Problem Set Solutions

Point totals are in the margin. Max total number of points was 24. Problem 4 was not graded.

1. Variations on Chernoff

(a) As indicated we start from Theorem 13.1 in Lecture 13, namely

\[ \Pr[X \geq \mu + \lambda] \leq \exp\{-nH_p(p + \frac{\lambda}{n})\} = \exp\{-n((p + \frac{\lambda}{n}) \ln\left(\frac{p+\lambda/n}{p}\right) + ((1-p) - \frac{\lambda}{n}) \ln\left(\frac{1-p-\lambda/n}{1-p}\right)\} \]

for \(0 \leq \lambda \leq n - \mu\). Our task is to prove that this final expression is at most \(\exp(-\frac{2\lambda^2}{n})\). Equivalently, writing \(z = \frac{\lambda}{n}\), we need to show that

\[ f(z) \equiv (p + z) \ln\left(\frac{p+z}{p}\right) + (1 - p - z) \ln\left(\frac{1-p-z}{1-p}\right) - 2z^2 \geq 0 \]

on the interval \(0 \leq z \leq 1 - p\). Note that, since the bounds in Theorem 13.1 are symmetric, this will also imply the same bound for the lower tail.

To get the above bound on \(f(z)\), some elementary calculus is required. Take derivatives twice w.r.t. \(z\) to get

\[ f'(z) = \ln \left( \frac{(1-p)(p+z)}{p(1-p-z)} \right) - 4z; \quad f''(z) = \frac{1}{(1-p-z)(p+z)} - 4. \]

Now note that \(f(0) = 0\), so it suffices to show that \(f(z)\) is non-decreasing for \(0 \leq z \leq 1 - p\), i.e., that \(f'(z) \geq 0\). But \(f'(0) = 0\) also, so it is enough to show that \(f''(z) \geq 0\). But it is clear that \(f''(z)\) is minimized for \((p+z) \in [0,1]\) at \(p+z = \frac{1}{2}\), and that this minimum value is \(4 - 4 = 0\). This completes the proof.

To compare this with Angluin’s bound, put \(\lambda = \beta \mu\) in the above to get the tail bound \(\exp(-2\beta^2 \mu^2 / n)\). Comparing this with \(\exp(-\beta^2 \mu / 2)\), we see that the first bound is better only when \(\mu > \frac{\beta}{2}\), and (since \(\mu \leq n\)) that it is better by at most a factor of 4 in the exponent. Similar observations hold for the upper tail. On the other hand, if \(\mu \ll n\), then Angluin’s bound can be much better. E.g., if \(\mu \sim \log n\) then Angluin’s bound gives a tail probability of order \(n^{-c}\) for any constant \(\beta\), while the above bound is vacuous.

(b) First we verify the fact claimed in the hint. Write \(p\) for the common expectation of \(Y\) and \(Z\). Then we have

\[ E(f(Z)) = (1-p)f(0) + pf(1) \]

\[ = E((1-Y)f(0) + Yf(1)) \]

\[ \geq E(f(Y)). \]

The first line here follows since \(Z\) is \((0,1)\)-valued, the second line from the expectation of \(Y\), and the third from convexity of \(f\).

To prove that the Chernoff/Hoeffding bound still holds, we just need to modify one line of the proof in Lecture 13, namely equation (13.1). Applying the above fact to the convex function \(f(x) = e^{tx}\), and adopting the notation of the proof, we have

\[ E(e^{tX_i}) \leq E(e^{tZ_i}) = e^{tp_i} + 1 - p_i, \]

where \(Z_i\) is \((0,1)\)-valued with mean \(p_i = E(X_i)\). Since this is the only step that depends on the distribution of \(X_i\), the rest of the proof carries over as before.
(c) We start with the observation that, for a Poisson r.v. $X$ with parameter $\mu$ and any $t > 0$, we have

$$E[e^{tx}] = \sum_{k=0}^{\infty} e^{-\mu} \frac{\mu^k}{k!} e^{tk} = e^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu e^t)^k}{k!} = \exp(\mu (e^t - 1)).$$

