1 Homework 1

1. Prove the union bound: Suppose $\{A_i\}_{i=1}^n$ is a collection of events. Then, $Pr[\cup_{i=1}^n A_i] \leq \sum_{i=1}^n Pr[A_i]$.

Proof. Let $t \in [n]$. Prove the above by induction on t. Certainly the inequality holds for the base case where $t = 1$, since $Pr[A_1] = Pr[A_1]$. Suppose the inequality holds for $n = t - 1$ and if I can prove that the inequality holds for $n = t$ then finish the proof.

$$\begin{aligned}
Pr[\cup_{i=1}^n A_i] &= Pr[\cup_{i=1}^{n-1} A_i \cup A_n] \\
&= Pr[\cup_{i=1}^{n-1} A_i] + Pr[A_n] - Pr[(\cup_{i=1}^{n-1} A_i) \cap A_n] \\
&\leq \sum_{i=1}^{n-1} Pr[A_i] + Pr[A_n] \\
&= \sum_{i=1}^{n} Pr[A_i]
\end{aligned}$$

□

2. Prove the linearity of expectation: Suppose for each $1 \leq i \leq n$, X_i is a random variable and a_i is a real number. Then, $E[\sum_{i=1}^n a_i X_i] = \sum_{i=1}^n a_i E[X_i]$. [Lin]

Definition 1.1 ([RM95]). The expectation of a random variable X with density function p is defined as $E[X] = \sum_x xp(x)$, where the summation is over the range of X.

Lemma 1.2. *For any random variable X and Y, $E[X + Y] = E[X] + E[Y]$.*

Proof. Prove by definition 1.1. $\sum_x$ and $\sum_y$ means the summation is over the range of X and Y.

$$\begin{aligned}
E[X+Y] &= \sum_x \sum_y (x+y) Pr[(X=x) \cap (Y=y)] \\
&= \sum_x \sum_y xPr[(X=x) \cap (Y=y)] + yPr[(X=x) \cap (Y=y)] \\
&= \sum_x x \sum_y Pr[(X=x) \cap (Y=y)] + \sum_y y \sum_x Pr[(X=x) \cap (Y=y)] \\
&= \sum_x xPr[X=x] + \sum_y yPr[Y=y] \\
&= \sum_x xp(x) + \sum_y yp(y) \\
&= E[X] + E[Y]
\end{aligned}$$

□

Lemma 1.3. *For any random variable X and any constant $a \in \mathbb{R}$, $E[aX] = aE[X]$.*

Proof. Prove by definition 1.1. $\sum_x$ means the summation is over the range of X.

$$\begin{aligned}
E[aX] &= \sum_x axp(x) \\
&= a \sum_x xp(x) \\
&= aE[X]
\end{aligned}$$

□

By combining Claim 1.2 and 1.3, we have:

$$E[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} E[a_i X_i]$$
$$= \sum_{i=1}^{n} a_i E[X_i]$$

3. Suppose X, $Y : \Omega \to \mathbb{Z}$ are independent random variables taking integer values. In this question, we shall prove the basic result $E[XY] = E[X]E[Y]$.

Proof. Prove by definition 1.1. $\sum_x$ and $\sum_y$ means the summation is over the range of X and Y.

$$\begin{aligned}
E[XY] &= \sum_x \sum_y xyPr[(X = x) \cap (Y = y)] \\
&= \sum_x \sum_y xyPr[X = x]Pr[Y = y] \\
&= \sum_x xPr[X = x] \sum_y yPr[Y = y] \\
&= \sum_x xp(x) \sum_y yp(y) \\
&= E[X]E[Y]
\end{aligned}$$

□

4. Let ϕ be a 3-CNF formula with m clauses and n variables.
(a) Let Y be the random variable denoting the number of unsatisfied clauses. Compute $E[Y]$.
(b) Find a suitable upper bound for $Pr[Y > \frac{3m}{16}]$, and conclude that the randomized algorithm finds an assignment satisfying at least $\frac{13m}{16}$ clauses with probability at least $\frac{1}{3}$.

