General methods for finding parameter Estimators

1) method of moments (MOM)
2.) maximum likelihood (MLE)

method of moments

Def

The k^{th} moment of a R.V. X is:

$$\mu_k = \mathbb{E}(X^k)$$

The k^{th} sample moment is

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

method of moment Estimators for model parameters are found by equating known sample moments to unknown population moments and solving for the parameters in terms of the data.

— In general, we need k equations to derive MOM estimators for k parameters. we need to solve the following system of equations

$$\begin{aligned} m_1 &= \mathbb{E}(X_1) \\ m_2 &= \mathbb{E}(X_1^2) \\ &\vdots \\ m_k &= \mathbb{E}(X_1^k) \end{aligned}$$

where X_1 is just a R.V. from our model $f_X(x)$

MOM can be biased and sometimes you can get Estimates outside of the parameter space, but typically they yield some kind of estimator

Ex

Let $X_1, \cdots, X_n \overset{iid}{\sim} geo(p)$

we only need an estimator for one parameter, p, so we only have to use the first moments

$$\left.\begin{aligned} m_1 &= \frac{1}{n}\sum X_i = \overline{X} \\ \mu_1 &= \mathbb{E}(X_1) = \frac{1}{p} \end{aligned}\right\} \Rightarrow \text{set } \overline{X} = \frac{1}{p} \text{ and solve for } p$$

$$\Rightarrow p = \frac{1}{\overline{X}} \Rightarrow \boxed{\hat{p}_{mom} = \frac{1}{\overline{X}}}$$

Ex

Let $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$

Now we have two parameters $\Rightarrow$ need first two moments

(1) $\frac{1}{n}\sum X_i = \mathbb{E}(X_1)$

(2) $\frac{1}{n}\sum X_i^2 = \mathbb{E}(X_1^2)$

from equation (1) we get $\boxed{\hat{\mu}_{mom} = \overline{X}}$

from equation (2) we have

$$\frac{1}{n}\sum X_i^2 = \sigma^2 + \mu^2$$

[Recall: $Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$
$\Rightarrow \mathbb{E}(X^2) = Var(X) + [\mathbb{E}(X)]^2 = \sigma^2 + \mu^2$]

$$\Rightarrow \sigma^2 = \frac{1}{n}\sum X_i^2 - [\overline{X}]^2$$

[plug in mom Estimator of μ]

$$\Rightarrow \boxed{\hat{\sigma}^2_{mom} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

["The Biased version of the Sample ~~star~~ Variance"]

Maximum likelihood Estimators (MLE's)

we have a random sample of data

$X_1 \cdots X_n \overset{iid}{\sim} f_X(x;\theta)$ where θ is an unknown parameter

The model for the entire sample is the joint distribution

$$f_X(x_1 \cdots x_n) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

Ex

Pois(λ): $\quad P_X(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

$$X_1 \cdots X_n \overset{iid}{\sim} pois(\lambda) \Rightarrow f_X(x_1 \cdots x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$$

$$\Rightarrow \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdot \frac{e^{-\lambda}\lambda^{x_2}}{x_2!} \cdot \cdots \cdot \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} = \boxed{\frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n} x_i!}}$$

A maximum likelihood estimator, $\hat{\theta}_{MLE}$, of θ is the value that "maximizes the probability of the data" i.e. maximizes our joint distribution model

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{n} f_X(x_i;\theta)$$

---

when we gather data, we can treat them as constants and plug them into the joint model. Then when viewed as a function of the parameter instead of the data, we call the joint distribution the <u>likelihood function</u>

$$\boxed{\text{likelihood function: } \quad L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)}$$

Ex

Flip a coin 10 times, $p = P(\text{heads})$ unknown

$X_i = \begin{cases} 1 & \text{heads} \\ 0 & \text{tails} \end{cases}$ we observe 1, 0, 0, 1, 0, 1, 0, 0, 0, 0

model for X_i: $X_1 \cdots X_n \overset{iid}{\sim} Bern(p)$

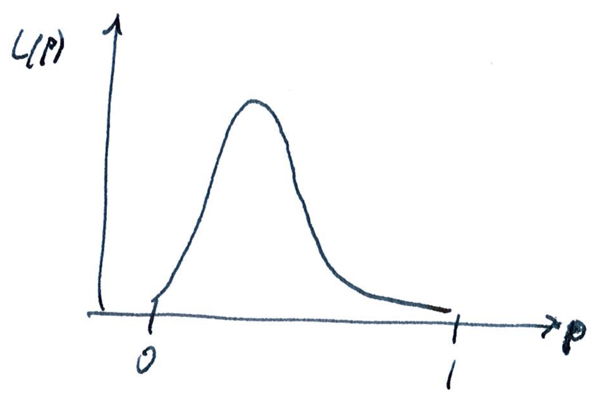
model for the entire Sample:

$$f_X(x_1 \cdots x_n) = \prod_{i=1}^{n} f_X(x_i; p) = \prod_{i=1}^{n} (1-p)^{1-x_i} p^{x_i} = (1-p)^{n-\sum x_i} p^{\sum x_i}$$

Plugging in the data and treating this as a function of p we get:

Likelihood function: $\boxed{L(p) = (1-p)^7 p^3}$

Plot of $L(p)$ vs $p \in [0,1]$



what value of p makes $L(p)$ the biggest?

