Math 262A/ CSE 291A
Midterm Solutions

1. What is the highest IQ in the world? More specifically, let $X_i, 1 \le i \le 5\times 10^9$ be independent, each normal with mean 100 and the standard deviation (square root of the variance) is 15. Let $X = \max_{1\le i\le 5\times 10^9} X_i$. Let $Y^{(\alpha)}$ denote the number of i with $X_i > \alpha$.

   (i) Find α with $E[Y^{(\alpha)}] \approx 1$
   (ii) Find β with $Pr[X < \beta] \approx .05$
   (iii) Find γ with $Pr[X < \gamma] \approx .95$

   Solution #1:

   We will use the following identity:

$$\int_a^\infty \frac{1}{15\sqrt{2\pi}} e^{-\frac{(t-100)^2}{2\times 15^2}} dt = \int_b^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

   where $b = \frac{a-100}{15}$.

   (1) Find α with $E[Y^{(\alpha)}] \approx 1$.

   Let $\lambda = \frac{\alpha-100}{15}$ and

$$Y_i = \begin{cases} 1 & \text{if } X_i > \alpha, \\ 0 & \text{otherwise.} \end{cases}$$

   Then,

$$\begin{aligned} E(Y^{(\alpha)}) &= E(\sum_{i=1}^{5\times 10^9} Y_i) \\ &= \sum_{i=1}^{5\times 10^9} E(Y_i) \\ &= 5\times 10^9 Pr(X_1 > \alpha) \\ &= 5\times 10^9 \int_\lambda^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \\ &\approx \frac{5\times 10^9}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} dt \\ &\approx 1 \end{aligned}$$

So,

$$\begin{aligned}\frac{\lambda^2}{2} + \log \lambda &= \log \frac{5 \times 10^9}{\sqrt{2\pi}} \\ &= 21.41\end{aligned}$$

Approximate λ using some reasonable technique (e.g., MAPLE) and find:

$$\begin{aligned}\lambda &\approx 6.26 \\ \rightarrow \alpha &= 100 + 15\lambda \\ &\approx 194\end{aligned}$$

(ii) Find β with $Pr[X < \beta] \approx .05$.
Let $\lambda = \frac{\beta - 100}{15}$. Then,

$$\begin{aligned}Pr(X > \beta) &= \left(\int_{-\infty}^{\lambda} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} dt \right)^{5 \times 10^9} \\ &= \left(1 - \int_{\lambda}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} dt \right)^{5 \times 10^9} \\ &\approx \left(1 - \frac{1}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} \right)^{5 \times 10^9} \\ &\approx e^{-\frac{5 \times 10^9}{\lambda\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}}} \\ &= .05\end{aligned}$$

This implies

$$\begin{aligned}\Rightarrow \lambda e^{\lambda^2/2} &\approx \frac{5 \times 10^9}{\sqrt{2\pi}(-\log(.05))} \\ &= 6.66 \times 10^8 \\ \Rightarrow \frac{\lambda^2}{2} + \log \lambda &\approx 20.3 \\ \Rightarrow \lambda &\approx 6.1 \\ \Rightarrow \beta &= 15\lambda + 100 \\ &= 191\end{aligned}$$

(iii) Find γ with $Pr[X < \gamma] \approx .95$.
using the same method as in (ii), we substitute $\lambda = \frac{\gamma - 100}{15}$. The resulting eqeuation becomes:

$$
\begin{aligned}
\lambda e^{\lambda^2/2} &\approx \frac{5\times 10^9}{\sqrt{2\pi}(-\log(.95))} \\
&= 3.89\times 10^8 \\
\Rightarrow \frac{\lambda^2}{2} + \log\lambda &\approx 24.4 \\
\Rightarrow \lambda &\approx 6.7 \\
\Rightarrow \beta &= 15\lambda + 100 \\
&= 201
\end{aligned}
$$

Solution #2.

1(i). Find α with $E[Y^{(\alpha)}] \approx 1$.

$$
\begin{aligned}
E\left[Y^{(\alpha)}\right] &= E\left[\#\{i : X_i > \alpha\}\right] \\
&= \sum_i Pr\left[X_i > \alpha\right] = 5\times 10^9 \cdot Pr\left[X_1 > \alpha\right].
\end{aligned}
$$

Setting this quantity to 1 we have:

$$Pr\left[X_1 > \alpha\right] \approx 2\times 10^{-10}.$$

Using a binary search algorithm to obtain convergence, we test values of α in the normal distribution for X_1 and obtain

$$\alpha \approx 193.810,$$

with a relative error of $\leq 10^{-6}\%$.

1(ii). Fine β with $Pr\left[X < \beta\right] \approx .05$.

