BIASED VS UNBIASED ESTIMATORS INTUITION.

Def $\underbrace{\hat{\theta} = u(X_1, \dots, X_m)}_{\text{function of the random sample}}$ is an unbiased estimator for θ if $E[u(X_1, X_2, \dots, X_m)] = \theta$

unknown

while is unknown, we aim to use $u(X_1, X_2, \dots, X_m)$ to estimate:

$\rightarrow$ collect $\underbrace{x_1, x_2, \dots, x_m}_{\text{actual values}}$ $\longrightarrow$ give an <u>estimate</u> $\theta = u(x_1, x_2, \dots, x_m) \leftarrow$ numerical value

Let us take $\underline{N}$ samples (N large or very large) all of which of size m (reasonably large!)

$\hookrightarrow m = 100$ or $m = 1000$

1st sample $x_1^1, x_2^1, \dots, x_m^1 \rightsquigarrow \theta_1$

2nd sample $x_1^2, x_2^2, \dots, x_m^2 \rightsquigarrow \theta_2$

3rd sample $x_1^3, x_2^3, \dots, x_m^3 \rightsquigarrow \theta_3$

$\vdots \qquad\qquad \vdots$

Nth sample $x_1^N, x_2^N, \dots, x_m^N \rightsquigarrow \theta_N$

$$\frac{1}{N} \sum_{i=1}^{N} \theta_i \xrightarrow[N \to \infty]{} \boxed{E[\hat{\theta}] = \theta} \text{ if } \hat{\theta} \text{ is unbiased}$$

strong law of large numbers

Math 4501 - Probability and Statistics II

6.4 - Maximum likelihood and method of moments estimation

techniques for point estimates!

TODAY!

Method of moments estimator

: NOTATION is $\tilde{\theta}$ is the method of moments estimator for θ

we $\hat{\theta}$ for MLE only

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with pmf or pdf $f(x; \theta_1, \ldots, \theta_k)$ depending on k unknown parameters $(\theta_1, \ldots, \theta_k) \in \Omega$, $k \geq 1$.

Recall that for $k = 1, 2, 3, \ldots$: (from 3501)

- the expectation $E(X^k)$ is often called the kth moment of the distribution.

raw k^{th} moment or k^{th} moment about 0

- the sum $M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$ is the kth moment of the sample

$$E[X^k] = \begin{cases} \int_{-\infty}^{\infty} x^k f(x)\,dx, & X \text{ cont.} \\ \sum_x x^k f(x), & X \text{ discrete} \end{cases}$$

pick a large enough number of

$E[X^k] = M_k, \quad k = 1, 2, \ldots, ??$ equations to determine all parameters

Method of moments:

Equate $E(X^k)$ to M_k beginning with $k = 1$ and continuing until there are enough equations to provide unique solutions for $\theta_1, \theta_2, \ldots, \theta_k$.

- the goal is equate sufficiently many distribution moments and sample moments to obtain a system of k independent equations in k unknowns.

$\theta_1, \theta_2, \ldots, \theta_k$

solution being the method of moments estimator

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the distribution with pdf

$$f(x;\theta) = \theta x^{\theta-1}\ , \quad 0 < x < 1\ ,$$

where $\theta \in \Omega = (0, \infty)$.

Determine the method of moments estimator of θ.

Single unknown parameter θ : take a single equation

$$\underbrace{E[X]}_{\text{function of } \theta} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{M_1} \implies \underbrace{\frac{\theta}{\theta+1}}_{E[X]} = \underbrace{\overline{X}}_{M_1}$$

solve for θ!

$$E[X] = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot \theta x^{\theta-1}\, dx = \theta \int_0^1 x^{\theta}\, dx = \theta \left[\frac{x^{\theta+1}}{\theta+1}\right]_{x=0}^{x=1} = \frac{\theta}{\theta+1}$$

$$\frac{\theta}{\theta+1} = \bar{X} \iff \theta = \bar{X}(\theta+1) \iff \theta - \bar{X}\theta = \bar{X}$$

$$\iff \theta(1-\bar{X}) = \bar{X} \iff \theta = \frac{\bar{X}}{1-\bar{X}}$$

CONCLUSION:

Method of moments estimator for θ is

$$\boxed{\tilde{\theta} = \frac{\bar{X}}{1-\bar{X}}} = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i}{1-\frac{1}{n}\sum_{i=1}^{n} X_i} = \frac{\sum_{i=1}^{n} X_i}{n-\sum_{i=1}^{n} X_i}$$

Example

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\overset{\mu}{\theta_1}, \overset{\sigma^2}{\theta_2})$ distribution, where

$$\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 < \infty,\ 0 < \theta_2 < \infty\} \ .$$

Determine the method of moments estimators for θ_1 and θ_2.

