MATH 503: Mathematical Statistics

Lecture 2: Dist. Theory, and Estimation

Readings: Sections 5.4-5.5, 6.3, 7.1-7.2

Kimberly F. Sellers

Department of Mathematics and Statistics

Today's Topics

- Delta Method
- Order Statistics
- (Point) Estimation Theory
  - Method of moments
  - Maximum likelihood estimation

Intro. to Statistical Inference

- Setup: have random variable X with unknown pdf (or pmf):
  1. $f(x)$ [or $p(x)$] completely unknown
  2. $f(x)$ [or $p(x)$] known, but based on θ unknown
- Goal: estimate θ
- Estimation based on sampling

Random Sample

- The random variables $X_1, \ldots, X_n$ constitute a random sample on a random variable X if they are independent and identically distributed (i.e. have the same distribution).
- Implications:

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F(x_i)$$

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$$

Statistic

- Suppose the n random variables $X_1, \dots, X_n$ constitute a sample from the distribution of a random variable X. Then any function $T = T(X_1, \dots, X_n)$ of the sample is a <u>statistic</u>.
- T is a random variable
- T is <u>unbiased</u> $\Leftrightarrow E(T) = \theta$
- T is <u>consistent</u> $\Leftrightarrow T \xrightarrow{p} \theta$
- $T(X_1, \dots, X_n)$ point estimator [$T(x_1, \dots, x_n)$ point estimate] of θ

Exercise

Let $X_1, \dots, X_n$ be iid $N(\mu, \sigma^2)$. Show that $\bar{X}$ and S^2 are both unbiased estimators for μ and σ^2, respectively.

Delta Method

- **Goal:** to derive large sample moments of a transformation of a (consistent) statistic

- **Theorem**: Let $\{Y_n\}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. Suppose the function g is differentiable at θ and $g'(\theta) \neq 0$. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2(g'(\theta))^2)$$

Slutsky's Theorem

- Let X_n, X, A_n, B_n be random variables and let a and b be constants. If $X_n \xrightarrow{d} X$, $A_n \xrightarrow{p} a$ and $B_n \xrightarrow{d} b$, then

$$A_n + B_n X_n \xrightarrow{d} a + bX.$$

Delta Method Derivation

- Let T be a statistic s.t. $E(T) = \mu$ and $V(T) = \sigma^2$. We want the moments for $g(T)$, where g is twice differentiable function

- Taylor expansion gives

$$g(t) \;=\; g(\mu) \;+\; g'(\mu)(t-\mu) \;+\; R_2(a),$$

where $R_2(a) = ½g''(a)(t-\mu)^2$ for some $a \in (t, \mu)$.

Delta Method Derivation (cont.)

- By consistency, remainder vanishes

$$\Rightarrow g(T) \;=\; g(\mu) \;+\; g'(\mu)(T-\mu),$$

- $E(g(T)) = \; g(\mu) \;+\; g'(\mu)E(T-\mu)$

  $= \; g(\mu)$

- $V\big(g(T)\big) = E[g(T) - g(\mu)]^2$

  $= E[g'(\mu)(T-\mu)]^{\,2}$

  $= \; [g'(\mu)]^2 V(T)$

Delta Method Derivation (cont.)

- Slutsky's Thm ⇒ asymptotic distribution of transformations of statistics of interest

- **Theorem**: Let $\{Y_n\}$ be a sequence of random variables such that $\sqrt{n}(Y_n - \theta) \xrightarrow{d} N(0, \sigma^2)$. Suppose the function g is differentiable at θ and $g'(\theta) \neq 0$. Then

$$\sqrt{n}(g(Y_n) - g(\theta)) \xrightarrow{d} N(0, \sigma^2(g'(\theta))^2)$$

Example

- Let $X_i, i = 1,2, \ldots$ be independent Bernoulli(p) random variables and let $Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then,
$$\sqrt{n}(Y_n - p) \xrightarrow{d} N(0, p(1-p))$$
- Use the delta method to find the asymptotic distribution of $\log(Y_n)$.

Example

- Let $X_1, \dots, X_n \sim$Poisson(λ) iid. By CLT,

$$\sqrt{n}(\bar{X} - \lambda) \xrightarrow{d} N(0, \lambda)$$

- Use the delta method to determine the asymptotic distribution using the function, $g(x) = \exp(.5x)$.

