Lecture 20: Two-parameter Estimation and Properties of Estimators

Slide Credit: Based on Stefano Tessaro’s slides for 312 19au incorporating ideas from Anna Karlin, Alex Tsun, Rachel Lin, Hunter Schafer & myself 😊
Agenda

• MLE Practice
• Two-parameter Estimation
• Properties of Estimators
  – Biased Estimators
  – Consistent Estimators
MLE for exponential distribution

Given \( n \) samples \( x_1, \ldots, x_n \) from an Exponential distribution with unknown parameter \( \theta \)

The **likelihood** function of independent observations \( x_1, \ldots, x_n \) is

\[
\mathcal{L}(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i}
\]

Find the MLE \( \hat{\theta} \)
log(ab) = log(a) + log(b)
log(a/b) = log(a) − log(b)
log(a^b) = blog(a)
General Recipe

1. **Input** Given $n$ iid samples $x_1, ..., x_n$ from parametric model with parameters $\theta$.

2. **Likelihood** Define your likelihood $\mathcal{L}(x_1, ..., x_n | \theta)$.
   - For discrete $\mathcal{L}(x_1, ..., x_n | \theta) = \prod_{i=1}^{n} \Pr(x_i ; \theta)$
   - For continuous $\mathcal{L}(x_1, ..., x_n | \theta) = \prod_{i=1}^{n} f(x_i ; \theta)$

3. **Log** Compute $\ln \mathcal{L}(x_1, ..., x_n | \theta)$

4. **Differentiate** Compute $\frac{\partial}{\partial \theta} \ln \mathcal{L}(x_1, ..., x_n | \theta)$

5. **Solve for** $\hat{\theta}$ by setting derivative to 0 and solving for max.

Generally, you need to do a second derivative test to verify it is a maximum, but we won’t ask you to do that in CSE 312.
Agenda

• MLE Practice
• **Two-parameter Estimation**
• Properties of Estimators
  – Biased Estimators
  – Consistent Estimators
Next: $n$ samples $x_1, \ldots, x_n \in \mathbb{R}$ from Gaussian $\mathcal{N}(\mu, \sigma^2)$. Most likely $\mu$ and $\sigma^2$?
**Two-parameter optimization**

Normal outcomes \(x_1, \ldots, x_n\)

**Goal:** estimate \(\theta_1 = \mu = \) expectation and \(\theta_2 = \sigma^2 = \) variance

\[
L(x_1, \ldots, x_n | \theta_1, \theta_2) = \left(\frac{1}{\sqrt{2\pi\theta_2}}\right)^n \prod_{i=1}^{n} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2}}
\]

\[
\ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = \log(ab) = \log(a) + \log(b)
\]

\[
\log(a/b) = \log(a) - \log(b)
\]

\[
\log(a^b) = b \log(a)
\]
Two-parameter optimization

Normal outcomes $x_1, \ldots, x_n$

**Goal:** estimate $\theta_1 = \mu = \text{expectation}$ and $\theta_2 = \sigma^2 = \text{variance}$

\[
L(x_1, \ldots, x_n | \theta_1, \theta_2) = \left( \frac{1}{\sqrt{2\pi \theta_2}} \right)^n \prod_{i=1}^{n} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2}}
\]

\[
\ln L(x_1, \ldots, x_n | \theta_1, \theta_2) =
\]

\[
= -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}
\]
Two-parameter estimation

\[ \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2} \]

We need to find a solution \( \hat{\theta}_1, \hat{\theta}_2 \) to

\[ \frac{\partial}{\partial \theta_1} \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = 0 \]

\[ \frac{\partial}{\partial \theta_2} \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = 0 \]
MLE for Expectation

\[ \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2} \]

\[ \frac{\partial}{\partial \theta_1} \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = \]
MLE for Expectation

\[
\ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \theta_1)^2}{2\theta_2}
\]

\[
\frac{\partial}{\partial \theta_1} \ln L(x_1, \ldots, x_n | \theta_1, \theta_2) = \frac{1}{\theta_2} \sum_{i}^{n} (x_i - \theta_1) = 0
\]

\[
\hat{\theta}_1 = \frac{\sum_{i}^{n} x_i}{n}
\]

In other words, MLE of expectation is the *sample mean* of the data, regardless of \( \theta_2 \)

