CS174 Midterm 2 Solutions

1. We throw m balls in 6 bins, uniformly and independently. Use a Chernoff bound on the tail of the binomial distribution to give an upper bound on the probability that the first bin contains more balls than the total number in the other five.

   If we let Z be a binomial(m,1/6) random variable, the required probability is

$$\begin{aligned}\Pr(Z > m - Z) &= \Pr\left(Z > \frac{m}{2}\right) \\ &= \Pr\left(Z > \frac{m}{6}\,(1+2)\right) \\ &\le \left(\frac{e^2}{3^3}\right)^{m/6}.\end{aligned}$$

2. A mailing list server accepts email requests for changes. These request messages accumulate in its mailbox, and at the end of each hour, it serves all of these requests, and empties its mailbox. Suppose that the list has n members, and for each period t, each member submits one of these requests with a certain probability p_t, independently of other members and of other requests. (Assume that the list server does not receive any other email.) The probability p_t depends on the time of day, but there is a p_0 such that for all t the probability satisfies $p_t \le p_0$. The mailbox is subject to a size limit of $N = np_0 + k$ emails, which we do not wish to exceed. Suppose that, for some $x > 0$, we want that, for each period, the probability that the mailbox reaches this size limit during that period is no more than e^{-x}. Show that it suffices to choose $k = \sqrt{3xnp_0}$.

For a particular period, the number of messages that arrive in that period has a binomial(n,p) distribution, for some $p \le p_0$. From Chernoff bounds,

$$\Pr(X \ge np_0 + k) \le \exp\left(\frac{-np(n(p_0 - p) + k)^2}{3(np)^2}\right) \le \exp\left(\frac{-(n(p_0 - p) + k)^2}{3np}\right) \le \exp\left(\frac{-k^2}{3np_0}\right).$$

and this is no more than e^{-x} for $k^2/(3np_0) \ge x$, which is equivalent to

$$k \ge \sqrt{3xnp_0}.$$

3. We throw m balls in n bins, uniformly and independently. Let X denote the number of bins that remain empty.

   (a) What is $\mathbb{E}X$?

      Let $X_1, \ldots, X_n$ denote the indicator variables for bin $1, \ldots, n$ remaining empty, respectively. Then we have

$$\begin{aligned} \mathbb{E}X &= \mathbb{E}\sum_{i=1}^{n} X_i \\ &= \sum_{i=1}^{n} \mathbb{E}X_i \\ &= n\left(1 - \frac{1}{n}\right)^m . \end{aligned}$$

(b) Using the Poisson approximation, or otherwise, give an upper bound on the probability that either $X \geq 2ne^{-m/n}$ or $X \leq ne^{-m/n}/2$.

Let Y denote the number of bins that remain empty under the Poisson distribution, and let $Y_1, \dots, Y_n$ denote the indicator variables for bin $1, \dots, n$ remaining empty, respectively. Then we can write $Y = \sum_{i=1}^{n} Y_i$. But each Y_i is a Bernoulli random variable with parameter p given by the probability that a discrete Poisson with parameter m/n has value 0. Thus,

$$p = e^{-m/n}.$$

Notice that $\mathbb{E}Y = np = ne^{-m/n}$. Since these Y_i are independent, we can apply Chernoff bounds:

$$\begin{aligned}
\Pr(Y \geq 2ne^{-m/n}) &= \Pr(Y \geq 2\mathbb{E}Y) \\
&\leq \exp\left(-\mathbb{E}Y/3\right) \\
&= \exp\left(-ne^{-m/n}/3\right).
\end{aligned}$$

Similarly,

$$\begin{aligned}
\Pr(Y \leq ne^{-m/n}/2) &= \Pr(Y \leq \mathbb{E}Y/2) \\
&\leq \exp\left(-\mathbb{E}Y/8\right) \\
&= \exp\left(-ne^{-m/n}/8\right).
\end{aligned}$$

Thus,

$$\Pr(Y \geq 2ne^{-m/n} \text{ or } Y \leq ne^{-m/n}/2) \leq \exp\left(-ne^{-m/n}/3\right) + \exp\left(-ne^{-m/n}/8\right).$$

Now, we use the Poisson approximation to relate this probability to the desired one.

$$\begin{aligned}
\Pr(X \geq 2ne^{-m/n} \text{ or } X \leq ne^{-m/n}/2) &\leq e\sqrt{m}\Pr(Y \geq 2ne^{-m/n} \text{ or } Y \leq ne^{-m/n}/2) \\
&\leq e\sqrt{m}\left(\exp\left(-ne^{-m/n}/3\right) + \exp\left(-ne^{-m/n}/8\right)\right).
\end{aligned}$$

4. In the travelling salesman problem, we must find a minimal length path that includes each of n cities, located at $x_1, \ldots, x_n \in [0,1]^2$. Define

$$J(x_1, \ldots, x_n) = \frac{1}{n} \min \left\{ \sum_{j=1}^{n-1} \|x_{i_{j+1}} - x_{i_j}\| : \{i_1, \ldots, i_n\} = \{1, \ldots, n\} \right\},$$

where we denote the Euclidean norm $\|a\| = \sqrt{a_1^2 + a_2^2}$. Consider a random travelling salesman problem: the n city locations $X_1, \ldots, X_n$ are random vectors chosen independently from $\{0, 1/n, 2/n, \ldots, 1\}^2$. Show that, with probability at least $1 - 1/n$,

$$|J(X_1, \ldots, X_n) - \mathbb{E}J(X_1, \ldots, X_n)| \leq 2\sqrt{\frac{\ln(2n)}{n}}.$$

(Hint: You might wish to use the following fact, and then apply the bounded differences inequality. Since the distance between any two points in this set is no more than $\sqrt{2}$, it is easy to see that, for any configuration of the cities $x_1, \ldots, x_n$, moving x_i to some other location y_i can change J by at most $2\sqrt{2}/n$.)

From the bounded differences inequality,

$$\begin{aligned} \Pr\{|J - \mathbb{E}J| \geq \epsilon\} &\leq 2\exp\left(\frac{-2\epsilon^2}{n(2\sqrt{2}/n)^2}\right) \\ &= 2\exp\left(\frac{-\epsilon^2 n}{4}\right). \end{aligned}$$

This is no more than $1/n$ for

$$\frac{\epsilon^2 n}{4} \geq \ln(2n) \iff \epsilon \geq \sqrt{\frac{4\ln(2n)}{n}}.$$