Chris Piech
CS109

CS109 Midterm Solutions

1. (a) The number of ways to order 10 distinct objects is 10!. Common mistake: treating songs by the same artist as indistinguishable.

   (b) Group the two songs by artist A. There are 2 ways to order this group. Then if this group is treated as a single object, we have to order 9 objects, which can be done in 9! ways. This gives a total of $2 \cdot 9!$ ways to order the songs such that the two songs from artist A are played in a row. Divide by the sample space (from part a) to get

$$\frac{2 \cdot 9!}{10!} = 1/5$$

   Common mistakes: forgetting to count the 2 ways of ordering the two songs by artist A. Only counting the number of ways to place the songs by artist A and ignoring the ordering of the rest of the songs. Only computing the probability of the two songs playing first (instead of in a row anywhere among the 10 songs).

   (c) Let A be the event that the two songs by artist A are played in a row, and let B be the event that the two songs by artist B are played in a row. Using the inclusion-exclusion principle, we have

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

   $P(A) = P(B) = 1/5$ from part (b). To compute $P(A \cap B)$, we follow a similar approach to part (b): First, group the two songs by A in one group and the two songs by B in another group. For each of these two groups, there are 2 ways to order the two songs. Then this leaves us with 8 distinct things to order (the 6 other songs, the group of the two A songs, and the group of the two B songs), which can be done 8! ways. So $P(A \cap B) = 2 \cdot 2 \cdot 8!/10!$, and

$$P(A \cup B) = 1/5 + 1/5 - 2 \cdot 2 \cdot 8!/10! = 2/5 - 2/45 = 16/45$$

   Common mistakes: just multiplying the answer to part (b) by 2. This only works if A and B are mutually exclusive, which they are not. Claiming $P(A \cap B) = P(A) \cdot P(B)$, which is not true here because A and B are not independent. Many people were slightly off in their computations of $P(A \cap B)$, miscounting the final number of things to order, or forgetting to order the 2-song groups, etc.

   (d) There are 5! ways to arrange the 5 artists in the first 5 songs. This determines the order of artists in the last five songs, so there is only one way to choose that order. For each artist, we must order the artist's two songs to decide which song to put in the first half ($2! = 2$ choices). Therefore, the total number of arrangements is $5! \cdot (2!)^5$.

   Common mistake: leaving out the ordering of the songs, or only multiplying by 2! once.

2. (a) 10^8. There are 10 options for each of the 8 digits in the passcode.

   (b) $\frac{10^6}{10^8} = 0.01$. They can try 10^6 passwords in one second, and each passcode is equally likely. Note that we are trying passwords sequentially, not randomly, so using the geometric distribution was not appropriate here.

   (c)

$$\sum_{i=1}^{\infty}\left(P(X=i)\left(\sum_{j=0}^{i-1} 2^j\right)\right) = \sum_{i=1}^{\infty}\left(1 - \frac{1}{1000}\right)^{i-1}\left(\frac{1}{1000}\right)(2^i - 1)$$

3. (a)

$$\binom{2050}{2} = \sum_{n=0}^{2049} n = \sum_{n=1}^{2050} 2050 - n$$

   Common mistakes: $2050 \cdot 2049$ (order shouldn't matter), off by 1 on summation bounds, product instead of summation.

(b)

$$\binom{50}{2} = \sum_{n=0}^{49} n = \sum_{n=1}^{50} 50 - n$$

Common mistakes: $50 \cdot 49$ (order shouldn't matter), off by 1 on summation bounds, product instead of summation.

(c)

$$\binom{2050}{2}/5 - \binom{50}{2} + 1 = \binom{2050}{2}/5 - \binom{50}{2}$$

We gave credit for correct setup and technique, even if you used incorrect values found in (a) and (b).

(d)

$$\sum_{i=d}^{n} \binom{n}{i} \left(\frac{1}{5}\right)^{i} \left(\frac{4}{5}\right)^{n-i} = 1 - \sum_{i=0}^{d} \binom{n}{i} \left(\frac{1}{5}\right)^{i} \left(\frac{4}{5}\right)^{n-i}$$

where $n = \binom{2050}{2} - \binom{50}{2}$. We accepted approximations with a normal distribution as long as it was mentioned that it was an approximation, and a justification was given for why a normal approximation is appropriate. We also accepted bounds beginning or ending at $d + 1$ instead of d.

Common mistakes: using Poisson or negative binomial; incorrect bounds; incorrect binomial form; finding the probability that a song receives exactly $1/5$ of votes instead of $> 1/5$.

