Estimation methods

Overview:

1. **Maximum likelihood**
2. **Bootstrap**

1. Maximum Likelihood

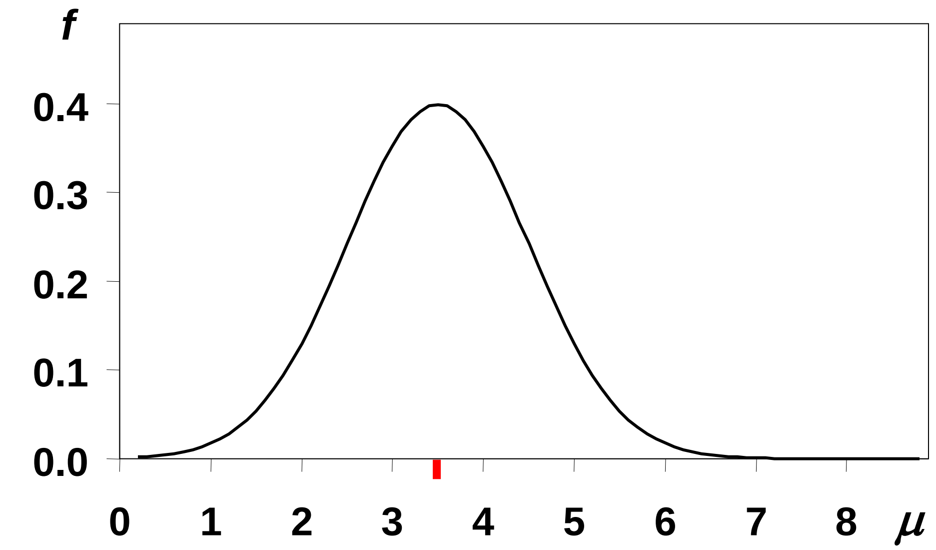
MAXIMUM LIKELIHOOD

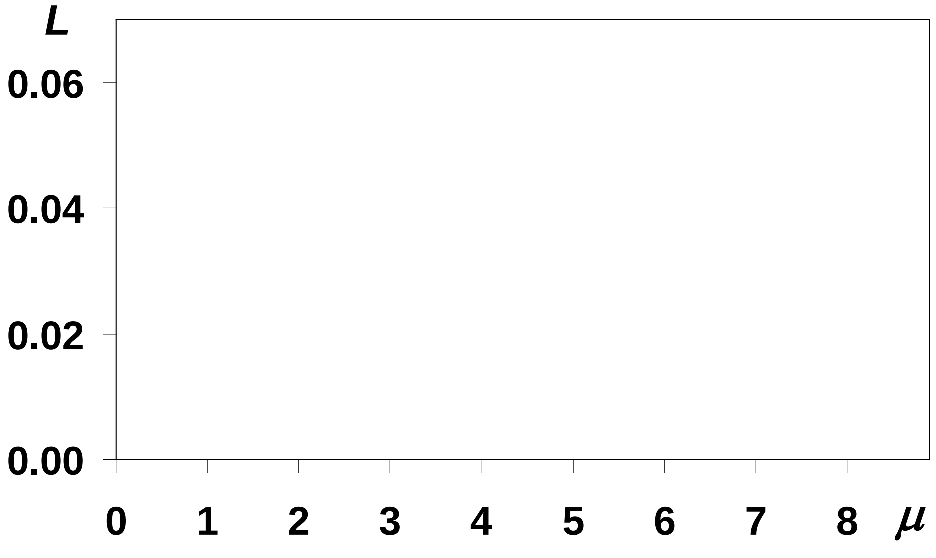
Suppose that:

• **you have a normally-distributed random variable *X* with unknown population mean μ and standard deviation 1 (for the time being)**

• **you have a sample of two observations, 4 and 6**

MAXIMUM LIKELIHOOD

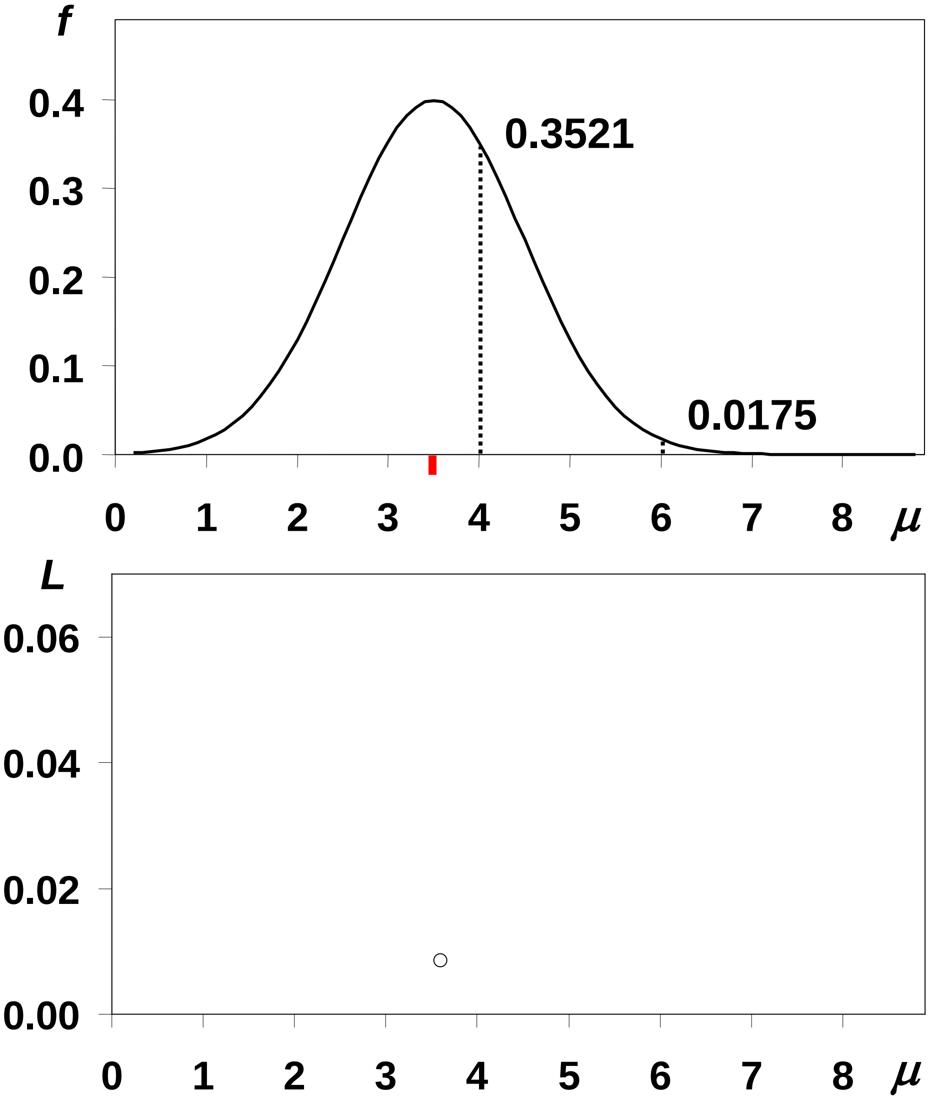




Suppose initially that μ = 3.5.

