CS109 Lecture 16 Worksheet: Bootstrapping

Samples vs. Populations, Sample Variance, the Bootstrap, and p-values

Summer 2026

Lecture 16 Worksheet: Bootstrapping

Topics: populations vs. samples, the unbiased sample variance, standard error of the mean, the bootstrap algorithm (resampling with replacement), estimating the distribution of *any* statistic, and p-values via the null hypothesis.

How to use this sheet. Work in small groups. A TA and the instructor will circulate to help. We will go over selected solutions on the board. You do not need to finish every part; aim for understanding over speed. Keep fractions exact where you can.

Problem 1 (Review of Lecture 15): Checkout Totals

An online store logs the number of items in each of 100 independent checkouts. Each checkout's item count has mean 3 and variance 4 (the counts are integer-valued and i.i.d., but *not* from any named distribution).

(a) What does the CLT say about the distribution of S, the total number of items across all 100 checkouts?

(b) Approximate $P(S \geq 320)$, using a continuity correction.

Problem 2: Ping Times

You measure the round-trip latency (in ms) of 8 pings to a server:

$$[12,\ 15,\ 11,\ 14,\ 13,\ 19,\ 12,\ 16]$$

(a) Compute the sample mean $\bar{X}$.

(b) Compute the (unbiased) sample variance S^2. Why do we divide by $n-1$ instead of n? Answer in one sentence, in terms of what plugging in $\bar{X}$ (instead of the true μ) does to the sum of squared deviations.

(c) Compute the standard error of the mean. In one sentence: what is this number an estimate of?

(d) Using the CLT, give an interval of the form $\bar{X} \pm 2 \cdot \mathrm{SE}$ and interpret it.

Problem 3: A Bootstrap You Can Enumerate

You have a (tiny!) sample of 3 file transfer times: $[2,\ 4,\ 9]$ seconds. A bootstrap resample draws 3 values from this sample *with replacement*, each draw uniform over the sample.

(a) How many equally likely ordered resamples are there?

(b) What is the probability that a resample's *maximum* equals 9?

(c) What is the probability that a resample's *mean* equals the original sample mean of 5?

(d) Your partner suggests using `replace = False` "to avoid duplicates." In one or two sentences, explain what goes wrong with the bootstrap if you sample n items from n without replacement.

Problem 4: Bootstrapping the Median

You have 50 user ratings of your app (integers 1–5). You report the sample *median*, and you want error bars. There is no CLT-style formula for the median.

(a) Write pseudocode (4–6 lines) that uses bootstrapping to estimate the standard deviation of the sample median.

(b) The brute-force version of this algorithm requires drawing fresh samples from the *true* rating distribution, which you do not have. What assumption does the bootstrap make so that it can proceed anyway?

(c) Give one situation from lecture where this assumption is dangerous.

Problem 5: Is the New Compiler Flag Faster?

You benchmark a program 40 times with compiler flag A (sample mean runtime 210 ms) and 45 times with flag B (sample mean 204 ms). The observed difference in means is 6 ms. You want to know if flag B is *really* faster or if this is sampling noise.

(a) State the null hypothesis for this experiment.

(b) Write pseudocode (5–8 lines) that uses bootstrapping to estimate the p-value of the observed difference. (Hint: under the null hypothesis, which pool should resamples be drawn from?)

(c) Suppose your procedure returns $p = 0.03$. Interpret this number in one sentence. Would your conclusion change if it returned $p = 0.4$?

Problem 6: Estimating Course Size (problem from the Winter 2025 Final)

In order to hire the correct number of TAs, Stanford needs to estimate final enrollment in CS109 two weeks before the start of the quarter. Two weeks before the start of the quarter, 300 students are enrolled. For the last 10 offerings of CS109, Stanford has recorded the ratio

$$r_i = \frac{\text{final enrollment for offering } i}{\text{enrollment two weeks before start for offering } i},$$

which you can access as a list of values $[r_1, r_2, \ldots, r_{10}]$. Assume each r_i is an i.i.d. sample from the true ratio random variable R. The number of students who end up taking the class this quarter will be $T = 300 \cdot R$. From historical analysis we know that both R and T are normally distributed.

(a) Estimate $E[T]$, the expected class size. Provide your answer as a math expression.

(b) Estimate $\text{Var}(T)$. Provide your answer as a math expression.

(c) Use *bootstrapping* to estimate the *standard deviation* of your estimate of $\text{Var}(T)$ from part (b) (i.e., the sampling variability due to having only 10 historical observations of R).

Challenge Problem (optional): Two Schools of Thought

You flip a coin of unknown probability p of heads 15 times and observe 5 heads, 10 tails.

(a) **Bayesian:** starting from the Laplace prior $X \sim \text{Beta}(2, 2)$, what is your posterior belief about p?

(b) **Frequentist:** describe how to use bootstrapping to build a distribution for $\hat{p}$ *without* using a prior.

(c) In one or two sentences: what is philosophically different about what the two resulting distributions represent?