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

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 $\hat{\theta}$ - 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?