CHAPTER 3

Estimators

- Statistical model = family P_θ

  Data generated from $\boxed{\text{one}}$ of the distributions in the model

Parametric model : $\{ P_\theta : \theta \in \Theta \}$
  if $\Theta \subseteq \mathbb{R}^d$

Estimation = process of determining the parameter θ that gives the best fitting model.

- Often we want to estimate a function $g(\theta)$ given the data x

ESTIMATOR or STATISTIC for $g(\theta)$
is a random vector $T(X)$
that depends ONLY on
$\underline{X} = (X_1, \dots, X_n)$

ESTIMATE
is the observed value $T(x)$

Both are indicated also by $\hat{\theta}$

Methods of Estimation

- maximum likelihood
- method of moments
- Bayes method

3.2 Mean Square Error

a good estimator must be "close" to the estimand $g(\theta)$

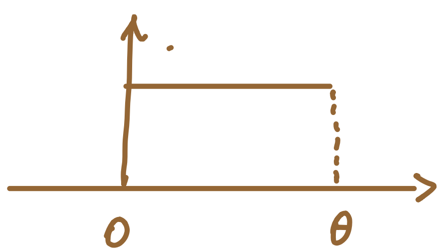
To measure the distance we consider the distribution of the distance

$$\| T(X) - g(\theta) \|^2$$

under the assumption the θ is the true parameter

Example 3.2.

Let $X = (X_1, \ldots, X_n) \overset{iid}{\sim} \mathrm{Unif}(0, \theta)$



$$E(X_i) = \frac{\theta}{2}$$

Idea : Estimate $\frac{\theta}{2}$ by $\bar{X}_n = \frac{\sum X_i}{n}$

or Estimate θ by $2\bar{X}_n$

Justified by the Law of large numbers

$$\bar{X}_n \xrightarrow{}_{p} \mu = \frac{\theta}{2} \quad \text{for } n \to \infty$$

For instance,

$x = (3.03,\ 2.7,\ 7,\ 1.59,\ 5.04,\ 5.92,$
$9.82,\ 1.11,\ 4.26,\ 6.96) \qquad n = 10$

$$2\bar{x} = 9.49$$

Note: this underestimates θ
because $9.49 < x_7 = 9.82$

Idea 2: take $\hat{\theta} = X_{(n)} = \max_i x_i$

But, again $X_{(n)} < \theta$.

We could correct $X_{(n)} \to \dfrac{n+2}{n+1} X_{(n)}$

$\Rightarrow$ which estimator is best ?

IDEA: the best estimator is the one with a sampling distribution most concentrated on θ "in the long run"; or that is best on average

See Notebook

Note: our simulation shows that if $\theta = 1$ $T_1 = \frac{n+2}{n+1} X_{(n)}$ is better than $T_2 = 2\bar{X}_n$

Mean square error

of an estimator T for $g(\theta)$ is

$$MSE(\theta; T) = E_\theta \|T - g(\theta)\|^2$$

this is a function of θ.

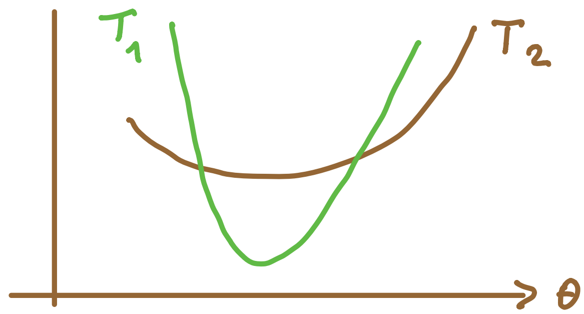
Note : the subscripted E_θ is essential. This means that we compute the expected square deviation of T from $g(\theta)$ under the assumption that θ is the true value of the parameter.

$$\boxed{\text{We prefer an estimator with MSE small for ALL values of } \theta.}$$