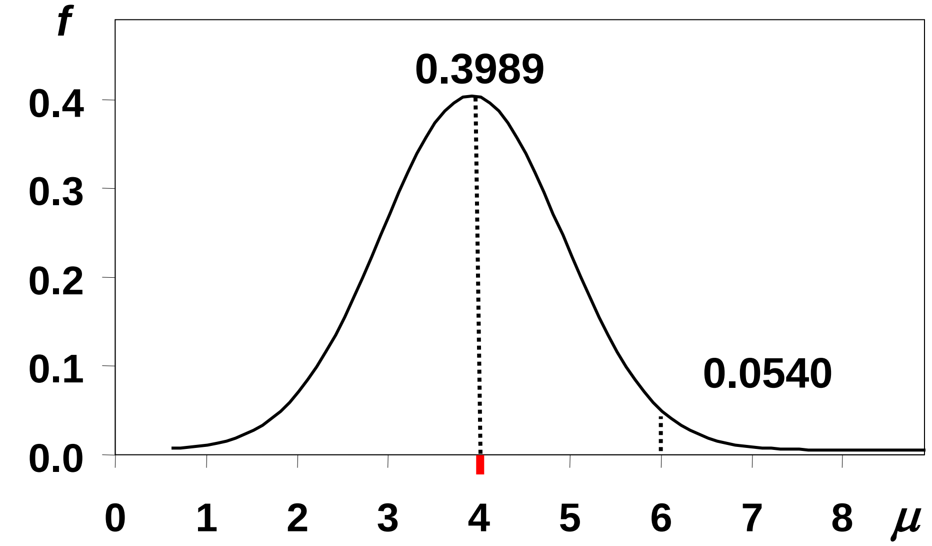
Under this hypothesis the probability density at 4 would be 0.3521 and that at 6 would be 0.0175.

