COSC 5010 - Randomness in Computation Notes

Libao Jin

March 13, 2020

Contents

1 Events and Probability **2**
- 1.1 Application: Verifying Polynomial Identities . . . 2
  - 1.1.1 Polynomial Identity Testing (for the reals) . . . 2
  - 1.1.2 Brute-force Approach . . . 2
  - 1.1.3 Rephrase the Question . . . 2
  - 1.1.4 Use Randomness . . . 3
  - 1.1.5 Improving the Probability . . . 3
- 1.2 Axioms of Probability . . . 3
  - 1.2.1 Probability Space . . . 3
  - 1.2.2 Events are sets . . . 3
  - 1.2.3 Conditional Probability . . . 4
  - 1.2.4 Game Show - Monty Hall Problem . . . 5
  - 1.2.5 Verifying Matrix Multiplication . . . 6
  - 1.2.6 Min-Cut Problem . . . 7
- 1.3 Binomial Random Variable . . . 9
  - 1.3.1 Linearity of Expectations . . . 10
  - 1.3.2 Jensen's Inequality . . . 10
- 1.4 The Bernoulli and Binomial Random Variables . . . 10
- 1.5 Conditional Expectation . . . 10
- 1.6 Geometric Distribution . . . 11
- 1.7 Coupon Collector's Problem . . . 12

2 Moments and Derivations **13**
- 2.1 Markov's Inequality . . . 13
- 2.2 Chernoff Bounds . . . 15
- 2.3 Satisfiability . . . 15
- 2.4 Moment Generating Functions (MGFs) . . . 15

1 Events and Probability

1.1 Application: Verifying Polynomial Identities

Definition 1.1. Two polynomials p and q are identical if $p(x) = q(x)$ for all x. Write $p(x) \equiv q(x)$ if they are identical.

1.1.1 Polynomial Identity Testing (for the reals)

Given two polynomials $p, q : \mathbb{R} \to \mathbb{R}$, is $p(x) \equiv q(x)$?

Example 1.1.

$$(3x - 4)(x + 2)(x - 5)(2x + 7) \equiv 280 - 46x - 127x^2 - 5x^3 + 6x^4.$$

1.1.2 Brute-force Approach

Convert each polynomial to canonical form

$$\sum_{i=0}^{d} c_i x^i$$

and compare coefficients. Running time: $O(d^2)$.

1.1.3 Rephrase the Question

Is $p(x) - q(x) = 0$? A polynomial of degree d has at most d roots. If $p(x) - q(x) \not\equiv 0$, then

$$|\{x|p(x) = q(x)\}| \leq d,$$

where $d = \deg(p(x) - q(x))$.

Algorithm 1: Check $d + 1$ values of x

```
Function checkPolynomials(p, q):
    Input: Polynomials p and q
    Output: Whether p ≡ q?
1   for i ← 1 to d + 1 do
2       if p(i) ≠ q(i) then
3           return false
4       end
5   end
6   return true
end
```

Let $T = \{1, 2, \ldots, d + 1\}$.

- If $p(x) \equiv q(x)$, then $p(i) = q(i)$ for all $i \in T$, output true.
- If $p(x) \not\equiv q(x)$, then $p(i) \neq q(i)$ for some $i \in T$, output false.

1.1.4 Use Randomness

Pick r at uniformly random from $\{1, 2, \ldots, 3d\}$.

- If $p(r) = q(r)$, output true.
- If $p(r) \neq q(r)$, output false.
- If $p(x) \equiv q(x)$, then the algorithm always gives the correct answer.
- If $p(x) \not\equiv q(x)$, there are at most d points r where $p(r) = q(r)$.

The algorithm outputs true with probability at most $d/3d = 1/3$, the algorithm outputs false (the correct answer) with probability greater than $2/3$. The running time is $O(d)$.

1.1.5 Improving the Probability

- Use a larger range: choose r from $\{1, ..., 100d\}$, the probability of incorrect answer is less than $\leq d/100d = 1/100$.
- Use repeated trials: Run the algorithm twice, if the answer is true both times, then output true. Otherwise, output false. Suppose that $p(x) \not\equiv q(x)$, the probability of getting true both times is less than $\frac{1}{3}\frac{1}{3} = \frac{1}{9}$. Therefore, get the correct answer with probability greater than $\frac{8}{9}$.
- In general, if we run the algorithm k times, the error probability is less than $1/3^k = 3^{-k}$. The running time is $O(kd)$. Take $k = 100$, $O(d)$ time with less than 3^{-100} error probability.

1.2 Axioms of Probability

1.2.1 Probability Space

Definition 1.2. A *probability space* has three components:

(a) a sample space Ω, the set of all possible outcomes,
(b) a family $\mathcal{F}$ of subsets of Ω, each set $E \in \mathcal{F}$ is an *event*.
(c) a probability function $\Pr : \mathcal{F} \to [0, 1]$ assigns $\Pr(\Omega) = 1$ and has countable additivity, i.e., for any sequence $E_1, E_2, \ldots$ of mutually disjoint events,

$$\Pr\left(\bigcup_{i\geq 1} E_i\right) = \sum_{i\geq 1} \Pr(E_i).$$

Example 1.2. $\Omega = \{1, 2, \ldots, 3d\}$, events $E_i = \{i\}$ for each $i \in \Omega$, $\Pr(E_i) = \frac{1}{3d}$ and $\Pr(\Omega) = 1$. Then

$$T = \{r \in \Omega | p(r) = q(r)\},$$
$$F = \{r \in \Omega | p(r) \neq q(r)\}.$$

- If $p(x) \equiv q(x)$, then $\Pr(T) = 1$ and $\Pr(F) = 0$.
- If $p(x) \not\equiv q(x)$, then $\Pr(T) \leq 1/3$ and $\Pr(F) \geq 2/3$.

