How to find estimators?

So far we have considered

$\bar{X} = 1/n \sum_{i=1}^{n} X_i$ — sample (estimating a mean)

$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$ — sample variance (estimating a variance).

↳ both are unbiased.

We don't have a general procedure for producing estimators.

Method of moments (very old).

Suppose $\theta_1, \theta_2, \ldots, \theta_l$ are l unknown parameters for some distribution.

We can instead try to estimate the moments.

The k^{th} moment of X is

$$\mu'_k = \mathbb{E}|X|^{k}$$

k^{th} power.

if negative

The estimator for μ'_k is the sample moment.

$$m'_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

just like estimating the mean.

Given a random sample $X_1, X_2, \dots, X_n$

<u>Goal</u>: Find $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_\ell$ such that

$$\mu'_k = m'_k \quad \forall\, k = 1, \dots, \ell.$$

function of $(\theta_1, \theta_2, \dots, \theta_\ell)$

System of equations for

Example (HW7)

Distribution is Uniform$(0, \theta)$

$\uparrow$ upper bound is unknown.

$X_1, X_2, \ldots, X_n$

one parameter, one equation.

$$\mu_1' = E(X) = \theta/2$$

$$m_1' = 1/n \sum_{i=1}^{n} X_i = \bar{X}$$

Find $\hat{\theta}$ s.t.

$$\mu_1' = m_1' \iff \frac{\hat{\theta}}{2} = \bar{X}.$$

$$\Downarrow$$

$$\hat{\theta} = 2\bar{X}.$$

$\swarrow$ Not $\hat{\theta} = \max\{X_1, X_2, \ldots, X_n\}$

$\hookrightarrow$ biased.

Easy to see unbiased.

$$E(\hat{\theta}) = 2E(\bar{X}) = 2\,\theta/2 = \theta.$$

The estimator is consistent by LLN.

$\bar{X} \to \theta/2$ in probability

by LLN

$\Rightarrow \hat{\theta} \to \theta$ in probability.

Remark: Method of moments always produces consistent estimators. May not be the best.

Ex Uniform$(0, \theta)$.

$$\hat{\theta}_1 = 2\bar{X} \quad , \quad \hat{\theta}_2 = \max\{X_1, X_2, \dots, X_n\}.$$

Which one is better?

$$MSE(\hat{\theta}_1) = Var(2\bar{X}).$$

$$= 4\,Var(\bar{X})$$

$$= 4\,\frac{Var(X_i)}{n}$$

Var of Uniform$(0, \theta)$

$$= \frac{4}{n}\left(\frac{1}{12}\theta^2\right) = \frac{1}{3n}\theta^2 \sim 1/n$$

$$MSE(\hat{\theta}_2) = \frac{2}{(n+1)(n+2)}\theta^2 \quad \text{By HW7.}$$

$$\sim 1/n^2$$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)} = \frac{\frac{2}{(n+1)(n+2)}\cancel{\theta^2}}{1/3n\,\cancel{\theta^2}}$$

$$= \frac{6n}{(n+1)(n+2)} < 1 \quad \text{if } n > 1$$

$$\| \quad 1 \text{ if } n = 1$$

Therefore $\hat{\theta}_2$ is better than $\hat{\theta}_1$.

Ex $N(\theta_1, \theta_2) \sim X$

$\uparrow$ mean, $\uparrow$ variance

$$\mu_1' = \mathbb{E}(X) = \theta_1, \quad \mu_2' = \mathbb{E}(X^2)$$

$$= Var(X) + (\mathbb{E}(X))^2$$

$$= \theta_2 + \theta_1^2$$

Given a sample $X_1, X_2, \dots, X_n$.

We want to solve

$$\mu_1' = m_1' \qquad\qquad \mu_2' = m_2'$$

$$\Downarrow$$

$$\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

unbiased

$$\theta_2 + \theta_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \theta_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

biased

sample population variance.

$$\hat{\theta}_1 = \bar{X}$$
$$\hat{\theta}_2 = \bar{S}^2$$

Both consistent.

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

WLN $\swarrow_{n\to\infty}$ $\qquad \downarrow n\to\infty$ in probability.

$$\mathbb{E}(X^2) \qquad (\mathbb{E}(X))^2$$

$$= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathrm{Var}(X).$$

Method of maximum likelihood.

Maximum likelihood estimators (MLEs)

Ex Container with 3 balls (red or white).