If: $MSE[\theta, T_1] \leq MSE[\theta, T_2]$ for all θ
with a strict inequality for at least one value of θ,

we prefer T_1 and we say that T_2 is inadmissible

- Sometimes this does not happen



Result: the MSE can be decomposed

$$\boxed{MSE(\theta, T) = \underbrace{var_\theta T}_{\text{variance of the estimator}} + \underbrace{\left[E_\theta T - g(\theta)\right]^2}_{\text{bias}}}$$

Proof (scalar): $E_\theta (T - \theta)^2 = E_\theta U^2$

We know that

$$var_\theta (U) = E_\theta U^2 - (E_\theta U)^2$$

$$\therefore \quad E_\theta U^2 = var_\theta U + (E_\theta U)^2$$

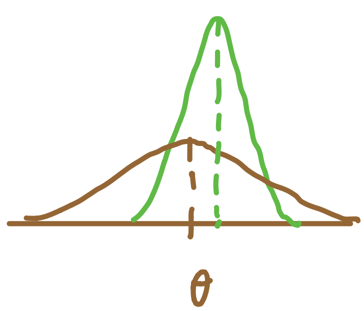
Thus:

$$E_\theta(T-\theta)^2 = \mathrm{var}_\theta(T-\theta) + [E_\theta(T-\theta)]^2$$
$$= \mathrm{var}_\theta T + [E_\theta T - \theta]^2$$

Unbiased estimator

T is unbiased for $g(\theta)$ if

$$E_\theta T = g(\theta) \text{ for all } \theta \in \Theta$$



While unbiased estimators look very desirable, there is a trade-off between variance and bias.

Standard error

The standard deviation of an estimator T: $\sigma_\theta(T) = \sqrt{\mathrm{var}_\theta T}$ is called the standard error

The standard error gives an idea of the quality of an estimate.

Example: $(X_1, \dots, X_n) \overset{iid}{\sim} N(\theta, \sigma_0^2)$

$\sigma_0^2 =$ known variance

$\bar{X}$ is an unbiased estimator of θ (the mean) because

$$E_\theta(\bar{X}) = \frac{1}{n} E_\theta \sum_{i=1}^{n} X_i = \frac{1}{n} \sum_i E X_i = \frac{n\theta}{n} = \theta$$

The standard error is

$$\sigma_\theta(\bar{X}) = \sqrt{\mathrm{var}_\theta \bar{X}} = \sqrt{\frac{1}{n^2} \mathrm{var}(\Sigma_i X_i)}$$

$$= \sqrt{\frac{\sigma_0^2}{n}} = \frac{\sigma_0}{\sqrt{n}}$$

This is the variability of $\bar{X}$ around θ in repeated sampling.

Example 3.6

1) $(X_1, \dots, X_n) \overset{iid}{\sim} \mathcal{U}(0, \theta)$

$2\bar{X}$ is unbiased for θ

$$\rightarrow \quad E_\theta(2\bar{X}) = 2E_\theta(\bar{X}) = \frac{2}{n}\sum_i E_\theta X_i$$

$$= \frac{2}{n}\sum_{i=1}^{n}\frac{\theta}{2} = \theta.$$

$$\rightarrow \quad MSE(\theta, 2\bar{X}) = var_\theta(2\bar{X}) = 4\, var_\theta(\bar{X})$$

$$= 4 \cdot \frac{var(X_1)}{n}$$

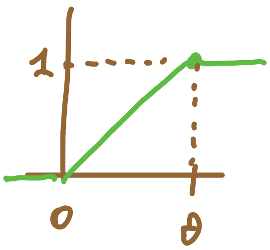
As $var\, X_1 = \dfrac{\theta^2}{12}$

$$= \frac{4\theta^2}{12n} = \frac{\theta^2}{3n}.$$

2) Is $X_{(n)}$ unbiased? The density of $X_{(n)}$ can be found in this way

$$F_{(n)}(x) = P(X_{(n)} \leq x)$$

$$= P(X_1 \leq x, X_2 \leq x, \dots X_n \leq x)$$



$$= \prod_{i=1}^{n} P(X_i \leq x)$$

$$= \prod_{i=1}^{n} \frac{x}{\theta} = \frac{x^n}{\theta^n}$$

So that

$$f_{(n)}(x) = \frac{d}{dx} \frac{x^n}{\theta^n} = \frac{n x^{n-1}}{\theta^n}.$$

The Expected value is

$$\int_0^\theta x \cdot \frac{n x^{n-1}}{\theta^n} dx = \frac{n}{\theta^n} \cdot \frac{x^{n+1}}{n+1}\Big|_0^\theta$$

$$= \frac{n}{n+1} \theta.$$

→ So $X_{(n)}$ is biased

→ The MSE is $\mathrm{var}_\theta X_{(n)} + \left(\frac{n}{n+1}\theta - \theta\right)^2$

