

Final Exam Appendix

Contents

1	Induction and Least Number Inference Rules	4
2	Functions	4
3	Relations	4
3.1	Equivalence	4
3.2	Partial Order	4
3.3	Operations	5
3.4	Dilworth Theorem	6
4	Graphs	6
4.1	Edges, Paths & Cycles	6
4.2	Complete & Connected Graphs	6
4.3	Euler and Hamiltonian Paths	6
4.4	Common graphs	7
4.5	Trees	7
4.6	Colorings	7
4.7	Hall's Theorem	7
4.8	Directed Graphs	8
5	State Machines & Invariants	8
6	Well-founded Partial Orders	8
7	Summations and Asymptotic Notation	9
7.1	Summation Formulae	9
7.2	Asymptotics	9

<i>Final Exam Appendix</i>	2
8 Combinatorics	10
8.1 Pigeonhole Principle	10
8.1.1 Pigeonhole Principle	10
8.1.2 Generalized Pigeonhole Principle	10
8.2 Counting	10
8.3 Combinatorial Identities	11
8.3.1 Sum Rule	11
8.3.2 Product Rule	11
8.3.3 Division Rule	11
8.3.4 Inclusion-Exclusion	11
8.3.5 Binomial Identity	11
8.3.6 Binomial Theorem	11
8.3.7 Multinomial Theorem	11
8.3.8 Stirling's Approximation	11
9 Probability	12
9.1 Probability Spaces	12
9.2 Events	12
9.3 Law of Total Probability	12
9.4 Conditional Probability	12
9.4.1 Conditional Total Probability	12
9.5 Independence	13
9.6 Random Variables	13
9.6.1 Indicator & Uniform RV's	13
9.6.2 Binomial Distribution	13
9.6.3 Independence	14
9.7 Expectation	14
9.7.1 Total Expectation	14
9.8 Wald's Theorem	14
9.9 Variance	15
9.10 Deviation from the Mean	15
9.10.1 Markov's Bound	15
9.10.2 Chebychev's Bound	15
9.10.3 Chernoff's Bound	15

<i>Final Exam Appendix</i>	3
9.11 Pairwise Independent Sampling	15
9.12 Weak Law of Large Numbers	16
9.13 Gambler's Ruin	16
9.14 Central Limit Theorem	16

1 Induction and Least Number Inference Rules

Rule (Induction).

$$\frac{P(0), \quad \forall m \in \mathbb{N} [P(m) \longrightarrow P(m + 1)]}{\forall n \in \mathbb{N} P(n)}.$$

Rule (Strong Induction).

$$\frac{P(0), \quad \forall n \in \mathbb{N} [[\forall m \leq n P(m)] \longrightarrow P(n + 1)]}{\forall n \in \mathbb{N} P(n)}$$

Rule (Least Number Principle).

$$\frac{\exists n \in \mathbb{N} P(n)}{\exists m \in \mathbb{N} [P(m) \wedge \forall n \in \mathbb{N} [P(n) \longrightarrow m \leq n]]}$$

2 Functions

A function, $f : A \rightarrow B$, is an *injection* iff $f(a_1) = f(a_2)$ implies $a_1 = a_2$, for all $a_1, a_2 \in A$.

It is a *surjection* iff $\forall b \in B \exists a \in A, f(a) = b$.

It is a *bijection* iff it is both an injection and a surjection.

3 Relations

3.1 Equivalence

A binary relation, R , on a set A is

- *reflexive* iff $\forall x \in A (xRx)$.
- *symmetric* iff $\forall x, y \in A (xRy \longrightarrow yRx)$.
- *transitive* iff $\forall x, y, z \in A (xRy \wedge yRz \longrightarrow xRz)$.

R is an *equivalence relation* iff it is reflexive, symmetric and transitive.

3.2 Partial Order

A binary relation, R , on a set A is

- *anti-symmetric* iff $xRy \wedge yRx \longrightarrow x = y$.
- a *partial order* iff it is transitive and anti-symmetric.

- a *total order* iff it is a partial order and for all $x \neq y \in A$ either xRy or yRx .

