

Midterm practice

This practice material is expected to take longer than 75 minutes. If one problem is incomplete after 75 minutes you're probably fine for pacing. If 2+ problems are unfinished, you may need to drill practice problems for speed so you have time to check answers during the midterm.

1 True/false

If false, explain why. For people simulating a midterm, don't spend too long on this section at the expense of working on other practice. If you're spending longer than 2 minutes on any T/F question, skip it and look up the solution later.

1. Asymptotically, the MLE is unbiased.
2. Asymptotically, the MLE has the lowest MSE of any estimator.
3. If a sequence converges in distribution to a constant, then it also converges in probability to that constant.
4. Given a sequence T_1, T_2, \dots of estimators for the estimand θ such that $\mathbb{E}[T_i] = \theta$ for any i , then the estimator is consistent, i.e., $T_n \xrightarrow{p} \theta$.
5. By the bias-variance tradeoff, for any given estimator, an alternate estimator with lower bias will have higher variance.
6. The MoM estimator is guaranteed to be unbiased.
7. If \mathbf{T} is a sufficient statistic for the data, then we can find the MLE using only \mathbf{T} without observing the entirety of the data.

8. A given likelihood function will have a unique sufficient statistic.
9. Removing multiplicative constants from log-likelihood calculation before finding the MLE will result in a change of the point estimate of the MLE.
10. For n observations, the Fisher information of the data is always n times the Fisher information of a single observation.
11. The MLE of the parameter θ for a Uniform distribution from $0, \theta$ is asymptotically Normal with variance given as the inverse of the Fisher info of one observation.
12. If X, Y are independent, then $g(X), h(Y)$ are also independent for functions g, h .

2 Exponential modeling

Emily sets up a sensor to detect radioactive decay from a sample. Assume that the times between radioactive particle detection can be modeled as Exponential with rate parameter λ , with units in minutes. Data is collected in the form y_1, y_2, \dots, y_n , where y_i is the interarrival time between the i th and $i - 1$ th particle (where $y_0 = 0$).

2.1 MLE

2.1.1 Likelihood function Write out the likelihood, log-likelihood, and solve for the MLE.

2.1.2 NEF Verify the above using properties of NEFs.

2.2 Broken sensor

When setting up the sensor, Emily finds that it automatically shuts down every hour. Emily can't figure out how to stop this, but can jerry-rig a setup to restart the sensor when it turns off. Emily can record the number of times the sensor restarts but not the times when it happens.

Rewrite the likelihood and solve for the MLE.

3 Sleep average

This is a pretty contrived setup, so ignore how this would be a questionable experiment IRL.

Sia is researching whether the average sleep time of Harvard students differs from the 2013 reported national average of 6.8 hours (Gallup). She (somehow) gets the average sleep data for n students and subtracts 6.8 from the collected data (to center the data around zero if average of Harvard students and the national average line up).

Sia doesn't know the standard deviation, but she can estimate it with the data (assume n is large enough).

3.1 Single CI

Let the sample mean of the data (post-subtraction) be the point estimate for the estimand. Use the appropriate pivot to construct a $100(1 - \alpha)$ percent confidence interval.

3.2 Multiple CIs

Sia wants to also look at whether certain concentrations deviate from the national average. So, she sorts her data into m concentrations. For this problem, assume that the concentrations are independent.

3.2.1 Independence assumptions In what ways should we expect the independence assumption to be violated?

3.2.2 Coverage probability Each CI collected on a concentration is constructed to have 95 percent coverage probability. Is it true that the *entire experiment* has the same coverage probability? That is, can we say “If we repeated this experiment, we would expect that *all of our CIs* would cover the true value of their respective estimands $100(1 - \alpha)$ percent of the time.”

If false, evaluate the coverage probability of the entire experiment.

3.2.3 Adjusted CIs Spoiler: 3.2.2 is false. Propose an adjusted CI for each of the m concentrations such that the entire experiment has $100(1 - \alpha)$ percent coverage.

4 Uniform CIs

Let $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Unif}(0, \theta)$.

4.1 Properties of the MLE

4.1.1 Regularity conditions (This is a review of material covered in homework and lecture examples). Solve for the MLE. Identify any violations of regularity conditions.

4.1.2 Bias of the MLE Solve for the bias given n . Is it possible to construct an unbiased estimator based on your solution? Also, is the MLE asymptotically unbiased?

4.2 Confidence intervals

Let $Y_{(n)}$ be the n th order statistic (i.e., the maximum of observations). Consider the two interval estimators:

1. $[aY_{(n)}, bY_{(n)}]$, where $1 \leq a \leq b$
2. $[c + Y_{(n)}, d + Y_{(n)}]$, where $0 \leq c \leq d$

In which of these cases does the coverage probability depend on θ ?

5 Poisson fun times

Let $Y_1, \dots, Y_n \stackrel{i.i.d.}{\sim} \text{Pois}(\lambda)$.

5.1 Poisson NEF

Show that the Poisson is a member of the NEF and use properties of the NEF to determine the MLE and its asymptotic distribution

5.2 Confidence interval using MLE

Using the results from (5.1), construct an asymptotic $100(1 - \alpha)$ percent confidence interval.

5.3 Confidence interval

Replace the estimand in (5.2) with an appropriate statistic to construct an alternate $100(1 - \alpha)$ percent CI.