Math 408 - Mathematical Statistics

Lecture 18. Estimation of a Ratio and the δ-method

March 4, 2013

Ratio and its Estimate

Suppose that for each member of a population, two values are measured:

$$i^{\text{th}} \text{ member} \rightsquigarrow (x_i, y_i)$$

We are interested in the following **ratio**:

$$\boxed{r = \frac{\sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} x_i}}$$

Ratios arise frequently in sample surveys.

Example:

Households are sampled. If y_i is the number of unemployed males in the i^{th} household, and x_i is the total number of males in the i^{th} household, then r is the proportion of unemployed males.

Estimate of a Ratio

Let $\begin{pmatrix} X_1 & \dots & X_n \\ Y_1 & \dots & Y_n \end{pmatrix}$ be a sample from a population.
Then the natural estimate of

$$r = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} = \frac{\frac{1}{N}\sum_{i=1}^N y_i}{\frac{1}{N}\sum_{i=1}^N x_i} = \frac{\mu_y}{\mu_x}$$

is

$$\boxed{R_n = \frac{\overline{Y}_n}{\overline{X}_n}}$$

Our goal: to derive expressions for $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$

Technical problem: since R_n a nonlinear function of $\overline{X}_n$ and $\overline{Y}_n$, we can't find $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$ in closed form.

Idea: To approximate $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$ using the δ-**method**.

The δ-method

In many applications, the following scenario is typical:

Problem

X is a random variable, μ_x and σ_X^2 are known. The problem is to find the mean and variance of $Y = f(X)$, where f is some (typically nonlinear) function.

The δ-**method** utilizes a strategy that is often used in applied mathematics: when confronted with a nonlinear problem that we can't solve, we linearize.

In the δ-method, the linearization is carried out through a Taylor series expansion of f about μ_X:

$$Y = f(X) \approx f(\mu_X) + (X - \mu_X) f'(\mu_X)$$

We thus obtain the first order approximations:

$$\boxed{\mu_Y \approx f(\mu_X)} \qquad \boxed{\sigma_Y^2 \approx (f'(\mu_X))^2 \sigma_X^2}$$

The δ-method

To obtain a better approximation for μ_Y, we can use the Taylor series expansion to the 2^{nd} order:

$$Y = f(X) \approx f(\mu_X) + (X - \mu_X)f'(\mu_X) + \frac{1}{2}(X - \mu_X)^2 f''(\mu_X)$$

Then the second order approximations for μ_Y is

$$\boxed{\mu_Y \approx f(\mu_X) + \frac{1}{2}\sigma_X^2 f''(\mu_X)}$$

We can similarly proceed in the case of two random variables X and Y:

Problem

Suppose that $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \sigma_{XY} = Cov(X, Y)$ are known.
The problem is to find μ_Z and σ_Z^2, where $Z = f(X, Y)$.

The δ-method

Using the Taylor series expansion to the first order:

$$Z = f(X,Y) \approx f(\mu) + (X-\mu_X)\frac{\partial f}{\partial x}(\mu) + (Y-\mu_Y)\frac{\partial f}{\partial y}(\mu), \qquad \mu = (\mu_X, \mu_Y)$$

Therefore,

$$\boxed{\mu_Z \approx f(\mu)} \qquad \boxed{\sigma_Z^2 \approx \sigma_X^2 \left(\frac{\partial f}{\partial x}(\mu)\right)^2 + \sigma_Y^2 \left(\frac{\partial f}{\partial y}(\mu)\right)^2 + 2\sigma_{XY}\frac{\partial f}{\partial x}(\mu)\frac{\partial f}{\partial y}(\mu)}$$

To obtain a better approximation for μ_Z, we can use the Taylor series expansion to the second order.

$$\boxed{\mu_Z \approx f(\mu) + \frac{1}{2}\sigma_X^2 \frac{\partial^2 f}{\partial x^2}(\mu) + \frac{1}{2}\sigma_Y^2 \frac{\partial^2 f}{\partial y^2}(\mu) + \sigma_{XY}\frac{\partial^2 f}{\partial x \partial y}(\mu)}$$

