

Mathematical Statistics

Anna Janicka

Lecture IV, 11.03.2019

POINT ESTIMATION

Plan for today

1. Estimation
2. Sample characteristics as estimators
3. Estimation techniques
 - method of moments
 - method of quantiles
 - maximum likelihood method



Point Estimation

- The choice, on the base of the data, of *the best* parameter θ , from the set of parameters which may describe P_θ
- An **Estimator** of parameter θ is any statistic $T = T(X_1, X_2, \dots, X_n)$ with values in Θ (we interpret it as an approximation of θ). Usually denoted by $\hat{\theta}$
- Sometimes we estimate $g(\theta)$ rather than θ .



Estimation: an example

Empirical frequency

Quality control example:

0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 1

□ Model: $X = \{0, 1, 2, \dots, n\}$ (here $n=50$),

$$P_{\theta}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x} \text{ for } \theta \in [0, 1]$$

□ parameter θ : probability of faulty element

□ an obvious estimator: $\hat{\theta} = X/n = 6/50$

n – sample size

X – number of faulty elements in sample



Problems with (frequency) estimators...

Example: three genotypes in a population, with frequencies $\theta^2 : 2\theta(1-\theta) : (1-\theta)^2$

In a population of size n , N_1 and N_2 and N_3 individuals of particular genotypes were observed.

Should we take $\hat{\theta} = \sqrt{N_1/n}$? or rather $\hat{\theta} = 1 - \sqrt{N_3/n}$? How about $\hat{\theta} = N_1/n + \frac{1}{2} N_2/n$?

Maybe something else?

→ How do we choose the best one?



Estimation – sample characteristics

Sample characteristics:

estimators based on the empirical distribution (empirical CDF)



Empirical CDF

- Let X_1, X_2, \dots, X_n be a sample from a distribution given by F (modeled by $\{P_F\}$)

(n -th) empirical CDF

$$\hat{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, t]}(X_i) = \frac{\text{number of observations } X_i : X_i \leq t}{n}$$

- For a given realization $\{X_i\}$ it is a function of t , the CDF of the empirical distribution (uniform over x_1, x_2, \dots, x_n). For a given t it is a statistic with a distribution

$$P(\hat{F}(t) = \frac{k}{n}) = \binom{n}{k} F(t)^k (1 - F(t))^{n-k}, \quad k = 0, 1, \dots, n$$



Empirical CDF: properties

1. $E_F \hat{F}_n(t) = F(t)$

2. $\text{Var} \hat{F}_n(t) = \frac{1}{n} F(t)(1 - F(t))$

3. from CLT: $\frac{\hat{F}_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \sqrt{n} \xrightarrow{n \rightarrow \infty} N(0,1)$

i.e., for any z : $P\left(\frac{\hat{F}_n(t) - F(t)}{\sqrt{F(t)(1 - F(t))}} \sqrt{n} \leq z\right) \rightarrow \Phi(z)$

4. Glivenko-Cantelli Theorem

$$\sup_{t \in \mathbb{R}} |\hat{F}_n(t) - F(t)| \xrightarrow{a.s.} 0$$

for $n \rightarrow \infty$

if sample size increases, we will approximate the unknown distribution with any given level of precision



Order statistics

- Let X_1, X_2, \dots, X_n be a sample from a distribution with CDF F . If we organize the observations in ascending order:

$X_{1:n}, X_{2:n}, \dots, X_{n:n} \leftarrow$ **order statistics**

($X_{1:n} = \min, X_{n:n} = \max$)

- An empirical CDF is a stair-like function, constant over intervals $[X_{i:n}, X_{i+1:n})$



Distribution of order statistics

- Let X_1, X_2, \dots, X_n be independent random variables from a distribution with CDF F . Then $X_{k:n}$ has a CDF equal to

$$F_{k:n}(x) = P(X_{k:n} \leq x) = \sum_{i=k}^n \binom{n}{i} (F(x))^i (1 - F(x))^{n-i}$$

- If additionally the distribution is continuous with density f , then $X_{k:n}$ has density

$$f_{k:n}(x) = n \binom{n-1}{k-1} f(x) (F(x))^{k-1} (1 - F(x))^{n-k}$$



Sample moments and quantiles as estimators

Sample moments and quantiles are moments and quantiles of the empirical distribution, so they are estimators of the corresponding theoretical values.

