

STAT 713 sp 2023 Lec 04 slides

Moment, maximum likelihood, and quantile estimators

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

Table of Contents

- 1 Moment estimators
- 2 Maximum likelihood estimators
- 3 Quantile estimators



Moment estimator

Let X_1, \dots, X_n be iid with a distribution characterized by θ .

- 1 The sample k th moment $\hat{m}_k = n^{-1} \sum_{i=1}^n X_i^k$ is an estimator of $m_k = \mathbb{E}X_1^k$.
- 2 If $\theta = g(m_1, \dots, m_k)$ then a *moment estimator* of θ is $\hat{\theta} = g(\hat{m}_1, \dots, \hat{m}_k)$.

The *method of moments* (MoMs) prescribes finding g by solving the system

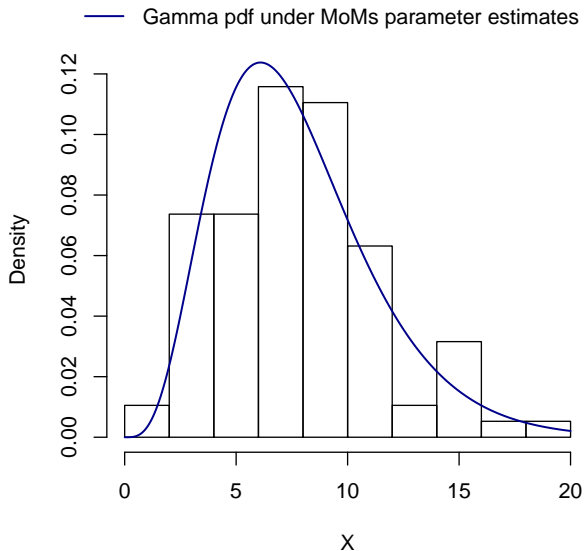
$$\hat{m}_j = m_j(\theta), \quad j = 1, \dots, \text{as many as you need.}$$

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$.

- 1 Find the MoMs estimators of μ and σ^2 .
- 2 Discuss whether better estimators might exist.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$.

- 1 Find the MoMs estimators of α and β .
- 2 Discuss whether better estimators might exist.
- 3 Compute MoMs estimators on [birth data set](#) from Davison (2003)



Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(0, \theta)$.

- 1 Find the MoMs estimator of θ .
- 2 Discuss whether better a estimator might exist.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Weibull}(a, b)$. The $\text{Weibull}(a, b)$ pdf is given by

$$f_X(x; a, b) = \left(\frac{a}{b}\right) \left(\frac{x}{b}\right)^{a-1} \exp\left[-\left(\frac{x}{b}\right)^a\right] \mathbf{1}(x > 0).$$

- 1 Find the MoMs estimators of a and b .
- 2 Discuss whether better estimators might exist.
- 3 Compute MoMs estimators on the [trees in Camden data set](#).

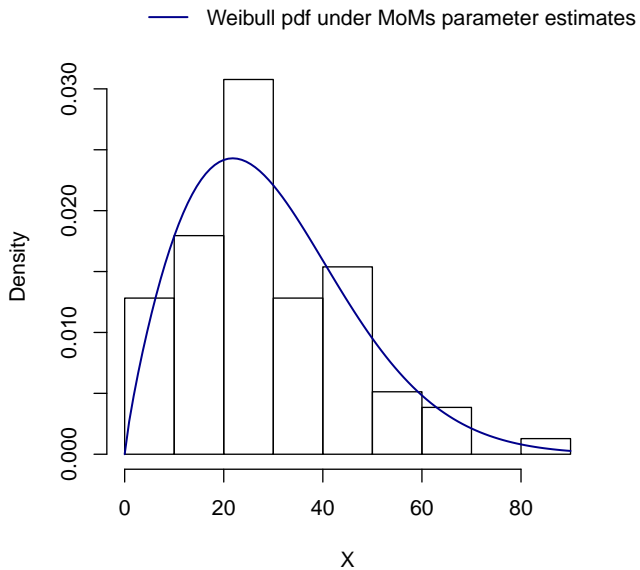


Table of Contents

- 1 Moment estimators
- 2 Maximum likelihood estimators
- 3 Quantile estimators



The *maximum-likelihood* approach asks:

Under which value of the parameter are the observed data the most “likely”?