$p = .2 \Rightarrow L(.2) = .201$

$p = .9 \Rightarrow L(.9) = .000008$

$\boxed{p = .3 \Rightarrow L(.3) = .267}$ *

$p = .4 \Rightarrow L(.4) = .215$

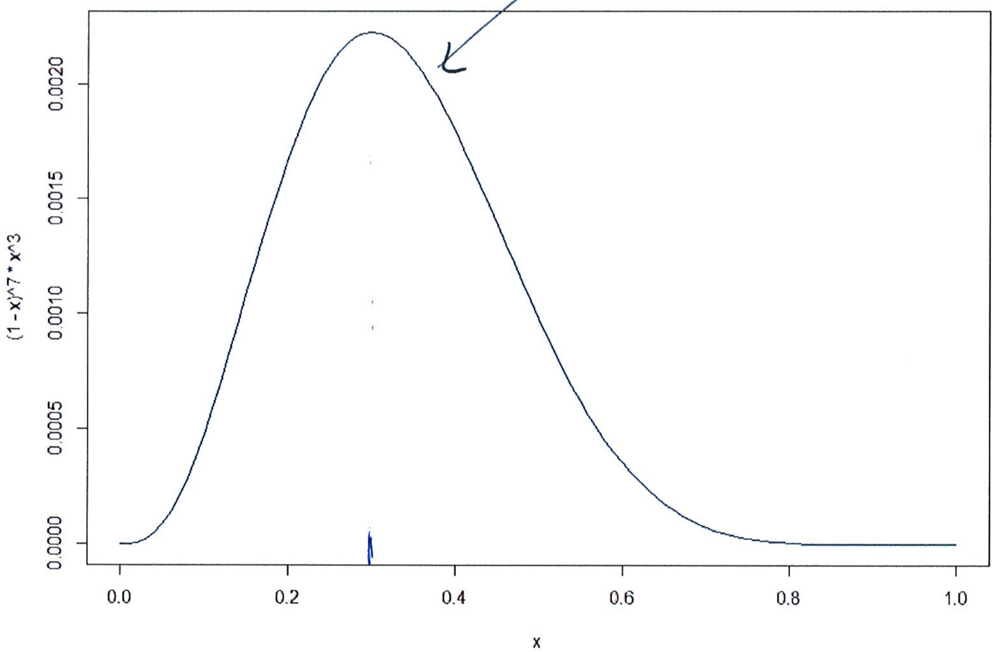
Based on our data, the MLE of p is .3

(See graphs on next page)

In practice, we work with the log-likelihood: $\ell(\theta) = \log(L(\theta))$ where $\log(\cdot) \equiv \ln(\cdot)$. working with the log-likelihood is computationally easier.

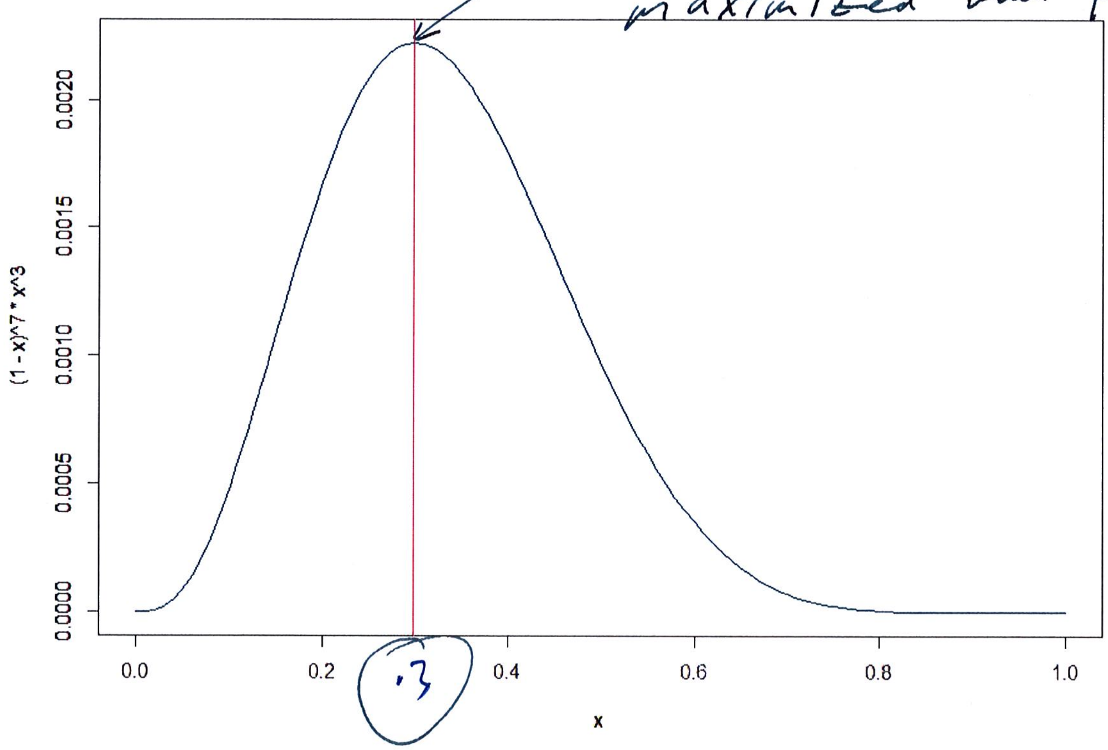
curve((1-x)^7*x^3,0,1,col="darkblue")