Problem description: There are n Boolean variables. $\phi(x_1, x_2, ..., x_n)$ is denoted as $C_1 \vee C_2 \vee ... \vee C_m$. Each clause C_i is a disjunction of 3 literals from 3 different variables. A literal is either a variable or its negation. A clause is satisfied if at least one of its 3 literals evaluates to TRUE. The goal is to find an assignment of the variables so that as many clauses as possible are satisfied.

Randomized algorithm: Independently for each variable, assign its value to be TRUE or FALSE, each with probability $\frac{1}{2}$.

(a) Define Y_i as a random variable which is related to clause C_i. Y_i takes value 1 if the clause C_i is unsatisfied and 0 otherwise. Then the number of unsatisfied clauses $Y = \sum_{i=1}^{m} Y_i$. For each clause C_i, there are 3 independent Boolean variables. And the value of each Boolean variable is independently assigned TRUE or FALSE, each with probability $\frac{1}{2}$. Since for each clause C_i is unsatisfied only if all the three literals evaluate to FALSE, for each Y_i, takes value 1 with probability $\frac{1}{8}$ and takes value 0 with probability $\frac{7}{8}$. By Lemma 1.2:

$$\begin{aligned}
E[Y] &= E[\sum_{i=1}^{m} Y_i] \\
&= \sum_{i=1}^{m} E[Y_i] \\
&= m(1 \cdot \frac{1}{8} + 0 \cdot \frac{7}{8}) \\
&= \frac{m}{8}
\end{aligned}$$

(b)By Markov's inequality, for all $\alpha > 0$ we have

$$Pr[Y \geq \alpha] \leq \frac{E[Y]}{\alpha}$$

If set $\alpha = \frac{3m}{16}$,we have

$$Pr[Y \geq \frac{3m}{16}] \leq \frac{m}{8} \cdot \frac{16}{3m} = \frac{2}{3}$$

Thus,

$$Pr[Y > \frac{3m}{16}] \leq Pr[Y \geq \frac{3m}{16}] \leq \frac{2}{3} \implies Pr[Y \leq \frac{3m}{16}] \geq \frac{1}{3}$$

Thus, the randomized algorithm finds an assignment satisfying at least $\frac{13m}{16}$ clauses with probability at least $\frac{1}{3}$.

5. Let $G = (V, E)$ be a graph. Recall the randomized algorithm mentioned in class for finding a cut $C \subset V$ for the graph G, namely, a point $v \in V$ is included in C independently with probability $\frac{1}{2}$. Assume that to make this decision takes 1 independent random bit for each point.

Let $E(C) := \{\{u, v\} \in E : u \in C, v \in V \setminus C\}$ be the edges in the cut. It is shown that $E[|E(C)|] = \frac{|E|}{2}$. The goal of this question is to design another randomized algorithm with better guarantees. Let $0 < \epsilon < 1$ and $0 < \delta < 1$. We shall design a randomized algorithm that, with failure probability at most δ, returns a cut C such that $|E(C)| \geq (\frac{1-\epsilon}{2}) \cdot |E|$.

(a) Given an upper bound on the failure probability that the above randomized procedure returns a cut such that the number of edges $|E(C)|$ is less than $(\frac{1-\epsilon}{2}) \cdot |E|$.

(b) Show that by repeating the above randomized procedure, it is possible to obtain a better randomized algorithm with failure probability at most δ. Compute the number of independent random bits used by your algorithm. [Max][Rev]

(a)

$$Pr[|E(C)| \leq (\frac{1-\epsilon}{2})|E|] = Pr[|E| - |E(C)| \geq |E| - \frac{1-\epsilon}{2}|E|]$$

By Markov's inequality, we have:

$$Pr[|E| - |E(C)| \geq |E| - \frac{1-\epsilon}{2}] \leq \frac{E[|E| - |E(C)|]}{|E| - \frac{1-\epsilon}{2}|E|}$$

By Lemma 1.2, we have:

$$E[|E| - |E(C)|] = |E| - E[|E(C)|] = |E| - \frac{|E|}{2} = \frac{|E|}{2}$$

Thus,

$$Pr[|E(C)| \leq (\frac{1-\epsilon}{2})|E|] \leq \frac{\frac{|E|}{2}}{|E| - \frac{1-\epsilon}{2}|E|} = \frac{1}{1+\epsilon}$$

(b)The random algorithm is described as follows.

