Chris Gregg
CS109

Summer 2026
August 3rd

CS109 Midterm 2 Practice Problems

These three problems are written to be similar in *scope and style* to what you will see on Midterm 2. They are practice, not a preview: the topics here are drawn from across the second half of the course, and the exam draws from the same pool but is not built from these problems. Do not read anything into which topics do or do not appear below.

You are welcome to use the course reader for this practice, but for the midterm you will be limited to five back/front handwritten pages of notes.

As on the exam, it is fine for your answers to include summations, products, factorials, exponentials, logarithms, and combinations. You may leave answers in terms of Φ (the CDF of the standard normal) or Φ^{-1}. Do not use a calculator — the numeric values in the solutions are there so you can check yourself afterwards, not because you need to produce them.

In the event of an incorrect answer, an explanation of how you got there can earn partial credit. Naming the distribution and its parameters is almost always worth writing down.

Topics covered: Poisson and Multinomial random variables, conditioning and the law of total expectation, maximum likelihood estimation, the Central Limit Theorem, p-values and resampling, and entropy and information gain.

1 Cache Tiers

A web service receives read requests independently at a rate of 12 per minute. Each arriving request is served, independently of every other request, by one of three tiers: the **L1** in-memory cache with probability $1/2$, the **L2** shared cache with probability $1/3$, and, if both caches miss, the **disk** with probability $1/6$.

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

(a) Suppose you are told that exactly 9 requests arrived during one particular minute. Given that, what is the probability that exactly 5 were served by L1, exactly 3 by L2, and exactly 1 by disk?

Once the total is fixed at 9, each request independently lands in one of three categories, which is exactly the Multinomial setup. Let $(N_1, N_2, N_D) \sim \text{Multinomial}\big(n = 9;\ 1/2, 1/3, 1/6\big)$. Using the Multinomial PMF,

$$P(N_1 = 5, N_2 = 3, N_D = 1) = \binom{9}{5,\,3,\,1} \left(\tfrac{1}{2}\right)^5 \left(\tfrac{1}{3}\right)^3 \left(\tfrac{1}{6}\right)^1 = \frac{9!}{5!\,3!\,1!} \cdot \frac{1}{2^5 \cdot 3^3 \cdot 6}.$$

Numerically, $\frac{9!}{5!3!1!} = \frac{362880}{720} = 504$ and $2^5 \cdot 3^3 \cdot 6 = 5184$, so the probability is

$$\frac{504}{5184} = \frac{7}{72} \approx 0.097.$$

(b) Now stop conditioning on the total. Let N be the number of requests in a one-minute window and let D be the number of those requests served by disk. State the distribution of N (with parameters) and compute $E[D]$ using the law of total expectation.

Events arriving independently at a constant rate with no upper bound on the count is the Poisson model, so $N \sim \text{Poi}(\lambda = 12)$ and $E[N] = 12$.
Conditioned on $N = n$, each of the n requests independently goes to disk with probability $1/6$, so $D \mid N = n \sim \text{Bin}(n, 1/6)$ and $E[D \mid N = n] = n/6$. By the law of total expectation,

$$E[D] = E\big[E[D \mid N]\big] = E\left[\frac{N}{6}\right] = \frac{E[N]}{6} = \frac{12}{6} = 2.$$

(c) It is a fact (which you may use without proof) that $D \sim \text{Poi}(2)$. What is the probability that at least 2 requests go to disk in a given minute?

Complement the two easy cases:

$$P(D \geq 2) = 1 - P(D = 0) - P(D = 1) = 1 - e^{-2}\frac{2^0}{0!} - e^{-2}\frac{2^1}{1!} = 1 - 3e^{-2} \approx 1 - 0.406 = 0.594.$$

Aside: the fact quoted in the problem is called *Poisson thinning*. If you take a $\text{Poi}(\lambda)$ stream and keep each event independently with probability p, what survives is $\text{Poi}(\lambda p)$. Notice that it is consistent with part (b): a $\text{Poi}(2)$ random variable has mean 2.

(d) An L1 hit takes 1 ms, an L2 hit takes 5 ms, and a disk read takes 40 ms. Let T be the service time of a single request. Compute $E[T]$ and $\text{Var}(T)$.