Likelihood and log-likelihood functions

Let \mathbf{X} have joint pdf/pmf $f(\mathbf{x}; \theta)$, where $\theta \in \Theta$.

- 1 The function $\mathcal{L}(\theta; \mathbf{X}) = f(\mathbf{X}; \theta)$ is called the *likelihood function*.
- 2 The function $\ell(\theta; \mathbf{X}) = \log \mathcal{L}(\theta; \mathbf{X})$ is called the *log-likelihood function*.

The likelihood is just the joint pdf/pmf of the rvs in the random sample.

Maximum likelihood estimator (MLE)

The statistic $\hat{\theta} := \hat{\theta}(\mathbf{X})$ is an *MLE* for θ if $\mathcal{L}(\theta; \mathbf{X}) \leq \mathcal{L}(\hat{\theta}; \mathbf{X})$ for all $\theta \in \Theta$.

We obtain a/the MLE for θ by maximizing the likelihood function.

The maximizer may not be unique, but it is for most of our favorite settings.

We sometimes write the MLE $\hat{\theta}$ as

$$\hat{\theta} \in \operatorname{argmax}_{\theta \in \Theta} \mathcal{L}(\theta; \mathbf{X}) = \operatorname{argmax}_{\theta \in \Theta} \ell(\theta; \mathbf{X}).$$

If the maximizer of $\mathcal{L}(\theta; \mathbf{X})$ is unique, then “ \in ” changes to “ $=$ ”.

In the above, $\operatorname{argmax}_{z \in A} g(z) = \{z \in A : g(z') \leq g(z) \text{ for all } z' \in A\}$.

We can often (not always) use calculus methods to find the MLEs...

Theorem (Finding MLEs with calculus)

Suppose $\ell(\theta; \mathbf{X})$ is differentiable and has a single maximum in the interior of Θ . If $\theta = (\theta_1, \dots, \theta_d)$, then the MLE $\hat{\theta} = (\hat{\theta}_1, \dots, \hat{\theta}_d)$ is the solution to the system

$$\frac{\partial}{\partial \theta_j} \ell(\theta; \mathbf{X}) = 0, \quad j = 1, \dots, d.$$

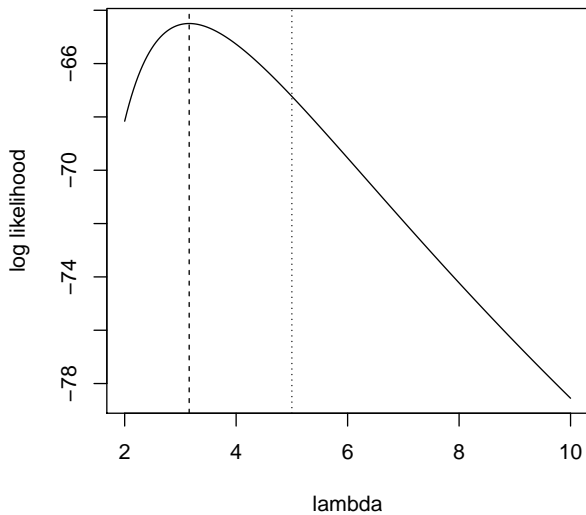
Steps for finding MLE via calculus:

- 1 Set 1st-order partial derivatives equal to zero and solve system of equations.
- 2 Check Hessian to see if a maximum (need neg-def: $\mathbf{a}^T \mathbf{H} \mathbf{a} < 0 \forall \mathbf{a}$).
- 3 Check for uniqueness or a solution on the boundary of Θ .