(e) Let S_2 be the event that Shazam predicts X_2, the Andy Williams song. $P(S_2 \mid X_1) = P(\text{Shazam is incorrect}) = 1 - q$, and $P(S_2|X_2) = P(\text{Shazam is correct}) = q$. Using Bayes' theorem:

$$\begin{aligned} P(X_1 \mid S_2) &= \frac{P(S_2 \mid X_1)P(X_1)}{P(S_2 \mid X_1)P(X_1) + P(S_2 \mid X_2)P(X_2)} \\ &= \frac{(1-q)0.8}{(1-q)0.8 + q \cdot 0.2} \\ P(X_2 \mid S_2) &= \frac{P(S_2 \mid X_2)P(X_2)}{P(S_2 \mid X_1)P(X_1) + P(S_2 \mid X_2)P(X_2)} \\ &= \frac{q \cdot 0.2}{(1-q)0.8 + q \cdot 0.2} \end{aligned}$$

Common mistake: trying to find $P(X_1 \mid \text{Shazam is correct})$. We don't know that the prediction is correct, just that it predicted the Andy Williams song.

4. (a) For $X \sim N(0, 2)$,

$$\begin{aligned} P(X \geq 1.2) &= P(\frac{X-0}{\sqrt{2}} \geq \frac{1.2-0}{\sqrt{2}}) \\ &= P(Z \geq \frac{1.2}{\sqrt{2}}) \\ &= 1 - P(Z \leq \frac{1.2}{\sqrt{2}}) \\ &= 1 - P(Z \leq 0.85) \\ &= 0.1977 \end{aligned}$$

(b) For $X \sim \text{Uni}(0, 10)$ and $Y \sim \text{Uni}(0, 10)$,

$$f_X(x) = \begin{cases} \frac{1}{10} & 0 \leq x \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

$f_Y(y) = \begin{cases} \frac{1}{10} & 0 \le y \le 10 \\ 0 & \text{otherwise} \end{cases}$

Since X and Y are independent,

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{100} & 0 \le x, y \le 10 \\ 0 & \text{otherwise} \end{cases}$

(c) Let R be the event that the user is a robot, C be the position that the user clicked, and D being the position's distance to the center.

$$\begin{aligned} P(C = (3,3)|R) &= \frac{P(C=(3,3)|R)P(R)}{P(C=(3,3)|R)P(R) + P(C=(3,3)|R^C)P(R^C)} \\ &= \lim_{\varepsilon \to 0} \frac{\varepsilon f_{D|R}(2)P(R)}{\varepsilon f_{D|R}(2)P(R) + \varepsilon f_{C|R^C}((3,3))P(R^C)} \\ &= \lim_{\varepsilon \to 0} \frac{f_{D|R}(2)P(R)}{f_{D|R}(2)P(R) + f_{C|R^C}((3,3))P(R^C)} \\ &= \frac{f_{C|R}(2)0.2}{f_{D|R}(2)0.2 + f_{C|R^C}((3,3))0.8} \\ &= \frac{\frac{1}{\sqrt{4\pi}} e^{\frac{(2-0)^2}{2\cdot 2}} 0.2}{\frac{1}{\sqrt{4\pi}} e^{\frac{(2-0)^2}{2\cdot 2}} 0.2 + \frac{1}{100} 0.8} \\ &= \frac{\frac{1}{\sqrt{4\pi}} e^{-1} 0.2}{\frac{1}{\sqrt{4\pi}} e^{-1} 0.2 + \frac{1}{100} 0.8} \end{aligned}$$

Since $D = 2$ is not exactly equivalent to $C = (3,3)$, we also gave credit if you divided $f_{D|R}(2)$ by 4π, the circumference of the circle with radius 2!

5. (a) Standard deviation = $\sqrt{\text{Var}(X)} = \frac{1}{\lambda}$
Mean = $E[X] = \frac{1}{\lambda}$
Using the CDF of an exponential $F_X(x) = 1 - e^{-\lambda x}$, we get $P(0 < x < \frac{2}{\lambda}) = F_X(\frac{2}{\lambda}) = 1 - e^{-2}$.

(b) Let M be the maximum of X_1 and X_2. The CDF of M, $F_M(k)$, is the probability that both X_1 and X_2 are at most k, so we need:

$$\begin{aligned} P(M \le k) &= P(X_1 \le k \cap X_2 \le k) \\ &= P(X_1 \le k) \cdot P(X_2 \le k) && \text{by independence} \\ &= (1 - e^{-k\lambda_1})(1 - e^{-k\lambda_2}) \end{aligned}$$

6. (a) $P(\text{no drop in any given minute}) = 1 - p$. The number of requests in each minute, and therefore the probability of dropping a request in each minute, are independent random variables, so

$$\begin{aligned} P(\text{no drops in whole hour}) = (1-p)^{60} &> 0.99 \\ p &< 1 - (0.99)^{\frac{1}{60}} \\ & \left[\text{or } 1 - \sqrt[60]{0.99}, \text{ or } 1 - e^{\left(\frac{1}{60}\ln 0.99\right)}\right] \end{aligned}$$

This is equal to $P(X = 0)$ for $X \sim \text{Bin}(60, p)$. (Common mistake: using 15 million as the number of trials for this binomial.) Since $n = 60$ is large enough and p is small, we gave almost full credit for a correct Poisson approximation ($p < -\frac{1}{60}\ln 0.99$). However, note that approximation is unnecessary here and doesn't really simplify the answer.