T is a discrete random variable taking the value 1 with probability $1/2$, 5 with probability $1/3$, and 40 with probability $1/6$. Directly from the definition of expectation (equivalently, by the law of total expectation conditioning on which tier served the request):

$$E[T] = \tfrac{1}{2}(1) + \tfrac{1}{3}(5) + \tfrac{1}{6}(40) = \tfrac{3}{6} + \tfrac{10}{6} + \tfrac{40}{6} = \frac{53}{6} \approx 8.83 \text{ ms}.$$

For the variance, use $\mathrm{Var}(T) = E[T^2] - (E[T])^2$:

$$E[T^2] = \tfrac{1}{2}(1) + \tfrac{1}{3}(25) + \tfrac{1}{6}(1600) = \tfrac{3}{6} + \tfrac{50}{6} + \tfrac{1600}{6} = \frac{1653}{6},$$

$$\mathrm{Var}(T) = \frac{1653}{6} - \left(\frac{53}{6}\right)^2 = \frac{9918}{36} - \frac{2809}{36} = \frac{7109}{36} \approx 197.5 \ \ (\text{ms}^2).$$

So the standard deviation is about 14.1 ms, considerably larger than the mean. That is what a heavy tail does: most requests are fast, but the rare 40 ms disk read dominates the spread.

2 Does the New Ranker Help?

For a single user session on your site, the time spent reading has mean $\mu = 12$ minutes and standard deviation $\sigma = 9$ minutes. Sessions are independent of one another. You do *not* know the shape of the per-session distribution.

(a) You collect $n = 400$ independent sessions and compute the sample mean $\bar{X}$. Approximate $P(\bar{X} > 13)$. Justify why the approximation is available to you even though the per-session distribution is unknown.

The Central Limit Theorem says that the sum (and hence the mean) of many i.i.d. random variables is approximately Normal regardless of the distribution of the individual terms, as long as it has finite mean and variance. That is exactly why not knowing the per-session shape is not a problem.

$$\bar{X} \approx \mathcal{N}\left(\mu = 12,\ \sigma^2 = \frac{9^2}{400} = \frac{81}{400}\right), \qquad \sigma_{\bar{X}} = \frac{9}{20} = 0.45.$$

Standardizing,

$$P(\bar{X} > 13) \approx P\left(Z > \frac{13-12}{0.45}\right) = 1 - \Phi\left(\frac{20}{9}\right) \approx 1 - \Phi(2.22) \approx 0.013.$$

(No continuity correction here: reading time is continuous.)

(b) You now run an experiment. 400 sessions use the old ranker and 400 different sessions use a new ranker, and the new group's mean is 0.8 minutes higher. State the null hypothesis you would test, and define, in one or two sentences, what the p-value for this experiment means.

Null hypothesis H_0: the ranker has no effect — the two groups' session times are draws from the same distribution, and the observed 0.8 minute gap is due to chance alone.
p-value: the probability, *assuming H_0 is true*, of observing a difference in group means at least as extreme as the 0.8 minutes we actually saw. It is emphatically *not* the probability that H_0 is true, and not the probability that the new ranker does nothing.

(c) The code below estimates that p-value with a permutation test. Line (i) is written for you; fill in the three remaining blanks. `np.concatenate([a, b])` joins two lists; `np.random.shuffle(a)` randomly reorders `a` in place; `np.mean(a)` returns the average of `a`.

```
import numpy as np
NUM_TRIALS = 10000

def p_value(old, new):
    obs_diff = np.mean(new) - np.mean(old)
    pooled   = np.concatenate([old, new])   # (i)
    n_new    = len(new)
    count = 0
    for i in range(NUM_TRIALS):
        np.random.shuffle(pooled)
        sim_new  = pooled[:n_new]
        sim_old  = pooled[n_new:]



        sim_diff = ________________________________________  # (ii)



        if ____________________________________________:  # (iii)

            count += 1



    return ________________________________________________  # (iv)
```

> (ii) `sim_diff = np.mean(sim_new) - np.mean(sim_old)`
>
> (iii) `if sim_diff >= obs_diff:`
>
> (iv) `return count / NUM_TRIALS`
>
> The idea: under H_0 the group labels are meaningless, so any reassignment of the 800 sessions into a group of 400 and a group of 400 is just as likely as the one we saw. Shuffling and re-splitting samples from the null distribution of the difference in means; the fraction of shuffles that beat `obs_diff` estimates the p-value.
>
> (Line (iii) as written gives a one-sided p-value, appropriate if you asked in advance only whether the new ranker *increases* session time. For a two-sided test you would compare `abs(sim_diff) >= abs(obs_diff)`.)