1.2.2 Events are sets

Roll a die, E_1 is the event that the die comes even $\{2, 4, 6\}$ and E_2 is the event that the die comes up less than 3, $\{1, 2, 3\}$.

- $E_1 \cap E_2 = \{2\}$.

- $E_1 \cup E_2 = \{1, 2, 3, 4, 6\}$.
- $E_1 - E_2 = \{4, 6\}$.
- $\overline{E_1} = \{1, 3, 5\}$.

Proposition 1.1. If E_1 and E_2 are disjoint, then $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2)$.

Lemma 1.1. $\Pr(E_1 \cup E_2) = \Pr(E_1) + \Pr(E_2) - \Pr(E_1 \cap E_2)$.

Corollary 1.1 (Union Bound)**.** For any sequence of events $E_1, E_2, \ldots$,

$$\Pr\left(\bigcup_{i \geq 1} E_i\right) \leq \sum_{i \geq 1} \Pr(E_i).$$

Definition 1.3 (Independence of Events)**.** Two events are *independent* if and only if $\Pr(E_1 \cap E_2) = \Pr(E_1)\Pr(E_2)$.

Definition 1.4 (Mutual Independence of Events)**.** Events $E_1, \ldots, E_k$ are *mutually independent* if and only if for all $I \subset \{1, \ldots, k\}$,

$$\Pr\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \Pr(E_i).$$

Back to the polynomial identity testing. Run the polynomial identity testing algorithm k times. Let E_i be the event that the algorithm returns the wrong answer on the ith run. Each run is independent of the others, so

$$\Pr\left(\bigcap_{i=1}^{k} E_i\right) = \prod_{i=1}^{k} \Pr(E_i) \leq \prod_{i=1}^{k} \frac{1}{3} = 3^{-k}.$$

1.2.3 Conditional Probability

Definition 1.5 (Conditional Probability)**.** The *conditional probability* that E occurs given that F occurs is

$$\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)},$$

where $\Pr(F) > 0$.

Proposition 1.2.

(a) $\Pr(E \cap F) = \Pr(E|F)\Pr(F)$.
(b) If E and F are independent, then $\Pr(E|F) = \Pr(E)$.

Proposition 1.3 (Bayes' Law)**.** Assume $E_1, \ldots, E_n$ are mutually disjoint with $\Omega = \bigcup_{i=1}^{n} E_i$. Let B be any event.

$$\Pr(E_j|B) = \frac{\Pr(E_j \cap B)}{\Pr(B)} = \frac{\Pr(B|E_j)\Pr(E_j)}{\sum_{i=1}^{n} \Pr(B|E_i)\Pr(E_i)}.$$

Example 1.3. Three coins: one coin is biased with probability of heads being 2/3 and probability of tails being 1/3, two coins are fair. The three coins are permuted randomly and then flipped. Suppose the outcome is $\{H, H, T\}$. What is the probability that the first coin is biased one?

Solution. For $i \in \{1,2,3\}$, let E_i be the event that the ith coin is biased,

$$\Pr(E_1) = \Pr(E_2) = \Pr(E_3) = \frac{1}{3}.$$

Let B be the event $\{H, H, T\}$. Then

$$\Pr(B|E_1) = \frac{2}{3}\frac{1}{2}\frac{1}{2}, \quad \Pr(B|E_2) = \frac{1}{2}\frac{2}{3}\frac{1}{2}, \quad \Pr(B|E_3) = \frac{1}{2}\frac{1}{2}\frac{1}{3}.$$

Then

$$\Pr(E_1|B) = \frac{\Pr(B|E_1)\Pr(E_1)}{\Pr(B|E_1)\Pr(E_1) + \Pr(B|E_2)\Pr(E_2) + \Pr(B|E_3)\Pr(E_3)} = \frac{2}{5}.$$

□

1.2.4 Game Show - Monty Hall Problem

There are three doors: D_1, D_2, D_3. Prize randomly put behind one of the doors. P_i is the event that prize is behind D_i and assume that $\Pr(P_1) = \Pr(P_2) = \Pr(P_3) = 1/3$. Contestant chooses D_1 (they can choose any door). The host knows which door has the prize and opens a door that does not have the prize, either D_2 or D_3. The host asks the contestant if they would like to switch to the other unopened door. Should they switch? Does it matter?

Solution. Let H_i be the event the host opens D_i.

$$\begin{aligned}
&\Pr(H_2|P_2) = 0, \quad \Pr(H_2|P_3) = 1, \quad \Pr(H_3|P_2) = 1,\\
&\Pr(H_3|P_3) = 0, \quad \Pr(H_2|P_1) = \alpha, \quad \Pr(H_3|P_1) = 1-\alpha.
\end{aligned}$$

Suppose that host open D_2. We want to know $\Pr(P_1|H_2)$ and $\Pr(P_3|H_2)$.

$$\begin{aligned}
\Pr(P_1|H_2) = \frac{\Pr(P_1 \cap H_2)}{\Pr(H_2)} = \frac{\Pr(H_2|P_1)\Pr(P_1)}{\Pr(H_2)} &= \frac{\Pr(H_2|P_1)\Pr(P_1)}{\sum_{i=1}^{3}\Pr(H_2 \cap \Pr_i)}\\
&= \frac{\Pr(H_2|P_1)\Pr(P_1)}{\sum_{i=1}^{3}\Pr(H_2|\Pr_i)\Pr_i}\\
&= \frac{\alpha/3}{\alpha/3 + 0 + 1/3}\\
&= \frac{\alpha}{\alpha+1}.
\end{aligned}$$

and

$$\begin{aligned}
\Pr(P_3|H_2) = \frac{\Pr(P_3 \cap H_2)}{\Pr(H_2)} = \frac{\Pr(H_2|P_3)\Pr(P_2)}{\Pr(H_2)} &= \frac{\Pr(H_2|P_3)\Pr(P_2)}{\sum_{i=1}^{3}\Pr(H_2 \cap \Pr_i)}\\
&= \frac{\Pr(H_2|P_3)\Pr(P_2)}{\sum_{i=1}^{3}\Pr(H_2|\Pr_i)\Pr_i}\\
&= \frac{1/3}{\alpha/3 + 0 + 1/3}\\
&= \frac{1}{\alpha+1}.
\end{aligned}$$

