ORIE 4580/5580 F18 Preliminary Exam
ORIE 5581 F18 Final Exam

Honor Code: I have neither given nor received any unauthorized aid on this exam.

Printed Name: ______________________________

Signature: ______________________________

NetID (e.g. sb2373): ______________________________

.

Instructions:

The exam script consists of this page of instructions, 13 pages of questions, and 3 pages for extra work or scratch work, for a total of 16 pages or 8 sheets. If you use the extra pages for answers to questions, then indicate that clearly in the space provided to answer the question. If you need extra paper then we will provide it.

- No cell phones/smartphones/computers/devices with a communication capability. Permitted: Graphical calculators, one page (front and back) of US-Letter notes.
- Time allowed: 120 minutes
- For numeric answers give either up to three significant digits, or simplified fractions.
- For 95% confidence interval calculations, **you can use** $z_{\alpha/2} = 2$ (instead of 1.96).

Question	Points	Out of
1		45
2		25
3		10
4		20
Total		100

1. (45 pts) Multiple choice questions. Circle the best/closest answer. Each question is worth 3 points. There is no need to show your work.

(a) Which of the following properties are *necessary* for a function $g(x), x \in [0,1]$ to be the CDF of a distribution?

A non-decreasing (i.e., $g(x) \geq g(y)$ for any $x > y$) **B** continuous

C $g(0) = 0, g(1) = 1$ **D** A and C **E** A, B and C

A (D also accepted)

A is essential for a valid CDF, since $g(x) - g(y) = \mathbb{P}[X \leq 0]$. B is not necessary, and in particular, not true for discrete distributions. C is also not necessarily true for discrete distributions (since $g(0) = p(0)$); what is true however is that $g(0^-) = 0$ (i.e, $g(x) = 0 \,\forall\, x < 0$) – however, since this is somewhat confusingly worded, we have given points for both A and D.

(b) You get tested for a rare genetic condition, for which you know the best testing method is correct 99% of the time (i.e., if you have the condition, it says that you do with prob 0.99; if you don't have the condition, it says that you do not with prob 0.99). Moreover, you know the condition occurs in the general population in only one of every 10,000 people. If your test results come back positive, what is the (approximate) chance that you actually have the disease?
Hint: Recall Bayes' Theorem $\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}$

A 99% **B** 90% **C** 50% **D** 10% **E** 1%

E

Let B denote the event that you are tested positive, and A denote the event that you actually have the condition (and $\bar{A}$ the event that you do not). Then we know that $\mathbb{P}[A] = 10^{-5}, \mathbb{P}[\bar{A}] = 0.9999$, $\mathbb{P}[B|A] = 0.99$ and $\mathbb{P}[B|\bar{A}] = 0.01$. Also, by the total probability theorem, we have $\mathbb{P}[B] = \mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\bar{A}]\mathbb{P}[\bar{A}]$. Now by Bayes' theorem, we have:

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B|A]\mathbb{P}[A] + \mathbb{P}[B|\bar{A}]\mathbb{P}[\bar{A}]} = \frac{0.99 \times 0.00001}{0.99 \times 0.00001 + 0.01 \times (0.9999)} \approx 0.01$$

(c) Suppose you have the LCG $x_{n+1} = ax_n \mod m$ and output $u_n = \frac{x_n+1}{m+1}$, with parameters $a = 2$, $m = 31$ and seed $x_0 = 12$. What is the second output u_2?

A 1/2 **B** 5/8 **C** 9/16 **D** 17/32 **E** 3/4

C

We have $x_1 = 12 \times 2mod(31) = 24, x_2 = 24 \times 2mod(31) = 17$, and thus $u_2 = \frac{[17+1]}{31+1} = \frac{9}{16}$.

(d) A random variable X has density $f(x) = 3x^2/8$ for $x \in [0,2]$. Given a Uniform[0, 1] rv U, which of the following quantities has the same distribution as X?

A $3U^2/8$ **B** $\sqrt{8U/3}$ **C** $U^3/8$ **D** $2U^{1/3}$ **E** none of the above

D

X has a CDF $F(x) = \int_0^x 3x^2 dx/8 = x^3/8$ for $x \in [0,2]$ (and $F(x) = 0 \, \forall \, x < 0$ and $F(x) = 1 \, \forall \, x > 2$. Moreover, $F^{-1}(u) = \sqrt[3]{8u} = 2\sqrt[3]{u}$. Thus, via the inversion method, given $U \sim$Uniform$[0,1]$, we get $2U^{1/3}$ is distributed according to $F(x)$.