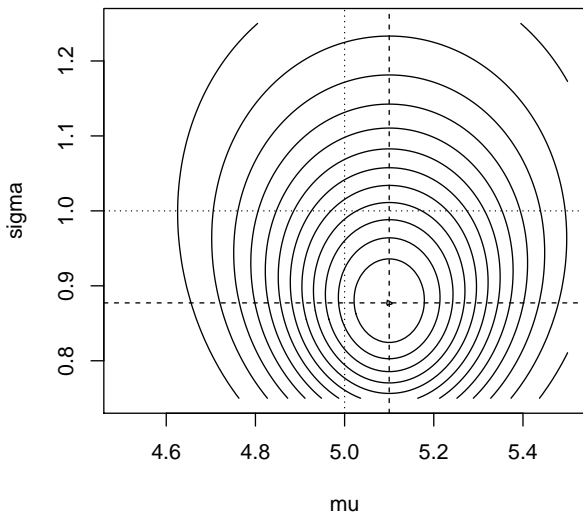
Most often easier to differentiate the log-likelihood than the likelihood.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$.

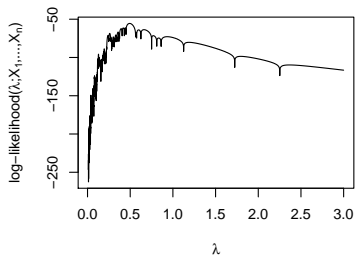
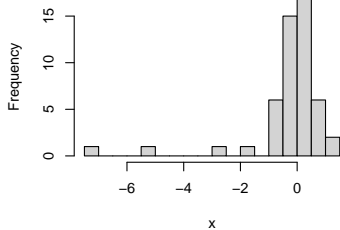
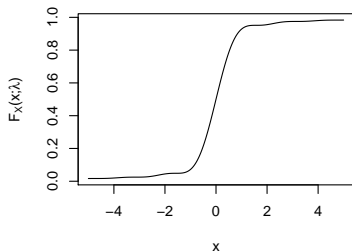
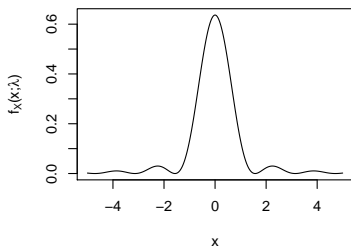
- 1 Find the MLE of λ .
- 2 Plot the log-likelihood function based on a simulated sample under $\lambda = 5$.



Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$. Find the MLEs of μ and σ^2 .



Example: Let $X_1, \dots, X_n \sim f_X(x; \lambda) = \frac{\lambda}{\pi} \frac{\sin^2(x/\lambda)}{x^2}$. Consider finding the MLE of λ .



Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(0, \theta)$. Find the MLE of θ .



Theorem (MLEs functions of minimal sufficient statistics)

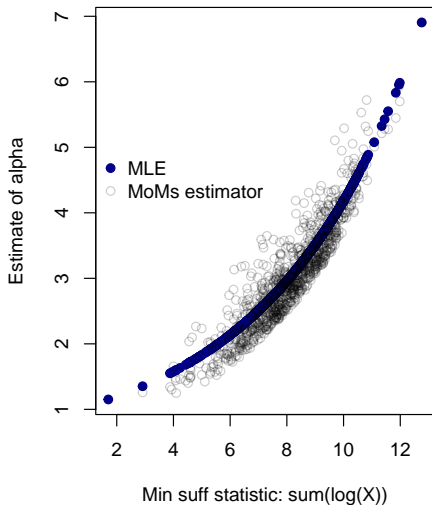
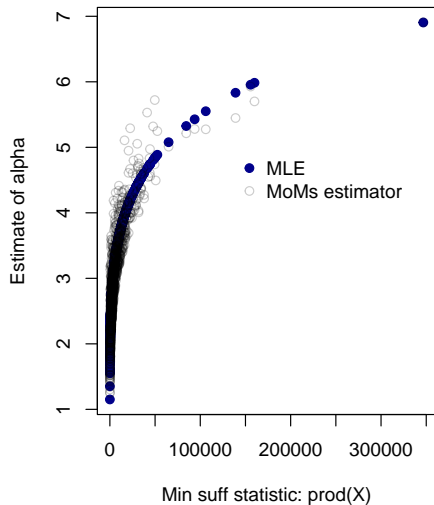
If the MLE is unique, it is a function of a minimal sufficient statistic.

