

Midterm practice (Solutions)

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.

1. TRUE
2. However, this does not mean the MLE is unbiased for any given number of observations.

2. Asymptotically, the MLE has the lowest MSE of any estimator.

1. FALSE
2. The MLE has the lowest MSE of any unbiased estimator. A biased estimator may have a sufficiently lower variance and thus lower MSE.

3. If a sequence converges in distribution to a constant, then it also converges in probability to that constant.

1. TRUE

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., $T_n \xrightarrow{p} \theta$.

1. FALSE
2. A counter example is taking only one observation. For example, if we observe i.i.d. X_1, \dots, X_n and we use X_1 to estimate $\mathbb{E}[X_1] = \mu$, then this estimator is unbiased but does not converge to a constant.

5. By the bias-variance tradeoff, for any given estimator, an alternate estimator with lower bias will have higher variance.

1. FALSE
2. This is not implied by bias-variance tradeoff. Use the same setup as (4). We can consider the (very bad) estimator $X_1 + 1$ and the (not as bad) estimator X_1 . The latter will have lower bias and the same variance.
3. The bias-variance tradeoff is not a theorem or quantitative result.
4. Though it relates to the MSE decomposition of MSE being bias squared plus variance, this decomposition does not imply any necessary relationship between bias and variance for any given estimator.

6. The MoM estimator is guaranteed to be unbiased.

1. FALSE
2. Estimating a moment by the sample moment is guaranteed to be unbiased (check this with linearity).
3. However, the MoM estimator is usually composed of functions of these estimated moments, which can introduce bias (through Jensen's inequality, for example).

7. If T is a sufficient statistic for the data, then we can find the MLE using only T without observing the entirety of the data.

1. TRUE
 2. This is Theorem 6.2.8, Corollary 1 in the lecture notes.
8. **A given likelihood function will have a unique sufficient statistic.**
1. FALSE
 2. A valid sufficient statistic is scaling another 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.**
1. FALSE
 2. It will result in a change of the shape of the log-likelihood function.
 3. This will result in problems with calculating Fisher info (and therefore variance) of the MLE.
10. **For n observations, the Fisher information of the data is always n times the Fisher information of a single observation.**
1. FALSE
 2. This only holds if the observations are independent.
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.**
1. FALSE
 2. The Uniform distribution violates regularity conditions (see 4.1.1) so the asymptotic distribution of the MLE does not have the form given here.
12. **If X, Y are independent, then $g(X), h(Y)$ are also independent for functions g, h .**
1. TRUE
 2. This is a property of independence. If you want the full proof, take STAT210

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.

We have that the PDF of an Exponential distribution is $f(y) = \lambda e^{-\lambda y}$. The likelihood, log-likelihood, and score are given as

$$L(\mathbf{y}; \lambda) = \prod_{i=1}^n \lambda e^{-\lambda y_i} = \lambda^n e^{-\lambda \sum_{i=1}^n y_i}$$

$$\ell(\mathbf{y}; \lambda) = n \log \lambda - \lambda \sum_{i=1}^n y_i$$

$$s(\mathbf{y}; \lambda) = \frac{n}{\lambda} - \sum_{i=1}^n y_i,$$

where setting the score equal to zero gives us $\hat{\lambda} = \frac{n}{\sum_{i=1}^n y_i}$.

2.1.2 NEF Verify the above using properties of NEFs.

On the midterm, you can work directly from the knowledge that the Binomial is an NEF. Here, we'll prove it by factorization, where

$$f(y) = \lambda e^{-\lambda y} = e^{\theta y + \log(-\theta)},$$

where $\theta = -\lambda$, $\psi(\theta) = -\log(-\theta)$, $h(y) = 1$.

So, the MLE of $\mu = E[Y_1] = \bar{Y}$ and by invariance, we have that $\hat{\lambda} = 1/\mu = 1/\bar{Y}$, which is identical to the above.

