MATH 340: Discrete Structures II. Winter 2017.

Assignment #4: Discrete Probability II.

1. *The Birthday problem.* Suppose that the birthdays of n people in the room are uniformly distributed among the 365 days of the year. Estimate how large should n be to guarantee that the probability of some two people sharing a birthday is at least 999/1000.

Solution: We estimate the probability of the event that all n people have different birthday:

$$\begin{aligned} p(B) &= \prod_{k=1}^{n-1}\left(1-\frac{k}{365}\right) \le \prod_{k=1}^{n-1} e^{-k/365} \\ &= e^{-\sum_{k=1}^{n-1} k/365} = e^{-n(n-1)/730} \le 1/1000. \end{aligned}$$

Thus we need n such that $n(n-1) \ge 730 \ln 1000$. $\boxed{n = 72}$ suffices.

2. *Quicksort.* Let $x_1, x_2, \dots, x_n$ be a permutation of numbers $1, \dots, n$ chosen uniformly at random.

a) Show that the probability that the numbers i and j, such that $1 \le i \le j \le n$, are compared to each other by the Quicksort algorithm is equal to

$$\frac{2}{j-i+1}.$$

b) Deduce that the expected number of comparisons made by the Quicksort algorithm is equal to

$$2\sum_{k=1}^{n-1}\frac{n-k}{k+1}.$$

Solution:
a): The numbers i and j, such that
$1 \le i \le j \le n$, are compared to each other by the Quicksort algorithm if and only if either i or j is the first number among the numbers $i, i+1, i+$

$2, \ldots, j$ that appears in the sequence $x_1, x_2, \ldots, x_n$. (Otherwise, one of the numbers strictly between i and j will be processed first, and they will not be compared.) Clearly the probability of this event is $\frac{2}{j-i+1}$.
b): By linearity of expectation the total number of comparisons will be

$$\sum_{1 \le i < j \le n} \frac{2}{j-i+1} = \sum_{k=1}^{n-1} \sum_{\substack{1 \le i < j \le n \\ j=i+k}} \frac{2}{j-i+1}$$

$$= \sum_{k=1}^{n-1} \left((n-k) \frac{2}{k+1} \right) = 2 \sum_{k=1}^{n-1} \frac{n-k}{k+1}.$$

3. *Balls and bins.* Suppose that we randomly drop $n^{3/2}$ balls into n bins. Give an upper bound on the expectation of the maximum number of balls in any bin.

Solution: Let X_i be the random variable equal to the number of balls in the ith bins. Then $E[X_i] = \sqrt{n}$. Let $\delta = 3\frac{\sqrt{\ln n}}{n^{1/4}}$. By the Chernoff bound,

$$p(X_i > (1+\delta)\sqrt{n}) \le e^{-\sqrt{n}\delta^2/3} = e^{-9/3 \ln n} = \frac{1}{n^3}$$

for $\delta \le 1$. Let M be the random variable equal to the maximum number of balls in a bin. Then

$$p(M > (1+\delta)\sqrt{n} \le \sum_{i=1}^{n} p(X_i > (1+\delta)\sqrt{n}) \le \frac{1}{n^2}.$$

Finally,

$$E[M] \le (1+\delta)\sqrt{n} p(M \le (1+\delta)n) + n^{3/2} p(M \ge (1+\delta)\sqrt{n})$$

$$\le (1+\delta)\sqrt{n} + 1 = \boxed{\sqrt{n} + 1 + 3n^{1/4}\sqrt{\ln n}}$$

4. *Balls and bins II.* Given n balls of each of n different colors (n^2 balls in total), we distribute them among n boxes, as follows. For each ball we choose a box at random. If the chosen box already contains the ball of the same color as the ball we are considering, we throw the current ball away. Otherwise, we put it in the box.

a) Show that the probability that a box contains a ball of given color is

$$1 - \left(1 - \frac{1}{n}\right)^n.$$

b) Find the expected number of balls that we throw away.

c) Show that with high probability no box contains more than

$$n\left(1 - \left(1 - \frac{1}{n}\right)^n\right) + 2\sqrt{n \ln n}$$

balls.

