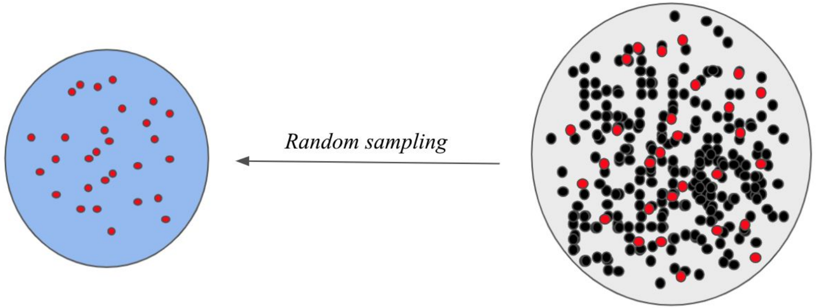
Sampling distributions of MM estimators

Chapter 8 of Rice - Method of Moments

06/24/2021

Berkeley
UNIVERSITY OF CALIFORNIA

In the previous lecture,



- Unbiased estimators:

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 \quad \textit{is an } \underline{\textit{unbiased estimator}} \textit{ of } \sigma^2.$$

- Central limit theorem under i.i.d assumption:
  No matter what the population distribution is, the sampling distribution of $\bar{X}_n$ *will always converge to a standard Normal distribution.*
- Confidence interval for μ:
  - Conservative CI
  - Bootstrap CI
- Method of moments estimators:
  - Oldest but very simple estimators;
  - Can give unrealistic estimations.

Method of moments

If unknown parameters $\theta_1, \ldots, \theta_k$ are not exactly moments of the population, they can be the solutions of a system of equations.

Population

$$\begin{aligned} \mu &= g_1(\theta_1, \ldots, \theta_k) \\ \mu_2 &= g_2(\theta_1, \ldots, \theta_k) \\ &\vdots \\ \mu_k &= g_k(\theta_1, \ldots, \theta_k) \end{aligned}$$

Samples

$$\begin{aligned} \hat{\mu} &= g_1\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \\ \hat{\mu}_2 &= g_2\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \\ &\vdots \\ \hat{\mu}_k &= g_k\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \end{aligned}$$

Method of Moments (MM): Match the sample moments with their population counterparts.

$\underline{\mathrm{Gamma}(\alpha, \beta)}$

$$\mu = \frac{\alpha}{\beta}$$
$$\sigma^2 = \frac{\alpha}{\beta^2}$$

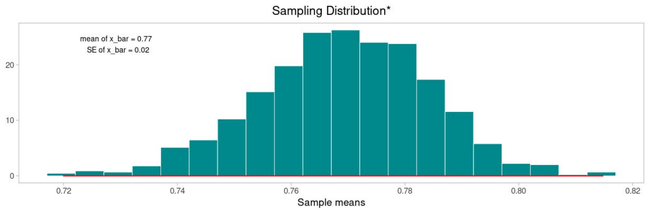
$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$
$$\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

What are the sampling distributions of $\hat{\alpha}_n$ and $\hat{\beta}_n$?

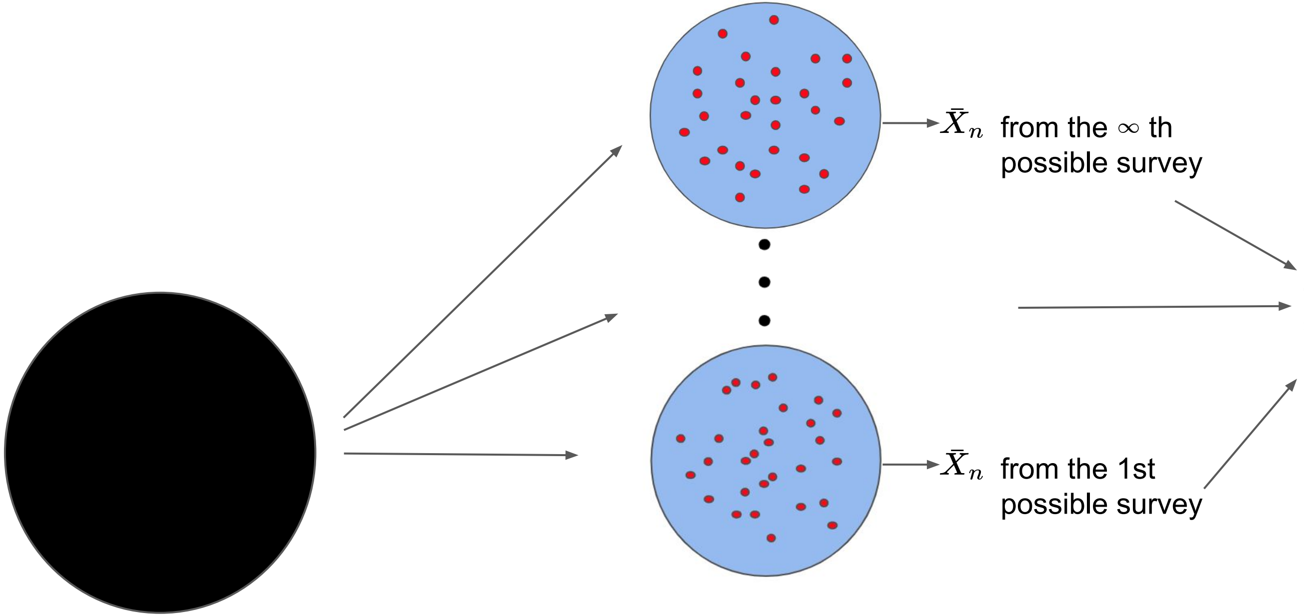
Sample distribution of $\bar{X}_n$

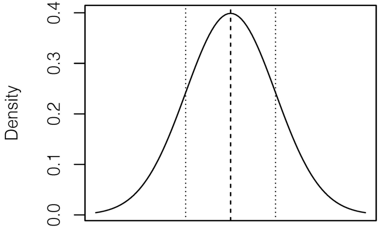
Population parameter is <u>known</u>

Simulation:
Draw as many samples of size n as possible, and plot the histogram of all $\bar{X}_n$.