(i) For each vertex $v \in V$, independently assign it a number 0 or 1, each with probability $\frac{1}{2}$. Then we can get a set of vertices C which consists of the vertices with number 1.

(ii) Independently repeat step (i) for k times, here, $k = -\frac{2}{\epsilon} \ln \delta$, which is the answer to the question. We can get k cuts of graph G, named $C_1, C_2, ..., C_k$.

(iii) For each cut C_i, compute $|E(C_i)|$, which is the number of edges in cut C_i. Then pick C_j from all C_is with the maximum $|E(C_i)|$. And then output C_j as the final result.

Analysis: By (a), for any C_i we have

$$Pr[|E(C_i)| \leq (\frac{1-\epsilon}{2}) \cdot |E|] \leq \frac{1}{1+\epsilon}$$

Since C_j is with the maximum $|E(C_i)|$ among all C_js, we have

$$\begin{aligned} Pr[|E(C_j)| \leq (\frac{1-\epsilon}{2}) \cdot |E|] &= \prod_{i=1}^{k} Pr[|E(C_i)| \leq (\frac{1-\epsilon}{2}) \cdot |E|] \\ &\leq (\frac{1}{1+\epsilon})^k \end{aligned}$$

Since for $0 < \epsilon < 1$, $1+\epsilon \geq e^{\frac{\epsilon}{2}}$, we have

$$\begin{aligned} Pr[|E(C_j)| \leq (\frac{1-\epsilon}{2}) \cdot |E|] &\leq e^{-\frac{\epsilon k}{2}} \\ &= e^{-\frac{\epsilon}{2}(-\frac{2}{\epsilon}\ln\delta)} \\ &= \delta \end{aligned}$$

6.(a) Suppose X is a discrete random variable that takes only a countable number of non-negative values. Prove that $E[X] = \int_0^\infty Pr[X \geq t]\, dt$. (If you do not like to deal with infinite support, you may assume X only takes a finite number of values.)
(b) Assume that for all non-negative random variables X, $E[X] = \int_0^\infty Pr[X \geq t]\, dt$. Derive a similar formula for $E[X]$ when X is any real-valued random variable.
(a) First assume discrete random variable X only takes a finite number n of non-negative values, $x_1, x_2, ..., x_n$ with probability $p_{x1}, p_{x2}, ..., p_{xn}$. Note that $\sum_{i=1}^{n} p_{xi} = 1$. By definition, $E[X] = \sum_{i=1}^{n} x_i p_{xi}$. For simplicity, assume $0 \leq x_1 < x_2 < ... < x_n$. Then we can get a piecewise function as follows.

$$Pr[X \geq t] = \begin{cases} 1 & 0 \leq t \leq x_1 \\ 1 - p_{x1} & x_1 < t \leq x_2 \\ 1 - \sum_{i=1}^{2} p_{xi} & x_2 < t \leq x_3 \\ 1 - \sum_{i=1}^{3} p_{xi} & x_3 < t \leq x_4 \\ \dots & \dots \\ 1 - \sum_{i=1}^{n-1} p_{xi} & x_{n-1} < t \leq x_n \\ 0 & x_n \leq t \end{cases}$$

Thus we have:

$$\begin{aligned} \int_0^\infty Pr[X \geq t]\, dt = {} & (x_1 - 0) \cdot 1 + (x_2 - x_1)(1 - p_{x1}) + (x_3 - x_2)(1 - \sum_{i=1}^{2} p_{xi}) + \\ & (x_4 - x_3)(1 - \sum_{i=1}^{3} p_{xi}) + ... + (x_n - x_{n-1})(1 - \sum_{i=1}^{n-1} p_{xi}) + 0 \cdot \infty \\ & = x_1(1 - 1 + p_{x1}) + x_2(1 - p_{x1} - (1 - \sum_{i=1}^{2} p_{x_i})) + x_3(1 - \sum_{i=1}^{2} p_{x_i} - (1 - \sum_{i=1}^{3} p_{x_i})) \\ & \quad + ... + x_n(1 - \sum_{i=1}^{n-1} p_{x_i} - (1 - \sum_{i=1}^{n} p_{x_i})) \\ & = x_1 p_{x1} + x_2 p_{x2} + x_3 p_{x3} + ... + x_n p_{xn} \\ & = E[X] \end{aligned}$$

