Lecture 2: Maximum Likelihood Estimation for a Random Sample from a Normal Population

MATH 667-01
Statistical Inference
University of Louisville

August 24, 2017
Last corrected: 8/30/2017

Introduction

- We begin by reviewing joint pdfs and independence concerning multivariate distributions from Section 4.6 in Casella and Berger (2001)[1].
- We will introduce some terminology regarding random samples which are discussed in Section 5.1.
- We will derive the maximum likelihood estimator of the parameters of a normal distribution based on a random sample. This is discussed in Example 7.2.11 and Example 7.2.12.

[1] Casella, G. and Berger, R. (2001). Statistical Inference, second edition. Duxbury Press.

Review

- *Definition L2.1* (p.177): If $\boldsymbol{X} = (X_1, \ldots, X_n)$ is a continuous random vector, then the *joint pdf* of $\boldsymbol{X}$ is a function $f(x_1, \ldots, x_n)$ such that

$$P(\boldsymbol{X} \in A) = \int \cdots \int_A f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n.$$

- *Definition L2.2* (p.178): If $(X_1, \ldots, X_n)$ is a continuous random vector with joint pdf $f(x_1, \ldots, x_n)$, then the *marginal pdf* of $(X_1, \ldots, X_k)$ is

$$f_{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k}(x_1, \ldots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_{k+1} \cdots dx_n.$$

- *Definition L2.3* (Def 4.6.5 on p.182): Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ be random vectors with joint pdf or pmf $f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Let $f_{\boldsymbol{X}_i}(\boldsymbol{x}_i)$ denote the marginal pdf or pmf of $\boldsymbol{X}_i$. Then $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ are called *mutually independent random vectors* if, for every $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$,

$$f(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) = f_{\boldsymbol{X}_1}(\boldsymbol{x}_1) \bullet \cdots \bullet f_{\boldsymbol{X}_n}(\boldsymbol{x}_n) = \prod_{i=1}^{n} f_{\boldsymbol{X}_i}(\boldsymbol{x}_i).$$

If the X_i's are all one-dimensional, then $X_1, \ldots, X_n$ are called *mutually independent random variables*.

Basic Concepts of Random Samples

- In statistics, we consider experimental situations where we want to model a population based on a sample (several observed data values from that population). To do so, we must model the data collection process used to collect the data.
- *Definition L2.4* (Def 5.1.1 on p.207): The random variables $X_1, \dots, X_n$ are called a *random sample of size* n *from a population* $f(x)$ if $X_1, \dots, X_n$ are mutually independent random variables and the marginal pdf or pmf of each X_i is $f(x)$. Alternatively, $X_1, \dots, X_n$ are called *independent and identically distributed* (iid) random variables with pdf or pmf $f(x)$.
- The definition above is sometimes referred to as sampling from an *infinite* population.

Basic Concepts of Random Samples

- If $X_1, \ldots, X_n$ is a random sample of size n from a population with a parametric pdf/pmf $f(x|\theta)$, then the joint pdf/pmf of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta).$$

- *Example L2.1*: Let $X_1, \ldots, X_n$ be a random sample from a normally distributed population with pdf

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}.$$

What is the joint pdf of $X_1, \ldots, X_n$?

- *Answer to Example L2.1*:

$$
\begin{aligned}
f(x_1,\ldots,x_n|\mu,\sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i-\mu)^2} \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \prod_{i=1}^{n} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) \\
&= \left((2\pi\sigma^2)^{-1/2}\right)^n \exp\left(-\sum_{i=1}^{n} \frac{1}{2\sigma^2}(x_i-\mu)^2\right) \\
&= (2\pi\sigma^2)^{-n/2} \exp\left(\frac{-\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right)
\end{aligned}
$$

- *Example L2.2*: Let $X_1, \ldots, X_n$ be a random sample from a normally distributed population with pdf

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

  where $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Find the MLE of (μ,σ^2) and show that it is a maximizer.

- *Answer to Example L2.2*: Since the natural logarithm is an increasing function, maximizing the likelihood function is equivalent to maximizing

$$\begin{aligned}
\ell(\mu, \sigma^2) &= \ln\left\{(2\pi\sigma^2)^{-n/2} \exp\left(\frac{-\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right)\right\} \\
&= \ln\left\{(2\pi\sigma^2)^{-n/2}\right\} + \ln\left\{\exp\left(\frac{-\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}\right)\right\} \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.
\end{aligned}$$

MLE for Normal Population

- *Answer to Example L2.2 continued*: Since

$$\ell(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

  is differentiable for all parameter values in its domain, any local extrema must satisfy $\frac{\partial \ell}{\partial \mu} = 0$ and $\frac{\partial \ell}{\partial \sigma^2} = 0$.
- Solving the system of equations

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0$$

$$\frac{-n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0,$$

  we obtain $\hat{\mu} = \bar{x}$ and $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

MLE for Normal Population

- *Answer to Example L2.2 continued*: One way to show that the solution to these equations maximizes ℓ is to use successive maximizations.
- For each fixed $\sigma^2 > 0$, note that $\ell(\bar{x}, \sigma^2) \geq \ell(\mu, \sigma^2)$ for all μ since $\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{n}{\sigma^2} < 0$. (Alternately, this can be done by showing $\sum_{i=1}^n (x_i - \mu)^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$, with equality if and only if $\mu = \bar{x}$, which is proven in Thm 5.2.4 on p.212.)
- Then we consider the *profile likelihood*

$$\ell^*(\sigma^2) = \ell(\bar{x}, \sigma^2) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \bar{x})^2.$$

- *Answer to Example L2.2 continued*: Since

$$\frac{d\ell^*}{d(\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \bar{x})^2$$
$$= -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(n\widehat{\sigma^2})$$
$$= -\frac{n}{2(\sigma^2)^2}\left(\sigma^2 - \widehat{\sigma^2}\right)$$

is positive if $\sigma^2 < \widehat{\sigma^2}$ and negative if $\sigma^2 > \widehat{\sigma^2}$, $\ell^*(\sigma^2)$ is maximized at $\sigma^2 = \widehat{\sigma^2}$.

- This proves that $(\hat{\mu}, \widehat{\sigma^2})$ is the MLE of (μ, σ^2) since

$$\ell(\hat{\mu}, \widehat{\sigma^2}) = \ell^*(\widehat{\sigma^2}) \geq \ell^*(\sigma^2) = \ell(\hat{\mu}, \sigma^2) \geq \ell(\mu, \sigma^2)$$

for all μ and σ^2.