Now, following the standard argument for such bounds, we have for any $t > 0$

$$\Pr[X \geq \mu + \lambda] = \Pr[e^{tx} \geq e^{(\mu + \lambda)t}] \leq e^{-(\mu + \lambda)t} E[e^{tx}] = e^{-(\mu + \lambda)t} \exp(\mu (e^t - 1)) = \exp(\mu (e^t - 1) - (\mu + \lambda)t).$$

By elementary calculus, the value of $t$ that minimizes this bound is given by $e^t = \frac{\mu + \lambda}{\mu}$. Plugging this in and tidying up gives the desired bound:

$$\Pr[X \geq \mu + \lambda] \leq \exp\left\{-[\mu + \lambda] \ln \left(\frac{\mu + \lambda}{\mu}\right) - \lambda\right\}.$$

Note for future reference that setting $\lambda = \beta \mu$ gives

$$\Pr[X \geq (1 + \beta)\mu] \leq \exp\{-\mu[(1 + \beta) \ln(1 + \beta) - \beta]\},$$

which is exactly the same as Angluin’s bound for the binomial distribution (Corollary 13.3). Moreover, if we follow the same argument as above for the lower tail, we obtain

$$\Pr[X \leq \mu - \lambda] \leq \exp\left\{-[\mu - \lambda] \ln \left(\frac{\mu - \lambda}{\mu}\right) + \lambda\right\},$$

and setting $\lambda = \beta \mu$ again recovers Angluin’s bound

$$\Pr[X \leq (1 - \beta)\mu] \leq \exp\{-\mu[(1 - \beta) \ln(1 - \beta) + \beta]\}.$$

2. The Quadratic Assignment problem

Since $E(C_x) = \frac{1}{2}\binom{n}{2} \equiv \mu$, and $C_x$ is the sum of $\binom{n}{2}$ independent r.v.’s on $[0, 1]$ each with mean $\frac{1}{2}$, we can apply the Chernoff bound of 2(a) (generalized as in 2(b)) to deduce

$$\Pr[C_x \geq \mu + n^{3/2} \sqrt{\log n \omega(n)}] \leq \exp\left(\frac{-2n^3 \log n \omega^2(n)}{\binom{n}{2}}\right) \leq \exp(-n \log n \omega'(n)),$$

where $\omega'(n) \to \infty$. Since the number of assignments $x$ is only $n! = \exp(n \ln n + O(n))$ by Stirling’s approximation, a union bound shows that

$$\Pr[\exists x \text{ s.t. } C_x \geq \mu + n^{3/2} \sqrt{\log n \omega(n)}] \to 0 \text{ as } n \to \infty.$$

[Actually, assuming log denotes base-2 or natural logarithm, we don’t really need the $\omega(n)$ as the constant in the Chernoff bound beats that in Stirling’s approximation.]

3. Codes in space

(a) Consider one pair of random strings of length $\ell$. The probability that they agree in more than $\ell \ell$ positions is the probability that the number of successes $X$ in $\ell$ independent trials, each with success probability $\frac{1}{a}$, exceeds $\ell \ell$. Using the Chernoff bound $\Pr[X > \mu + \lambda] \leq \exp(-2\lambda^2/n)$ from Corollary 13.2 of Lecture 13, with $\mu = \ell \ell$ and $\lambda = (\epsilon - \frac{1}{a})\ell$, we get

$$\Pr[X > \ell \ell] \leq \exp\left(-2(\epsilon - \frac{1}{a})^2 \ell\right) \leq \exp\left(-2(\epsilon - \frac{1}{a})^2 C \ln m\right) = m^{-2(\epsilon - 1/a)^2 C}.$$
(b) Using the same analysis as above, but the stronger Chernoff bound quoted in the hint, with \( \beta = \frac{\lambda}{\mu} = 1 \), we get

\[
\Pr[X > \epsilon \ell] \leq \exp \left\{ -\frac{\ell}{\alpha}(\alpha \ln(\alpha e) - \alpha e + 1) \right\} \leq \exp \left\{ -\ell(\epsilon \ln(\alpha e) - \epsilon) \right\}.
\]

Now for any fixed \( \epsilon > 0 \), by choosing \( a \) large enough, we can make this probability less than \( \exp(-K\ell) \) for any desired \( K \). So if \( \ell = \delta \ln m \) for fixed \( \delta > 0 \), by arranging for \( K > \frac{2}{\delta} \) we can still make the above \( o(m^{-2}) \), and then apply a union bound over pairs of strings as in part (a). (Specifically, we just need that \( a > e^{-\epsilon} \exp(1 + 2/(\epsilon \delta)) \).)