Order Statistics

- Consider ordering random sample $X_1, \ldots, X_n$, by denoting $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$

  $X_{(1)} =$ minimum of X_is

  $X_{(n)} =$ maximum of X_is

  $X_{(k)} = k$th order statistic (i.e. kth smallest) of X_is

- Assume X_is are continuous with density $f(x)$, and cdf $F(x)$. What are the pdf and cdf for the order statistics?

CDF and PDF for $X_{(n)}$

CDF and PDF for $X_{(1)}$

More generally….

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} F^j(x)[1 - F(x)]^{n-j}$$

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!\,(n-k)!} F^{k-1}(x)[1 - F(x)]^{n-k} f(x)$$

- Thm: The joint pdf of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ is given by

$$f_{X_{(1)},X_{(2)},\ldots,X_{(n)}}\left(x_{(1)}, x_{(2)}, \ldots, x_{(n)}\right) = n! \prod_{i=1}^{n} f\left(x_{(i)}\right), \quad x_{(1)} < \cdots < x_{(n)}$$

Estimation Theory

- Two common approaches to estimating θ
  - Method of moments (MOM)
  - Maximum likelihood estimation (MLE)

Method of Moments

- Applies only for $X_1, \dots, X_n$ iid with distribution depending on unknown parameter θ
- Idea:

  $E(X) = g(\theta)$, thus $g(\tilde{\theta}) = \bar{X}$.

  Solve for $\tilde{\theta}$.

- More generally, find first nontrivial moment to determine estimator

Example

- Let $X_1, \ldots, X_n$ iid ~ Unif$(0, \theta)$. Find the MOM for θ.

Example

- Let $X_1, \ldots, X_n$ iid ~ Beta$(\theta + 1, 1)$. Find the MOM for θ.

Example

- Let $X_1, \ldots, X_n$ iid ~ Unif$(-\theta, \theta)$. Find the MOM for θ.

Notes Regarding MOMs

- For $X_1, \ldots, X_n$ iid with distribution depending on unknown parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$

$$
\begin{aligned}
E(X) &= g_1(\boldsymbol{\theta}) & \qquad g_1\left(\widetilde{\boldsymbol{\theta}}\right) &= \bar{X} \\
E(X^2) &= g_2(\boldsymbol{\theta}) & \qquad g_2\left(\widetilde{\boldsymbol{\theta}}\right) &= \overline{X^2} \\
&\vdots & &\vdots \\
E\left(X^k\right) &= g_k(\boldsymbol{\theta}) & \qquad g_k\left(\widetilde{\boldsymbol{\theta}}\right) &= \overline{X^k}
\end{aligned}
$$

Example

- Let $X_1, \ldots, X_n$ iid ~ Gamma(α, β). Find the MOM for α, β.

Example

- Let $X_1, \ldots, X_n$ iid ~ Beta(α, β). Find the MOM for α, β.

Maximum Likelihood Estimation

- Let $X_1, \dots, X_n$ iid ~ $f(x; \theta)$, $\theta \in \Omega$ unknown scalar
- Let $L(\theta; \boldsymbol{x}) = L(\theta; x_1, \dots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$.
- $L(\theta; \boldsymbol{x})$ is called a likelihood function
- The estimator, $\hat{\theta}$, that maximizes $L(\theta; \boldsymbol{x})$ over all θ is called the maximum likelihood estimator of θ.
- When dealing with differentiation where $L(\theta; \boldsymbol{x}) \neq 0$, it is generally better to consider

$$\underbrace{\frac{L'(\theta)}{L(\theta)}}_{\frac{\partial}{\partial\theta}(\log L(\theta))} = 0$$

Steps for Determining MLEs[*]

1. Determine $L(\theta; \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i; \theta)$
2. Transform to get $\ln L(\theta; \boldsymbol{x})$.
3. Differentiate $\ln L(\theta)$ wrt θ and set equal to 0. These are sometimes referred to as the estimating equation(s).
4. Solve for θ. This solution is then labeled as the MLE, $\hat{\theta}$.

*Note: this algorithm doesn't work when support space depends on θ.

Example

- Let $X_1, \dots, X_n$ iid ~ Exponential(θ). Find the MLE of θ.

