CS109 Lecture 11 Worksheet: ANSWER KEY

Inference

Summer 2026

Lecture 11 Worksheet: Answer Key

Instructor / TA copy. Problem-set solutions are hidden in the standard build; run `./convert-to-pdf.sh 11 solutions` for the full instructor/TA edition.

Problem 1 (Review): One Bayes Update

(a)

$$P(\text{Trick} \mid \text{heads}) = \frac{P(\text{heads} \mid \text{Trick})\, P(\text{Trick})}{P(\text{heads} \mid \text{Trick})\, P(\text{Trick}) + P(\text{heads} \mid \text{Fair})\, P(\text{Fair})}.$$

(b)

$$= \frac{0.9(0.5)}{0.9(0.5) + 0.5(0.5)} = \frac{0.45}{0.45 + 0.25} = \frac{0.45}{0.70} \approx 0.643.$$

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

Prior $P(A) = \{0.2 \to 0.2,\ 0.5 \to 0.3,\ 0.9 \to 0.5\}$; likelihood of **wrong** is $1 - a$.

(a) Unnormalized posteriors $P(A = a)(1 - a)$:

$$a = 0.2 :\ 0.2(0.8) = 0.16, \qquad a = 0.5 :\ 0.3(0.5) = 0.15, \qquad a = 0.9 :\ 0.5(0.1) = 0.05.$$

(b) Sum $= 0.16 + 0.15 + 0.05 = 0.36$. This equals $P(Y = 0)$, the total probability of the observation (the normalization constant / denominator of Bayes' theorem).

(c) Normalize:

$P(A = 0.2 \mid Y = 0) = \frac{0.16}{0.36} \approx 0.444, \quad P(A = 0.5 \mid Y = 0) = \frac{0.15}{0.36} \approx 0.417, \quad P(A = 0.9 \mid Y = 0) = \frac{0.05}{0.36} \approx 0.139.$

Belief shifted toward **lower** ability (0.2 rose from 0.20 to 0.44), which makes sense: getting the letter wrong is evidence of worse vision.

Problem 3: Multiple Observations

New prior = Problem 2 posterior $\{0.444, 0.417, 0.139\}$; likelihood of **correct** is a.

(a) Unnormalized: $0.444(0.2) = 0.0889,\ 0.417(0.5) = 0.2083,\ 0.139(0.9) = 0.125$. Sum $= 0.4222$. Posterior:

$$P(A = 0.2) \approx 0.211, \qquad P(A = 0.5) \approx 0.493, \qquad P(A = 0.9) \approx 0.296.$$

(b) Multiplying the original prior by both likelihoods: $0.2(0.8)(0.2) = 0.032$, $0.3(0.5)(0.5) = 0.075$, $0.5(0.1)(0.9) = 0.045$; sum $= 0.152$; normalized gives $\{0.211, 0.493, 0.296\}$, identical. They must agree because multiplying likelihoods (for conditionally independent observations) and normalizing once is the same as updating sequentially, normalizing each time; the intermediate normalization is just an overall constant.

Problem 4: Why Normalize?

(a) The denominator is $P(\text{observation}) = \sum_a P(A = a)P(Y \mid a)$, the total probability of the evidence. It does not depend on the particular value a (it sums a out), so it is a single constant that rescales all the unnormalized values to sum to 1.

(b) No. The normalization constant is the same positive number for every a, so it does not change which value is largest. To find only the mode you can compare the unnormalized products $P(A = a)P(Y \mid a)$ directly.

Problem 5: Discretizing a Continuous Belief

(a) A dictionary maps each candidate (discretized) age to its posterior probability. With enough closely spaced ages, this table approximates the continuous belief arbitrarily well, and it is easy to update and normalize in code. Each entry is $P(A = i \mid \text{observation})$.

(b) `calc_likelihood(m, age)` is $P(M = m \mid A = \text{age})$: the probability of observing m remaining C14 molecules given that the sample is that age (computed from exponential decay).

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

(a) A single base pair mutates as a Poisson process with rate 6.67×10^{-6}/yr, so the probability it has mutated **at least once** by time t is

$$p(t) = 1 - e^{-6.67\times10^{-6}\,t}.$$

(b) With 10,000 independent base pairs, the number that have mutated is $\text{Bin}(10000, p(t))$, so

$$P(X = 10 \mid T = t) = \binom{10000}{10} p(t)^{10}\big(1 - p(t)\big)^{10000-10}.$$

(c) With a uniform prior over $t \in \{0, \dots, 200\}$, the prior cancels in the normalization, leaving

$$P(T = 150 \mid X = 10) = \frac{P(X = 10 \mid T = 150)}{\sum_{t=0}^{200} P(X = 10 \mid T = t)} \approx 0.011.$$

Evaluating in code, the posterior peaks at $T = 150$ (its mode), consistent with $p(150) \approx 0.001$ giving an expected $10000 \cdot 0.001 = 10$ mutations.

Challenge Problem: The Baby That Won't Arrive

(a) The observation is "no baby yet as of day -17." If the true delivery day were $d \le -17$, the baby would already have arrived, so that observation would be impossible (likelihood 0). If $d > -17$, the baby has not arrived yet, exactly matching the observation (likelihood 1). Hence the likelihood is the indicator $\mathbb{1}[d > -17]$.

(b) The posterior is the prior times this indicator, renormalized: all days ≤ -17 get probability 0, and the surviving days (those > -17) keep their prior values but are divided by the smaller total (the remaining probability mass). Because we divide by a number less than 1, every surviving day's probability goes **up**; ruling out early days redistributes their mass onto the days still possible.