STAT 390 A
Statistical Methods in Engineering and Science
Week 7 Lectures – Part 3 – Spring 2023
Properties of the MLE & Standard Errors

Alexander Giessing

Department of Statistics

University of Washington

May 12, 2023

Outline

Properties of the MLE

Invariance Property

Let $g : \mathbb{R}^m \to \mathbb{R}^m$ be an invertible function. If T is the MLE for a parameter $\theta \in \mathbb{R}^m$, then $g(T)$ is the MLE of $g(\theta)$.

- Recall the Gaussian error model:

$$X_i = \mu + \varepsilon_i, \quad \varepsilon_i \sim_{iid} N(0, \sigma^2) \qquad 1 \leq i \leq n.$$

- We have shown that

$$\hat{\sigma}_{MLE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

  is the MLE of the standard deviation σ of the errors.

- Hence, $\hat{\sigma}^2_{MLE}$ is the MLE of the variance σ^2 of the error (because the map $g(x) = x^2$ is invertible on the positive real numbers).

Properties of the MLE

Asymptotic Unbiasedness

If T_n is the MLE of a parameter θ based on a random sample $X_1, \ldots, X_n$ of size n, then T_n is asymptotically unbiased for θ, i.e.

$$\lim_{n\to\infty} \mathrm{E}[T_n] = \theta.$$

- We have shown that an unbiased estimate for the variance σ^2 in the Gaussian error model is

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

- We compute

$$\mathrm{E}[\hat{\sigma}^2_{MLE}] = \mathrm{E}\left[\frac{n-1}{n}S_n^2\right] = \frac{n-1}{n}\mathrm{E}\left[S_n^2\right] = \frac{n-1}{n}\sigma^2.$$

Hence,

$$\lim_{n\to\infty} \mathrm{E}[\hat{\sigma}^2_{MLE}] = \lim_{n\to\infty}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

Properties of the MLE

Asymptotic minimum variance (Heuristic)

If T_n is the MLE of a parameter θ based on a random sample $X_1, \ldots, X_n$ of size n, then T_n has asymptotically (as $n \to \infty$) the smallest variance among all unbiased estimators for θ.

- This is a very deep result *(for those of you who are interested, check: "Cramér-Rao lower bound")*.
- **Note:** The theorem does not rule out the existence of biased estimators with smaller variance than the MLE.

Example: MME, MLE for upper endpoint of $Unif(0, \theta)$.

Let $X_1, \ldots X_n$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Compute the MME and MLE of θ and compare their variances/ standard errors.

- Recall that $\mathrm{E}[X] = \theta/2$. Thus, the MME $\hat{\theta}_{MME}$ solves

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\hat{\theta}_{MME}}{2} \Leftrightarrow \hat{\theta}_{MME} = \frac{2}{n}\sum_{i=1}^{n} X_i.$$

- The variance is

$$\mathrm{Var}(\hat{\theta}_{MME}) = \mathrm{Var}\left(\frac{2}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{4}{n}\frac{\theta^2}{12} = \frac{\theta^2}{3n}.$$

- The standard error (aka standard deviation of an estimator) is

$$SE(\hat{\theta}_{MME}) = \sqrt{\mathrm{Var}(\hat{\theta}_{MME})} = \frac{\theta}{\sqrt{3n}}.$$

Example: MME, MLE (Cont.)

Let $X_1, \ldots X_n$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Compute the MME and MLE of θ and compare their variances.

- Recall that the pdf of $Unif(0, \theta)$ is given by

$$f(x; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \leq x \leq \theta \\ 0 & o/w. \end{cases}$$

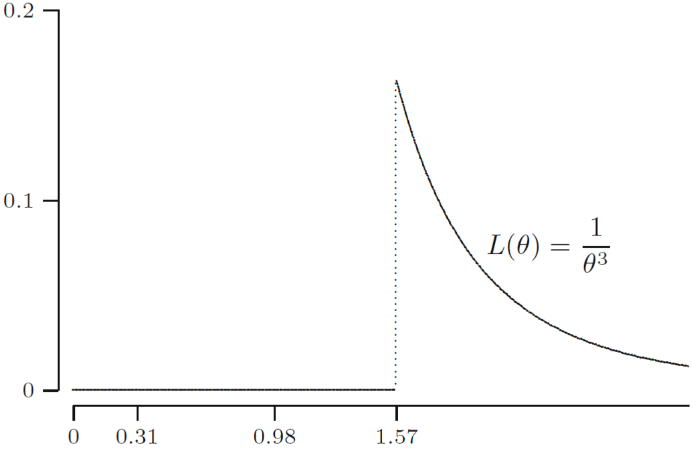
- Thus, the likelihood function based on $X_1, \ldots, X_n$ is given by

