Topic 10 - Point Estimation

STAT 511

Professor Bruce Craig

Background Reading

Devore : Section 6.1 - 6.2

General Concepts

- **Inference**: Drawing some type of conclusion about one or more parameters (i.e., parameter characteristics)

Characteristic	Population	Sample
Mean	μ	$\overline{x}$
Variance	σ^2	s^2
Proportion	p	$\widehat{p}$

- A **point estimate** of parameter θ
  - Based on selecting a suitable statistic
  - Single numeric value computed from sample data
- Chosen statistic called a **point estimator**

Choice of Estimator

- Numerous estimators for each parameter
- For example

Problem #1: Interested in estimating the average flexural strength of concrete beams. If strength distribution relatively symmetric we could use the following estimators

  - Sample mean : $\overline{X}$
  - Sample median : $\tilde{X}$
  - .5Max(X) + .5Min(X)
- How do we choose the "most suitable"?

Unbiased Estimator

- Estimator $\widehat{\theta}$ said to be unbiased if $E(\widehat{\theta}) = \theta$
- **Bias** defined to be $E(\widehat{\theta}) - \theta$
- Proposition

If X_i's from distribution with mean μ, then $\overline{X}$ is unbiased estimator of μ. If distribution **continuous and symmetric**, then $\tilde{X}$, .5Max(X) + .5Min(X), and trimmed means also unbiased estimators.

- Bias looks only at accuracy
- Likely want to factor in precision

Unbiased Estimator

- Have shown $E(\hat{p}) = E(X/n) = p$ ($X \sim$ Bin(n,p))

- What about S^2?

If X_i have mean μ and variance σ^2

$$\begin{aligned} E(S^2) &= E\left(\frac{\sum (X_i - \overline{X})^2}{n-1}\right) \\ &= E\left(\frac{1}{n-1}\left[\sum X_i^2 - n\overline{X}^2\right]\right) \\ &= \frac{1}{n-1}\left[\sum E(X_i^2) - nE(\overline{X}^2)\right] \\ &= \frac{1}{n-1}\left[\sum (\sigma^2+\mu^2) - n(\sigma^2/n+\mu^2)\right] \\ &= \sigma^2 \end{aligned}$$

Precision of Estimator

- Estimator a random variable
  - Has mean and variance

- Use precision to choose estimator

- Among all unbiased estimators, choose the one that has minimum variance

- Estimator known as minimum variance unbiased estimator **(MVUE)**

Precision of Estimator

- Back to Problem # 1 : If X_i's a sample from N(μ, σ). Can show $\overline{X}$ is MVUE

- Does **not** mean $\overline{X}$ always the best
  - If $X_i \sim$ Uniform(a,b), then .5Max(X)+.5Min(X) is the MVUE
  - If $X_i \sim$ Cauchy, then $\tilde{X}$ is the MVUE
  - Trimmed mean does well in all cases but not MVUE

Standard Error

- The standard deviation of a estimator is known as the **standard error**

$$\sigma_\theta = \sqrt{V(\hat{\theta})}$$

- If σ_θ depends on unknown parameters, plug in estimates

- Known as estimated standard error

$$\hat{\sigma}_\theta = s_\theta = \sqrt{\hat{V}(\hat{\theta})}$$

- Example: If X_i Normal with μ unknown, standard error of $\overline{X}$ is $\sigma/\sqrt{n}$. If σ also unknown, then the estimated standard error is $s/\sqrt{n}$

Bootstrapping

- When $f(x)$ unknown or $\widehat{\theta}$ sufficiently complicated, may not be able to obtain an expression for s_θ
- The bootstrap is a computer intensive method to address this problem
  - Parametric bootstrap
  - Nonparametric bootstrap
- Generate B "new" samples of size n and compute $\widehat{\theta}$ for each sample
- Standard error estimated with

$$S_\theta = \sqrt{\frac{1}{B-1}\sum(\widehat{\theta}_i^\star - \overline{\theta}^\star)^2}$$