Contestant should switch if

$$\Pr(P_3|H_2) = \frac{1}{1+\alpha} \geq \frac{\alpha}{1+\alpha} = \Pr(P_1|H_2), \forall \alpha \in [0,1].$$

- When $\alpha = 1/2$, $\Pr(P_1|H_2) = 1/3$, $\Pr(P_3|H_2) = 2/3$.
- When $\alpha = 0$, $\Pr(P_1|H_2) = 0$, $\Pr(P_3|H_2) = 1$.
- When $\alpha = 1$, $\Pr(P_1|H_2) = 1/2$, $\Pr(P_3|H_2) = 1/2$.

□

1.2.5 Verifying Matrix Multiplication

Given: three $n \times n$ matrices A, B, C. Assume that the matrices are boolean and arithmetic is done mod 2. Can just compute $A \cdot B$ and compare to C. Standard matrix multiplication takes $O(n^3)$ time and $\approx O(n^{2.37})$ is best known.

Proposition 1.4 (Law of Total Probability)**.** Let $E_1, \ldots, E_n$ be mutually disjoint events that partition Ω.

$$\Omega = \bigcup_{i=1}^{n} E_i.$$

For any event B,

$$\Pr(B) = \sum_{i=1}^{n} \Pr(B \cap E_i) = \sum_{i=1}^{n} \Pr(B|E_i) \Pr(E_i).$$

Theorem 1.1. If $AB \neq C$ and $\mathbf{r} \in \{0,1\}^n$ is chosen at random, then

$$\Pr(AB\mathbf{r} = C\mathbf{r}) \leq \frac{1}{2}.$$

Proof. Let $D = C - AB$. Then $D \neq O$, i.e., D has some nonzero entry. Without loss of generality (WLOG), let that entry be d_{11}. Let $\mathbf{r} \in \{0,1\}^n$ and suppose $AB\mathbf{r} = C\mathbf{r}$. Then $D\mathbf{r} = \mathbf{0}$. In particular, the first entry of $D\mathbf{r}$ is 0, that is,

$$\sum_{j=1}^{n} d_{1j} r_j = 0.$$

Equivalently,

$$r_1 = -\frac{1}{d_{11}} \sum_{j=2}^{n} d_{1j} r_j. \tag{1.1}$$

Consider first choosing $r_2, \ldots, r_n \in \{0,1\}$ at random. This determines the right-hand side of (1.1). There is then one value of r_1 that makes (1.1) true. If we now choose r_1 at random, (1.1) holds with probability less than 1/2 (Principle of Differed Decisions). For any $(x_2, x_3, \ldots, x_n) \in \{0,1\}^{n-1}$,

$$\begin{aligned}\Pr[AB\mathbf{r} = C\mathbf{r}|(r_2, \ldots, r_n) = (x_2, \ldots, x_n)] &\leq \Pr\left[r_1 = -\sum_{j=2}^{n} \frac{d_{1j} r_j}{d_{11}} | (r_2, \ldots, r_n) = (x_2, \ldots, x_n)\right] \\ &\leq \frac{1}{2}.\end{aligned}$$

Then

$$\begin{aligned}
\Pr[AB\mathbf{r} = C\mathbf{r}] &= \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr[AB\mathbf{r} = C\mathbf{r} \text{ and } (r_2,\ldots,r_n) = (x_2,\ldots,x_n)] \\
&\leq \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\left[r_1 = -\sum_{j=2}^{n} \frac{d_{1j}r_j}{d_{11}} \text{ and } (r_2,\ldots,r_n) = (x_2,\ldots,x_n)\right] \\
&\leq \sum_{(x_2,\ldots,x_n)\in\{0,1\}^{n-1}} \Pr\left[r_1 = -\sum_{j=2}^{n} \frac{d_{1j}r_j}{d_{11}} \middle| (r_2,\ldots,r_n) = (x_2,\ldots,x_n)\right] \Pr[(r_2,\ldots,r_n) = (x_2,\ldots,x_n)] \\
&\leq \frac{1}{2}\cdot 2^{n-1}\cdot\frac{1}{2^{n-1}} \\
&= \frac{1}{2}.
\end{aligned}$$

□

By associativity, $(AB)\mathbf{r} = A(B\mathbf{r})$ can be computed efficiently in $O(n^2)$.

Algorithm 2: Randomized Algorithm for Verifying Matrix Multiplication

```
Function VerifyMatrixMultiplication(A, B, C):
    Input: A, B, and C are three matrices.
    Output: Whether A · B = C?
1   Choose r ∈ {0,1}^n at random
2   Compute y ← B · r, z ← C · r, x ← A · y
3   if x = z then
4       return true
5   else
6       return false
7   end
end
```

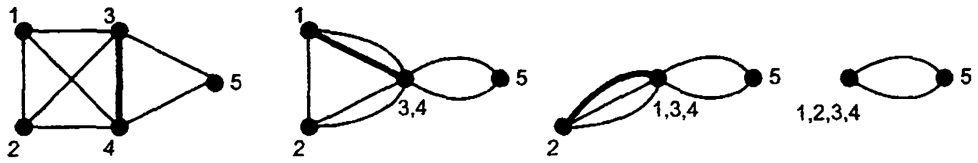
1.2.6 Min-Cut Problem

A *cut-set* is a graph is a set of edges whose removal breaks the graph into two or more connected component.

