

# STA 506 - Midterm Examination

Spring 2026

Due: Sunday, 03/08 at 11:59 PM

## Contents

<b>1</b>	<b>Midterm Exam Objectives</b>	<b>1</b>
<b>2</b>	<b>Policies of Using AI Tools</b>	<b>1</b>
<b>3</b>	<b>Policies of the Midterm Exam</b>	<b>2</b>
<b>4</b>	<b>Submission Preparation</b>	<b>2</b>
<b>5</b>	<b>Exam Questions</b>	<b>2</b>
5.1	<b>Log-normal Distribution</b> . . . . .	2
5.2	<b>Question 1:</b> Method of Moments Estimation (MME) of $\mu$ and $\sigma^2$ . . . . .	3
5.3	<b>Question 2:</b> Finding the MLE of $\mu$ and $\sigma^2$ . . . . .	4
5.4	<b>Question 3:</b> Sampling Distribution of Lognormal <b>Sample Mean</b> $\widehat{\mu}_{LN}$ . . . . .	4
5.5	<b>Question 4:</b> Extra Bonus Credit . . . . .	5

## 1 Midterm Exam Objectives

- Understand the definition and relationship between PDFs and CDFs, including their non-parametric estimators: the empirical distribution function and kernel density estimation (KDE).
- Estimate sampling distributions using simulation-based methods, specifically the bootstrap.
- Derive point estimates of parameters using the method of moments and maximum likelihood estimation (MLE).
- Describe the asymptotic (normal) and bootstrap sampling distributions of maximum likelihood estimators.
- Apply all the above inferential procedures in a programming environment to perform numerical data analysis.

## 2 Policies of Using AI Tools

- **Policy on AI Tool Use:** Students must adhere to the AI tool policy specified in the course syllabus. The direct copying of AI-generated content is strictly prohibited. All submitted work must reflect your

own understanding; where external tools are consulted, content must be thoroughly rephrased and synthesized in your own words.

- **Code Inclusion Requirement:** Any code included in your essay must be properly commented to explain the purpose and/or expected output of key code lines. Submitting AI-generated code without meaningful, student-added comments will not be accepted.

### 3 Policies of the Midterm Exam

- This exam must be completed independently. Collaboration with others is not permitted.
- You may consult class notes and online resources. However, please ensure that your use of AI tools complies with the AI policy established for this course.
- All work must be submitted by the deadline. Submissions (including partial work) received after the deadline will not be accepted.

### 4 Submission Preparation

Since this exam involves derivations, you can use the following approaches to prepare your submission and save time:

1. Handwrite the derivation on paper and take a photo. From there, you can either use AI tools to convert it into LaTeX code and paste it into RMarkdown, or insert the image directly into your RMarkdown document, following the example shown in class this week.
2. Prepare the formulas using other non-programmatic editors and take a screenshot to embed the image in your RMarkdown document using `knitr::include_graphics`.
3. Compile all handwritten derivations into a single file and submit it as an attachment.

Please follow the same submission procedure as used for the weekly assignments.

### 5 Exam Questions

**Caution:** *Please follow the suggested expressions and guided steps to complete the exam. Other approaches, such as transformations that trivialize the problems, will not meet the exam objectives.*

#### 5.1 Log-normal Distribution

The log-normal distributions have been widely used in many different fields such as finance and economics, engineering and reliability, environmental science, medical and biology etc.

If  $X$  follows a lognormal distribution:

$$f(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0$$

where  $\mu \in \mathbb{R}$  is the mean of  $\ln X$  and  $\sigma^2 > 0$  is the variance of  $\ln X$ .

**Caution:**  $\mu$  and  $\sigma^2$  are not population mean and variance of the lognormal distribution!. Instead, we will use  $\mu_{LN}$  and  $\sigma_{LN}^2$  to denote the lognormal mean and variance respectively.

For an i.i.d. random sample  $x_1, x_2, \dots, x_n$ :

$$\begin{aligned}\ell(\mu, \sigma^2) &= \sum_{i=1}^n \ln f(x_i; \mu, \sigma^2) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2\end{aligned}$$

The  $k$ -th moment of log normal distribution given by

$$\mathbb{E}[X^k] = \exp\left(k\mu + \frac{1}{2}k^2\sigma^2\right).$$

Particularly, the mean of the above lognormal distribution, denoted by  $\mu_{LN}$ , is given by

$$\mu_{LN} = \mathbb{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

Once the parameters were estimated, say  $(\bar{\mu}, \bar{\sigma}^2)$ , we can use **the plug-in Principle** to estimate the lognormal mean by

$$\bar{\mu}_{LN} = \exp\left(\bar{\mu} + \frac{1}{2}\bar{\sigma}^2\right).$$

**Working Dataset:** A reservoir engineer at an oil exploration company has recently drilled 50 new wells in a shale formation. According to geological theory, the size of oil reserves in this type of formation is often modeled using a log-normal distribution. The engineer is responsible for estimating the typical reserve size and quantifying the uncertainty of these estimates for a presentation to senior management. The estimated ultimate recovery (EUR) for the 50 wells, measured in thousands of barrels (Mbbl), is as follows:

17.3, 24.8, 8.2, 31.5, 14.1, 42.7, 11.9, 55.3, 21.4, 9.7, 36.2, 18.6, 63.1, 13.2, 28.9, 47.5, 19.8, 7.5, 33.4, 52.1, 16.0, 25.9, 38.7, 10.3, 44.2, 22.5, 58.6, 12.8, 30.3, 48.9, 20.1, 35.4, 15.7, 60.2, 26.7, 41.3, 18.1, 53.8, 23.9, 46.2, 29.4, 37.6, 14.9, 50.5, 32.8, 19.3, 56.7, 11.2, 39.5, 27.1

This assignment focuses on finding the asymptotic sampling distributions of the MLE  $\hat{\mu}$  and  $\hat{\sigma}^2$  corresponding to  $\mu$  and  $\sigma^2$  based on an environmental study data.

