MAT2377

Catalin Rada

Version May 26, 2009

Lectures Covering 7-1; 7-2; 7-3

Point estimation

Statistical inference consists of methods used to make conclusions about a **population** based on a **random sample**. What is a random sample?

DEF: The rv $X_1, X_2, \ldots, X_n$ is called a random sample if the $X_1, X_2, \ldots, X_n$ are independent rv, AND each X_i has the same p.d.f. **Terminology:** independent and identically distributed (i.i.d.)

In particular, we want to estimate an unknown **parameter**, say θ, using a single number called **point estimate**. Examples of such parameters are:

— the mean μ of a population;

— the variance σ^2 of a population;

— the proportion p of items in a population that belong a certain class of interest.

This point estimate is obtained using a **statistic**, which is simply a function of a random sample ($\hat{\Theta} = h(X_1, \ldots, X_n)$). An example: Suppose that from a population we selected a random sample X_1, X_2, X_3, X_4. Then a statistic is $h(X_1, X_2, X_3, X_4) = \frac{X_2+X_3}{67}$.

Any statistic is a rv! The observed value of the rv $\hat{\Theta}$ is $\hat{\theta} = h(x_1, x_2, \ldots, x_n)$ and it is called a point estimate. (The statistic $\hat{\Theta}$ is called point estimator.)

The probability distribution of a statistics is called a **sampling distribution**.

Example: If we want to estimate parameter μ (the population mean), we may take sample $X_1, \ldots, X_n$ and compute a statistic $\overline{X} = \frac{X_1+X_2+\cdots+X_n}{n}$.

— $\overline{X}$ is a point estimator of the mean μ of population;

— the observed value of $\overline{X}$ (denoted by $\overline{x}$) is a point estimate of μ.

NOTE: If we consider another random sample, we may get different values for $X_1, X_2, \ldots, X_n$, and $\overline{X}$ may change from sample to sample!

EXP: Given the following numbers (representing the life time of CDs exposed to gas): 4, 87, 134, 45, 59 find a point estimate of the mean of lifetime of CDs exposed to gas! SOL: We just need to compute $\overline{x} = \frac{4+87+134+45+59}{5} = 65.8$.

Another statistic: sample variance is given by $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{(\sum_{i=1}^{n} X_i^2) - n\overline{X}^2}{n-1}$.

— S^2 is a point estimator of the variance σ^2 of population;

— the observed value of S^2 (denoted by s^2) is a point estimate of σ^2.

One more statistic: the sample proportion $\hat{P} = \frac{X}{n}$, where X is the number of items in the sample $X_1, X_2, \ldots, X_n$ that belong to a certain class of interest.

— $\hat{P} = \frac{X}{n}$ is a point estimator of the proportion p of items in population that belong a certain class of interest;

— the observed value of $\frac{x}{n}$ (denoted by $\hat{p}$) is a point estimate of p.

Central Limit Theorem

If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population with mean μ and finite variance σ^2 let $\overline{X}$ be the sample mean. Then the limiting form of the distribution of the rv $Z_n = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$, as $n \to \infty$ is the standard normal distribution.

In other words: $\lim_{n\to\infty} F_{Z_n}(x) = \Phi(x)$. For us $n > 30$ it is sufficiently large!!! So we can apply the theorem (it is an approximation theorem).

Applications: 1 The records of the Ministry of Health from planet MathematiX show that the mean of medical expenses of a student during 2089 is 5000 dollars, and the standard deviation is 800 dollars. Compute the probability that the mean of medical expenses of 64 students picked at random is:

i) more than 4820 dollars; ii) between 4800 and 5120 dolars.

Sol: i) So $\mu = 5000$, $\sigma = 800$, $n = 64 > 30$. We need to compute $P(\overline{X} > 4820)$. By the CLT we get: $P(\overline{X} > 4820) = P(\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{64}}} > \frac{4820-5000}{\frac{800}{\sqrt{64}}}) = P(Z > -1.8) = 1 - P(Z \leq -1.8) \cong 1 - \Phi(-1.8) = 1 - 0.035930 = 0.96407$;

ii) We need to compute $P(4800 < \overline{X} < 5120) = P(\frac{4800-5000}{\frac{800}{\sqrt{64}}} < \frac{\overline{X}-5000}{\frac{800}{\sqrt{64}}} < \frac{5120-5000}{\frac{800}{\sqrt{64}}}) = P(-2 < Z < 1.2) \cong \Phi(1.2) - \Phi(-2) = 0.884930 - 0.022750 = 0.86218$.

2 Suppose that the amount of time a student spends watching soccer is a random variable with mean 8.2 minutes and standard deviation 1.5 minutes. Assume a random sample of $n = 49$ students is observed. Compute the probability that the average time of soccer watching for these students is:

a) less than 10 minutes; b) between 5 and 10 minutes; c) less than 6 minutes.

