What to do today (Jan 17, 2023)?

Part 2. Basic Statistical Inference (Chp 6-9)

Reminder: Homework 2 is assigned and due on Monday 5:00pm.

§2.1 Point Estimation (Chp6)

§2.1.1 Some General Concepts

2.1.1A. What does point estimation do?

Suppose r.v. $X \sim F(\cdot; \theta)$ *(***population***) with unknown* θ *(***parameter***).*

- *Use the available information (data, a sample from the population) to compute a 'good guess' (***point estimate***) for the true value of* θ
  - *The formula used to obtain a point estimate is called the* **point estimator** *of* θ*, denoted by* $\hat{\theta}$*.*
  - *A point estimator is a suitable statistic: it is often referred to as a realization or an evaluation of the corresponding point estimator.*

Example 2.2 (Devore 9th: p249) Reconsider the accompanying 20 observations on dielectric breakdown voltage for pieces of epoxy resin first introduced in textbook's Example 4.30 (Section 4.6). The pattern in the normal probability plot given there is quite straight, so we now assume that the distribution of breakdown voltage X is normal with mean value μ.

Example 2.3 (Devore 9th: p250) The article "Is a Normal Distribution the Most Appropriate Statistical Distribution for Volumertric Properties in Asphalt Mixtures?" first cited in the textbook's Example 4.26, reported 52 observations on $X =$ voids filled with asphalt (%) for 52 specimens of a certain type of hot-mix asphalt.

§2.1 Point Estimation (Chp6)

§2.1.1 Some General Concepts

§2.1.1B. Criteria for selecting a good estimator

Unbiased Estimator.

- **definition:** An estimator $\hat{\theta}$ of θ is **unbiased** if $E(\hat{\theta}) = \theta$.

- eg, with a random sample from a population: the sample mean $\bar{X}$ is an unbiased estimator of the population mean;
- eg, with a random sample from a population: the sample variance $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ is an unbiased estimator of the population variance.

  (*Check the results, assuming* $X \sim N(\mu, \sigma^2)$.)

$\Longrightarrow$ Use an unbiased estimator for a parameter in general, when possible.

Standard Error.

- **definition:** $\sigma_{\hat{\theta}} = \sqrt{Var(\hat{\theta})}$
  a measure on how well an unbiased estimator $\hat{\theta}$ estimates θ

- If both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased, $\hat{\theta}_1$ is better (i.e., more efficient) than $\hat{\theta}_2$ if $\sigma_{\hat{\theta}_1} \leq \sigma_{\hat{\theta}_2}$.

  (*Equivalently, use* $\hat{\theta}_1$ *if* $Var(\hat{\theta}_1) \leq Var(\hat{\theta}_2)$)

Example 2.2 (cont'd) If the breakdown voltage $X \sim N(\mu, \sigma^2)$, $\hat{\mu}'s$ s.e.:

Example 2.1 (cont'd) $\hat{p}'s$ s.e.:

Minimum Variance Unbiased Estimator (MVUE)

- **definition:**
  An estimator $\hat{\theta}$ is called **MVUE** if $\hat{\theta}$ is unbiased and $Var(\hat{\theta}) \leq Var(\hat{\theta}^*)$ for any unbiased estimator $\hat{\theta}^*$ of θ.

- If available, use the MVUE of θ to estimate θ.

- *Proposition*: If $X_1, \ldots, X_n$ are iid observations (a random sample) from $N(\mu, \sigma^2)$, then $\hat{\mu} = \bar{X}$ is the MVUE of μ.

Remark: in general, one wants to minimize $\hat{\theta} - \theta$.

An Often-Used Approach: to minimize **mean squared error** (MSE)

$$E(\hat{\theta} - \theta)^2 = Var(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2$$

- If $\hat{\theta}$ is unbiased?
  $\Longrightarrow$ to check only its variance.

- If $\hat{\theta}$ is biased: the bias $= E(\hat{\theta}) - \theta$?