- sample mean = estimator of the expected value
- sample variance = estimator of variance
- sample median = estimator of median
- sample quantiles = estimators of quantiles



Method of Moments Estimation (MM)

- We compare the theoretical moments (depending on unknown parameter(s)) to their empirical counterparts.
- Justification: limit theorems
- We need to solve a (system of) equation(s).



EMM – cont.

- If θ is single-dimensional, we use one equation, usually: $E_{\theta}X = \bar{X}$
- If θ is two-dimensional, we use two equations, usually:
$$\begin{cases} E_{\theta}X = \bar{X}, \\ \text{Var}_{\theta}X = \hat{S}^2 \end{cases}$$
- If θ is k -dimensional, we use k equations, usually
$$\begin{cases} E_{\theta}X = \bar{X}, \\ \text{Var}_{\theta}X = \hat{S}^2, \\ E_{\theta}(X - E_{\theta}X)^3 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3, \\ \dots E_{\theta}(X - E_{\theta}X)^k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k \end{cases}$$



MME – Example 1.

- Exponential model: X_1, X_2, \dots, X_n are a sample from an exponential distr. $\text{Exp}(\lambda)$.

we know: $E_{\lambda} X = \frac{1}{\lambda}$

equation: $\frac{1}{\lambda} = \bar{X}$

solution:

$$\hat{\lambda} = MME(\lambda) = \hat{\lambda}_{MM} = \frac{1}{\bar{X}}$$



MME – Example 2.

□ Gamma model: X_1, X_2, \dots, X_n are a sample from distr. $\text{Gamma}(\alpha, \lambda)$.

We know: $E_{\alpha, \lambda} X = \frac{\alpha}{\lambda}$, $\text{Var}_{\alpha, \lambda} X = \frac{\alpha}{\lambda^2}$

System of equations:

$$\frac{\alpha}{\lambda} = \bar{X}, \quad \frac{\alpha}{\lambda^2} = \hat{S}^2$$

Solution:

$$\hat{\lambda}_{MM} = \frac{\bar{X}}{\hat{S}^2}, \quad \hat{\alpha}_{MM} = \frac{\bar{X}^2}{\hat{S}^2}$$



$$f_{\alpha, \lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \text{ dla } x > 0$$

Method of Quantiles Estimation (MQ)

- If moments are hard to calculate or formulae are complicated, we can use quantiles instead of moments. We choose as many levels of p as we have parameters, and we put

$$q_p(\theta) = \hat{q}_p$$

or equivalently

$$F_\theta(\hat{q}_p) = p$$



MQE – Example 1.

- Exponential model: X_1, X_2, \dots, X_n are a sample from an exponential distr. $\text{Exp}(\lambda)$.

CDF: $F_\lambda = 1 - \exp(-\lambda x)$ for $\lambda > 0$

one parameter \rightarrow one equation, usually for the median

$$1 - \exp(-\lambda \hat{q}_{1/2}) = \frac{1}{2}$$

solution:

$$MQE(\lambda) = \hat{\lambda}_{MQ} = -\frac{\ln \frac{1}{2}}{\hat{q}_{1/2}} = \frac{\ln 2}{\text{Med}}$$



MQE – Example 2.

- Weibull Model: X_1, X_2, \dots, X_n are a sample from a distribution with CDF

$$F_{b,c} = 1 - \exp(-cx^b)$$

for $b=1$
exponential
distr. with
parameter c

where $b, c > 0$ are unknown parameters.