Then consider the situation when discrete random variable X takes infinite number of non-negative values (but the number of values is still countable). Assume X can take values $0 \leq x_1 < x_2 < x_3 < ...$ with probability $p_{x1}, p_{x2}, p_{x3}, ...$ and $\int_0^\infty Pr[X = t]\,dt = 1$. By definition, $E[X] = \sum_{k=0}^{\infty} kPr[X = k]$ and $Pr[X \geq t] = \sum_{k=t}^{\infty} Pr[X = k]$. Thus, we have:

$$\begin{aligned}
\int_0^\infty Pr[X \geq t]\,dt &= \int_0^\infty \sum_{k=t}^{\infty} Pr[X = k]\,dt \\
&= \int_0^{x_1} \sum_{k=x_1}^{\infty} Pr[X = k]\,dt + \int_{x_1}^{x_2} \sum_{k=x_2}^{\infty} Pr[X = k]\,dt + \int_{x_2}^{x_3} \sum_{k=x_3}^{\infty} Pr[X = k]\,dt + ... \\
&= (x_1 - 0) \sum_{k=x_1}^{\infty} Pr[X = k] + (x_2 - x_1) \sum_{k=x_2}^{\infty} Pr[X = k] + (x_3 - x_2) \sum_{k=x_3}^{\infty} Pr[X = k] + ... \\
&= x_1 (\sum_{k=x_1}^{\infty} Pr[X = k] - \sum_{k=x_2}^{\infty} Pr[X = k]) + x_2 (\sum_{k=x_2}^{\infty} Pr[X = k] - \sum_{k=x_3}^{\infty} Pr[X = k]) + \\
&\quad x_3 (\sum_{k=x_3}^{\infty} Pr[X = k] - \sum_{k=x_2}^{\infty} Pr[X = k]) + ... \\
&= x_1 Pr[X = x_1] + x_2 Pr[X = x_2] + x_3 Pr[X = x_3] + ... \\
&= \sum_{k=0}^{\infty} kPr[X = k] \\
&= E[X]
\end{aligned}$$

(b) Let X be a continuous real-valued random variable. The probability density function of X is $f(x) = \lim_{h \to 0} \frac{1}{h} \cdot Pr[x \leq X \leq x + h]$. By definition, $\int_{\mathbb{R}} f(x)\,dx = 1$, $E[X] = \int_{\mathbb{R}} xf(x)\,dx$. $E[X]$ can also be written as $E[X] = -\int_{-\infty}^{0} Pr[X \leq t]\,dt + \int_0^\infty Pr[X \geq t]\,dt$

Proof.

$$\begin{aligned}
\int_0^\infty Pr[X \geq t]\,dt &= \int_0^\infty \int_t^\infty f(k)\,dk\,dt \\
&= \int_0^\infty kf(k)\,dk \\
&= \int_0^\infty xf(x)\,dx
\end{aligned}$$

$$\begin{aligned}
-\int_{-\infty}^{0} Pr[X \leq t]\,dt &= -\int_{-\infty}^{0} \int_{-\infty}^{t} f(k)\,dk\,dt \\
&= -\int_0^{-\infty} \int_t^{-\infty} f(k)\,dk\,dt \\
&= -\int_0^{-\infty} kf(k)\,dk \\
&= \int_{-\infty}^{0} xf(x)\,dx
\end{aligned}$$

$$
\begin{aligned}
E[X] &= \int_{\mathbb{R}} xf(x)\,dx \\
&= \int_{-\infty}^{0} xf(x)\,dx + \int_{0}^{\infty} xf(x)\,dx \\
&= -\int_{-\infty}^{0} Pr[X \leq t]\,dt + \int_{0}^{\infty} Pr[X \geq t]\,dt
\end{aligned}
$$

□

References

[Lin] https://proofwiki.org/wiki/Linearity_of_Expectation_Function.

[Max] https://www.cs.ubc.ca/ nickhar/W15/Lecture2Notes.pdf.

[Rev] https://www.planetmath.org/reversemarkovinequality.

[RM95] Prabhakar Raghavan Rajeev Motwani. *Randomized Algorithms*, chapter APPENDIX C: Basic Probability Theory, pages 438–446. Cambridge University Press, 1995.