How to construct good estimators?

§2.1.2 Methods of point estimation

§2.1.2A The method of moments estimation (MME)

Recall sample mean $\bar{X}$ to estimate population mean μ: how about to extend the idea to estimating kth population moment, with k an integer (eg, $k = 2$)?

Estimation for population moments: Suppose $X \sim F(\cdot; \theta_1, \dots, \theta_m)$ and iid observations $X_1, \dots, X_n$.

- kth population moment of X: $\mu_k = E(X^k)$
- kth sample moment with $X_1, \dots, X_n$:

$$\hat{\mu}_k = \frac{1}{n}(X_1^k + \dots + X_n^k)$$

- **Use $\hat{\mu}_k$ to estimate μ_k!** (*unbaised estimator*)

eg, $\mu_2 = E(X^2)$ is estimated by

$$\hat{\mu}_2 = \frac{1}{n}(X_1^2 + \dots + X_n^2).$$

Further, what if $X \sim F(\cdot; \theta_1, \ldots, \theta_m)$ with $\theta_1, \ldots, \theta_m$ not all population moments? For example,

$$X \sim N(\mu, \sigma^2): \quad \theta_1 = \mu; \theta_2 = \sigma^2.$$

How to estimate μ and σ^2?

Recall that

$$\mu_2 = E(X^2) = \sigma^2 + \mu^2 = \theta_2 + \theta_1^2$$

How about use the following?

$$\begin{cases} \hat{\mu}_1 = \bar{X} \text{ to estimate } \mu_1 = \mu; \\ \hat{\mu}_2 \text{ to estimate } \sigma^2 + \mu^2 \end{cases}$$

If so, then

$$\begin{cases} \hat{\mu}_1 = \bar{X} \text{ as } \hat{\mu}, \\ \hat{\sigma}^2 = \hat{\mu}_2 - \bar{X}^2 \text{ to estimate } \sigma^2 \end{cases}$$

MME Procedure:

- $X_1, \ldots, X_n$ are iid observations from the population $X \sim F(\cdot; \theta_1, \ldots, \theta_m)$.
- Denote the kth population mean μ_k by $\mu_k = \mu_k(\theta_1, \ldots, \theta_m)$ with $k = 1, \ldots$.
- The **MME** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are the solution to the equations jointly:

$$\begin{cases} \hat{\mu}_1 = \mu_1(\theta_1, \ldots, \theta_m), \\ \ldots\ldots \\ \hat{\mu}_m = \mu_m(\theta_1, \ldots, \theta_m) \end{cases}$$

Revisit to the example of estimating μ and σ^2 with $X \sim N(\mu, \sigma^2)$:
Solve the equations

$$\begin{cases} \bar{X} = \mu, \\ \hat{\mu}_2 = \sigma^2 + \mu^2 \end{cases}$$

and obttain

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \hat{\mu}_2 - \bar{X}^2.$$

Remarks:

- $\hat{\sigma}^2$ *is* $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$.
  $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$, approximately unbiased as $n \to \infty$.

Example 2.4 (p265) Let $X_1, \ldots, X_n$ be a random sample from the population with distribution $Gamma(\alpha, \beta)$.

§2.1.2 Methods of point estimation

§2.1.2B Maximum Likelihood Estimation (MLE)

by R.A. Fisher (geneticist and statistician), 1920

Likelihood Function:

- Let the joint distribution (pmf, or pdf) of r.v.s. $X_1, \ldots, X_n$ be $p(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$.
  When $x_1, \ldots, x_n$ are the observed values (realizations) of the r.v.s., the **likelihood function** of $\theta_1, \ldots, \theta_m$ given the data is

$$L(\theta_1, \ldots, \theta_m \mid \text{ data }) = f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

- **interpretation**: a measure on how likely the observed sample is overall with the values of $\theta_1, \ldots, \theta_m$.