The δ-method: special case $Z = Y/X$

Example

If $Z = Y/X$, *then*

$$\boxed{\mu_Z \approx \frac{\mu_Y}{\mu_X} + \frac{1}{\mu_X^2}\left(\sigma_X^2 \frac{\mu_Y}{\mu_X} - \sigma_{XY}\right)} \tag{1}$$

$$\boxed{\sigma_Z^2 \approx \frac{1}{\mu_X^2}\left(\sigma_X^2 \frac{\mu_Y^2}{\mu_X^2} + \sigma_Y^2 - 2\sigma_{XY}\frac{\mu_Y}{\mu_X}\right)} \tag{2}$$

Approximations of $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$

The estimate of $r = \frac{\mu_y}{\mu_x}$ is

$$R_n = \frac{\overline{Y}_n}{\overline{X}_n}$$

To use the δ-method to approximate $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$, we need to know $\mu_{\overline{X}_n}, \mu_{\overline{Y}_n}, \sigma^2_{\overline{X}_n}, \sigma^2_{\overline{Y}_n}$, and $Cov(\overline{X}_n, \overline{Y}_n)$. In previous Lectures, we found that

- $\mu_{\overline{X}_n} = \mu_x$
- $\mu_{\overline{Y}_n} = \mu_y$
- $\sigma^2_{\overline{X}_n} = \frac{\sigma_x^2}{n}\left(1 - \frac{n-1}{N-1}\right)$
- $\sigma^2_{\overline{Y}_n} = \frac{\sigma_y^2}{n}\left(1 - \frac{n-1}{N-1}\right)$

It can be shown that

- $Cov(\overline{X}_n, \overline{Y}_n) = \frac{\sigma_{xy}}{n}\left(1 - \frac{n-1}{N-1}\right)$, where σ_{xy} is the population covariance of x and y, $\sigma_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)$.

Approximations of $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$

Using approximations (1) and (2) from the δ-method, we obtain

Theorem

The expectation and variance of R_n are given by

$$\boxed{\mathbb{E}[R_n] \approx r + \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r\sigma_x^2 - \sigma_{xy})} \tag{3}$$

$$\boxed{\mathbb{V}[R_n] \approx \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r^2\sigma_x^2 + \sigma_y^2 - 2r\sigma_{xy})} \tag{4}$$

In applications, population parameters $\mu_x, \sigma_x, \sigma_y, \sigma_{xy}$ are unknown. To compute the **estimated** values of $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$, we use (3) and (4) together with

- $r \approx R_n \qquad \mu_x \approx \overline{X}_n$
- $\sigma_x^2 \approx \hat{\sigma}^2_{x,\text{unbiased}} = \frac{N-1}{Nn-N}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
- $\sigma_y^2 \approx \hat{\sigma}^2_{y,\text{unbiased}} = \frac{N-1}{Nn-N}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$
- $\sigma_{xy} \approx \frac{N-1}{Nn-N}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$

Summary

- Ratios $r = \mu_y/\mu_x$ arise frequently in sample surveys
- The natural estimate of r is $R_n = \overline{Y}_n/\overline{X}_n$
- We can find expressions for $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$ using the δ-method:

$$\boxed{\mathbb{E}[R_n] \approx r + \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r\sigma_x^2 - \sigma_{xy})}$$

$$\boxed{\mathbb{V}[R_n] \approx \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_x^2}(r^2\sigma_x^2 + \sigma_y^2 - 2r\sigma_{xy})}$$

- To compute the estimated values of $\mathbb{E}[R_n]$ and $\mathbb{V}[R_n]$, we use:
  - $r \approx R_n \qquad \mu_x \approx \overline{X}_n$
  - $\sigma_x^2 \approx \hat{\sigma}^2_{x,\text{unbiased}} = \frac{N-1}{Nn-N}\sum_{i=1}^n (X_i - \overline{X}_n)^2$
  - $\sigma_y^2 \approx \hat{\sigma}^2_{y,\text{unbiased}} = \frac{N-1}{Nn-N}\sum_{i=1}^n (Y_i - \overline{Y}_n)^2$
  - $\sigma_{xy} \approx \frac{N-1}{Nn-N}\sum_{i=1}^n (X_i - \overline{X}_n)(Y_i - \overline{Y}_n)$