What about the variance?
MLE for Variance

\[
\ln L(x_1, \ldots, x_n | \hat{\theta}_1, \theta_2) = -n \frac{\ln(2\pi \theta_2)}{2} - \sum_{i=1}^{n} \frac{(x_i - \hat{\theta}_1)^2}{2\theta_2}
\]

\[= -n \frac{\ln 2\pi}{2} - n \frac{\ln \theta_2}{2} - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2
\]

\[
\frac{\partial}{\partial \theta_2} \ln L(x_1, \ldots, x_n | \theta_1, \hat{\theta}_1) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2 = 0
\]

\[
\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_1)^2
\]

In other words, MLE of variance is what’s called the *population variance* of the data set.
Likelihood – Continuous Case

**Definition.** The *likelihood* of independent observations \(x_1, \ldots, x_n\) is

\[
L(x_1, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta)
\]

Normal outcomes \(x_1, \ldots, x_n\)

\[
\hat{\theta}_\mu = \frac{\sum_{i}^{n} x_i}{n}
\]

MLE estimator for expectation

\[
\hat{\theta}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\theta}_\mu)^2
\]

MLE estimator for variance
Agenda

• MLE Practice
• Two-parameter Estimation
• Properties of Estimators
  – Biased Estimators
  – Consistent Estimators
When is an estimator good?

\[ \theta = \text{unknown parameter} \]

**Definition.** An estimator of parameter \( \theta \) is an **unbiased estimator**

\[ \mathbb{E}(\hat{\theta}_n) = \theta. \]
Example – Coin Flips

Coin-flip outcomes $x_1, \ldots, x_n$, with $n_H$ heads, $n_T$ tails

Recall: $\hat{\theta} = \frac{n_H}{n}$

Fact. $\hat{\theta}$ is unbiased

i.e., $\mathbb{E}(\hat{\theta}) = p$, where $p$ is the probability that the coin turns out heads.
Consistent Estimators & MLE

Definition. An estimator is unbiased if $\mathbb{E}(\hat{\theta}_n) = \theta$ for all $n \geq 1$.

Definition. An estimator is consistent if $\lim_{n \to \infty} \mathbb{E}(\hat{\theta}_n) = \theta$.

Theorem. MLE estimators are consistent. (But not necessarily unbiased)
Example – Consistency

Normal outcomes $X_1, \ldots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$  

Assume: $\sigma^2 > 0$

$$
\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2
$$

**MLE – Biased!**

$\hat{\sigma}^2$ converges to $\sigma^2$, as $n \to \infty$.

$\hat{\sigma}^2$ is “consistent”
Why is the estimator consistent, but biased?

\[
\mathbb{E}(\hat{\Theta}_{\sigma^2}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \bar{\Theta}_\mu)^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \right] \\
= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ X_i^2 - \frac{2}{n} X_i \sum_{j=1}^{n} X_j + \frac{1}{n^2} \sum_{j=1}^{n} X_j \sum_{k=1}^{n} X_k \right]
\]

...
Why is the estimator consistent, but biased?

\[
\mathbb{E}(\hat{\Theta}_{\sigma^2}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \hat{\Theta})^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \right]
\]

... 

\[
= \left( 1 - \frac{1}{n} \right) \sigma^2 = \frac{n-1}{n} \sigma^2
\]
Why is the estimator consistent, but biased?

\[ \mathbb{E}(\hat{\Theta}_{\sigma^2}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \hat{\Theta}_1)^2 \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E} \left[ \left( X_i - \frac{1}{n} \sum_{j=1}^{n} X_j \right)^2 \right] \]

... 

\[ = \left( 1 - \frac{1}{n} \right) \sigma^2 = \frac{n-1}{n} \sigma^2 \to \sigma^2 \text{ for } n \to \infty \]

Therefore: 

\[ \frac{1}{n-1} \sum_{i=1}^{n} \mathbb{E} \left[ (X_i - \hat{\Theta}_1)^2 \right] = \frac{n}{n-1} \mathbb{E}(\hat{\Theta}_{\sigma^2}) = \sigma^2 \]

Bessel's correction
Example – Consistency

Normal outcomes $X_1, \ldots, X_n$ iid according to $\mathcal{N}(\mu, \sigma^2)$  \hspace{1cm} Assume: $\sigma^2 > 0$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

MLE – Biased!

$$\hat{\sigma}^2$$ converges to $\sigma^2$, as $n \to \infty$.

$$\hat{\sigma}^2$$ is “consistent”

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu})^2$$

Sample variance – Unbiased!