

CS109 Midterm Exam 2

This is a closed calculator/computer/phone/smart-watch/smart-toothbrush exam. You are, however, allowed to use 5 pages (front and back) of handwritten notes in the exam. You have 75 minutes to take the exam. The exam is 65 points. You may want to use the point allocation for each problem as a rough indicator for pacing yourself on the exam.

In the event of an incorrect answer, any explanation you provide of how you obtained your answer can potentially allow us to give you partial credit for a problem. For example, describe the distributions and parameter values you used, where appropriate. It is fine for your answers to include summations, products, factorials, exponentials, and combinations. You can leave your answer in terms of Φ (the CDF of the standard normal) or Φ^{-1} . For example $\Phi(3/4)$ is an acceptable final answer. You can only leave your answer in **terms of code** when the problem explicitly says so.

Stanford Honor Code: The Honor Code is an undertaking of the Stanford academic community, individually and collectively. Its purpose is to uphold a culture of academic honesty. Students will support this culture of academic honesty by neither giving nor accepting unpermitted academic aid on this examination.

This course is participating in the proctoring pilot overseen by the Academic Integrity Working Group (AIWG), therefore proctors will be present in the exam room. The purpose of this pilot is to determine the efficacy of proctoring and develop effective practices for proctoring in-person exams at Stanford.

I acknowledge and accept the letter and spirit of the honor code. I pledge to write more neatly than I have in my entire life:

Signature: _____

First and Last Name (print): _____

Stanford Email (@stanford.edu): _____

Exam Break Sign-out

I pledge that during my exam break:

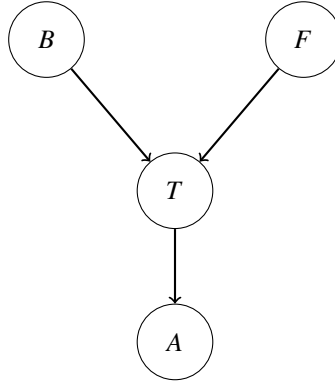
- I will not bring any paper, electronic devices (phone, smart watch, smart glasses, etc), or aid (permitted or unpermitted) *out of or into* the exam room, nor access any aid during the break.
- I will not communicate with anyone other than the course instructional staff about the content of the exam.

Signature Confirming Honor Code	Exit Time	Return Time	Proctor Initial	Length (mins)

If you are feeling unwell and are not able to complete the exam, please speak with the proctor.

1 Continuous Integration [13 Points]

Your team's continuous integration (CI) system is modeled by the Bayesian network below. All four random variables are binary (1 = true, 0 = false): B = the commit under test contains a **bug**, F = the test machine is running **flaky** infrastructure, T = the **test** suite fails, and A = the on-call engineer is **al**erted.



$$P(B = 1) = 0.1$$

$$P(F = 1) = 0.2$$

$$P(T = 1 \mid B = 1, F = 1) = 0.99 \quad P(T = 1 \mid B = 1, F = 0) = 0.90$$

$$P(T = 1 \mid B = 0, F = 1) = 0.50 \quad P(T = 1 \mid B = 0, F = 0) = 0.02$$

$$P(A = 1 \mid T = 1) = 0.95$$

$$P(A = 1 \mid T = 0) = 0.01$$

- (a) (3 points) Write the joint distribution $P(B, F, T, A)$ as a product of conditional probabilities implied by the network, and use it to compute $P(B = 0, F = 1, T = 0, A = 1)$.

- (b) (4 points) The test suite just failed. What is $P(B = 1 \mid T = 1)$, the probability the commit really contains a bug?

- (c) (2 points) Suppose you already know that the test suite failed ($T = 1$). If you additionally learn that the commit contains a bug ($B = 1$), does the probability that the on-call engineer is alerted change? Answer yes or no and justify your answer in one sentence using the structure of the network.
- (d) (2 points) You could also have estimated the probability in part (b) using *rejection sampling*. State one advantage and one disadvantage of rejection sampling in general.
- (e) (2 points) For *this particular* network and query, would you prefer to compute the answer by hand as in part (b), or to run rejection sampling? Justify your answer in one or two sentences.

2 A Suspicious Coin [13 Points]

You come across a special coin, which has a heads side and a tails side. You can tell just by looking at it that it is weighted unusually, but you have no idea in which direction. Let p be the (unknown) probability that this coin comes up heads.

You may not use code to answer any part of this problem.

(a) (3 points) Define a random variable, *including its parameters*, to represent your prior belief about p , before you have made any flips.

(b) (2 points) Write a sentence justifying your choice of random variable.

(c) (3 points) You flip the coin 8 times and observe 5 heads and 3 tails. What is the distribution for your *posterior* belief about p ?