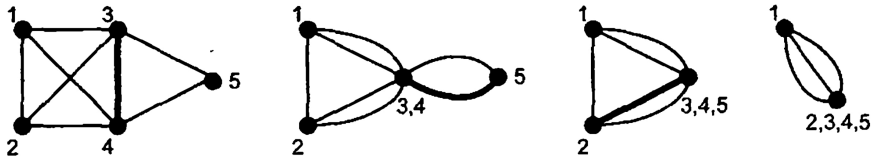
- *Goal*: Given a graph, find a minimum size cut-set (min-cut).
- *Edge contraction* for a edge (u, v) merge u and v into one vertex, and eliminate all edges between u and v.

Algorithm 3:

Function ():
    Input:
    Output:
1   **for** $i \leftarrow 1$ **to** $n-2$ **do**
2   randomly pick an edge and contract it
3   **end**
4   **return** the edges between the remaining two vertices
end



(a) A successful run of min-cut.



(b) An unsuccessful run of min-cut.

Figure 1: An example of two executions of min-cut in a graph with minimum cut-set of size 2.

Theorem 1.2. The algorithm outputs a min-cut set with probability greater than $\frac{2}{n(n-1)}$.

Proof. Let k be the size of min-cut set in G. Let C be any min-cut set. We'll compute the probability of finding C. C partitions vertices V into two sets S and $V-S$. The algorithm outputs C if it never contracts an edge (u,v) where $u \in S$ and $v \in V-S$. Let E_i be the event that in iteration i, the algorithm contracts an edge not in C. Let $F_i = \bigcap_{j=1}^{i} E_j$ and F_{n-2} be the event the algorithm outputs C. Because the min-cut set has K edges, every vertex in G has degree greater than k. Therefore,

$$|E| = \frac{1}{2}\sum_{v \in V} \deg(v) \geq \frac{1}{2}\sum_{v \in V} k = \frac{nk}{2}.$$

Then

$$\Pr(F_1) = \Pr(E_1) = \frac{|E|-|C|}{|E|} = 1 - \frac{|C|}{|E|} \geq 1 - \frac{k}{nk/2} = 1 - \frac{2}{n} = \frac{n-2}{n}.$$

Next, we calculate $\Pr(E_2|F_1)$. Assuming F_1 the graph has $n-1$ vertices and has a min-cut size of K.

$$P(E_2|F_1) \geq 1 - \frac{k}{(n-1)k/2} = \frac{n-3}{n-1}.$$

In general,

$$\Pr(E_{i+1}|F_i) \geq \frac{n-i-2}{n-i}.$$

Note that

$$\Pr(F_{n-2}) = \Pr(E_{n-2} \cap F_{n-3}) = \Pr(E_{n-2}|F_{n-3}) \Pr(F_{n-3}).$$

In general,

$$\Pr(F_{n-i}) = \Pr(E_{n-i} \cap F_{n-i-1}) = \Pr(E_{n-i}|F_{n-i-1})P(F_{n-i-1}).$$

Then

$$\begin{aligned}
\Pr(F_{n-2}) &= \Pr(E_{n-2}|F_{n-3}) \Pr(E_{n-3}|F_{n-4}) \cdots \Pr(E_2|F_1) \Pr(F_1) \\
&= P(F_1) \prod_{i=1}^{n-3} \Pr(E_{i+1}|F_i) \\
&= \frac{n-2}{n} \prod_{i=1}^{n-3} \frac{n-i-2}{n-i} \\
&= \frac{n-2}{n} \frac{n-3}{n-1} \frac{n-4}{n-2} \cdots \frac{3}{5} \frac{2}{4} \frac{1}{3} \\
&= \frac{2}{n(n-1)}
\end{aligned}$$

The success probability for fining min-cut is $\frac{2}{n(n-1)}$. Run the algorithm m times, select the smallest cut set. The probability of not finding min-cut set is less than

$$\left(1 - \frac{2}{n(n-1)}\right)^m \leq \left\{e^{-2/[n(n-1)]}\right\}^m = e^{-2m/[n(n-1)]}$$

Let $m = n(n-1)$, then the failure probability is less than

$$e^{-2} = \frac{1}{e^2} < \frac{1}{4}.$$

Let $m = n(n-1)\ln n$, then the failure probability is less than

$$e^{-2} = \frac{1}{e^{2\ln n}} = \frac{1}{n^2}.$$

Let $m = 100n(n-1)$, then the failure probability is less than

$$e^{-2} = \frac{1}{e^{200}} < \frac{1}{4^{100}}.$$

□

1.3 Binomial Random Variable

Definition 1.6. A random variable X on a sample space Ω is a real-valued (measurable) function on Ω; that is, $X : \Omega \to \mathbb{R}$. A discrete random variable is a random variable that takes on only a finite or countably infinite number of values.

Definition 1.7. Two random variables X and Y are independent if and only if

$$\Pr(X = x \cap Y = y) = \Pr(X = x) \cdot P(Y = y)$$

for all values x and y. Similarly, random variables $X_1, X_2, \ldots, X_k$ are mutually independent if and only if, any subset $I \subset [1, k]$ and any values $x_i, i \in I$,

$$\Pr(\bigcap_{i \in I} X_i = x_i) = \prod_{i \in I} \Pr(X_i = x_i).$$

Definition 1.8. The expectation of a discrete random variable X, denoted by $E[X]$, is given by

Definition 1.9.

$$E[X] = \sum_i i \Pr(X = i),$$

where the summation is over all values in the range of X. The expectation is finite if $\sum_i |i| \Pr(X = i)$ converges; otherwise, the expectation is unbounded.

