STAT 512 su 2021 Lec 11 slides

MoMs and MLEs

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.

Let $X_1, \ldots, X_n$ be a rs from a dist. with parameters $(\theta_1, \ldots, \theta_d) \in \Theta \subset \mathbb{R}^d$, $d \geq 1$.

The *method of moments* sets population moments equal to sample moments.

Finds $\theta_1, \ldots, \theta_d$ which make the first d population and sample moments equal.

Method of moments (MoMs) estimators

The *MoMs* estimators of $\theta_1, \dots, \theta_d$ are the solutions to the system of equations

$$m_1' := \frac{1}{n}\sum_{i=1}^n X_i = \mathbb{E}X =: \mu_1'(\theta_1, \dots, \theta_d)$$

$$\vdots$$

$$m_d' := \frac{1}{n}\sum_{i=1}^n X_i^d = \mathbb{E}X^d =: \mu_d'(\theta_1, \dots, \theta_d),$$

provided $\mathbb{E}X, \mathbb{E}X^2, \dots, \mathbb{E}X^d$ are all finite.

The $m_1', \dots, m_d'$ are the sample moments.

The $\mu_1', \dots, \mu_d'$ are the population moments, which depend on $\theta_1, \dots, \theta_d$.

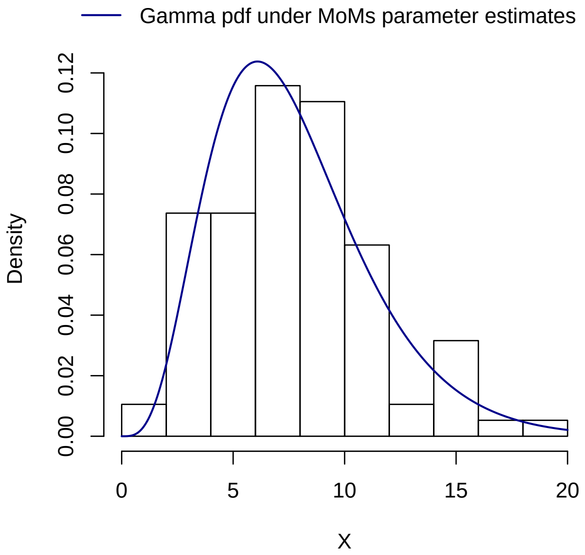
MoMs are param. vals for which 1st d pop. moments equal 1st d sample moments.

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

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

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

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)


Gamma pdf under MoMs parameter estimates
Density
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0
5
10
15
20
X

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

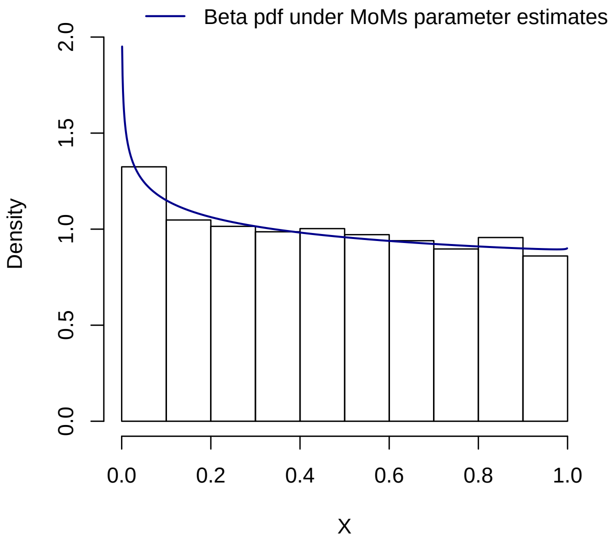
1. Find the MoMs estimator of θ.
2. Discuss whether better a estimator might exist.

Exercise: Let $Y_1, \dots, Y_n \overset{\text{ind}}{\sim} \text{Geometric}(p)$.

1. Find the MoMs estimator of p.
2. Compare to estimator $\tilde{p} = \dfrac{n-1}{\sum_{i=1}^{n} Y_i - 1}$, for $n \geq 2$.
3. Run a simulation to compare the MSE of the two estimators at $p = 1/2$.