2.2 Broken sensor

When setting up the sensor, Emily finds that it automatically shuts down every hour if it observes no data. Emily can't figure out how to stop this, but can jerry-rig a setup to restart the sensor when it turns off.

Assume the time to restart is negligible.

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.

We can solve this based on memorylessness of the Exponential. We have that for an r.v. X distributed Exponential,

$$P(X > a + b | X \geq b) = P(X \geq a).$$

Therefore, for an observation where we have the sensor time out at 60 minutes and make an observation of a after the restart,

$$P(X \geq a + 60) = P(X \geq a + 60 | X \geq 60)P(X \geq 60) = P(X \geq a)P(X \geq 60).$$

The CDF of an Exponential is $1 - e^{-\lambda x}$, so we can find $P(X \geq 60) = 1 - F(60) = e^{-60\lambda}$.

Because observations are independent, knowing when the sensor times out isn't important. Let m be the number of times we count a restart. We have

$$\begin{aligned} L(\mathbf{y}; \lambda) &= \lambda^n e^{-\lambda \sum_{i=1}^n y_i} (e^{-60\lambda})^m = \lambda^n e^{-\lambda(60m + \sum_{i=1}^n y_i)} \\ \ell(\mathbf{y}; \lambda) &= n \log \lambda - \lambda \left(60m + \sum_{i=1}^n y_i \right) \\ s(\mathbf{y}; \lambda) &= \frac{n}{\lambda} - \left(\sum_{i=1}^n y_i + 60m \right), \end{aligned}$$

and then we can find $\hat{\lambda} = \frac{n}{\sum_{i=1}^n y_i + 60m}$.

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.

Because Sia is estimating the standard deviation using the data, we can use the t distribution as our pivot and perform the Student t -test. We go over the Student t distribution in Lecture 09 and Example 5.4.2 in the Stat 111 draft textbook.

Let X_1, \dots, X_n be the collected data and let μ . We have that

$$\frac{\bar{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}.$$

Let Q be the quantile function of the t_{n-1} distribution. We can then write

$$C(\mathbf{X}) = [\bar{X} - Q(1 - \alpha/2)\hat{\sigma}/\sqrt{n}, \bar{X} + Q(1 - \alpha/2)\hat{\sigma}/\sqrt{n}].$$

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?

Answers here will vary, so if a reasonable example is difficult for you to think of (or if you're not sure if what you're considering is reasonable), then you should check with a TF.

If concentrations share classes and certain classes are known to cause sleep deprivation, this will lead to dependence in the data.

Note that the question is not asking for whether the concentrations are identically distributed. So, for example, noting that certain concentrations have very different workloads is not on its own an indication of dependence, only difference in distribution.

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.

Let X_i be the event that the CI for the i th concentration covers the estimand for that concentration and let X be the event that the entire experiment covers all the sub-estimands. We have

$$\Pr(X) = \prod_{i=1}^n \Pr(X_i) = \Pr(X_i)^n = 100(1 - \alpha)^n \%.$$

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.

Algebra time! Let $1 - \alpha_0$ be the coverage probability of a sub-experiment. We can solve for

$$\begin{aligned} 1 - \alpha &= \Pr(X_i)^n \\ (1 - \alpha)^{1/n} &= 1 - \alpha_0 \\ \alpha_0 &= 1 - (1 - \alpha)^{1/n}. \end{aligned}$$

So our confidence intervals should be constructed from $\hat{\mu} \pm Q(1 - \alpha_0/2)$, with α_0 as defined above.

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.**

This is the continuous variation of the German tank problem in Lecture 08. The likelihood function is given as

$$L(\mathbf{y}; \theta) = \prod_{i=1}^n \mathbb{I}(y_i \leq \theta) \frac{1}{\theta},$$

which, by observation, gives us that the MLE is $Y_{(n)}$, or the largest value of the observations.