If R is a partial order, then the set A is called a *partially ordered set (poset)*. In this case,

- a is a *minimal element* of A if $\neg \exists b \in A [bRa \wedge b \neq a]$
- a is a *maximal element* of A if $\neg \exists b \in A [aRb \wedge b \neq a]$
- a subset of A is a *chain* iff it is totally ordered by R .
- elements $a_1, a_2 \in A$ are *incomparable* iff neither a_1Ra_2 nor a_2Ra_1 holds.
- a subset of A is an *anti-chain* iff its elements are pairwise incomparable.

3.3 Operations

The *identity relation*, Id_A , on a set, A , is

$$\text{Id}_A ::= \{(a, a) \mid a \in A\}.$$

If $R \subseteq A \times B$, then, R^{-1} , the *inverse of R* , is the relation on $B \times A$ given by

$$R^{-1} ::= \{(b, a) \mid (a, b) \in R\}.$$

The *composition* of relations $R_1 \subseteq A \times B$ and $R_2 \subseteq B \times C$ is the relation $R_2 \circ R_1 \subseteq A \times C$ given by

$$R_2 \circ R_1 ::= \{(a, c) \mid \exists b (a, b) \in R_1 \wedge (b, c) \in R_2\}.$$

A *path* from a_0 to a_k in a relation $R \subseteq A \times A$ is a sequence a_0, \dots, a_k with $k \geq 0$ such that $(a_i, a_{i+1}) \in R$ for every $i < k$. The *length* of the path is k . For $n \in \mathbb{N}$,

$$R^n ::= \begin{cases} \text{Id}_A & \text{if } n = 0, \\ R^{n-1} \circ R & \text{if } n > 0. \end{cases}$$

Lemma. $R^n = \{(a, b) \mid \exists \text{ a length } n \text{ path from } a \text{ to } b \text{ in } R\}$.

For $R \subseteq A \times A$,

- The *reflexive closure* of R is $R \cup \text{Id}_A$.
- The *symmetric closure* of R is $R \cup R^{-1}$.
- The *transitive closure*, R^+ , of R is

$$R^+ ::= \bigcup_{n=1}^{\infty} R^n.$$

- The *reflexive, transitive closure*, R^* , (also called the *connectivity relation*) of R is

$$R^* ::= \bigcup_{n=0}^{\infty} R^n = R^+ \cup \text{Id}_A.$$

Lemma. If A is finite, then $R^* = (R \cup \text{Id}_A)^{|A|}$.

3.4 Dilworth Theorem

Theorem. If (A, R) is a finite poset with longest chain of length t , then A can be partitioned into t antichains.

Theorem (Dilworth). For all t , every poset with n elements must have either a chain of size greater than t or an antichain of size at least n/t .

4 Graphs

4.1 Edges, Paths & Cycles

A simple graph G is a pair (V, E) where V is an arbitrary nonempty set whose elements are called vertices or nodes, and E is a set of simple edges on V . A simple edge on V is a set of two vertices.

Theorem. The sum of the degrees of the vertices in a simple graph equals twice the number of edges.

Theorem (Handshake). In every graph, there are an even number of vertices of odd degree.

Two vertices $u, v \in V$ are adjacent iff there is an edge of G between them, viz., $\{u, v\} \in E$.

A path in G is any sequence v_0, v_1, \dots, v_n of $n \geq 0$ vertices such that v_i and v_{i+1} are adjacent for $0 \leq i < n$. The length of the path is defined to be n . The path is simple if $v_i \neq v_j$ for $0 \leq i < j \leq n$.

A cycle is a path that begins and ends at the same vertex.

Two vertices $u, v \in V$ are connected iff there is a path between them, i.e., $v_0 = u$ and $v_n = v$.

4.2 Complete & Connected Graphs

G is complete iff every two vertices are adjacent.

G is connected iff every two vertices is connected.

A subset, C , of vertices is a connected component of G iff there is a vertex, $v \in C$, such that C is the set of vertices connected to v . So G is connected iff it has exactly one connected component.

4.3 Euler and Hamiltonian Paths

Let G be an undirected graph. An Euler path in G is a path that traverses every edge of G exactly once. A Hamiltonian path is a simple path that includes every vertex.

Theorem. G has an Euler cycle iff G is finite, connected, and every vertex has even degree.

4.4 Common graphs

The *empty graph* or *anticlique*, A_n , on n vertices is the graph $A_n ::= (\{v_1, v_2, \dots, v_n\}, \emptyset)$.