- Often $X_1, \ldots, X_n$ are iid observations (a random sample) from the population with distribution $f(x;\theta)$. If the observed values are $x_1, \ldots, x_n$, then the likelihood function is

$$L(\theta \mid \text{ data }) = \prod_{i=1}^{n} f(x_i;\theta) = f(x_1;\theta) \ldots f(x_n;\theta).$$

For example, iid $X_1, \ldots, X_{100} \sim N(\mu, \sigma^2)$ with observed values $x_1, \ldots, x_{100}$. The likelihood function of μ, σ^2 is

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{100} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ -\frac{1}{2\sigma^2}(x_i - \mu)^2 \Big\} \\ &= \Big(\frac{1}{\sqrt{2\pi\sigma^2}} \Big)^{100} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{100} (x_i - \mu)^2 \Big\}. \end{aligned}$$

Maximum Likelihood Estimator (MLE):

- The **MLE** $\hat{\theta}_1, \ldots, \hat{\theta}_m$ are the values of $\theta_1, \ldots, \theta_m$ that maximize the likelihood function:

$$L(\hat{\theta}_1, \ldots, \hat{\theta}_m \mid \text{ data }) = \max L(\theta_1, \ldots, \theta_m \mid \text{ data }).$$

- **interpretation**: The MLE $\hat{\theta}_1, \ldots, \hat{\theta}_m$ give the parameter values that agree most closely with the observed sample (the data).
- Often used **procedures**: (*Why?*)
(1) to maximize $\ln L(\theta_1, \ldots, \theta_m)$
(2) to obtain the solution to

$$\begin{cases} \frac{\partial \ln L(\theta_1, \ldots, \theta_m)}{\partial \theta_1} = 0, \\ \ldots\ldots \\ \ldots\ldots \\ \frac{\partial \ln L(\theta_1, \ldots, \theta_m)}{\partial \theta_m} = 0 \end{cases}$$

Example 2.5 (textbook p266)

- **Study.** To examine the quality of a new bike helmet: a sample of 10 new helmets are tested and the 1st, 3rd and 10th are found flawed.

- **Statistical Formulation.** $X = 1$ if a helmet is flawed; $X = 0$, otherwise: $X \sim B(1, p)$, the Bernoulli distn with
  $p = P(\text{flawed helmet})$
  To estm p with a random sample $X_1, \ldots, X_{10}$?

- **MME.** Recall $p = P(X = 1)$ and $E(X) = \mu_1(p) = p$, the population mean. The MME is

  $$\text{to solve the equation: } \bar{X} = \mu_1(p), \text{ with respect to } p.$$

  Thus MME $\hat{p} = \bar{X}$, and $\hat{p}_{obs} = 3/10$.

- **MLE.** Recall $f(x; p) = p^x(1-p)^{1-x}$. The likelihood function of p is

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}.$$

With the current data, $L(p) = p^3(1-p)^7$.

$$\ln L(p) = (\sum_i x_i)\ln p + (n - \sum_i x_i)\ln(1-p),$$

$$\frac{d\ln L(p)}{dp} = \frac{\sum_i x_i}{p} - \frac{n-\sum_i x_i}{1-p} = 0.$$

The MLE $\hat{p} = \sum_i X_i \big/ 10$. With the current data, $\hat{p}_{obs} = 3/10$.

The MLE and MME are the same.

What will we do next?

Part 1. Introduction and Review (Chp 1-5)

Part 2. Basic Statistical Inference (Chp 6-9)

- **2.1 Point Estimation**
  - *2.1.1 Some General Concepts*
  - **2.1.2 Methods of Point Estimation**
- **2.2 Confidence Interval**
- 2.3 One-Sample Test
- 2.4 Inference Based on Two-Samples

Part 3. Important Topics in Statistics (Chp 10-13)

Part 4. Further Topics (Selected from Chp 14-16)

Homework 2 is due on Monday 5:00pm.