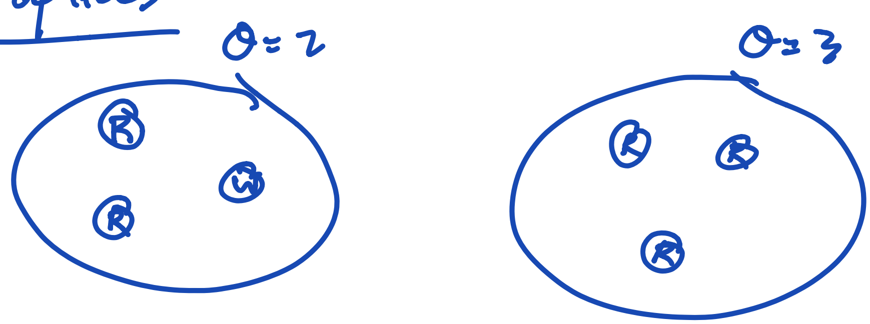
Estimate θ = # of red balls.

Suppose you sample 2 twice (w/o replacement)

Suppose I get two red balls.

What is the best estimate for θ?

Two options



$$P_{\theta=2}(2 \text{ red}) = \frac{1}{\binom{3}{2}} = \frac{1}{3} \quad \Big| \quad P_{\theta=3}(2 \text{ red}) = 1$$

The most likely scenario is $\theta = 3$.

Ex How about sampling with replacement.

Now we have independent samples

Sample $n = 4$ times.

$$X_i = \begin{cases} 1 & i^{th} \text{ sample is red} \\ 0 & i^{th} \text{ sample is white.} \end{cases}$$

$$X_i \sim \text{Bernoulli}(\theta/3) \rightarrow$$ fraction of red balls in container

Suppose my experiment gives

$X_1 = 1$, $X_2 = 0$, $X_3 = 1$, $X_4 = 1$
red white red red.

$(1, 0, 1, 1)$

What is the most probable # of red balls θ

Likelihood.

$P(x_1, x_2, x_3, x_4 | \theta)$

} independence

$= P(x_1|\theta)\, P(x_2|\theta)\, P(x_3|\theta)\, P(x_4|\theta)$

$P(1, 0, 1, 1) = \theta/3 \left(1 - \theta/3\right) \theta/3 \cdot \theta/3$

$= \left(\frac{\theta}{3}\right)^3 \left(1 - \theta/3\right)$

θ	0	1	2	3
$P(1,0,1,1\|\theta)$	0	0.0247	0.0988	0

↑ most likely.

$\theta = 2$ is the most likely

# of cards.

The method

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution w/ parameter θ.

The likelihood function

$$L(x_1, x_2, \ldots, x_n | \theta) = \begin{cases} p(x_1, x_2, \ldots, x_n | \theta) & \text{discrete case} \\ f(x_1, x_2, \ldots, x_n | \theta) & \text{continuous case.} \end{cases}$$

How likely we are to see $x_1, x_2, \ldots, x_n$ for the given θ.

The maximum likelihood estimator (MLE)

$\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$. is an θ

that maximized $L(x_1, x_2, \ldots, x_n | \theta)$.

Remark If $f(x|\theta)$ is the density for the distribution

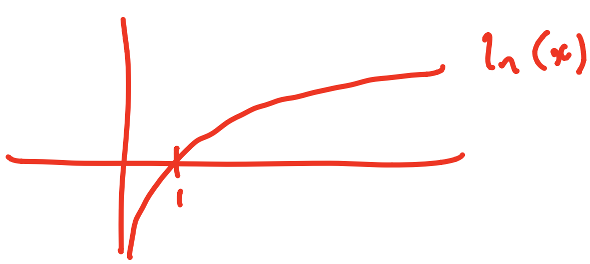
$$L(x_1, x_2, \dots, x_n | \theta) = f(x_1|\theta) f(x_2|\theta) \dots f(x_n|\theta)$$
$$= \prod_{i=1}^{n} f(x_i|\theta).$$
$$\shortparallel$$
$$\text{“}p(x_i|\theta).$$

Log-likelihood.

$$\ln\left(L(x_1, x_2, \dots, x_n|\theta)\right) = \sum_{i=1}^{n} \ln\left(f(x_i|\theta)\right).$$

↳ easier to maximize

Since $\ln(x)$ is one-to-one and increasing



Any maximizer of $\ln(L(\theta))$ also maximizes $L(\theta)$.

Ex Exponential(θ). $f(x|\theta) = \begin{cases} \theta e^{-\theta x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

$X_1, X_2, \ldots X_n$ random sample

Likelihood function.

$$L(x_1, x_2, \ldots, x_n | \theta) = \prod_i \theta e^{-\theta x_i}$$

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

sufficient statistic

take ln.

$$\ln(L(x_1, x_2, \ldots, x_n | \theta)) = \ln\left(\theta^n e^{-\theta \sum_{i=1}^n x_i}\right)$$

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

To maximize, take derivative in θ.

$$\frac{\partial}{\partial \theta}\left[n \ln \theta - \theta \sum_{i=1}^n x_i\right] = \frac{n}{\theta} - \sum_{i=1}^n x_i = 0$$

solve for θ

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

$$\hat{\theta} = \frac{1}{\bar{X}} \checkmark \quad \text{MLE}$$

↳ Consistent.