Sol: a) The mean is 8.2, the standard deviation is 1.5, and $n = 49 > 30$. So again CLT! We compute $P(\overline{X} < 10) = P(\frac{\overline{X}-\mu}{\frac{1.5}{\sqrt{49}}} < \frac{10-8.2}{\frac{1.5}{\sqrt{49}}}) = P(Z < \frac{1.8}{0.2143}) = P(Z < 8.4) = \Phi(8.4) = 1$ (look in the tables, 8.4 is not there...) b) We need to compute $P(5 < \overline{X} < 10) = P(\frac{5-8.2}{\frac{\sigma}{\sqrt{49}}} < \frac{\overline{X}-8.2}{\frac{\sigma}{\sqrt{49}}} < \frac{10-8.2}{\frac{\sigma}{\sqrt{49}}}) = P(\frac{-3.2}{0.2143} < Z < \frac{1.8}{0.2143}) \cong \Phi(8.4) - \Phi(-14.93) = 1 - 0$ (again -14.93 is not in the table!); c) We need $P(\overline{X} < 6) = P(\frac{\overline{X}-\mu}{\frac{1.5}{\sqrt{49}}} < \frac{6-8.2}{\frac{1.5}{\sqrt{49}}}) = P(Z < \frac{-2.2}{0.2143}) = P(Z < -10.26598) \cong \Phi(-10.26598) = 0$.

Question: What if we have 2 independent populations? Say that the mean of the first population is μ_1, and the variance of the first population is σ_1^2, and say that the mean of the second population is μ_2, and the variance of the second population is σ_2^2. Pick a random sample (of size n_1) from the

first population, and pick a random sample (of size n_2) from the second population.

Then $\overline{X_1}$ and $\overline{X_2}$ follow normal distributions (by CLT!). So, the distribution of $\overline{X_1} - \overline{X_2}$ is approximately normal with mean and variance:

$$\mu_{\overline{X_1}-\overline{X_2}} = \mu_{\overline{X_1}} - \mu_{\overline{X_2}} = \mu_1 - \mu_2;$$

$$\sigma^2_{\overline{X_1}-\overline{X_2}} = \sigma^2_{\overline{X_1}} + \sigma^2_{\overline{X_2}} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$ In other words: $\frac{\overline{X_1}-\overline{X_2}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$ is approximately standard normal (if the conditions in CLT apply, i.e., n_1 and n_2 are greater than 30) In the case the 2 populations are normal, then Z is exactly standard normal! (of course: as we have seen — if the 2 populations are normal, then $\overline{X_1}$ and $\overline{X_2}$ are normal)

EXC: Suppose that the amount of time a boy spends watching soccer is normaly distributed with mean 5000 hours and with standard deviation 40

hours. Suppose that the amount of time a girl spends watching soccer is normaly distributed with mean 5050 hours and with standard deviation 30 hours. Let X_1 be the average of time spent by 20 boys watching soccer, Let X_2 be the average of time spent by 15 girls watching soccer. Compute $P(X_1 - X_2 < -10)$.

Sol: $P(X_1 - X_2 < -10) = P(\frac{X_1 - X_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{40^2}{20} + \frac{30^2}{15}}}) < \frac{-10 - (\mu_1 - \mu_2)}{\sqrt{\frac{40^2}{20} + \frac{30^2}{15}}}) = P(Z < \frac{40}{\sqrt{140}}) = P(Z < 3.38) = 0.999638.$

Bias of an Estimator

The point estimator $\hat{\Theta}$ of θ is **unbiased** if

$$\mathrm{E}[\hat{\Theta}] = \theta.$$

The bias is $\mathrm{E}[\hat{\Theta}] - \theta$. When the estimator is unbiased, the bias is zero.

Example: A sample mean $\overline{X}$ is an unbiased estimator of the population mean μ, since $\mathrm{E}[\overline{X}] = \mu$.

EXC/Example: Show that the sample variance $S^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} = \frac{(\sum_{i=1}^{n} X_i^2) - n\overline{X}^2}{n-1}$ is an unbiased estimator of the variance of population σ^2.

Sol: $E(S^2) = E(\frac{(\sum_{i=1}^{n} X_i^2) - n\overline{X}^2}{n-1}) = \frac{1}{n-1}\{\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\} = \frac{1}{n-1}\{n(\sigma^2 + \mu^2) - n(\mu^2 + \sigma^2/n)\} = \sigma^2$.

Variance of an estimator and standard error

Example: If data $X_1, \ldots, X_n$ come from a population with unknown mean μ and known variance σ^2, then $\mathrm{Var}(\overline{X}) = \sigma^2/n$. Thus

$$\text{Standard error :} \qquad \sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}.$$

If the variance is unknown σ^2, then

$$\text{Estimated Standard error :} \qquad \hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}},$$

where S^2 is sample variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

The mean squared error

$$\mathrm{MSE}(\hat{\Theta}) = \mathrm{E}(\hat{\Theta} - \theta)^2.$$

We have

$$\begin{aligned} \mathrm{E}(\hat{\Theta} - \theta)^2 &= \mathrm{E}(\hat{\Theta} - \mathrm{E}(\hat{\Theta}))^2 + (\theta - \mathrm{E}(\hat{\Theta}))^2 \\ &= \mathrm{Var}(\hat{\Theta}) + (\mathrm{bias})^2. \end{aligned}$$

