

Math 4261: Mathematical Statistics I

Lecture 9: Method of moments

Ref: Rice 8.4, 10.4; DeGroot–Schervish 7.6, 10.7

Fall 2026

Outline

Bias and mean-square error

Method of Moments

Bootstrap standard errors

Recap: Quantile estimation

For i.i.d. data with CDF F and $0 < \alpha < 1$, estimate the population quantile $q_\alpha = \inf\{x : F(x) \geq \alpha\}$ by the sample quantile

$$\hat{q}_{n,\alpha} = \widehat{F}_n^{-1}(\alpha) = X_{(\lceil n\alpha \rceil)}.$$

$\hat{q}_{n,\alpha}$ is the **first crossing** of level α by the empirical CDF $\widehat{F}_n$.

Consistency

If $F(q_\alpha - \epsilon) < \alpha < F(q_\alpha + \epsilon)$ for every $\epsilon > 0$, then $\hat{q}_{n,\alpha} \xrightarrow{p} q_\alpha$.

Sampling error

If F has a continuous density f near q_α with $f(q_\alpha) > 0$, then

$$\sqrt{n}(\hat{q}_{n,\alpha} - q_\alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f(q_\alpha)^2}\right).$$

Recap: Parameter estimation and bias

Model: $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} f(x \mid \theta)$, where $\theta \in \Omega \subseteq \mathbb{R}^k$ is unknown and Ω is the **parameter space**.

An **estimator** $\hat{\theta} = T(X_1, \dots, X_n)$ is a function of the data, so it is random. For a scalar parameter, its bias is

$$\text{Bias}_{\theta}(\hat{\theta}) = \mathbb{E}_{\theta}[\hat{\theta}] - \theta.$$

It is **unbiased** if $\mathbb{E}_{\theta}[\hat{\theta}] = \theta$ for **every** $\theta \in \Omega$.

Example: Gaussian variance

For i.i.d. $\mathcal{N}(\mu, \sigma^2)$ data, the empirical variance satisfies

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad \mathbb{E}[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2.$$

Its bias is $-\sigma^2/n$: it underestimates σ^2 on average.

Outline

Bias and mean-square error

Method of Moments

Bootstrap standard errors

Nonlinear transformations of estimators

Suppose we know that the estimator T is unbiased for some parameter θ so $\mathbb{E}[T] = \theta$.

Suppose we want to estimate $g(\theta)$ for some function g . In general, a **nonlinear** transformation g **need not preserve unbiasedness**:

$$\mathbb{E}[g(T)] \neq g(\mathbb{E}[T])$$

Example: Suppose $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$. The sample mean is unbiased for μ : $\mathbb{E}[\overline{X}_n] = \mu$.

To estimate μ^2 , consider the **plug-in** estimator $T_n = \overline{X}_n^2$. Then

$$\mathbb{E}[T_n] = \text{Var}[\overline{X}_n] + (\mathbb{E}[\overline{X}_n])^2 = \frac{\sigma^2}{n} + \mu^2.$$

Nevertheless, $\overline{X}_n^2 \xrightarrow{p} \mu^2$ by the LLN and continuous mapping theorem. Notice that **bias and consistency are different questions**.

Example: Sample standard deviation

Suppose $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$, with $n \geq 2$ and $\sigma > 0$. Correct the earlier empirical variance $\hat{\sigma}^2$ by defining

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

Then $\mathbb{E}[S^2] = \sigma^2$.

Does taking the square root give an unbiased estimator of σ ?

Let $S = \sqrt{S^2}$. The square-root function is **strictly concave**.

Jensen's inequality

For concave g , $\mathbb{E}[g(Y)] \leq g(\mathbb{E}[Y])$. If g is **strictly concave**, equality holds if and only if Y is constant almost surely.