So MLEs use all the information in the sample about the target parameter.

Exercise: Prove sufficiency part of above result.

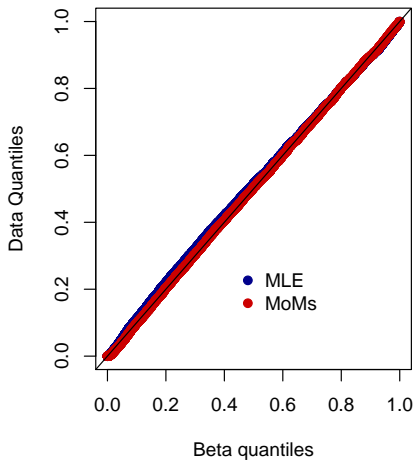
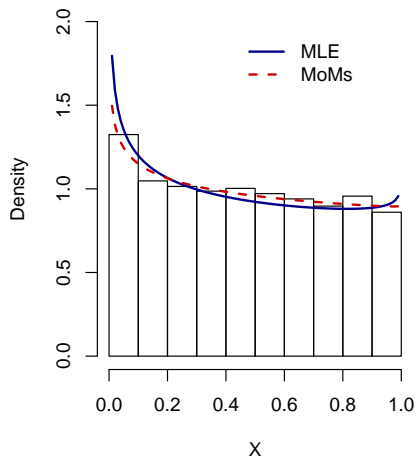
Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha, 2)$.

- 1 Find the MLE $\hat{\alpha}$ of α .
- 2 Plot many values of the pair $(\hat{\alpha}, \prod_{i=1}^n X_i)$ for simulated data under $\alpha = 3$.
- 3 Compare the MSE of $\hat{\alpha}$ to that of the MoMs estimator $\bar{\alpha}$.



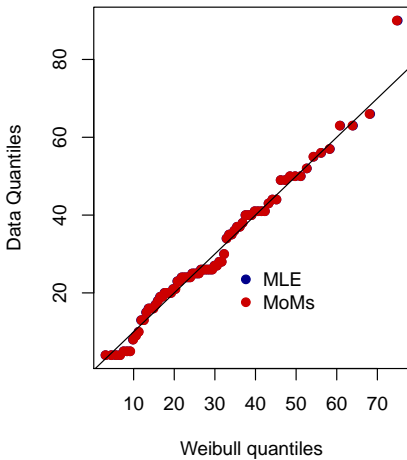
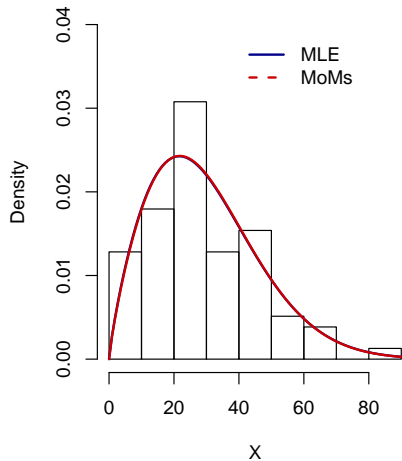
Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Beta}(\alpha, \beta)$.

- 1 Find the MLEs $\hat{\alpha}$ and $\hat{\beta}$ of α and β .
- 2 Compute MLEs on [prostate cancer p-values data](#) from [Efron \(2012\)](#).
- 3 Compare to MoMs estimates.



Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Weibull}(a, b)$.

- 1 Find the MLEs \hat{a} and \hat{b} of a and b .
- 2 Compute MLEs on the [trees in Camden data set](#).
- 3 Compare to MoMs estimates.



Suppose we are interested in a function $\eta = \tau(\theta)$ of the parameter(s).