The variance is

$$\mathrm{var}_\theta\left(X_{(n)}\right) = E_\theta X_{(n)}^2 - \left[E_\theta X_{(n)}\right]^2$$

$$= \int_0^\theta x^2 \cdot \frac{n x^{n-1}}{\theta^n} dx - \left[\frac{n\theta}{n+1}\right]^2$$

$$= \frac{n}{\theta^n} \int_0^\theta x^{n+1} dx - \frac{n^2\theta^2}{(n+1)^2}$$

$$= \frac{n}{\theta^n} \left. \frac{x^{n+2}}{n+2} \right|_0^\theta - \frac{n^2\theta^2}{(n+1)^2}$$

$$= \frac{n\,\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}$$

$$= \theta^2 \frac{n}{(n+2)(n+1)^2}.$$

And finally:

$$MSE(\theta, X_{(n)}) = \theta^2 \frac{n}{(n+2)(n+1)^2} + \left(\frac{\theta n}{n+1} - \theta\right)^2$$

$$= \frac{2\theta^2}{(n+2)(n+1)}.$$

Compare

$MSE(\theta, 2\bar{X})$ $\qquad$ $MSE(\theta, X_{(n)})$

$\dfrac{\theta^2}{3n}$ $\qquad$ $\dfrac{2\theta^2}{(n+2)(n+1)}$

See notebook

Other estimators

→ $\dfrac{n+1}{n} X_{(n)}$ unbiased

→ $\dfrac{n+2}{n+1} X_{(n)}$ better

MSE $= \dfrac{\theta^2}{(n+1)^2}$

Example 3.7

In general $(X_1, \dots, X_n) \overset{iid}{\sim} F$
with mean and variance

$$\mu = E_F(X_i) \qquad \sigma^2 = var_F(X_i)$$

: a nonparametric model

→ Two estimators of μ and σ^2

$$\bar{X} = \frac{1}{n} \sum_i X_i \qquad S_X^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

→ $\bar{X}$ is unbiased for μ

→ $MSE(F; \bar{X}) = var_F(\bar{X})$

$$= \frac{1}{n^2} \sum_i var_F(X_i) = \frac{\sigma^2}{n}$$

$\boxed{\text{The precision of the estimator } \bar{X} \text{ increases by a factor of } \sqrt{n}}$

→ The sample variance is unbiased for σ^2. Proof:

$$s_x^2 = \frac{1}{n-1}\left[\sum_i (X_i-\mu)^2 - n(\bar{X}-\mu)^2\right]$$

ESERCIZIO

$$E_F(s_x^2) = \frac{1}{n-1}\left[\sum_i E_F(X_i-\mu)^2 - n\, E_F(\bar{X}-\mu)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} \sigma^2 - n\, var_F(\bar{X})\right]$$

$$= \frac{1}{n-1}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \sigma^2. \quad //$$

Note that unbiasedness is not preserved under transformation

$$E_F(S_x) \neq \sigma$$

$$E_F(\bar{X}^2) \neq \mu^2$$

3.3. Maximum likelihood

Is the most common method for finding estimators.

→ Is based on a function called the likelihood function

Given the data $\underline{x} = (x_1, \dots, x_n)$ and a parametric model $P_\theta \leftrightarrow p_\theta$ the likelihood is the density

$$p_\theta(\underline{x})$$

of the observation, as a function of $\theta \in \Theta$.

Example 3.9 – Binomial

Data after tossing a biased coin 10 times:

(H T T H T T H T T T)

So we get 3 heads.