Empirical





Theoretical

$\mathrm{Gamma}(2, 0.5)$

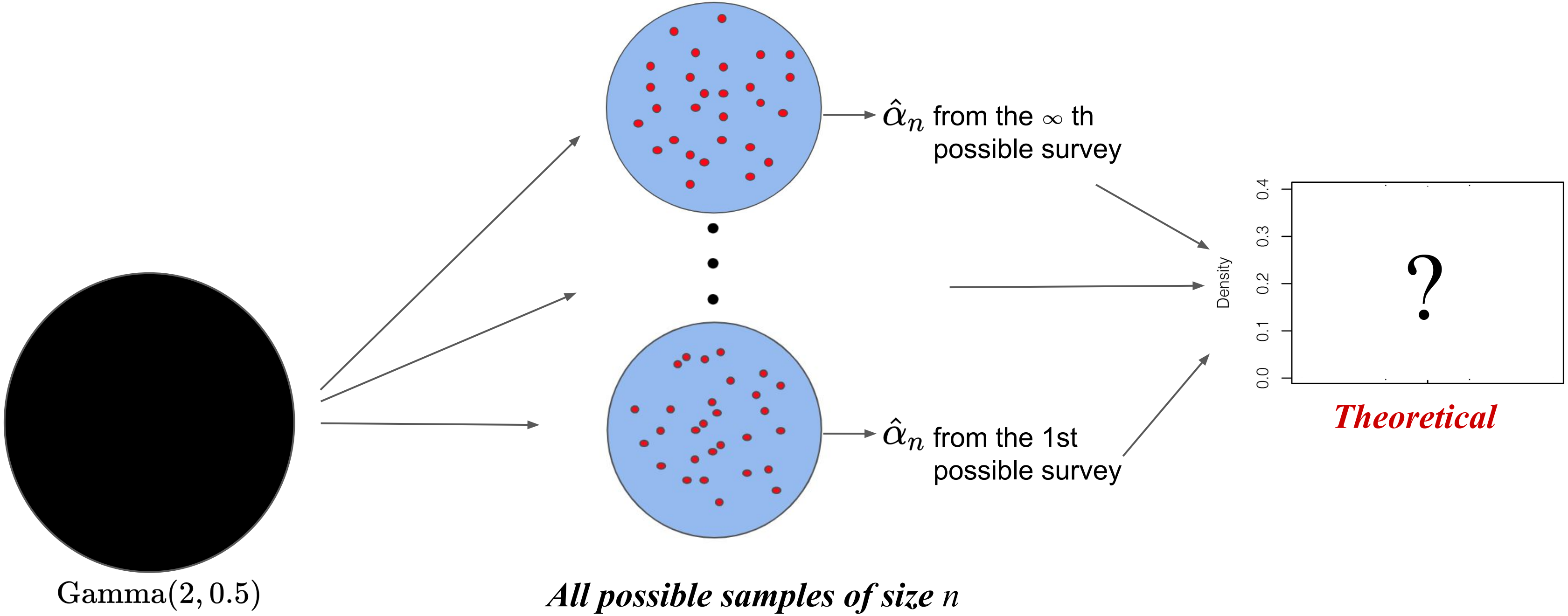
***All possible samples of size** n*

Sample distribution of $\hat{\alpha}_n$

Population parameter is *<u>known</u>*

$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

Simulation:
Draw as many samples of size n as possible, and plot the histogram of all $\hat{\alpha}_n$.