$$L(\theta) = \prod_{i=1}^{n} f(X_i; \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \leq X_i \leq \theta \ \text{ for all } 1 \leq i \leq n, \\ 0 & o/w. \end{cases}$$

- How do we maximize this (odd) likelihood function ...?

Example: MME, MLE (Cont.)

Likelihood function corresponding to a sample of size $n = 3$ *with* $x_1 = 0.98$, $x_2 = 1.57$, $x_3 = 0.31$. *We easily see that* $L(\theta)$ *attains its maximum at* $\max\{x_1, x_2, x_3\} = 1.57$.



Example: MME, MLE (Cont.)

Let $X_1, \dots X_n$ be a random sample from the uniform distribution on $(0, \theta)$, where $\theta > 0$ is unknown. Compute the MME and MLE of θ and compare their variances.

- Recall that the pdf of $Unif(0, \theta)$ is given by

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x \le \theta \\ 0 & o/w. \end{cases}$$

- Thus, the likelihood function based on $X_1, \dots, X_n$ is given by

$$L(\theta) = \prod_{i=1}^{n} f(X_i;\theta) = \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le X_i \le \theta \text{ for all } 1 \le i \le n, \\ 0 & o/w. \end{cases}$$

- General solution:
  - $L(\theta) = 0$ if θ is smaller than at least of the X_i.
  - $L(\theta) = 1/\theta^n$ if θ is greater than or equal to the largest X_i.

$$\Longrightarrow \hat{\theta}_{MLE} = \max\{X_1, \dots, X_n\}.$$

Example: MME, MLE (Cont.)

Let $X_1, \dots X_n$ *be a random sample from the uniform distribution on* $(0, \theta)$*, where* $\theta > 0$ *is unknown. Compute the MME and MLE of* θ *and compare their variances.*

- Recall that if $X_1, \dots, X_n \sim_{iid} F$ then $V = \max\{X_1, \dots X_n\}$ has cdf $F_V(x) = (F(x))^n$. Hence, cdf and pdf or $\hat{\theta}_{MLE}$ are

$$F_{\hat{\theta}_{MLE}}(x) = \left(\frac{x}{\theta}\right)^n \qquad \text{and} \qquad f_{\hat{\theta}_{MLE}}(x) = n\frac{x^{n-1}}{\theta^n}.$$

- Thus, we compute

$$\begin{aligned}
\mathrm{E}[\hat{\theta}_{MLE}] &= \int_0^\theta xn\frac{x^{n-1}}{\theta^n}dx = \frac{n}{n+1}\theta \to \theta \quad \text{as} \quad n \to \infty, \\
\mathrm{E}[\hat{\theta}^2_{MLE}] &= \int_0^\theta x^2 n\frac{x^{n-1}}{\theta^n}dx = \frac{n}{n+2}\theta^2, \\
\mathrm{Var}(\hat{\theta}^2_{MLE}) &= \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 < \frac{1}{3n}\theta^2.
\end{aligned}$$

Outline

Estimating the Standard Error

In the preceding example on the uniform distribution the standard error (SE) depends on the unknown parameter θ. How should we to estimate the SE?

- Recall from the preceding example that

$$SE(\hat{\theta}_{MME}) = \frac{\theta}{\sqrt{3n}} \qquad \text{and} \qquad SE(\hat{\theta}_{MLE}) = \sqrt{\frac{n}{n+2}}\frac{\theta}{n+1}.$$

- We can simply estimate the SE by plugging in the estimates of θ, i.e,

$$\widehat{SE}(\hat{\theta}_{MME}) = \frac{\hat{\theta}_{MME}}{\sqrt{3n}} = \frac{2\bar{X}_n}{\sqrt{3n}}$$

$$\widehat{SE}(\hat{\theta}_{MLE}) = \sqrt{\frac{n}{n+2}}\frac{\hat{\theta}_{MLE}}{n+1} = \sqrt{\frac{n}{n+2}}\frac{\max\{X_1, \ldots, X_n\}}{n+1}.$$

- This "plug-in approach" works whenever we can obtain a closed-form expression for the SE.