Hence

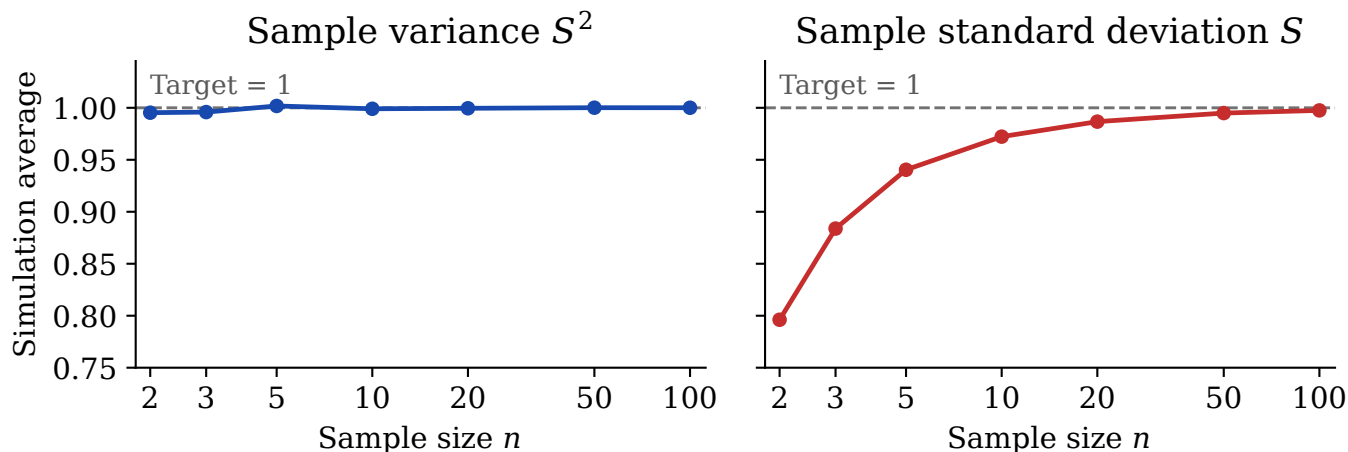
$$\mathbb{E}[S] = \mathbb{E}[\sqrt{S^2}] < \sqrt{\mathbb{E}[S^2]} = \sigma.$$

Strict because S^2 is nonconstant. S^2 is unbiased for the variance, but S is **biased** (downward) for the standard deviation.

Example: Sample standard deviation

Suppose $X_1, \dots, X_n$ are i.i.d. $\mathcal{N}(0, 1)$. Population $\sigma^2 = \sigma = 1$.
Compute S^2 and $S = \sqrt{S^2}$. Repeat **200,000 times** for each n .

200,000 independent samples from $N(0, 1)$ at each n



The average of S is **below 1**, especially for small samples¹.

¹Possible to work out analytically that $\mathbb{E}[S] = c_n \sigma$, where $c_n \rightarrow 1$

Bias-Variance decomposition

The mean-square error (MSE) of an estimator $\hat{\theta}$ is [Let θ true param, $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ estimator.]

$$\mathbb{E}_{\theta}[(\hat{\theta} - \theta)^2]$$

It can be decomposed as $\text{MSE} = \text{bias}^2 + \text{variance}$.

$$\begin{aligned} \text{MSE} &= \mathbb{E}[(\theta - \hat{\theta})^2] \\ &= \mathbb{E}[(\underbrace{\theta - \mathbb{E}[\hat{\theta}]}_{\text{a deterministic}} + \underbrace{\mathbb{E}[\hat{\theta}] - \hat{\theta}}_{b \text{ has randomness}})^2] \\ &= a^2 + a \mathbb{E}[b] + \mathbb{E}[b^2] \\ &= (\theta - \mathbb{E}[\hat{\theta}])^2 + \underbrace{(\theta - \mathbb{E}[\hat{\theta}]) \mathbb{E}[\hat{\theta} - \mathbb{E}[\hat{\theta}]]}_{=0} + \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] \\ &= \text{bias}^2(\hat{\theta}) + \text{Var } \hat{\theta}. \end{aligned}$$

Possible to be biased but still have good MSE.

Bias, variance, and MSE generally depend on the true parameter value θ : the same estimator can be more accurate at some values than others.

Example: more bias? (Wasserman, Example 12.3)

Flip a coin independently four times, with unknown heads probability p . Let Y be the number of heads, so $Y \sim \text{Binomial}(4, p)$. Compare the estimators

$$\hat{p} = \frac{Y}{4}, \quad \tilde{p} = \frac{Y+1}{6}.$$

1. Find the bias, variance, and MSE of each estimator.
2. Which has **smaller** MSE when $p = 1/4$? Is it better for every p ?

$$\mathbb{E} \hat{p} = \mathbb{E} \frac{Y}{4} = p, \quad \mathbb{E} \tilde{p} = \mathbb{E} \frac{Y+1}{6} = \frac{4p+1}{6}$$

$$\text{Bias}(\hat{p}) = 0, \quad \text{Bias}(\tilde{p}) = \frac{4p+1}{6} - p = \frac{1-2p}{6}$$

$$\begin{aligned} \text{Variance: } \text{Var} \hat{p} &= \frac{1}{16} 4p(1-p) & \text{Var} \tilde{p} &= \frac{1}{36} 4p(1-p) \\ &= \frac{1}{4} p(1-p) & &= \frac{1}{9} p(1-p) \end{aligned}$$