Let $x = 3$ and the model is

$$X \sim \text{Bin}(10, p)$$

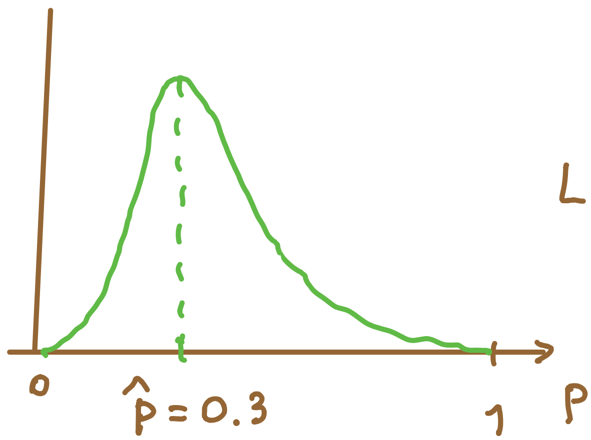
with $p =$ prob. of success unknown.

What's the probability of getting 3 heads on 10 tosses?

$$P_p(X=3) = \binom{10}{3} p^3 (1-p)^7$$

is a function of p

See the notebook



Likelihood

$$L(\theta) = \binom{10}{3} p^3 (1-p)^7$$

$\hat{p} = 0.3$ is the max probability of getting 3 heads

If $\underline{X}$ is a random vector with density* $p_\theta(\underline{x})$, $\theta \in \Theta$

the likelihood function is the function

$$\theta \longmapsto L(\theta; \underline{x}) = p_\theta(\underline{x})$$

for $\underline{x}$ fixed at the observed value

* we talk of density both for continuous / discrete

i.i.d Sample

If $\underset{\sim}{X} = (X_1, \dots X_n) \overset{i.i.d}{\sim} p_\theta(x)$

the likelihood is

$$L(\theta; \underline{x}) = \prod_{i=1}^{n} p_\theta(x_i)$$

Here $p_\theta(x_i)$ is

the marginal density of X_i.

$$\boxed{\begin{array}{l}\text{Maximum likelihood estimate (MLE)} \\ \quad \text{Is the value } T(\underline{x}) \in \Theta \text{ that} \\ \quad \text{maximizes } L(\theta; \underline{x}).\end{array}}$$

The maximum likelihood
is an intuitively reasonable
principle for finding estimators.
MLE are not necessarily the best ones.

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, L(\theta; \underline{x}) = \underset{\theta}{\operatorname{argmax}}\, \log L(\theta; \underline{x})$$

likelihood log likelihood

If $L(\theta)$ is differentiable in $\Theta \subset \mathbb{R}^k$ and it takes the maximum in an interior point of Θ then

$$\left.\frac{\partial}{\partial \theta_j} \log L(\underline{\theta}; \underline{x})\right|_{\underline{\theta} = \underline{\hat{\theta}}} = 0 \qquad j = 1, \dots, k$$

System of likelihood equations

NB - cannot be solved always explicitly

→ Check the form of the likelihood

To verify if a solution is the max.

→ If the solution is the maximum the 2nd derivative is negative

→ All the eigenvalues of the Hessian are negative

Score function $= \frac{\partial}{\partial \theta_j} \log L(\underline{\theta}, \underline{x})$

the likelihood equation for an i.i.d sample

$$\boxed{\sum_{i=1}^{n} \frac{\partial}{\partial \theta_j} \log p_\theta(x_i) = 0 \qquad j = 1, \dots, k}$$

Example 3.12 - Exponential

$(X_1, \dots, X_n) \overset{iid}{\sim} Exp(\lambda)$

$\lambda =$ rate $\qquad E_\lambda(X_i) = \frac{1}{\lambda} > 0$

$$\frac{d}{d\lambda} \log p_\theta(x_i) = \frac{d}{d\lambda} \log\left(\lambda e^{-\lambda x_i}\right)$$

$$= \frac{d}{d\lambda}\left(\log \lambda - \lambda x_i\right)$$

$$= \frac{1}{\lambda} - x_i$$

Likelihood equations

$$\sum_{i=1}^{n}\left(\frac{1}{\lambda} - x_i\right) = \frac{n}{\lambda} - \sum x_i = 0$$

1 Solution $= \dfrac{1}{\bar{x}}$

2nd derivative

$$\frac{d}{d\lambda}\left(\frac{1}{\lambda} - x_i\right) = -\frac{1}{\lambda^2}$$

So $\dfrac{d^2}{d\lambda^2} \log L(\underline{x}; \lambda) = -\dfrac{n}{\lambda^2} < 0, \ \forall \lambda$

so $\hat{\lambda}_{ML} = \dfrac{1}{\bar{x}}$.

