CS109 Lecture 11 Worksheet: Inference

Updating Belief Distributions with Bayes' Theorem

Summer 2026

Lecture 11 Worksheet: Inference

Topics: inference as updating a belief distribution over a random variable, the prior $\rightarrow$ posterior update (multiply prior by likelihood, then normalize), representing beliefs as a dictionary/table (non-parametric), handling multiple observations (multiply the likelihoods / posterior becomes the new prior), and discretizing continuous variables.

How to use this sheet. Work in small groups. A TA and the instructor will circulate to help. We will go over selected solutions on the board. You do not need to finish every part; aim for understanding over speed. Keep fractions exact where you can.

Problem 1 (Review of Lecture 10): One Bayes Update

A bag is equally likely to be **Fair** ($P(\text{heads}) = 0.5$) or **Trick** ($P(\text{heads}) = 0.9$). You draw a coin and flip it once; it lands **heads**.

(a) Write Bayes' theorem for $P(\text{Trick} \mid \text{heads})$.

(b) Compute it. (Prior 0.5 each.)

Problem 2: Belief as a Dictionary, and the Update Loop

The Stanford Acuity Test models a patient's vision ability A (on a $[0, 1]$ scale, where 1 is standard vision). We use a discrete belief over three ability values with this **prior**:

a	0.2	0.5	0.9
$P(A = a)$	0.2	0.3	0.5

Model of an observation: when shown a letter, the patient reads it **correctly** with probability equal to their ability a (so **wrong** with probability $1 - a$).

The patient is shown one letter and gets it **wrong** ($Y = 0$).

(a) For each a, compute the *unnormalized* posterior $P(A = a) \cdot P(Y = 0 \mid a) = P(A = a)\,(1 - a)$.

(b) Sum the three unnormalized values to get the normalization constant. What quantity does this sum equal, in Bayes' theorem terms?

(c) Normalize to get the posterior belief $P(A = a \mid Y = 0)$ for each a. Which way did the belief shift, and does that make sense?

Problem 3: Multiple Observations

Continuing Problem 2, the patient is next shown a (larger) letter and gets it **correct** ($Y = 1$), where $P(Y = 1 \mid a) = a$.

(a) Update again, using your Problem 2 posterior as the **new prior** (multiply by the likelihood a and renormalize).

(b) Verify you get the same answer by instead multiplying the *original* prior by **both** likelihoods, $P(A = a)\,(1-a)\,(a)$, and normalizing. Why must these two routes agree?

Problem 4: Why Normalize?

In the update `belief[a] = prior[a] * likelihood(a)` followed by `normalize(belief)`:

(a) In Bayes' theorem, what does the normalization constant (the denominator) represent, and why is it the same for every value of a?

(b) If you only ever needed the **most likely** value of A (the mode), would you have to normalize? Explain.

Problem 5: Discretizing a Continuous Belief

In Bayesian carbon dating, the age A of a sample is continuous, but the code loops over integer ages $100, 101, \ldots, 10000$ and stores a belief in a dictionary.

(a) Explain why representing the belief as a dictionary over discretized ages is reasonable, and what the entries represent.

(b) The update multiplies a uniform prior by a likelihood `calc_likelihood(m, age)` and normalizes. In one sentence, what does `calc_likelihood(m, age)` represent here?

Problem 6: Full Posterior over Time (pset4: mutation_clock)

A strand of mitochondrial DNA has 10,000 base pairs, each mutating at rate 6.67×10^{-6} per year. You observe that exactly 10 base pairs have mutated at least once. With a uniform prior over integer years $T \in \{0, 1, \ldots, 200\}$, you infer the posterior $P(T \mid X = 10)$.

(a) For a fixed age t, what is the probability $p(t)$ that a single base pair has mutated at least once? (Model each base pair's mutation as a Poisson process.)

(b) Write the likelihood $P(X = 10 \mid T = t)$ in terms of $p(t)$ (hint: 10,000 independent base pairs).

(c) Write the expression for the posterior $P(T = 150 \mid X = 10)$. (You would evaluate this in code; a quick check gives about 0.011, and $T = 150$ is the most probable age.)

Challenge Problem (optional): The Baby That Won't Arrive

A due-date model gives a prior PMF over the delivery day D (relative to the due date, day 0), spanning days -50 to 30. Today is day -17 and **no baby yet**.

(a) The likelihood of "no baby yet by day -17" given a true delivery day d is an indicator: it is 0 if $d \leq -17$ and 1 if $d > -17$. Explain why.

(b) Describe how the posterior PMF is obtained from the prior and this likelihood, and why the probability mass on the remaining days must go **up** even though you gained only "negative" information.