W1211 Introduction to Statistics
Lecture 13

Wei Wang

Oct 17, 2012

Independent rv's

- Recall the definition of independence of two random events A and B.

$$P(A \cap B) = P(A)\, P(B)$$

- We say two random variables X and Y are independent if and only if

$$P(X=x, Y=y) = P(X=x)\, P(Y=y), \text{ for any } x \text{ and } y.$$

- More specifically, two random variables X and Y are said to be independent if for every pair x and y values,

$$p(x, y) = p_X(x)\, p_Y(y), \text{ when X and Y are discrete;}$$

  or

$$f(x, y) = f_X(x)\, f_Y(y), \text{ when X and Y are continuous.}$$

<u>Ex.</u> The second case of the previous example.

Multiple Random Variables

- If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function
$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$
If the variables are continuous, the joint pdf of $X_1, X_2, \ldots, X_n$ is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$,
$$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n) dx_1 \ldots dx_n$$

- What should be the regularity conditions for $p(x_1, x_2, \ldots, x_n)$ and $f(x_1, x_2, \ldots, x_n)$?

- How do get the marginal distributions of $X_1, X_2, \ldots$ by using $p(x_1, x_2, \ldots, x_n)$ and $f(x_1, x_2, \ldots, x_n)$?

Independence

- Proposition:

  The random variables $X_1, X_2, \ldots, X_n$, are said to be independent if for every subset $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$, of the variables (each pair, each triple, and so on), the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

- $p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$

- $f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$

Conditional dist.

- Using the marginal distributions, one can calculate the conditional distribution of one rv given the other.

- Let X and Y be two conditional rv's with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x)>0$, the conditional probability density function of Y given that X=x is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} \quad -\infty < y < \infty.$$

- If X and Y are discrete, replace pdf's by pmf's in this definition gives the conditional probability mass function of Y when X=x.

Expectation of Functions

- Recall how we compute E[*h*(X)]. A similar result also holds for a function *h*(X, Y) of two jointly distributed rv's.

- Let X and Y be jointly distributed rv's with pmf *p*(*x*, *y*), if they are discrete; or pdf *f* (*x*, *y*), if they are continuous. The expected value of a function *h*(X, Y), denoted by E[*h*(X, Y)] is given by

$$\mathrm{E}[h(\mathrm{X},\mathrm{Y})] = \begin{cases} \sum_x \sum_y h(x,y) \cdot p(x,y) & \text{if X and Y are discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x,y) \cdot f(x,y) dx dy & \text{if X and Y are continuous} \end{cases}$$

- This result can also be extended to multiple (>2) rv case.

Expectation of Linear Function of Multiple RV's

- Linearity is well preserved in expectation.

$$E(a \cdot X + b \cdot Y + c) = a \cdot E(X) + b \cdot E(Y) + c$$

Expectation of Product of Multiple RV's

- Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

Expectation of Product of Multiple RV's

- Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

- But if X and Y are independent, then

$$\begin{aligned} E(XY) &= \int \int xyf(x,y)dxdy = \int \int xyf_X(x)f_Y(y)dxdy \\ &= \int xf_X(x)dx \int yf_Y(y)dy = E(X)E(Y) \end{aligned}$$

Expectation of Product of Multiple RV's

- Unlike the linear case, expectation of product in general doesn't equal to the product of expectations

$$E(XY) \neq E(X)E(Y)$$

- But if X and Y are independent, then

$$\begin{aligned} E(XY) &= \int\int xyf(x,y)dxdy = \int\int xyf_X(x)f_Y(y)dxdy \\ &= \int xf_X(x)dx \int yf_Y(y)dy = E(X)E(Y) \end{aligned}$$

- And for independent RV's, in general

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

Covariance

- When two random variables X and Y are not independent, it is often of interest to assess how strongly they are related to one another.

- A popular measurement to characterize the dependence of two rv's is called correlation. To calculate correlation of two rv's, we'll have calculate the covariance of the two rv's.