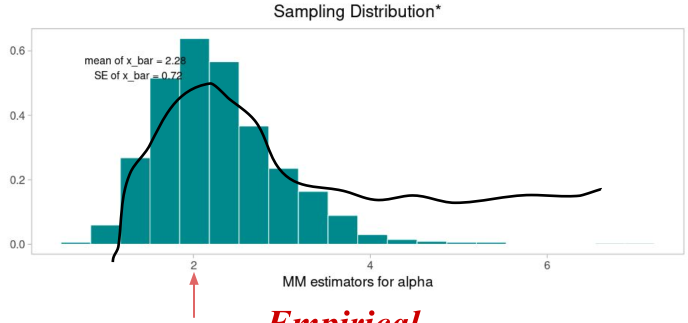


Sample distribution of $\hat{\alpha}_n$

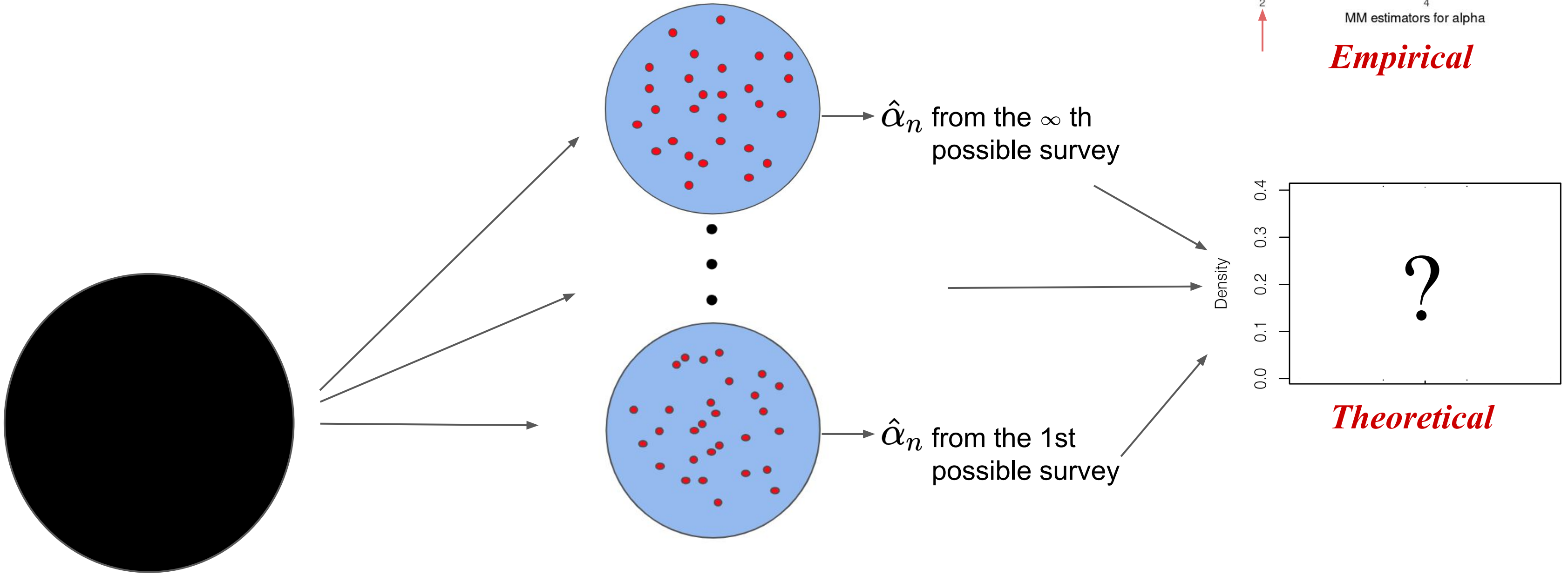
Population parameter is <u>known</u>

$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

Simulation:
Draw as many samples of size n as possible, and plot the histogram of all $\hat{\alpha}_n$.



Empirical



$\mathrm{Gamma}(2, 0.5)$

Theoretical

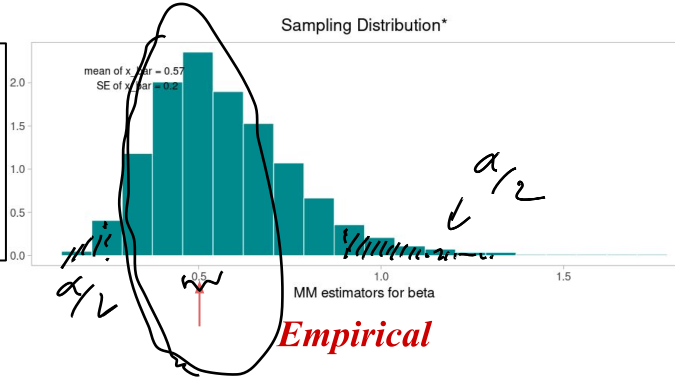
All possible samples of size n

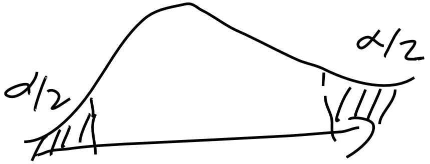
Sample distribution of $\hat{\beta}_n$

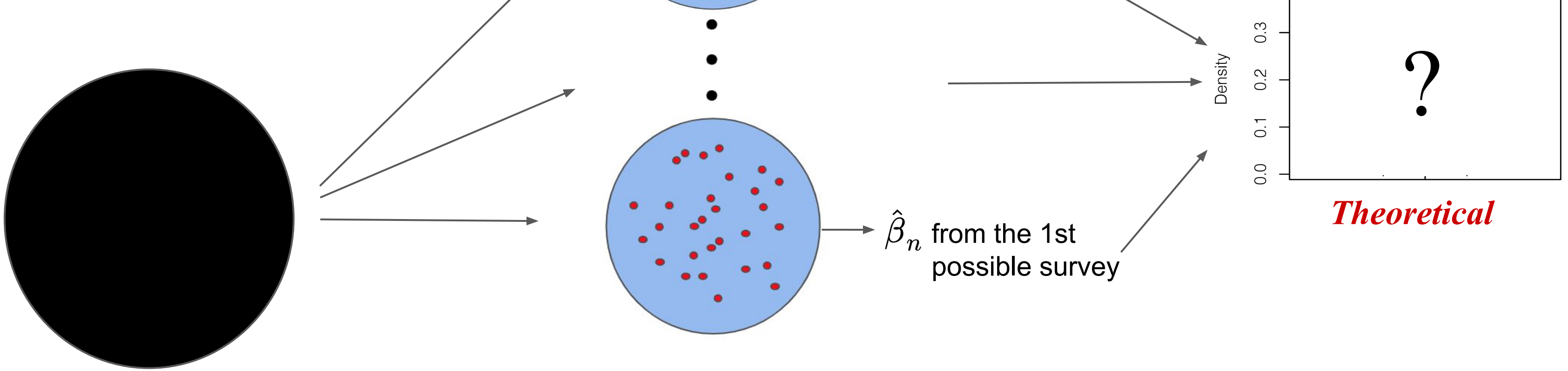
Population parameter is *__known__*

$$\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

Simulation:
Draw as many samples of size n as possible, and plot the histogram of all $\hat{\beta}_n$.