Example: $\widehat{SE}$ for $\hat{\mu}_{MLE}$ in the Gaussian error model

- Recall the Gaussian error model:

$$X_i = \mu + \varepsilon_i, \quad \varepsilon_i \sim_{iid} N(0, \sigma^2) \qquad 1 \le i \le n.$$

- We have shown that

$$\hat{\mu}_{MLE} = \bar{X}_n \qquad \text{and} \qquad \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

- The SE of $\hat{\mu}_{MLE}$ is

$$SE(\hat{\mu}_{MLE}) = \sqrt{\mathrm{Var}(\hat{\mu}_{MLE})} = \frac{\sigma}{\sqrt{n}}.$$

- We can estimate the SE of $\hat{\mu}_{MLE}$ with the "plug-in" estimate based on $\hat{\sigma}^2_{MLE}$ or the unbiased estimate of the variance S_n^2,

$$\widehat{SE}(\hat{\mu}_{MLE}) = \frac{\hat{\sigma}_{MLE}}{\sqrt{n}} = \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2},$$

$$\widehat{SE}(\hat{\mu}_{MLE}) = \frac{S_n}{\sqrt{n}} = \sqrt{\frac{1}{n(n-1)}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}.$$

Optional Example: Population Genetics

- Consider a population of organisms which are either male or female, reproduce by combining male and female gametes, and which have two alleles a and A at each gene locus.

- The Hardy-Weinberg principle postulates that when sampling from such a population (in equilibrium), one can observe three genotypes (aa, aA, AA) with the following frequencies

	aa	aA	AA
	$p_1 = \theta^2$	$p_2 = 2\theta(1-\theta)$	$p_3 = (1-\theta)^2$
$\theta = 0.5$	0.25	0.5	0.25

- Biologists are interested in estimating the parameter θ, which measures the frequency of the gene allele a in the population.

- What is the role of a statistician in this process?

  1. Build a probability model that includes the parameter θ.
  2. Estimate the parameter θ via MLE (or MME).
  3. Compute and estimate the standard error of the estimate.

Optional Example: Population Genetics (Cont.)

1. Build a probability model that includes the parameter θ.

- Consider drawing n elements from a population with r categories. Let X_i be the number of "successes" in ith category and p_i the corresponding "success" probability. Then, its pmf is given by

$$\begin{aligned} p(x_1, \ldots, x_r) &= \binom{n}{x_1}\binom{n-x_1}{x_2}\cdots\binom{n-x_1-\ldots-x_{r-1}}{x_r} p_1^{x_1}\cdots p_r^{x_r} \\ &= \frac{n!}{x_1!\cdots x_r!} p_1^{x_1}\cdots p_r^{x_r}, \qquad x_1+\ldots+x_r = n. \end{aligned}$$

- Applied to our problem: Denote by N_1, N_2, N_3 be the number of geneotypes aa, aA, AA, respectively. Then,

$$p(N_1, N_2, N_3) = \frac{n!}{N_1!N_2!N_3!}\theta^{2N_1}\left(2\theta(1-\theta)\right)^{N_2}\left(1-\theta\right)^{N_3},$$

  where $N_1 + N_2 + N_3 = n$.

Optional Example: Population Genetics (Cont.)

2. Estimate the parameter θ via MLE (or MME).

- In a sample of size n, let (n_1, n_2, n_3) be the outcome of genotypes aa, aA, AA. Then, the likelihood function is given as

$$L(\theta) = \frac{n!}{n_1!n_2!n_3!}\theta^{2n_1}\left(2\theta(1-\theta)\right)^{n_2}(1-\theta)^{2n_3},$$

- The log-likelihood function is given by

$$\log L(\theta) = c + (2n_1 + n_2)\log\theta + (n_2 + 2n_3)\log(1-\theta),$$

  where c is some constant independent of θ.

