Lecture 20: Covariance and correlation

Statistics 251

Yi Sun and Zhongjian Wang

Department of Statistics
The University of Chicago

Lecture Outline

Covariance

Correlation

Where are we?

Covariance

Correlation

Expected value of multiplication of functions of independent r.v.

If X and Y are independent, then, for any functions h and g,

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)]$$

Continuous case proof:

Definition of Covariance

Recall definition of $\operatorname{Var} X$:

Definition of Covariance

Recall definition of $\mathrm{Var}\, X$:

The covariance between X and Y, denoted by Cov (X, Y), is defined by

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Alternative form

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

$$\Updownarrow$$

$$\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Properties

(i) $\text{Cov}(X, Y) = \text{Cov}(Y, X)$

(ii) $\text{Cov}(X, X) = \text{Var}(\mathrm{X})$

(iii) $\text{Cov}(aX, Y) = a\,\text{Cov}(X, Y)$

(iv) $\text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{m} Y_j\right) = \sum_{i=1}^{n} \sum_{i=1}^{m} \text{Cov}\left(X_i, Y_j\right)$

Variance of sum

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right) + 2\sum_{i<j} \text{Cov}\left(X_i, X_j\right)$$

Example: Sample Variance

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2, Llet $\bar{X} = \sum_{i=1}^{n} X_i/n$ be the *sample mean*. Then what is $\text{Var}(\bar{X})$?

Example: Sample Variance

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having expected value μ and variance σ^2, Llet $\bar{X} = \sum_{i=1}^{n} X_i/n$ be the *sample mean*. Then what is $\text{Var}(\bar{X})$? The quantities $X_i - \bar{X}, i = 1, \ldots, n$, are called *deviations*, as they equal the differences between the individual data and the sample mean. The random variable

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

is called the *sample variance*. Find $\mathbb{E}\left[S^2\right]$.

Where are we?

Covariance

Correlation

Definition

The correlation of two random variables X and Y, denoted by $\rho(X, Y)$, is defined, as long as $\text{Var}(X)\,\text{Var}(Y)$ is positive, by

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}$$

Definition

The correlation of two random variables X and Y, denoted by $\rho(X, Y)$, is defined, as long as $\operatorname{Var}(X)\operatorname{Var}(Y)$ is positive, by

$$\rho(X, Y) = \frac{\operatorname{Cov}(X, Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

Then $-1 \leq \rho(X, Y) \leq 1$.

Example: Derivation and sample mean are uncorrelated

Let $X_1, \ldots, X_n$ be independent and identically distributed random variables having variance σ^2. Then

$$\text{Cov}\left(X_i - \bar{X}, \bar{X}\right) = 0$$

Example: if $Y = a + bX$

Given mean and variance of X to be μ and σ^2, calculate $\rho(X, Y)$, where $Y = a + bX$.