STA 522, Spring 2022
Introduction to Theoretical Statistics II

Lecture 5

Department of Biostatistics
University at Buffalo

AGENDA

- Comments on Exam 1
- Point Estimation
- Method of Moments
- Method of Maximum Likelihood
- Review for Exam 1

Review: Likelihood Function

- Let $f(\underline{x} \mid \theta)$ denote the joint pdf or pmf of the sample $\underline{X} = (X_1, X_2, \ldots, X_n)$. Then, given that $\underline{X} = \underline{x}$ is observed, the function of θ defined by $L(\theta \mid \underline{x}) = f(\underline{x} \mid \theta)$ is called the **likelihood function**.

- **Example (Poisson Likelihood):** Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ denote a random sample from a Poisson distribution with mean λ. The likelihood function for $0 < \lambda < \infty$ is given by:

$$L(\lambda \mid \underline{x}) = P_\lambda(\underline{X} = \underline{x}) = \exp(-n\lambda)\, \frac{\lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n x_i!}$$

- **Example (Normal Likelihood):** Let $\underline{X} = (X_1, X_2, \ldots, X_n)$ denote a random sample from a $\mathrm{N}(\mu, \sigma^2)$ distribution. The likelihood function for $-\infty < \mu < \infty$ and $\sigma > 0$ is given by

$$L(\mu, \sigma \mid \underline{x}) = f(\underline{x} \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right)$$

Point Estimation

Background: When sampling is from a population with pdf/pmf $f(\underline{x}|\theta)$, knowledge of θ yields knowledge of the entire population. Given a sample we to find a meaningful reasonable "estimator" of the point θ.

- **Example:** Suppose you have a random sample $X_1, X_2, \ldots, X_n$ from a $N(\mu, \sigma^2)$ population. How do we determine μ and σ^2 from $X_1, X_2, \ldots, X_n$?

Definition: A point estimator is *any function* $W(X_1, X_2, \ldots, X_n)$ of a sample; that is, any statistic is a point estimator.

- **NOTE:** There is no mention in the definition of any correspondence between the estimator and the parameter. Also there is no mention in the definition of the range of the statistic $W(X_1, X_2, \ldots, X_n)$. This ensures that we do not eliminate any candidates from consideration.

Estimate vs. Estimator

- An estimator is a function of the random sample, while an estimate is the *realized value* of an estimator that is obtained when a sample is actually taken.
- Thus, an estimator is a random variable whereas an estimate is its observed value.

Method of Finding Point Estimators

- Method of moments
- Method of maximum likelihood
- Bayesian Methods (later)

Moments

Definition: The r-th moment about the origin (or raw moment) of a random variable X, denoted by μ'_r , is defined as $\mu'_r = E(X^r)$.

- Note that $\mu'_1 = E(X) = \mu =$ population mean.

Definition: The r-th moment about the mean (or central moment) of a random variable X, denoted by μ_r , is defined as $\mu_r = E[(X-\mu)^r]$ where $\mu = \mu'_1$ is the population mean (first raw moment).

- Note that $\mu_1 = E[(X-\mu)] = 0$ and $\mu_2 = E(X-\mu)^2 = \sigma^2 =$ population variance.

Definition: The kth sample (raw) moment of a random sample $X_1, X_2, \ldots, X_n$ is the mean of the kth powers, denoted by m_k and defined as $m_k = \dfrac{1}{n}\sum_{i=1}^{n} X_i^k$. Note that $m_1 = \overline{X}$ and

$$\tilde{S}^2 := m_2 - m_1^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Method of Moments

Let $X_1, X_2, \ldots, X_n$ be a sample from a population with pdf/pmf $f(\underline{x} \mid \theta_1, \ldots, \theta_k)$ ·

Method of moments estimators of $\theta_1, \ldots, \theta_k$ are obtained by equating the first k sample *raw* moments to the corresponding k population *raw* moments, then solving to get $\hat{\theta}_1, \ldots, \hat{\theta}_k$.

- i.e., set $m_j = \mu'_j = \mu'_j(\theta_1, \ldots, \theta_k)$ for $j = 1, \ldots, k$ (as many equations as we have parameters) and solve for the parameters.

Note that $m_j = m_{j,n} \xrightarrow{a.s.} \mu'_j$ for $j \geq 1$ under standard regularity conditions, using SLLN.

- i.e., in large samples the sample moment get arbitrarily close to the population moments, so the estimation will be reasonable.
- Estimation can be sub-optimal in small samples

NOTE: When there are just two parameters in the distribution (e.g., $\mathrm{N}(\mu, \sigma^2)$), it is sometimes easier to solve the equations $\mu = \overline{X}$ and $\sigma^2 = \tilde{S}^2$. Must write μ and σ^2 in terms of the model parameters first.