- Solving the first order condition for θ (i.e. setting the first derivative of $\log L(\theta)$ equal to zero) we find

$$\hat{\theta}_{MLE} = \frac{2n_1 + n_2}{2n}.$$

- Find the MME yourself! *(Optional practice exercise.)*

Optional Example: Population Genetics (Cont.)

3. Compute and estimate the standard error of the estimate.

$$\begin{aligned}\mathrm{Var}(\hat{\theta}_{MLE}) &= \mathrm{Var}\left(\frac{2N_1+N_2}{2n}\right)\\ &= \frac{1}{n^2}\mathrm{Var}(N_1) + \frac{1}{4n^2}\mathrm{Var}(N_2) + \frac{2}{n^2}\mathrm{Cov}(N_1, N_2).\end{aligned}$$

- First, let's compute the marginal pmf of X_1, i.e for $k = 0, \ldots, n$

$$\begin{aligned}p(X_1 = k) &= \sum_{x_2+x_3=n-k} p(X_1 = k, X_2 = x_2, X_3 = x_3)\\ &= \sum_{x_2+x_3=n-k} \frac{n!}{k!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}\\ &= \frac{n!}{k!(n-k)!} p_1^{x_1} \sum_{x_2+x_3=n-k} \frac{(n-k)!}{x_2!x_3!} p_2^{x_2} p_3^{x_3}\\ &= \frac{n!}{k!(n-k)!} p_1^{x_1} \sum_{j=0}^{n-k} \frac{(n-k)!}{j!(n-k-j)!} p_2^{j} p_3^{n-k-j}\\ &= \frac{n!}{k!(n-k)!} p_1^{x_1} (p_2+p_3)^{n-k} = \frac{n!}{k!(n-k)!} p_1^{x_1} (1-p_1)^{n-k}.\end{aligned}$$

Optional Example: Population Genetics (Cont.)

- The preceding calculations show that $X_i \sim Bin(n, p_i)$ for $i = 1, 2, 3$.
- Second, let's compute the covariance between X_1 and X_2. Recall the identity

$$2\mathrm{Cov}(X_1, X_2) = \mathrm{Var}(X_1 + X_2) - \mathrm{Var}(X_1) - \mathrm{Var}(X_2).$$

What is the variance of $X_1 + X_2$? Intuitively, $X_1 + X_2$ means that we combine category 1 and 2, hence the success probability of the combined category is $p_1 + p_2$. In particular for all k, x_3 such that $k + x_3 = n$,

$$p(X_1 + X_2 = k, X_3 = x_3) = \frac{n!}{k!x_3!}(p_1 + p_2)^k p_3^{x_3},$$

and by calculations as on the previous slide

$$p(X_1 + X_2 = k) = \frac{n!}{k!(n-k)!}(p_1 + p_2)^k (1 - p_1 - p_2)^{n-k}.$$

Thus, $X_i + X_j \sim Bin(n, p_i + p_j)$ for all $i \neq j$.

Optional Example: Population Genetics (Cont.)

- We can now compute the variance of $\hat{\theta}_{MLE}$ as

$$\begin{aligned}
\mathrm{Var}(\hat{\theta}_{MLE}) &= \frac{1}{n^2}\mathrm{Var}(N_1) + \frac{1}{4n^2}\mathrm{Var}(N_2) + \frac{2}{n^2}\mathrm{Cov}(N_1, N_2) \\
&= \frac{np_1(1-p_1)}{n^2} + \frac{np_2(1-p_2)}{4n^2} \\
&\quad + \frac{1}{n^2}\big(n(p_1+p_2)(1-p_1-p_2) - np_1(1-p_1) - np_2(1-p_2)\big) \\
&= \frac{1}{n}\big(p_1(1-p_1) + p_2(1-p_2)/4 - 2p_1p_2\big) \\
&= \frac{\theta(1-\theta)}{2n} - \frac{2\theta^3(1-\theta)}{n}.
\end{aligned}$$

- By the "plug-in" principle, we estimate the standard error of the MLE as

$$\widehat{SE}(\hat{\theta}_{MLE}) = \sqrt{\frac{\hat{\theta}_{MLE}(1-\hat{\theta}_{MLE})}{2n} - \frac{2\hat{\theta}^3_{MLE}(1-\hat{\theta}_{MLE})}{n}}.$$