To find the method of moments estimators for θ_1 and θ_2, we use two equations:

$$\begin{cases} E[X] = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{M_1} \\ E[X^2] = \underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{M_2} \end{cases}$$

$E[X]$ and $E[X^2]$ may depend on θ_1 and θ_2

We need to find $E[X]$ and $E[X^2]$

$E[X] = \theta_1$

$E[X^2] = \theta_2 + \theta_1^2$

we $\mathrm{Var}(X) = E[X^2] - \underbrace{\left(E[X]\right)^2}$

$\theta_2 = E[X^2] - \theta_1^2$

$E[X^2] = \theta_2 + \theta_1^2$

$$\underbrace{\theta_1}_{E[X]} = \overbrace{\frac{1}{m}\sum_{i=1}^{m} X_i}^{M_1}$$

$$\underbrace{\theta_2 + \theta_1^2}_{E[X^2]} = \underbrace{\frac{1}{m}\sum_{i=1}^{m} X_i^2}_{M_2}$$

$\longrightarrow$ solve for θ_1 and θ_2

$$\left\{ \begin{aligned} \tilde{\theta}_1 &= \overline{X} \\ \tilde{\theta}_2 &= \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \overbrace{\overline{X}^2}^{\theta_1^2} \\ &= \frac{1}{m}\sum_{i=1}^{m} (X_i - \overline{X})^2 \end{aligned} \right\}$$ Method of moments estimators for θ_1, θ_2

same as obtained using MLE

NOTE : $\tilde{\theta}_2 = \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \overline{X}^2 = \frac{1}{m}\sum_{i=1}^{m} (X_i - \overline{X})^2$ } discrete analogue to $Var(X) = E[X^2] - (E[X])^2$

$$\underbrace{\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i}_{m\overline{X} = \sum_{i=1}^{m} X_i}$$

Observe that $\frac{1}{m}\sum_{i=1}^{m}(X_i - \overline{X})^2 = \frac{1}{m}\sum_{i=1}^{m}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right)$

$$= \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \frac{2}{m}\sum_{i=1}^{m} X_i \boxed{\overline{X}} + \frac{1}{m}\underbrace{\sum_{i=1}^{m}\overline{X}^2}_{m\overline{X}^2} = \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \frac{2\overline{X}}{m}\underbrace{\sum_{i=1}^{m} X_i}_{m\cdot\overline{X}} + \frac{1}{m} m\overline{X}^2$$

does not depend on i

$$= \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \frac{2\overline{X}}{m} m\overline{X} + \overline{X}^2 = \frac{1}{m}\sum_{i=1}^{m} X_i - 2\overline{X}^2 + \overline{X}^2 = \frac{1}{m}\sum_{i=1}^{m} X_i^2 - \overline{X}^2$$

Math 4501 - Probability and Statistics II

6.5 - Regression

we will employ MLE technique to determine the regression parameters

Regression

Broader problem: predict the value of a random variable Y corresponding to an observed value x of another random variable X

Simpler problem: estimate the conditional mean of Y given that $X = x$:

$$E[Y|x] = \mu(x) \; ,$$

find what μ should be!!!

where $\mu(x)$ is assumed to be of a given form:

$$\mu(x) = \alpha + \beta x \quad \text{or} \quad \mu(x) = \alpha + \beta x + \gamma x^2 \quad \text{or} \quad \mu(x) = \alpha e^{\beta x}$$

Strategy: to estimate the model parameters (such as α, β and γ):

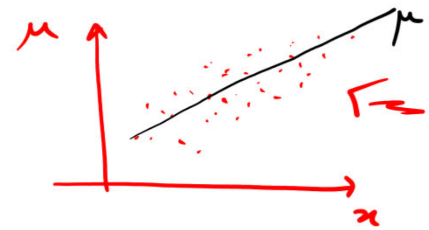
- observe the random variable Y for each of n possibly different values of x:

$$x_1, x_2, \ldots, x_n \; .$$

- once n independent experiments have been performed, use the n pairs

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

to estimate $E[Y|x]$.



Terminology:

- Problems like this are often classified under regression because

$$E[Y|x] = \mu(x)$$

  is frequently called a *regression curve*.