The *line graph*, L_n , on n vertices is the graph $L_n ::= (\{v_1, v_2, \dots, v_n\}, \{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}\})$.

The *cycle graph*, C_n , on n vertices is L_n plus the edge $\{v_n, v_1\}$.

The *wheel*, W_n , on $n \geq 4$ vertices is the graph consisting of C_{n-1} plus one additional vertex connected to all the other vertices.

The *complete graph* or *clique*, K_n , on n vertices has an edge between every pair of vertices.

4.5 Trees

A *tree* is a (possibly infinite) simple graph $G = (V, E)$, such that any of these equivalent conditions hold:

- G is connected, but removing any edge from G leaves a disconnected graph.
- G is connected and acyclic.
- There is a unique simple path between any two distinct vertices of G .

In addition, if $|V|$ is finite, then another equivalent condition is

- G is connected and $|E| = |V| - 1$.

A *rooted tree* is a tree T together with a distinguished node, r , called the *root* of T .

A rooted tree is called *finite-path* iff it has no infinite paths away from the root.

4.6 Colorings

A k -*coloring* of G is mapping from V to a k -element set, whose element are called *colors*, such that no two adjacent vertices are mapped to the same color. G is k -*colorable* iff it has a k -coloring. G is *bipartite* if it is 2-colorable. The *chromatic number* of G is the smallest k such that G is k -colorable.

4.7 Hall's Theorem

Definition. A *bipartite graph*, $G = (V_1, V_2, E)$, is a simple graph whose vertices are the disjoint union of V_1 and V_2 and whose edges go between V_1 and V_2 , viz.,

$$E \subseteq \{\{v_1, v_2\} \mid v_1 \in V_1 \text{ and } v_2 \in V_2\}.$$

A *perfect matching* in G is an injection $f : V_1 \rightarrow V_2$ such that $\{v, f(v)\} \in E$ for all $v \in V_1$.

For any set, A , of vertices, define the neighbor set,

$$N(A) ::= \{v \mid \exists a \in A \{a, v\} \in E\}.$$

A set $A \subseteq V_1$ is called a *bottleneck* if $|A| > |N(A)|$.

Theorem (Hall). A bipartite graph has a perfect matching iff it has no bottlenecks.

4.8 Directed Graphs

A *directed graph* or *digraph* are defined in the same way as undirected graphs, except that edges are *ordered* pairs of vertices instead of an (unordered) set of two vertices. The definitions of *directed paths*, *etc.* carry over straightforwardly from the undirected case, but now successive vertices in the path must be connected by a directed edge from each vertex to the next one in the path.

5 State Machines & Invariants

A *state machine* has three parts:

1. a nonempty set, Q , whose elements are called *states*,
2. a nonempty subset $Q_0 \subseteq Q$, called the set of *start states*,
3. a binary relation, δ , on Q , called the *transition relation*.

An *invariant* for a state machine is a predicate, P , on states, such that whenever $P(q)$ is true of a state, q , and $q \rightarrow r$ for some state, r , then $P(r)$ holds.

Theorem (Invariant Theorem). *Let P be an invariant predicate for a state machine. If P holds for all start states, then P holds for all reachable states.*

A *derived variable* is a function with domain Q . If $f : Q \rightarrow P$ is a derived variable with partially ordered codomain, (P, \preceq) , then f is *strictly decreasing* iff

$$q \rightarrow q' \text{ implies } f(q') \prec f(q),$$

and f is *weakly decreasing* iff

$$q \rightarrow q' \text{ implies } f(q') \preceq f(q).$$

Theorem. *If $f : Q \rightarrow \mathbb{N}$ is a strictly decreasing derived variable of a state machine, then the length of any execution starting at a start state q is at most $f(q)$.*

6 Well-founded Partial Orders

If (P_1, \preceq_1) and (P_2, \preceq_2) are posets, then the *lexicographic partial order*, \preceq_{lex} , on $P_1 \times P_2$ is defined by the condition that

$$(p_1, p_2) \preceq_{\text{lex}} (q_1, q_2) ::= p_1 \prec_1 q_1 \text{ OR } (p_1 = q_1 \wedge p_2 \preceq_2 q_2).$$

The *coordinatewise partial order*, \preceq_c , on $P_1 \times P_2$ is defined by the condition that

$$(p_1, p_2) \preceq_c (q_1, q_2) ::= (p_1 \preceq_1 q_1 \wedge p_2 \preceq_2 q_2).$$

A poset (P, \preceq) is *well-founded* iff every nonempty subset $S \subseteq P$ has a *minimal element*.