(e) Which of the following random variable pairs are uniformly distributed *on* the unit circle (i.e., on the curve $\{x^2 + y^2 = 1\}$)?

A $(\sin(U_1), \cos(U_2))$, where $U_i \sim$Uniform$[0,1]$

B $\left(\frac{U_1}{\sqrt{U_1^2+U_2^2}}, \frac{U_2}{\sqrt{U_1^2+U_2^2}}\right)$, where U_1, U_2 are independent Uniform$[0,1]$

C $\left(\frac{N_1}{\sqrt{N_1^2+N_2^2}}, \frac{N_2}{\sqrt{N_1^2+N_2^2}}\right)$, where N_1, N_2 are independent Normal$(0,1)$ rv

D All of the above **E** None of the above

C

A is wrong since the angle only spans over $[0,1]$ instead of $[0, 2\pi]$. B is also clearly wrong since points in $[0,1]^2$ are only in the positive orthant, so can not be uniform over the circle. Finally, to see that C is correct, recall that independent Normal$(0,1)$ are spherically symmetric (we proved this as part of the Box-Muller method).

(f) Let X be a random variable with with CDF $F(x) = x^a$ on the interval $[0,1]$ (with parameter $a \geq 0$). Suppose you have a sample you suspect to come from this distribution. What is the method of moments estimate of a?

A $\log m_1$ **B** $\frac{1-m_1}{m_1}$ **C** $\frac{1}{m_1} - 2$ **D** $\frac{m_1}{1+m_1}$ **E** $\frac{m_1}{1-m_1}$

E

For the given family, we have $f(x) = ax^{a-1}$ and thus $\mathbb{E}[X] = \int_0^1 x.ax^{a-1}dx = \frac{a}{a+1}$. Setting $\frac{a}{a+1} = m_1$, we get $a = \frac{m_1}{1-m_1}$.

(g) For the same distribution as in the previous part, what is the MLE estimate of a?

A $-n/(\ln \prod_{i=1}^n X_i)$ **B** $-n \sum_{i=1}^n \ln X_i$ **C** $\max_i |X_i|$

D $\sqrt{-2\sum_i \ln X_i}$ **E** $\frac{\max_i X_i - \min_i X_i}{2}$

A

The log-likelihood function for any data $(X_1, X_2, \ldots, X_n)$ is $l(a) = \sum_{i=1}^n ln(aX_i^{a-1}) = n \log a + (a-1)\sum_{i=1}^n lnX_i$ for $a \geq 0$. Thus we have:

$$\frac{dl(a)}{da} = \frac{n}{a} + \sum_{i=1}^n lnX_i = \frac{n}{a} + ln\prod_{i=1}^n X_i$$

Moreover, $\frac{dl(a)}{da} = -n/a^2 < 0$. Thus, setting $\frac{dl(a)}{da} = 0$, we get $a^* = -n/(\ln \prod_{i=1}^n X_i)$

(h) Consider an alternate grading policy for assignments, to speed up how fast they are graded: given an assignment with 5 questions, each worth 20 points, the grader chooses a *random* question for each student, and then sets the overall score of that student to be 5 times their score on that problem.

Let x_i be the score a student would have got for question $i \in \{1, 2, \ldots, 5\}$ if they were all graded, and $\hat{x} = \sum_{i=1}^{5} x_i$ be the student's true total score. What is the expected score the student gets under the new grading policy?

A Less than $\hat{x}$ **B** $= \hat{x}$ **C** More than $\hat{x}$

D Depends on if $\hat{x} \geq 50$ or < 50 **E** $= \hat{x}$ as long as $x_i \neq 0$ for any question

B

The expected score $\mathbb{E}[X]$ for the given student is $5 \times \left(\sum_{i=1}^{5} \frac{1}{5} \times x_i\right) = \hat{x}$.

(i) Define $s^2 = \sum_{i=1}^{5} x_i^2$. What is the variance of the score that the student receives under this new grading policy?

