Math 4501 - Probability and Statistics II

6.4 - Maximum likelihood and method of moments estimation

techniques for point estimates!

Invariance property of maximum likelihood estimators

Theorem

If $\widehat{\theta}$ is the maximum likelihood estimator of θ based on a random sample from the distribution with pdf or pmf $f(x;\theta)$, and g is a one-to-one function, then $g(\widehat{\theta})$ is the maximum likelihood estimator of $g(\theta)$.

In short, $\left.\begin{array}{l}\hat{\theta} \text{ MLE for } \theta \\ g \text{ is 1-to-1}\end{array}\right\} \Rightarrow \widehat{g(\theta)}$ MLE for $g(\theta)$ is $\widehat{g(\theta)} = g(\hat{\theta})$

CONSEQUENCE: if we know $\hat{\theta}$, then we know $\hat{\Psi}$ for any Ψ of the form $\Psi = g(\theta)$ with g a 1-to-1 function,

MOTIVATION / Proof sketch: $\overbrace{L(\theta)}^{\text{likelihood function}}$ is maximized by $\hat{\theta}$. Set $\Psi = g(\theta)$

Since g is 1-to-1, then g has an inverse h so that $\theta = h(\Psi)$

Then $L(\theta) = L(h(\Psi))$ is maximized at $\hat{\theta} = h(\hat{\Psi}) \Rightarrow \boxed{\hat{\Psi} = g(\hat{\theta})}$

Example

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

$$f(x;p) = p(1-p)^{x-1}, \quad x = 1, 2, \ldots ,$$

unknown parameter

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

Determine the maximum likelihood estimator for the mean of the population.

Recall that the mean of a geometric distribution is $\mu = E[X] = \frac{1}{p}$

Using the invariance principle for MLE, we know that the MLE of μ is

$$\hat{\mu} = \frac{1}{\hat{p}}$$

(because the function $g(x) = \frac{1}{x}$ is 1-to-1 in $(0, \infty)$)

All we have to do to find $\hat{\mu}$ is then $\underbrace{\text{to find } \hat{p}}_{\text{we already know how to do!}}$ and then take its reciprocal

Let us find the MLE for the parameter p (review of what we've done before!)

Define the likelihood function:

$$L(p) = \underbrace{\prod_{i=1}^{m} f(x_i; p)}\ = \underbrace{\prod_{i=1}^{m} p(1-p)^{x_i - 1}}\ = p^m (1-p)^{\sum_{i=1}^{m}(x_i - 1)}$$

$$p(1-p)^{x_1-1} \cdot p(1-p)^{x_2-1} \cdot p(1-p)^{x_3-1} \cdots\cdots p(1-p)^{x_m-1}$$

$$\underbrace{p \cdot p \cdot p \cdots p}_{m \text{ times}} \cdot (1-p)^{x_1-1} \cdot (1-p)^{x_2-1} \cdots\cdots (1-p)^{x_m-1}$$

$$p^m \quad (1-p)^{x_1-1+x_2-1+\cdots+x_m-1}$$

We obtain that

$$L(p) = p^m (1-p)^{\sum_{i=1}^{m} x_i - m}$$

Before differentiating $L(p)$, apply natural logarithm:

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

using
$$\begin{cases} \ln(x \cdot y) = \ln x + \ln y, & x, y > 0 \\ \ln x^p = p \ln x, & x > 0, p \in \mathbb{R} \end{cases}$$

We know find the derivative of $\ln(L(p))$:

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

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

The 1^{st} order condition is then:

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

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

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

We found out that $p = \dfrac{n}{\sum_{i=1}^{n} x_i}$ is a critical pt for $\ln(L(p))$ (and $L(p)$ as well !!!)