//

Example 3.13 — Binomial.

x = # successes n = # trials

$$L(p; x) = \binom{n}{x} p^x (1-p)^{n-x}$$

$$\log L(p; x) = \ell(p) =$$

$$\log \binom{n}{x} + x \log p + (n-x) \log(1-p)$$

If $0 < x < n$,

$$\lim_{p \to 0} \ell(p) = -\infty \qquad \lim_{p \to 1} \ell(p) = -\infty$$

so the maximum is in $(0,1)$

Likelihood equation

$$\frac{d}{dp} \log L(p; x) = \frac{x}{p} - \frac{n-x}{1-p} = 0$$

Single solution $\hat{p} = \dfrac{x}{n} = \dfrac{\text{\# successes}}{\text{\# trials}}$

If

$x = 0$, $\ell(p) = n \log(1-p) \Rightarrow \hat{p} = 0$

$x = n$, $\ell(p) = n \log p \Rightarrow \hat{p} = 1$

Both can be written as $\hat{p} = \dfrac{x}{n}$. //

Example 3.14 - Normal distribution

$$(X_1, \dots, X_n) \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$$

$$(\mu, \sigma^2) \mapsto \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}(x_i - \mu)^2\right]$$

$$\Theta = \mathbb{R} \times (0, \infty)$$

log-likelihood

$$(\mu, \sigma^2) \mapsto -\frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2$$

Likelihood equations

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

$$\mu = \frac{\sum x_i}{n} = \bar{x}$$

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

for $\mu = \bar{x}$ likelihood has a maximum for every $\sigma > 0$

$$\text{Hessian} = \begin{pmatrix} \dfrac{\partial^2 \ell}{\partial \mu^2} & \dfrac{\partial^2 \ell}{\partial \mu \, \partial \sigma^2} \\[2ex] \dfrac{\partial^2 \ell}{\partial \sigma^2 \, \partial \mu} & \dfrac{\partial^2 \ell}{\partial (\sigma^2)^2} \end{pmatrix}$$

$$= \begin{pmatrix} \dfrac{-n}{\sigma^2} & -\dfrac{1}{\sigma^4} \sum_i (x_i - \mu) \\[2ex] -\dfrac{1}{\sigma^4} \sum_i (x_i - \mu) & \dfrac{n}{2\sigma^4} - \dfrac{1}{\sigma^6} \sum_i (x_i - \mu)^2 \end{pmatrix}$$

Substitute $\mu \to \bar{x}$ and $\sigma^2 \to \hat{\sigma}^2$

$$H = \begin{pmatrix} \dfrac{n}{\hat{\sigma}^2} & 0 \\[2ex] 0 & \dfrac{n}{2\hat{\sigma}^4} - \dfrac{n\hat{\sigma}^2}{\hat{\sigma}^6} \end{pmatrix}.$$

$$= \begin{pmatrix} -\dfrac{n}{\hat{\sigma}^2} & 0 \\[2ex] 0 & -\dfrac{n}{2\hat{\sigma}^4} \end{pmatrix}.$$

Both eigenvalues are negative so

$\left(\bar{x}, \dfrac{n-1}{n} S_x^2 \right)$ is the MLE //

Notes

1)... MLE can be taken outside the interior of Θ

2)... likelihood equations do not hold

3)... likelihood not everywhere differentiable

4)... MLE has several local maxima.

