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

SDS 321: Introduction to Probability and Statistics
Lecture 21: Frequentist statistics: introduction

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

www.cs.cmu.edu/∼psarkar/teaching

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).
- Note that this is not a conditional probability like in Bayesian statistics, simply because θ is no longer random.
- In Bayesian statistics we were dealing with one probabilistic model.
- 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.

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 $\mathrm{var}_\theta(\hat{\Theta}_n)$.
- 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 $\lim_{n\to\infty} E[\hat{\Theta}_n] = \theta$

Bias Variance decomposition

- An estimator is a random quantity whose distribution depends on θ. One can have many estimators for the same parameter.
- A better estimator is one whose MSE (mean square error) is smaller.
- It can be shown that $E[(\hat{\theta} - \theta)^2] = b_\theta(\hat{\theta})^2 + \text{var}(\hat{\theta})$. This is the famous bias-variance decomposition. Can we derive it?

Bias Variance decomposition

- An estimator is a random quantity whose distribution depends on θ. One can have many estimators for the same parameter.
- A better estimator is one whose MSE (mean square error) is smaller.
- It can be shown that $E[(\hat{\theta}-\theta)^2] = b_\theta(\hat{\theta})^2 + \mathrm{var}(\hat{\theta})$. This is the famous bias-variance decomposition. Can we derive it?

  $E[(\hat{\theta}-\theta)^2] = E[(\hat{\theta}-E[\hat{\theta}]+E[\hat{\theta}]-\theta])^2] = E[((\hat{\theta}-E[\hat{\theta}])+(E[\hat{\theta}]-\theta))^2]$
- $$= \underbrace{E[(\hat{\theta}-E[\hat{\theta}])^2]}_{\mathrm{var}_\theta(\hat{\theta})} + \underbrace{(E[\hat{\theta}]-\theta)^2}_{\text{squared bias}} + 2\underbrace{E[(\hat{\theta}-E[\hat{\theta}])(E[\hat{\theta}]-\theta)]}_{0}$$

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \dots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \ldots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i / n$.

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \ldots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i / n$.
- Another estimator is to always use 0, no matter what the data is.

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \ldots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i/n$.
- Another estimator is to always use 0, no matter what the data is.
- $MSE(\bar{X}) = \mathrm{var}(\bar{X}) + b(\bar{X})^2 = \frac{\sigma^2}{n} + (E[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n}$

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \dots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i / n$.
- Another estimator is to always use 0, no matter what the data is.
- $MSE(\bar{X}) = \text{var}(\bar{X}) + b(\bar{X})^2 = \frac{\sigma^2}{n} + (E[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n}$
- $MSE(0) = \text{var}(0) + b(0)^2 = 0 + \mu^2$

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \ldots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i / n$.
- Another estimator is to always use 0, no matter what the data is.
- $MSE(\bar{X}) = \text{var}(\bar{X}) + b(\bar{X})^2 = \frac{\sigma^2}{n} + (E[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n}$
- $MSE(0) = \text{var}(0) + b(0)^2 = 0 + \mu^2$
- If $\mu = 0$, this is a great predictor, but otherwise its pretty crappy.

Bias variance decomposition

- Lets say we want to estimate the mean of a normal distribution from n iid random variables $X_1, \ldots, X_n$ drawn from a $N(\mu, \sigma^2)$ distribution.
- A reasonable choice is $\bar{X} = \sum_i X_i / n$.
- Another estimator is to always use 0, no matter what the data is.
- $MSE(\bar{X}) = \text{var}(\bar{X}) + b(\bar{X})^2 = \frac{\sigma^2}{n} + (E[\bar{X}] - \mu)^2 = \frac{\sigma^2}{n}$
- $MSE(0) = \text{var}(0) + b(0)^2 = 0 + \mu^2$
- If $\mu = 0$, this is a great predictor, but otherwise its pretty crappy.
- We want an estimator which works well for all values of μ.