A $\frac{s^2}{5}$ **B** $\frac{s^2}{5} - \hat{x}^2$ **C** $5s^2 - \hat{x}^2$ **D** $\frac{s^2 - \hat{x}^2}{5}$ **E** $s^2 - \hat{x}^2$

C

Define $X = 5Y$, where $Y = x_i, i \in \{1, 2, \ldots, 5\}$ with prob 1/5. Then $Var(X) = 25Var(Y)$, $\mathbb{E}[Y] = \hat{x}/5$ and $\mathbb{E}[Y^2] = \frac{1}{5}\sum_{i=1}^{n} x_i^2 = s^2/5$. Thus we have that $Var(X) = 25(\mathbb{E}[Y^2] - \mathbb{E}[Y]^2) = 25 \times (s^2/5 - \hat{x}^2/25) = 5s^2 - \hat{x}^2$.

(j) The TAs like this policy (call it policy A), but propose a modification (policy B) which they like more: rather than choosing a different random question for each student, they propose choosing a single question, uniformly at random, to grade for all students, after which they scale the points by 5 to get the assignment grade. You wonder what this does to the class average.

Let μ be the true class average (if all questions were graded), and X_A and X_B be the class averages under the two new grading policies. Then:

A $\mathbb{E}[X_B] < \mathbb{E}[X_A] = \mu$ **B** $\mu = \mathbb{E}[X_A] < \mathbb{E}[X_B]$ **C** $\mu = \mathbb{E}[X_A] = \mathbb{E}[X_B]$

D $\mu, \mathbb{E}[X_A], \mathbb{E}[X_B]$ are all different **E** Different distributions result in (A) or (B)

C

Let $X_A(s), X_B(s)$ be the random grades obtained by a student s under the policies A and B, and her true total grade be $\hat{x}(s)$. Since both policies seem identical to a single student (both comprise grading a single random question), therefore $\mathbb{E}[X_A(s)] = \mathbb{E}[X_B(s)] = \hat{x}(s)$. Moreover, by linearity of expectation, the class average under policy A is $\mathbb{E}[X_A] = \mathbb{E}[\frac{1}{n}\sum_{s=1}^{n} X_A(s)] = \frac{1}{n}\sum_{s=1}^{n} \mathbb{E}[X_A(s)] = \frac{1}{n}\sum_{s=1}^{n} \hat{x}(s)$, and the similarly for policy B, we have $\mathbb{E}[X_B] = \mathbb{E}[\frac{1}{n}\sum_{s=1}^{n} X_B(s)] = \frac{1}{n}\sum_{s=1}^{n} \mathbb{E}[X_B(s)] = \frac{1}{n}\sum_{s=1}^{n} \hat{x}(s)$. Thus, both methods, in expectation, preserve the class average.

(k) An internet troll likes to be the first person to reply whenever a celebrity tweets. The troll estimates from past data that the celebrity's tweets follow a homogeneous Poisson process between 5pm and 10pm, and also, that the probability no tweet is published during this 5 hour period is 0.4. Suppose the troll logs on at

5pm, and plans to monitor Twitter till 6pm. What is the chance that at least one tweet is published by the celebrity during this time?

A $1 - e^{-0.1}$ **B** $1 - (0.4)^{1/5}$ **C** 0.92 **D** $1 - (0.6)^{1/5}$ **E** 0.12

B

Let the rate of tweeting be λ per hour - then we have $\mathbb{P}[\text{No tweet in 5 hours}] = e^{-5\lambda} = 0.4$. Thus, $\mathbb{P}[\text{At least one tweet in 1 hours}] = 1 - \mathbb{P}[\text{No tweet in 1 hours}] = 1 - e^{-\lambda} = 1 - (0.4)^{1/5}$.

(l) During winter in Ithaca, meteorologists have observed that each day is sunny with probability p_{sun}, independent of other days. You find a copy of this study, which contains a plot based on sunshine data for 100 separate days, showing the estimated value of p as well as a 95% confidence interval which has *halfwidth* equal to 0.1. Unfortunately, the value of p is unreadable as someone dropped coffee over it. However, based on the above data, what is your best guess for the estimated value of p? *For 95% confidence interval calculations, assume* $z_{\alpha/2} = 2$.