Example: more bias? (Wasserman, Example 12.3)

$$MSE(\hat{p}) = \text{bias}^2 + \text{Variance} = 0^2 + \frac{1}{4} p(1-p)$$

$$MSE(\tilde{p}) = \left(\frac{1-2p}{6}\right)^2 + \frac{1}{9} p(1-p) = \frac{1-4p+4p^2}{36} + \frac{4p-4p^2}{36} = \frac{1}{36}$$

$$\text{At } p = \frac{1}{4}, \quad MSE(\hat{p}) = \frac{3}{64} > MSE(\tilde{p}) = \frac{1}{36} = \frac{3}{108}$$

At $p \approx 0$ or $p \approx 1$, $MSE(\hat{p})$ is better.

(We will encounter this $\tilde{p}$ later in Bayesian approach)

Outline

Bias and mean-square error

Method of Moments

Bootstrap standard errors

Sample moments

Recall that the k^{th} moment of a r.v. X is

$$\mu_k = \mathbb{E}[X^k].$$

Given data $X_1, \dots, X_n$, define the k^{th} sample moment to be

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$$

Think of $\hat{\mu}_k$ as estimating μ_k . Note $\hat{\mu}_k \xrightarrow{p} \mu_k$ by LLN

Procedure

Method of Moments (MoM)

Setup: Given data $X_1, \dots, X_n$ assumed to come i.i.d. from a distribution with parameter vector θ .

- (1) **Compute moments** μ_1, μ_2 , expressing them in terms of the parameters θ . Typically only need as many moments as parameters.
- (2) **Invert the expressions** found in (1), expressing the parameters in terms of the moments.

$$\theta_1 = f_1(\mu_1, \mu_2)$$

$$\theta_2 = f_2(\mu_1, \mu_2)$$

- (3) **Substitute sample moments** for moments in (2) to obtain parameter estimates

$$\hat{\theta}_1 = f_1(\hat{\mu}_1, \hat{\mu}_2)$$

$$\hat{\theta}_2 = f_2(\hat{\mu}_1, \hat{\mu}_2).$$

Example: Poisson

Develop the MoM estimator for $X_1, \dots, X_n$ assumed to come i.i.d. from $\text{Poisson}(\lambda)$.

① Let $X \sim \text{Poisson}(\lambda)$. Then

$$\mu_1 = \mathbb{E}X = \lambda$$

② Thus parameter $\lambda = \mu_1$

③ MoM estimate $\hat{\lambda} = \hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n X_i$

Example: Poisson

Steel et al. (1980) counted asbestos fibers on filters to develop measurement standards for asbestos concentration. An operator counted the **number of fibers** on each of 23 grid squares giving counts:

31	29	19	18	31	28
34	27	34	30	16	18
26	27	27	18	24	22
28	24	21	17	24	

Let's try fitting a **Poisson** to model the counts in each grid square. Such a model will indicate the inherent variability in future measurements.

Write down a model and give the MoM estimate.

That is, our model is:

*$X_i = \# \text{ counts in each cell, } i=1, \dots, 23$
are iid $\text{Poisson}(\lambda)$*

Example: Poisson

If we repeat the experiment, the counts will change, so $\hat{\lambda}$ will change.
How **stable** is the MoM estimate?

□ Let $S = X_1 + \dots + X_n$, $X_i \stackrel{iid}{\sim} \text{Pois}(\lambda)$
MoM est. $\hat{\lambda} = \frac{S}{n}$ is a random variable.
randomness is sampling distrib.

Recall:

$$\mathbb{E} \hat{\lambda} = \lambda$$
$$\text{Var} \hat{\lambda} = \frac{\lambda}{n}$$

Often, this is all we really know. The
sampling dist. (dist. of estimator) may be complicated
in general.

Example: Poisson

If we repeat the experiment, the counts will change, so $\hat{\lambda}$ will change.
How **stable** is the MoM estimate?

