Jan 23rd

Ratio Estimation

$\frac{\mu_y}{\mu_x} = r$ we use $\frac{\bar{Y}}{\bar{X}} = R$ to estimate r

Application: Use R to estimate μ_y

An estimate of μ_y is $\mu_x R$

$\hat{Y} = \mu_x R$

$\downarrow$

$\hat{Y}$ estimates μ_y

According to last lecture notes

$$E\hat{Y} \approx \mu_y + \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{\mu_x}\left(r\sigma_x^2 - \rho\sigma_x\sigma_y\right)$$

σ , σ "sigma"

$$Var(\hat{Y}) \approx \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\left(\mu^2\sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y\right)$$

$$Var(\bar{Y}) = \frac{\sigma_y^2}{n}\left(\frac{N-n}{N-1}\right)$$ (lecture 1)

We know that $\bar{Y} = \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)$ is also an estimate of μ_y

$\hat{Y}$ and $\bar{Y}$ which one better?

$Var(\hat{Y}) < Var(\bar{Y})$

$\Leftrightarrow \frac{1}{n}\left(\frac{N-n}{N-1}\right)\left(r^2\sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y\right) < \frac{1}{n}\left(\frac{N-n}{N-1}\right)\sigma_y^2$

$\Leftrightarrow r^2\sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y < \sigma_y^2$

$\Leftrightarrow r^2\sigma_x^2 < 2r\rho\sigma_x\sigma_y$ Now assume that $r > 0$

$\Leftrightarrow r < 2\rho\frac{\sigma_y}{\sigma_x}$

Hence $\hat{Y}$ reduces the variance of $\bar{Y}$ if

$r < 2\rho\frac{\sigma_y}{\sigma_x}$

or equivalently $\rho > \frac{1}{2} r \frac{\sigma_x}{\sigma_y}$

EX: There are 393 hospitals we want to estimate the average number of beds occupied in each hospital.

$\bar{X} = 274.8$

$\bar{Y} = 814.6$

$r = 2.96$

$\sigma_x = 213.2$

$\sigma_y = 589.7$

$\rho = 0.91$

estimate of μ_y: $\sigma_{\hat{Y}} = 30.0$, $\sigma_{\bar{Y}} \approx 66.3$

Chapter 7 is done. §7.5 is not required.

Chp 8. Estimation of Parameters and fitting of probability distributions

THE Method of Moment

The method of Moment

Review of some distributions

Normal Distribution (Gaussian) $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$ $x \in R$

$N(\mu, \sigma^2)$

Gamma Distribution Gamma(α, λ) $f(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda}$, $x > 0$

Example 1: Suppose we observe the # of calls per minute for 10 different 1-minute slots, And the observed # of calls are

$\{0, 2, 5, 3, 8, 3, 4, 3, 6, 2\}$

Guess: in the population, what is the # of calls per minute?

Example (b): Sps we observe 10 observations from a uniform (a, b) population which are $\{0.11, 0.35, 1.2, 1.5, 0.06, 0.72, 0.42, 1.11, 0.32, 1.33\}$

What is your best guess for a & b?

(a). One obvious solution: Take the sample mean.

(b). One solution: $\hat{a}$ = minimum, $\hat{b}$ = maximum   Is it the best way?

First, let's define

$m_k = E[X^k]$. Here X is a random variable of interest

Let's assume that X has a CDF $F(x)$ which depends on unknown parameter(s) θ.

↳ may be multivariate

In this case we write $F(X|\theta)$ to emphasize the dependence of F on θ.

Now suppose $m_i = f_i(\theta)$ ①

Then under some conditions θ can be solved from ①

Let's say $\theta = G(m_1, m_2, \dots, m_p)$

Finally, since $\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$ is a good estimate of m_k for each k.

Weak Law of large numbers

Suppose $Z_1, Z_2, \dots, Z_n$ are iid random variables

$E|Z_i| < \infty$

Then $\frac{1}{n}\sum_{i=1}^{n} Z_i \longrightarrow E[Z_i]$

We have $\hat{\theta}_{mm} = G(\hat{m}_1, \hat{m}_2, \dots \hat{m}_p)$ is a good estimate of θ. each k

Suppose X *is the # of calls/min*
$X \sim Poisson(\lambda)$ $\qquad P(X=k) = \frac{\lambda^k}{K!} e^{-\lambda}$

$m_1 = \lambda$
$\hat{\lambda}_{mm} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}[\cdots \quad] = ?$

$Var(\hat{\lambda}_{mm}) = Var(\frac{1}{n}\sum x_i) = \frac{1}{n^2}\sum Var(x_i) = \frac{\lambda}{n}$

$E[\lambda_{mm}] = \frac{1}{n}\sum E[X_i] = \lambda$

Based on the discussion above a 95% Confidence Interval for λ is
$$\hat{\lambda}_{mm} \pm 1.96\sqrt{\frac{\hat{\lambda}_{mm}}{n}}$$

Example 2: Sps $X \sim uniform(a, b)$
$m_1 = \frac{a+b}{2}$
$$m_2 = E[x^2] = \int_a^b x^2 \frac{1}{b-a}\,dx = \frac{1}{3}x^3\Big|_a^b = \frac{1}{3}\frac{b^3-a^3}{b-a} = \frac{b^2+a^2+ab}{3} \quad (*)$$