(c) Any embedding of a pair of strings in \( \mathbb{Z}^3 \) defines a “set of adjacencies” consisting of those pairs of symbols (from different strings, or non-consecutive positions on the same string) that are adjacent in the embedding. We will bound the number of possible sets of adjacencies, and the probability that any given set has a large score, then use a union bound.

To bound the number of possible sets of adjacencies for a pair of strings, note that each one can be realized by embedding both strings in a cube of side length \( 2\ell \). Within such a cube, the number of embeddings of the two strings is (very crudely) at most \( (2\ell)^3(6\ell-1)^2 \), where the first factor counts the starting points and the second the number of walks in \( \mathbb{Z}^3 \) of length \( \ell - 1 \), and we pessimistically ignore the constraint that the two strings be non-overlapping. Thus the number of possible sets of adjacencies is at most \( \exp(\alpha \ell) \) for some universal constant \( \alpha \).

Now fix a particular set of adjacencies. The maximum number of possible pairs in the set is at most \( 5\ell \) (since each of the \( 2\ell \) symbols on the strings can be adjacent to at most 5 others, and each adjacency gets counted twice). For each pair, we get a score of 1 iff the two symbols in it are assigned the same value. Thus, the probability of getting a score of more than \( \epsilon \ell \) is bounded by the probability of more than \( \epsilon \ell \) successes in \( 5\ell \) trials with success probability \( \frac{1}{\alpha} \). However, the events that different pairs score more than \( \epsilon \ell \) are not necessarily independent. (E.g., consider a cycle \((p_1, p_2, p_3, p_4)\) of length 4 in \( \mathbb{Z}^3 \), and suppose the pairs \((p_1, p_2)\), \((p_2, p_3)\), \((p_3, p_4)\) and \((p_4, p_1)\) are adjacencies; if any three of these score, then so must the fourth.) To get around this, we can partition the adjacencies into three disjoint sets according to the direction (in \( \mathbb{Z}^3 \)) of the adjacency. Now it’s easy to see that the scoring events within each set of adjacencies are independent. Moreover, it is sufficient to bound the probability of a score of more than \( \frac{\epsilon \ell}{3} \) within each such set, and multiply by 3. Using the same Chernoff bound as in part (b), but now with \( \mu = \frac{5\ell}{a} \) and \( \beta = \frac{ae}{15} - 1 \), we see that this probability is bounded by

\[
3 \exp \left\{ -\frac{5\ell}{a} \left( \frac{ae}{15} \ln \left( \frac{ae}{15} \right) - \frac{ae}{15} + 1 \right) \right\} \leq 3 \exp \left\{ -\epsilon \ell \left( \frac{\epsilon}{3} \ln \left( \frac{ae}{15} \right) - \frac{\epsilon}{3} \right) \right\}.
\]

Now, as in part (b), we can choose \( a \) large enough so that this probability is less than \( \exp(-K\ell) \) for any desired \( K \). Thus, taking the union bound over sets of adjacencies, the probability of two strings having a score of more than \( \epsilon \ell \) in any embedding is at most \( \exp(\alpha \ell) \exp(-K\ell) = o(m^{-2}) \) if we set \( \ell = \delta \ln m \) and choose \( a \) (and thus \( K \)) large enough. Again, a union bound over pairs of strings finishes the job.

**Common Issues:** To partition the scoring events into several sets, within which each event is independent, many people used Vizing’s theorem, which states that a graph with maximum degree \( d \) can be edge-colored using \( d + 1 \) colors. This is perfectly correct, though using such a powerful theorem was not necessary.

4. More on the power of two choices

Following the hint, we will define a decreasing sequence of values \( \alpha_i \) (to be determined shortly) and events \( \mathcal{E}_i = “after (1 - \frac{1}{2i})n balls have been thrown, the number of bins with load \geq i is at least \alpha_i” \). Our aim is to bound the probability \( \Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i] \).

