

S1211Q Introduction to Statistics

Lecture 13

Wei Wang

July 23, 2012

CLT

- Theorem:

The Central Limit Theorem (CLT)

Let X_1, X_2, \dots, X_n , be an i.i.d. sequence from a distribution with mean μ and variance σ^2 . Then if n is sufficiently large, the sample mean \bar{X} has approximately a normal distribution with $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}}^2 = \sigma^2/n$; And the sample total has approximately a normal distribution with $\mu_T = n\mu$, $\sigma_T^2 = n\sigma^2$. The larger the value of n , the better the approximation.

- Rule of Thumb: if $n > 30$, the CLT can be used.

Distribution of a Linear Combination

- ▶ Sample mean is a particular case of linear combinations.
- ▶ The expectation and variance of a general linear combination

$$a_1X_1 + a_2X_2 + \dots + a_nX_n$$

is given by the following result.

A key result ***

Let X_1, X_2, \dots, X_n , have mean values $\mu_1, \mu_2, \dots, \mu_n$, respectively, and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$, respectively.

- Whether or not the X_i 's are independent,

$$\begin{aligned} E(a_1X_1 + a_2X_2 + \dots + a_nX_n) &= a_1E(X_1) + a_2E(X_2) + \dots + a_nE(X_n) \\ &= a_1\mu_1 + a_2\mu_2 + \dots + a_n\mu_n \end{aligned}$$

- For any X_1, X_2, \dots, X_n ,

$$\text{Var}(a_1X_1 + a_2X_2 + \dots + a_nX_n) = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \text{Cov}(X_i, X_j)$$

If they are independent, then

$$\begin{aligned} &\text{Var}(a_1X_1 + a_2X_2 + \dots + a_nX_n) \\ &= a_1^2 \text{Var}(X_1) + a_2^2 \text{Var}(X_2) + \dots + a_n^2 \text{Var}(X_n) \\ &= a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \dots + a_n^2 \sigma_n^2 \end{aligned}$$

Special Cases

- $E(X+Y) = E(X) + E(Y)$;
- $E(X-Y) = E(X) - E(Y)$;
- $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$
- $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y)$
- If X and Y are independent, then $\text{Cov}(X, Y) = 0$, and
 $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$
 $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$

Example

Ex. Show that if $X \sim \text{Bin}(n, p)$, then $E(X) = np$, and $\text{Var}(X) = np(1 - p)$.

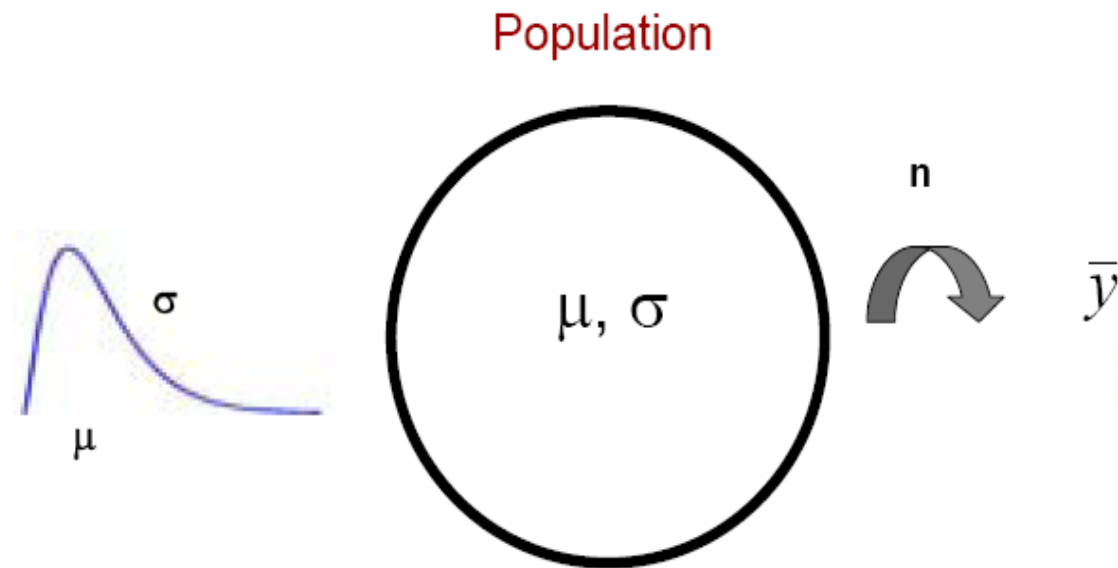
Ex. Show that if X is a negative binomial rv with pmf $nb(x; r, p)$, then $E(X) = r(1-p)/p$,
 $\text{Var}(X) = r(1 - p)/p^2$.