Lemma. Suppose (P_1, \preceq_1) and (P_2, \preceq_2) are posets. Then

1. so are $(P_1 \times P_2, \preceq_{lex})$ and $(P_1 \times P_2, \preceq_c)$. Moreover,
2. if (P_1, \preceq_1) and (P_2, \preceq_2) are both well-founded, then so are $(P_1 \times P_2, \preceq_{lex})$ and $(P_1 \times P_2, \preceq_c)$.
3. if (P_1, \preceq_1) and (P_2, \preceq_2) are both totally ordered, then so is $(P_1 \times P_2, \preceq_{lex})$.

Lemma. A poset is well-founded iff it has no infinite decreasing chain.

Theorem (Fundamental Theorem for two-person games of perfect information). For games in which every play is finite and ends in win or lose, there is a winning strategy for one of the players.

7 Summations and Asymptotic Notation

7.1 Summation Formulae

$$\begin{aligned} \sum_{i=0}^n i^2 &= \frac{n(n+1)(2n+1)}{6} \\ \sum_{i=0}^n x^i &= \frac{1-x^{n+1}}{1-x} \\ \sum_{i=0}^{\infty} x^i &= \frac{1}{1-x} \\ \sum_{i=1}^n ix^i &= \frac{x - (n+1)x^{n+1} + nx^{n+2}}{(1-x)^2} \\ \sum_{i=1}^{\infty} ix^i &= \frac{x}{(1-x)^2} \\ \sum_{i=1}^{\infty} i^2 x^i &= \frac{x(1+x)}{(1-x)^3} \\ e^x &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\ 1 + x &\leq e^x \quad \text{for all real } x \\ H_n &::= \sum_{i=1}^n \frac{1}{i} \quad (\text{Harmonic Numbers}) \\ H_n &\sim \ln n \end{aligned}$$

7.2 Asymptotics

For functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$, we say

- $f \sim g$, or f is asymptotically equal to g , iff $\lim_{x \rightarrow \infty} f(x)/g(x) = 1$

- $f(x) = o(g(x))$, or f is asymptotically smaller than g , iff $\lim_{x \rightarrow \infty} f(x)/g(x) = 0$
- with g nonnegative, we say that ¹ $f = O(g)$ iff $\limsup_{x \rightarrow \infty} |f(x)|/g(x) < \infty$.
equivalently, there exists a constant $c \geq 0$ and an x_0 such that $\forall x \geq x_0, |f(x)| \leq cg(x)$
- $f = \Theta(g)$ iff $f = O(g) \wedge g = O(f)$

8 Combinatorics

8.1 Pigeonhole Principle

8.1.1 Pigeonhole Principle

If there are more pigeons than pigeonholes, then there must be at least two pigeons in one hole.

8.1.2 Generalized Pigeonhole Principle

If there are m pigeons and n pigeonholes, then there must be at least $\lceil m/n \rceil$ pigeons in some hole.

8.2 Counting

Given a set of n elements:

- number of r -permutations without replacement:

$$P(n, r) ::= n(n-1) \dots (n-r+1) = \frac{n!}{(n-r)!}$$

- r -combinations without replacement:

$$C(n, r) ::= \binom{n}{r} = \frac{n!}{(n-r)! r!}$$

- r -permutations with replacement (r distinct balls, n distinct bins):

$$n^r$$

- r -combinations with replacement (r identical balls, n distinct bins):

$$\binom{n+r-1}{r}$$

- permutations with repetition:

$$\binom{n}{r_1, r_2, \dots, r_k} ::= \frac{n!}{r_1! r_2! \dots r_k!}$$

where $r_1 + r_2 + \dots + r_k = n$.

¹ $\limsup_{x \rightarrow \infty} h(x) ::= \lim_{x \rightarrow \infty} \sup_{y \geq x} h(y)$.