1.3.1 Linearity of Expectations

Theorem 1.3 (Linearity of Expectations)**.** For any finite collection of discrete random variables $X_1, X_2, \ldots, X_n$ with finite expectations,

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i].$$

Lemma 1.2. For any constant c and discrete random variable X,

$$E[cX] = cE[X].$$

1.3.2 Jensen's Inequality

Definition 1.10. A function $f : \mathbb{R} \to \mathbb{R}$ is said to be convex if, for any x_1, x_2 and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda) f(x_2).$$

Lemma 1.3. If f is a twice differentiable function, then f is convex if and only if $f''(x) \geq 0$.

Theorem 1.4 (Jensen's Inequality)**.** If f is a convex function, then

$$E[f(X)] \geq f(E[X]).$$

1.4 The Bernoulli and Binomial Random Variables

Definition 1.11. A binomial random variable X with parameters n and p, denoted by $B(n, p)$, is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$\Pr(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}.$$

That is, the binomial random variable X equals j when there are exactly j successes and $n - j$ failures in n independent experiments, each of which is successful with probability p.

1.5 Conditional Expectation

Definition 1.12.

$$E[Y|Z = z] = \sum_y y \Pr(Y = y|Z = z),$$

where the summation is over all y in the range of Y.

Lemma 1.4.

$$E[X] = \sum_y \Pr(Y = y) E[X|Y = y]$$

Example 1.4. Two dice X_1 is the value of first die, X_2 is the value of the second die, $X = X_1 + X_2$. Then

$$E[X|X_1 = 3] = \sum_x x \Pr(X = x|X_1 = 3) = \cdots = \frac{13}{2},$$
$$E[X_1|X = 4] = \sum_x x \Pr(X_1 = x|X = 4) = \cdots = 2.$$

Definition 1.13. For two random variables Y and Z, $E[Y|Z]$ is the random variable $f(Z)$ that takes the value $E[Y|Z = z]$ when $Z = z$,

$$f(z) = E[Y|Z = z].$$

Example 1.5. Two dice, $X = X_1 + X_2$,

$$E[X|X_1] = \sum_x x \Pr(X = x|X_1) = \sum_{x=x_1+1}^{x_1+6} x\frac{1}{6} = X_1 + \frac{7}{2}.$$

$E[X|X_1]$ is a random variable, so we can take its expectation:

$$E[E[X|X_1]] = E[X_1 + 7/2] = E[X_1] + 7/2 = 7 = E[X].$$

Theorem 1.5.

$$E[Y] = E[E[Y|Z]].$$

1.6 Geometric Distribution

A geometric random variable X with parameter p is given by the distribution

$$\Pr(X = n) = (1 - p)^{n-1}p.$$

for $n = 1, 2, \ldots$. Suppose $\Pr(H) = p$ and $\Pr(T) = 1 - p$ where H stands for the events that we flip heads while T is for tails. $\Pr(X = n)$ is the probability it takes n trials to obtain the first head.

$$\sum_{n=1}^{\infty} \Pr(X = n) = \sum_{n=1}^{\infty}(1 - p)^{n-1}p = p\sum_{n=1}^{\infty}(1 - p)^{n-1} = p\frac{1}{1 - (1 - p)} = 1.$$

Example 1.6 (Brancing Process)**.** Process S that recursively spawns new copies of S. The number of processes spawned by S is a binomial random variable with parameters n and p. Suppose we start with one call to S. What is the expected number of trial of total calls to S?

Algorithm 4:

```
Function S():
    for i = 1 to n do
        flips p-binomial coin
        if heads then
            call S()
    end
end
```

Let Y_i be the number of processes in generation i.

(a) Generation 0: 1 process, $Y_0 = 1$, $E[Y_0] = 1$.
(b) Generation 1: $E[Y_1] = np$.
(c) Generation $i, i \geq 2$: Suppose $Y_{i-1} = y$. There are y processes. For each k, $1 \leq k \leq y$, $k + Z_k$ be the number of processes spawned by the kth process. $E[Z_k] = np$, $Y_i = \sum_{k=1}^{Y_{i-1}} Z_k$. Then

$$\begin{aligned}
E[Y_i|Y_{i-1} = y] &= E\left[\sum_{k=1}^{y} Z_k | Y_{i-1} = y\right] \\
&= \sum_{k=1}^{y} E[Z_k|Y_{i-1} = y] \\
&= \sum_{k=1}^{y}\sum_{j=0}^{n} j \Pr(Z_k = j|Y_{i-1} = y) \\
&= \sum_{k=1}^{y}\sum_{j=0}^{n} j \Pr(Z_k = j) \\
&= \sum_{k=1}^{y} E[Z_k] \\
&= \sum_{k=1}^{y} np \\
&= ynp.
\end{aligned}$$

It follows that $E[Y_i|Y_{i-1}] = Y_{i-1}np$, then

$$E[Y_i] = E[EY_i|Y_{i-1}] = E[Y_i np] = E[Y_i]np.$$

By induction, $E[Y_i] = (pn)^i$. Let Y be the total number of processes spawned,

$$Y = \sum_{i=1}^{\infty} Y_i.$$

Then

$$E[Y] = E\left[\sum_{i=1}^{\infty} Y_i\right] = \sum_{i=1}^{\infty} E[Y_i] = \sum_{i=1}^{\infty} (np)^i = \begin{cases} \infty & \text{if } np \geq 1, \\ \frac{1}{1-np} & \text{if } np < 1. \end{cases}$$

1.7 Coupon Collector's Problem

Each box of cereal contains one of N coupons, chosen uniformly at random. How many boxes do you need to collect all n coupon? (How many times would an Nsided die be rolled before we set all n sides?) Let X be the number of boxes bought to obtain all n boxes, X_i be the number of boxes bought to obtain exactly $i - 1$ coupons. Then

$$X = \sum_{i=1}^{n} X_i.$$

It is known that $X_1 = 1$, while we have $i-1$ coupons, the probability of obtaining a new coupon is

$$\frac{n-(i-1)}{n} = 1 - \frac{i-1}{n} = p_i.$$

Then X_i is a geometric random variable with parameter p_i. Then

$$E[X_i] = \frac{1}{p_i} \implies E[X] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum j = 1^n \frac{1}{j} = nH_n,$$

where $\ln n \leq H_n \leq \ln n + 1$.