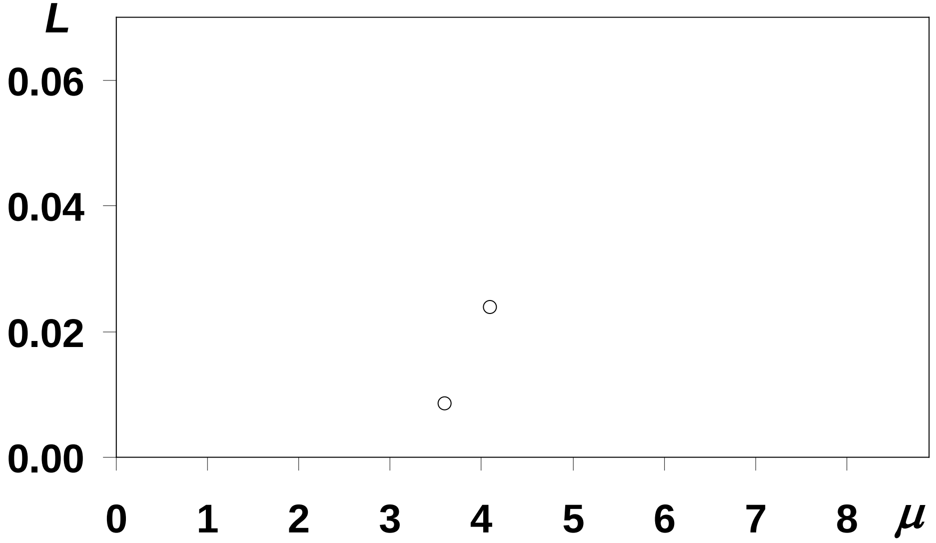
MAXIMUM LIKELIHOOD



μ	$f(4)$	$f(6)$	L
3.5	0.3521	0.0175	0.0062

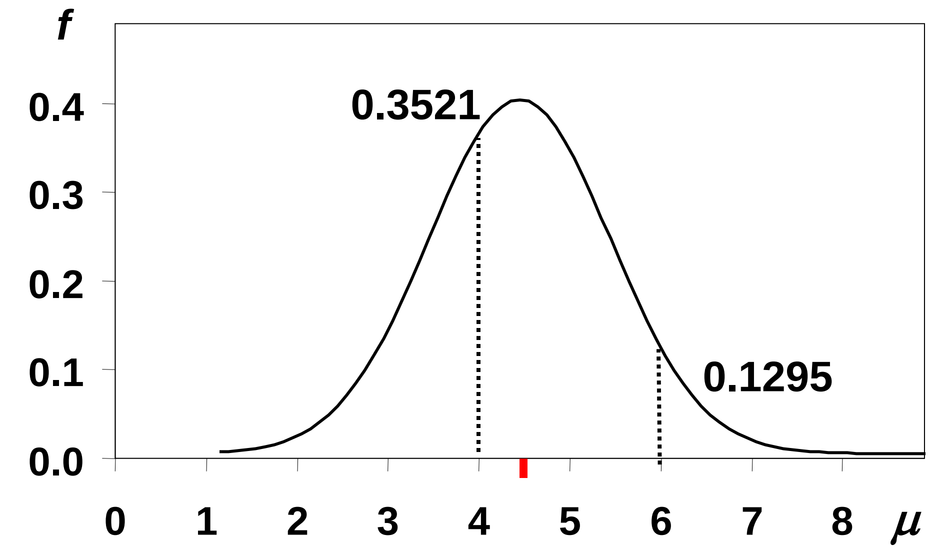
MAXIMUM LIKELIHOOD

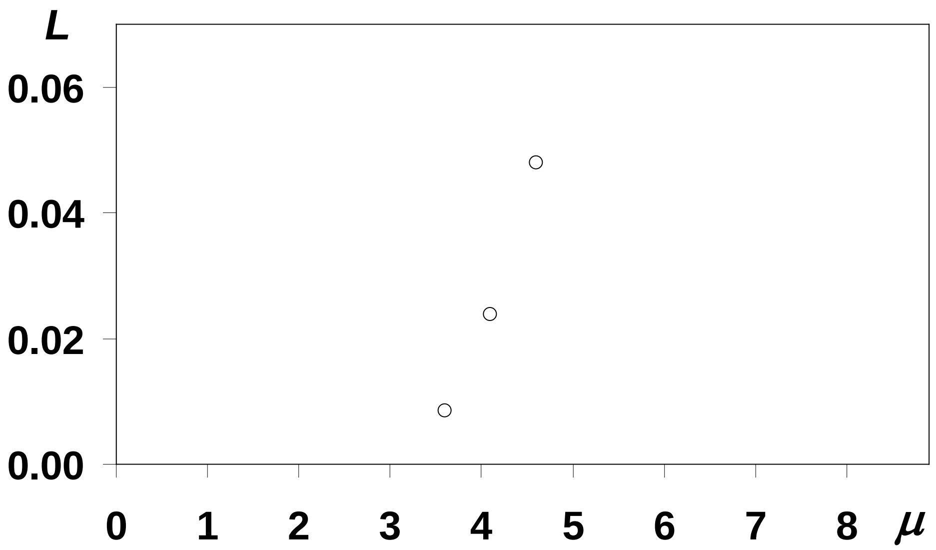




μ	$f(4)$	$f(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215

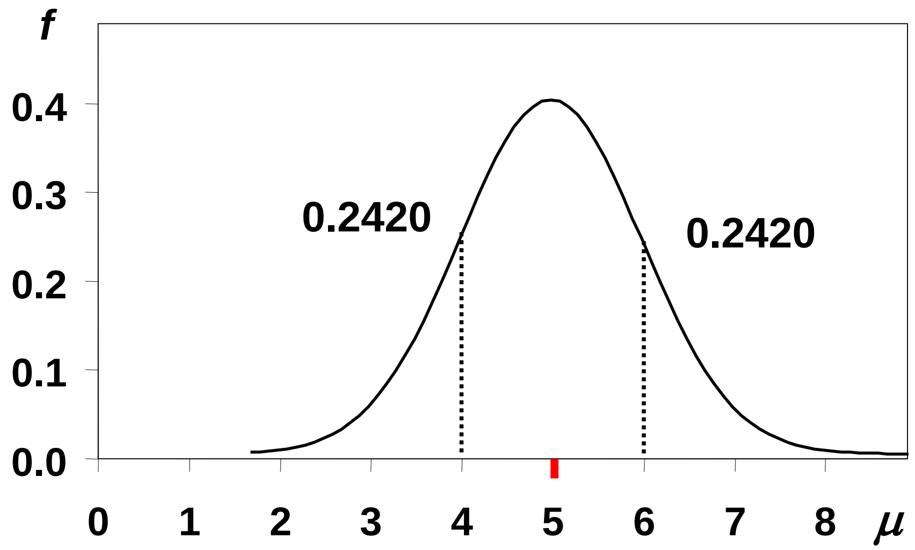
MAXIMUM LIKELIHOOD

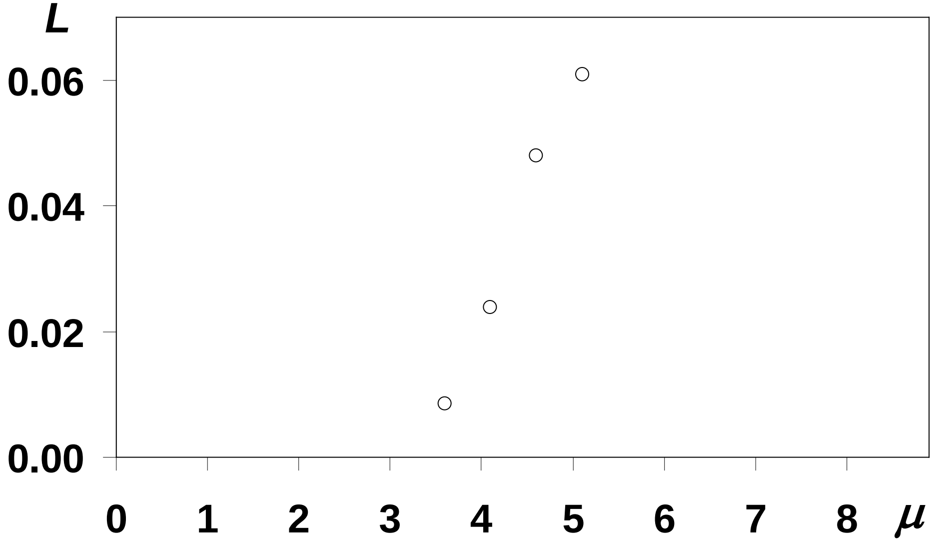




μ	$f(4)$	$f(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456

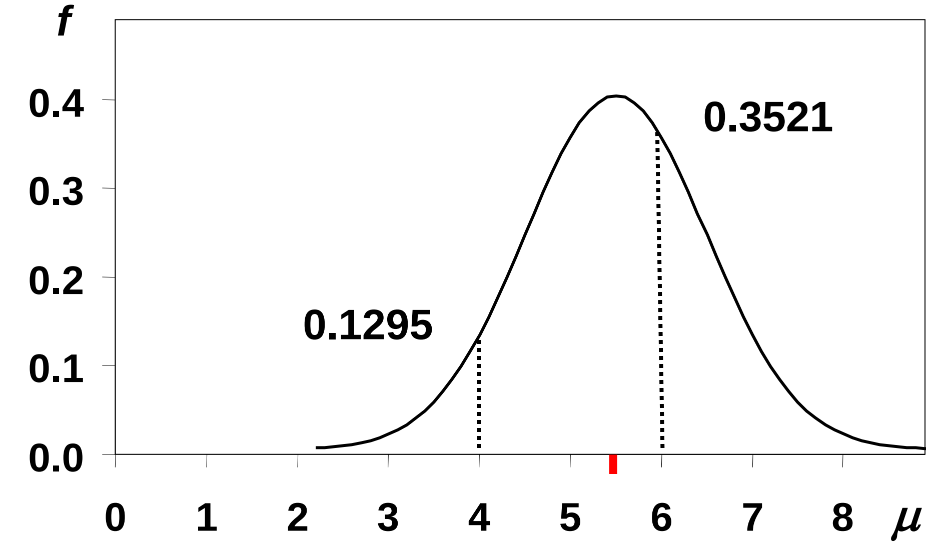
MAXIMUM LIKELIHOOD

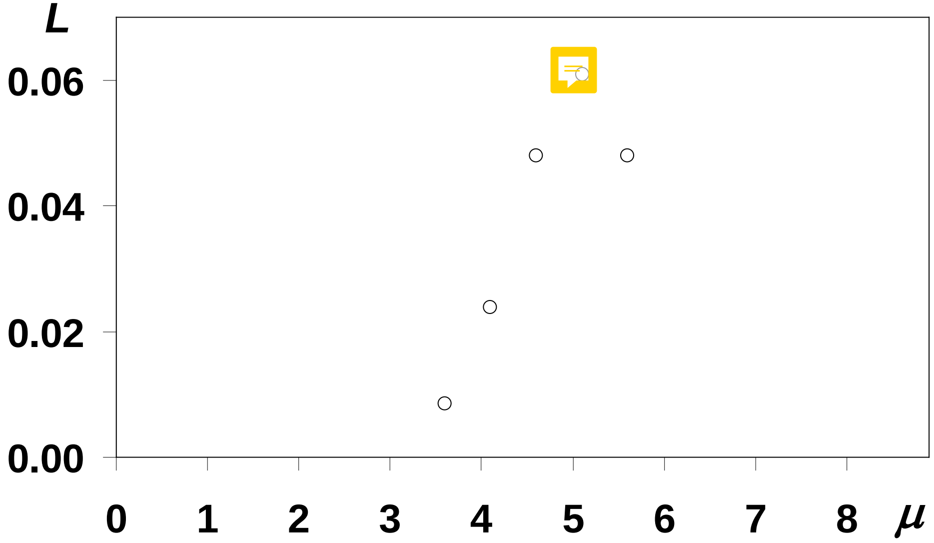




μ	$f(4)$	$f(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585

MAXIMUM LIKELIHOOD





μ	$f(4)$	$f(6)$	L
3.5	0.3521	0.0175	0.0062
4.0	0.3989	0.0540	0.0215
4.5	0.3521	0.1295	0.0456
5.0	0.2420	0.2420	0.0585
5.5	0.1295	0.3521	0.0456

MAXIMUM LIKELIHOOD

Let's do it analytically:

MAXIMUM LIKELIHOOD

Let's do it analytically:

If X is normally distributed with mean μ and standard deviation 1, its density function is

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

MAXIMUM LIKELIHOOD

Let's do it analytically:

If X is normally distributed with mean μ and standard deviation 1, its density function is

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

Thus at 4 and 6 it will be:

$$f(4) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \qquad f(6) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}$$

MAXIMUM LIKELIHOOD

Let's do it analytically:

If X is normally distributed with mean μ and standard deviation 1, its density function is

$$f(X) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X-\mu)^2}$$

Thus at 4 and 6 it will be:

$$f(4) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \qquad f(6) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}$$

$$\text{joint density} = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

MAXIMUM LIKELIHOOD

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

MAXIMUM LIKELIHOOD

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

MAXIMUM LIKELIHOOD

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}\right)$$

$$\log L = \log\left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2}\right)\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}\right)\right]$$

$$= \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2}\right) + \log\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2}\right)$$

MAXIMUM LIKELIHOOD

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

MAXIMUM LIKELIHOOD

$$L(\mu \mid 4,6) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$\log L = \log \left[\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right) \right]$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(4-\mu)^2} \right) + \log \left(\frac{1}{\sqrt{2\pi}} \right) + \log \left(e^{-\frac{1}{2}(6-\mu)^2} \right)$$

$$= 2\log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

MAXIMUM LIKELIHOOD

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

MAXIMUM LIKELIHOOD

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

Taking the derivative and setting it to zero leads to

$$\frac{d\log L}{d\mu} = (4-\mu) + (6-\mu)$$

$$\frac{d\log L}{d\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = 5$$

MAXIMUM LIKELIHOOD

$$\log L = 2\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{1}{2}(4-\mu)^2 - \frac{1}{2}(6-\mu)^2$$

Taking the derivative and setting it to zero leads to