~~Gamma(2, 0.5)~~

All possible samples of size n

Sample distribution of $\hat{\alpha}_n$ and $\hat{\beta}_n$

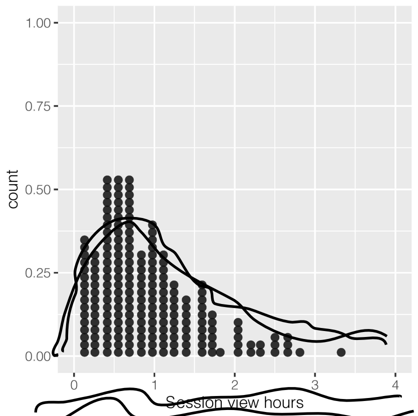
Population parameter is unknown

How to simulate the sampling distributions of $\hat{\alpha}_n$ and $\hat{\beta}_n$?

$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

$$\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

Sample mean = 0.916
Sample variance = 0.398



Design A

$n = 200$

$$\hat{\alpha}_n = 2.104$$
$$\hat{\beta}_n = 2.298$$

Bootstrap simulation:
Pretend that the population is Gamma(2.104, 2.298), and repeat the simulation process on Page 6 and 7.

Goodness of fit

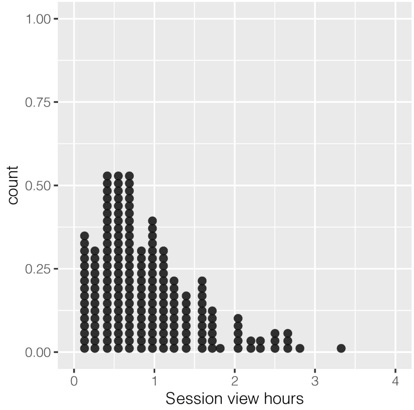
Sample distribution of $\hat{\alpha}_n$ and $\hat{\beta}_n$

Population parameter is *unknown*

$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

$$\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

Sample mean = 0.916
Sample variance = 0.398



Design A

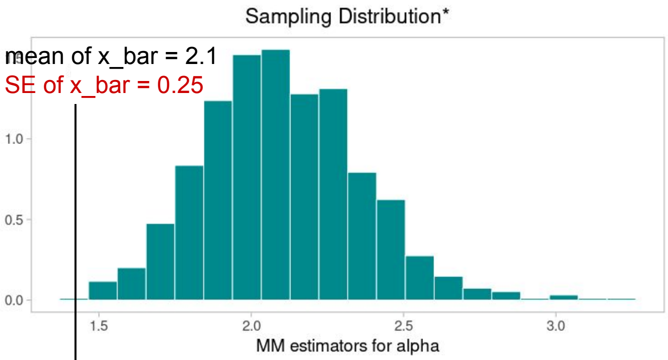
$n = 200$

$\hat{\alpha}_n = 2.104$

$\hat{\beta}_n = 2.298$

How to simulate the sampling distributions of $\hat{\alpha}_n$ and $\hat{\beta}_n$?

Bootstrap simulation: Pretend that the population is Gamma(2.104, 2.298), and repeat the simulation process on Page 6 and 7.



*Distribution of means of 1000 random samples, each consisting of 200 observations from a right skewed population

Standard error of the estimate

Sample distribution of MM estimators

Population parameter is unknown in $f(x \mid \theta)$

Bootstrap simulation:

- Construct a system of equations relating the parameters of interest to the sample moments;
- Solve the equations for the MM estimators $\hat{\theta}_{MM}$;
- Pretend that the population is $f\left(x \mid \hat{\theta}_{MM}\right)$, and repeat the simulation process similar to Page 6 and 7;
- Plot the histogram to approximate the sampling distribution of $\hat{\theta}_{MM}$, and calculate the standard error (SE) of $\hat{\theta}_{MM}$ from the sampling distribution.

Sample distribution of MM estimators

$SE(\hat{\theta}_{MM}) = \sqrt{Var(\hat{\theta}_{MM})}$

Population parameter is unknown in $f(x \mid \theta)$

If you only want to know the SE of $\hat{\theta}_{MM}$ ***, ...***

Bootstrap estimation:

- Construct a system of equations relating the parameters of interest to the sample moments;
- Solve the equations for the MM estimators $\hat{\theta}_{MM}$;
- Calculate $\mathrm{Var}(\hat{\theta}_{MM}) = h(\theta, n)$, which is difficult to calculate sometimes;
- Estimate the SE $\sqrt{h(\theta, n)}$ by $\sqrt{h\left(\hat{\theta}_{MM}, n\right)}$.

Sample distribution of MM estimators

Population parameter is unknown in $f(x \mid \theta)$

Example 5. Suppose the population follows a discrete distribution

X	0	1	2	3
Probability	$\frac{2}{3}\theta$	$\frac{1}{3}\theta$	$\frac{2}{3}(1-\theta)$	$\frac{1}{3}(1-\theta)$

$\bar{X}_n = \frac{3}{2}$

The following are i.i.d observations of this distribution: 3, 0, 2, 1, 3, 2, 1, 0, 2, 1. Find the Method of Moments estimate for θ, and approximate SE for this estimate.

$\mu = EX = 0 \times \frac{2}{3}\theta + 1 \times \frac{1}{3}\theta + 2 \times \frac{2}{3}(1-\theta) + \frac{3}{3}(1-\theta) = \frac{7}{3} - 2\theta.$

$\sigma^2 = EX^2 - (EX)^2$

$\hookrightarrow EX^2 = 0 \times \frac{2}{3}\theta + 1^2 \times \frac{1}{3}\theta + 2^2 \times \frac{2}{3}(1-\theta) + \frac{3^3}{3}(1-\theta)$

$= \frac{17 - 16\theta}{3}$

Thus $\sigma^2 = \frac{17-16\theta}{3} - \left(\frac{7}{3} - 2\theta\right)^2 := g(\theta)$

Sample distribution of MM estimators

Population parameter is unknown in $f(x \mid \theta)$

Example 5. Cont'd.