8.3 Combinatorial Identities

8.3.1 Sum Rule

If A_i are disjoint finite sets, then $|A_1 \cup A_2 \cup \dots \cup A_n| = |A_1| + |A_2| + \dots + |A_n|$.

8.3.2 Product Rule

If A_i are finite sets, then $|A_1 \times A_2 \times \dots \times A_n| = |A_1| \cdot |A_2| \cdot \dots \cdot |A_n|$.

8.3.3 Division Rule

If $f : A \rightarrow B$ maps exactly k elements of A to each element of B , then $|A| = k|B|$.

8.3.4 Inclusion-Exclusion

$$\begin{aligned} \left| \bigcup_{i=1}^n A_i \right| &= \sum_{\emptyset \neq S \subseteq \{1, \dots, n\}} (-1)^{|S|+1} \left| \bigcap_{i \in S} A_i \right| \\ &= \sum_{k=1}^n (-1)^{k+1} \sum_{S \subseteq \{1, \dots, n\}, |S|=k} \left| \bigcap_{i \in S} A_i \right| \end{aligned}$$

8.3.5 Binomial Identity

$$\binom{n}{m} = \binom{n}{n-m}$$

8.3.6 Binomial Theorem

$$(x_1 + x_2)^n = \sum_{i=0}^n \binom{n}{i} x_1^{n-i} x_2^i$$

8.3.7 Multinomial Theorem

$$(x_1 + x_2 + \dots + x_k)^n = \sum_{r_1+r_2+\dots+r_k=n} \binom{n}{r_1, r_2, \dots, r_k} x_1^{r_1} x_2^{r_2} \dots x_k^{r_k}$$

8.3.8 Stirling's Approximation

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

9 Probability

9.1 Probability Spaces

A *sample space* is a non-empty set S whose elements are called *outcomes*. The *events* are subsets of S .

A *probability space* consists of a sample space S together with a *probability function*, $\Pr\{\cdot\} : P(S) \rightarrow [0, 1]$ satisfying the Sum Rule and such that $\Pr\{S\} = 1$.

9.2 Events

$$\Pr\left\{\bigcup_{n \in \mathbb{N}} A_n\right\} = \sum_{n \in \mathbb{N}} \Pr\{A_n\} \quad \text{for pairwise disjoint } A_n \quad (\text{Sum Rule})$$

$$\Pr\{A - B\} = \Pr\{A\} - \Pr\{A \cap B\} \quad (\text{Difference Rule})$$

$$\Pr\{\overline{B}\} = 1 - \Pr\{B\} \quad (\text{Complement Rule})$$

$$\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\} \quad (\text{Inclusion-Exclusion})$$

$$\Pr\{A \cup B\} \leq \Pr\{A\} + \Pr\{B\} \quad (\text{Boole's inequality})$$

$$\Pr\{A\} \leq \Pr\{A \cup B\} \quad (\text{Monotonicity})$$

9.3 Law of Total Probability

Suppose the sample space is the disjoint union of B_0, B_1, \dots . Then for all events A ,

$$\Pr\{A\} = \sum_{i \in \mathbb{N}} \Pr\{A \cap B_i\}.$$

9.4 Conditional Probability

$$\Pr\{A \mid B\} ::= \frac{\Pr\{A \cap B\}}{\Pr\{B\}}$$

$$\Pr\{A \cap B\} = \Pr\{A \mid B\} \Pr\{B\} \quad (\text{Product Rule})$$

$$\Pr\{A \mid B\} = \frac{\Pr\{B \mid A\} \Pr\{A\}}{\Pr\{B\}} \quad (\text{Bayes Rule})$$

9.4.1 Conditional Total Probability

Suppose the sample space is the disjoint union of B_0, B_1, \dots . Then for all events A ,

$$\Pr\{A\} = \sum_{i \in \mathbb{N}} \Pr\{A \mid B_i\} \Pr\{B_i\}.$$

9.5 Independence

Definition. Events A and B are *independent* iff

$$\Pr \{A \cap B\} = \Pr \{A\} \Pr \{B\}.$$

Events A_0, A_1, A_2, \dots are *mutually independent* iff for all subsets $J \subset \mathbb{N}$,

$$\Pr \left\{ \bigcap_{i \in J} A_i \right\} = \prod_{i \in J} \Pr \{A_i\}.$$

Events A_0, A_1, A_2, \dots are *k-wise independent* iff $\{A_i \mid i \in J\}$ are mutually independent for all subsets $J \subset \mathbb{N}$ with $|J| = k$.