<u>Ex</u> $X \sim \text{Binomial}(m, \theta)$

$$p(x \mid \theta) = \binom{m}{x} \theta^{x} (1-\theta)^{m-x}$$ ↳ try to figure out how biased the coin is.

$X_1, X_2, \ldots, X_n$ – random sample

$$L(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} \binom{m}{x_i} \theta^{x_i} (1-\theta)^{m-x_i}$$

Want to maximize in θ.

$$= \left[\prod_{i=1}^{n} \binom{m}{x_i}\right] \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{mn - \sum_{i=1}^{n} x_i}$$

$$\shortparallel$$
$$h(x_1, x_2, \dots, x_n).$$

↑ doesn't depend on θ

Take a $\ln$.

$$\ln\left(L(x_1, x_2, \dots, x_n \mid \theta\right) = \ln(h) + \left(\sum_{i=1}^{n} x_i\right)\ln\theta + \left(mn - \sum_{i=1}^{n} x_i\right)\ln(1-\theta)$$

the derivative is

$$\frac{\partial}{\partial\theta}\ln\left(L(\theta)\right) = \left(\sum_{i=1}^{n} x_i\right)\frac{1}{\theta} + \left(mn - \sum_{i=1}^{n} x_i\right)\frac{-1}{1-\theta} = 0$$

Solve for θ.

$$\Rightarrow (1-\theta)\left(\sum_{i=1}^{n} x_i\right) - \left(mn - \sum_{i=1}^{n} x_i\right)\theta = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i = mn\theta$$

$$\Rightarrow \theta = \frac{1}{mn}\sum_{i=1}^{n} x_i$$

$$\boxed{\hat{\theta} = \frac{1}{mn}\sum_{i=1}^{n} X_i}$$

Generalizes to multiple parameters $\theta_1, \theta_2, \dots, \theta_l$

Maximize
$$L(x_1, x_2, \dots, x_n \mid \theta_1, \theta_2, \dots, \theta_l).$$

Ex $X_1, X_2, \dots, X_n \sim N(\theta_1, \theta_2)$

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

Likelihood function.

$$L(x_1, x_2, \dots, x_n \mid \theta_1, \theta_2) = \frac{1}{(2\pi)^{n/2}(\theta_2)^{n/2}} e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i-\theta_1)^2}$$

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

take derivatives

$$\frac{\partial}{\partial\theta_1}\ln(L(\theta_1, \theta_2)) = \frac{1}{\theta_2}\sum_{i=1}^{n}(x_i - \theta_1) = 0$$

$$\frac{\partial}{\partial \theta_2} \ln\left(L(\theta_1, \theta_2)\right) = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \sum_{i=1}^{n} (x_i - \theta_1)^2 = 0$$

$\downarrow$ system of equations for θ_1, θ_2:

$$\theta_1 = 1/n \sum_{i=1}^{n} x_i$$

θ_2:

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

$$\Rightarrow \frac{n}{\cancel{2}}\, \theta_2 = 1/\cancel{2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

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

Summarize

$$\hat{\theta}_1 = \overline{X}, \quad \hat{\theta}_2 = \overline{S}^2$$

What if you can't take the derivative?

Ex $X_1, X_2, \ldots, X_n \sim \text{Uniform}(0, \theta)$

$$L(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

$$= \begin{cases} \frac{1}{\theta^n} & x_1, x_2, \ldots, x_n \in [0, \theta] \\ 0 & \text{otherwise.} \end{cases}$$

To maximize $L(x_1, x_2, \ldots, x_n \mid \theta)$

we want to choose θ big enough so that,

$x_1, x_2, \ldots, x_n \in [0, \theta]$, but no bigger.

Choose $\theta = \max\{x_1, x_2, \ldots, x_n\}$

We see

$\hat{\theta} = \max\{X_1, \ldots, X_n\}$ is the MLE.

Invariance property.

If $u = u(X_1, X_2, \dots, X_n)$ is sufficient

$$L(x_1, x_2, \dots, x_n | \theta) = g(u, \theta) \cdot h(x_1, x_2, \dots, x_n).$$

$$\ln(L(\theta)) = \ln(g(u, \theta)) + \ln(h).$$

Maximizing $L(\theta)$ is the same as maximizing $g(u;\theta)$.

$\Rightarrow$ Any MLE is a function of a sufficient statistic.

$$\hat{\theta}_{MLE} = t(u(X_1, X_2, \dots, X_n))$$

The invariance property

If $t(\theta)$ is one-to-one. then

then $\widehat{t(\theta)} = t(\hat{\theta})$.

can view $\hat{\theta}$ as a sufficient statistic for $\tau(\theta)$.