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
  $$P(X_1, \ldots, X_n; p) = \prod_{i=1}^{n} p^{X_i}(1-p)^{1-X_i}$$
- Its often more convenient to maximizes the logarithm of a product form.
  $$\log P(X_1, \ldots, X_n; p) = \sum_i \log P(X_i; p)$$
- $$= \sum_i \log\left(p^{X_i}(1-p)^{1-X_i}\right) = \sum_i X_i \log p + \sum_i (1-X_i)\log(1-p)$$
- $\hat{p} = \arg\max_p \log P(X_1, \ldots, X_n; p)$

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.
  $\log P(X_1, \ldots, X_n; p) = \sum_i \log P(X_i; p)$
- $$= \sum_i \log\left(p^{X_i}(1-p)^{1-X_i}\right) = \sum_i X_i \log p + \sum_i (1-X_i)\log(1-p)$$
- $\hat{p} = \arg\max_p \log P(X_1, \ldots, X_n; p)$
- $$\frac{\sum_i X_i}{\hat{p}} - \frac{n - \sum_i X_i}{1-\hat{p}} = 0 \rightarrow \hat{p} = \frac{\sum_i X_i}{n}$$

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

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.

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

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

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.

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

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

I have n iid random variables from a $N(\mu, \sigma^2)$ distribution. I 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.

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

- First derive w.r.t μ and set to zero to solve for $\hat{\mu}$. $-\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$

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.
  $$\log f_X(x; \mu, \sigma) = -\log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$$
- 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 $\hat{\sigma}$.
  $$-\frac{1}{\sigma} + \sum_i 2\frac{(x_i - n\mu)^2}{2\sigma^3} = 0$$

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.
  $$\log f_X(x; \mu, \sigma) = -\log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$$
- 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 $\hat{\sigma}$.
  $$-\frac{1}{\sigma} + \sum_i 2\frac{(x_i - n\mu)^2}{2\sigma^3} = 0$$
- Now you have two equations and two unknowns!

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

$$-\frac{1}{\sigma} + \sum_i \frac{(x_i - \mu)^2}{\sigma^3} = 0$$

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.
  $$\log f_X(x; \mu, \sigma) = -\log(\sqrt{2\pi\sigma^2}) - \sum_i \frac{(x_i - \mu)^2}{2\sigma^2}$$
- 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 $\hat{\sigma}$.
  $$-\frac{1}{\sigma} + \sum_i 2\frac{(x_i - n\mu)^2}{2\sigma^3} = 0$$
- Now you have two equations and two unknowns!
  $$\frac{\sum_i (x_i - \mu)}{\sigma^2} = 0$$
  $$-\frac{1}{\sigma} + \sum_i \frac{(x_i - \mu)^2}{\sigma^3} = 0$$
- Solving we see:
  $$\hat{\mu} = \sum_i x_i/n = \bar{x} \qquad \hat{\sigma}^2 = \frac{\sum_i (x_i - \bar{x})^2}{n}$$

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. $n\frac{1}{\hat{\lambda}} - \sum_i x_i = 0 \rightarrow \hat{\lambda} = \frac{1}{\sum_i x_i/n}$.

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 $f(X_i; a) = \dfrac{1\{a \leq X_i \leq 1\}}{(1-a)}$

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 $f(X_i; a) = \dfrac{1\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} 1\{a \leq X_i \leq 1\} =$
  $\dfrac{1\{a \leq \min(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$
- 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.

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 $f(X_i; a) = \dfrac{1\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} 1\{a \leq X_i \leq 1\} = \dfrac{1\{a \leq \min(X_1, X_2, \ldots) \leq 1\}}{(1-a)^n}$
- 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 $1/(1-\hat{a})^n$?

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 $f(X_i; a) = \dfrac{1\{a \leq X_i \leq 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \dots X_n; a) = \dfrac{1}{(1-a)^n} \prod_{i=1}^{n} 1\{a \leq X_i \leq 1\} =$ $\dfrac{1\{a \leq \min(X_1, X_2, \dots) \leq 1\}}{(1-a)^n}$
- 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 $1/(1-\hat{a})^n$?
- For all $\hat{a} \leq \min(X_1, \dots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.