We may rewrite the event $X < .05$ as

$$X_1 < .05 \quad \text{and} \quad X_2 < .05 \quad \text{and} \quad \cdots \quad \text{and} \quad X_N < .05$$

where $N = 5\times 10^9$. Thus

$$
\begin{aligned}
Pr[X < \beta] &= \prod_{i=1}^{N} Pr[X_i < \beta] \\
&= Pr[X_i < \beta]^{5\times 10^9}
\end{aligned}
$$

Setting the resulting expression to .05, we have

$$Pr[X_i < \beta] = .05^{1/5\times 10^9} = .05^{2\times 10^{-10}}.$$

Thus β is the *quantile* corresponding to the value on the right. That is, an IQ of less than β occurs with the probability $.05^{2\times 10^{-10}}$. Using mathematica, this quantile is

$$\beta \approx 191.207.$$

1(iii). Find γ with $Pr\left[X < \gamma\right] \approx .95$.

Similarly to 1(ii), we determine

$$\gamma \approx 200.534.$$

These three facts 1(i)-(iii) blend together. (ii) and (iii) tell us that the smartest person will have between 191.207 and 200.534 IQ with about 90% probability. Thus at some point α in that range, there will only be one person smarter than α. This is reflected by the value of $\alpha = 193.810$ in (i).

2. A set D of vertices in an Internet graph G is called dominating if every $v \in G$ is either in D or is adjacent to a vertex of D (or both). State (i.e. find the α in the statement) and prove a result of the following form.
   Theorem. If G has n vertices and all vertices have degree at least d then there exists a dominating set D with $|D| \leq \alpha$.

   To do this, let C be a random set of vertices with $Pr[v \in C] = p$. Let N be the set of vertices neither in C nor with any neighbors in C. Then $C \cup N$ is dominating. Adjust p so that $E[|C \cup N|]$ is small. Use the inequality $1-p \leq e^{-p}$ to simplify the analysis and give a cleaner statement.

   Solution:

   Fix the graph G. Now, construct a vertex subset C as follows. Let C_v be the event that we choose v to be in C. Let $Pr[C_v] = p$, where the C_v's are identically distributed, independent random variables. Then we have

$$E[|C|] = \sum_v Pr[C_v] = n \cdot p.$$

   For every vertex v in $V(G)$, either v is in C, v is adjacent to a vertex in C, or v is neither in C or adjacent to a vertex in C. Define

$$N = \{v \in V(G) : v \notin C \text{ and v not adjacent to a vertex in } C\}.$$

Let Y_v be the random variable of the event that $v \in N$. Then $Pr[Y_v] \leq (1-p) \cdot (1-p)^d$, since v must both not be in C and none of its at least d neighbors may be in C either. We have

$$E[|N|] = \sum_v Pr[Y_v] \leq n \cdot (1-p)^{d+1} \leq n \cdot e^{-p(d+1)}.$$

Since C and N are disjoint, by linearity of expectation we have

$$E[|C \cup N|] = E[|C| + |N|] = E[|C|] + E[|N|] \leq n \cdot p + n \cdot e^{-p(d+1)}.$$

There is a choice of C which meets or falls below this expected size of $|C \cup N|$. Minimizing this expression for p, we have

$$\begin{aligned} n - (d+1)n \cdot e^{-p(d+1)} &= 0 \\ e^{-p(d+1)} &= \frac{1}{d+1} \\ -p(d+1) &= \ln \frac{1}{d+1} = -\ln(d+1) \\ p &= \frac{\ln(d+1)}{d+1}. \end{aligned}$$

And so with this p we have

$$\begin{aligned} E[|C \cup N|] &= n \cdot \frac{\ln(d+1)}{d+1} + n \cdot e^{-\ln(d+1)} \\ &= n \cdot \frac{\ln(d+1)}{d+1} + \frac{n}{d+1} = \frac{n}{d+1}\left[1 + \ln(d+1)\right], \end{aligned}$$

an upper bound on the smallest possible size of $C \cup N$.

3. Call a $0-1$ matrix A *ninefree* if there is no 3×3 submatrix with all entries one. (Rows and columns need not be consecutive.) Let $f(n)$ denote the maximal number of ones in an $n \times n$ ninefree A. Find a lower bound for $f(n)$ – i.e., show, for α as large as possible, that there exists a ninefree $n \times n$ A with at least α ones. Use the Deletion Method, first letting $Pr[a_{ij} = 1] = p$ and then changing a one to zero in every 3×3 submatrix with all entries one.

