

MATH 4750 / MSSC 5750

Instructor: Mehdi Maadooliat

Method of Moments

Maximum Likelihood Estimate

Bayesian Estimate

MOM

MLE

Bayes



Department of Mathematical and Statistical Sciences

7.3 Point Estimation Techniques

- Two methods are considered: the method of moments and the method of maximum likelihood.
- the information in a random sample X_1, X_2, \dots, X_n is used to make inferences about the unknown θ .
- The observed values of the random sample are denoted x_1, x_2, \dots, x_n .
- Further, a random sample X_1, X_2, \dots, X_n is referred to with the boldface \mathbf{X} the observed values in a random sample x_1, x_2, \dots, x_n with the boldface \mathbf{x} .
- The joint **pdf** of X_1, X_2, \dots, X_n is given by

$$\begin{aligned} f(\mathbf{x}|\theta) &= f(x_1, x_2, \dots, x_n|\theta) \\ &= f(x_1|\theta) \times f(x_2|\theta) \times \dots \times f(x_{\mathbf{n}}|\theta) = \prod_{i=1}^n f(x_i|\theta). \end{aligned} \tag{7.17}$$

7.2.3.1 Method of Moments Estimators

- The idea behind the **method of moments** is to equate population moments about the origin to their corresponding sample moments, where the r^{th} **sample moment about the origin**, denoted m_r , is defined as

$$m_r = \frac{1}{n} \sum_{i=1}^n X_i^r, \quad (7.18)$$

and subsequently to solve for estimators of the unknown parameters.

- Recall that the r^{th} population moment about the origin of a random variable X denoted α_r , was defined as $E[X^r]$.
- It follows that $\alpha_r = E[X^r] = \sum_{i=1}^{\infty} x_i^r \mathbb{P}(X = x_i)$ for discrete X , and that $\alpha_r = E[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx$ for continuous X .

- Specifically, given a random sample X_1, X_2, \dots, X_n from a population with **pdf** $f(x|\theta_1, \theta_2, \dots, \theta_k)$, the method of moments estimators, denoted $\tilde{\theta}_i$ for $i = 1, \dots, k$ are found by equating the first k population moments about the origin to their corresponding sample moments and solving the resulting system of simultaneous equations given in Equation (7.19).

$$\begin{cases} \alpha_1(\theta_1, \dots, \theta_k) = m_1 \\ \alpha_2(\theta_1, \dots, \theta_k) = m_2 \\ \quad \quad \quad \vdots \quad \quad \quad \vdots \\ \alpha_k(\theta_1, \dots, \theta_k) = m_k \end{cases} \quad (7.19)$$

Example 7.10 Given a random sample of size n from a $Bin(1, \pi)$ population, find the method of moments estimator of π .

Solution: The first sample moment m_1 is \bar{X} and the first population moment about zero for the binomial random variable is $\alpha_1 = E[X^1] = 1 \cdot \pi$. By equating the first population moment to the first sample moment,

$$\alpha_1(\pi) = \pi \stackrel{\text{set}}{=} \bar{X} = m_1,$$

which implies that the method of moments estimator for π , is $\tilde{\pi} = \bar{X}$. ■



Example 7.11 Given a random sample of size m from a $Bin(n, \pi)$ population, find the method of moments estimator of π .

Solution: The first sample moment m_1 is \bar{X} and the first population moment about zero for the binomial random variable is $\alpha_1 = E[X^1] = n \cdot \pi$. By equating the first population moment to the first sample moment,

$$\alpha_1(\pi) = n\pi \stackrel{\text{set}}{=} \bar{X} = m_1,$$

which implies that the method of moments estimator for π , is $\tilde{\pi} = \frac{\bar{X}}{n}$. ■

Example 7.12 Given a random sample of size n from a $Pois(\lambda)$ population, find the method of moments estimator of λ .

Solution: The first sample moment m_1 is \bar{X} and the first population moment about zero for a Poisson random variable is $\alpha_1 = E[X^1] = \lambda$. By equating the first population moment to the first sample moment,

$$\alpha_1(\pi) = \lambda \stackrel{\text{set}}{=} \bar{X} = m_1,$$

which implies that the method of moments estimator for λ , is $\tilde{\lambda} = \bar{X}$. ■



Example 7.13 Given a random sample of size n from a $N(\mu, \sigma)$ population, find the method of moments estimators of μ and σ^2 .

Solution: The first and second sample moments m_1 and m_2 are \bar{X} and $\frac{1}{n} \sum_{i=1}^n X_i^2$ respectively. The first and second population moments about zero for a normal random variable are $\alpha_1 = E[X^1] = \mu$ and $\alpha_2 = E[X^2] = \sigma^2 + \mu^2$. By equating the first two population moments to the first two sample moments,

$$\begin{cases} \alpha_1(\mu, \sigma^2) = \mu \stackrel{\text{set}}{=} \bar{X} = m_1 \\ \alpha_2(\mu, \sigma^2) = \sigma^2 + \mu^2 \stackrel{\text{set}}{=} \frac{1}{n} \sum_{i=1}^n X_i^2 = m_2. \end{cases} \quad (7.20)$$

Solving the system of equations in (7.20) yields $\tilde{\mu} = \bar{X}$ and $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = S_u^2$ as the method of moments estimators for μ and σ^2 respectively. ■



7.3.2 Likelihood and Maximum Likelihood Estimators

When sampling from a population described by a **pdf** $f(x|\theta)$, knowledge of θ provides knowledge of the entire population. The idea behind maximum likelihood is to select the value for θ that makes the observed data most likely under the assumed probability model.