[2] Here, sampling dist. can actually be found

$$\mathbb{E} \hat{\lambda} = \lambda, \quad \text{Var} \hat{\lambda} = \frac{\lambda}{n}$$

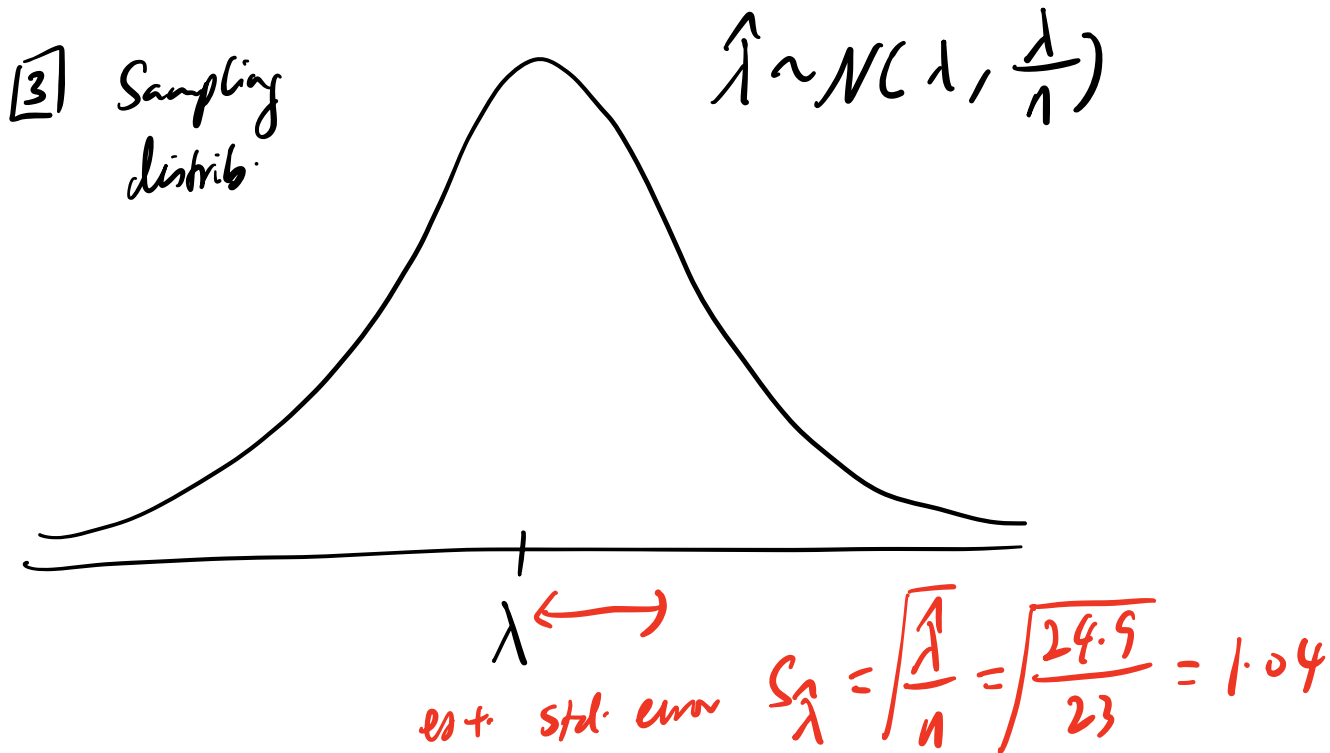
Fact: $\sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda)$. So $S \sim \text{Poisson}(n\lambda)$
 $\mathbb{E} S = \text{Var} S = n\lambda$

By CLT, $S \sim \mathcal{N}(n\lambda, n\lambda)$ approximately. (hw 2!)

Hence $\hat{\lambda} = \frac{S}{n}$ approx. normal too!

Example: Poisson

If we repeat the experiment, the counts will change, so $\hat{\lambda}$ will change.
How **stable** is the MoM estimate?



Expect error to be mostly $\pm 2s_{\hat{\lambda}} = \pm 2.08$

Example: Poisson

Summary

In this Poisson model $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Pois}(\lambda)$. We found the MoM $\hat{\lambda} = \bar{X}$.

For $X_i \sim \text{Pois}(\lambda)$, we have $\mathbb{E}[X_i] = \text{Var}[X_i] = \lambda$. So the sample mean satisfies

$$\mathbb{E}_\lambda[\hat{\lambda}] = \lambda, \quad \text{and} \quad \text{Var}_\lambda[\hat{\lambda}] = \frac{\lambda}{n}.$$

In particular $\hat{\lambda}$ is unbiased for λ . The **MSE** is $\frac{\lambda}{n}$.