Table 1: caption

n	nH_n
2	3
3	5.5
4	8.3
5	11.3

2 Moments and Derivations

2.1 Markov's Inequality

Let X be a random variable that assumes only nonnegative values, then

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

Proof. Define an indicator random variable I by

$$I = \begin{cases} 1 & \text{if } x \geq a, \\ 0 & \text{otherwise.} \end{cases}$$

Then $I \leq X/a$ because $X \geq 0$:

- If $X \geq a$, then $I = 1$ and $1 \leq X/a$.
- If $X < a$, then $I = 0$ and $0 \leq X/a$.

Therefore,

$$\Pr(X \geq a) = \Pr(I = 1) = E[I] \leq E[X/a] = E[X]/a.$$

□

Example 2.1 (Jack Lutz)**.** If the average man is 6 feet tall, then at most half of all men are 12 feet tall.

Example 2.2. Let X be a binomial$(n, 1/2)$ random variable, $E[X] = n/2$, then

$$\Pr(X \geq 3n/4) \leq \frac{n/2}{3n/4} = \frac{2}{3}.$$

Definition 2.1 (Moment)**.** The kth moment of a random variable X is $E[X^k]$.

Definition 2.2 (Variance)**.** The *variance* of a random variable is

$$\mathrm{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2.$$

Definition 2.3 (Standard deviation)**.** The *standard deviation* of a random variable is

$$\sigma[X] = \sqrt{\mathrm{Var}[X]}.$$

Note that $\mathrm{Var} \geq 0$ by Jensen's inequality.

Example 2.3. Let $X = \text{binomial}(n, p)$.

$$E[X^2] = \sum_{i=1}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = \cdots = n(n-1)p^2 + np, \quad E[X] = np.$$

Then we have

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = np[(n-1)p + 1 - np] = np(1-p).$$

Theorem 2.1 (Chebyshev's inequality)**.** For any $a > 0$,

$$\Pr(|X - E[X]| \geq a) \leq \frac{\mathrm{Var}[X]}{a^2}.$$

Proof. Observe that

$$|X - E[X]| \geq a \iff (X - E[X])^2 \geq a^2.$$

Let $Y = (X - E[X])^2$. Then $E[Y] = \mathrm{Var}[X]$. By Markov's Inequality,

$$\Pr(|X - E[X]| \geq a) = \Pr((X - E[X])^2 \geq a^2) = \Pr(Y \geq a^2) \leq \frac{E[Y]}{a^2} = \frac{\mathrm{Var}[X]}{a^2}.$$

□

Example 2.4. Let X be a binomial$(n, 1/2)$ random variable, $E[X] = n/2$, then

$$\Pr(X \geq 3n/4) \leq \Pr(|X - E[X]| \geq n/4) \leq \frac{\mathrm{Var}[X]}{(n/4)^2} = \frac{n/4}{(n/4)^2} = \frac{4}{n}.$$

Consider a BPP (bounded-error probabilistic polynomial time) algorithm M for a decision problem A.

- If $x \in A$, $\Pr[M \text{ accepts } x] \geq 2/3 \iff \Pr[M \text{ rejects } x] \leq 1/3$.
- If $x \notin A$, $\Pr[M \text{ accepts } x] \leq 1/3 \iff \Pr[M \text{ rejects } x] \geq 2/3$.

Consider running the algorithm n times:

$$X = X_1 + X_2 + \cdots + X_n,$$

$$\Pr(X_i = 1) = p = \Pr[M \text{ accepts } x], \quad \Pr(X_i = 0) = 1 - p = \Pr[M \text{ rejects } x].$$

Then

$$E[X] = np = \begin{cases} \geq 2n/3 & \text{if } x \in A, \\ \leq n/3 & \text{if } x \notin A. \end{cases}$$

Take a majority vote,

- If $X \geq n/2$, we declare $x \in A$.
- If $X \leq n/2$, we declare $x \notin A$.

What is the probability that we are correct? X is a binomial(n, p) random variable with mean $E[X] = np$ and variance $\text{Var}[X] = np(1-p)$. Suppose $p \geq 2/3$,

$$\Pr(X < n/2) \leq \Pr(|X - E[X]| \geq n/6) \leq \frac{\text{Var}[X]}{(n/6)^2} \leq \frac{np(1-p)}{(n/6)^2} = \frac{36p(1-p)}{n} \leq \frac{8}{n}.$$

Similar for $p \leq 1/3$.

2.2 Chernoff Bounds

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\delta^2/3}, \quad \Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2},$$

where $\mu = E[X]$. Suppose $p \geq 2/3$,

$$\begin{aligned}
\Pr(X < n/2) &= \Pr(X < \frac{np}{2p}) \\
&= \Pr(X < \frac{\mu}{2p}) \\
&= \Pr(X < \frac{1}{2p}\mu) \\
&= \Pr(X < (1 - \frac{2p-1}{2p})\mu) \\
&= \Pr(X < (1-\delta)\mu) \\
&\leq e^{-\mu\delta^2/2} \\
&\leq e^{-n/48},
\end{aligned}$$

where $\delta = (2p-1)/(2p) = 1 - 1/(2p) \geq 1/4$.