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 $f(X_i; a) = \frac{1\{a \le X_i \le 1\}}{(1-a)}$
- Now write $f(X_1, X_2, \ldots X_n; a) = \frac{1}{(1-a)^n} \prod_{i=1}^{n} 1\{a \le X_i \le 1\} =$
  $\frac{1\{a \le \min(X_1, X_2, \ldots) \le 1\}}{(1-a)^n}$
- 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 $\hat{a} \le \min(X_1, \ldots, X_n)$, what maximizes $1/(1-\hat{a})^n$?
- For all $\hat{a} \le \min(X_1, \ldots, X_n)$, what is the largest $\hat{a}$? (larger $\hat{a}$ smaller $1/(1-\hat{a})^n$.
- $\hat{a} = \min(X_1, \ldots, X_n)$!

Desirable properties of the MLE

- **Invariance principle:** For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$
- **Unbiased**: we want $E[\hat{\theta}] = \theta$, i.e. the bias $b_\theta(\hat{\theta}) = 0$.
- If not **Unbiased**, the estimator should be at least **asymptotically unbiased**, i.e. $\lim_{n\to\infty} b_\theta(\hat{\theta}) = 0$
- **Consistency:** As $n \to \infty$, $\hat{\theta}$ converges to θ

Invariance principle

For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$. We worked with an exponential distribution before. Lets look at another parametrization of it. We have iid r.v's $X_1, \ldots, X_n$ from the Exponential distribution.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$. But another parametrization of this is: $f_{X_i}(x_i; \theta) = 1/\theta e^{-x_i/\theta}$
- We want MLE of θ. If the MLE obeys the invariance principle, then we should have: $\hat{\theta} = 1/\hat{\lambda}$ since $\theta = 1/\lambda$ is a one-to-one function.

Invariance principle

For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$. We worked with an exponential distribution before. Lets look at another parametrization of it. We have iid r.v's $X_1, \ldots, X_n$ from the Exponential distribution.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$. But another parametrization of this is: $f_{X_i}(x_i; \theta) = 1/\theta e^{-x_i/\theta}$
- We want MLE of θ. If the MLE obeys the invariance principle, then we should have: $\hat{\theta} = 1/\hat{\lambda}$ since $\theta = 1/\lambda$ is a one-to-one function.
- Now write the joint likelihood $f_X(x; \theta) = \prod_i 1/\theta e^{-x_i/\theta} = \theta^{-n} \prod_i e^{-x_i/\theta}$

Invariance principle

For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$. We worked with an exponential distribution before. Lets look at another parametrization of it. We have iid r.v's $X_1, \ldots, X_n$ from the Exponential distribution.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$. But another parametrization of this is: $f_{X_i}(x_i; \theta) = 1/\theta e^{-x_i/\theta}$
- We want MLE of θ. If the MLE obeys the invariance principle, then we should have: $\hat{\theta} = 1/\hat{\lambda}$ since $\theta = 1/\lambda$ is a one-to-one function.
- Now write the joint likelihood $f_X(x; \theta) = \prod_i 1/\theta e^{-x_i/\theta} = \theta^{-n} \prod_i e^{-x_i/\theta}$
- Now take logarithm of the joint likelihood. $\log f_X(x; \theta) = -n \log \theta - \sum_i x_i/\theta$.

Invariance principle

For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$. We worked with an exponential distribution before. Lets look at another parametrization of it. We have iid r.v's $X_1, \ldots, X_n$ from the Exponential distribution.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$. But another parametrization of this is: $f_{X_i}(x_i; \theta) = 1/\theta e^{-x_i/\theta}$
- We want MLE of θ. If the MLE obeys the invariance principle, then we should have: $\hat{\theta} = 1/\hat{\lambda}$ since $\theta = 1/\lambda$ is a one-to-one function.
- Now write the joint likelihood $f_X(x; \theta) = \prod_i 1/\theta e^{-x_i/\theta} = \theta^{-n} \prod_i e^{-x_i/\theta}$
- Now take logarithm of the joint likelihood. $\log f_X(x; \theta) = -n \log \theta - \sum_i x_i/\theta$.
- Differentiate and set to zero to get the MLE. $-n\frac{1}{\hat{\theta}} + \frac{\sum_i x_i}{\hat{\theta}^2} = 0 \to \hat{\theta} = \sum_i x_i/n = 1/\hat{\lambda}$.