(b) Let Y be a random variable representing the number of requests received in a particular minute. $Y \sim \text{Poi}(\lambda)$, with $\lambda = 15{,}000{,}000$. Each server can handle $C = 10{,}000$ requests per minute, so the probability of dropping a request in any given minute is

$$\begin{aligned} p = P(Y > CK) &= 1 - P(Y \leq CK) \\ &= 1 - \sum_{i=0}^{CK} p_Y(i) \\ &= 1 - \sum_{i=0}^{CK} e^{-\lambda} \frac{\lambda^i}{i!} \end{aligned}$$

From part (a) we have $p < 1 - (0.99)^{\frac{1}{60}}$, so

$$\begin{aligned} p = 1 - \sum_{i=0}^{CK} e^{-\lambda} \frac{\lambda^i}{i!} &< 1 - (0.99)^{\frac{1}{60}} \\ \sum_{i=0}^{10\,000K} e^{-15\,000\,000} \frac{15\,000\,000^i}{i!} &> (0.99)^{\frac{1}{60}} \end{aligned}$$

(There isn't a closed-form solution for K in terms of elementary operations; we gave full credit for getting to this inequality.)

Part (d) explores the use of the normal approximation, which is applicable for this value of λ, but we asked for an exact solution in part (a), so we took off a bit more for using an approximation here.

(c) We start by approximating $X \sim P(1000)$ with $Y \sim N(1000, 1000)$:

$$\begin{aligned} p(990 < X < 1000) &= P(990.5 < Y < 999.5) && \text{continuity correction} \\ &= P(Y < 999.5) - P(Y < 990.5) \\ &= P\left(\frac{Y - 1000}{\sqrt{1000}} < \frac{999.5 - 1000}{\sqrt{1000}}\right) - \\ &\qquad P\left(\frac{Y - 1000}{\sqrt{1000}} < \frac{990.5 - 1000}{\sqrt{1000}}\right) \\ &= P\left(Z < \frac{-0.5}{\sqrt{1000}}\right) - P\left(Z < \frac{-9.5}{\sqrt{1000}}\right) \\ &= \Phi\left(\frac{-0.5}{\sqrt{1000}}\right) - \Phi\left(\frac{-9.5}{\sqrt{1000}}\right) \\ &= 1 - \Phi\left(\frac{0.5}{\sqrt{1000}}\right) - \left(1 - \Phi\left(\frac{9.5}{\sqrt{1000}}\right)\right) \\ &= \Phi\left(\frac{9.5}{\sqrt{1000}}\right) - \Phi\left(\frac{0.5}{\sqrt{1000}}\right) \end{aligned}$$

(d) Let X be a random variable which represents the number of requests in a minute. Since we receive on average 15 million requests per minute, $X \sim \text{Poi}(\lambda = 15 \text{ million})$.

Following from part (c), since the value of λ for X is larger than 1,000, we can approximate X using a normal that matches its mean and variance. Since both the mean and variance of a Poisson are equal to λ, $X \approx Y$ where $Y \sim N(\mu = 15 \text{ million}, \sigma^2 = 15 \text{ million})$.

We want to choose number of servers K such that the probability that we get more than $10{,}000 \cdot K$ requests in one minute is equal to p from part (a):

$$
\begin{aligned}
p &= P(X \le 10\,000K) \\
&\approx P(Y < 10\,000K + 0.5) && \text{continuity correction} \\
&= P\left(Z < \frac{10\,000K + 0.5 - \lambda}{\sqrt{\lambda}}\right) && \text{transform to standard normal} \\
&= \Phi\left(\frac{10\,000K + 0.5 - \lambda}{\sqrt{\lambda}}\right) && \text{since } \Phi \text{ is the CDF of } Z \\
\Phi^{-1}(p) &= \frac{10\,000K + 0.5 - \lambda}{\sqrt{\lambda}} && \text{using the probit function} \\
K &= \frac{\Phi^{-1}(p)\sqrt{\lambda} + \lambda - 0.5}{10\,000}
\end{aligned}
$$

The expression above would get full credit. For curiosity, let's find out the numeric answer:

$$
\begin{aligned}
K &= \frac{\Phi^{-1}(p)\sqrt{\lambda} + \lambda - 0.5}{10\,000} \\
K &= \frac{\Phi^{-1}(\sqrt[60]{0.99})\sqrt{15\,000\,000} + 15\,000\,000 - 0.5}{10\,000} \\
K &= \frac{3.5866 \cdot 3872.98 + 15\,000\,000 - 0.5}{10\,000} \\
K &= \lceil 1501.4 \rceil = 1502
\end{aligned}
$$

Interestingly, with 1500 servers, you have a 0.5 chance of dropping a request. With just two more servers the probability of not-dropping a request increases to 0.9998 (which is the numeric value for p).