$\mu = EX = \frac{7}{3} - 2\theta$. To get the MM estimator for θ, solve

$$\hat{\mu} = \frac{7}{3} - 2\hat{\theta}_{MM}$$

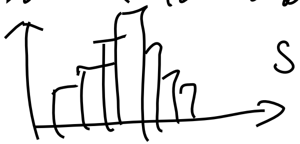
which will give: $\hat{\theta}_{MM} = \frac{1}{2}\left(\frac{7}{3} - \bar{X}_n\right) = \frac{1}{2}\left(\frac{7}{3} - \frac{3}{2}\right) = \frac{5}{12}$.

To get the SE for $\hat{\theta}_{MM}$, we can:

	0	1	2	3
prob	$\frac{2}{3}\times\frac{5}{12}$	$\frac{1}{3}\times\frac{5}{12}$	...	...

Conduct 1000 experiments, each experiment will contain 10 obs from

$\Rightarrow$ 1000 $\hat{\theta}_{MM}$'s



SE for $\hat{\theta}_{MM}$

$$var(\hat{\theta}_{MM}) = var\left[\frac{1}{2}\left(\frac{7}{3} - \bar{X}_n\right)\right]$$

$$= \frac{1}{4} var(\bar{X}_n) = \frac{1}{4}\frac{\sigma^2}{n}$$

$$= \frac{g(\theta)}{40} \approx \frac{g(\hat{\theta}_{MM})}{40} \quad \cdots \quad \|_2$$

$$SE(\hat{\theta}_{MM}) = \sqrt{\frac{g(\hat{\theta}_{MM})}{40}} = \frac{17 - 16\times\frac{5}{12}}{3} - \left(\frac{7}{3} - 2\times\frac{5}{12}\right)^2$$

Sample distribution of MM estimators

if $X \sim Exp(\lambda)$, $f(x|\lambda) = \lambda e^{-\lambda x}$

Population parameter is unknown in $f(x \mid \theta)$

Example 6. Suppose the population follows the exponential distribution $\mathrm{Exp}(\lambda)$. Find the MM estimator $\hat{\lambda}_{MM}$ and SE($\hat{\lambda}_{MM}$).

Solution. The population $\mu = E(X) = \frac{1}{\lambda}$, $var(X) = \frac{1}{\lambda^2}$.

$$\hat{\mu} = \frac{1}{\hat{\lambda}_{MM}} \Rightarrow \hat{\lambda}_{MM} = \frac{1}{\hat{\mu}} = \frac{1}{\overline{X}_n}.$$

Computer Simulation

$$var\left(\frac{1}{\overline{X}_n}\right) = var\left(\frac{n}{\underbrace{X_1 + \cdots + X_n}}\right)$$

$a N(0, b^2) = N(0, a^2 b^2)$ $\qquad \frac{1}{\bar{X}_n} \qquad \bar{X}_n^2 \qquad \sin \bar{X}_n$

Sample distribution of MM estimators

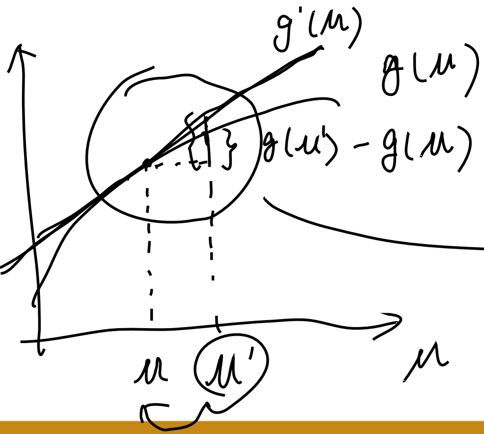
Population parameter is unknown in $f(x \mid \theta)$

Delta Method. Under the i.i.d assumption, $\sqrt{n}(\bar{X}_n - \mu) \to N(0, \sigma^2)$ as $n \to \infty$ by Central Limit Theorem. If given a differentiable function g, it follows that

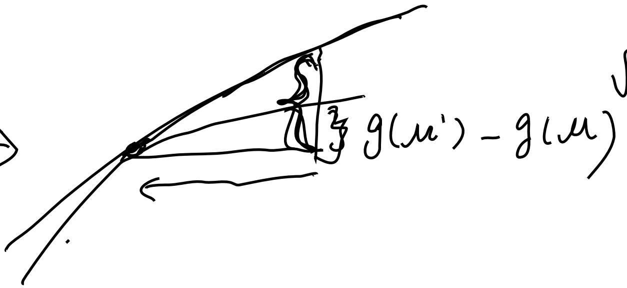
$$\sqrt{n}\{g(\bar{X}_n) - g(\mu)\} \to g'(\mu) N(0, \sigma^2) \text{ as } n \to \infty.$$

$g(\bar{X}_n) \approx N\left(g(\mu), \frac{[g'(\mu)]^2 \sigma^2}{n}\right)$

Something that immediately follows: $\operatorname{Var}[g(\bar{X}_n)] \approx \frac{[g'(\mu)]^2 \sigma^2}{n}$.



$g(\mu') - g(\mu) \approx g'(\mu)(\mu' - \mu)$



$\sqrt{n}(g(\bar{X}_n) - g(\mu)) = g'(\mu)\sqrt{n}(\bar{X}_n - \mu)$

$N(0, \sigma^2)$

Sample distribution of MM estimators

Population parameter is unknown in $f(x \mid \theta)$

$\mu = \frac{1}{\lambda}$

$b^2 = \frac{1}{\lambda^2}$

Example 6 *cont'd*. Suppose the population follows the exponential distribution $\mathrm{Exp}(\lambda)$. Find the MM estimator $\hat{\lambda}_{MM}$ and SE($\hat{\lambda}_{MM}$).

