Math 408 - Mathematical Statistics

Lecture 16. Estimation of the Population Variance σ

February 27, 2013

Agenda

- Why do we need to estimate σ?
- How can we estimate σ?
- Summary

The Need of Estimation of σ

We know that the sample mean $\overline{X}_n$ is an unbiased estimate of the population mean μ:

$$\mathbb{E}[\overline{X}_n] = \mu$$

Moreover, the accuracy of the approximation $\overline{X}_n \approx \mu$ can be measured by the standard deviation of $\overline{X}_n$ (also called "standard error"):

$$\sigma_{\overline{X}_n} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)}, \qquad \sigma_{\overline{X}_n} \approx \frac{\sigma}{\sqrt{n}}, \quad \text{if } n \ll N \tag{1}$$

where σ is the population variance

$$\sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Q: What is the main drawback of (1)?

A: We can't use (1) since σ is unknown.
To use (1), σ must be estimated from the sample $X_1, \ldots, X_n$.

Estimation of σ

It seems natural to use the following estimate

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

However, this estimate is biased.

Theorem

The expected value of $\hat{\sigma}_n^2$ is given by

$$\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2 \frac{Nn - N}{Nn - n}$$

Important Remark:

- Since $\frac{Nn-N}{Nn-n} < 1$, we have $\mathbb{E}[\hat{\sigma}_n^2] < \sigma^2$
  Therefore, $\hat{\sigma}_n^2$ tends to underestimate σ^2

Corollaries

Corollary

Since $\mathbb{E}[\hat{\sigma}_n^2] = \sigma^2 \frac{Nn-N}{Nn-n}$,

$$\hat{\sigma}^2_{n,\text{unbiased}} = \frac{Nn-n}{Nn-N}\hat{\sigma}_n^2$$

is an unbiased estimate of σ^2

Recall that

$$\mathbb{V}[\overline{X}_n] = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

In practice, σ is unknown, so we need to estimate it.

Corollary

An unbiased estimate of $\mathbb{V}[\overline{X}_n]$ *is*

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

Summary

Let us summarize what we have learned about estimation of population parameters:

- **Population mean** μ
  - Unbiased estimate:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

  - Variance of estimate

$$\mathbb{V}[\overline{X}_n] \equiv \sigma^2_{\overline{X}_n} = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$$

  - Estimated variance

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

- **Population variance** σ
  - Unbiased estimate:

$$\hat{\sigma}^2_{n,\text{unbiased}} = \frac{Nn-n}{Nn-N}\hat{\sigma}^2_n, \qquad \hat{\sigma}^2_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

Conclusion

In simple random sampling, we can not only form estimate of unknown population parameter (e.g. μ), but also obtain the likely size of errors of these estimates.In other words, we can obtain the estimate of a parameter as well as the estimate of the error of that estimate