$$\frac{d\log L}{d\mu} = (4-\mu)+(6-\mu)$$

$$\frac{d\log L}{d\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = 5$$

Thus from the first order condition we confirm that 5 is the value of μ that maximizes the log-likelihood function, and hence the likelihood function.

MAXIMUM LIKELIHOOD

We will generalize this result to a sample of *n* observations $X_1, \ldots, X_n$ normally distributed with mean μ and standard deviation σ. The probability density for X_i is given by:

$$f(X_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_i - \mu}{\sigma}\right)^2}$$

The joint density (likelihood) function for a sample of *n* observations is the product of their individual densities.

$$L(\mu, \sigma \mid X_1, \ldots, X_n) = \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1 - \mu}{\sigma}\right)^2} \right) \times \ldots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n - \mu}{\sigma}\right)^2} \right)$$

MAXIMUM LIKELIHOOD

$$L(\mu,\sigma \mid X_1,\ldots,X_n) = \left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right)\times\ldots\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$\log L = \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right)\times\ldots\times\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)\right]$$

MAXIMUM LIKELIHOOD

$$L(\mu,\sigma \mid X_1,\dots,X_n) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \dots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$\log L = \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \dots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)\right]$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) + \dots + \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

MAXIMUM LIKELIHOOD

$$L(\mu,\sigma \mid X_1,\ldots,X_n) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \ldots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$\log L = \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \ldots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)\right]$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) + \ldots + \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2 - \ldots - \frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2$$

MAXIMUM LIKELIHOOD

$$L(\mu,\sigma \mid X_1,\ldots,X_n) = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \ldots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$\log L = \log\left[\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) \times \ldots \times \left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)\right]$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2}\right) + \ldots + \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2}\right)$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2}\left(\frac{X_1-\mu}{\sigma}\right)^2 - \ldots - \frac{1}{2}\left(\frac{X_n-\mu}{\sigma}\right)^2$$

$$= n\log\left(\frac{1}{\sigma}\right) + n\log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{1}{\sigma^2}\left(-\frac{1}{2}(X_1-\mu)^2 - \ldots - \frac{1}{2}(X_n-\mu)^2\right)$$

MAXIMUM LIKELIHOOD

$$\log L = n\log\left(\frac{1}{\sigma}\right) + n\log\left(\frac{1}{\sqrt{2\pi}}\right) + \frac{1}{\sigma^2}\left(-\frac{1}{2}\left(X_1-\mu\right)^2 - \ldots - \frac{1}{2}\left(X_n-\mu\right)^2\right)$$

$$= -n\log\sigma + n\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2}\sum\left(X_i-\mu\right)^2$$

To maximize it, we will set the partial derivatives with respect to μ and σ equal to zero.