A 0.07 **B** 0.31 **C** 0.5 **D** 0.64 **E** 0.91

C

For a Bernoulli(p) random variable, we know that $\sigma^2 = p(1-p)$ and thus a 95% CI has halfwidth approximately $2\sigma/\sqrt{n} = 2\sqrt{p(1-p)/n}$. Squaring, we get $p(1-p) \approx 0.1^2 \times 100/4 = 1/4$ and thus $p \approx 0.5$.

(m) A pollster uses *independent* phone surveys, one for California and one for Texas, to obtain 95% prediction intervals for the number of people who plan to vote in the midterms in California (X) and Texas (Y) respectively. The probability that her intervals *simultaneously* capture the actual number of people who will vote in the two states is? *Hint: Note that the polls are independent*

A > 0.95 **B** $= 0.95$ **C** $= 0.95^2$ **D** $= 0.90$ **E** Either 0 or 1.

C

By definition of a 95% prediction interval, we know that $\mathbb{P}[\text{interval for CA captures actual number of voters}] = \mathbb{P}[\text{interval for Texas captures actual number of voters}] = 0.95$. Moreover, since the polls are independent, we have that $\mathbb{P}[\text{Interval for CA is correct AND Interval for TX is correct}] = \mathbb{P}[\text{Interval for CA is correct}]\mathbb{P}[\text{Interval for TX is correct}] = 0.95^2$. Note that here we are ok with multiplying the two (instead of using the union bound) because we believe the two events are independent.

(n) In order to get a desired half-width for pricing a complex financial derivative, an analyst at Silverman Punts found that they needed $n = 10000$ replications, each of which took 36 seconds on average. Using your new simulation skills, you manage to find a new estimator which took the same time to compute, but has half the standard deviation of the old estimator. How much time did you *save* the company for this calculation?

A 5 hours **B** 15 hours **C** 30 hours **D** 50 hours **E** 75 hours

E

The original simulation took $36 \times 10000sec = 100hrs$. If we reduce σ by 0.5, then in order to keep the same CI width, we need $1/4^{th}$ the replications (since we want $\sigma/\sqrt{n}$ to be constant) – thus, we only need $25hrs$, resulting in a saving of $75hrs$.

(o) You use thinning to generate a non-homogeneous Poisson arrival process with rate function $(\lambda(t) : 0 \leq t \leq 10)$ on the interval $[0, 10]$. You observe that $\lambda(t) \leq 8$ for all $t \in [0, 10]$, so choose $\lambda^* = 8$. Unfortunately, in the thinning step you accidentally code "if $\lambda^* U > \lambda(T^*)$", instead of the intended "if $\lambda^* U < \lambda(T^*)$." The resulting process of accepted arrival times

A cannot be described as any standard stochastic process
B is a Poisson process with rate function $(\lambda^* - \lambda(t) : 0 \leq t \leq 10)$
C is a Poisson process with rate function $(\lambda(t) + \lambda^* : 0 \leq t \leq 10)$
D is a Poisson process with rate function $(\lambda(t)/\lambda^* : 0 \leq t \leq 10)$
E is a Poisson process with rate function $((\lambda^* - \lambda(t))/\lambda^* : 0 \leq t \leq 10)$

B
This corresponds to accepting an arrival at time t with probability $\frac{\lambda^* - \lambda(t)}{\lambda^*}$ – this is the same as simulating a Poisson process with rate $\lambda^* - \lambda(t)$.

2. (30 pts) Short-answer questions. **Please provide justifications for your answers**.

(a) (7 pts) TCAT has been receiving complaints that Route 10 buses are getting bunched up together at the Collegetown stop, leading to traffic. The city authorities recommend the spacing between buses should be at least 7 minutes. To understand the problem, TCAT engineers decide to model the bus inter-arrival time T to be normally distributed with mean μ and variance σ^2, where μ and σ^2 are unknown. Based on 10,000 past trips, they compute a 95% confidence interval of $[11.9, 12.1]$ for μ, the gap between bus arrivals at Collegetown. Using this information, give an approximation for $\mathbb{P}[T < 7]$ in terms of the standard Normal cdf $\Phi(x)$.
The mean $\mu \approx 12$ (midpoint of CI). Also, $2\sigma/\sqrt{10000} \approx 0.1$ so that $\sigma \approx 5$. Then, standardizing, $\mathbb{P}[T < 7] = \mathbb{P}[(T - 12)/5 < (7 - 12)/5] \approx \mathbb{P}[N < -1] = \Phi(-1)$.