To check that the critical pt is indeed a maximum of $L(p)$, we check the 2^{nd} derivative

$$\frac{d^2}{dp^2} \ln(L(p)) = \frac{d}{dp}\Bigg[\underbrace{\frac{n}{p} - \frac{\sum_{i=1}^{n} x_i - n}{1-p}}_{\text{expression for } 1^{st} \text{ derivative found earlier}} \Bigg] =$$

$$= -\frac{n}{p^2} - \frac{\overbrace{\sum_{i=1}^{n} x_i - n}^{\geq 0 \text{ because } x_i \geq 1 \text{ for each } i}}{(1-p)^2} < 0 \quad \text{for all } p \in (0,1)$$

CONCLUSION : $\hat{p} = \dfrac{n}{\sum_{i=1}^{n} x_i}$ is the MLE of p

We conclude that (using the invariance principle):

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

Unbiased estimators

Recall: a statistic is a function of the random sample that does not depend on any unknown parameter

Definition

The statistic $u(X_1, X_2, \ldots, X_n)$ is called an *unbiased* estimator of θ if

$$E\left[u(X_1, X_2, \ldots, X_n)\right] = \theta \ .$$

Otherwise, $u(X_1, X_2, \ldots, X_n)$ is said to be *biased*.

Example (previous example continued)

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

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

where $\underbrace{\theta \in \Omega = (0, \infty)}^{\text{unknown!}}$.

Show that the maximum likelihood estimator of θ determined earlier is unbiased.

Recall that in a previous class we found that the MLE of θ is

$$\rightarrow \boxed{\hat{\theta} = \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i}$$

To show that $\hat{\theta}$ is unbiased, we need to check that $E[\hat{\theta}] = \theta$

$$E[\hat{\theta}] = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \underbrace{\overbrace{E[X_i]}^{\theta}}_{\text{expected value of exponential distr.}} = \frac{1}{n} \overbrace{\sum_{i=1}^{n} \theta}^{n \cdot \theta} = \frac{1}{n} \cdot n\theta = \theta$$

linearity

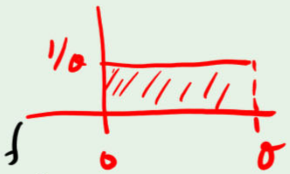
CONCLUSION: $\hat{\theta}$ is unbiased!

Example (Very important example – one reason why order statistics are important)

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

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

support of f depends on θ



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

Determine the maximum likelihood estimator of θ and show it is biased.

To find the MLE of θ, define the likelihood function:

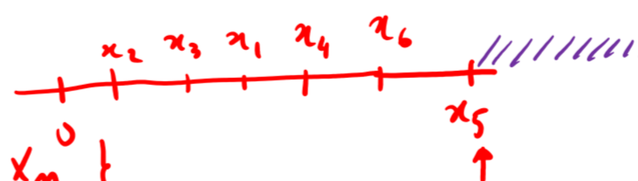
$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} = \frac{1}{\theta^n},$$

for θ such that $\theta \geq x_i > 0$, $i = 1, 2, \ldots, n$

To maximize $L(\theta)$, we want to take θ as small as possible.

However, we must have that $\theta \geq x_i$ for each $i = 1, 2, \ldots, n$.

The smallest value θ can take is $\theta = \max\{X_1, \ldots, X_n\} = Y_n$



Hence, the MLE for θ is $\hat{\theta} = Y_n = \max\{X_1, \ldots, X_n\}$

To show that $\hat{\theta} = Y_n$ is biased, we need to check that $E[\hat{\theta}] \neq \theta$

Let us find the pdf of $\hat{\theta} = Y_n$ so that we can compute $E[\hat{\theta}]$.

Let $y \geq 0$, and note that:

$$G_n(y) = P(Y_n \leq y) = P(\max\{X_1, \dots, X_n\} \leq y) =$$

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

same argument as done in Sec. 6.3

independence

$$= P(X_1 \leq y) \cdot P(X_2 \leq y) \cdots P(X_n \leq y)$$

$X_1, \dots, X_n$ identically distributed

$$= \Big(\underbrace{P(X_i \leq y)}_{\frac{y}{\theta}}\Big)^n = ?$$

(next slide)

$$F_{X_i}(y) = P(X_i \leq y) = \int_{-\infty}^{y} f(x,\theta)\,dx = \int_0^y \frac{1}{\theta}\,dx = \left[\frac{x}{\theta}\right]_{x=0}^{x=y}$$

$$= \frac{y}{\theta}, \qquad 0 \leq y \leq \theta$$

Thus, we find that

$$G_n(y) = \left[P(X_i \leq y)\right]^n = \left[\frac{y}{\theta}\right]^n = \begin{cases} 0 & y < 0 \\ \dfrac{y^n}{\theta^n}, & 0 \leq y \leq \theta \\ 1, & y > \theta \end{cases}$$

