THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 321: Introduction to Probability and Statistics

Lecture 24: Maximum Likelihood Estimation

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

www.cs.cmu.edu/∼psarkar/teaching

Roadmap

- Frequentist Statistics introduction
- Biased, Unbiased and Asymptotically unbiased estimators
- M.L.E and how to find it.

Frequentist Statistics

- The parameter(s) θ is fixed and unknown
- Data is generated through the likelihood function $p(X;\theta)$ (if discrete) or $f(X;\theta)$ (if continuous).
- Now we will be dealing with multiple candidate models, one for each value of θ
- We will use $E_\theta[h(X)]$ to define the expectation of the random variable $h(X)$ as a function of parameter θ

Problems we will look at

- **Parameter estimation**: We want to estimate unknown parameters from data.
  - **Maximum Likelihood estimation (section 9.1):** Select the parameter that makes the observed data most likely.
  - i.e. maximize the probability of obtaining the data at hand.
- **Hypothesis testing:** An unknown parameter takes a finite number of values. One wants to find the best hypothesis based on the data.
  - **Significance testing:** Given a hypothesis, figure out the rejection region and reject the hypothesis if the observation falls within this region.

Classical parameter estimation

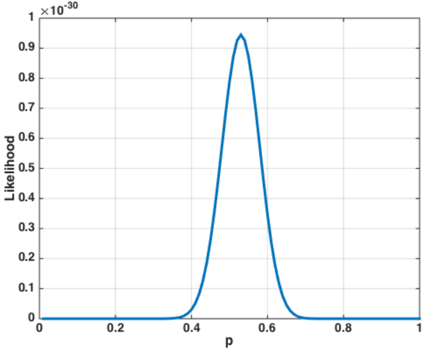
We are given observations $X = (X_1, \ldots, X_n)$. An **estimator** is a random variable of the form $\hat{\Theta} = g(X)$ (sometime also denoted by Θ_n).

- Since the distribution of X depends on θ, so does the distribution $\hat{\Theta}_n$
- The mean and variance of $\hat{\Theta}_n$ can be defined as $E_\theta[\hat{\Theta}_n]$ and $\text{var}_\theta(\hat{\Theta}_n)$.
- For simplicity we will also use $E[.]$ and var[.] and drop the θ from the notation
- The **estimation error** denoted by $\tilde{\Theta}_n = \hat{\Theta}_n - \theta$.
- **Bias** of an estimator is given by $b_\theta(\hat{\Theta}_n) = E_\theta[\hat{\Theta}_n] - \theta$
- An **Unbiased** estimator is one for which $E[\hat{\Theta}_n] = \theta$
- An asymptotically unbiased estimator is one for which $\lim_{n\to\infty} E[\hat{\Theta}_n] = \theta$

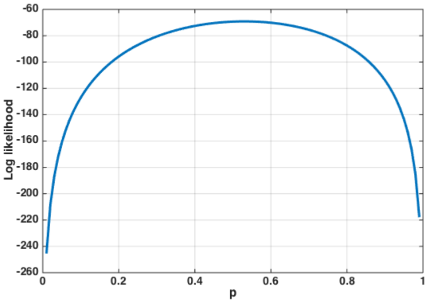
Maximum Likelihood Estimation

- Find the θ that maximizes the joint likelihood $p(X_1, \ldots, X_n; \theta)$ (of the joint pdf for continuous random variables).
- We have $P(X_i; \theta)$ for random variable X_i
- Often X_i are independent and so $P(X_1, \ldots, X_n; \theta) = \prod_i P(X_i; \theta)$
- We want to calculate the **M**aximum **L**ikelihood **E**stimate $\hat{\theta}$
- First calculate $P(X_1, \ldots, X_n; \theta) = \prod_i P(X_i; \theta)$
- Now calculate the logarithm. $\log P(X_1, \ldots, X_n; \theta) = \sum_i \log P(X_i; \theta)$
- Now take a derivative and set it to zero. $\frac{d}{d\theta} \sum_i \log P(X_i; \theta) = 0$ and solve for θ

Likelihood vs. Log likelihood



(A) Likelihood



(B) Log-likelihood

[Takeaway]

- Loglikelihood is a monotonic function of the likelihood.
- So the maximum is achieved at the same point, albeit, in most cases with a lot less computation.

Estimating the parameter of the Exponential

n customers arrive at a mall at times Y_i. We take $Y_0 = 0$. The inter arrival times are $X_i = Y_i - Y_{i-1}$ are often modeled as i.i.d Exponential(λ) r.v's. We want to the MLE of λ.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$
- Now write the joint likelihood $f_X(x; \lambda) = \prod_i \lambda e^{-\lambda x_i} = \lambda^n \prod_i e^{-\lambda x_i}$
- Now take logarithm of the joint likelihood.
  $\log f_X(x; \lambda) = n \log \lambda - \lambda \sum_i x_i$.
- Differentiate and set to zero to get the MLE.
  $n \frac{1}{\hat{\lambda}} - \sum_i x_i = 0 \to \hat{\lambda} = \frac{1}{\sum_i x_i / n}$.