$\hat{\lambda}_{MM} = \frac{1}{\bar{X}_n}$ $\qquad g(\mu) = \frac{1}{\mu}$, $g'(\mu) = -\frac{1}{\mu^2}$

Therefore, $\mathrm{var}(\hat{\lambda}_{MM}) = \mathrm{var}\left(\frac{1}{\bar{X}_n}\right) = \frac{[g'(\mu)]^2 b^2}{n}$

$= \frac{b^2}{n\mu^4} = \frac{1/\lambda^2}{n\, 1/\lambda^4} = \frac{\lambda^2}{n}.$

$SE(\hat{\lambda}_{MM}) = \sqrt{\mathrm{var}(\hat{\lambda}_{MM})} = \frac{\lambda}{\sqrt{n}} \approx \frac{\hat{\lambda}_{MM}}{\sqrt{n}}.$

Bootstrap estimate.

The consistency of MM estimators

06/24/2021

$\varepsilon = 0.00001$

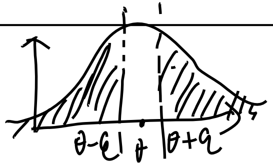
How good are the MM estimators?

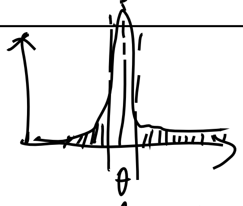
Definition. $\hat{\theta}_n = g(X_1, \ldots, X_n)$ is said to be a consistent estimator of θ if

$$\hat{\theta}_n \xrightarrow{p} \theta, \text{ as } n \to \infty.$$

That is, $\hat{\theta}_n$ converges in probability to θ.

Sampling distr of $\hat{\theta}_n$





degenerate

*Convergence in probability**:

200

$\uparrow$

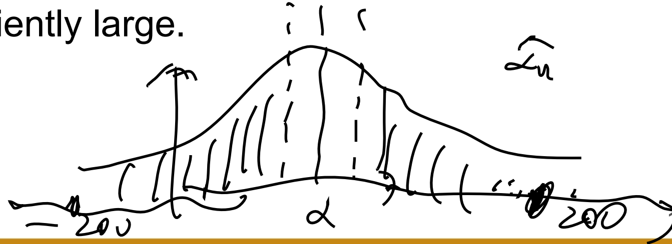
$$\text{For any } \epsilon > 0,\ P\left(\left|\hat{\theta}_n - \theta\right| > \epsilon\right) \to 0 \text{ as } n \to \infty.$$

$E(\hat{\theta}_n) = \theta$

Connection with unbiasedness:

$$\text{If } \hat{\theta}_n \xrightarrow{p} \theta, \text{ then } E\left(\hat{\theta}_n\right) \to \theta \text{ as } n \to \infty.$$

But consistency is more than the average behavior. It assures that the sampling distribution $\hat{\theta}_n$ of is **highly concentrated** around θ when the sample size is sufficiently large.



MM estimators are consistent!

Law of large number. Let $X_1, \ldots, X_n$ be i.i.d samples from a population $f(x)$. Then for any integer k,

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{p} E\left(X^k\right) = \mu_k \text{ as } n \to \infty.$$

Sample moments are consistent estimators of the population moments!

Continuous mapping theorem. For any continuous function $g(\cdot)$, if $\hat{\theta}_n \xrightarrow{p} \theta$, then

$$g\left(\hat{\theta}_n\right) \xrightarrow{p} g(\theta)$$

MM estimators are consistent!

Theorem E. Under the i.i.d assumption, MM estimators are consistent as long as functions relating the estimates to the sample moments are continuous.

solution

$$\boxed{\begin{aligned} \hat{\mu} &= g_1\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \\ \hat{\mu}_2 &= g_2\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \\ &\vdots \\ \hat{\mu}_k &= g_k\left(\hat{\theta}_1, \ldots, \hat{\theta}_k\right) \end{aligned}} \Rightarrow \begin{aligned} \hat{\theta} &= h_1(\hat{\mu}_1, \ldots, \hat{\mu}_k) \\ \hat{\theta}_2 &= h_2(\hat{\mu}_1, \ldots, \hat{\mu}_k) \\ &\vdots \\ \hat{\theta}_k &= h_k(\hat{\mu}_1, \ldots, \hat{\mu}_k) \end{aligned}$$

$\underline{\mathrm{Gamma}(\alpha, \beta)}$

$$\begin{aligned} \hat{\mu} &= \frac{\hat{\alpha}}{\hat{\beta}} \\ \hat{\mu}_2 &= \frac{\hat{\alpha}}{\hat{\beta}^2} + \frac{\hat{\alpha}^2}{\hat{\beta}^2} \end{aligned} \Rightarrow \begin{aligned} \hat{\alpha} &= \frac{\hat{\mu}^2}{\hat{\mu}_2 - \hat{\mu}^2} \\ \hat{\beta} &= \frac{\hat{\mu}}{\hat{\mu}_2 - \hat{\mu}^2} \end{aligned}$$

This justifies the bootstrap (plug-in) estimates.

The maximum likelihood estimators (MLE)

“What! You have solved it already?”
“Well, that would be too much to say. I have discovered a suggestive fact, that is all.”

Dr. Watson and Sherlock Holmes
The Sign of Four

06/24/2021

The disadvantages of MM estimators

- The consistency is a limiting result but we always only have finite sample size n;
- With small sample size n, it is not uncommon that MM estimators give unrealistic estimates.
- They sometimes fail to take into account all relevant information in the sample.