Example 3.15 – Uniform

$(X_1, \dots, X_n) \overset{iid}{\sim}$ Uniform on $[0, \theta]$

$\underset{\sim}{x} = (x_1, \dots, x_n) \to x_1 \leq \theta, x_2 \leq \theta, \dots, x_n \leq \theta$

$\Rightarrow \boxed{x_{(n)} \leq \theta}$

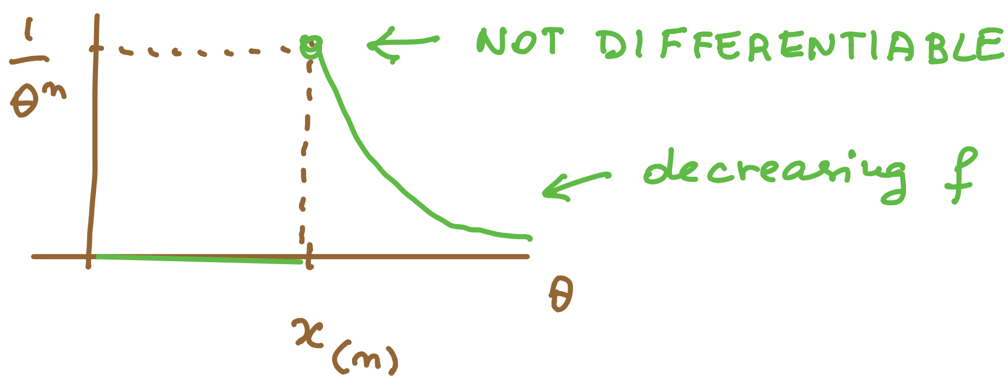
$$L(\theta ; \underset{\sim}{x}) = \prod_{i=1}^{n} p_\theta(x_i)$$

where $p_\theta(x_i) = \frac{1}{\theta} \cdot 1_{0 \leq x_i \leq \theta}$

Therefore:

$$L(\theta; \underset{\sim}{x}) = \left(\frac{1}{\theta}\right)^n \cdot 1_{x_1 \geq 0} \cdot 1_{x_{(n)} \leq \theta}$$

$$= \begin{cases} \left(\frac{1}{\theta}\right)^n & \text{if} \quad \theta \geq x_{(n)} \\ 0 & \text{if} \quad \theta < x_{(n)} \end{cases}$$



→ MLE for θ: $\quad \hat{\theta} = X_{(n)} = \max_i X_i$. //

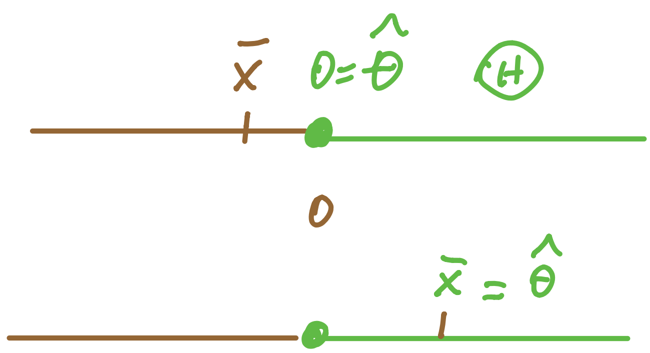
3.16 – Normal distrib. with restrictions

$(X_1, \dots, X_n) \overset{iid}{\sim} N(\mu, \sigma^2 = 1)$

under the restriction $\mu \geq 0$

$\Theta = [0, \infty)$

Given $\underline{x} = (x_1, \dots, x_n)$ it can happen that the unrestricted MLE $\hat{\theta} = \bar{x}$ is negative while $\mu \geq 0$. So this is NOT the restricted MLE



$$\text{MLE} = \max(0, \bar{x}).$$

$\left.\begin{array}{l}\text{The statistical model} \\ \text{the MLE}\end{array}\right\}$ depend on the form of Θ.

Equivariance of the MLE

What's the MLE of $g(\theta)$?

Let $g : \Theta \to H$ a bijective fun.

We parametrize the model with

$\eta = g(\theta)$, $\eta \in H$

$\hat{\theta}$ is MLE of θ $\Rightarrow$ $\hat{\eta} = g(\hat{\theta})$ is MLE of η

For any function g

Def.