9.6 Random Variables

A *random variable* over a given sample space, \mathcal{S} , is a function from $\mathcal{S} \rightarrow \mathbb{R}$.

The *density function*, f_R , of a random variable, R , is

$$f_R(a) ::= \Pr \{R = a\}.$$

The *cumulative distribution function*, F_R , is

$$F_R(a) ::= \Pr \{R \leq a\}.$$

9.6.1 Indicator & Uniform RV's

For any event A , its *indicator variable*, I_A , is the 0-1 valued variable such that the event $[I_A = 1]$ is the same as the event A .

A random variable, U , is *uniform* iff all its values are equally likely. That is

$$\Pr \{U = a\} = \Pr \{U = b\}$$

for all $a, b \in \text{range}(U)$.

9.6.2 Binomial Distribution

A random variable, B , is *binomial* with parameters (n, p) iff its pdf is the function $f_{n,p} : \mathbb{N} \rightarrow [0, 1]$ defined by

$$f_{n,p}(k) ::= \binom{n}{k} p^k (1-p)^{n-k}$$

where parameter $n \in \mathbb{N}$ and $0 < p < 1$. Equivalently, $B = \sum_{k=1}^n B_k$ where B_1, B_2, \dots, B_n are mutually independent indicator variables with $\Pr \{B_i = 1\} = p$.

$$F_{n,p}(\alpha n) \leq \frac{1-\alpha}{1-\alpha/p} f_{n,p}(\alpha n)$$

where $0 < \alpha < p$.

9.6.3 Independence

Random variables R_1, R_2, \dots are *mutually independent* iff

$$\Pr \left\{ \bigcap_i [R_i = x_i] \right\} = \prod_i \Pr \{R_i = x_i\},$$

for all $x_1, x_2, \dots \in \mathbb{R}$. They are *k-wise independent* iff $\{R_i \mid i \in J\}$ are mutually independent for all subsets $J \subset \mathbb{N}$ with $|J| = k$.

9.7 Expectation

$$E[R] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr \{R = r\} \quad \text{definition in terms of random variables}$$

$$E[R] ::= \sum_{s \in \mathcal{S}} R(s) \cdot \Pr \{s\} \quad \text{definition in terms of sample space}$$

$$E[R] = \sum_{r \in \mathbb{N}} \Pr \{R > r\} \quad R \text{ natural number valued}$$

$$E \left[\sum_{n \in \mathbb{N}} R_n \right] = \sum_{n \in \mathbb{N}} E[R_n] \quad \text{providing } \sum_{n=1}^{\infty} E[|R_n|] \text{ converges}$$

$$E \left[\prod_{n \in \mathbb{N}} R_n \right] = \prod_{n \in \mathbb{N}} E[R_n] \quad \text{for mutually independent } R_0, R_1, \dots$$

For an random variable, R , and event, A , the *conditional expectation* of R given A is

$$E[R \mid A] ::= \sum_{r \in \text{range}(R)} r \cdot \Pr \{R = r \mid A\}$$

9.7.1 Total Expectation

Let B_0, B_1, \dots be disjoint events whose union is the entire sample space. Then for any random variable, X ,

$$E[X] = \sum_{i \in \mathbb{N}} E[X \mid B_i] \Pr \{B_i\}$$

provided that all the expectations exist and are finite.

9.8 Wald's Theorem

Let C_1, C_2, \dots , be a sequence of nonnegative random variables, and let Q be a positive integer-valued random variable, all with finite expectations. Suppose that

$$E[C_i \mid Q \geq i] = \mu$$

for some $\mu \in \mathbb{R}$ and for all $i \geq 1$. Then

$$E[C_1 + C_2 + \dots + C_Q] = \mu E[Q].$$

9.9 Variance

$$\begin{aligned} \text{Var}[R] &::= \text{E}[(R - \text{E}[R])^2] = \text{E}[R^2] - \text{E}^2[R], \\ \text{Var}[I_A] &= \text{Pr}\{A\}(1 - \text{Pr}\{A\}) \\ \text{Var}[aR + b] &= a^2 \text{Var}[R], && \text{for } a, b \in \mathbb{R}, \\ \text{Var}[R_1 + \dots + R_n] &= \text{Var}[R_1] + \dots + \text{Var}[R_n] && \text{for pairwise independent } R_1, \dots, R_n \\ \sigma_R &::= \sqrt{\text{Var}[R]} && \text{(the standard deviation of } R) \end{aligned}$$