Induced likelihood or profile likelihood

For \mathbf{X} with likelihood $\mathcal{L}(\theta; \mathbf{X})$ and a new parameter $\eta = \tau(\theta)$, the function

$$\mathcal{L}^*(\eta; \mathbf{X}) = \sup_{\{\theta \in \Theta: \tau(\theta) = \eta\}} \mathcal{L}(\theta; \mathbf{X}).$$

is called the *induced likelihood* for η .

Exercise: Give the induced likelihood for η in the following settings:

- ① $\mathbf{X} \stackrel{\text{ind}}{\sim} \text{Uniform}(a, b)$, with $\eta = b - a$.
- ② $\mathbf{X} \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$ with $\eta = \alpha$ (discuss *profile likelihood*, *nuisance param*).

Theorem (Invariance property of MLEs, cf. Thm 7.2.10 of CB)

Let $\hat{\theta}$ be an MLE for θ and suppose $\eta = \tau(\theta)$ induces the likelihood $\mathcal{L}^*(\eta; \mathbf{X})$. Then $\hat{\eta} = \tau(\hat{\theta})$ is a maximizer of $\mathcal{L}^*(\eta; \mathbf{X})$.

Says if $\hat{\theta}$ is an MLE for θ , then $\hat{\eta} = \tau(\hat{\theta})$ is an MLE for $\eta = \tau(\theta)$.

Exercise: Prove the result.

Exercise: Let $\varepsilon_1, \dots, \varepsilon_n$ be ind. Normal rvs with mean 0 and standard dev. σ .

- 1 Find the MLE of σ .
- 2 Find the MLE of $\tau = \sigma^2$.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$.

- 1 Find the MLE of $\eta = 1/\lambda$.
- 2 Find the MLE of the median of the $\text{Exponential}(\lambda)$ distribution.

Table of Contents

- 1 Moment estimators
- 2 Maximum likelihood estimators
- 3 Quantile estimators**

Quantile estimators

Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} F(x; \theta)$, with pop. quantiles τ_1, \dots, τ_k denoted by $\xi_{\tau_1}, \dots, \xi_{\tau_k}$.

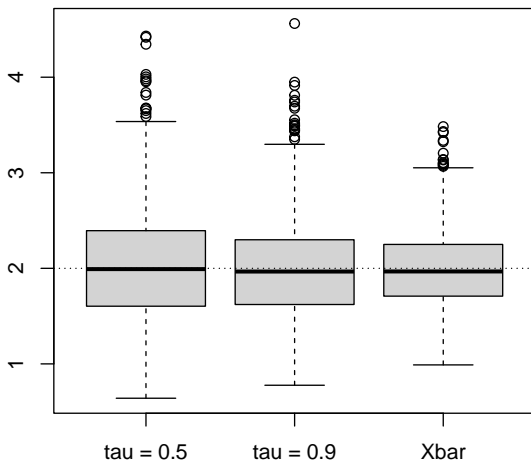
If $\theta = g(\xi_{\tau_1}, \dots, \xi_{\tau_k})$ for some function g , a *quantile estimator* for θ is

$$\hat{\theta} = g(\hat{\xi}_{\tau_1}, \dots, \hat{\xi}_{\tau_k}), \quad \text{where } (\hat{\xi}_{\tau_1}, \dots, \hat{\xi}_{\tau_k}) = (X_{(\lceil \tau_1 n \rceil)}, \dots, X_{(\lceil \tau_k n \rceil)}).$$

One can define various quantile estimators for a parameter in a given setting.

Exercise: Let $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} F(x; \lambda) = 1 - e^{-x/\lambda}$ for $x > 0$, for some $\lambda > 0$.

- 1 Give a two quantile estimators for λ .
- 2 Run a simulation to compare their performance.

Quantile estimation of exponential parameter with $n = 25$ 

Anthony Christopher Davison. *Statistical models*, volume 11. Cambridge University Press, 2003.

Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press, 2012.