The pdf of Y_n is

$$g_n(y) = G_n'(y) = \frac{n\, y^{n-1}}{\theta^n}, \qquad \underline{0 < y < \theta}, \quad [\text{and zero otherwise}]$$

$$E[\hat{\theta}] = E[Y_m] = \int_{-\infty}^{\infty} y \cdot \underbrace{g_m(y)}_{\frac{m\,y^{m-1}}{\theta^m}} \, dy = \int_0^{\theta} y \cdot \frac{m\,y^{m-1}}{\theta^m} \, dy$$

$$= \frac{m}{\theta^m} \int_0^{\theta} y^m \, dy = \frac{m}{\theta^m} \cdot \left[\frac{y^{m+1}}{m+1} \right]_{y=0}^{y=\theta} =$$

$$= \frac{m}{\theta^m} \cdot \frac{\theta^{m+1}}{m+1} = \frac{m\,\theta}{m+1} \neq \theta$$

$\Longrightarrow$ $\hat{\theta}$ is biased!

Example (continued from an example of a previous class)

We have seen that when sampling from $N(\theta_1, \theta_2)$, one finds that the maximum likelihood estimators of θ_1 and θ_2 are

μ σ^2

$$\hat{\mu} = \widehat{\theta}_1 = \bar{X} \quad \text{and} \quad \underbrace{\widehat{\theta}_2}_{\overset{\shortparallel}{\sigma^2}} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \ .$$

} from the previous example!

Show that $\widehat{\theta}_1$ is an unbiased estimator of θ_1, but $\widehat{\theta}_2$ is not an unbiased estimator of θ_2.

↳ $\hat{\theta}_2$ is biased!

$\hat{\theta}_1$ is an unbiased estimator for $\theta_1 = \mu$.

$$E[\hat{\theta}_1] = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{} = \frac{1}{n}\sum_{i=1}^{n}\theta_1 =$$

def of $\bar{X}$

linearity

since $X_i \sim N(\theta_1, \theta_2)$ then $E[X_i] = \theta_1$

$$= \frac{1}{n} \cdot n \cdot \theta_1 = \theta_1 \Rightarrow \hat{\theta}_1 \text{ is unbiased!}$$

Let us now show that $\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is a biased estimator for θ_2

What can we say about the distribution of $\hat{\theta}_2$?

Recall: we have seen that if $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

with $X_1, X_2, \dots, X_n \sim N(\mu, \sigma^2)$, then

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

θ_2 is our σ^2

Gamma with $\alpha = \frac{n}{2}$ and $\theta = 2$

Note that $\frac{n\,\hat{\theta}_2}{\theta_2} = \sum_{i=1}^{n}\frac{(X_i - \bar{X})^2}{\theta_2} \sim \chi^2(n-1)$

Recall that if $X \sim \chi^2(n)$ then $E[X] = \frac{n}{2}\cdot 2 = n$

and so $E\left[\frac{n\,\hat{\theta}_2}{\theta_2}\right] = n-1 \Rightarrow \frac{n}{\theta_2}E[\hat{\theta}_2] = n-1$

$$E[\hat{\theta}_2] = \frac{n-1}{n} \cdot \theta_2 \neq \theta_2$$

$\Rightarrow$ $\hat{\theta}_2$ is unbiased

reason why we do not use $\hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$

as an estimator for σ^2

Example

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

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

Show that the sample variance

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

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

is an unbiased estimator of θ_2.

Note that $S^2 = \frac{n}{n-1} \hat{\theta}_2$ and so

$$E[S^2] = E\left[\frac{n}{n-1}\hat{\theta}_2\right] = \frac{n}{n-1} E[\hat{\theta}_2] = \frac{n}{n-1} \cdot \frac{n-1}{n} \theta_2 = \theta_2$$

σ^2