(d) The function returns 0.031. State what you conclude at a significance threshold of $\alpha = 0.05$, and give one reason why this conclusion could still be wrong.

> Since $0.031 < 0.05$, we reject H_0 and call the difference statistically significant.
>
> Any one of the following is a fine reason for doubt: (1) a p-value of 0.031 still means that, if the ranker does nothing, we would see a gap this large about 3% of the time — rejecting at $\alpha = 0.05$ has a 5% false-positive rate by construction; (2) if this is one of many metrics or variants tested, the chance that *some* test crosses 0.05 by luck is much higher than 5%; (3) statistical significance is not practical significance — 0.8 minutes may not be worth shipping; (4) the permutation test assumes the two groups are otherwise exchangeable, which fails if assignment was confounded (say, one group ran during a holiday).

3 One Word of Evidence

An email is spam ($Y = 1$) with probability 0.25 and legitimate ($Y = 0$) otherwise. Let $W = 1$ if the email contains the word "urgent" and $W = 0$ otherwise, with

$$P(W = 1 \mid Y = 1) = 0.6, \qquad P(W = 1 \mid Y = 0) = 0.1.$$

Use $\log_2$ throughout parts (a)–(c), so that entropies are in bits. It is fine to leave answers as expressions involving logarithms.

(a) Compute $H(Y)$.

Entropy is expected surprise, $H(Y) = \sum_y P(y) \log_2 \frac{1}{P(y)}$:

$$H(Y) = \tfrac{1}{4}\log_2 4 + \tfrac{3}{4}\log_2 \tfrac{4}{3} = \tfrac{1}{2} + \tfrac{3}{4}\big(2 - \log_2 3\big) \approx 0.500 + 0.311 = 0.811 \text{ bits}.$$

Less than 1 bit, as it must be: a biased coin is more predictable than a fair one.

(b) Compute $P(W = 1)$, and then $P(Y = 1 \mid W = 1)$ and $P(Y = 1 \mid W = 0)$.

By the law of total probability,

$$\begin{aligned} P(W = 1) &= P(Y = 1)P(W = 1 \mid Y = 1) + P(Y = 0)P(W = 1 \mid Y = 0) \\ &= (0.25)(0.6) + (0.75)(0.1) = 0.225. \end{aligned}$$

By Bayes' theorem,

$$P(Y = 1 \mid W = 1) = \frac{(0.25)(0.6)}{0.225} = \frac{0.15}{0.225} = \frac{2}{3} \approx 0.667,$$

$$P(Y = 1 \mid W = 0) = \frac{(0.25)(0.4)}{1 - 0.225} = \frac{0.10}{0.775} = \frac{4}{31} \approx 0.129.$$

One word moves the belief from 0.25 to either 0.667 or 0.129 — real evidence, but far from decisive in either direction.

(c) Compute the conditional entropy $H(Y \mid W)$ and the information gain $H(Y) - H(Y \mid W)$. Interpret the information gain in one sentence.

Write $H(p)$ for the entropy of a Bernoulli with parameter p. Then

$$H(Y \mid W) = P(W=1)\,H\left(\tfrac{2}{3}\right) + P(W=0)\,H\left(\tfrac{4}{31}\right).$$

The two pieces:

$$H\left(\tfrac{2}{3}\right) = \tfrac{2}{3}\log_2 \tfrac{3}{2} + \tfrac{1}{3}\log_2 3 = \log_2 3 - \tfrac{2}{3} \approx 0.918,$$

$$H\left(\tfrac{4}{31}\right) = \tfrac{4}{31}\log_2 \tfrac{31}{4} + \tfrac{27}{31}\log_2 \tfrac{31}{27} \approx 0.381 + 0.174 = 0.555.$$

So

$$H(Y \mid W) \approx (0.225)(0.918) + (0.775)(0.555) \approx 0.207 + 0.430 = 0.637 \text{ bits},$$

$$\text{information gain} = H(Y) - H(Y \mid W) \approx 0.811 - 0.637 = 0.174 \text{ bits}.$$

Interpretation: on average, learning whether the email contains "urgent" removes about 0.17 bits of the 0.81 bits of uncertainty about whether it is spam — roughly a fifth of the total. (This quantity is the mutual information $I(Y;W)$, and it is what a decision tree maximizes when choosing a split. Note that $H(Y \mid W) \le H(Y)$ always: conditioning never increases entropy *on average*, even though an individual outcome — here $W = 1$, with $0.918 > 0.811$ — can leave you more uncertain than you started.)