The mean squared error is an important tool to check, which estimator is *more efficient*. If $\mathrm{MSE}(\hat{\Theta}_1)$ and $\mathrm{MSE}(\hat{\Theta}_2)$ are mean squared errors of estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$ (of the same parameter θ), then the relative efficiency is

$$\frac{\mathrm{MSE}(\hat{\Theta}_1)}{\mathrm{MSE}(\hat{\Theta}_2)}.$$

If this is less than 1, we can conclude that $\hat{\Theta}_1$ is more efficient. In particular, if we consider unbiased estimators only, computing the relative frequency is equivalent to comparison of variances. WHAT ESTIMATOR IS MORE EFFECTIVE?

Typically, there are many unbiased estimators. If we have two unbiased estimators, we choose the one with a smaller variance. WHAT ESTIMATOR IS BETTER? DEF: If we consider all unbiased estimators of a parameter θ, the one with the smallest variance is called MINIMUM VARIANCE UNBIASED ESTIMATOR (MVUE). As an example: if the rv $X_1, X_2, \ldots, X_n$ form a random sample of size n from normal distribution (mean μ and variance σ^2), then $\overline{X}$ is the MVUE for μ.

EXC: 7-15; 7-16 (answer: $\hat{\Theta}_2$); 7-17 (try it home) PLAN: 1) DECIDE WHAT ARE THE BIASED ESTIMATORS AND WHAT ARE THE UNBIASED ESTIMATORS; 2) COMPUTE THE BIAS (IF APPLICABLE); 3) FOR UNBIASEDNESS: CHOOSE THE UNBIASED ESTIMATOR WITH

THE SMALLEST VARIANCE — SO YOU'LL GET the better ESTIMATOR; 4) COMPUTE Relative Efficiencies AND THEN DECIDE THEN WHICH ONE IS THE MOST EFFICIENT (among all of them)!

EXC 15. a) $E(\hat{\Theta}_1) = E(\overline{X}) = \frac{1}{7}E(X_1 + \cdots + X_7) = \frac{1}{7}\{E(X_1) + \cdots + E(X_7)\} = \frac{1}{7}\{\mu + \cdots + \mu\} = \mu$, so $\hat{\Theta}_1$ is unbiased. Now we have $E(\hat{\Theta}_2) = E(\frac{2X_1 - X_6 + X_4}{2}) = E(X_1 - (1/2)X_6 + (1/2)X_4) = 1 \times \mu + (-1/2) \times \mu + (1/2) \times \mu = \mu$, so $\hat{\Theta}_2$ is unbiased. b) We just need to compute the variances (since bias $= 0$ for both of them): $Var(\hat{\Theta}_1) = Var(\overline{X}) = Var(\frac{1}{7}X_1 + \cdots + \frac{1}{7}X_7) = (\frac{1}{7})^2 \times \sigma^2 + \cdots + (\frac{1}{7})^2 \times \sigma^2 = \frac{7}{7^2} \times \sigma^2 = \frac{\sigma^2}{7}$. For the other estimator we do have: $Var(\hat{\Theta}_2) = Var(X_1 + (-1/2)X_6 + (1/2)X_4) = 1^2 \times \sigma^2 + (-1/2)^2 \times \sigma^2 + (1/2)^2 \times \sigma^2 = \sigma^2\{1 + \frac{1}{4} + \frac{1}{4}\} = \sigma^2\frac{3}{2}$. Since $\sigma^2\frac{3}{2} > \frac{\sigma^2}{7}$ we decide that $\hat{\Theta}_1$ is a better estimator!

EXC 16. Since $4 < 10$ we say $\hat{\Theta}_2$ is better than $\hat{\Theta}_1$. We use only the variance when the estimators are unbiased (see the Plan)! Of course

$\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \frac{10}{4} = 2.5 > 1.$

EXC 17. Since $\hat{\Theta}_2$ is biased and $\hat{\Theta}_1$ is unbiased we use the relative efficiency. So we compute $\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \frac{\text{Var}(\hat{\Theta}_1)+(\text{bias})^2}{\text{Var}(\hat{\Theta}_2)+(\text{bias})^2} = \frac{\text{Var}(\hat{\Theta}_1)+0}{\text{Var}(\hat{\Theta}_2)+(\text{bias})^2} = \frac{10}{4+(\text{bias})^2}$, where $(\text{bias})^2 = (\frac{\theta}{2} - \theta)^2 = \frac{\theta^2}{4}$; hence $\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} = \frac{10}{4+\frac{\theta^2}{4}}$. This fraction is ≤ 1 if and only if $\theta \in (-\infty, -\sqrt{24}] \cup [\sqrt{24}, \infty)$. So, in this case $\hat{\Theta}_1$ is more effective! Of course if $\theta \in (-\sqrt{24}, \sqrt{24})$, then $\hat{\Theta}_2$ is more effective!