For these data, $\hat{\lambda} = 573/23 \approx 24.91$, so the **plug-in standard error** is

$$\sqrt{\frac{\hat{\lambda}}{23}} \approx 1.04.$$

Example: Poisson model standard error $\sqrt{\widehat{\text{Var}} \hat{\lambda}}$

We wish to report the standard error of $\hat{\lambda}$. Since the true standard error $\sqrt{\lambda/n}$ depends on λ , which is unknown, we typically report a plug-in estimate $\sqrt{\hat{\lambda}/n}$ for this standard error.

Should we further account for the uncertainty of *this estimate* $\sqrt{\hat{\lambda}/n}$?

We usually don't, because this additional error is much smaller than the standard error itself for large sample sizes n . For fixed $\lambda > 0$, $\hat{\lambda} - \lambda$ has typical scale $1/\sqrt{n}$. A Taylor expansion then shows that

$$\sqrt{\frac{\hat{\lambda}}{n}} - \sqrt{\frac{\lambda}{n}} \quad \text{has typical scale} \quad \frac{1}{n}.$$

For example: If $n = 100$ and we estimate $\hat{\lambda} = 1$, we may report the standard error as $\sqrt{\hat{\lambda}/n} = 0.1$. The difference between this and the true standard error should be on the scale of $1/n = 0.01$, which is small compared to our reported standard error of 0.1.

Estimating the standard error

$$\hat{\lambda} = \bar{X}, \quad \text{Var } \hat{\lambda} = \frac{\lambda}{n}$$

$$\sqrt{\frac{\hat{\lambda}}{n}} - \sqrt{\frac{\lambda}{n}} \quad \text{has typical scale } \frac{1}{n}.$$

$$\left. \begin{aligned} f(x) &= \sqrt{x}, \quad f'(x) = \frac{1}{2\sqrt{x}}, \quad f''(x) = -\frac{1}{4x^{3/2}} \\ f(x) &= f(a) + (x-a)\frac{1}{2\sqrt{a}} + O\left(\frac{(x-a)^2}{a^{3/2}}\right) \end{aligned} \right\} \quad \begin{aligned} \text{set } x &= \frac{\hat{\lambda}}{n} \\ a &= \frac{\lambda}{n} \end{aligned}$$

$$\sqrt{\frac{\hat{\lambda}}{n}} = \sqrt{\frac{\lambda}{n}} + \underbrace{\frac{\hat{\lambda}-\lambda}{n} \cdot \frac{1}{2\sqrt{\lambda/n}}}_{\text{std. dev.}} + \underbrace{O\left(\left(\frac{\hat{\lambda}-\lambda}{n}\right)^2 \cdot n^{3/2}\right)}_{\text{typically } \frac{1}{n} \text{ (fluctuations of } \hat{\lambda}-\lambda)^2}$$

$$\text{std. dev.} : \frac{1}{2\sqrt{n\lambda}} \text{SD}(\hat{\lambda}-\lambda)$$

$$= \frac{1}{2\sqrt{n\lambda}} \sqrt{\lambda} = \frac{1}{2n}$$

$$\begin{aligned} &\text{typically } \frac{1}{n} \text{ (fluctuations of } \hat{\lambda}-\lambda)^2 \\ &= \frac{1}{n} \left(\frac{1}{n}\right)^2 = \frac{1}{n^{3/2}} \end{aligned}$$

$$\text{SD}[aX+b] = |a| \text{SD}[X]$$

Example

Develop the MoM estimator for $X_1, \dots, X_n$ assumed to come i.i.d. from $\mathcal{N}(\mu, \sigma^2)$.

$$\begin{aligned} (1) \quad \mu_1 &= \mathbb{E}X = \mu \\ \mu_2 &= \mathbb{E}[X^2] = \text{Var}X + (\mathbb{E}X)^2 = \sigma^2 + \mu_1^2 \end{aligned}$$

$$(2) \text{ Invert } \quad \begin{aligned} \mu &= \mu_1 \\ \sigma^2 &= \mu_2 - \mu_1^2 \end{aligned}$$

$$\begin{aligned} (3) \text{ MoM est. } \quad \hat{\mu} &= \hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \\ \hat{\sigma}^2 &= \hat{\mu}_2 - \hat{\mu}_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2 \\ &\vdots \\ &= \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (\text{biased}) \end{aligned}$$

Consistency of estimators

Let $\hat{\theta}_n$ be an estimator of a parameter θ based on a sample of size n . $\hat{\theta}_n$ is a **consistent estimator** of θ if for any $\epsilon > 0$,

$$\mathbb{P} [|\hat{\theta}_n - \theta| \geq \epsilon] \xrightarrow{n \rightarrow \infty} 0. \quad \text{i.e. } \hat{\theta}_n \xrightarrow{p} \theta$$

Consistency of MoM

Under reasonable conditions, the method of moments estimator is consistent.²

Example: (Bernoulli model) If $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(\theta)$ then MoM $\hat{\theta}_n = \bar{X}_n$ (check this). Hence by LLN

$$\hat{\theta}_n \xrightarrow{p} \theta.$$

²Wasserman, *All of Statistics* (2004), §9.2, Theorem 9.6(2), p. 122.