(d) (3 points) For parts (d) and (e), suppose instead that you had observed an arbitrary x heads and y tails, with $x + y \geq 1$. Compared to your prior, did the *variance* of your belief distribution increase, decrease, or can you not say? Justify your answer.

(e) (2 points) Did the *entropy* of your belief distribution increase, decrease, or can you not say? Justify your answer.

3 Ride Sharing [12 Points]

You run a ride-sharing app. Each driver receives ride requests at a rate of 5 requests per hour, and each driver earns \$10 per ride fulfilled. Assume that requests arrive independently, that the number of requests is unbounded, and (for the sake of this problem) that a driver can fulfill an unlimited number of requests.

You may not use code to answer any part of this problem.

- (a) (3 points) What are the expected earnings over the course of one hour for an individual driver?

- (b) (3 points) What is the variance of an individual driver's earnings in one hour?

- (c) (6 points) Your fleet includes 100 drivers, who receive requests independently of one another. What is the probability that your fleet earns *at least* \$5,200 in one hour? You may use an approximation.

4 A Delicate Poll [14 Points]

You want to estimate the probability p that a randomly selected person believes the answer to a single delicate true/false question is “true.” Because the topic is sensitive, people will not answer it honestly if asked directly.

So instead you hand each participant 7 true/false questions and ask them only to report *how many* of the 7 they believe are true. One of the 7 is the delicate question. The other 6 are unimportant, non-sensitive questions for which each person is equally likely to answer true or false, independently of every other answer.

Let X_i be the number of questions person i believes are true. You survey $n = 100$ people and observe $X_1, \dots, X_{100}$, stored in a list `samples`.

- (a) (3 points) Derive $E[X_i]$ in terms of p , and use it to give an estimator $\hat{p}$ for p in terms of the sample mean $\bar{X}$.

- (b) (6 points) The code below uses bootstrapping to estimate the probability that $\hat{p}$ is within 0.1 of the true value of p . Line (i) is written for you; fill in the three remaining blanks. `np.random.choice(a, size, replace)` draws `size` values from the list `a`; `np.mean(a)` returns the average of the list `a`.

```
import numpy as np
NUM_BOOTSTRAPS = 10000

def estimate_p(samples):
    return np.mean(samples) - 3

def bootstrap_p_within(samples):
    our_estimate = estimate_p(samples)
    n = len(samples)
    count = 0
    for i in range(NUM_BOOTSTRAPS):
        resample = np.random.choice(samples, size=n, replace=True) # (i)

        p_hat = _____ # (ii)

        if _____: # (iii)
            count += 1

    return _____ # (iv)
```

(c) (3 points) Explain what would go wrong if line (i) were changed to use `replace=False`, and what would go wrong if it instead used `size=n//2`.

(d) (2 points) The number returned approximates $P(|\hat{p} - p| < 0.1)$. What assumption does the bootstrap rely on to make this valid?

5 Load Balancing [13 Points]

A load balancer routes each incoming request, independently, to one of three servers: server A with probability $1/2$, server B with probability $1/3$, and server C with probability $1/6$.

You may not use code to answer any part of this problem.

- (a) (3 points) Twelve requests arrive. What is the probability that exactly 6 go to A , exactly 4 go to B , and exactly 2 go to C ?

- (b) (2 points) Still with 12 requests, let N_C be the number of requests routed to server C . State the distribution of N_C (including its parameters), and give $E[N_C]$ and $\text{Var}(N_C)$.

- (c) (3 points) Let $S \in \{A, B, C\}$ be a random variable representing the server that handles a single request. Compute the entropy $H(S)$, using $\log_2$. Then state which routing probabilities would *maximize* $H(S)$, and what that maximum value is.

- (d) (2 points) Server A processes a request in 2 ms, server B in 3 ms, and server C in 10 ms. What is the expected processing time of a single request?

- (e) (3 points) Server C is overloaded. With probability $1/4$ it drops the request after spending its full 10 ms on it, and the dropped request is then sent back through the load balancer and routed again from scratch, independently of everything that came before. (Servers A and B never drop requests.) Let R be the total time from arrival until the request is finally processed. Compute $E[R]$.

You're done! A few parting thoughts. The CI network in Problem 1 is a real pattern: "the tests failed" is much weaker evidence of a bug than it feels like in the moment, because flaky infrastructure explains failures too. Problem 2 is the whole Bayesian story in five parts — a prior, some data, a sharper posterior — and parts (d) and (e) say the same thing twice, once in the language of variance and once in the language of information. The delicate-polling design in Problem 4 is a genuine technique from survey statistics called randomized response, and the bootstrap is what lets you put error bars on it without knowing the population.