MAXIMUM LIKELIHOOD

$$\log L = -n\log\sigma + n\log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2}\sum\left(X_i-\mu\right)^2$$

When differentiating with respect to μ:

$$\frac{\partial \log L}{\partial\mu} = \frac{1}{\sigma^2}\frac{\partial}{\partial\mu}\left(-\frac{1}{2}(X_1-\mu)^2 - \ldots - \frac{1}{2}(X_n-\mu)^2\right)$$

$$= \frac{1}{\sigma^2}\left[(X_1-\mu)+\ldots+(X_n-\mu)\right]$$

$$= \frac{1}{\sigma^2}\left(\sum X_i - n\mu\right)$$

$$\frac{\partial \log L}{\partial\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = \overline{X}$$

MAXIMUM LIKELIHOOD

$$\log L = -n \log \sigma + n \log\left(\frac{1}{\sqrt{2\pi}}\right) - \frac{\sigma^{-2}}{2}\sum\left(X_i - \mu\right)^2$$

When differentiating with respect to σ:

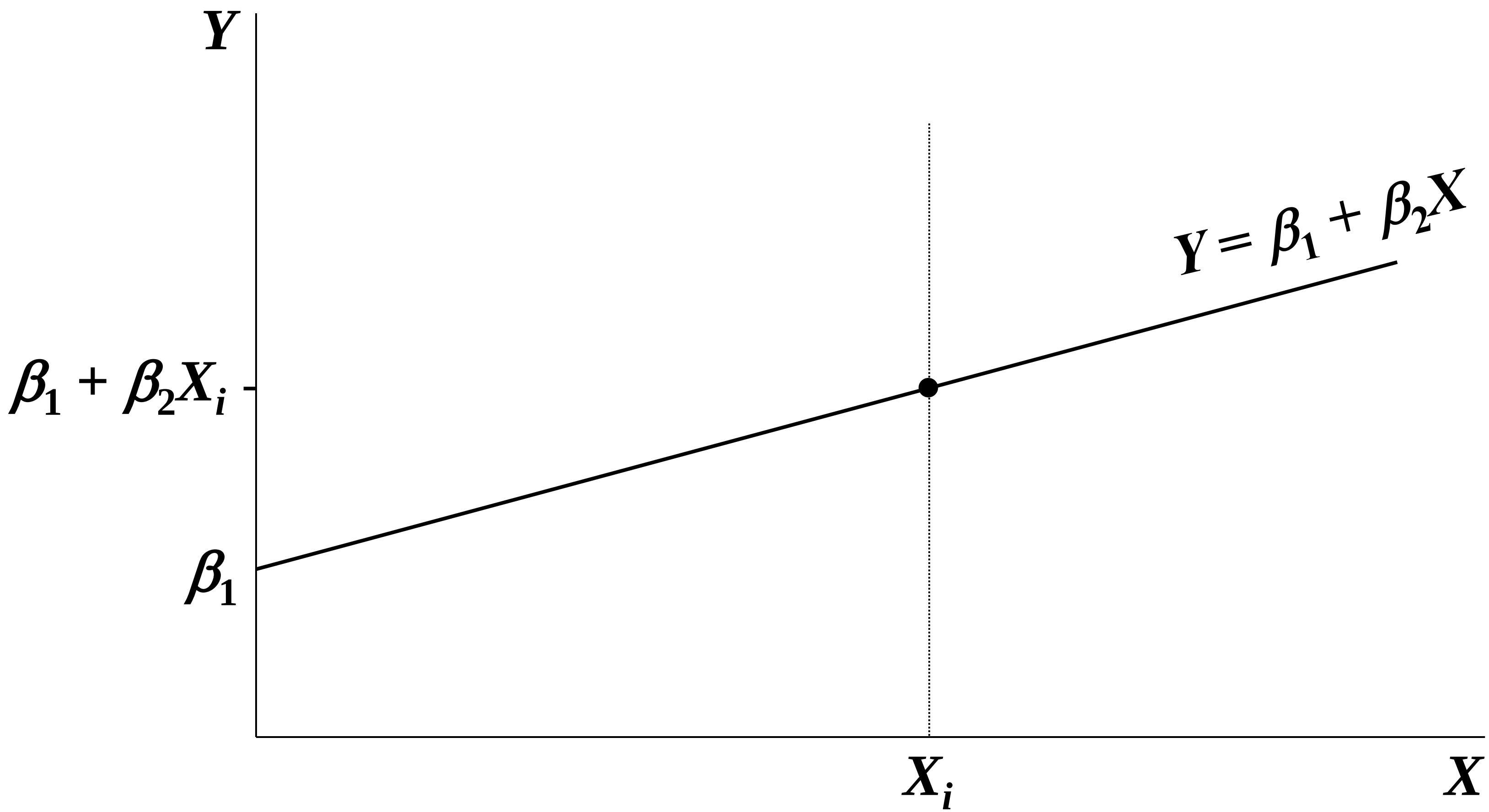
$$\frac{\partial \log L}{\partial \sigma} = -\frac{n}{\sigma} + \sigma^{-3}\sum(X_i - \mu)^2$$

$$\frac{\partial \log L}{\partial \sigma} = 0 \quad \Rightarrow \quad -\frac{n}{\hat{\sigma}} + \hat{\sigma}^{-3}\sum(X_i - \hat{\mu})^2 = 0$$