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p
- First write the likelihood of r.v i conveniently as $P(X_i; p) = p^{X_i}(1-p)^{1-X_i}$

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p
- First write the likelihood of r.v i conveniently as $P(X_i; p) = p^{X_i}(1-p)^{1-X_i}$
- Now note that the r.v's are all independent and so

  $$P(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p
- First write the likelihood of r.v i conveniently as $P(X_i; p) = p^{X_i}(1-p)^{1-X_i}$
- Now note that the r.v's are all independent and so
  $$P(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$
- Its often more convenient to maximizes the logarithm of a product form.
  $$\log P(X_1, \ldots, X_n; p) = \sum_i \log P(X_i; p)$$
- $$= \sum_i \log \left(p^{X_i}(1-p)^{1-X_i} \right) = \sum_i X_i \log p + \sum_i (1 - X_i) \log(1-p)$$

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p
- First write the likelihood of r.v i conveniently as $P(X_i; p) = p^{X_i}(1-p)^{1-X_i}$
- Now note that the r.v's are all independent and so
  $$P(X_1, \dots, X_n; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$
- Its often more convenient to maximizes the logarithm of a product form.
  $$\log P(X_1, \dots, X_n; p) = \sum_i \log P(X_i; p)$$
- $$= \sum_i \log\left(p^{X_i}(1-p)^{1-X_i}\right) = \sum_i X_i \log p + \sum_i (1-X_i)\log(1-p)$$
- $\hat{p} = \arg\max_p \log P(X_1, \dots, X_n; p)$

Maximum Likelihood Estimation

Lets start with an example.

- You have a n i.i.d. bernoulli random variables $X_i \sim Bernoulli(p)$.
- Find the Maximum Likelihood Estimate of p
- First write the likelihood of r.v i conveniently as $P(X_i; p) = p^{X_i}(1-p)^{1-X_i}$
- Now note that the r.v's are all independent and so
  $$P(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$
- Its often more convenient to maximizes the logarithm of a product form.
  $$\log P(X_1, \ldots, X_n; p) = \sum_i \log P(X_i; p)$$
- $$= \sum_i \log \left(p^{X_i}(1-p)^{1-X_i}\right) = \sum_i X_i \log p + \sum_i (1-X_i)\log(1-p)$$
- $\hat{p} = \arg\max_p \log P(X_1, \ldots, X_n; p)$
- $$\frac{\sum_i X_i}{\hat{p}} - \frac{n - \sum_i X_i}{1-\hat{p}} = 0 \rightarrow \hat{p} = \frac{\sum_i X_i}{n}$$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I know σ^2 but not μ. Whats the MLE of μ?

- Notation: $X = (X_1, \ldots, X_n)$ and $x = (x_1, \ldots, x_n)$.
- First write $f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$
- Now write the joint PDF $f_X(x; \mu, \sigma) = \prod_i f_{X_i}(x_i; \mu, \sigma)$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I know σ^2 but not μ. Whats the MLE of μ?

- Notation: $X = (X_1, \ldots, X_n)$ and $x = (x_1, \ldots, x_n)$.
- First write $f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$
- Now write the joint PDF $f_X(x; \mu, \sigma) = \prod_i f_{X_i}(x_i; \mu, \sigma)$
- Now take the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I know σ^2 but not μ. Whats the MLE of μ?

- Notation: $X = (X_1, \ldots, X_n)$ and $x = (x_1, \ldots, x_n)$.
- First write $f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$
- Now write the joint PDF $f_X(x; \mu, \sigma) = \prod_i f_{X_i}(x_i; \mu, \sigma)$
- Now take the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$
- Derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I know σ^2 but not μ. Whats the MLE of μ?

- Notation: $X = (X_1, \ldots, X_n)$ and $x = (x_1, \ldots, x_n)$.
- First write $f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$
- Now write the joint PDF $f_X(x; \mu, \sigma) = \prod_i f_{X_i}(x_i; \mu, \sigma)$
- Now take the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$
- Derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$
- Derive w.r.t μ and set to zero to solve for $\hat{\mu}$.
  $\sum_i x_i - n\hat{\mu} = 0 \rightarrow \hat{\mu} = \frac{\sum_i x_i}{n}$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I know σ^2 but not μ. Whats the MLE of μ?

- Notation: $X = (X_1, \ldots, X_n)$ and $x = (x_1, \ldots, x_n)$.
- First write $f_X(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$
- Now write the joint PDF $f_X(x; \mu, \sigma) = \prod_i f_{X_i}(x_i; \mu, \sigma)$
- Now take the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$
- Derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$
- Derive w.r.t μ and set to zero to solve for $\hat{\mu}$.
  $\sum_i x_i - n\hat{\mu} = 0 \to \hat{\mu} = \frac{\sum_i x_i}{n}$
- The MLE is just the sample mean, which is often denoted by $\bar{X}$

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

*Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

- We start with taking the logarithm of the joint likelihood.