(b) (10 pts) You are given a binary random number generator (RNG) which returns 0 with probability p, else returns 1; you want to use this to generate a Bernoulli(0.5) rv. Consider the following protocol (the *von Neumann extractor*):

1. Create 2 copies of the RNG, and give one to your friend. Both of you initialize your RNGs with different seeds.
2. The protocol now proceeds in rounds, where in each round, both you and your friend generate a number from your RNGs.
3. In the current round:
– If your RNG returns 1, and your friend's returns 0, then output 1 and stop.
– If your RNG returns 0, and your friend's returns 1, then output 0 and stop.
4. If however both of you draw 0s, or both draw 1s, then you discard the samples and start a new round (i.e., go to step 2).

Assume on each draw, the RNG returns independent samples, and that the two generators are independent.

(i) Prove that the above protocol generates a Bernoulli(0.5) rv. (5 pts)
Hint: It is sufficient to argue that conditioned on stopping in round number k (for any $k \in \{1, 2, \dots\}$), you output 1 and 0 with equal probability.

Suppose you stop in the k^{th} round – then the probability you output a 1 is $(1-p)p$, which corresponds to you getting an odd number and your friend getting an even number; by symmetry, the probability of outputting 0 is $p(1-p)$, and thus the two are equal.

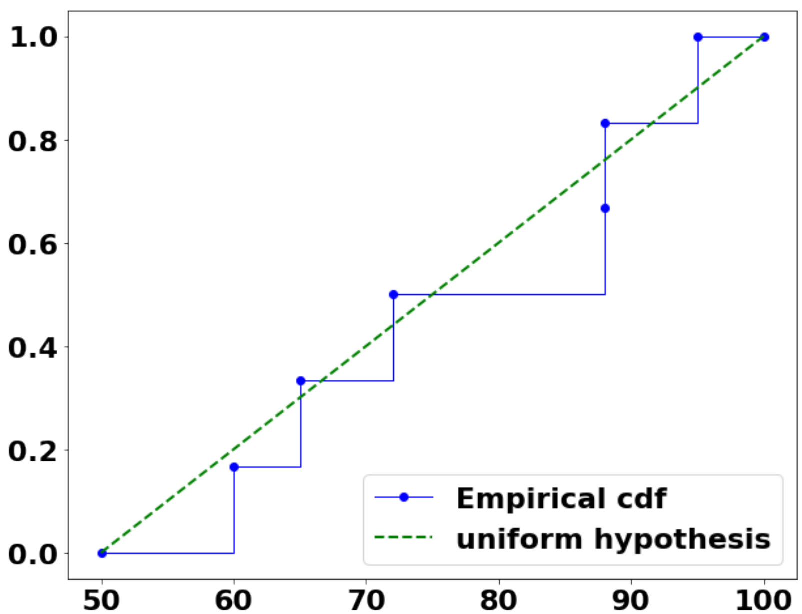
(ii) On average, how many draws (in terms of p) do you need from the RNG in order to obtain a single Bernoulli(0.5) sample? (5 pts)

Each 'round' requires 2 draws, and each round successfully terminates with probability $2p(1-p)$, else continues. Thus, the number of rounds is a Geometric($2p(1-p)$) rv, and hence on average, we need $2 \times \frac{1}{2p(1-p)}$ draws from the RNG.

(c) (8 pts) You believe that the midterm grades of students in Simulation are uniformly distributed in $[50, 100]$ (to keep things simple, let's assume the grades are continuous). To test this, you ask 6 of your friends from last years' batch, and learn that their grades were $\{88, 95, 65, 88, 72, 60\}$. Plot the empirical CDF of these samples (without the continuity correction) and compute the Kolmogorov-Smirnoff test statistic $D_{KS} = \max_{x\in[50,100]} |F(x) - \hat{F}(x)|$.