This violates the regularity conditions of the support not depending on the estimand and the estimate not being on the edge of the parameter space.

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?

By properties of the Uniform order statistic, $Y_{(n)} \sim \theta \text{Beta}(n, 1)$. So, the expectation is $\theta n/(n + 1)$, which gives us a bias of $\theta/n + 1$.

The MLE is asymptotically unbiased as the expectation goes to θ as n becomes large.

It is possible to construct an unbiased estimator because there is an easy correction of multiplying the estimator by $(n + 1)/n$. So an unbiased estimator would be $Y_{(n)}(n + 1)/n$.

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 ?

Only in case (2).

For case 1, we can write

$$\begin{aligned}\Pr(aY_{(n)} \leq \theta \leq bY_{(n)}) &= \Pr(a \leq \theta/Y_{(n)} \leq b) \\ &= \Pr(1/a \geq Y_{(n)}/\theta \geq 1/b),\end{aligned}$$

which can be easily evaluated because of the known distribution of $Y_{(n)}/\theta \sim \text{Beta}(n, 1)$.

If we attempt to perform the same procedure on case (2), we have

$$\begin{aligned}\Pr(c + Y_{(n)} \leq \theta \leq d + Y_{(n)}) &= \Pr(c \leq \theta - Y_{(n)} \leq d) \\ &= \Pr(c \leq \theta(1 - X) \leq d) : X \sim \text{Beta}(n, 1),\end{aligned}$$

where the distribution of $\theta(1 - X)$ is not solvable without θ .

For example, fix $c = 0$ and some d . For $\theta < d$, the coverage is 100 percent. For $\theta \gg d$, the coverage is less than 100 percent as our highest order statistic may still be more than d away from 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.

We can factor the Poisson as

$$f(y) = \frac{1}{y!} e^{-\lambda} \lambda^y = \frac{1}{y!} e^{\theta y - e^\theta}$$

where $\theta = \log(\lambda)$, $h(y) = 1/y!$, $\psi(\theta) = e^\theta$.

We have $\psi'(\theta) = e^\theta$, $\psi''(\theta) = e^\theta$.

The MLE of λ is $\hat{\lambda} = \bar{Y}$. The Fisher info per observation is $e^{\ln \lambda} = \lambda$.

The asymptotic distribution is

$$\sqrt{n}(\bar{Y} - \lambda) \xrightarrow{d} \mathcal{N}(0, \lambda).$$

5.2 Confidence interval using MLE

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

Our asymptotic pivot is $\frac{\sqrt{n}(\bar{Y} - \lambda)}{\sqrt{\lambda}}$, which is asymptotically standard Normal. Let Q be the quantile function for the standard normal. Let $c = Q(1 - \alpha/2)$. We want

$$\begin{aligned}C(\mathbf{Y}) &= \left\{ \lambda : \left| \frac{\sqrt{n}(\bar{Y} - \lambda)}{\sqrt{\lambda}} \right| \leq c \right\} \\ &= \{ \lambda : n\lambda^2 - (2n\bar{Y} + c^2)\lambda + n\bar{Y}^2 \leq 0 \},\end{aligned}$$

where we can find the zeroes of the LHS using the quadratic formula

$$\frac{2n\bar{Y} + c^2 \pm \sqrt{4n\bar{Y}c^2 + c^4}}{2n}.$$

5.3 Confidence interval

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

Instead of λ in the pivot above, we can use \bar{Y} , which is consistent for λ . We can follow a similar procedure as above, finding

$$\begin{aligned} C(\mathbf{Y}) &= \left\{ \lambda : \left| \frac{\sqrt{n}(\bar{Y} - \lambda)}{\sqrt{\bar{Y}}} \right| \leq c \right\} \\ &= \left\{ \lambda : \bar{Y} - \sqrt{\bar{Y}/nc} \leq \lambda \leq \bar{Y} + \sqrt{\bar{Y}/nc} \right\}. \end{aligned}$$