So assume that \( \mathcal{E}_i \) holds, and consider the placement of balls \((1 - \frac{1}{2i})n + 1 \) through \((1 - \frac{1}{2i+1})n \) (of which there are \( \frac{n}{2i+1} \)). For any such ball, if both of its choices are bins with load at least \( i \), and at least one of the
two has load exactly $i$, then a new bin with load $i + 1$ will be created. Assuming that we still have fewer than $\alpha_{i+1}$ bins of load at least $i + 1$ when the ball is thrown, the probability that the ball makes such choices is at least
\[
\left( \frac{\alpha_i}{n} \right)^{\left( \frac{\alpha_i+1}{n} - \left( \frac{\alpha_i+1}{n} \right) \right)} \geq \frac{1}{2} \left( \frac{\alpha_i}{n} \right)^2,
\]
where we have assumed that $\alpha_i \geq \frac{1}{2} \alpha_{i+1}$. (This is for algebraic convenience only; in fact $\alpha_i$ will decrease quite a bit faster than this; see below.) To justify (*), note that the number of bins with load exactly $i$ remains at least $(\alpha_i - \alpha_{i+1})$ throughout, as long as the number of bins with load at least $i + 1$ remains below $\alpha_{i+1}$; and the the number of bins with load at least $i$ is certainly at least $\alpha_i$.

Noting that if the number of bins with load at least $i + 1$ reaches $\alpha_{i+1}$ then event $\mathcal{E}_{i+1}$ certainly holds, we see from (*) that the desired probability $\Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i]$ is bounded above by the probability that $\text{Bin} \left( \frac{n}{2^{i+1}}, \frac{1}{2} \left( \frac{\alpha_i}{n} \right)^2 \right)$ is less than $\alpha_{i+1}$. The expectation of the above random variable is $\mu_i = \frac{1}{2^{i+1}} \alpha_i^2$. To get a small tail probability, this suggests that we should take (say) $\alpha_{i+1} = \frac{1}{2} \mu_i$, or equivalently, $\alpha_{i+1} = \frac{1}{2^{i+1}} \alpha_i^2$. Of course, $\alpha_0 = n$.

With this choice of $\alpha_i$, a Chernoff bound (Angluin’s version, Corollary 13.3) with $\beta = \frac{1}{2}$ tells us that
\[
\Pr[\neg \mathcal{E}_{i+1} | \mathcal{E}_i] \leq \exp(-\mu_i/8) = O(n^{-1})
\]
provided $\mu_i \geq 8 \ln n$. Let $i^*$ be the largest integer $i$ for which this still holds. By unwinding the recurrence $\frac{\alpha_{i+1}}{n} = \frac{1}{2^{i+1}} \left( \frac{\alpha_i}{n} \right)^2$, we see that $\alpha_i = n 2^{-\Theta(2^i)}$ and hence $i^* = \frac{\ln \ln n}{\ln 2} - O(1)$. Finally we have
\[
\Pr[\mathcal{E}_{i^*}] \geq \Pr[\mathcal{E}_0] \times \prod_{i=0}^{i^*-1} \Pr[\mathcal{E}_{i+1} | \mathcal{E}_i] \geq \left( 1 - \frac{1}{n} \right)^{O(\ln \ln n)} = 1 - o(1).
\]
But since $\alpha_{i^*} \geq 1$, $\mathcal{E}_{i^*}$ implies that the maximum load is at least $i^*$ and we are done.

5. Random Geometric Graphs

(a) Following the hint, partition the unit square into small squares of area $\frac{\log n}{n}$. Let $S$ be any such square. 2pts

Then
\[
\Pr[S \text{ contains no points}] = (1 - \text{area}(S))^n = (1 - \frac{\log n}{n})^n \leq e^{-\log n} = n^{-1}.
\]
Thus taking a union bound over all $\frac{n}{\log n}$ choices of $S$, we see that $\Pr[\text{some S contains no points}] \to 0$ as $n \to \infty$.

