CS109 Lecture 12: LLM Learning Guide

General Inference

Summer 2026

Lecture 12: LLM Learning Guide

Before lecture: read the Lecture 12 slides (general inference: Bayesian networks, generative models, and sampling). Then open your favorite LLM and work through the concepts below **in order**. Each concept has a *Learn* prompt and a *Test me* prompt. Attempt the "Test me" questions yourself before asking the LLM to grade you.

How to get the most out of this:

- Tell the LLM your level: "I'm a student in Stanford's CS109 (probability for computer scientists). Explain at that level."
- After any answer, ask "why?" at least once and ask it to show each step.
- When it states a formula, ask it to define every symbol before continuing.
- Attempt practice yourself first, then say: "Here is my reasoning: _____. Where exactly am I wrong, if anywhere?"

Concept 1: Why we need models (the joint is too big)

Learn: > "Explain why naively storing a joint distribution is intractable: for n variables each taking k values, the joint has k^n entries. Explain how this motivates structured models that factor the joint into smaller pieces."

Test me: > "Give me a number of variables and values-per-variable and make me compute the size of the full joint table, then explain why that is a problem. Check my arithmetic and reasoning."

Concept 2: Bayesian networks and the factorized joint

Learn: > "Explain a Bayesian network: a directed acyclic graph where each node is a random variable and edges point from parents (causes) to children (effects). State the key factorization: the joint equals the product over nodes of $P(\text{node} \mid \text{its parents})$. Work a small example (e.g. Flu causing Fever and Tired) and compute a joint probability of one full assignment."

Test me: > "Give me a small Bayesian network with conditional probability tables and make me (1) write the joint as a product and (2) compute the probability of a specific full assignment. Check my factorization and arithmetic."

Concept 3: Conditional independence

Learn: > “Explain the independence assumption a Bayesian network encodes: each variable is conditionally independent of its non-descendants given its parents. Contrast conditional independence (given a parent) with marginal independence, using an example where two effects of a common cause are dependent marginally but independent given the cause.”

Test me: > “Give me a small network and make me state a conditional independence it encodes, and decide whether two specific variables are marginally independent. Check my reasoning about the common-cause structure.”

Concept 4: Generative models and ancestral sampling

Learn: > “Explain what makes a model ‘generative’: it tells a step-by-step story for producing a sample, sampling each variable after its parents (ancestral sampling). Explain why writing a program that samples from the joint is equivalent to defining the joint. Give pseudocode for sampling a small network.”

Test me: > “Give me a small Bayesian network and make me write the ancestral-sampling procedure (in what order, and each variable’s sampling distribution). Check that I sample parents before children.”

Concept 5: Rejection sampling

Learn: > “Explain rejection sampling for conditional inference: draw many joint samples, discard those inconsistent with the observed evidence, and estimate the query probability as a frequency among the kept samples. Explain why this works (probability as long-run frequency) and work a small example.”

Test me: > “Give me a network, an observation, and a query, and make me describe the reject/keep rule and how to form the estimate. Check that I condition on exactly the evidence and estimate the right fraction.”

Concept 6: Limits of rejection sampling and the idea of MCMC

Learn: > “Explain why rejection sampling becomes inefficient when the evidence is rare (most samples get discarded), and introduce MCMC at a conceptual level: it holds the observed variables fixed and repeatedly re-samples the unobserved ones, so no samples are wasted. Name that this is how large probabilistic models are queried in practice.”

Test me: > “Ask me what happens to rejection sampling as the observed evidence gets rarer, and how MCMC helps. Check that I connect rare evidence to wasted samples and describe MCMC as sampling with evidence fixed.”

Wrap-up prompt (after all six concepts)

> “Give me one multi-part CS109-style problem with a small Bayesian network where I must (1) write the joint as a product of conditionals, (2) compute a specific joint and a conditional probability, (3) describe ancestral sampling, and (4) describe rejection sampling for a query and explain when it becomes inefficient. Let me solve every part, then grade each part, tell me which concept it tested, and tell me the single concept I should review most.”

Note: this completes the inference arc of CS109. From here the course moves toward learning model parameters from data (maximum likelihood estimation and machine learning).

Bring any sticking points to lecture; the TAs and instructor will help you work through them on the worksheet.