2.3 Satisfiability

SAT $\in$ NP, SAT $\in$ P, given ϕ with boolean variables, $\vee$, $\wedge$, $\neg$.

$$\phi \in \text{SAT}, \phi \notin \text{SAT}.$$

2.4 Moment Generating Functions (MGFs)

Definition 2.4. The MGF of a random variable X is

$$M_X(t) = E[e^{tX}].$$

Let $f(t) = e^{tX}$, then

$$\begin{aligned}
f'(t) &= Xe^{tX} \\
f^{(2)}(t) &= X^2e^{tX} \\
f^{(3)}(t) &= X^3e^{tX} \\
&\vdots \\
f^{(n)}(t) &= X^ne^{tX}
\end{aligned}$$

Therefore, $f^{(n)}(0) = X^n$.

Theorem 2.2.

$$E[X^n] = M_X^{(n)}(0).$$

Example 2.5. Let X be geometric with parameter p. Then

$$\Pr[X = n] = p(1-p)^{n-1}.$$

Then the moment generating function is

$$\begin{aligned} M_X(t) &= E[e^{tX}] \\ &= \sum_{n=1}^{\infty} \Pr[X = n] e^{tn} \\ &= \sum_{n=1}^{\infty} p(1-p)^{n-1} e^{tn} \\ &= \frac{p}{1-p} \sum_{n=1}^{\infty} [(1-p)e^t]^n \\ &= \frac{p}{1-p} \frac{1}{e^{-t} - 1 + p}. \end{aligned}$$

Example 2.6. Indicator random variable, X, with $\Pr(X = 1) = p$ and $\Pr(X = 0) = 1 - p$, then

$$M_X(t) = E[e^{tX}] = pe^{t\cdot 1} + (1-p)e^{t\cdot 0} = pe^t + (1-p) = 1 + p(e^t - 1).$$

That implies

$$\begin{aligned} M_X'(t) = pe^t &\implies M_X'(t) = p = E[X] \\ M_X''(t) = pe^t &\implies M_X''(t) = p = E[X^2]. \end{aligned}$$

Theorem 2.3. If X and Y aer independent random variables, then

$$M_{X+Y}(t) = M_X(t) M_Y(t).$$

Example 2.7 (Poisson Trials)**.** Let $X_1, X_2, \ldots, X_n$ be independent with $\Pr(X_i = 1) = p_i$ and $\Pr(X_i = 0) = 1 - p_i$. Let $X = \sum_{i=1}^n X_i$ and $\mu = E[X] = \sum_{i=1}^n p_i$. The moment generating function is

$$M_{X_i}(t) = 1 + p_i(e^t - 1) \le e^{p_i(e^t-1)} \implies M_X(t) = \prod_{i=1}^n M_{X_i}(t) \le \prod_{i=1}^n e^{p_i(e^t-1)} = e^{\sum_{i=1}^n p_i(e^t-1)} = e^{(e^t-1)\mu}.$$

Proposition 2.1 (Moment General Bounds)**.** Let $X \sim \text{Poisson}(\theta)$. For all $\delta > 0$,

$$\Pr(X \ge (1+\delta)\mu) \le \left[\frac{e^\delta}{(1+\delta)^{(1+\delta)}} \right]^\mu.$$

For all $\delta \in (0, 1)$,

$$\Pr(X \le (1-\delta)\mu) \le \left[\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \right]^\mu.$$

Proof.

$$\Pr(X \geq (1+\delta)\mu) = \Pr(e^{tX} \geq e^{t(1+\delta)\mu}) \leq \frac{E[e^{tX}]}{e^{t(1+\delta)\mu}} = \frac{e^{(e^t-1)\mu}}{e^{t(1+\delta)\mu}} = \left[\frac{e^{(e^t-1)}}{e^{t(1+\delta)}}\right]^{\mu} = e^{[(e^t-1)-t(1+\delta)]\mu}$$

Choose $t > 0$ to minimize $f(t) = e^t - 1 - t(1+\delta)$, $f'(t) = e^t - (1+\delta)$, let $f'(t) = 0 \implies t = \ln(1+\delta)$. Therefore,

$$\Pr(X \geq (1+\delta)\mu) \leq e^{[(e^t-1)-t(1+\delta)]\mu}\Big|_{t=\ln(1+\delta)} = \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu}.$$

Similar for the other bound by letting $t = \ln(1-\delta)$. □

Corollary 2.1. For $\delta \in (0, 1]$,

$$\Pr(X \geq (1+\delta)\mu) \leq e^{-\mu\sigma^2/3}.$$

For $\delta \in (0, 1)$,

$$\Pr(X \leq (1-\delta)\mu) \leq e^{-\mu\sigma^2/2}.$$

Proof. Calculus to show

$$\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}} \leq \frac{1}{e^{\delta^2/3}}$$

and

$$\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}} \leq \frac{1}{e^{\delta^2/2}}.$$

□

Corollary 2.2. For $R = \alpha\mu$ where $\alpha > e$,

$$\Pr(X \geq R) \leq (\frac{e}{\alpha})^R.$$

In particular, if $R \geq 6\mu$,

$$\Pr(X \geq R) \leq 2^{-R}.$$

Proof.

$$\Pr(X \geq R) \leq \left[\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} \leq \left[\frac{e^{1+\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} = \left[\frac{e^{1+\delta}}{(1+\delta)^{(1+\delta)}}\right]^{\mu} = \left[\frac{e}{(1+\delta)}\right]^{(1+\delta)\mu} = \left[\frac{e}{\alpha}\right]^{\alpha\mu}.$$

□

Corollary 2.3.

$$\Pr(|X - \mu| \geq \delta\mu) \leq 2e^{-\mu\sigma^2/2}.$$

Example 2.8 (Uniform case)**.** Let $X \sim \text{binomial}(n, 1/2)$, $\Pr(X_i = 1) = 1/2$, $\mu = n/2$. Then

$$\Pr(X \geq (1+\delta)n/2) \leq e^{-n\delta^2/6}, \Pr(X \leq (1-\delta)n/2) \leq e^{-n\delta^2/4}.$$