Now let $D$ denote the disc of radius $\sqrt{\frac{10 \log n}{n}}$ centered at some point. W.l.o.g. we may assume that $D$ is contained entirely within the unit square; otherwise the argument below only gets better. The expected number of points in $D$ is $\mu := n \times \text{area}(D) = 10 \pi \log n$. Thus using Angluin’s version of the Chernoff bound (upper tail with $\beta = 1$), we get
\[
\Pr[D \text{ contains more than } 2\mu \text{ points}] \leq \exp(-\mu/3) = \exp(-\frac{10 \pi}{3} \log n) = n^{-1(1+\delta)},
\]
for some constant $\delta > 0$. Taking a union bound over all $n$ points ensures that
\[
\Pr[\text{some D contains more than } 2\mu \text{ points}] \to 0 \quad \text{as } n \to \infty.
\]

Now take $c > 20 \pi$, and form the graph $G$ in which every point is connected to its $c \log n$ closest neighbors). From the above arguments we have that, with probability $1 - o(1)$, every square contains a point, and every point is connected to points in all its immediately neighboring squares. But these conditions are clearly enough to ensure that $G$ is connected.

[NOTE: Clearly the constant 10 is pretty arbitrary here, and leads to a fairly large value for $c$. If you are interested, you might like to find the minimum constant that works; this of course will give a better bound on $c$ in the final result.]
Note first that if our point set contains a “bad” system of three concentric discs (as defined in the original hint, with condition (iii) modified as in the supplementary hint posted on 11/8) then the resulting graph $G$ is disconnected. To see this, note that condition (i) implies that the $k$ nearest neighbors of all points in $D_1$ also lie in $D_1$, and condition (ii) says that there are no points in $D_3 \setminus D_1$. We now argue that conditions (ii) and (iii) imply that the $k$ nearest neighbors of any point outside $D_3$ all lie outside $D_3$. To see this, for any point $p$ outside $D_3$, let $r_p$ denote the distance to the boundary of $D_1$ (thus $r_p$ is a lower bound on the distance to the nearest point in $D_1$). Consider the disc of radius $1.5r$ (as defined in (iii)) centered at a point on $D_3$ that is closest to the line connecting $p$ with the boundary of $D_1$. By the triangle inequality, the distance from $p$ to this disc’s center will be at most $r_p - 2r + .01r$, and by condition (iii) and the triangle inequality, there will be at least $k$ points within distance $r_p - 2r + .01r + 1.5r = r_p - 49r < r_p$ from the point $p$, and thus $p$ will not be connected to any point inside $D_1$, as desired.

Proceeding as suggested in the hint, we now consider that the points are distributed according to a Poisson point process (PPP) of intensity $n$. The main gain from this translation is that disjoint regions behave independently. (In the original model, of course, the only dependencies came from the fact that the number of points was fixed to be $n$.)

Now let’s consider an arbitrary system of three concentric discs as specified in the hint. We analyze the probability of each of the events (i), (ii), (iii) under the Poisson model.

(i) The expected number of points in disc $D_1$ is $\mu := n \times \text{area}(D) = k + 1$, so we are looking for the probability that a Poisson r.v. is at least equal to its mean. This is at least $1/4$. (Actually it is close to $1/2$.)

(ii) The expected number of points in $D_3 \setminus D_1$ is $n \times \text{area}(D_3 \setminus D_1) = 8\pi r^2 n = 8(k + 1)$. Thus the probability that no points fall in this area is $\exp(-8(k + 1)) = \exp(-8(c_2 \log n + 1))$. If we choose $c_2 > (1 - \epsilon)/8$ for some small $\epsilon > 0$, this probability is greater than $n^{-1+\epsilon}$.

(iii) The expected number of points lying in each circle of radius $1.5r$, excluding the portion within $D_3$, is at least $\frac{\pi(1.5r)^2}{2} = 1.125(k + 1)$, and thus the probability that a given circle contains fewer than $k + 1$ points, by the Chernoff bound for Poissons in Q1 above, is $\exp(-c_3 k)$, for some constant $c_3$. Applying the union bound to the $\frac{\pi 6r}{3\sqrt{\pi}} \leq 2000$ such circles on the boundary of $D_3$ yields that this condition is satisfied with probability at least $1 - 2000 \exp(-c_3 k) > .99$, for sufficiently large $k$.