The disadvantages of MM estimators

Example 4. Suppose $X_1, \ldots, X_n$ are i.i.d samples from $X \sim$ Binomial(k, p). Find the Method of Moments estimators for k and p.

$$\hat{p}_{MM} = \frac{\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}{\bar{X}_n}, \quad < 0$$

$$\hat{k}_{MM} = \frac{\bar{X}_n^2}{\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2} \quad < 0$$

Example 3. Suppose $X_1, \ldots, X_n$ are i.i.d samples from $\mathrm{Gamma}(\alpha, \beta)$, find Method of Moments for *α* and *β.*

$$\mu = \frac{\alpha}{\beta} \qquad \sigma^2 = \frac{\alpha}{\beta^2}$$

$$\Rightarrow$$

$$\hat{\alpha}_n = \frac{\bar{X}_n^2}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

$$\hat{\beta}_n = \frac{\bar{X}_n}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2}$$

$$\begin{cases} \mu = \dfrac{\alpha}{\beta} \\ \mu_3 = \dfrac{\alpha(\alpha+1)(\alpha+2)}{\beta^3} \end{cases}$$

Does the third moment also contain information about α *and* β*?*

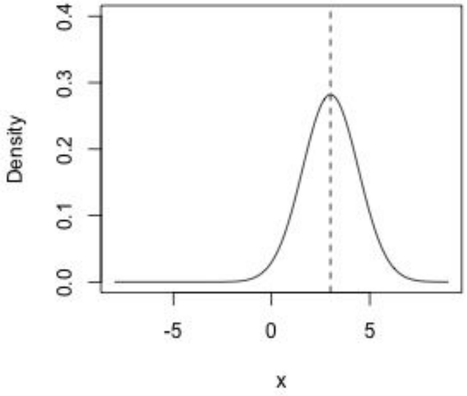
The joint density or frequency function

Let $X_1, \ldots, X_n$ be i.i.d samples from a population $f(x \mid \theta)$. The joint density of $X_1, \ldots, X_n$ can be written as

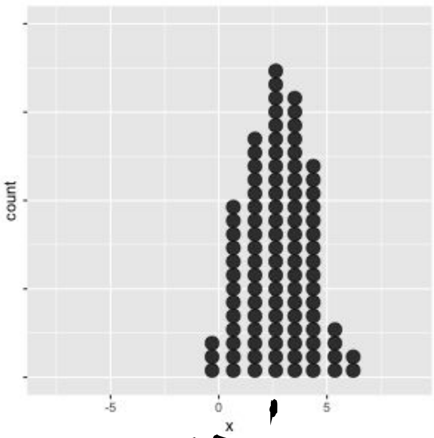
$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

If the population is a normal distribution, then the joint density is $f\left(x_1, \ldots, x_n \mid \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - \mu)^2}$.

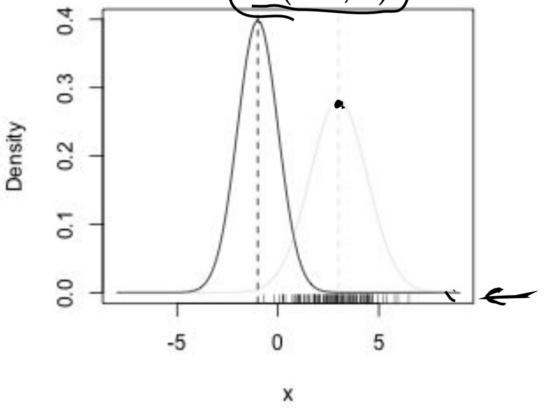
$X \sim N(3, 2)$



$\xrightarrow{n=100}$



$X \sim N(-1, 1)$



$f(x_1, \ldots, x_{100} \mid 3, 2)$ or $f(x_1, \ldots, x_{100} \mid -1, 1)$?

The joint density or frequency function

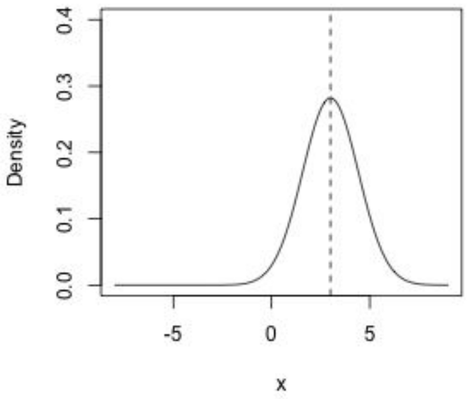
Let $X_1, \ldots, X_n$ be i.i.d samples from a population $f(x \mid \theta)$. The joint density of $X_1, \ldots, X_n$ can be written as

$$f(x_1, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta).$$

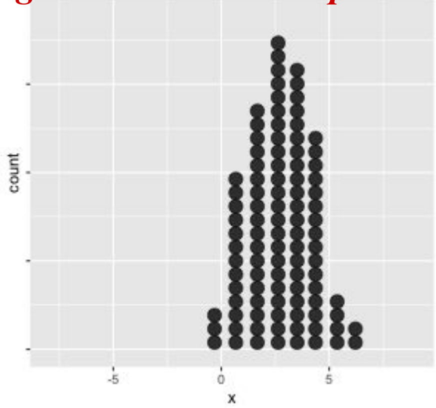
If the population is a normal distribution, then the joint density is $f\left(x_1, \ldots, x_n \,\middle|\, \mu, \sigma^2\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2}$.

Which one has more chance to generate such a sample?