Example

Let $X_1, \ldots, X_n$ iid ~ N(θ, σ^2), σ^2 known. Find the MLE of θ.

Example

Let $X_1, \ldots, X_n$ iid ~ Unif$(0, \theta)$. Find the MLE of θ.

Theorem 1

- Let θ_0 be the true parameter. Under the following regularity conditions, namely
  1. Pdfs are distinct, i.e. $\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x_i;\theta')$
  2. Pdfs have common support for all θ

$$\lim_{n\to\infty} P_{\theta_0}[L(\theta_0, \boldsymbol{X}) > L(\theta, \boldsymbol{X})] = 1, \text{for all } \theta \neq \theta_0$$

- The point: asymptotically, $L(\theta)$ is maximized at the true value θ_0

Proof to Theorem 1

$$L(\theta_0, \boldsymbol{X}) > L(\theta, \boldsymbol{X})$$

$$\Leftrightarrow \prod_{i=1}^{n} f(x_i; \theta_0) > \prod_{i=1}^{n} f(x_i; \theta)$$

$$\Leftrightarrow \sum_{i=1}^{n} \log f(x_i; \theta_0) > \sum_{i=1}^{n} \log f(x_i; \theta)$$

$$\Leftrightarrow \sum_{i=1}^{n} [\log f(x_i; \theta) - \log f(x_i; \theta_0)] < 0$$

$$\Leftrightarrow \frac{1}{n} \sum_{i=1}^{n} \left[\log\left(\frac{f(x_i; \theta)}{f(x_i; \theta_0)}\right)\right] < 0$$

Proof to Theorem 1 (cont.)

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[\log\left(\frac{f(x_i;\theta)}{f(x_i;\theta_0)}\right)\right] \underbrace{\xrightarrow{p}}_{\text{by WLLN}} & \; E_{\theta_0}\left[\log\left(\frac{f(x_i;\theta)}{f(x_i;\theta_0)}\right)\right] \\
\underbrace{\leq}_{\text{Jensen's Ineq.}} & \; \log E_{\theta_0}\left(\frac{f(x_i;\theta)}{f(x_i;\theta_0)}\right) \\
= & \; \log\left[\underbrace{\int \frac{f(x_i;\theta)}{f(x_i;\theta_0)} f(x_i;\theta_0)\mathrm{d}x}_{1}\right] \\
= & \; 0.
\end{aligned}
$$

Theorem 2

Let $X_1, \dots, X_n$ iid with pdf $f(x;\theta), \theta \in \Omega$. For a specified function g, let $\eta = g(\theta)$ be a parameter of interest. Suppose $\hat{\theta}$ is the mle of θ. Then $g(\hat{\theta})$ is the mle of $\eta = g(\theta)$.

Proof: For g a 1-1 function,

$$\max L\big(g(\theta)\big) = \max_{\eta = g(\theta)} L(\eta) = \max_{\eta} L\big(g^{-1}(\eta)\big).$$

Maximum occurs when $g^{-1}(\eta) = \hat{\theta} \Rightarrow \hat{\eta} = g(\hat{\theta})$. For g not 1-1, define set $g^{-1}(\eta) = \{\theta : g(\theta) = \eta\}$. Maximum occurs at $\hat{\theta}$, and domain of g is Ω which covers $\hat{\theta}$. Thus, $\hat{\theta}$ lies in (only one of) the preimages. Thus, choose $\hat{\eta}$ s.t. $g^{-1}(\hat{\eta})$ is that unique preimage containing $\hat{\theta} \therefore \hat{\eta} = g\big(\hat{\theta}\big)$.

Theorem 3

- Assume $X_1, \dots, X_n$ satisfy the following regularity conditions, where θ_0 is the true parameter:
  1. Pdfs are distinct, i.e. $\theta \neq \theta' \Rightarrow f(x_i;\theta) \neq f(x_i;\theta')$
  2. Pdfs have common support for all θ
  3. The point θ_0 is an interior point in Ω
- Further, assume $f(x;\theta)$ differentiable wrt $\theta \in \Omega$. Then the likelihood equation has a solution $\hat{\theta}$ s.t. $\hat{\theta} \xrightarrow{p} \theta_0$
- Corollary: If $\hat{\theta}$ is unique, then $\hat{\theta}$ is a consistent estimator of θ_0.