

CS109 Lecture 12 Worksheet: ANSWER KEY

General Inference

Summer 2026

Lecture 12 Worksheet: Answer Key

Instructor / TA copy. Problem-set solutions are hidden in the standard build; run `./convert-to-pdf.sh 12 solutions` for the full instructor/TA edition.

Problem 1 (Review): One More Update

$$P(\text{Spam} \mid \text{"free"}) = \frac{0.6(0.3)}{0.6(0.3) + 0.1(0.7)} = \frac{0.18}{0.18 + 0.07} = \frac{0.18}{0.25} = 0.72.$$

Problem 2: The Joint as a Product

- (a) Following the arrows (Flu is the parent of both):

$$P(\text{Flu}, \text{Fever}, \text{Tired}) = P(\text{Flu}) P(\text{Fever} \mid \text{Flu}) P(\text{Tired} \mid \text{Flu}).$$

- (b)

$$\begin{aligned} P(\text{Flu} = 1, \text{Fever} = 1, \text{Tired} = 0) &= P(\text{Flu} = 1) P(\text{Fever} = 1 \mid \text{Flu} = 1) P(\text{Tired} = 0 \mid \text{Flu} = 1) \\ &= 0.2 \times 0.9 \times (1 - 0.8) = 0.2 \times 0.9 \times 0.2 = 0.036. \end{aligned}$$

- (c) A full joint over 3 binary variables has $2^3 = 8$ rows; over n binary variables, 2^n rows. The factorization stores a small conditional table per node (sized by its number of parents) instead of one giant 2^n table, which is exponentially smaller when each node has few parents.

Problem 3: Conditional Independence

- (a) It asserts that **Fever and Tired are conditionally independent given Flu**: once you know the flu status, knowing whether the person is tired tells you nothing more about fever (and vice versa).
- (b) $P(\text{Fever} = 0 \mid \text{Flu} = 1) = 1 - 0.9 = 0.1$.
- (c) They are **not** marginally independent. Both depend on Flu, so observing Tired = 1 raises the chance of Flu, which in turn raises the chance of Fever. Conditional independence given Flu does **not** imply marginal independence. (Formally $P(\text{Fever} = 1 \mid \text{Tired} = 1) \neq P(\text{Fever} = 1)$.)

Problem 4: Sampling from the Joint (Generative Model)

- (a) **Ancestral sampling:** sample parents before children. (1) Sample $\text{Flu} \sim \text{Bern}(0.2)$. (2) Given the drawn Flu value, sample $\text{Fever} \sim \text{Bern}(P(\text{Fever} = 1 \mid \text{Flu}))$. (3) Given the same Flu value, sample $\text{Tired} \sim \text{Bern}(P(\text{Tired} = 1 \mid \text{Flu}))$. Output (Flu, Fever, Tired).
- (b) A program that samples from the joint implicitly assigns the correct probability to every joint outcome (its long-run frequency of producing that outcome equals the joint probability). So specifying the sampler is equivalent to specifying the joint distribution; anything you could compute from the joint you could also estimate by sampling.

Problem 5: Rejection Sampling

- (a) Repeatedly draw full joint samples by ancestral sampling. **Discard** any sample in which $\text{Fever} \neq 1$. Among the **kept** samples (those with $\text{Fever} = 1$), the estimate is the fraction that also have $\text{Flu} = 1$.
- (b) By the frequency definition of probability, the fraction of $\text{Fever} = 1$ samples that have $\text{Flu} = 1$ approaches $P(\text{Flu} = 1 \mid \text{Fever} = 1)$ as the number of kept samples grows; conditioning on the evidence is exactly “restrict to the sub-population where the evidence holds.”

(c)

$$P(\text{Flu} = 1 \mid \text{Fever} = 1) = \frac{P(\text{Fever} = 1 \mid \text{Flu} = 1)P(\text{Flu} = 1)}{P(\text{Fever} = 1)},$$

where $P(\text{Fever} = 1) = 0.9(0.2) + 0.05(0.8) = 0.18 + 0.04 = 0.22$. So

$$P(\text{Flu} = 1 \mid \text{Fever} = 1) = \frac{0.18}{0.22} \approx 0.818.$$

Problem 6: Rejection Sampling on a Bigger Network (pset4: web_md)

- (a) Sample in causal order (parents first): the roots $\text{Tick} \sim \text{Bern}(0.85)$ and $\text{Stress} \sim \text{Bern}(0.95)$; then the diseases Lyme and Flu using probabilities that depend on their parents; then each symptom (fever, rash, tired, cough, runny nose, headache) using its parents’ (Lyme, Flu) values. Each node is Bernoulli with a probability looked up from its conditional table given the already-sampled parent values.
- (b) **Keep** only samples in which $\text{fever} = 1$, $\text{tick} = 1$, and $\text{cough} = 0$; discard all others. Among kept samples, estimate $P(\text{Lyme} = 1 \mid \text{obs})$ as the fraction with $\text{Lyme} = 1$, and $P(\text{Flu} = 1 \mid \text{obs})$ as the fraction with $\text{Flu} = 1$. They need not sum to 1 because Lyme and Flu are separate variables: a person can have both or neither, so the two events are not a partition. Running rejection sampling (100k+ samples) gives roughly $P(\text{Lyme} = 1 \mid \text{obs}) \approx 0.215$ and $P(\text{Flu} = 1 \mid \text{obs}) \approx 0.79$.

Challenge Problem: When Rejection Sampling Breaks

- (a) If the evidence occurs in only 1 of every 10,000 joint samples, then about 9,999 of every 10,000 samples are thrown away, so you must generate enormous numbers of samples to accumulate enough kept ones for a stable estimate. Efficiency collapses as the evidence becomes rarer (and it gets worse with more observed variables).
- (b) MCMC (e.g. Gibbs sampling) never throws samples away: it holds the observed variables **fixed** at their evidence values and repeatedly re-samples the unobserved variables from their conditional distributions, producing a stream of samples already consistent with the evidence. This lets it estimate posteriors even when the evidence would almost never occur under naive forward sampling.