$g(\hat{\theta})$ is the MLE of $g(\theta)$

Example 3.17 Exponential (again)

$(X_1, \dots, X_n) \overset{iid}{\sim} Exp(\lambda)$

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

$$\hat{\lambda}_{ML} = \frac{1}{\bar{X}} \Rightarrow \hat{\mu} = \bar{X}. \quad //$$

Example 3.18 - gamma

$(X_1, \dots, X_n) \overset{i.i.d}{\sim} \text{Gamma}(\alpha, \lambda)$

$$p_{\alpha,\lambda}(x) = \frac{x^{\alpha-1}}{\Gamma(\alpha)} \lambda^{\alpha} e^{-\lambda x}$$

λ = inverse scale

α = shape

Gamma function

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} \, ds$$

$$\log p_{\alpha,\lambda}(x_i) = (\alpha-1)\log x_i + \alpha \log \lambda - \\ - \lambda x_i - \log \Gamma(\alpha)$$

$$\ell(\alpha,\lambda) = (\alpha-1)\sum_i \log x_i + n\alpha \log\lambda - \\ - \lambda \sum_i x_i - n \log \Gamma(\alpha)$$

Parameter $\boxed{\begin{aligned} \theta &= (\alpha,\lambda) \in \Theta \\ \Theta &= [0,\infty)\times[0,\infty) \end{aligned}}$

$$\frac{\partial \ell}{\partial \alpha} = \sum_i \log x_i + n\log\lambda - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n\alpha}{\lambda} - \sum_i x_i = 0$$

From the 2nd equation

$$\lambda = \frac{\alpha}{\bar{X}}$$

substituting into the 1st eq.
we solve this equation

$$\sum_i \log x_i + n \log\left(\frac{\alpha}{\bar{x}}\right) - n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = 0$$

numerically.

See the notebook

Example 3.19 – Counting bacteria

With Petri dishes it is possible to see by naked eye colonies of bacteria.

Petri dish
(capsula di Petri)

Assumption : # colonies $\sim$ Poisson(μ)
in 1 cc

of contaminated water.

Dilute the water in 100 buckets so that

$$X_1, \ldots, X_{100} \overset{iid}{\sim} \text{Poisson}\left(\mu/_{100}\right)$$

And record

$Y_1, \ldots Y_{100}$ where $Y_i = \begin{cases} 0 & \text{if no coeonies} \\ 1 & \text{otherwise} \end{cases}$

Estimate μ from $(Y_1 \ldots Y_{100})$

$$Y_i \overset{iid}{\sim} \text{Bernoulli}\left(p = P(X_i = 1)\right)$$

So that $p = 1 - \dfrac{e^{-\lambda^0}\lambda^0}{0!} = 1 - e^{-\frac{\mu}{100}}$.

Therefore, $\hat{p} = \dfrac{\Sigma Y_i}{100}$

and $\left(1 - e^{-\frac{\mu}{100}}\right) = \hat{p}$

$$\hat{\mu} = -100 \log(1 - \Sigma Y_i/_{100})$$

3.4 — Method of moments estimator

- Simpler alternative to maximum likelihood

→ Requires only the theoretical form of the moments.

→ Is based on imposing the matching of theoretical and sample moments

$$E_\theta(X^j) \qquad\qquad \overline{X^j} = \frac{1}{n}\sum_{i=1}^{n} X_i^j$$

Method of moments estimator

is the value $\hat{\theta}$ where

$$E_{\hat{\theta}}(X^j) = \overline{X^j}$$

take j as small as possible

The MME for $g(\theta)$ is taken to be $g(\hat{\theta})$

Example 3.27 – Exponential

$$X_i \sim \lambda e^{-\lambda x_i} \qquad \lambda > 0$$

$$E(X_i) = \frac{1}{\lambda} \Rightarrow \frac{1}{\lambda} = \bar{X} \Rightarrow \hat{\lambda}_{MM} = \frac{1}{\bar{X}}.$$

$\hat{\lambda}_{MM} = \hat{\lambda}_{ML}$.