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

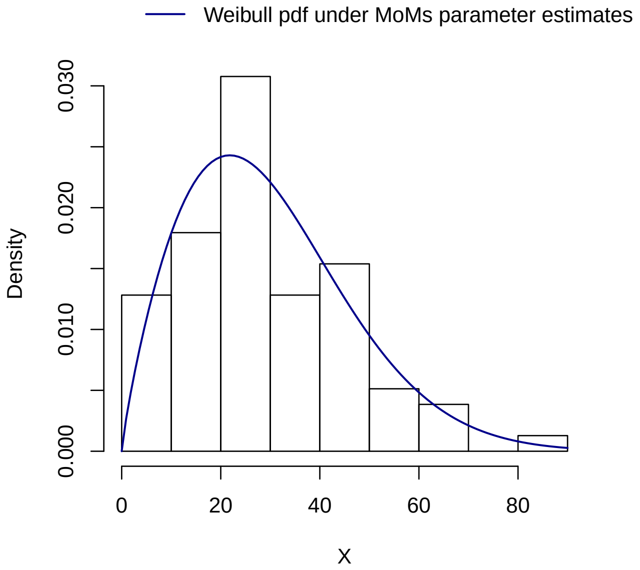
1. Find the MoMs estimators of α and β.
2. Discuss whether better estimators might exist.
3. Compute MoMs est'rs on prostate cancer p-values data from Efron (2012).


Beta pdf under MoMs parameter estimates
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
X

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim}$ Weibull(a, b). The 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 trees in Camden data.



Weibull pdf under MoMs parameter estimates
Density
0.000
0.010
0.020
0.030
0
20
40
60
80
X

The *maximum-likelihood method* takes another approach:

Find $\theta_1, \ldots, \theta_d$ which maximize joint pdf/pmf when evaluated at observed data.

Asks: Under which parameter values are the observed data the most "likely"?

Likelihood and log-likelihood functions

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} f_X(x; \theta_1, \dots, \theta_d)$, for some $(\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$, $d \geq 1$.

Then the function

$$\mathcal{L}(\theta_1, \dots, \theta_d; X_1, \dots, X_n) = \prod_{i=1}^{n} f_X(X_i; \theta_1, \dots, \theta_d)$$

is called the *likelihood function* and the function

$$\ell(\theta_1, \dots, \theta_d; X_1, \dots, X_n) = \log \mathcal{L}(\theta_1, \dots, \theta_d; X_1, \dots, X_n)$$

is called the *log-likelihood function*.

Same for discrete and continuous: $f_X(x; \theta_1, \dots, \theta_d)$ represents a pdf or a pmf.

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

Let $X_1, \dots, X_n$ have likelihood $\mathcal{L}(\theta_1, \dots, \theta_d; X_1, \dots, X_n)$, $(\theta_1, \dots, \theta_d) \in \Theta \subset \mathbb{R}^d$.

Maximum likelihood estimator (MLE)

The *maximum likelihood estimators (MLEs)* of $\theta_1, \dots, \theta_d$ are the values $\hat{\theta}_1, \dots, \hat{\theta}_d$ given by

$$(\hat{\theta}_1, \dots, \hat{\theta}_d) = \underset{(\theta_1, \dots, \theta_d) \in \Theta}{\operatorname{argmax}} \mathcal{L}(\theta_1, \dots, \theta_d; X_1, \dots, X_n).$$

The $\underset{(\theta_1, \dots, \theta_d) \in \Theta}{\operatorname{argmax}}$ returns the value of $(\theta_1, \dots, \theta_d)$ which maximizes the function.

We get the same estimator if we maximize the log-likelihood function:

$$(\hat{\theta}_1, \dots, \hat{\theta}_d) = \underset{(\theta_1, \dots, \theta_d) \in \Theta}{\operatorname{argmax}} \ell(\theta_1, \dots, \theta_d; X_1, \dots, X_n).$$

We can often use calculus methods to find the MLEs...

