CSCI 145 Problem 25: Audit Batch Centering

August 27, 2026

Submission Instructions

Please upload your work to Gradescope by **11:59pm on DATE**.

- Turn in a single PDF of *handwritten* solutions. Typed (including LaTeX) solutions will not be accepted.
- Problems are grouped into three-week **cycles**. Once per cycle, you will meet with the grader and present the solution to *one* problem from that cycle, selected at random when you arrive; bring your work and be ready to explain any of them.

Problem 25: Audit Batch Centering 2 points

Let $(G_i, \mathbf{u}_i)$ be N independent trajectory samples, where $\mathbf{u}_i = \nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\tau_i)$, $\mathbb{E}[\mathbf{u}_i] = \mathbf{0}$, and the desired policy gradient is $\mathbf{g} = \mathbb{E}[G_i \mathbf{u}_i]$. An implementation uses the same batch to form $\bar{G} = N^{-1} \sum_j G_j$ and reports

$$\hat{\mathbf{g}}_{\text{center}} = \frac{1}{N} \sum_{i=1}^{N} (G_i - \bar{G}) \mathbf{u}_i.$$

The code comment says "subtracting any action-independent baseline is unbiased." Audit that comment. Compute $\mathbb{E}[\hat{\mathbf{g}}_{\text{center}}]$ exactly, including the case $N = 1$, and explain why the usual baseline proof does not apply without modification.

For $N \geq 2$, repair the estimator without collecting more trajectories. Your repair must use, for each sample, only returns from the other samples to construct its baseline. Prove unbiasedness and find an exact algebraic relation between your estimator and $\hat{\mathbf{g}}_{\text{center}}$.

Optional starred extension ($\star$). This extension is not required for full credit. Translate the lesson to a learned state-dependent critic. Give a concrete data-leakage or action-dependence condition that invalidates the standard unbiasedness argument, then repair it by cross-fitting: split the trajectories into folds and evaluate each trajectory using a critic trained only on other folds.

Results you may use. You may use independence across trajectories and the score identity $\mathbb{E}[\mathbf{u}_i] = \mathbf{0}$.