Example 3.28 – Uniform

$$X_i \sim \frac{1}{\theta}$$

$$E(X_i) = \theta/2$$

$$\bar{X} = \theta/2 \Rightarrow \hat{\theta}_{MM} = 2\bar{X}.$$

$\hat{\theta}_{MM} \neq \hat{\theta}_{ML}$.

Example 3.29 – Normal

$X_i \sim N(\mu, \sigma^2)$

$E(X_i) = \mu \qquad E(X_i^2) = \sigma^2 + \mu^2$

$$\begin{cases} \mu = \bar{X} \\ \sigma^2 + \mu^2 = \overline{X^2} \end{cases}$$

$$\hat{\mu}_{MM} = \bar{X} \qquad \hat{\sigma}^2_{MM} = \overline{X^2} - \bar{X}^2$$

$$= \frac{1}{n}\sum_i (X_i - \bar{X})^2.$$

$\hat{\sigma}^2_{MM} = \hat{\sigma}^2_{ML}$. //

Example 3.30 – Gamma distribution

$X_i \sim \text{Gamma}(\alpha, \lambda)$

$E(X_i) = \frac{\alpha}{\lambda} \qquad E(X_i^2) = \frac{\alpha}{\lambda^2} + \frac{\alpha^2}{\lambda^2}$

$$= \frac{\alpha}{\lambda^2}(1 + \alpha)$$

$$\begin{cases} \overline{X} = \dfrac{\alpha}{\lambda} \\ \overline{X^2} = \dfrac{\alpha}{\lambda^2}(1+\alpha) \end{cases} \rightarrow \begin{cases} \hat{\alpha}_{MM} = \dfrac{(\overline{X})^2}{\overline{X^2} - (\overline{X})^2} \\ \hat{\lambda}_{MM} = \dfrac{\overline{X}}{\overline{X^2} - (\overline{X})^2} \end{cases}$$

Are different from MLE and in
<u>closed form</u>

3.5 Bayes estimators

A different approach to inference

- θ is fixed and unknown

- θ is a random variable described by a density $\pi(\theta)$ representing the information prior to observing the data.

1) The Bayesian approach starts by defining the prior distribution $\pi(\theta)$ on Θ in addition to the statistical model.

2) the Bayesian approach uses data and Bayes formula to transform the prior into a posterior $\pi(\theta \mid \underline{x})$

Bayes risk of $T(x)$ given $\pi(\theta)$

is the weighted average of the MSE

$$R(\pi, T) = \int E_\theta (T - g(\theta))^2 \pi(\theta)\, d\theta$$

with weight $\pi(\theta)$.

Bayes estimator

is the estimator T that minimizes $R(\pi, T)$ over all estimators.

Bayes formula

Discrete form.

$$X \quad\quad \Theta$$
$$\downarrow \quad\quad \downarrow$$
$$x \quad\quad \theta$$

$P_\theta(x) = p(x \mid \Theta = \theta)$ likelihood

$\pi(\theta)$ Prior

$$\boxed{\begin{aligned} &\textbf{Posterior} \\ &P(\Theta = \theta \mid X = x) = \\ &= \frac{P(X = x \mid \Theta = \theta)\, P(\Theta = \theta)}{\sum_\theta P(X = x \mid \Theta = \theta)\, P(\Theta = \theta)} \end{aligned}}$$

Continuous form

Posterior

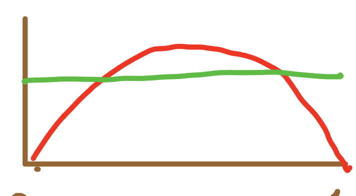
$$p_{\Theta|X}(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int p(x|\vartheta)\pi(\vartheta)}$$

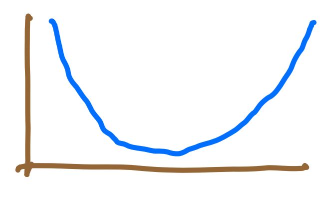
- The denominator is a normalizing constant.

Example 3.38 – Binomial

$X \sim Bin(n, p)$

Prior on $p \in (0,1)$





A useful family of priors is the Beta

$\boxed{\text{Beta density}}$

See the notebook