Statistical Inference

- From the previous two examples, we know that quite often, we need to infer the truth (**population**) from some partial information (**sample**).
- Question: why do we need a model?
- **Statistical inference** comprises the use of statistics and random sampling to make inferences concerning some unknown aspect of a population.
- A **point estimate** of a parameter θ is a single number that can be regarded as a sensible value for θ . A point estimate is obtained by selecting a suitable statistic and computing its value from the given sample data. The selected statistic is called the **point estimator** of θ .

Sampling scheme for a Mean

- Usually our problem set up will be as illustrated in the graph.



- The actual sample observations y_1, y_2, \dots, y_n (**realizations**) are assumed to be the result of a random sample Y_1, Y_2, \dots, Y_n (**random variables**) from a certain distribution.

Estimating probability

Ex. A biased coin has probability p of having heads and p is unknown. Suppose we flipped the coin for 100 times and had 73 heads. What is your best guess for p ?

Naturally, people would use estimator $\hat{p} = \frac{\text{number of heads}}{\text{number of flips}} = \frac{73}{100} = 0.73$

In other words, we are using the **sample proportion** to estimate the **population probability**.

Is this a good estimator? Are there any other estimators?

Measure of a good Estimator

- Our estimator $\hat{\theta}$ is in fact a function of the sample x_i 's, therefore, it is also a random variable. For some samples, $\hat{\theta}$ may yield a value larger than θ , whereas for other samples $\hat{\theta}$ may underestimate θ .
- The quantity $\hat{\theta} - \theta$ characterize the error of estimation. A good estimator should result in small estimation errors.
- A commonly used measure of accuracy is the **mean square error**.

$$\text{MSE} = E(\hat{\theta} - \theta)^2$$

- However, since MSE will generally depend on the value of θ , finding an estimator with smallest MSE is typically **NOT** possible.

Unbiased Estimators

- One way to find good estimators, is to restrict our attention just to estimators that have some specified desirable properties and then find the best in this restricted group.
- One popular property is *unbiasedness*.
- A point estimator $\hat{\theta}$ is said to be an *unbiased estimator* of θ if $E(\hat{\theta}) = \theta$ for every possible value of θ . If $\hat{\theta}$ is not unbiased, the difference $E(\hat{\theta}) - \theta$ is called the *bias* of $\hat{\theta}$.

Example

Ex. Recall the unbiased coin example. Is the sample proportion an unbiased estimator of the population probability?

$$\text{estimator } \hat{p} = \frac{\text{number of heads}}{\text{number of flips}} = \frac{73}{100} = 0.73$$

What distribution does “number of heads” follow? What is its expectation?

General Result

- Proposition:

When X is a binomial rv with parameters n and p , the sample proportion $\hat{p} = X/n$ is an unbiased estimator of p .

General Result

- Proposition:

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a distribution with mean μ and variance σ^2 . Then the estimator

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

is an unbiased estimator of σ^2 .

General Result

- Proposition:

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a distribution with mean μ . Then the sample mean \bar{X} is an unbiased estimator of μ . If in addition the distribution is continuous and symmetric, then the sample median M and any trimmed mean are also unbiased estimators of μ .

MVUE

- For unbiased estimators, what are their MSE's?

$$E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}))^2 = \text{Var}(\hat{\theta})$$

- Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the **minimum variance unbiased estimator (MVUE)** of θ .
- One needs more knowledge to actually identify if some estimator is really MVUE. But in a special case, we have the following theorem.

Let X_1, X_2, \dots, X_n be an i.i.d. sequence of random samples from a **normal distribution** with mean μ and σ . Then the estimator $\hat{\mu} = \bar{X}$ is the **MVUE** for μ .

The Standard Error

- When reporting a point estimator, one also reports the **standard error** associated with it.
- The **standard error** of an estimator $\hat{\theta}$ is its standard deviation $\sigma_{\hat{\theta}} = \sqrt{\text{Var}(\hat{\theta})}$. If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the **estimated standard error** of the estimator, which we denote as $\hat{\sigma}_{\hat{\theta}}$.
- The associated standard error gives us an idea of how good/accurate the estimators are.

Examples

Ex. What is the standard error of the sample proportion?

Ex. What is the standard error of the sample mean? Given that we know the variance.

Ex. What is the standard error of the sample mean? Given that we don't know the variance.