  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$ *

*Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

- We start with taking the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$ *
- First derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$

*Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

- We start with taking the logarithm of the joint likelihood.
  $\log f_X(x; \mu, \sigma) = -n \log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$*
- First derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$
- Next derive w.r.t σ and set to zero to solve for $\hat{\sigma}$.
  $-\frac{n}{\sigma} + \sum_i 2 \frac{(x_i - n\mu)^2}{2\sigma^3} = 0$

*Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

[†] Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

- Now you have two equations and two unknowns![†]

$$\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$$

$$-\frac{n}{\sigma} + \sum_i \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

[†] Thanks to Aneesh Adhikari-Desai for correcting this!

MLE estimate of mean of Gaussian r.v.'s

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I do not know σ^2 or μ. Whats the MLE of μ and σ^2?

- Now you have two equations and two unknowns![†]

$$\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$$

$$-\frac{n}{\sigma} + \sum_i \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

- Solving we see:

$$\hat{\mu} = \sum_i x_i / n = \bar{x} \qquad \hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

[†] Thanks to Aneesh Adhikari-Desai for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \frac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$

‡Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} = \dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \ldots) \leq \max(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$ ‡
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \ldots, X_n)$, otherwise the likelihood will be zero.

‡ Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \dots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} = \dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \dots) \leq \max(X_1, X_2, \dots) \leq 1\}}{(1-a)^n}$ ‡
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \dots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \leq \min(X_1, \dots, X_n)$, what maximizes $1/(1-\hat{a})^n$?

‡ Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a,1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} = \dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \ldots) \leq \max(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$ ‡
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \ldots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \leq \min(X_1, \ldots, X_n)$, what maximizes $1/(1-\hat{a})^n$?
- For all $\hat{a} \leq \min(X_1, \ldots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.

‡ Thanks Miles Hutson for correcting this!

Estimating the lower limit of the Uniform

I have n independent Uniform($[a, 1]$) random variables. Whats the MLE of a? Although most time taking the log helps, sometimes its easier to work with the joint likelihood directly.

- First write $f(X_i; a) = \dfrac{\mathbf{1}\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} \mathbf{1}\{a \leq X_i \leq 1\} =$
  $\dfrac{\mathbf{1}\{a \leq \min(X_1, X_2, \ldots) \leq \max(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$ ‡
- How do I maximize this? Well, $\hat{a}$ has to be less than $\min(X_1, \ldots, X_n)$, otherwise the likelihood will be zero.
- For all $\hat{a} \leq \min(X_1, \ldots, X_n)$, what maximizes $1/(1-\hat{a})^n$?
- For all $\hat{a} \leq \min(X_1, \ldots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.
- $\hat{a} = \min(X_1, \ldots, X_n)$!

‡ Thanks Miles Hutson for correcting this!

Unbiased and asymptotically unbiased estimators

Recall that the MLE of σ^2 obtained using n iid random variables from a $N(\mu, \sigma^2)$ distribution are given by:

$$\hat{\sigma^2} = \frac{\sum_i (x_i - \bar{x})^2}{n} \qquad \hat{\sigma}^2 = \sum_i \frac{(x_i - \hat{\mu})^2}{n}$$

This is also the sample variance. Is it unbiased?

Unbiased and asymptotically unbiased estimators

Recall that the MLE of σ^2 obtained using n iid random variables from a $N(\mu, \sigma^2)$ distribution are given by:

$$\hat{\sigma^2} = \frac{\sum_i (x_i - \bar{x})^2}{n} \qquad \hat{\sigma}^2 = \sum_i \frac{(x_i - \hat{\mu})^2}{n}$$

This is also the sample variance. Is it unbiased?

- First note that

$$\begin{aligned}\sum_i \frac{(x_i - \bar{x})^2}{n} &= \sum_i \frac{x_i^2 - 2x_i\bar{x} + \bar{x}^2}{n} \\ &= \sum_i \frac{x_i^2}{n} - 2\bar{x}^2 + \bar{x}^2 \\ &= \sum_i \frac{x_i^2}{n} - \bar{x}^2\end{aligned}$$

Unbiased and asymptotically unbiased estimators

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$

Unbiased and asymptotically unbiased estimators

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \mathrm{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$

Unbiased and asymptotically unbiased estimators

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$

Unbiased and asymptotically unbiased estimators

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \text{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$
- So the MLE of the variance is not unbiased. However it is asymptotically unbiased!

Unbiased and asymptotically unbiased estimators

- $E\left[\sum_i \frac{(x_i - \bar{x})^2}{n}\right] = E[X_1^2] - E[\bar{X}^2]$
- $E[X_1^2] = \sigma^2 + \mu^2$ and $E[\bar{X}^2] = \mathrm{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$
- So the MLE of the variance is not unbiased. However it is asymptotically unbiased!
- Also you can have another estimator $\sum_i (x_i - \bar{x})^2/(n-1)$, which is not the MLE, but it is unbiased.