W1211 Introduction to Statistics
Lecture 18

Wei Wang

Nov 7, 2012

Distribution of a Linear Combination

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

$$a_1X_1 + a_2X_2 + \ldots + a_nX_n$$

is given by the following result.

A key result ***

Let X_1, X_2, …, X_n, have mean values μ_1, μ_2, …, μ_n, respectively, and variances σ_1^2, σ_2^2, …, σ_n^2, respectively.

- Whether or not the Xi's are independent,

$$\begin{aligned} \mathrm{E}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) &= a_1\mathrm{E}(\mathrm{X}_1) + a_2\mathrm{E}(\mathrm{X}_2) + \cdots + a_n\mathrm{E}(\mathrm{X}_n) \\ &= a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n \end{aligned}$$

- For any X_1, X_2, …, X_n,

$$\mathrm{Var}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mathrm{Cov}(\mathrm{X}_i, \mathrm{X}_j)$$

  If they are independent, then

$$\begin{aligned} & \mathrm{Var}(a_1\mathrm{X}_1 + a_2\mathrm{X}_2 + \cdots + a_n\mathrm{X}_n) \\ = \ & a_1^2\mathrm{Var}(\mathrm{X}_1) + a_2^2\mathrm{Var}(\mathrm{X}_2) + \cdots + a_n^2\mathrm{Var}(\mathrm{X}_n) \\ = \ & a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2 \end{aligned}$$

Special Cases

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

Example

- Show that if $X \sim Bin(n, p)$, then $EX = np$ and $Var(X) = np(1-p)$

Example

- Show that if $X \sim Bin(n,p)$, then $EX = np$ and $Var(X) = np(1-p)$
- Since X can be seen as the sum of n IID Bernoulli random variables, i.e.,

$$X = \sum_{i=1}^{n} Y_i, \text{ in which } Y_i \sim Bern(p)$$

- Recall that $E(Y_i) = p$ and $Var(Y_i) = p(1-p)$.
- Then

$$E(X) = E(\sum_{i=1}^{n} Y_i) = nE(Y_1) = np,$$

and

$$Var(X) = Var(\sum_{i=1}^{n} Y_i) = nVar(Y_i) = np(1-p)$$

.

Overview of Statistical Inference

- Sampling Distributions of statistics enable us to infer characteristics of populations from samples.

Overview of Statistical Inference

- Sampling Distributions of statistics enable us to infer characteristics of populations from samples.
- Specifically, there are three topics in statistical inference that we will touch on
  - Point Estimation (Ch 6)
  - Confidence Interval (Ch 7)
  - Hypothesis Testing based on A Single Sample (Ch 8) and Two Samples (Ch 9)

Overview of Statistical Inference

- Sampling Distributions of statistics enable us to infer characteristics of populations from samples.
- Specifically, there are three topics in statistical inference that we will touch on
  - Point Estimation (Ch 6)
  - Confidence Interval (Ch 7)
  - Hypothesis Testing based on A Single Sample (Ch 8) and Two Samples (Ch 9)
- 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.

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

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 $\hat{\theta}$ - θ 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.

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

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

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

is an unbiased estimator of σ^2.

General Result

- Proposition:

Let $X_1, X_2, \ldots, X_n$ 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?

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

- 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 $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence of random samples from a normal distribution with mean μ and σ. Then the estimator $\hat{\mu} = \bar{\mathrm{X}}$ 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 $\sigma_{\hat{\theta}} = \sqrt{\mathrm{Var}(\hat{\theta})}$. 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.

Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip X_i is a Bernoulli RV with parameter p, the estimator of parameter p is the sample mean/proportion

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip X_i is a Bernoulli RV with parameter p, the estimator of parameter p is the sample mean/proportion

$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- If we flipped 100 times and observed 75 heads, then our estimate of p is

$$\hat{p} = \frac{75}{100} = 0.75$$

Estimator, Its Standard Error and Estimated Standard Error

- Now we are trying to estimate the probability of getting heads of a biased coin, so each flip X_i is a Bernoulli RV with parameter p, the estimator of parameter p is the sample mean/proportion
$$\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$$
- If we flipped 100 times and observed 75 heads, then our estimate of p is
$$\hat{p} = \frac{75}{100} = 0.75$$
- Also, we need to report how good our estimator is through its Standard Error. This is also related to the Interval Estimation.

Estimator, Its Standard Error and Estimated Standard Error

- The standard error is $Var(\hat{p}) = \frac{p(1-p)}{n}$, but we cannot report it since we don't know what p is.
- So we can only report the estimated standard error of the estimator $\hat{p}$

$$\widehat{Var(\hat{p})} = \frac{\hat{p}(1-\hat{p})}{n}$$

Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean μ and variance σ^2, what's the estimator of μ?

Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean μ and variance σ^2, what's the estimator of μ?
- $$\hat{\mu} = \bar{X}$$

Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean μ and variance σ^2, what's the estimator of μ?
- $$\hat{\mu} = \bar{X}$$
- The standard error of $\hat{\mu}$ is $\sqrt{Var(\hat{\mu})} = \frac{\sigma}{\sqrt{n}}$. Can we report $\frac{\sigma}{\sqrt{n}}$?

Another Example

- Now we have $X_1, X_2, \ldots, X_N$ IID with mean μ and variance σ^2, what's the estimator of μ?
- $$\hat{\mu} = \bar{X}$$
- The standard error of $\hat{\mu}$ is $\sqrt{Var(\hat{\mu})} = \frac{\sigma}{\sqrt{n}}$. Can we report $\frac{\sigma}{\sqrt{n}}$?
- It really depends on whether or not we know σ. If we know it, then we can report $\frac{\sigma}{\sqrt{n}}$; otherwise, we can only report $\frac{\hat{\sigma}}{\sqrt{n}}$.

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.