$$\therefore -n\hat{\sigma}^2 + \sum(X_i - \overline{X})^2 = 0$$

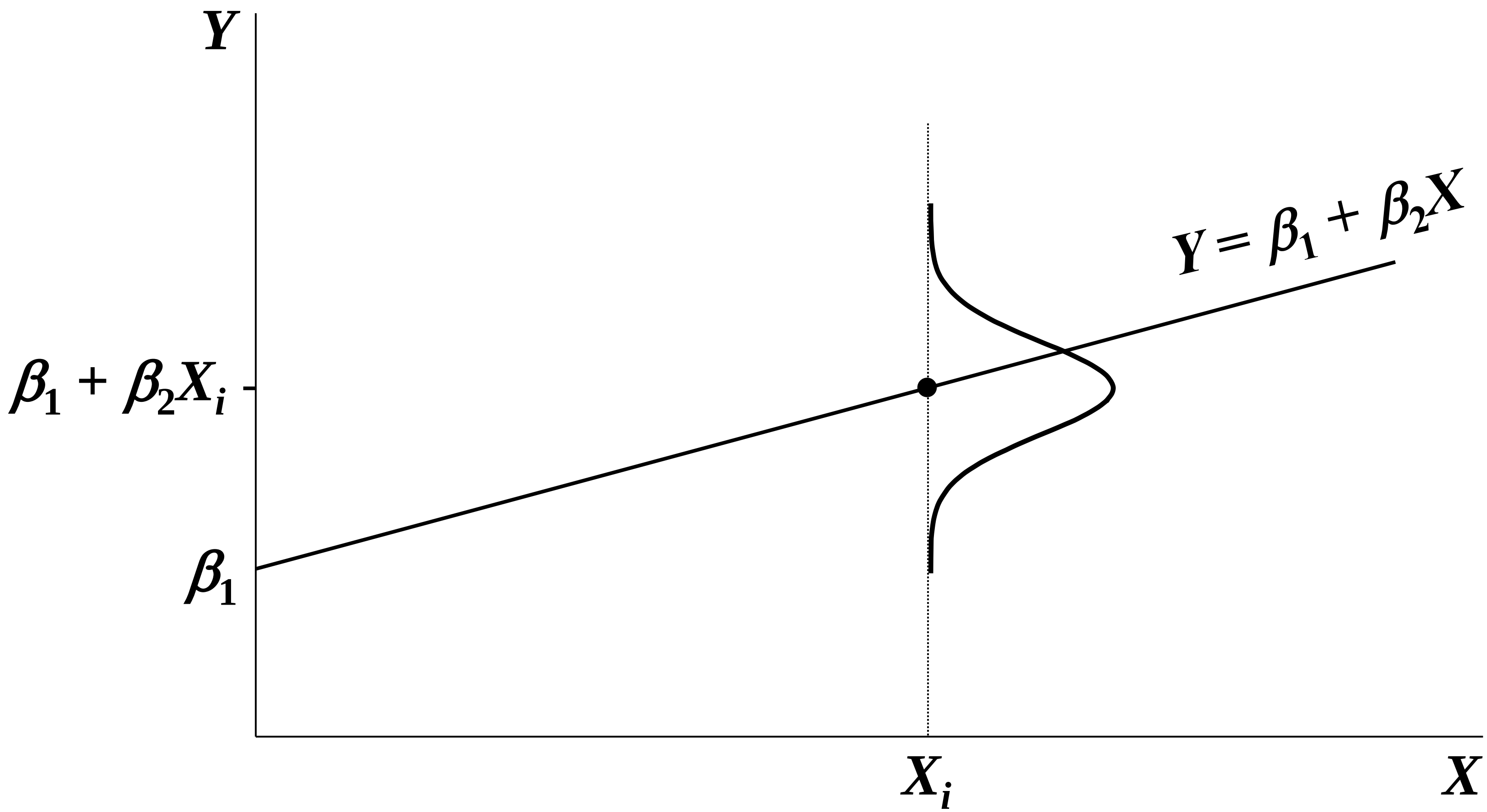
$$\therefore \hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \overline{X})^2 = \mathrm{Var}(X)$$

REGRESSION COEFFICIENTS



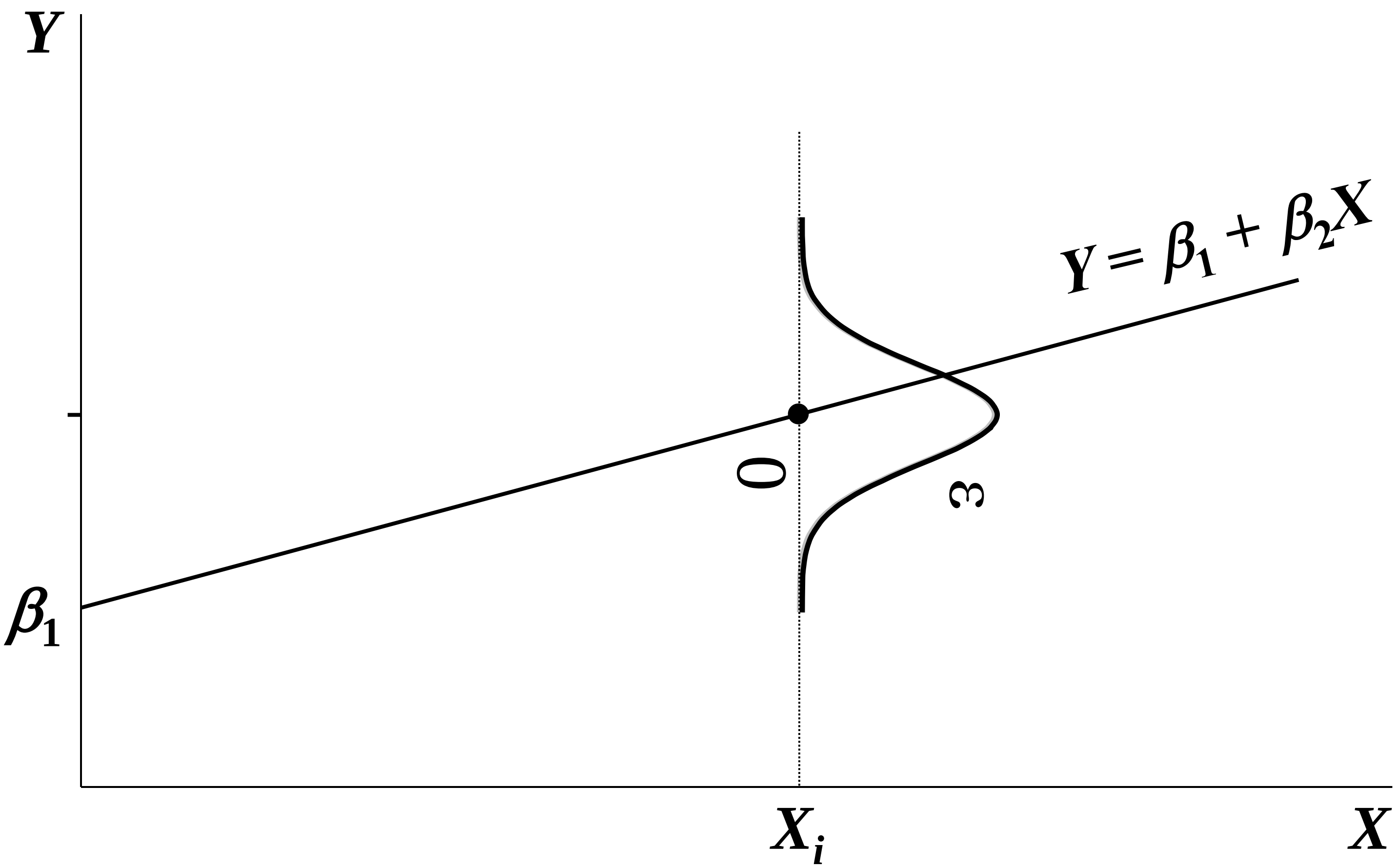
We apply the maximum likelihood principle to regression analysis, using the model $Y = b_1 + b_2X + u$.