likelihood function:
$L(p) = (1-p)^7 p^3$



abline(v=.3,col="red")

likelihood function is maximized when $p = .3$



Finding MLEs

1.) Find likelihood function: $L(\theta) = \prod_{i=1}^{n} f(x_i)$

2.) Take log: $\ell(\theta) = \log(L(\theta))$

3.) take derivative of $\ell(\theta)$ with respect to θ

4.) set equal to 0

5.) Solve for θ

[technically we need to make sure 2nd derivative < 0 at $\hat{\theta}$]

Ex Roll a die untill a six. Do it 100 times

X_i = # of Rolls till a six

Data

x	1	2	...	29
# of times	18	20	...	1

} use data to Estimate p = probability of rolling a six

model for data

$X_1 \cdots X_n \sim geo(p)$ $\qquad p(x) = p(1-p)^{x-1}$

1.) $L(p) = \prod_{i=1}^{n} p(1-p)^{x-1} = p^n (1-p)^{\sum x_i - n}$

2.) $\ell(p) = \log(L(p)) = n\log(p) + (\sum x_i - n)\log(1-p)$

3.) $\frac{\partial \ell(p)}{\partial p} = \frac{n}{p} - \frac{\sum x_i - n}{1-p}$

4.) Set $= 0$, $\Rightarrow \frac{n}{p} = \frac{\sum x_i - n}{1-p}$

5.) Solve for p:

$\frac{1-p}{p} = \frac{\sum x_i - n}{n} \Rightarrow \frac{1}{p} - 1 = \frac{\sum x_i}{n} - 1 \Rightarrow \frac{1}{p} = \frac{\sum x_i}{n}$

$\Rightarrow \boxed{\hat{p}_{ml} = \frac{n}{\sum x_i} = \frac{1}{\bar{X}}}$

Then plug in data

$n = 100$, $\bar{X} = 5.68$

$\sum x_i = 568$

Thus $\boxed{\hat{p}_{ml} = .176}$

We may have more than 1 parameter to Estimate

EX

$X_1 \cdots X_n \sim N(\mu, \sigma^2)$ μ & σ^2 unknown

$$f_X(x_i) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

1.) likelihood function:

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f_X(x_i) = (2\pi)^{-n/2}\sigma^{-n} \exp\left\{-\frac{\Sigma(x_i-\mu)^2}{2\sigma^2}\right\}$$

2.) log-likelihood:

$$\ell(\mu,\sigma^2) = \log(L(\mu,\sigma^2)) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{\Sigma(x_i-\mu)^2}{2\sigma^2}$$

3.) take derivative, this time with respect to both parameters

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \mu} = -\left[\frac{2\sigma^2(-2)\Sigma(x_i-\mu) - 0}{4\sigma^4}\right] = \frac{\Sigma(x_i-\mu)}{\sigma^2}$$

$$\frac{\partial \ell(\mu,\sigma^2)}{\partial \sigma} = \frac{-n}{\sigma} - \left[\frac{2\sigma^2(0) - \Sigma(x_i-\mu)^2(4\sigma)}{4\sigma^4}\right]$$

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

NOW we have to find the values that simultaneously give zeros.

Start with the first equation.

i.) $\Sigma(x_i - \mu) = 0$

$$\Rightarrow \Sigma x_i - n\mu = 0 \Rightarrow \Sigma x_i = n\mu$$

$$\Rightarrow \hat{\mu}_{ML} = \frac{\Sigma x_i}{n} = \boxed{\bar{X}}$$

Now plug in the maximizer of the first equation into the second

(ii) $$-\frac{n}{\sigma} + \frac{\sum(x_i - \bar{x})^2}{\sigma^3} = 0$$

$$\Rightarrow \frac{\sum(x_i-\bar{x})^2}{\sigma^3} = \frac{n}{\sigma}$$

$$\Rightarrow \hat{\sigma}^2_{MLE} = \boxed{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}$$