Example (Normal Method of Moments): Suppose $X_1, X_2, \ldots, X_n \sim \text{iid}\, \text{N}(\mu, \sigma^2)$. Use the method of moments to find estimators of the parameters μ and σ^2.

For a $\text{N}(\mu, \sigma^2)$ distribution, $\mu_1' = \mu$ and $\mu_2 = \sigma^2$. Therefore the method of moment estimators are given by $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Example (Uniform Method of Moments): Let $X_1, X_2, \ldots, X_n \sim \text{iidUniform}(0, \theta)$ for $\theta > 0$. Use the method of moments to find an estimator of the parameter θ.

Here $\mu_1' = E(X) = \frac{\theta}{2}$. Therefore the method of moments estimator for θ is obtained as $\hat{\theta} = 2\overline{X}$.

Example (Gamma Method of Moments): Let $X_1, X_2, \ldots, X_n \sim \text{iid Gamma}(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$. Use the method of moments to find estimators of the parameters α and β.

Note at the outset that for the Gamma(α, β) distribution, $\mu_1' = \alpha\beta$ and $\mu_2 = \alpha\beta^2$. Write $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Then the method of moment estimators are: $\hat{\beta} = \tilde{S}^2/\overline{X}$ and $\alpha = \overline{X}^2/\tilde{S}^2$

Example (Binomial Method of Moments): Let $X1, X2, ..., Xn \sim$ iid Binomial(k, p), k is a positive integer, $p > 0$ and both k and p are unknown.

Exemplary Situation: want to estimate crime rates for crimes that are known to have many unreported occurrences. For such a crime, both the true reporting rate, p, and the total number of occurrences, k, are both unknown.

Write $\tilde{S}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$. Method of moment estimators are obtained from the system of equations:

$$\overline{X} = \hat{k}\hat{p}$$
$$\tilde{S}^2 = \hat{k}\hat{p}(1 - \hat{p})$$

Solving, we get

$$\hat{k} = \frac{\overline{X}^2}{\overline{X} - \tilde{S}^2} \text{ and } \hat{p} = \frac{\overline{X}}{\hat{k}}$$

NOTE: $\hat{k}$ can be negative.

Maximum Likelihood Estimation

Definition: For each sample point $\underline{x}$, let $\hat{\theta}(\underline{x})$ be a parameter value at which the likelihood function $L(\theta \mid \underline{x})$ attains its maximum as a function of θ, with $\underline{x}$ held fixed. A maximum likelihood estimator (MLE) of the parameter θ based on a sample $\hat{X}$ is $\hat{\theta}(\underline{X})$.

Notes: - Since the logarithm function is strictly increasing on $(0, \infty)$ (and so one-to-one), the value which maximizes $\log L(\theta \mid \underline{x})$ is the same value that maximizes $L(\theta \mid \underline{x})$.

- $l(\theta \mid \underline{x}) := \log L(\theta \mid \underline{x})$ is called the *log-likelihood* function
- Often maximizing $\log L(\theta \mid \underline{x})$ is easier than maximizing $L(\theta \mid \underline{x})$.

Example: (Binomial MLE) Let $X_1, X_2, \ldots, X_n \sim$ iid Bernoulli(p) where $0 \leq p \leq 1$. Find the MLE for p.

The likelihood function is

$$L(p \mid \underline{x}) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^y(1-p)^{n-y}$$

where $y = \sum_{i=1}^{n} x_i$.

First consider $0 < y < n$. We'll maximize the log-likelihood:

$$l(p \mid \underline{x}) := \log L(p \mid \underline{x}) = y \log p + (n-y)\log(1-p)$$

$$\frac{\partial}{\partial p} l(p \mid \underline{x}) = \frac{y}{p} - \frac{n-y}{1-p} \overset{\text{set}}{=} 0 \implies p = \frac{y}{n}$$

Straightforward to verify that $\left.\dfrac{\partial^2}{\partial p^2} l(p \mid \underline{x})\right|_{p=y/n} < 0$ meaning $p = y/n$ maximizes $l(p \mid \underline{x})$ when $0 < y < n$.

If $y = 0$ or $y = n$ then

$$l(p \mid \underline{x}) := \log L(p \mid \underline{x}) = \begin{cases} n \log(1-p) & \text{if } y = 0 \\ n \log p & \text{if } y = n \end{cases}$$

In each case $l(p \mid \underline{x})$ is a monotone function of p, and is maximized at $p = y/n$.

Thus, $\hat{p} = \dfrac{y}{n}$ is the MLE of $\hat{p}$.

Homework:

1. Find the method of moment estimator of p.
2. Consider $X_1, X_2, \ldots, X_n \sim$ iid Binomial(m, p) where m is a fixed, known positive integer, and p is unknown. What are the MLE and method of moments estimator of p?