- A model for the conditional mean of the form

$$\mu(x) = \alpha + \beta x + \gamma x^2$$

  ← linear on parameters

  is called a *linear model* because it is linear in the parameters α, β, and γ.
  - Note that a linear model may be nonlinear in x!

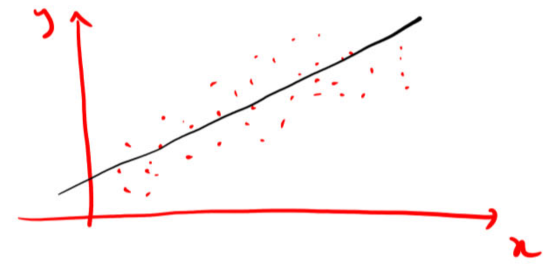
- A model for the conditional mean of the form

$$\mu(x) = \alpha e^{\beta x}$$

  nonlinear dependance on β

  is not a linear model, because it is not linear in the parameters α and β.

Simplest regression problem



Given the data points

$$\to \quad (x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$$

estimate the parameters α and β of the linear model

$$\boxed{E[Y|x] = \underbrace{\alpha_1 + \beta x}_{\mu(x)} \, ,}$$

that is, fit a straight line to the given set of data.

$E[Y|x] = \mu(x) = \alpha_1 + \beta x$

Assumptions:

- for each particular value of x, the value of Y differs from its mean by a random amount ε.
- the distribution of ε is $N(0, \sigma^2)$.

$Y = \alpha_1 + \beta x + \varepsilon$

σ^2 is another parameter to estimate

Consequence: For the linear model described above, we have

$$\boxed{Y_i = \overbrace{\alpha_1 + \beta x_i}^{\mu(x_i)} + \varepsilon_i \, ,}$$

$\varepsilon_i \sim N(0, \sigma^2)$

where ε_i, $i = 1, 2, \ldots, n$, are independent $N(0, \sigma^2)$ random variables.

GOAL: Estimate $\alpha_1, \beta, \sigma^2$

$\Rightarrow \quad Y_i \sim N(\alpha_1 + \beta x_i, \sigma^2)$

- For convenience, we set

$$\boxed{\alpha_1 = \alpha - \beta\bar{x}\ ,}$$

$y = mx + b \leftarrow$

$\updownarrow$

$y = y_0 + m(x - x_0)$

where $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of the observations $x_1, \ldots, x_n$.

- For each $i = 1, 2, \ldots, n$, we have that

$$Y_i = \alpha_1 + \beta x + \varepsilon_i \xrightarrow{\alpha_1 = \alpha - \beta\bar{x}} \boxed{Y_i = \underbrace{\alpha + \beta\,(x_i - \bar{x})} + \varepsilon_i}$$

is equal to a nonrandom quantity $\alpha + \beta\,(x_i - \bar{x})$ plus a mean-zero normal random variable ε_i.

- The random variables $Y_1, Y_2, \ldots, Y_n$ are mutually independent normal variables with respective means

$$\alpha + \beta\,(x_i - \bar{x})\ , \quad i = 1, 2, \ldots, n$$

and unknown variance σ^2.

estimate α, β, σ^2

$$\Bigg\}\ Y_i \sim N(\alpha + \beta(x_i - \bar{x}), \sigma^2)$$

Proposition

Under the conditions described above, the maximum likelihood estimators of α, β and σ^2 are given by:

$$\boxed{\widehat{\alpha} = \bar{Y}}$$

$$\widehat{\beta} = \frac{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \widehat{\alpha} - \widehat{\beta}\,(x_i - \bar{x})\right]^2$$

Proof (yet another example of MLE):

θ stands for σ^2

Recall that $Y_i \sim N(\alpha + \beta(x_i - \bar{x}), \theta)$ independent!

Define the likelihood function:

pdf of $X \sim N(\mu, \sigma^2)$ is
$$\boxed{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}$$

$$L(\alpha, \beta, \theta) = \prod_{i=1}^{m} f(y_i ; \alpha, \beta, \sigma^2) =$$

$$= \prod_{i=1}^{m} \frac{1}{\sqrt{2\pi\,\underset{\theta \text{ is } \sigma^2}{\theta}}} \exp\left(- \frac{\Big(y_i - \overbrace{(\alpha + \beta(x_i - \bar{x}))}^{\mu(x)}\Big)^2}{2\,\underset{\theta \text{ is } \sigma^2}{\theta}} \right)$$

$$= (2\pi\theta)^{-m/2} \exp\left(- \frac{1}{2\theta} \sum_{i=1}^{m} \big(y_i - (\alpha + \beta(x_i - \bar{x}))\big)^2 \right)$$

Apply natural log:

$$\ln\left(L(\alpha, \beta, \sigma^2)\right) = \overbrace{-\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\underbrace{\sum_{i=1}^{n}\left(y_i - \alpha - \beta(x_i - \bar{x})\right)^2}_{H(\alpha, \beta)}}$$

$\theta = \sigma^2 > 0$

To maximize $\ln\left(L(\alpha, \beta, \sigma^2)\right)$, we need to make $H(\alpha, \beta)$ as small as possible!

We start by minimizing $H(\alpha, \beta) = \sum_{i=1}^{n}\left(y_i - \alpha - \beta(x_i - \bar{x})\right)^2$

First order conditions are:

$$\begin{cases} \dfrac{\partial H}{\partial \alpha} = 0 \\[2ex] \dfrac{\partial H}{\partial \beta} = 0 \end{cases} \iff \begin{array}{l} -2\sum_{i=1}^{n}\left(y_i - \alpha - \beta(x_i - \bar{x})\right) = 0 \\[2ex] -2\sum_{i=1}^{n}(x_i - \bar{x})\left(y_i - \alpha - \beta(x_i - \bar{x})\right) = 0 \end{array}$$

Divide both sides of each equation by -2:

$$\begin{cases} \sum_{i=1}^{m}\left(y_i - \alpha - \beta(x_i - \bar{x})\right) = 0 \\ \sum_{i=1}^{m}(x_i - \bar{x})\left(y_i - \alpha - \beta(x_i - \bar{x})\right) = 0 \end{cases}$$

1st equation $\sum_{i=1}^{m} y_i - \sum_{i=1}^{m}\alpha - \sum_{i=1}^{m}\beta(x_i - \bar{x}) = 0$

$$\Rightarrow \sum_{i=1}^{m} y_i - m\alpha - \beta \underbrace{\sum_{i=1}^{m}(x_i - \bar{x})}_{=0} = 0$$

$$\sum_{i=1}^{m} x_i - \sum_{i=1}^{m}\bar{x} = m\bar{x} - m\bar{x} = 0$$

$$\Rightarrow \sum_{i=1}^{m} y_i - m\alpha = 0 \quad \Rightarrow \quad \alpha = \frac{1}{m}\sum_{i=1}^{m} y_i = \overline{Y}$$

2^{nd} equation: (replacing α by $\overline{y}$)

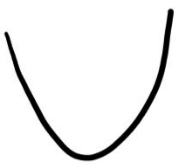
$$\sum_{i=1}^{n} (x_i - \overline{x}) \left((y_i - \overline{y}) - \beta (x_i - \overline{x}) \right) = 0$$

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) - \beta (x_i - \overline{x})^2 = 0$$

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) - \beta \sum_{i=1}^{n} (x_i - \overline{x})^2 = 0$$

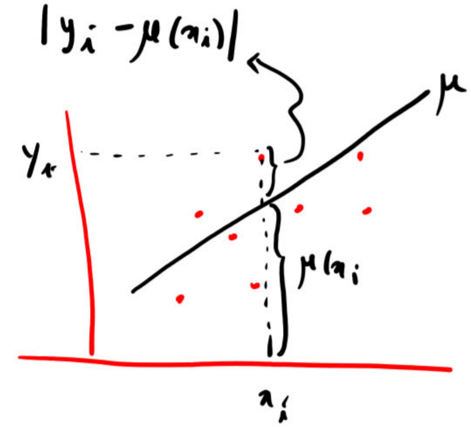
$$\beta = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

Relation with method of least squares



- The parameters α and β minimize the quantity

$$H(\alpha,\beta)=\sum_{i=1}^{n}\left[y_i \overbrace{-\alpha-\beta\left(x_i-\bar{x}\right)}^{\mu(x_i)}\right]^2.$$



- Since

$$\left|y_i-\alpha-\beta\left(x_i-\bar{x}\right)\right|=\left|y_i-\mu\left(x_i\right)\right|$$

  is the vertical distance from the point (x_i, y_i) to the line $y=\mu(x)$, then $H(\alpha,\beta)$ represents the sum of the squares of those distances.

- Selecting α and β so that the sum of the squares is minimized means that we are fitting the straight line to the data by the *method of least squares*.

- Thus, the maximum likelihood estimates of α and β are also called *least squares estimates*.