Theorem (Finding MLEs with calculus)

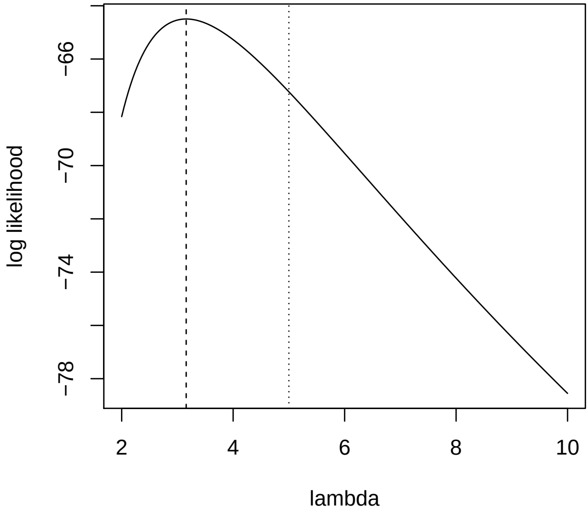
If $\ell(\theta_1, \dots, \theta_d; X_1, \dots, X_n)$ is differentiable and has a single maximum in the interior of Θ, the MLEs $\hat{\theta}_1, \dots, \hat{\theta}_d$ are the solutions to the system of equations

$$\begin{aligned}\frac{\partial}{\partial \theta_1} \ell(\theta_1, \dots, \theta_d; X_1, \dots, X_n) &= 0 \\ &\vdots \\ \frac{\partial}{\partial \theta_d} \ell(\theta_1, \dots, \theta_d; X_1, \dots, X_n) &= 0\end{aligned}$$

We often prefer using the log-likelihood, because it is easier to differentiate it.

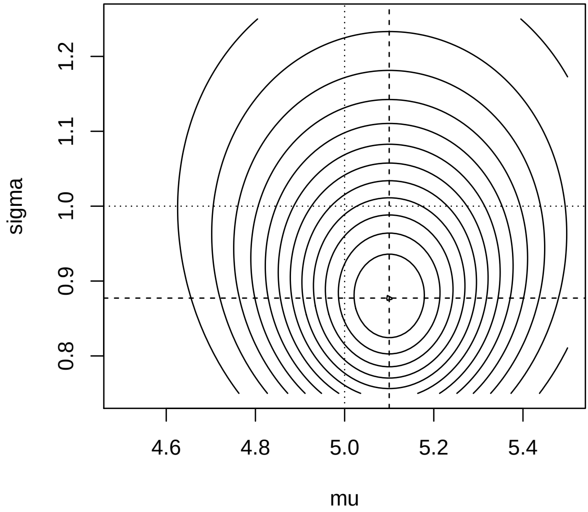
Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponental}(\lambda)$.

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


log likelihood
−66
−70
−74
−78
2
4
6
8
10
lambda

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Poisson}(\lambda)$. Find the MLE of λ.

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


sigma
1.2
1.1
1.0
0.9
0.8
4.6
4.8
5.0
5.2
5.4
mu

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

Theorem (MLEs always functions of sufficient statistics)

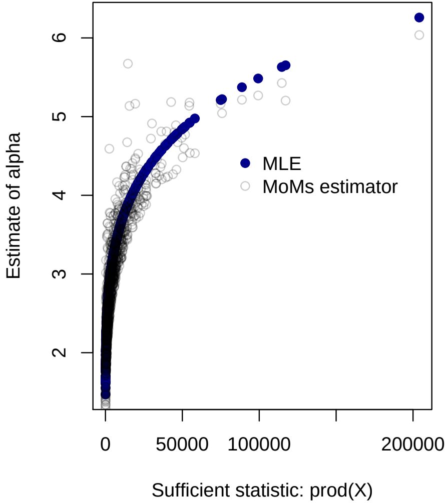
The MLE is always a function of a sufficient statistic.