$\Rightarrow a+b = 2m_1$
$\Rightarrow b = 2m_1 - a$
Plug in to $(*)$ we have
$3m_2 = a^2 + a(2m_1 - a) + (2m_1 - a)^2 = 2m_1 a + 4m_1^2 - 4m_1 a + a^2$
$= 4m_1^2 - 2m_1 a + a^2$
$a^2 - 2m_1 a + 4m_1^2 - 3m_2 = 0$
$a = \frac{2m_1 \pm \sqrt{4m_1^2 - 4(4m_1^2 - 3m_2)}}{2} = m_1 \pm \sqrt{3m_2 - 3m_1^2}$
$= m_1 \pm \sqrt{3}\cdot\sqrt{m_2 - m_1^2}$

Since $a + b = 2m_1$
If $a = m_1 + \sqrt{3}\cdot\sqrt{m_1 - m_2^2}$ Then
$b = m_1 - \sqrt{3}\cdot\sqrt{m_1 - m_2^2}$
$\because b > a$ for a uniform distri. $\Rightarrow$ This is not acceptable
There $\hat{a}_{mm} = m_1 - \sqrt{3}\cdot\sqrt{m_2 - m_1^2}$
$\hat{b}_{mm} = m_1 + \sqrt{3}\cdot\sqrt{m_2 - m_1^2}$

Example 3: *Normal distribution*
If $X \sim N(\mu, \sigma^2)$ We observe $X_1, X_2, \cdots, X_n$
Goal: To estimate μ and σ^2 from $X_1, \cdots, X_n$
$m_1 = E[X] = \mu$
$m_2 = E[X^2] = \mu^2 + \sigma^2$ $\quad \Rightarrow \quad \begin{cases} \mu = m_1 \\ \sigma^2 = m_2 - m_1^2 \end{cases}$

$\hat{\mu}_{mm} = \hat{m}_1 = \overline{X}$

$$\hat{\sigma}^2_{mm} = \hat{m}_2 - \hat{m}_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Confidence Interval constructing for MM estimate

For the normal case

$\hat{\mu} = \bar{X}$ so a 95% CI for μ is just $\bar{X} \pm 1.96 s/\sqrt{n}$

How about σ^2?

$$\hat{\sigma}^2_{mm} = \frac{1}{n}\sum_{i=1}^{k}(X_i - \bar{X})^2$$

* Bootstrap

① Rationale: If we know μ & σ^2 then to assess the variability (or sampling distribution) of $\hat{\sigma}^2_{mm}$ we could just generate a large repetition of iid $N(\mu, \sigma^2)$ samples of size n.
For each of those samples we can obtain MM estimates of σ^2.
Then we actually obtain many $\hat{\sigma}^2_{mm}$. The key step is to use the variability or sampling distribution of all these $\hat{\sigma}^2_{mm}$'s to approximate $Var(\hat{\sigma}^2_{mm})$ and the distribution of $\hat{\sigma}^2_{mm}$.

② The Rationale is not really applicable since in practice we don't know μ & σ^2. Remedy. Replace μ, and σ^2 by $\hat{\mu}_{mm}$ and $\hat{\sigma}^2_{mm}$ and pretend that those are the time population parameters $\longrightarrow$ This remedy works in many (but not all) cases.

The method of Maximum Likelihood

Likelihood. Def: Suppose that $X_1, X_2, \cdots, X_n$ have a joint density or pmf in the form of $f(X_1, X_2, \cdots X_n|\theta)$

Then $f(X_1, \cdots X_n|\theta)$ at observed values $x_1, x_2, \cdots, x_n$ is called the likelihood when viewed as a function of θ.

Ex: Suppose $x_1, \cdots, x_n$ are from a $\exp(\lambda)$ distribution X's are independent.
Try to write down the likelihood.

$$f(x_1, \cdots x_n|\lambda) = \prod_{i=1}^{n} \frac{1}{\lambda} e^{-x/\lambda} = \frac{1}{\lambda^n} e^{-(\sum_{i=1}^{n} x_i/\lambda)}$$ $\longleftarrow$ viewed as a fn if λ is the likelihood

** The principle of Maximum Likelihood

If we want to estimate a parameter θ then we just find $\hat{\theta}$ s.t. $f(x_1, \cdots, x_n|\theta)$ is maximized at $\hat{\theta}$.

Def: $l(x_1, \cdots, x_n|\theta)$ which we call the log likelihood is equal to $\log f(x_1, \cdots, x_n|\theta)$

In the $\text{Exp}(\lambda)$ case

$$\underline{l(x_1, \cdots, x_n|\theta)} = n\log\lambda - \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

maximize this

$$l'(\theta) = \frac{-n}{\lambda} + \frac{1}{\lambda^2}\sum_{i=1}^{n} x_i = 0 \implies \lambda = \frac{1}{n}\sum_{i=1}^{n} x \qquad \hat{\lambda}_{MLE} = \bar{X}$$

$$Var(\hat{\lambda}_{MLE}) = Var(\overline{X}) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i)$$

$$Var(X_i) = E[X_i^2] - \underbrace{[E[X_i]]^2}_{\lambda^2}$$

$$= \int_0^\infty x^2 \frac{1}{\lambda} e^{-x/\lambda}\, dx \xrightarrow{z=\frac{x}{\lambda}} = \int_0^\infty \lambda^2 z^2 e^{-z} \lambda \cdot dz \cdot \frac{1}{\lambda} = \lambda^2 \int_0^\infty z^2 e^{-z} dz$$

$$= \lambda^2 \Gamma(3) = 2\lambda^2$$