CS109 Lecture 14: LLM Learning Guide

Beta: The Random Variable for Probabilities

Summer 2026

Lecture 14: LLM Learning Guide

Before lecture: read the Lecture 14 slides (Beta: the random variable for probabilities). 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: A probability can itself be a random variable

Learn: > "Explain the idea that an unknown probability (like a coin's chance of heads, or a drug's chance of working) can be treated as a random variable X with values in $[0, 1]$. Contrast a point estimate ($p = 0.75$) with a full belief distribution over p, and explain why representing uncertainty matters for decisions (e.g. a drug tested 4 times vs. 10,000 times can have the same point estimate but very different uncertainty)."

Test me: > "Give me two scenarios with the same observed success rate but very different sample sizes, and ask me why a point estimate treats them identically and what a belief distribution captures that the point estimate misses. Push back on vague answers."

Concept 2: Bayesian updating with a continuous parameter

Learn: > "Walk me through Bayes' theorem when the unknown is continuous: posterior density $f(x \mid \text{data}) \propto P(\text{data} \mid X = x)\, f(x)$, where the denominator is a normalization constant. Explain the roles of prior, likelihood, and posterior, and why we often work 'up to proportionality' and normalize at the end."

Test me: > "Give me a prior and a simple likelihood and ask me to identify: the prior, the likelihood, what the posterior is proportional to, and what normalization means. Check that I never treat the likelihood as a density over x that must integrate to 1."

Concept 3: Uniform prior + coin flips = Beta

Learn: > "Derive the posterior for a coin's heads probability X starting from $X \sim \text{Uni}(0,1)$, after observing h heads and t tails: the likelihood is binomial, so the posterior density is proportional to $x^h(1-x)^t$, which is $\text{Beta}(h+1, t+1)$. Emphasize the '+1' pattern: $a = \text{successes} + 1$, $b = \text{failures} + 1$, and that $\text{Beta}(1,1) = \text{Uni}(0,1)$."

Test me: > "Give me a handful of scenarios (successes and failures observed, uniform prior) and make me state the posterior Beta parameters, and also the parameters for a case with zero observations. Check my '+1's."

Concept 4: The Beta distribution itself

Learn: > "Describe the $\text{Beta}(a,b)$ random variable: support $[0,1]$, PDF proportional to $x^{a-1}(1-x)^{b-1}$, mean $\frac{a}{a+b}$, mode $\frac{a-1}{a+b-2}$ (for $a, b > 1$), and variance $\frac{ab}{(a+b)^2(a+b+1)}$. Explain how the shape changes with a and b: symmetric when $a = b$, skewed otherwise, and more concentrated as $a + b$ grows."

Test me: > "Give me a few Beta distributions and make me compute mean, mode, and variance, and describe each shape (skew and spread) without plotting. Then ask me which of two Betas represents more certainty and why."

Concept 5: Conjugacy and Laplace smoothing

Learn: > "Explain conjugacy: if the prior on a coin's probability is $\text{Beta}(a,b)$ and we observe h heads and t tails, the posterior is $\text{Beta}(a+h, b+t)$ — same family, just updated parameters. Interpret $a-1$ and $b-1$ as imagined prior successes/failures. Then explain Laplace smoothing as using a $\text{Beta}(2,2)$ prior, giving the estimate $\frac{s+1}{n+2}$, and why that avoids estimating probability 0 or 1 from small data."

Test me: > "Give me a prior Beta and new observations, and make me write the posterior instantly (no integration). Then give me a case with 0 successes out of 3 trials and ask what the unsmoothed and Laplace-smoothed estimates are and why the smoothed one is safer."

Concept 6: Decisions with Betas: CDFs and Thompson sampling

Learn: > "Explain how to answer questions like $P(X > 0.6)$ for $X \sim \text{Beta}(a,b)$ using the Beta CDF (e.g. `stats.beta.cdf` in scipy), since the integral has no clean closed form. Then explain Thompson sampling for choosing between two options (e.g. two drugs): keep a Beta belief per option, sample one value from each Beta, pick the option with the larger sample, observe, update. Explain how this balances exploration and exploitation."

Test me: > "Give me beliefs for two options as Beta distributions (one well-tested, one barely tested) and ask me (1) how to compute the probability each option's success rate exceeds a threshold, and (2) to walk through one round of Thompson sampling and explain why the uncertain option sometimes gets picked. Check my reasoning about exploration."

Wrap-up prompt (after all six concepts)

"Give me one multi-part CS109-style problem in which I must (1) turn a uniform prior plus observed successes/failures into a Beta posterior, (2) compute its mean, mode, and variance, (3) update it again with more data using conjugacy, and (4) describe how to use the posterior for a decision (a CDF probability and one round of Thompson sampling). 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: Beta is the first distribution in the course whose job is to represent belief about a parameter rather than data. This "distribution over parameters" idea returns when the course covers 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.