   Proof:

   Consider an $n \times n$ matrix A as a random variable from the space of $0-1$ matrices such that each matrix has a 1 in a particular entry with probability p; each entry is an identically distributed, independent random variable. If $f(n)$

is the maximal number of ones in a $n \times n$ ninefree matrix A, then $f(n) > \alpha$, where

$$\alpha = \sqrt[8]{4} \cdot \frac{8}{9} \cdot n^{3/2}.$$

Let $X_{i,j}$ be the random variable for the i,j-entry of A. Define

$$X = \sum_{i,j} X_{i,j}.$$

Then the expected number of 1's in the matrix A is $E[X]$. By linearity of expectation,

$$E[X] = E\left[\sum_{i,j} X_{i,j}\right] = \sum_{i,j} E[X_{i,j}] = p \cdot n^2.$$

Now, let S range over all of the 3×3 submatrices of A. Since rows and columns do not have to be consecutive, there are $\binom{n}{3} \cdot \binom{n}{3}$ such submatrices. Given S, let $Y_S = 1$ if the submatrix S is all 1's, and let $Y_S = 0$ otherwise. Thus Y_S is the indicator function of the event that S is not a *ninefree* matrix. Define

$$Y = \sum_S Y_S.$$

Then the expected number of non-ninefree submatrices S is, by linearity of expectation,

$$E[Y] = E\left[\sum_S Y_S\right] = \sum_S E[Y_S] = \binom{n}{3}\binom{n}{3} \cdot p^9 < \frac{n^6}{36} \cdot p^9.$$

Now, consider the expected value of the number of 1's in a matrix minus the number of non-ninefree submatrices S. This is

$$E[X - Y] = E[X] - E[Y] > p \cdot n^2 - \frac{n^6}{36} \cdot p^9.$$

To make this expectation as large as possible, differentiate the expression in p, equate to 0, and solve. This gives

$$\begin{aligned} n^2 - \frac{9 \cdot n^6}{36} \cdot p^8 &= 0 \\ n^2 &= \frac{n^6}{4} \cdot p^8 \\ p^8 &= 4 \cdot n^{-4} \\ p &= \sqrt[8]{4} \cdot n^{-1/2} \end{aligned}$$

Since there exists a matrix A that meets or exceeds the expectation $E[X-Y]$, we simply choose a 1 to change to 0 in every non-ninefree submatrix of A, and we are left with

$$E[X] - E[Y] = \sqrt[8]{4} \cdot n^{3/2} - \frac{\sqrt[8]{4}}{9} \cdot n^{3/2} = \sqrt[8]{4} \cdot \frac{8}{9} \cdot n^{3/2}$$

1's in A. Therefore if $\alpha = \sqrt[8]{4} \cdot \frac{8}{9} \cdot n^{3/2}$, $f(n) > \alpha$.

4. Let $A_1, \ldots, A_n \subset \{1, 2, \ldots, n\}$. We showed that (as a Army-Navy game) there exists $\chi : \{1, \ldots, n\} \rightarrow \{-1, 1\}$ so that all $|\chi(A_i)| \leq \sqrt{2n \log(2n)}$. Assume, for convenience, n even. Improve this result by using the following "pairing" distribution. Let $\chi(2i-1)$ be uniform and independent, $1 \leq i \leq n/2$ and set $\chi(2i) = -\chi(2i-1)$.

   Proof:

   Choose χ randomly according to the prescription. Consider the distribution of χ on some A_j. If both of the members of the corresponding pair $\{2i-1, 2i\}$ are present, then the contribution to $\chi(A_j)$ is 0 with probability 1. We are interested in finding out the maximum number of non-cancelling elements of A_j. By the pigeon-hole principle, as soon as A_i is larger than $n/2$, the excess elements all cancel. Therefore $\chi(A_j)$ has distribution S_m, where $m \leq n/2$. Therefore using the fact

   $$Pr[S_n > \lambda] < e^{-\lambda^2/2n},$$

   we have

   $$Pr[|\chi(A_j)| > \lambda] < 2e^{-\lambda^2/n}.$$

   Looking at all $A_j \in \mathcal{A}$, we have

   $$Pr[\exists j \text{ with } |\chi(A_j)| > \lambda] < 2n \cdot e^{-\lambda^2/n}.$$

   This value is less than 1 provided that

   $$\begin{aligned} 2n \cdot e^{-\lambda^2/n} &< 1 \\ e^{-\lambda^2/n} &< \frac{1}{2n} \\ -\lambda^2/n &< -\log(2n), \quad \text{yielding} \\ \lambda &= \sqrt{n \cdot \log(2n)}. \end{aligned}$$

   For this λ, the probability that $\chi(\mathcal{A})$ is greater than λ is less than 1, and so there exists a χ such that $\chi(\mathcal{A}) \leq \sqrt{n \cdot \log(2n)}$. Therefore $disc(\mathcal{A}) \leq \sqrt{n \cdot \log(2n)}$.