CS109 Lecture 13: LLM Learning Guide

Multinomial

Summer 2026

Lecture 13: LLM Learning Guide

Before lecture: read the Lecture 13 slides (the multinomial coefficient, multinomial random variables, and probabilistic text analysis). 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: The multinomial coefficient (counting with repeated types)

Learn: > "Explain the multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\, n_2! \cdots n_r!}$ as the number of distinct orderings of n objects when there are r types with n_i identical objects of each type. Show how it generalizes the binomial coefficient (which is the special case $r = 2$), and work an example like counting orderings of a string with repeated letters."

Test me: > "Give me a short string with repeated letters and make me count its distinct orderings with the multinomial coefficient. Then ask me why we divide by each $n_i!$. Check my arithmetic and my explanation."

Concept 2: The multinomial random variable and its joint PMF

Learn: > "Explain the multinomial random variable: n independent trials, each landing in one of r categories with probabilities $p_1, \ldots, p_r$ summing to 1, and X_i counting the trials in category i. Derive the joint PMF $P(X_1 = c_1, \ldots, X_r = c_r) = \binom{n}{c_1, \ldots, c_r} p_1^{c_1} \cdots p_r^{c_r}$ the same way the binomial PMF is derived: probability of one specific ordering, times the number of orderings. State the constraint on the c_i."

Test me: > "Give me a problem with a die (fair or weighted) rolled a handful of times and specific target counts for each face, and make me compute the probability. Check that I use both the multinomial coefficient and the right powers of the probabilities."

Concept 3: When does a multinomial apply?

Learn: > "Explain the assumptions behind the multinomial model: a fixed number of trials, each trial independent, each trial taking exactly one of r outcomes with the same probabilities every trial. Give two real examples that fit (e.g. routing requests to servers, drawing loot-box rarities) and one example that does not fit (e.g. drawing cards without replacement) and explain why."

Test me: > "Describe three scenarios and make me decide for each whether a multinomial model applies, and if not, which assumption fails. Check my reasoning."

Concept 4: Marginal counts are binomial

Learn: > "Explain why each single count X_i in a multinomial is marginally distributed as $\mathrm{Bin}(n, p_i)$: collapse the categories into 'type i' versus 'everything else.' Also explain intuitively why the counts are *not* independent of each other (they must sum to n)."

Test me: > "Give me a multinomial setup and ask me for the distribution and expectation of one single count, and then whether two of the counts are independent. Check my justification, not just my answers."

Concept 5: Text as a multinomial (bag of words) and authorship

Learn: > "Explain how a document can be modeled as a multinomial over words (the 'bag of words' model): the document is n word draws, and each author or class (e.g. spam vs. not spam) has its own word probabilities. Then show how Bayes' theorem uses $P(\text{document} \mid \text{author})$ and a prior to compute $P(\text{author} \mid \text{document})$, as in the Federalist Papers authorship analysis. Explain why the multinomial coefficient cancels when comparing two authors."

Test me: > "Give me a tiny vocabulary (3 words), word probabilities for two authors, and word counts for a mystery document, and make me set up the posterior comparison between the authors. Check my setup before I compute anything."

Concept 6: Log probabilities

Learn: > "Explain why multiplying thousands of small probabilities underflows to zero in floating point, and why we compute with log probabilities instead. State the identities $\log(ab) = \log a + \log b$ and $\log(a^n) = n \log a$, explain why log being monotonically increasing means comparing log-scores is the same as comparing probabilities, and explain what it means when a log probability is a large negative number."

Test me: > "Give me two products of several small probabilities and make me compare them using logs instead of multiplying. Then ask me: if $\ln P(D \mid A) = -1200$ and $\ln P(D \mid B) = -1190$, which is more likely and by what factor? Check my use of the log identities."

Wrap-up prompt (after all six concepts)

> "Give me one multi-part CS109-style problem in which I must (1) count orderings with a multinomial coefficient, (2) compute a multinomial probability with unequal category probabilities, (3) find the marginal distribution of one count, and (4) set up a Bayesian authorship comparison using log probabilities. 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: the bag-of-words multinomial model is the engine behind Naive Bayes classifiers — the same idea you will see again when the course turns to machine learning.

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