Consistency of estimators

For the Gaussian model we just used, the MoM estimators are

$$\hat{\mu} = \overline{X}_n, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \overline{X}_n^2.$$

1. The **LLN** estimates the population moments.

$$\overline{X}_n \xrightarrow{p} \mu, \quad \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} \mathbb{E}[X_1^2] = \mu^2 + \sigma^2.$$

2. **Continuous mapping** gives parameter consistency.

$$\hat{\mu} \xrightarrow{p} \mu, \quad \hat{\sigma}^2 \xrightarrow{p} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.$$

Note: the empirical variance is **biased at finite n** , but **consistent**.

The same idea applies to other MoM estimators when the sample moments converge and the formulas recovering the parameters are continuous at the population moments.

Example: Sampling from Uniform

Develop the MoM estimator for $\text{Unif}[\theta, \theta + 1]$.

What happens when we observe a sample of size three:

$X_1 = 0.2, X_2 = 0.99, X_3 = 0.01$ (so $\bar{X} = 0.4$)?

① Let $X \sim \text{Unif}[\theta, \theta + 1]$

$$\mu_1 = EX = \theta + \frac{1}{2}$$

② Invert $\theta = \mu_1 - \frac{1}{2}$

③ MoM $\hat{\theta} = \hat{\mu}_1 - \frac{1}{2} = \bar{X} - \frac{1}{2}$

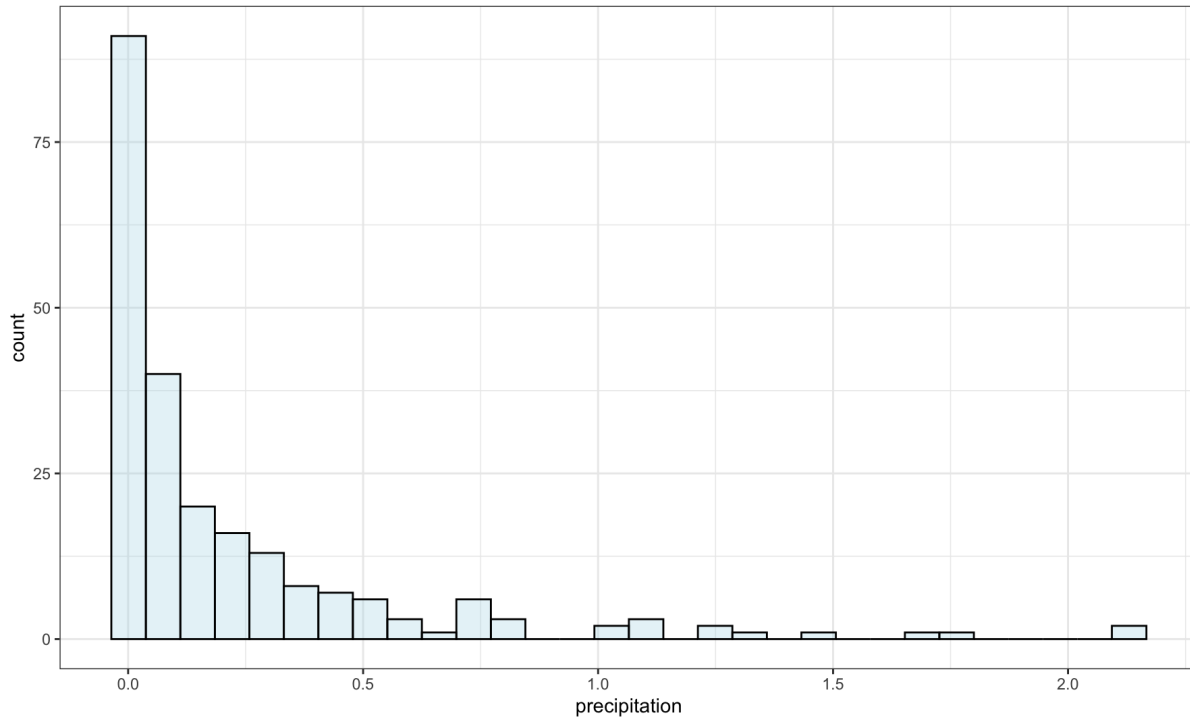
If $\bar{X} = 0.4$, $\hat{\theta} = -0.1$, implies $X_i \stackrel{\text{iid}}{\sim} \text{Unif}[-0.1, 0.9]$

impossible b/c $X_2 = 0.99$

Thus: need more samples.