Note: CDF of a Unif$[a, b]$ distrib is $F(x) = (x - a)/(b - a)$ for $x \in [a, b]$

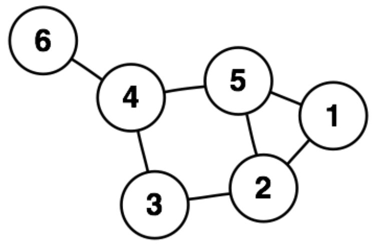
Note that one of the values (88) repeat – this corresponds to a jump of $2/n = 0.33$ at 88. This gives us the following empirical CDF (drawn along with the CDF of Uniform$[50, 100]$, which is the distribution we are proposing).



The KS statistic is $D_{KS} = 0.26$ which is achieved at $x = 88$

3. (10 pts) In 1991, the sociologist Scott Feld wrote an article about social networks titled *"Why your friends have more friends than you do"*. Some of you may recognize this sentiment, and perhaps feel this *friendship paradox* is because you spend too much time working on Simulation…however, we will now show this is not the case!

(a) (6 pts) Consider the following simple social network with 6 users (denoted by the nodes), with undirected edges between 'friends'.



Compute (and compare) the 'average number of friends' in two ways:

(*i*) Ask each user to report the number of friends they have, and find the 'average number of friends' $\hat{x}$.

(*ii*) Ask each user to report *the number of friends that each of their friends' has* (for example, user 5 reports $\{2, 3, 3\}$ as the number of friends that user 1, 2 and 4 have respectively). Concatenate these reports and find the average number of 'friends of friends' $\hat{y}$.

Note that the number of data points for computing $\hat{x}$ is 6 (one per user), while for computing $\hat{y}$ is more (in particular, there are 14 data points, since there are 7 edges and each edge corresponds to two reports).

First we compute the average number of friends as

$$\hat{x} = \frac{2+3+2+3+3+1}{6} = \frac{14}{6} = 2.33$$

On the other hand, we can compute the average number of friends-of-friends as

$$\hat{y} = \frac{(3+3)+(2+2+3)+(3+3)+(2+3+1)+(2+3+3)+(3)}{14} = \frac{36}{14} = 2.57$$

Thus we see that $\hat{y} > \hat{x}$ and hence, the average number of friends-of-friends is more than the average number of friends.

(b) (4 pts) Now we prove this holds in general. Consider any undirected social network with n users, where the user i has d_i friends: the vector $\{d_1, d_2, \ldots, d_n\}$ is known as the *degree sequence* of the network. Let μ and σ^2 denote the mean and variance of the degree sequence if we pick users uniformly at random, and $m = \sum_{i=1}^{n} d_i = n\mu$ denote the sum of degrees (which also corresponds to twice the number of edges in the network, since each edge contributes to two user degrees).

Argue that for any such network, we have that the average number of friends $\hat{x} = \mu$, and average number of friends-of-friends $\hat{y} = \mu + \frac{\sigma^2}{\mu}$.

Hint: To compute $\hat{y}$, try to figure out how many users report the number of friends d_i for any particular user i.

As before, we compute the average number of friends as

$$\hat{x} = \frac{\sum_{i=1}^{n} d_i}{n} = \mu$$

On the other hand, to compute the average number of friends-of-friends, note that each user i reports d_i separate values of friends-of-friends; moreover, any particular user j has d_j friends, all of whom report that j has d_j friends. Thus, we have

$$\hat{y} = \frac{\sum_{j=1}^{n} d_j^2}{\sum_{i=1}^{n} d_i} = \frac{\sum_{j=1}^{n} d_j^2}{n} \times \frac{n}{\sum_{i=1}^{n} d_i} = \left(\mu^2 + \sigma^2\right) \times \frac{1}{\mu} = \mu + \frac{\sigma^2}{\mu}$$

Thus we see that $\hat{y} > \hat{x}$ for any social network.

4. (20 pts) An important use of probability is in ensuring *privacy* in data collection. In this question, we will explore this idea in the context of conducting sensitive surveys.

Suppose a researcher wants to find the fraction of students in the Cornell senior-year batch who have blacked out after excessive drinking at some time in the past. The researcher knows that if asked directly, all students will deny it; to get around this, she asks each student to do the following:

- First, the student tosses a fair coin twice
- If the first toss comes up heads, the student reports the truth (i.e., *Yes* if blacked out in the past, *No* otherwise)
- If the first toss comes up tails, the student reports *Yes* if the second toss returns heads, and *No* if the second toss returns tails

(a) (5 pts) The reason the above protocol may be acceptable to the students is the idea of *plausible deniability* – the idea that even if students report *Yes*, they can claim that it was because of the coin toss. To formalize this, suppose before the survey, people believed that the probability a particular student may have blacked out was 10% – prove that even if the student reports *Yes* on the survey, the new belief on the student having blacked out is at most 30%.