## 5.2 Question 1: Method of Moments Estimation (MME) of $\mu$ and $\sigma^2$

Estimate the parameters  $\mu$  and  $\sigma^2$  (*caution: not  $\sigma$ !*) from the raw oil reserve data using the given  $k$ -th moment function. Organize your work into three parts:

- Derive the system of two equations for the method of moments estimators (MMEs) of  $\mu$  and  $\sigma^2$ .
- Derive the closed-form expressions for the MMEs of  $\mu$ ,  $\sigma^2$ , and  $\mu_{LN}$ , denoted by  $\tilde{\mu}$ ,  $\tilde{\sigma}^2$ , and  $\tilde{\mu}_{LN}$ , respectively.

(c) Write an R function that outputs the MMEs  $\widetilde{\mu}$ ,  $\widetilde{\sigma^2}$ , and  $\widetilde{\mu_{LN}}$  based on the given data set.

### 5.3 Question 2: Finding the MLE of $\mu$ and $\sigma^2$

This question involves two parts:

(1). Derive the score functions (i.e., the gradient of the log-likelihood) for the parameters of the log-normal distribution, and set them to zero to obtain the system of score equations. Begin with the full log-likelihood and compute its partial derivatives with respect to  $\mu$  and  $\sigma^2$ . [Hint: letting  $\beta = \sigma^2$  may simplify differentiation.]

Assume that  $\{x_1, x_2, \dots, x_n\} \rightarrow \text{LN}(\mu, \sigma^2)$  with density function given by

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \sum_{i=1}^n \ln x_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln x_i - \mu)^2$$

(2). Perform some algebra to find the closed form of the MLE of  $\mu$  and  $\sigma^2$ , denoted by  $\widehat{\mu}$  and  $\widehat{\sigma^2}$ .

(3). Recall that lognormal population mean (i.e., the first moment) is

$$\mu_{LN} = \mathbb{E}[X] = \exp\left(\mu + \frac{1}{2}\sigma^2\right).$$

Using the relationship between variance and the first and second moments,  $\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ , we derive the population variance of the lognormal distribution, denoted as  $\sigma_{LN}^2$ . This result will be used in applying the Central Limit Theorem in subsequent questions

### 5.4 Question 3: Sampling Distribution of Lognormal Sample Mean $\widehat{\mu_{LN}}$

We have derived two different estimators of lognormal population mean  $\mu_{LN}$ . This Question focuses on the sampling distribution of lognormal sample means based on MLE. The following are detailed instructions.

(1). Write an R function to estimate the MLE of  $\mu$  and  $\sigma^2$ , denoted by  $\widehat{\mu}$  and  $\widehat{\sigma^2}$  and return the MLE of  $\mu_{LN}$ , denoted by  $\widehat{\mu_{LN}}$ .

(2). Take 1000 Bootstrap samples from the raw oil data set. For each bootstrap sample, call the above R function to find the **bootstrap MLE** of  $\mu_{LN}$ , denoted by  $\widehat{\mu_{LN}}^{*(1)}$ ,  $\widehat{\mu_{LN}}^{*(2)}$ ,  $\dots$ ,  $\widehat{\mu_{LN}}^{*(1000)}$ .

(3). Using kernel density estimation approach to estimate the bootstrap density curves based on the bootstrap MLEs:  $\widehat{\mu_{LN}}^{*(1)}$ ,  $\widehat{\mu_{LN}}^{*(2)}$ ,  $\dots$ ,  $\widehat{\mu_{LN}}^{*(1000)}$ .

(4). In order to find the asymptotic sampling distribution using the **central limit theorem (CLT)**. That is, when sample size is large,

$$\bar{X} \rightarrow N\left(\mu_{LN}, \frac{\sigma_{LN}}{\sqrt{n}}\right)$$

where  $\mu_{LN}$  and  $\sigma_{LN}$  can be estimated by using the **plug-in principle of MLE**. Based on the asymptotic sampling distribution of **lognormal sample mean**  $\bar{X}$ , add the asymptotic density curve to the above Bootstrap kernel density curve.

(5). Write a short summary to compare the two sampling distributions of the sample mean of the lognormal distribution.

### 5.5 Question 4: Extra Bonus Credit

Derive the Hessian matrix by computing the second-order partial derivatives of the log-likelihood function (with respect to  $\mu$  and  $\sigma^2$ ). Then, translate the derived Hessian into an R function that takes the MLEs of the parameters and the data as inputs and returns the Hessian matrix. Finally, compare your analytically derived Hessian with the numerical Hessian returned by `optim()`.