Bootstrapping

- Consider Example 6.10

1. Parametric bootstrap
   - Estimate λ from sample of $n = 10$ obs
   - Generate B new samples from Exp($\widehat{\lambda}$)
2. Nonparametric bootstrap
   - Don't assume distribution $f(x)$
   - Use data sample to represent distribution
   - Sample **with replacement** from sample
   - Generate B new samples

Methods of Estimation

- Method of Moments

  The kth moment of $f(x)$ is $E(X^k)$

  The kth sample moment is $\sum x^k/n$

  If m unknown parameters, equate the first m sample moments with the first m moments

- Example: $X_i \sim N(\mu, \sigma)$

$$\begin{aligned} \sum X/n &= E(X) = \mu \\ \sum X^2/n &= E(X^2) = \sigma^2 + \mu^2 \end{aligned}$$

$\widehat{\mu} = \overline{X}$ and $\widehat{\sigma} = \sqrt{\sum(X_i - \overline{X})^2/n}$

Methods of Estimation

- Maximum Likelihood

  Strongly recommended when n large

  Define $f_1 = f(x_1)$

  Likelihood function $L = f_1 \times f_2 \times \cdots \times f_n$

  Find θ's which maximize L

- Example: $X_i \sim N(\mu, \sigma)$

$$L = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\sum(x_i - \mu)^2/2\sigma^2\right)$$

$\widehat{\mu} = \overline{X}$ and $\widehat{\sigma} = \sqrt{\sum(X_i - \overline{X})^2/n}$

MLE Properties

- Invariance property
  - Let $\widehat{\theta}_1$, $\widehat{\theta}_2$,…,$\widehat{\theta}_m$ be the MLEs for parameters θ_1, θ_2,…,θ_m.
  - The MLE of $h(\theta_1, \theta_2, ..., \theta_m)$ is the function $h(\widehat{\theta}_1, \widehat{\theta}_2, ..., \widehat{\theta}_m)$.
- Large sample behavior
  - In many situations, the MLE is approximately unbiased and has variance nearly as small as the MVUE
  - Distribution of MLE approximately normal

Example : Problem #22

Let X denote the proportion of allotted time that a student spends on an aptitude test. Suppose the pdf of X is

$$f(x; \theta) = (\theta + 1)x^\theta \text{ for } 0 \le x \le 1$$

1. Use the method of moments to estimate θ

$$\begin{aligned} E(X) &= \int_0^1 x(\theta+1)x^\theta dx \\ &= \left(\frac{\theta+1}{\theta+2}\right) x^{\theta+2}\Big|_0^1 \\ &= \left(\frac{\theta+1}{\theta+2}\right) \end{aligned}$$

$$\begin{aligned} \overline{X} &= \left(\frac{\theta+1}{\theta+2}\right) \\ &= 1 - 1/(\theta+2) \\ &\downarrow \\ \widehat{\theta} &= \frac{1}{1-\overline{X}} - 2 \end{aligned}$$

Example : Problem #22

Let X denote the proportion of allotted time that a student spends on an aptitude test. Suppose the pdf of X is

$$f(x; \theta) = (\theta + 1)x^\theta \text{ for } 0 \le x \le 1$$

2. Obtain the MLE of θ

$$\begin{aligned} L &= (\theta+1)^n \prod X_i^\theta \\ &\downarrow \\ \log(L) &= n \log(\theta+1) + \theta \sum \log X_i \end{aligned}$$

To maximize $\log(L)$, take derivative with respect to θ and set equal to zero

$$\begin{aligned} \frac{\partial \log(L)}{\partial \theta} &= 0 \\ &= \frac{n}{\theta+1} + \sum \log X_i \\ &\downarrow \\ \widehat{\theta} &= \frac{n}{\sum \log X_i} - 1 \end{aligned}$$