Example

In a study of the natural variability of rainfall, the **rainfall of summer storms** was measured by a network of rain gauges in southern Illinois for the years 1960–1964 (Changnon and Huff, Le Cam and Neyman 1967).



Example

We are going to fit the **Gamma distribution** to the data. This has two parameters α and λ :

$$f(x|\alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x}, \quad 0 < x < \infty$$

where Γ is the gamma function. The first two moments are

$$\mu_1 = \frac{\alpha}{\lambda}, \quad \mu_2 = \frac{\alpha(\alpha + 1)}{\lambda^2}.$$

Develop the MoM estimator.

Example

(1)

$$\mu_1 = \frac{\alpha}{\lambda}, \quad \mu_2 = \frac{\alpha(\alpha + 1)}{\lambda^2}.$$

(2) Invert: From (b.) $\mu_2 = \frac{\alpha^2}{\lambda^2} + \frac{\alpha}{\lambda^2} = \mu_1^2 + \frac{\mu_1}{\lambda}$

$$\text{So } \lambda = \frac{\mu_1}{\mu_2 - \mu_1^2}$$

From (a.)

$$\alpha = \lambda \mu_1 = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

(3)

MOM

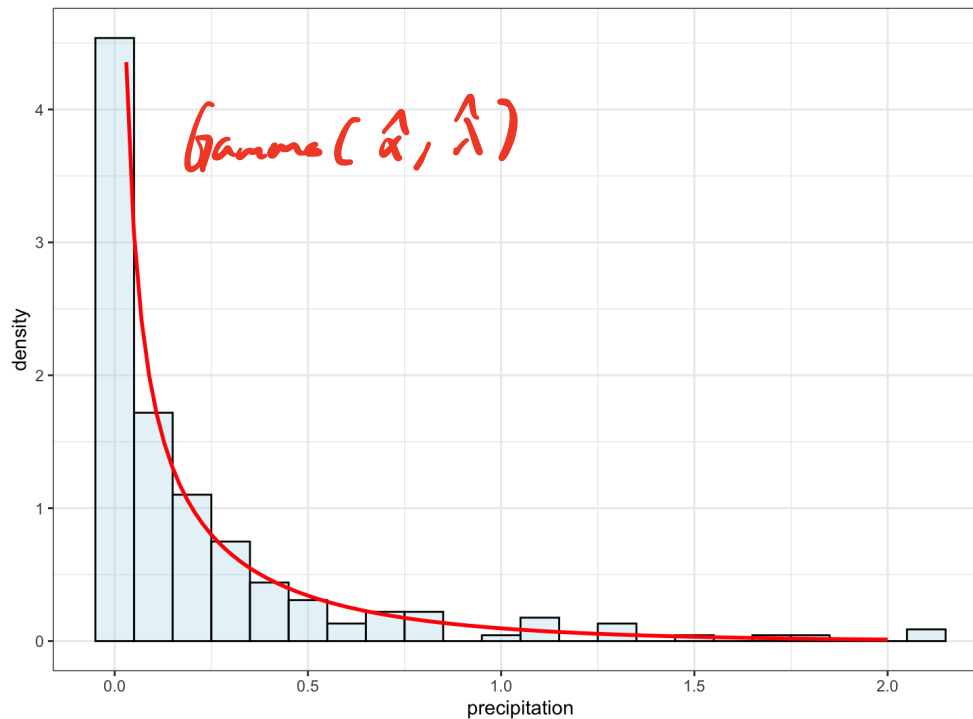
$$\hat{\lambda} = \frac{\hat{\mu}_1}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}}{\hat{\sigma}^2}$$

$$\hat{\alpha} = \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1^2} = \frac{\bar{X}^2}{\hat{\sigma}^2}$$

$$[\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2]$$

Example

Here $\hat{\alpha} = 0.375$, $\hat{\lambda} = 1.674$



Outline

Bias and mean-square error

Method of Moments

Bootstrap standard errors

Example: Rainfall in Illinois–Bootstrapping

Motivation:

After estimating the parameters $\hat{\alpha}$ and $\hat{\lambda}$ from the data, we want to understand their sampling distributions.