9.10 Deviation from the Mean

9.10.1 Markov's Bound

If R is a nonnegative random variable, then for all $x > 0$

$$\Pr\{R \geq x\} \leq \frac{\text{E}[R]}{x}.$$

9.10.2 Chebychev's Bound

Let R be a random variable, and let x be a positive real number. Then

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

9.10.3 Chernoff's Bound

Let T_1, T_2, \dots, T_N be mutually independent Bernoulli variables, and let $T ::= T_1 + T_2 + \dots + T_N$. Then for all $c \geq 1$,

$$\Pr\{T \geq c \text{E}[T]\} \leq \exp(-(c \ln c - c + 1) \text{E}[T]). \quad (1)$$

9.11 Pairwise Independent Sampling

Let G_1, \dots, G_n be pairwise independent random variables with deviations bounded by a constant $b > 0$. Let

$$A_n ::= \frac{\sum_{i=1}^n G_i}{n}.$$

Then

$$\Pr\{|A_n - \text{E}[A_n]| \geq x\} \leq \frac{1}{n} \left(\frac{b}{x}\right)^2. \quad (2)$$

9.12 Weak Law of Large Numbers

Let G_1, \dots, G_n, \dots and A_n be as in the Pairwise Independent Sampling Theorem. Then for any $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr \{|A_n - E[A_n]| < \epsilon\} = 1.$$

9.13 Gambler's Ruin

Theorem. In the Gambler's Ruin game with probability p of winning each individual bet, with initial capital, n , and goal, T ,

$$\Pr \{\text{the gambler is a winner in the fair game}\} = \frac{n}{T}, \quad (3)$$

$$\Pr \{\text{the gambler is a winner a biased game}\} = \frac{(q/p)^n - 1}{(q/p)^T - 1}, \quad (4)$$

where $q ::= (1 - p)$.

$$\Pr \{\text{the gambler is a winner in an unfair game}\} \leq (p/q)^{T-n}. \quad (5)$$

Let Q be the number of bets till the game ends.

$$\begin{aligned} E[Q \text{ in an unfair game}] &= \frac{\Pr \{\text{gambler is a winner}\} T - n}{2p - 1}, \\ E[Q \text{ in a fair game}] &= n(T - n). \end{aligned} \quad (6)$$

To obtain formula (4), let p and T be fixed, and let w_n be the gambler's probability of winning when his initial capital is n dollars. So $w_0 = 0$ and $w_T = 1$. For $0 < n < T$, the Total Expectation Theorem implies

$$w_n = pw_{n+1} + qw_{n-1}.$$

Solving this simple linear recurrence yields (4).

9.14 Central Limit Theorem

Definition. The normal density function is the function

$$\eta(x) ::= \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

and the normal distribution function is its integral

$$N(y) ::= \int_{-\infty}^y \eta(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-x^2/2} dx.$$

The function $\eta(x)$ defines the standard *Bell curve*, centered about the origin with height $1/\sqrt{2\pi}$ and about two-thirds of its area within unit distance of the origin. The normal distribution function $N(y)$ approaches 0 as $y \rightarrow -\infty$. As y approaches zero from below, $N(y)$ grows rapidly towards $1/2$. Then as y continues to increase beyond zero, $N(y)$ rapidly approaches 1.

Theorem (Central Limit). Let $S_n = \sum_{i=1}^n G_i$ where G_1, \dots, G_i, \dots are mutually independent variables with the same mean, μ , and deviation, σ . Let $\mu_n ::= E[S_n] = n\mu$, and $\sigma_n ::= \sigma_{S_n} = n\sigma$. Now let $S_n^* ::= (S_n - \mu_n)/\sigma_n$ be the normalized version of S_n . Then

$$\lim_{n \rightarrow \infty} \Pr \{S_n^* \leq \beta\} = N(\beta)$$

for any real number β .