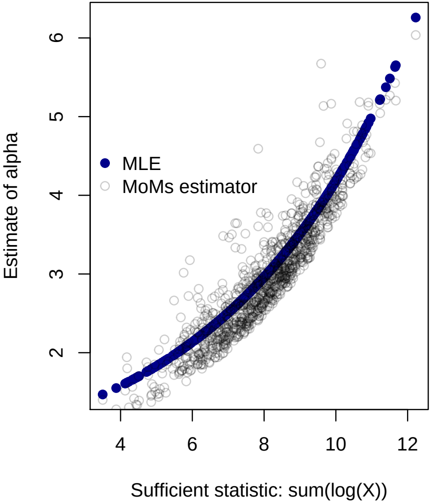
So MLEs use all the information in the sample about the target parameter.

Instructor: Show proof.

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

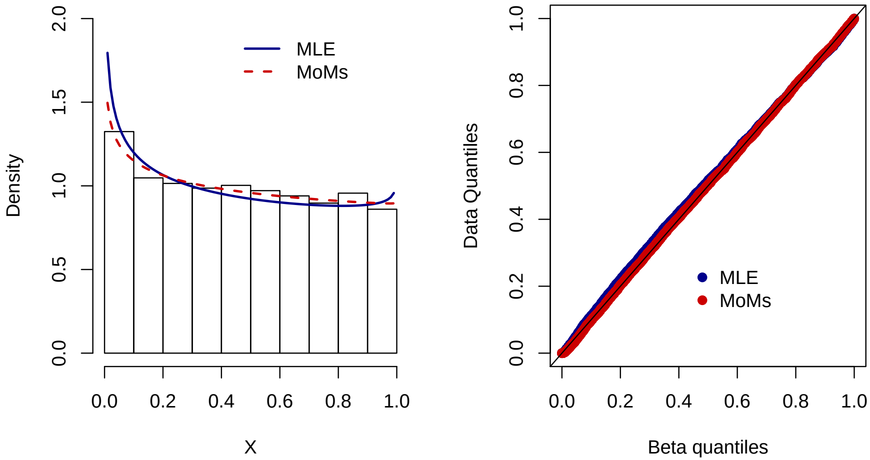
1. Find the MLE $\hat{\alpha}$ of α.
2. Plot many values of the pair $(\hat{\alpha}, \prod_{i=1}^n X_i)$ for simulated data under $\alpha = 3$.
3. Compare the MSE of $\hat{\alpha}$ to that of the MoMs estimator $\bar{\alpha}$.


Estimate of alpha
6
5
4
3
2
MLE
MoMs estimator
0
50000
100000
200000
Sufficient statistic: prod(X)


Estimate of alpha
6
5
4
3
2
MLE
MoMs estimator
4
6
8
10
12
Sufficient statistic: sum(log(X))

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

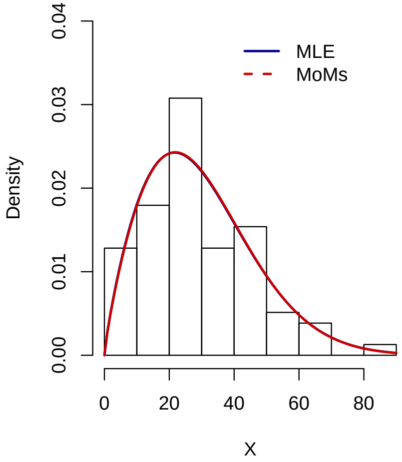
1. Find the MLEs $\hat{\alpha}$ and $\hat{\beta}$ of α and β.
2. Compute MLEs on prostate cancer p-values data from Efron (2012).
3. Compare to MoMs estimates.



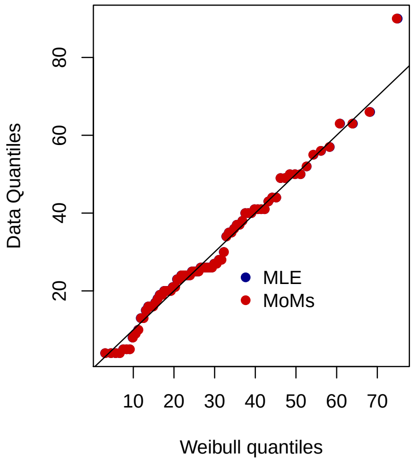
MLE
MoMs
Density
X
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
MLE
MoMs
Data Quantiles
Beta quantiles

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

1. Find the MLEs $\hat{a}$ and $\hat{b}$ of a and b.
2. Compute MLEs on trees in Camden data.
3. Compare to MoMs estimates.



