chernoff’s inequality was first proved by Rubin.
Conditional Expectation
Conditional Expectation given an Event

**Definition**

The **conditional expectation** of a discrete random variable $X$ given an event $A$ is denoted as $E[X \mid A]$ and is defined by

$$E[X \mid A] = \sum_x x \Pr[X = x \mid A].$$
Computing Expectations

We can compute the expected value of $X$ as a sum of conditional expectations. This is similar to the law of total probability.

**Proposition**

*If $X$ and $Y$ are discrete random variables, then*

$$
E[X] = \sum_y E[X \mid Y = y] \Pr[Y = y].
$$
Definition

Let $X$ and $Y$ be two discrete random variables.

The **conditional expectation** $E[X \mid Y]$ of $X$ given $Y$ is the random variable defined by

$$E[X \mid Y](\omega) = E[X \mid Y = Y(\omega)].$$
Law of the Iterated Expectation

Proposition

\[ E[E[X \mid Y]] = E[X]. \]

Proof.

\[
E[E[X \mid Y]] = \sum_y E[E[X \mid Y] \mid Y = y] \Pr[Y = y] \\
= \sum_y E[X \mid Y = y] \Pr[Y = y] \\
= E[X]
\]
Wald’s Theorem

**Theorem**

Suppose that $X_1, X_2, \ldots$ are independent random variables, all with the same mean. Suppose that $N$ is a nonnegative, integer-valued random variable that is independent of the $X_i$’s. If $E[X_1] < \infty$ and $E[N] < \infty$, then

$$E \left[ \sum_{k=1}^{N} X_i \right] = E[N]E[X_1].$$
Probability Generating Functions
Definition
Let $X$ be a discrete random variable defined on a probability space with probability measure $\Pr$. Assume that $X$ has non-negative integer values. The **probability generating function** of $X$ is defined by

$$G_X(z) = E[z^X] = \sum_{k=0}^{\infty} \Pr[X = k]z^k.$$ 

This series converges for all $z$ with $|z| \leq 1$. 


The expectation value can be expressed by

\[ E[X] = \sum_{k=1}^{\infty} k \Pr[X = k] = G'_X(1), \quad (1) \]

where \( G'_X(z) \) denotes the derivative of \( G_X(z) \).

Indeed, \( G'_X(z) = \sum_{k=0}^{\infty} k \Pr[X = k]z^{k-1} = \sum_{k=1}^{\infty} k \Pr[X = k]z^{k-1} \).
Complexity Classes
The Class **RP** of Randomized Polynomial Time DP

**Definition**

Let $\varepsilon$ be a constant in the range $0 \leq \varepsilon \leq 1/2$.

The class **RP** consists of all languages $L$ that do have a polynomial-time randomized algorithm $A$ such that

1. $x \in L$ implies $\Pr[A(x) \text{ accepts}] \geq 1 - \varepsilon$,
2. $x \notin L$ implies $\Pr[A(x) \text{ rejects}] = 1$.

**One-Sided Error**

Randomized algorithms in **RP** may err on 'yes' instances, but not on 'no' instances.
The Class **co-RP** of Randomized Polynomial Time DP

**Definition**

Let $\varepsilon$ be a constant in the range $0 \leq \varepsilon \leq 1/2$. The class **co-RP** consists of all languages $L$ whose complement $\overline{L}$ is in **RP**. In other words, $L$ is in **co-RP** if and only if there exists a polynomial-time randomized algorithm $A$ such that

1. $x \in L$ implies $\Pr[A(x) \text{ accepts}] = 1$,
2. $x \notin L$ implies $\Pr[A(x) \text{ rejects}] \geq 1 - \varepsilon$.

**One-Sided Error**

Randomized algorithms in **co-RP** may err on 'no' instances, but not on 'yes' instances.
The class **ZPP** consists of all languages $L$ such that there exists a randomized algorithm $A$ that always decides $L$ correctly and runs in expected polynomial time.
The class **BPP** of Bounded-Error Probabilistic Polynomial Time DP

**Definition**

Let \( \varepsilon \) be a constant in the range \( 0 \leq \varepsilon < 1/2 \).

The class **BPP** consists of all languages \( L \) such that there exists a polynomial-time randomized algorithm \( A \) such that

1. \( x \in L \) implies \( \Pr[A(x) \text{ accepts}] \geq 1 - \varepsilon \),
2. \( x \notin L \) implies \( \Pr[A(x) \text{ rejects}] \geq 1 - \varepsilon \).
Overview

- **P** → **ZPP** → **RP** → **NP** → **Σ₂ ∩ Π₂** → **PSPACE**
- **P** → **ZPP** → **BPP** → **Σ₂ ∩ Π₂** → **PSPACE**
- **P** → **ZPP** → **coRP** → **coNP** → **PSPACE**
Randomized Algorithms
Karger’s Minimum Cut Algorithm

**Contract**$(G)$

**Require:** A connected loopfree multigraph $G = (V, E)$ with at least 2 vertices.

1. while $|V| > 2$ do
2.   Select $e \in E$ uniformly at random;
3.   $G := G/e$;
4. end while
5. return $|E|$.

**Ensure:** An upper bound on the minimum cut of $G$. 
Iterated conditional probabilities:

\[
\Pr \left[ \bigcap_{\ell=1}^{n} E_{\ell} \right] = \left( \prod_{m=2}^{n} \Pr \left[ E_{m} \mid \bigcap_{\ell=1}^{m-1} E_{\ell} \right] \right) \Pr[E_1].
\]
Karger’s contraction algorithm is the prototypical example of a Monte Carlo type algorithm. Study it carefully!
Quicksort

Suppose that we want to sort an array $A[1..n]$ of length $n$.

Quicksort picks a pivot element $p$ uniformly at random.

Then partitions the array $A$ into three parts: left, pivot, and right.

Partition requires $n - 1$ comparisons with the pivot element $p$.

Then quicksort recursively sorts left and right parts.
Proposition

The expected number of comparisons made by randomized quicksort on an array of size $n$ is at most $2n \ln n$. 
Randomized quicksort is the prototypical example of a Las Vegas algorithm. Study the analysis carefully!
Randomized Data Structures
Skip Lists