REGRESSION COEFFICIENTS



The mean value of the distribution of Y_i is $b_1 + b_2X_i$. Its standard deviation is σ, the standard deviation of the disturbance term.

REGRESSION COEFFICIENTS



Hence $f(Y_i) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_i - \beta_1 - \beta_2 X_i}{\sigma}\right)^2}$

REGRESSION COEFFICIENTS

The joint density function for the observations on *Y* is the product of their individual densities.

$$f(Y_1) \times \ldots \times f(Y_n) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2 X_n}{\sigma}\right)^2}$$

$$L\left(\beta_1, \beta_2, \sigma \mid Y_1, \ldots, Y_n\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2 X_n}{\sigma}\right)^2}$$

$$\log L = \log\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2 X_1}{\sigma}\right)^2} \times \ldots \times \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2 X_n}{\sigma}\right)^2}\right)$$

REGRESSION COEFFICIENTS

$$\log L = \log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2}\times\ldots\times\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2}\right)$$

$$= \log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2}\right)+\ldots+\log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2}\right)$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{1}{2}\left(\frac{Y_1-\beta_1-\beta_2X_1}{\sigma}\right)^2-\ldots-\frac{1}{2}\left(\frac{Y_n-\beta_1-\beta_2X_n}{\sigma}\right)^2$$

$$= n\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right)-\frac{\sigma^{-2}}{2}Z$$

$$\text{where } Z = \left[(Y_1-\beta_1-\beta_2X_1)^2+\ldots+(Y_n-\beta_1-\beta_2X_n)^2\right]$$

To maximize the log$L$, we need to minimize Z.

2. Bootstrap

BASIC IDEA OF BOOTSTRAP

The use of the term “bootstrap” comes from the phrase “To pull oneself up by one's bootstraps” - generally interpreted as succeeding in spite of limited resources.

This phrase comes from the adventures of Baron Munchausen - Raspe (1786).

In one of his many adventures, Baron Munchausen had fallen to the bottom of a lake and just as he was about to succumb to his fate he thought to pick himself up by his own bootstraps!

BASIC IDEA OF BOOTSTRAP

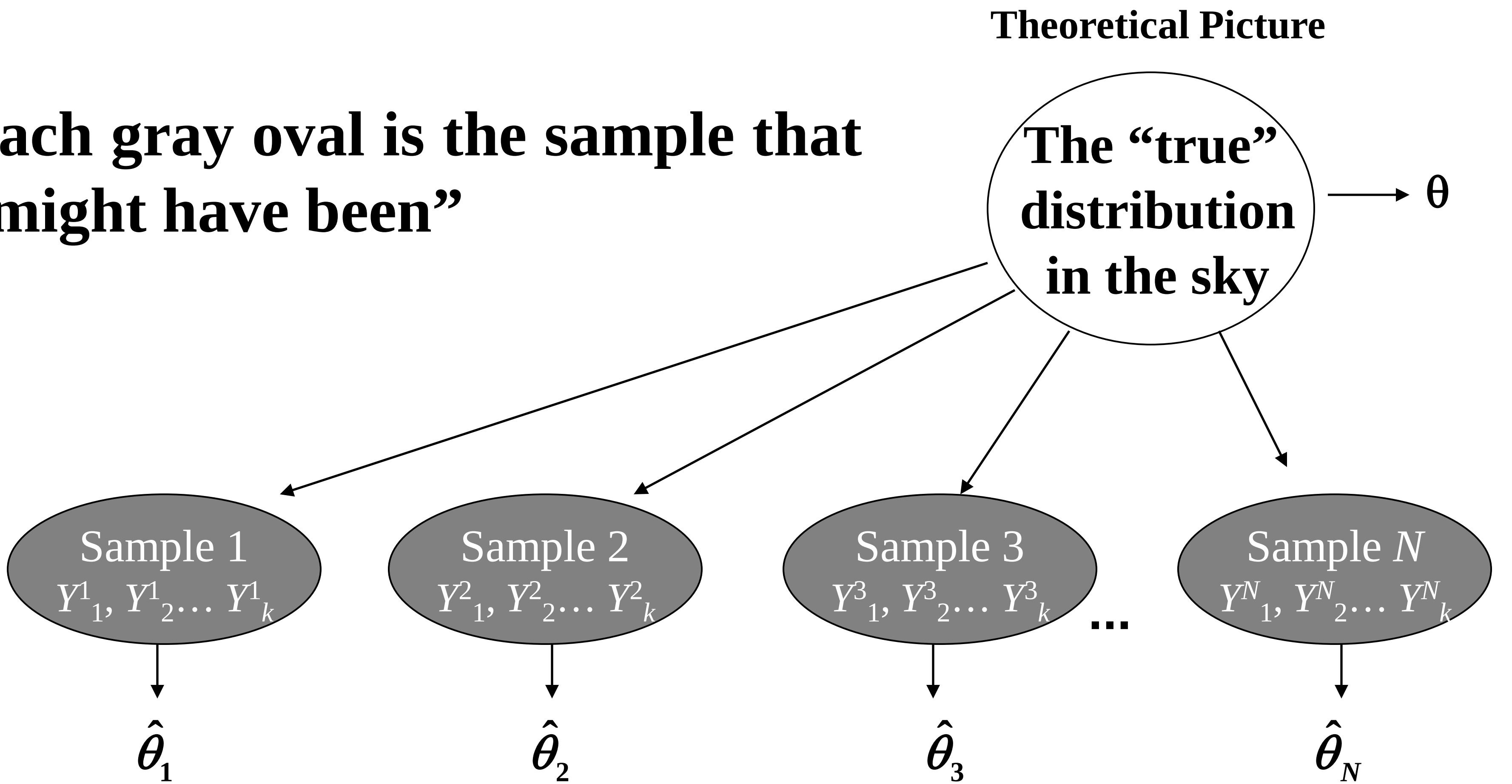
In the late 70's the statistician Brad Efron made an ingenious suggestion.

Most (sometimes all) of what we know about the "true" probability distribution comes from the data. So let's treat the data as a proxy for the true distribution.

We draw multiple samples from this proxy…This is called "resampling". And compute the statistic of interest on each of the resulting pseudo-datasets.