- ▶ This is the distribution of the random variables $\hat{\alpha}$ and $\hat{\lambda}$, which are functions of the sample $X_1, \dots, X_{227}$.
(Note: this is **not** the distribution of X_i 's, which we already assumed to be Gamma).
- ▶ Allows us to understand variability in the estimates and estimate their standard errors.

Option 1: work out theoretically the sampling distribution (may be difficult depending on model).

(Poisson asbestos example)

Option 2: Parametric bootstrap. Pretend the true parameters were $\hat{\alpha}$ and $\hat{\lambda}$ and do many simulations of the dataset to get an indication of the variability inherent in estimating parameters from a sample of this size.

Example: Rainfall in Illinois–Bootstrapping

Parametric bootstrap

1. Draw $X_1^{(1)}, \dots, X_{227}^{(1)}$ i.i.d. from the fitted $\text{Gamma}(\alpha = 0.375, \lambda = 1.674)$
2. Compute the MoM estimates $\lambda^{(1)}$ and $\alpha^{(1)}$ and record them.
3. Repeat with $X_1^{(2)}, \dots, X_{227}^{(2)}$ to get $\lambda^{(2)}$ and $\alpha^{(2)}$, and so on for 1000 times.

Use the 1000 parameter estimates to compute **estimated standard errors** for say $\hat{\alpha}$:

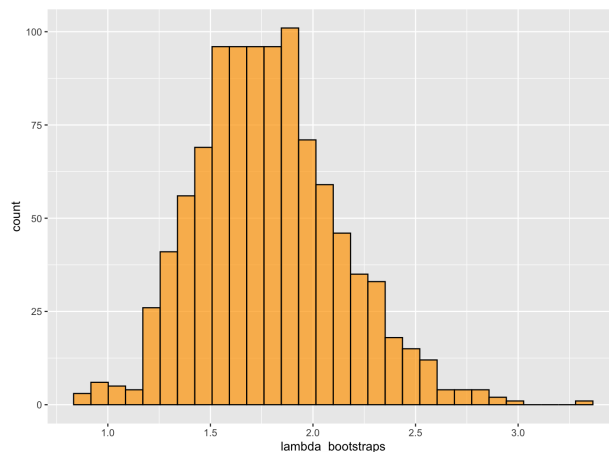
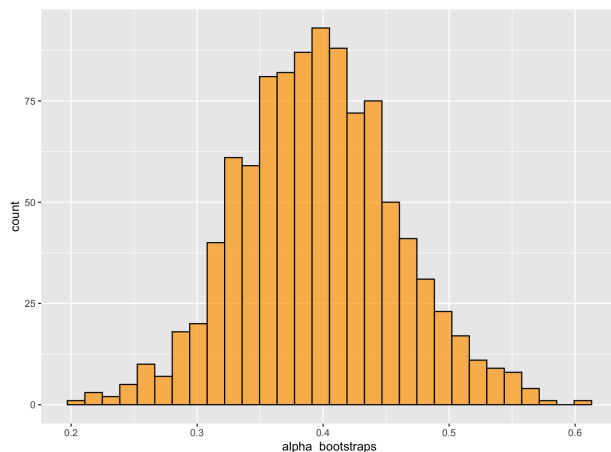
$$s_{\hat{\alpha}} = \sqrt{\frac{1}{999} \sum_{j=1}^{1000} (\alpha^{(j)} - \bar{\alpha})^2}, \quad \text{where} \quad \bar{\alpha} = \frac{1}{1000} \sum_{j=1}^{1000} \alpha^{(j)}$$

Similarly for $\hat{\lambda}$.

Example: Rainfall in Illinois–Bootstrapping

The **bootstrap estimates** for the standard errors were found to be $s_{\hat{\alpha}} = 0.06$ and $s_{\hat{\lambda}} = 0.34$.

Histograms suggest **Gaussian approximation** is appropriate.



For instance, if the true α were 0.375, a rough two-standard-error range for its estimate would be 0.375 ± 0.12 .

Summary

- ▶ The MSE is a measure of the accuracy of an estimator. It decomposes into bias^2 plus variance. This indicates potential **tradeoffs**.
- ▶ Method of moments constructs estimators by **matching sample moments to the true model moments**.
- ▶ In many cases, the LLN and continuous mapping theorem yield **consistency of MoM estimators**.
- ▶ **Parametric bootstrap** estimates standard errors by simulating from the fitted model and recomputing the estimates.