Solution:
a): If a box does not contain a ball of some color, then for all the balls of this color we chose one of the other boxes to place it in. As these choices are independent this happens with probability $\left(1 - \frac{1}{n}\right)^n$. Thus the probability that a box contains a ball of given color is

$$1 - \left(1 - \frac{1}{n}\right)^n.$$

b): Let X_{ij} be the number of balls in box i of color j. (Let us assume that boxes and colors are numbered $1, \ldots, n$.) By a) $E(X_{ij}) = 1 - \left(1 - \frac{1}{n}\right)^n$. Thus the expected total number of balls we keep is

$$E\left(\sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}\right) = n^2\left(1 - \left(1 - \frac{1}{n}\right)^n\right).$$

All the other balls we throw away, giving the answer

$$n^2 - n^2\left(1 - \left(1 - \frac{1}{n}\right)^n\right) = n^2\left(1 - \frac{1}{n}\right)^n.$$

c): Let $Z_i = \sum_{j=1}^{n} X_{ij}$ be the number of balls in ith box. Let $c = 1 - \left(1 - \frac{1}{n}\right)^n$ for brevity. From b) we have $\mu = E(Z_i) = cn$ As X_{ij} are independent Bernoulli random variables by Chernoff bound we have

$$p(Z_i > nc + 2\sqrt{n \ln n}) = p\left(Z_i > \mu\left(1 + 2\sqrt{\frac{\ln n}{nc^2}}\right)\right)$$
$$\leq e^{-4\mu \ln n/(3c^2 n)} \leq e^{-4 \ln n/3} = \frac{1}{n^{4/3}}.$$

Thus by the union bound probability that some box contains more than $nc + 2\sqrt{n \ln n}$ is at most $\frac{n}{n^{4/3}} = \frac{1}{n^{1/3}} \to_{n\to\infty} 0$, as desired.

5. *Deviation below the mean.* Prove the following variant of the Chernoff bound for the deviation below the mean. Let X be the sum of independent Bernoulli random variables, and let $\mu = \mathrm{E}[X]$. Show that, if $0 < \delta < 1$, then

$$p(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \le e^{-\mu\delta^2/2}$$

Solution: For any $t \ge 0$ we have

$$\begin{aligned} p(X \le (1-\delta)\mu) &= p(e^{tX} \le e^{(1-\delta)t\mu}) \\ &= p(e^{-tX} \ge e^{-(1-\delta)t\mu}) \le \frac{E[e^{-tX}]}{e^{-(1-\delta)t\mu}}, \end{aligned}$$

where the last inequality is by Markov's inequality. Let X be the sum of the independent Bernoulli random variables $X_1, X_2, \ldots, X_n$ such that $E[X_i] = p_i$. We have

$$\begin{aligned} E[e^{-tX}] = E[e^{-t\sum_{i=1}^n X_i}] &= \prod_{i=1}^n E[e^{-tX_i}] \\ &= \prod_{i=1}^n ((1-p_i)\cdot e^0 + p_i e^{-t}) = \prod_{i=1}^n (1 - p_i(1-e^{-t})) \\ &\le \prod_{i=1}^n e^{-p_i(1-e^{-t})} = e^{-(1-e^{-t})\sum_{i=1}^n p_i} = e^{-\mu(1-e^{-t})}. \end{aligned}$$

Thus

$$p(X \le (1-\delta)\mu) \le \frac{e^{-\mu(1-e^{-t})}}{e^{-(1-\delta)t\mu}}.$$

Setting $t = -\ln(1-\delta)$, we get

$$p(X \le (1-\delta)\mu) \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \le e^{-\mu\delta^2/2}.$$

To verify the last inequality we need to show that

$$(1-\delta)^{(1-\delta)} \ge e^{-\delta+\delta^2/2},$$

or, equivalently,

$$(1-\delta)\ln(1-\delta)+\delta-\delta^2/2 \geq 0.$$

For $\delta = 0$ the above inequality holds with equality, and the derivative of the left side is $-\ln(1-\delta)-\delta$, which is non-negative for $\delta < 1$. Thus the inequality holds for all $0 < \delta < 1$.

6. *Random graphs.* In a *random graph* on n vertices for each pair of vertices i and j we independently include the edge $\{i, j\}$ in the graph with probability $1/2$. Show that with high probability every two vertices have at least $n/4 - \sqrt{n \log n}$ common neighbors.

Solution: Given a pair of vertices i and j, let X_{ij} be the random variable equal to the number of their common neighbors. As each of the remaining $n-2$ vertices is a common neighbor of i and j with probability $1/4$, we have $E[X_{ij}] = (n-2)/4$. Let $\delta = 4\sqrt{n \ln n}/(n-2)$ and

$$p\left(X_{ij} < (1+\delta)\frac{n-2}{4}\right) \leq e^{-(n-2)\delta^2/8} = e^{2n\ln n/(n-2)} \rightarrow_{n\to\infty} 0,$$

as desired.