two parameters \rightarrow two equations, usually

quartiles
$$\begin{cases} 1 - \exp(-c\hat{q}_{1/4}^b) = 1/4 \\ 1 - \exp(-c\hat{q}_{3/4}^b) = 3/4 \end{cases}$$

solution:

$$MQE(b) = \hat{b}_{MQ} = \ln(\ln 4 / (\ln 4 - \ln 3)) / (\ln \hat{q}_{3/4} - \ln \hat{q}_{1/4}),$$

$$MQE(c) = \hat{c}_{MQ} = \ln 4 \hat{q}_{3/4}^{-\hat{b}}$$



Properties of MME and MQE estimators

- ❑ Simple conceptually
- ❑ Not too complicated calculations
- ❑ BUT: sometimes not optimal (large errors, bad properties for small samples)
- ❑ Better method (usually): maximum likelihood



Maximum Likelihood Estimation (MLE)

We choose the value of θ for which the obtained results have the highest probability

Likelihood – describes the (joint) probability f (density or discrete probability) treated as a function of θ , for a given set of observations;

$$L: \Theta \rightarrow \mathbb{R}$$

$$L(\theta) = f(\theta; x_1, x_2, \dots, x_n)$$



Maximum Likelihood Estimator

$\hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$ is the MLE of θ , if

$$f(\hat{\theta}(x_1, x_2, \dots, x_n); x_1, x_2, \dots, x_n) =$$
$$= \sup_{\theta \in \Theta} f(\theta; x_1, x_2, \dots, x_n)$$

for any x_1, x_2, \dots, x_n .

Denoted:

$$\hat{\theta} = \hat{\theta}_{ML} = MLE(\theta)$$

$$MLE(g(\theta)) = g(MLE(\theta))$$



MLE – practical problems

- Usually: sample of independent obs.

Then:

$$L(\theta) = f_{\theta}(x_1)f_{\theta}(x_2)\dots f_{\theta}(x_n)$$

- If $L(\theta)$ is differentiable, and θ is k -dimensional, then the maximum may be found by solving: $\frac{\partial L(\theta)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, k$
 - very frequently: instead of $\max L(\theta)$ we look for $\max l(\theta) = \ln(L(\theta))$
-



MLE – Example 1.

□ Quality control, cont. We maximize

$$L(\theta) = P_{\theta}(X = x) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

or equivalently maximize

$$l(\theta) = \ln \binom{n}{x} + \ln(\theta^x) + \ln((1 - \theta)^{n-x}) = \ln \binom{n}{x} + x \ln(\theta) + (n - x) \ln(1 - \theta)$$

i.e. solve $l'(\theta) = \frac{x}{\theta} - \frac{n-x}{1-\theta} = 0$

solution: $MLE(\theta) = \hat{\theta}_{ML} = \frac{x}{n}$



MLE – Example 2.

□ Exponential model: X_1, X_2, \dots, X_n are a sample from $\text{Exp}(\lambda)$, λ unknown.

We have: $L(\lambda) = f_\lambda(x_1, x_2, \dots, x_n) = \lambda^n e^{-\lambda \sum x_i}$

we maximize

$$l(\lambda) = \ln L(\lambda) = n \ln \lambda - \lambda \sum x_i$$

we solve

$$l'(\lambda) = \frac{n}{\lambda} - \sum x_i = 0$$

we get

$$\hat{\lambda}_{ML} = \frac{1}{\bar{X}}$$



MLE – Example 3.

□ Normal model: X_1, X_2, \dots, X_n are a sample from $N(\mu, \sigma^2)$. μ, σ unknown.

$$\begin{aligned} l(\mu, \sigma) &= \ln\left(\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{1}{2\sigma^2} \sum (x_i - \mu)^2\right)\right) \\ &= -\frac{n}{2} \ln(2\pi) - n \ln \sigma - \frac{1}{2\sigma^2} \left(\sum x_i^2 - 2\mu \sum x_i + n\mu^2\right) \end{aligned}$$

we solve

$$\begin{cases} \frac{\partial l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} (\sum x_i^2 - 2\mu \sum x_i + n\mu^2) = 0 \\ \frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2} \sum x_i - \frac{n\mu}{\sigma^2} = 0 \end{cases}$$

we get:

$$\hat{\mu}_{ML} = \bar{X}, \quad \hat{\sigma}_{ML}^2 = \frac{1}{n} \sum (x_i - \bar{X})^2$$





WARSAW UNIVERSITY
Faculty of Economic Sciences