Invariance principle

For a one-to-one function h of θ, the MLE of $\xi = h(\theta)$ should be given by $\hat{\xi} = h(\hat{\theta})$. We worked with an exponential distribution before. Lets look at another parametrization of it. We have iid r.v's $X_1, \ldots, X_n$ from the Exponential distribution.

- First write $f_{X_i}(x_i; \lambda) = \lambda e^{-\lambda x_i}$. But another parametrization of this is: $f_{X_i}(x_i; \theta) = 1/\theta e^{-x_i/\theta}$
- We want MLE of θ. If the MLE obeys the invariance principle, then we should have: $\hat{\theta} = 1/\hat{\lambda}$ since $\theta = 1/\lambda$ is a one-to-one function.
- Now write the joint likelihood $f_X(x; \theta) = \prod_i 1/\theta e^{-x_i/\theta} = \theta^{-n} \prod_i e^{-x_i/\theta}$
- Now take logarithm of the joint likelihood. $\log f_X(x; \theta) = -n \log \theta - \sum_i x_i/\theta$.
- Differentiate and set to zero to get the MLE. $-n\frac{1}{\hat{\theta}} + \frac{\sum_i x_i}{\hat{\theta}^2} = 0 \to \hat{\theta} = \sum_i x_i/n = 1/\hat{\lambda}$.
- It can be shown that the MLE follows the invariance principle.

Unbiased and asymptotically unbiased estimators

Recall that the MLE estimate of μ obtained using n iid random variables from a $N(\mu, \sigma^2)$ distribution are given by: $\hat{\mu} = \frac{\sum_i x_i}{n}$

- Using linearity of expectation $E[\hat{\mu}] = \mu$. So this MLE of μ is unbiased.
- How about the Bayesian MAP estimate? For a prior of $N(\mu_0, \sigma_0^2)$ on μ, we saw that

$$\mu_{MAP} = \left(\sum_i \frac{x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right) \Big/ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)$$

Unbiased and asymptotically unbiased estimators

Recall that the MLE estimate of μ obtained using n iid random variables from a $N(\mu, \sigma^2)$ distribution are given by: $\hat{\mu} = \frac{\sum_i x_i}{n}$

- Using linearity of expectation $E[\hat{\mu}] = \mu$. So this MLE of μ is unbiased.
- How about the Bayesian MAP estimate? For a prior of $N(\mu_0, \sigma_0^2)$ on μ, we saw that

$$\mu_{MAP} = \left(\sum_i \frac{x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2} \right) \Big/ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2} \right)$$

- $E[\mu_{MAP}] = \frac{n\mu/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} \neq \mu$. So this is biased.

Unbiased and asymptotically unbiased estimators

Recall that the MLE estimate of μ obtained using n iid random variables from a $N(\mu, \sigma^2)$ distribution are given by: $\hat{\mu} = \frac{\sum_i x_i}{n}$

- Using linearity of expectation $E[\hat{\mu}] = \mu$. So this MLE of μ is unbiased.
- How about the Bayesian MAP estimate? For a prior of $N(\mu_0, \sigma_0^2)$ on μ, we saw that

$$\mu_{MAP} = \left(\sum_i \frac{x_i}{\sigma^2} + \frac{\mu_0}{\sigma_0^2}\right) \Bigg/ \left(\frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}\right)$$

- $E[\mu_{MAP}] = \frac{n\mu/\sigma^2 + \mu_0/\sigma_0^2}{n/\sigma^2 + 1/\sigma_0^2} \neq \mu$. So this is biased.
- But $E[\mu_{MAP}] = \frac{\mu + \mu_0/\sigma_0^2 \times \sigma^2/n}{1 + 1/\sigma_0^2 \times \sigma^2/n} \to \mu \qquad$ as $n \to \infty$ So it is asymptotically 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?

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 $E[\bar{X}^2] = \mathrm{var}(\bar{X}) + E[\bar{X}]^2 = \frac{\sigma^2}{n} + \mu^2$
- And so, $E[\hat{\sigma}^2] = \sigma^2 - \sigma^2/n = \sigma^2(1 - 1/n)$

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!

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.