When x_1, x_2, \dots, x_n

are the observed values of a random variable X from a population with parameter θ , the notation $L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta)$ will be used to indicate that the distribution depends on the parameter θ , and \mathbf{x} to indicate the distribution is dependent on the observed values from the sample. Once the sample values are observed, $L(\theta|\mathbf{x})$ can still be evaluated in a formal sense, although it no longer has a probability interpretation (in the discrete case) as does (7.17).

$L(\theta|\mathbf{x})$ is the **likelihood function** of θ for \mathbf{x} and is denoted by

$$L(\theta|\mathbf{x}) = f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta) = f(x_1|\theta) \times f(x_2|\theta) \times \cdots \times f(x_n|\theta). \quad (7.22)$$

The key difference between (7.17) and (7.22) is that the joint **pdf** given in (7.17) is a function of \mathbf{x} for a given θ and the likelihood function given in (7.22) is a function of θ for given \mathbf{x} .

The value of θ that maximizes $L(\theta|\mathbf{x})$ is called the **maximum likelihood estimate** (mle) of θ . Another way to think of the mle is the mode of the likelihood function. The maximum likelihood estimate is denoted as $\hat{\theta}(\mathbf{x})$, and the maximum likelihood estimator (MLE), a statistic, as $\hat{\theta}(\mathbf{X})$.

LOG-LIKELIHOOD VS LIKELIHOOD

In general, the likelihood function may be difficult to manipulate, and it is usually more convenient to work with the natural logarithm of $L(\theta|\mathbf{x})$, called the **log-likelihood function**, since it converts products into sums. Finding the value θ that maximizes the log-likelihood function ($\ln L(\theta|\mathbf{x})$) is equivalent to finding the value of θ that maximizes $L(\theta|\mathbf{x})$ since the natural logarithm is a monotonically increasing function. If $L(\theta|\mathbf{x})$ is differentiable with respect to θ , a possible mle is the solution to

$$\frac{\partial(\ln L(\theta|\mathbf{x}))}{\partial\theta} = 0. \quad (7.23)$$

```
> L <- logL <- NULL; par(mfrow=c(1,2));
> n <- 10; mus <- seq(0,10,length=100); mu <- 5; x <- rnorm(n,mean=mu);
> Like <- function(mu, data=x) {prod(dnorm(x,mean=mu))}
> logLike <- function(mu, data=x) {sum(dnorm(x,mean=mu,log=TRUE))}
> max.L <- optim(1, Like, data=x, control=list(fnscale=-1))
> # max.L <- optim(1, Like, data=x, control=list(fnscale=-1), method="Brent",lower=0,upper=10)
> max.logL <- optim(1, logLike, data=x, control=list(fnscale=-1))
> for (i in 1:length(mus)) {L[i] <- Like(mus[i], x); logL[i] <- logLike(mus[i], x)}
> plot(mus,L,type="l"); abline(v=max.L$par, col=2)
> plot(mus,logL,type="l"); abline(v=max.logL$par, col=2)
```

BAYESIAN METHODS

- In the Bayesian paradigm, information brought by
 - the data x , realization of

$$X \sim f(x|\theta),$$

- combined with prior information specified by *prior distribution* with density $\pi(\theta)$
- Summary in a probability distribution, $\pi(\theta|x)$, called the **posterior distribution**
- Derived from the *joint* distribution $f(x|\theta)\pi(\theta)$, according to

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int f(x|\theta)\pi(\theta)d\theta},$$

[Bayes Theorem]

- where

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta$$

is the *marginal density* of X

EXAMPLE: BINOMIAL BAYES ESTIMATOR

- For an observation X from the binomial distribution $\text{Binomial}(n, p)$ the (so-called) conjugate prior is the family of beta distributions $\text{Beta}(a, b)$
- The classical Bayes estimator δ^π is the posterior mean

$$\begin{aligned}\delta^\pi &= \frac{\Gamma(a+b+n)}{\Gamma(a+x)\Gamma(n-x+b)} \int_0^1 p p^{x+a-1} (1-p)^{n-x+b-1} dp \\ &= \frac{n}{a+b+n} \left(\frac{x}{n}\right) + \frac{a+b}{a+b+n} \left(\frac{a}{a+b}\right).\end{aligned}$$

- A Biased estimator of p

THE VARIANCE/BIAS TRADE-OFF

- Bayes Estimators are biased
- Mean Squared Error (MSE) = Variance + Bias²
 - $\text{MSE} = \text{E}(\delta^\pi - p)^2$
 - Measures average closeness to parameter
- Small Bias \uparrow can yield large Variance \downarrow .

$$\delta^\pi = \frac{n}{a+b+n} \left(\frac{x}{n} \right) + \frac{a+b}{a+b+n} \left(\frac{a}{a+b} \right)$$

$$\text{Var} \delta^\pi = \left(\frac{n}{a+b+n} \right)^2 \text{Var} \left(\frac{x}{n} \right)$$