THE BASIC IDEA

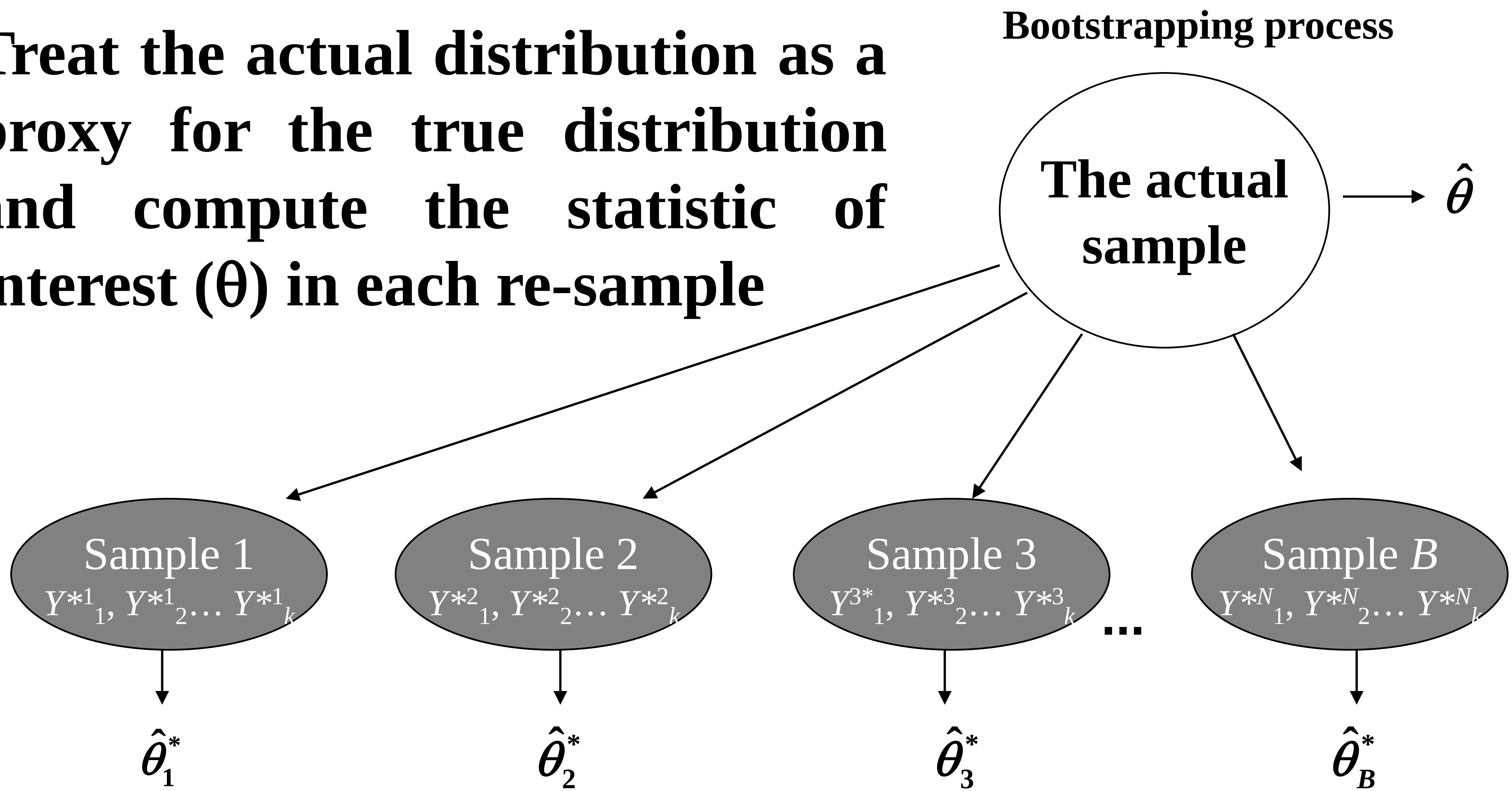
Each gray oval is the sample that "might have been"



The distribution of our estimator (θ) depends on both the true distribution and the size (*n*) of our sample

THE BASIC IDEA

Treat the actual distribution as a proxy for the true distribution and compute the statistic of interest (θ) in each re-sample



The distribution of θ* constitutes an estimate of the distribution of θ.

BOOTSTRAP

In a very simple form it works as follows.

We have a sample of size *n*. We want to estimate some parameter θ. For each sample point we assign probability 1/n.

From this sample we draw another random sample (with replacement) of size *n* and estimate θ. Bootstrap estimator for θ and its variance is calculated as:

$$\theta_B^* = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}_j^* \quad \text{and the variance} \quad V_B(\theta_B^*) = \frac{1}{B-1}\sum_{j=1}^{B}(\hat{\theta}_j^* - \theta_B^*)^2$$

Having this, we can do inference …

BOOTSTRAP

While resampling we did not use any assumption about the population distribution. This bootstrap is called non-parametric bootstrap. If we have some idea about the population distribution then we can use it in resampling to improve the performance of the estimation.

For example if we know that population distribution is normal, we can estimate its parameters using our sample (e.g. mean and variance) and approximate population distribution with this sample distribution to draw new samples.

BALANCED BOOTSTRAP

Makes sure that number of occurrences of each sample point (in aggregate) is the same. It can be achieved as follows (for a sample size *n*):

1) Repeat numbers from 1 to n, B times

2) Take a random permutation of numbers from 1 to nB.

3) Take the first n points and the corresponding sample points (x_n). Estimate the parameters of interest. Then take the second k points (from k+1 to 2k) and estimate the parameters of interest. Repeat it B times and find bootstrap estimators and distributions.

BALANCED BOOTSTRAP: EXAMPLE

Suppose we have 3 sample points and number of bootstraps we want is 3. Our observations are: (x_1, x_2, x_3)

Then we repeat numbers from 1 to 3 three times:

1 2 3 1 2 3 1 2 3

Then we take one of the random permutations of numbers from 1 to 3x3=9. E.g. 4 3 9 5 6 1 2 8 7

First we take observations x1, x3, x3 estimate the parameter

Then we take x2, x3, x1 and estimate the parameter

Then we take x2, x2, x1 and we estimate parameter

As it can be seen each observation is present 3 times

This technique is meant to improve the results of bootstrap resampling.

BOOTSTRAP PARADIGM

Bootstrap Paradigm

$$F(\hat{\theta}-\theta)=F(\theta_B^*-\hat{\theta})$$

The distribution of the difference of the estimated beta and the true beta is the same as the distribution of the difference of the bootstrapped beta and the estimated beta.