Putting together the above three events, and noting that they are independent since they refer to disjoint areas, we get that

$$\Pr[\text{set of discs is bad}] \geq (1/4)(.99)n^{-1+\epsilon} \geq cn^{-1+\epsilon}.$$

Finally, note that we can pack a total of $c' \frac{n}{\log n}$ disjoint systems of three discs into our unit square, for some absolute constant $c'$. Each of these is bad independently with probability at least $cn^{-1+\epsilon}$, so the probability that our graph is connected is at most

$$(1 - cn^{-1+\epsilon})^{c'n} / \log n \leq \exp(-c''n^{\epsilon} / \log n).$$

To complete the proof, we note that in the above argument we assumed that the points are chosen according to a Poisson Point Process of intensity $n$. However, in this process the probability that the total number of points chosen is exactly $n$ is $\Pr[\text{Po}(n) = n] \geq \frac{1}{2\sqrt{n}}$, where $\text{Po}(n)$ denotes a random variable distributed according to the Poisson distribution with expectation $n$. Thus even if the entire failure probability of $\exp(-c''n^{\epsilon} / \log n)$ occurs when the number of points chosen by the Poisson Point Process is exactly $n$, we still have that, when the number of points is exactly $n$, the failure probability is at most $\exp(-c''n^{\epsilon} / \log n)(2\sqrt{n}) \to 0$, as desired.
6. Concentration of the Longest Common Subsequence

(a) We let \( Z_i = (a_i, b_i) \), and let \( L = L(Z_1, \ldots, Z_n) \) denote the length of a lcs of \( a, b \). Then \( X_i = \frac{2}{2} \text{E}(L|Z_1, \ldots, Z_i) \) is a martingale (the Doob martingale of \( L \) w.r.t. \( (Z_i) \)). It is easy to check that \( L \) is 2-Lipschitz (if we remove \( a_i, b_i \) and their partners from any common subsequence of \( a, b \), we get a subsequence at most two shorter; and by reversing the argument we get a similar bound on the increase caused by changing \( a_i, b_i \)). Since the \( Z_i \) are also independent, we can apply Azuma’s inequality with bounded differences of 2 to deduce that

\[
\Pr[|X_n - \mu_n| \geq \lambda] \leq 2 \exp(-\lambda^2/8n).
\]

NOTE: We can actually do slightly better by considering instead the filter \( Z_{2i-1} = a_i, Z_{2i} = b_i \), which makes \( L \) 1-Lipschitz with a difference sequence of length \( 2n \) and hence replaces the above bound by

\[
2 \exp(-\lambda^2/4n).
\]

Common Issues: many people argued that \( L \) is 2-Lipschitz, and thus the Doob martingale has bounded difference. The fact that the the different elements of the sequences are independent is crucial to being able to make this statement, as is illustrated in part ii below.

(b) (i) No difference; the argument above is oblivious to the alphabet size.

(ii) In the absence of independence, we can’t claim any non-trivial concentration. (For example, suppose we have the following values for \( a, b \), each with probability \( \frac{1}{2} \): \( (a = b = 0^n), (a = b = 1^n), (a = 0^n, b = 1^n) \) and \( (a = 1^n, b = 0^n) \). Then \( \text{E}(L) = n/2 \), but \( |L - \text{E}(L)| = n/2 \) with probability 1.)

Common Issues: As some of you pointed out, the arguments of part (a) still work if \( a_i \) and \( b_i \) are dependent, provided that \( a_i, a_j \) are independent, and \( b_i, b_j \) are independent, for \( i \neq j \).

(iii) Here the argument above still holds, but the function \( L \) becomes 3-Lipschitz. Thus we get the slightly weaker bound

\[
\Pr[|X_n - \mu_n| \geq \lambda] \leq 2 \exp(-\lambda^2/18n).
\]

NOTE: The alternative argument also extends, making \( L \) 1-Lipschitz over a sequence of length \( 3n \) and giving the better bound \( 2 \exp(-\lambda^2/6n) \).