Another approach. Let $Y = Y_1 + Y_2 + \cdots + Y_n$, where $\Pr(Y_i = 1) = \frac{1}{2} = \Pr(Y_i = -1)$. Then $E[Y] = 0$.

$$Y_i = 2X_i - 1 \implies X_i = \frac{Y_i + 1}{2} \implies Y = 2X - n, X = \frac{Y}{2} + \frac{n}{2}.$$

Proposition 2.2 (Bound)**.** For any $a > 0$,

$$\Pr(Y \geq a) \leq e^{-a^2/(2n)}$$

and

$$\Pr(Y \leq -a) \leq e^{-a^2/(2n)}$$

and

$$\Pr(|Y| \geq a) \leq 2e^{-a^2/(2n)}.$$

Proof.

$$E[e^{tY_i}] = \frac{1}{2}e^t + \frac{1}{2}e^{-t}. - \frac{1}{2}(e^t + e^{-t}).$$

Recall that $e^x = \sum_{i=0}^{\infty} \frac{x^i}{x!}$, then

$$\begin{aligned}
e^t &= 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots \\
e^{-t} &= 1 - t + \frac{t^2}{2!} - \frac{t^3}{3!} + \frac{t^4}{4!} + \cdots \\
e^t + e^{-t} &= 2(1 + \frac{t^2}{2!} + \frac{t^4}{4!} + \frac{t^6}{6!} + \cdots) \\
&= 2\sum_{i=0}^{\infty} \frac{t^{2i}}{(2i)!} \\
&\leq 2\sum_{i=0}^{\infty} \frac{t^{2i}}{i!\, 2^i} \\
&= 2\sum_{i=0}^{\infty} \frac{(t^2/2)^i}{i!} \\
&= 2e^{t^2/2}.
\end{aligned}$$

Therefore, $E[e^{tY_i}] \leq e^{t^2/2}$. Then

$$E[e^{tY}] = \prod_{i=1}^{n} E[e^{tY_i}] \leq e^{nt^2/2}.$$

Then

$$\Pr(Y \geq a) = \Pr(e^{tY} \geq e^{ta}) \leq \frac{E[e^{tY}]}{e^{ta}} \leq e^{nt^2/2 - ta} = e^{t(nt/2 - a)}$$

It can be minimized when $t = a/n$. It follows that

$$\Pr(Y \geq a) \leq e^{a/n(-a/2)} = e^{-a^2/(2n)}.$$

Similar to $\Pr(Y \leq -a) \leq e^{-a^2/(2n)}$. □

Corollary 2.4. For any $a > 0$,

$$\Pr(X \geq \mu + a) \leq e^{-a^2/\mu} = e^{-2a^2/n}, \Pr(X \geq \mu - a) \leq e^{-a^2/\mu}.$$

For any $\delta > 0$,

$$\Pr[X \geq (1+\delta)\mu] \leq e^{-\delta^2\mu} = e^{-n\delta^2/2}, \Pr[X \geq (1-\delta)\mu] \leq e^{-\delta^2\mu}.$$

Proof. Given that $Y = 2X - n$. We have

$$X \geq \mu + a = \frac{n}{2} + a \iff Y \geq 2(\frac{n}{2} + a) - n = 2a.$$

- Then
$$\Pr(X \geq \mu + a) = \Pr(Y \geq 2a) \leq e^{-(2a)^2/(2n)} = e^{-2a^2/n} = e^{-a^2/\mu}.$$
- Then
$$\Pr(X \geq (1+\delta)\mu) = \Pr(X \geq \mu + \delta\mu) \leq e^{-(\delta\mu)^2/\mu} = e^{-\delta^2\mu}.$$

□

Example 2.9 (Estimating a Probability)**.** Suppose we have a large population of size N. Each individual satisfies a property ϕ or does not. Let P be the fractional of the population that satisfies ϕ. If we choose an individual I uniformly at random,

$$\Pr(I \text{ satisifies } \phi) = p.$$

Choose n individuals uniformly at random,

$$X = \begin{cases} 1 & \text{if } i\text{th individual satisfies } \phi \\ 0 & \text{otherwise.} \end{cases}$$

The $S_n = \sum_{i=1}^{n} X_i$. Let $\tilde{P}_n = S_n/n$ be the *empirical probability* observed with n individuals.

$$E[\tilde{P}_n] = p.$$

How many samples n should we take to be fairly certain that $\tilde{p}_n = p$.

Definition 2.5. A $1-\gamma$ *confidence interval* is an interval $[\tilde{p}_n - \delta, \tilde{p}_n + \delta]$ such that

$$\Pr(p \in [\tilde{p}_n - \delta, \tilde{p}_n + \delta]) = \Pr(|p - \tilde{p}_n| \leq \delta) \geq 1 - \gamma.$$

with three parameters γ, δ, n. Then

$$\begin{aligned} \Pr(|\tilde{p}_n - p| > \delta) &= \Pr(|S_n/n - p| > \delta) \\ &= \Pr(|S_n - np| > n\delta) \\ &= \Pr(|S_n - np| > \delta/n \cdot np) \\ &\leq 2e^{-np(\delta/p)^2/3} \\ &= 2e^{-n\delta^2/(3p)}. \end{aligned}$$

Fix δ and n. What is the confidence $1-\gamma$? Given that $\gamma \leq 2e^{-n\delta^2/(3p)}$. But p is unknown. Suppose we know $p \geq 1/3$, then $\gamma \leq 2e^{-n\delta^2/p}$. Fix γ and δ. How many samples? It is known that $\gamma \leq 2e^{-n\delta^2/(3p)}$ if $\ln\gamma \geq \ln 2 + \frac{-n\delta^2}{3p}$. Need $n \geq 3p/\delta^2 \ln(2/\delta)$. So $3/\delta^2 \ln(2/\gamma)$ samples suffice, i.e., $\Theta(\frac{1}{\delta^2}\ln(1/\gamma))$ samples.