Let p be the prior belief (i.e., before the survey) that a particular student may have blacked out (in this case, $p = 0.1$). From Bayes' Theorem, we have:

$$\begin{aligned}
&\mathbb{P}[\text{Student has blacked out}|\text{student reports } \textit{Yes}] \\
&\qquad = \frac{\mathbb{P}[\text{Student has blacked out AND student reports } \textit{Yes}]}{\mathbb{P}[\text{student reports } \textit{Yes}]} \\
&\qquad = \frac{p \times 0.75}{p \times 0.75 + (1-p) \times 0.25} = \frac{3p}{1+2p} \leq 3p
\end{aligned}$$

Thus, after the survey, the belief that the student may have blacked out is at most 30% (and in general, the probability at most doubles).

(b) (8 pts) Now suppose the batch consists of n students, of which k have blacked out due to excessive drinking. Let X denote the (random) number of students that say *Yes* in the survey. Assuming everyone follows the above protocol, what is the expected value and variance of X?

Let Y_i be the indicator that the i^{th} student has blacked out, and X_i be the indicator that the i^{th} student says *Yes*. Moreover, let X be the total number of students who report *Yes*. Then by definition, we have $\sum_{i=1}^{n} Y_i = k$, and by linearity of expectation, we have $\mathbb{E}[X] = \mathbb{E}[\sum_{i=1}^{n} X_i] = n\mathbb{E}[X_1]$, and moreover:

$$\begin{aligned}\mathbb{E}[X_1] &= \mathbb{P}[\text{First coin is H}] \times Y_i + \mathbb{P}[\text{First coin is H}] \times \mathbb{P}[\text{Second coin is H}] \\ &= \frac{2Y_i + 1}{4}\end{aligned}$$

Thus $\mathbb{E}[X] = \frac{2\sum_i Y_i + n}{4} = \frac{2k+n}{4}$.

Similarly, we have $Var(X) = \sum_{i=1}^{n} Var(X_i)$, and since $X_i \sim Ber(\frac{2Y_i+1}{4})$, we have

$$\begin{aligned}Var(X) &= \sum_{i=1}^{n} \frac{1+2Y_i}{4}\left(1 - \frac{1+2Y_i}{4}\right) \\ &= \sum_{i=1}^{n} \frac{3+4Y_i-4Y_i^2}{16} = \sum_{i=1}^{n} \frac{3}{16} = \frac{3n}{16}\end{aligned}$$

Here, we use the fact that $Y_i^2 = Y_i$ since $Y_i \in \{0, 1\}$.

Another way to see that this must be the variance is to observe that students who have blacked out in the past will say *Yes* with probability $3/4$ else *No*, while students who have not blacked out will say *Yes* with probability $1/4$ else say *No*. Thus, in both cases, the variance of the indicator variable that a given student says *Yes* is the same $(= \frac{3}{4} \times \frac{1}{4} = \frac{3}{16})$. Moreover, since student reports are independent, therefore the $Var(X) = \frac{3n}{16}$.

(c) (7 pts) Argue that $\hat{\gamma} = \frac{2X - 0.5n}{n}$ is an unbiased estimator for the fraction $\gamma = k/n$ of students who have blacked out, and moreover, $Var(\hat{\gamma})$ is at most $1.25/n$.

From the previous part, we have that $\mathbb{E}[X] = \frac{n+2k}{4}$, and hence we have $\frac{k}{n} = \frac{2\mathbb{E}[X]}{n} - 0.5$. Thus, $\hat{\gamma} = \frac{2X}{n} - \frac{1}{2}$ gives an unbiased estimator for k/n, and we have $Var(\gamma) = \frac{4Var(X)}{n^2} = \frac{3}{4n} \leq \frac{5}{4n}$.

Extra work for Question ______ only:
Extra work for Question ______ only:
Extra work for Question ______ only: