Math 4501 - Probability and Statistics II

6.4 - Maximum likelihood and method of moments estimation

techniques for point estimates!

POINT ESTIMATOR : function $\hat{\theta} = \theta(x_1, \ldots, x_n)$

discrete case : pmf $f(x;\theta) = P(X = x)$

Maximum likelihood estimator

→ main technique we will use to find point estimators!

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution depending on one or more unknown parameters $\theta_1, \ldots, \theta_m$.

Denote the distribution pmf or pdf by $\underbrace{f(x; \theta_1, \ldots, \theta_m)}_{\text{pdf/pmf of each of } x_1, x_2, \ldots, x_n}$, with $\overbrace{(\theta_1, \ldots, \theta_m)}^{\text{unknown parameters}} \in \Omega$.

parameter space

The function

$$\begin{aligned} L(\theta_1, \ldots, \theta_m) &= \prod_{i=1}^{n} f(x_i; \theta_1, \ldots, \theta_m) \\ &= f(x_1, \theta_1, \ldots, \theta_m) \cdots f(x_n, \theta_1, \ldots, \theta_m), \quad (\theta_1, \ldots, \theta_m) \in \Omega, \end{aligned}$$

when regarded as a function of $\theta_1, \ldots, \theta_m$, is called the *likelihood function*.

$$\to L(\theta_1, \theta_2, \ldots, \theta_m) = P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) =$$

independency

$$= \underbrace{P(X_1 = x_1)}_{f(x_1, \theta_1, \ldots, \theta_m)} \cdot P(X_2 = x_2) \cdots \underbrace{P(X_m = x_m)}_{f(x_m, \theta_1, \ldots, \theta_m)} = \prod_{i=1}^{m} f(x_i, \theta_1, \ldots, \theta_m)$$

The functions

$$\left\{\begin{array}{c} \widehat{\theta_1} = u_1\left(X_1, \ldots, X_n\right) \\ \vdots \\ \widehat{\theta_m} = u_m\left(X_1, \ldots, X_n\right) \end{array}\right\}$$

that maximize $L(\theta_1, \ldots, \theta_m)$ are the *maximum likelihood estimators* of $\theta_1, \theta_2, \ldots, \theta_m$, respectively.

The corresponding observed values of these statistics

$$u_1\left(x_1, \ldots, x_n\right), u_2\left(x_1, \ldots, x_n\right), \ldots, u_m\left(x_1, \ldots, x_n\right)$$

are called *maximum likelihood estimates*.

Notes:
1) In many practical cases, these estimators (and estimates) are unique.
2) For many applications, there is just one unknown parameter θ. In such cases, the likelihood function is given by

$$L(\theta) = \prod_{i=1}^{n} f\left(x_i; \theta\right) = f(x_1, \theta) \cdots f(x_n, \theta), \quad \theta \in \Omega . \leftarrow$$

like a Calculus I optimization problem!

Remark

Since the natural logarithm function is a strictly increasing function, the maxima of $L(\theta)$ and of $\ln(L(\theta))$, whenever they exist, are attained at the same value of θ.

It is often easier to maximize $\ln(L(\theta))$ than it is to maximize $L(\theta)$, since

$$\ln(L(\theta)) = \ln\left(\prod_{i=1}^{n} f(x_i;\theta)\right) = \sum_{i=1}^{n} \ln(f(x_i;\theta)) .$$

A similar comment applies to the case of more than one unknown parameter.

$\ln(x.y) = \ln x + \ln y$

Example

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

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

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1-p & \text{if } x = 0 \end{cases}$$

unknown parameter

where $p \in \Omega = (0,1)$.

Determine the maximum likelihood estimator of p.

Define the likelihood function

$$L(p) = \prod_{i=1}^{n} f(x_i, p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n - \sum_{i=1}^{n} x_i}$$

$$p^{x_1} \cdot p^{x_2} \cdots p^{x_n}$$

$$(1-p)^{1-x_1} \cdot (1-p)^{1-x_2} \cdots (1-p)^{1-x_2}$$

Take natural log of $L(p)$ and simplify

$$\ln(L(p)) = \ln\Big(\underbrace{p^{\sum_{i=1}^{n} x_i} \cdot (1-p)^{n-\sum_{i=1}^{n} x_i}}_{L(p)}\Big) = \ln\left(p^{\sum_{i=1}^{n} x_i}\right) + \ln\left((1-p)^{n-\sum_{i=1}^{n} x_i}\right)$$

Recall another property of $\ln x$: $\ln(x^y) = y \ln x$

to get

$$\ln(L(p)) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p)$$

Take the derivative w.r.t p and set to zero (first order condition) to get

$$\frac{d}{dp} \ln(L(p)) = \left(\sum_{i=1}^{n} x_i\right) \frac{1}{p} + \left(n - \sum_{i=1}^{n} x_i\right) \cdot \frac{1}{1-p} \cdot (-1)$$

$$= \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}, \quad p \in (0,1)$$

$$\frac{d}{dp}\ln(L(p)) = 0 \iff \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p} = 0$$

and solve for p:

$$\frac{\sum_{i=1}^{n} x_i}{p} = \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\Rightarrow (1-p)\sum_{i=1}^{n} x_i = p\left(n - \sum_{i=1}^{n} x_i\right) \Rightarrow \sum_{i=1}^{n} x_i = np \Rightarrow$$

$$\Rightarrow \underbrace{p = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}}_{\text{our candidate to the maximizer of } L(p)}$$

We still need to check that $p = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is actually a maximizer.

Use the 2nd derivative test:

$$\frac{d^2}{dp^2}\ln(L(p)) = \frac{d}{dp}\left[\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}\right] =$$

$$= -\frac{\sum_{i=1}^{n} x_i}{p^2} - \frac{n - \sum_{i=1}^{n} x_i}{(1-p)^2} \leq 0 \text{ for all } \underline{p \in (0,1)}$$

$\sum_{i=1}^{n} x_i \in \{0,1,2,\dots,n\}$

$p \in (0,1)$

$$\Rightarrow \hat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \text{ is a Maximizer of } L(p)$$

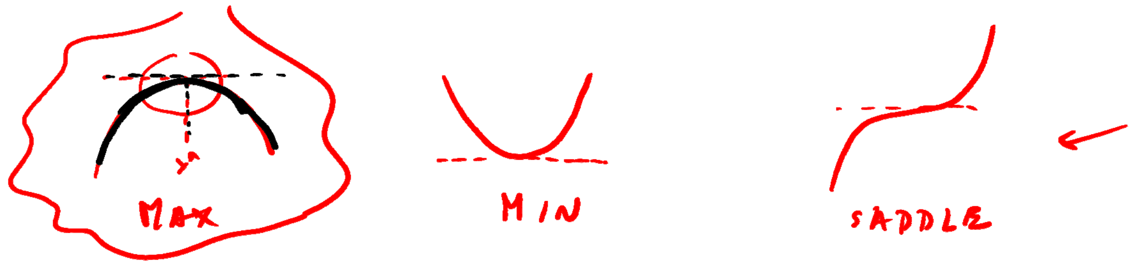
$$\Rightarrow \hat{p} = \overline{X} \text{ is the } \underbrace{\text{MLE}}_{\text{maximum likelihood estimator!}}$$

Review: Optimization for functions of a single variable

Goal: find a maximizer for $f(x)$

(1) Find critical pts x^* of $f(x)$ using the FOC: $f'(x)=0$

critical pts:



2ND DER. TEST $\quad f''(x^*)<0 \quad\quad f''(x^*)>0$

Example

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

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

unknown parameter

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

Determine the maximum likelihood estimator of θ.

Define the likelihood function:

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-x_i/\theta} = \left(\frac{1}{\theta}\right)^n \cdot e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}$$

$$\underbrace{\frac{1}{\theta} \cdot \frac{1}{\theta} \cdots \frac{1}{\theta}}_{n \text{ times}}$$

$$e^{-x_1/\theta} \cdot e^{-x_2/\theta} \cdots e^{-x_n/\theta}$$

Apply natural log:

$$\ln(L(\theta)) = \ln\left(\frac{1}{\theta^n} e^{-\left(\sum_{i=1}^{n} x_i\right)/\theta}\right) = \ln\left(\frac{1}{\theta^n}\right) + \ln\left(e^{-\sum_{i=1}^{n} x_i/\theta}\right)$$

$\ln(xy) = \ln x + \ln y$

$$= \ln\left(\theta^{-n}\right) + \ln\left(e^{-\sum_{i=1}^{n} x_i/\theta}\right)$$

$\ln x^y = y \ln x$

$\ln e^x = x$

$$= -n \ln \theta - \frac{\sum_{i=1}^{n} x_i}{\theta}$$

First order condition:

$$\frac{d}{d\theta} \ln(L(\theta)) = \frac{d}{d\theta}\left[-n\ln\theta - \frac{\sum_{i=1}^{n} x_i}{\theta}\right] = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

$$\frac{d}{d\theta} \ln(L(\theta)) = 0 \implies -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} = 0 \implies \frac{n}{\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

cross multiply

$$\implies \underbrace{\theta = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{X}}_{\text{critical pt for } L(\theta)}$$

Let us check that the critical pt is indeed a maximum:

Study the 2^{nd} derivative:

$$\frac{d^2}{d\theta^2} \ln(L(\theta)) = \frac{d}{d\theta}\left[-\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \right] = \frac{n}{\theta^2} - 2\,\frac{\sum_{i=1}^{n} x_i}{\theta^3}$$

We need to check that $\frac{d^2}{d\theta^2} \ln(L(\theta))$ is negative when $\underbrace{\theta = \frac{1}{n}\sum_{i=1}^{n} x_i}_{\text{critical pt determined earlier}}$

$$\frac{d^2}{d\theta^2} \ln(L(\theta)) \Big|_{\theta = \frac{1}{n}\sum_{i=1}^{n} x_i} = \frac{n}{\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^2} - 2\,\frac{\sum_{i=1}^{n} x_i}{\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)^3}$$

$$= \frac{n^3}{\left(\sum_{i=1}^{n} x_i\right)^2} - 2\,\frac{n^3 \sum_{i=1}^{n} x_i}{\left(\sum_{i=1}^{n} x_i\right)^3} =$$

$$= \frac{n^3}{\left(\sum_{i=1}^{n} x_i\right)^2} - \frac{2\, n^3}{\left(\sum_{i=1}^{n} x_i\right)^2} = -\frac{n^3}{\left(\sum_{i=1}^{n} x_i\right)^2} < 0$$

$\Rightarrow \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ maximizes $L(\theta)$

$\Rightarrow \quad \hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is the MLE of θ

Example

stands for μ stands for σ^2

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $N(\theta_1, \theta_2)$ distribution, where

$$\underbrace{\Omega}_{\text{parameter space}} = \{(\theta_1, \theta_2) : -\infty < \underline{\underline{\theta_1}} < \infty,\ \underline{\underline{0 < \theta_2 < \infty}}\}\ .$$

Determine the maximum likelihood estimators for θ_1 and θ_2.

Recall that the pdf of $N(\theta_1, \theta_2)$ is

$$f(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi\,\theta_2}}\, e^{-\frac{(x-\theta_1)^2}{2\theta_2}}$$

look up formula sheet and replace μ by θ_1 and σ^2 by θ_2

Define the likelihood function:

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2) = \prod_{i=1}^{n} \frac{1}{(2\pi\,\theta_2)^{1/2}}\, e^{-\frac{(x_i-\theta_1)^2}{2\theta_2}}$$

$$= \left[\frac{1}{(2\pi\theta_2)^{1/2}}\right]^n e^{-\sum_{i=1}^{n}\frac{(x_i-\theta_1)^2}{2\theta_2}}$$

$$= (2\pi\theta_2)^{-\frac{n}{2}} e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2}$$

Apply natural log:

$$\ln(L(\theta_1,\theta_2)) = \ln\left[(2\pi\theta_2)^{-n/2} e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2}\right]$$

$$= \ln\left[(2\pi\theta_2)^{-n/2}\right] + \ln\left[e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2}\right]$$

$$= -\frac{n}{2}\ln(2\pi\theta_2) - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

$$\Rightarrow \quad \ln\left(L(\theta_1,\theta_2)\right) = \underbrace{-\frac{n}{2}\ln 2\pi}\ \underline{-\frac{n}{2}\ln\theta_2} - \frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2$$

First order conditions (one for each variable!)

$$\begin{cases} \dfrac{\partial}{\partial\theta_1}\ln\left(L(\theta_1,\theta_2)\right) = 0 \\[2ex] \dfrac{\partial}{\partial\theta_2}\ln\left(L(\theta_1,\theta_2)\right) = 0 \end{cases} \Rightarrow \begin{cases} -\dfrac{1}{2\theta_2}\displaystyle\sum_{i=1}^{m}(-1)\cdot 2(x_i-\theta_1) = 0 \\[2ex] -\dfrac{m}{2}\dfrac{1}{\theta_2} + \dfrac{1}{2\theta_2^2}\displaystyle\sum_{i=1}^{m}(x_i-\theta_1)^2 = 0 \end{cases}$$

$$\begin{cases} \dfrac{1}{\theta_2}\displaystyle\sum_{i=1}^{m}(x_i-\theta_1) = 0 \\[2ex] \dfrac{m}{\cancel{2}}\dfrac{1}{\theta_2} = \dfrac{1}{\cancel{2}\theta_2^2}\displaystyle\sum_{i=1}^{n}(x_i-\theta_1)^2 \end{cases} \Rightarrow \begin{cases} \displaystyle\sum_{i=1}^{m}(x_i-\theta_1) = 0 \\[2ex] \dfrac{\theta_2^2}{\theta_2} = \dfrac{1}{m}\displaystyle\sum_{i=1}^{n}(x_i-\theta_1)^2 \end{cases}$$

$$\begin{cases} \sum_{i=1}^{m} x_i - m\theta_1 = 0 \\ \theta_2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \theta_1)^2 \end{cases} \implies \begin{cases} m\theta_1 = \sum_{i=1}^{m} x_i \\ \text{———} \end{cases}$$

$$\implies \begin{cases} \theta_1 = \frac{1}{m} \sum_{i=1}^{m} x_i \\ \theta_2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \theta_1)^2 \end{cases} \implies$$

CRITICAL PTS ARE

$$\hat{\theta}_1 = \overline{X}$$

$$\hat{\theta}_2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \overline{x})^2$$

Using the 2^{nd} derivative test for functions of two variables (Calculus III) we could check that $\underbrace{\hat{\theta}_1 \text{ and } \hat{\theta}_2 \text{ are indeed maximizers of } L(\theta_1, \theta_2)}_{\text{check the notes posted on Blackboard}}$

CONCLUSION: The MLE for the $N(\mu, \sigma^2)$ distribution:

$$\hat{\mu} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

when we've defined S^2, it was

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

why $n-1$ and not n

After the midterm!!