MATH E-156 Mathematical Statistics

Harvard Extension School

Dmitry Kurochkin

Fall 2020
Lecture 10

Contents

Contents

Expectation of $\hat{\sigma}^2$

Thm.

Assuming simple random sampling from a population with the population variance σ^2,

$$\mathrm{E}[\hat{\sigma}^2] = \sigma^2 \left(\frac{n-1}{n}\right) \frac{N}{N-1},$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

Remark

If

$$\mathrm{E}[\hat{\sigma}^2] \neq \sigma^2,$$

we will say that $\hat{\sigma}^2$ is *biased* estimator of σ^2.

Expectation of $s^2_{\bar{X}}$

<u>Corollary</u>

Assuming simple random sampling from a population with the population variance σ^2,

$$s^2_{\bar{X}} = \frac{\hat{\sigma}^2}{n}\left(\frac{n}{n-1}\right)\left(\frac{N-1}{N}\right)\left(\frac{N-n}{N-1}\right)$$

is an unbiased estimate of $\mathsf{Var}(\bar{X})$, that is,

$$\mathrm{E}[s^2_{\bar{X}}] = \mathsf{Var}(\bar{X}).$$

<u>Remark</u>

Note that

$$s^2_{\bar{X}} = \frac{S^2}{n}\Bigg(1 - \underbrace{\frac{n}{N}}_{\text{sampling fraction}}\Bigg), \quad \text{where} \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2.$$

Contents

Estimation of Population Proportions

Corollary

Assume simple random sampling from a population with $x_1, x_2, \ldots, x_N \in \{0, 1\}$ (dichotomous case).

Let

$$\hat{p} = \bar{X}.$$

Then

$$s_{\hat{p}}^2 = \frac{\hat{p}(1-\hat{p})}{n-1}\left(1 - \frac{n}{N}\right)$$

is an unbiased estimate of $\mathsf{Var}(\hat{p})$, that is,

$$\mathrm{E}[s_{\hat{p}}^2] = \mathsf{Var}(\hat{p}).$$

Contents

Normal Approximation to the Sampling Distribution of $\bar{X}$

Assume simple random sampling. If sample size n is large, but small relative to population size N, then

$$\frac{\bar{X}_n - \mu}{s_{\bar{X}}} \quad \text{is approximately} \quad \text{Normal}(0, 1),$$

that is,

$$P\left(\frac{\bar{X}_n - \mu}{s_{\bar{X}}} \leq x\right) \approx \Phi(x) \quad \text{for all} \quad x \in \mathbb{R},$$

where

$$\Phi(z) \doteq \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

is the cdf of a standard normal random variable.

Confidence Interval for the Population Mean μ

Let $z_{\alpha/2}$ denote a number such that $\Phi(z_\alpha) = 1 - \alpha/2$. Then

$$P\left(-z_{\alpha/2} < \frac{\bar{X}_n - \mu}{s_{\bar{X}}} < z_{\alpha/2}\right) \approx 1 - \alpha$$

and therefore

$$P\left(\bar{X}_n - z_{\alpha/2} s_{\bar{X}} < \mu < \bar{X}_n + z_{\alpha/2} s_{\bar{X}}\right) \approx 1 - \alpha.$$

Terminology

The interval $\left(\bar{X}_n - z_{\alpha/2}\, s_{\bar{X}}, \bar{X}_n + z_{\alpha/2}\, s_{\bar{X}}\right)$, denoted by

$$\bar{X}_n \pm z_{\alpha/2}\, s_{\bar{X}},$$

is called $100(1-\alpha)\%$ *confidence interval* for μ.

Contents

Consistency of Parameter Estimators

Def.:

Let $\hat{\theta}_n$ be an estimator of a parameter θ of some distribution based on a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from this distribution.

$\hat{\theta}_n$ is called *consistent* (or *asymptotically consistent*) estimator of θ if for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

Contents

Sample Moments

Def:

Given a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from some distribution,

$$\hat{\mu}_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

is called kth *sample moment.*

Method of Moments (MM)

Method of Moments:
Assume that a distribution can be parametrized by $\theta_1, \theta_2, \ldots, \theta_d$, then given a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from this distribution, the unknown θ_k can be obtained as the solution to the following system:

$$\begin{aligned}
\mathrm{E}[X] &\overset{\text{set}}{=} \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \\
\mathrm{E}[X^2] &\overset{\text{set}}{=} \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \\
&\vdots \\
\mathrm{E}[X^d] &\overset{\text{set}}{=} \hat{\mu}_d = \frac{1}{n}\sum_{i=1}^{n} X_i^d,
\end{aligned}$$

where $\hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_k^i$ is the kth *sample moment.*

Contents

MM for Bernoulli Distribution

Example:

Consider a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from Bernoilli(p), i.e. the pmf is

$$P(X = k|p) = \begin{cases} p^k(1-p)^{1-k}, & \text{if } k \in \{0, 1\}, \\ 0, & \text{otherwise.} \end{cases}$$

Then the Method of Moments estimator of p is

$$\hat{p}_{\text{MM}} = \bar{X}.$$

MM for Exponential Distribution

Example:
Consider a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from Exponential(λ), i.e. the pdf is

$$f(x|\lambda) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0, \\ 0, & \text{otherwise }. \end{cases}$$

Then the Method of Moments estimator of λ is

$$\hat{\lambda}_{\mathsf{MM}} = \frac{1}{\bar{X}}.$$

MM for Normal Distribution

Example:

Consider a sample of i.i.d. $X_1, X_2, \ldots, X_n$ from Normal(μ, σ^2), i.e. the pdf is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R},$$

Then the Method of Moments estimators of μ and σ^2 are

$$\begin{aligned} \hat{\mu}_{\text{MM}} &= \bar{X}, \\ \hat{\sigma}^2_{\text{MM}} &= \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \bar{X}^2. \end{aligned}$$