MLE
MoMs
Density
0.00
0.01
0.02
0.03
0.04
0
20
40
60
80
X



MLE
MoMs
Data Quantiles
20
40
60
80
10
20
30
40
50
60
70
Weibull quantiles

Now consider estimating a function $\tau(\theta)$ of θ with ML.

Theorem

If $\hat{\theta}$ is the MLE for θ, then for any function τ, $\tau(\hat{\theta})$ is the MLE for $\tau(\theta)$.

Instructor: Prove the result.

Exercise: Let $\varepsilon_1, \ldots, \varepsilon_n$ 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 $X_1, \dots, X_n \overset{\text{ind}}{\sim}$ Exponential(λ).

1. Find the MLE of $\eta = 1/\lambda$.
2. Find the MLE of the median of the Exponential(λ) distribution.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Bernoulli}(p)$.

1. Find the MLE of p.
2. Find the MLE of $p(1-p)$.

In certain "nice" settings, MLEs are asymptotically Normal.

Theorem (Asymptotic distribution of MLEs)

Let $f_X(x;\theta)$ for $\theta \in \Theta \subset \mathbb{R}$ be a family of pdfs/pmfs. Suppose

(i) $X_1, \dots, X_n \overset{ind}{\sim} f_X(x;\theta_0)$, *so that* θ_0 *denotes the true value of* θ.

(ii) $\theta = \theta' \iff f_X(x;\theta) = f_X(x;\theta')$ *for all* $x \in \mathbb{R}$.

(iii) $f_X(x;\theta)$ *is differentiable with respect to* θ *for all* $x \in \mathbb{R}$.

(iv) θ_0 *is not on the boundary of* Θ.

(v),(vi) See Casella and Berger (2021), 2nd Ed, pg. 516.

Then the MLE $\hat{\theta}_n$ *satisfies*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} \text{Normal}(0, 1/\mathcal{I}(\theta_0)) \quad \text{as} \quad n \to \infty$$

where

$$\mathcal{I}(\theta_0) = \mathbb{E}[(\tfrac{\partial}{\partial\theta} \log f_X(X;\theta)|_{\theta=\theta_0})^2], \quad \text{with } X \sim f_X(x;\theta_0).$$

The quantity $\mathcal{I}(\theta_0)$ is called the *Fisher information*.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Exponential}(\lambda_0)$.

1. Get the Fisher information $\mathcal{I}(\lambda_0)$.
2. Make a statement about the asymptotic behavior of $\sqrt{n}(\hat{\lambda}_n - \lambda_0)$.

Exercise: Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} \text{Beta}(1, \theta_0)$.

1. Find the MLE for θ_0.
2. Get the Fisher information $\mathcal{I}(\theta_0)$.
3. Make a statement about the asymptotic behavior of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.
4. Run a simulation to check the coverage of the confidence intervals

$$\hat{\theta}_n \pm z_{\alpha/2} \cdot n^{-1/2} \cdot \mathcal{I}^{-1/2}(\theta_0) \quad \text{and} \quad \hat{\theta}_n \pm z_{\alpha/2} \cdot n^{-1/2} \cdot \mathcal{I}^{-1/2}(\hat{\theta}_n)$$

Coverages under $\alpha = 0.10$ over $1{,}000$ simulated data sets under $\theta = 10$,

n	2	4	8	16	32	64	128	256
$\mathcal{I}(\theta_0)$	0.76	0.84	0.86	0.88	0.90	0.91	0.89	0.89
$\mathcal{I}(\hat{\theta}_n)$	0.92	0.91	0.90	0.92	0.91	0.91	0.89	0.89

George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

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.