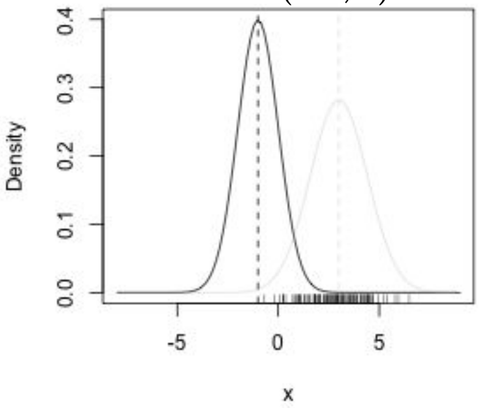
$X \sim N(3, 2)$



n=100



$X \sim N(-1, 1)$



$f(x_1, \ldots, x_{100} \mid 3, 2)$ or $f(x_1, \ldots, x_{100} \mid -1, 1)$?

The likelihood function

$k = 1, 2, 3, \ldots, \; p \in (0,1)$

Let $X_1, \ldots, X_n$ be i.i.d samples from a population $f(x \,|\, \theta)$. The joint density of $X_1, \ldots, X_n$ can be written as

$$f(x_1, \ldots, x_n \,|\, \theta) = \prod_{i=1}^{n} f(x_i \,|\, \theta).$$

Binomial(k, p)

$$L(k,p) = \prod_{i=1}^{n} f(x_i \,|\, k, p)$$

To find θ that makes the observed data "most probable" or "most likely".

Definition. For the i.i.d samples $X_1, \ldots, X_n$, we vary the value of θ *in a meaningful set* to evaluate its likelihood

= parameter space

$$L(\theta) = \prod_{i=1}^{n} f(X_i \,|\, \theta),$$

and $L(\theta)$ is called the likelihood function. The maximum likelihood estimator of θ is the particular value that maximizes the likelihood.

More often than not, it is easier to deal with the log-likelihood function:

maximizing $l(\theta)$ is equivalent to maximizing $L(\theta)$.

$$l(\theta) = \log\left(\prod_{i=1}^{n} f(X_i \,|\, \theta)\right) = \sum_{i=1}^{n} \log f(X_i \,|\, \theta).$$

Maximum likelihood estimators

Hessian matrix:

$$\nabla^2 l = \begin{pmatrix} \frac{\partial^2 l}{\partial \mu^2} & \frac{\partial^2 l}{\partial \mu \partial \sigma} \\ \frac{\partial^2 l}{\partial \sigma \partial \mu} & \frac{\partial^2 l}{\partial \sigma^2} \end{pmatrix}$$

$|\nabla^2 l| > 0 \to$ convex (minimum)

$|\nabla^2 l| < 0 \to$ concave (maximum)

If (μ, σ) maximizes $l(\mu, \sigma)$, or achieves a maximum, then $\left(\frac{\partial l}{\partial \mu}, \frac{\partial l}{\partial \sigma}\right) = (0, 0)$.

Example 6. Let $X_1, \ldots, X_n$ be i.i.d $N(\mu, \sigma^2)$. Find the MLE for μ and σ.

Solution. The likelihood function is $L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\mu)^2}$.

The log-likelihood would be:

$$l(\mu, \sigma) = \sum_{i=1}^{n} \log \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(X_i-\mu)^2} \right\}$$

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

$$\begin{cases} \frac{\partial l}{\partial \mu} = -\frac{1}{2\sigma^2}\sum_{i=1}^{n}\{-2(X_i-\mu)\} = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n} X_i - n\mu\right) \\ \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 \end{cases}$$

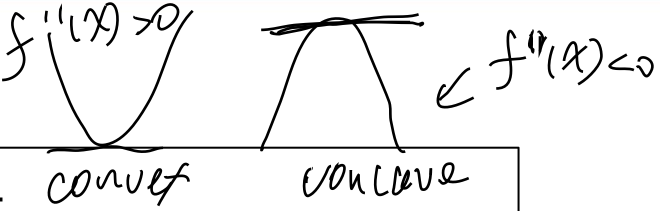
If $\hat{\mu}_n, \hat{\sigma}_n$ are MLE, $\left(\frac{\partial l}{\partial \mu}, \frac{\partial l}{\partial \sigma}\right) = (0, 0)$. $\Rightarrow$

$$\begin{cases} \frac{1}{\hat{\sigma}_n^2}\left(\sum_{i=1}^{n} X_i - n\hat{\mu}_n\right) = 0 \\ -\frac{n}{\hat{\sigma}_n} + \frac{1}{\hat{\sigma}_n^3}\sum_{i=1}^{n}(X_i-\hat{\mu}_n)^2 = 0 \end{cases}$$

As a result, $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$, $\hat{\sigma}_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$. $\to$ MLE, MM are the same.

Maximum likelihood estimators

$\mu = \lambda$



Example 7. Let $X_1, \ldots, X_n$ be i.i.d $\text{Poisson}(\lambda)$. Find the MLE for λ.

Solution. The probability mass function is $P(X = x \mid \lambda) = \dfrac{\lambda^x e^{-\lambda}}{x!}$.

The likelihood function can be written as

$$L(\lambda) = \prod_{i=1}^{n} p(x_i \mid \lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}.$$

Thus, the log-likelihood function is:

$$l(\lambda) = \log L(\lambda) = \left(\sum x_i\right) \log \lambda - n\lambda - \sum_{i=1}^{n} \log(x_i!).$$

$$\frac{\partial l}{\partial \lambda} = \frac{\sum x_i}{\lambda} - n \underbrace{= 0} \Rightarrow \hat{\lambda}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$= \bar{X}_n.$$

$$\frac{\partial^2 l}{\partial \lambda^2} = -\frac{\overset{>0}{\sum x_i}}{\lambda^2} < 0 \quad (\text{concave})$$

Next Tuesday...

More on MLE: examples and limiting theory