- The covariance between two rv's X and Y is

$$
\begin{aligned}
\mathrm{Cov}(\mathrm{X}, \mathrm{Y}) &= \mathrm{E}[(\mathrm{X} - \mu_X)(\mathrm{Y} - \mu_Y)] \\
&= \begin{cases} \sum_x \sum_y (x - \mu_X)(y - \mu_Y) \cdot p(x, y) & \text{X, Y discrete} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) \cdot f(x, y) dx dy & \text{X, Y continuous} \end{cases}
\end{aligned}
$$

Short cut

- Proposition:

$$\boxed{\mathrm{Cov}(X, Y) = E(XY) - E(X)E(Y)}$$

- What happens if we set Y=X?

Covariance and Variance

- As we can see, variance is a special case of covariance, where $X = Y$.

Covariance and Variance

- As we can see, variance is a special case of covariance, where $X = Y$.
- Variance of linear function of multiple RV's is given by

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2ab \cdot Cov(X, Y)$$

Example

Ex. Suppose the joint distribution of X and Y are

$$f(x,y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x+y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

What is the covariance of X and Y?

$$f_X(x) = \int_y f(x,y)dy = \int_0^{1-x} 24xydy = 12x(1-x)^2$$

$$f_Y(y) = 12y(1-y)^2$$

$$\mathrm{E(X)} = \int_0^1 x \cdot 12x(1-x)^2 dx = \frac{2}{5} = \mathrm{E(Y)}$$

$$\mathrm{E(XY)} = \int\int_{x,y} xyf(x,y)dxdy = \int_0^1 \int_0^{1-y} 24x^2y^2 dxdy = \frac{2}{15}$$

$$\mathrm{Cov(X,Y)} = \mathrm{E(XY)} - \mathrm{E(X)E(Y)} = \frac{2}{15} - \left(\frac{2}{5}\right)^2 = -\frac{2}{75}$$

Correlation

- The correlation coefficient of X and Y, denoted by Corr(X, Y) or $\rho_{X,Y}$ is defined by

$$\rho_{X,Y} = \frac{\mathrm{Cov}(\mathrm{X}, \mathrm{Y})}{\sigma_X \cdot \sigma_Y}$$

- Because of Cauchy-Schwarz inequality, we have

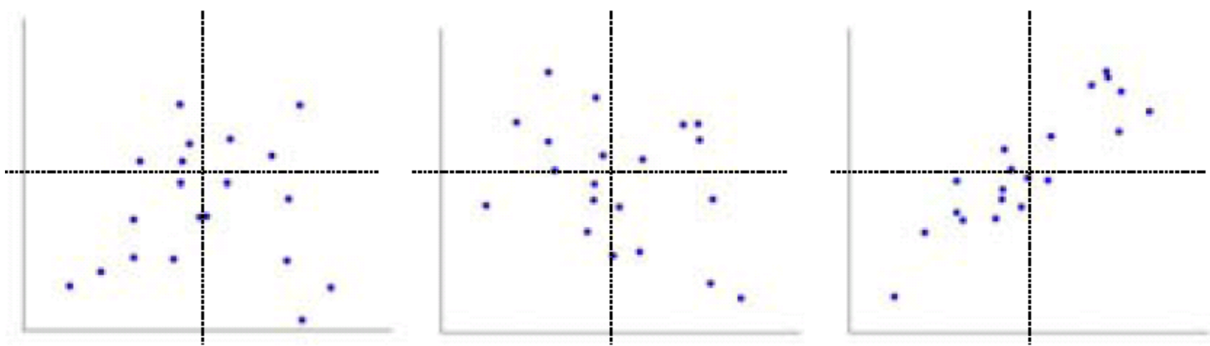
$$\mathrm{Cov}^2(\mathrm{X}, \mathrm{Y}) \leq \mathrm{Var}(\mathrm{X})\mathrm{Var}(\mathrm{Y}) \Longrightarrow |\rho_{X,Y}| \leq 1$$

- The correlation coefficient $\rho_{X,Y}$ is NOT a completely general measure of the strength of a relationship. $\rho_{X,Y}$ is actually a measure of the degree of *linear* relationship between X and Y.

Remarks

- If X and Y are independent, then $\rho_{X,Y} = 0$ (why?). But $\rho_{X,Y} = 0$ does NOT imply independence.

- $\rho_{X,Y} = 1$ or -1 iff $Y = aX+b$ for some numbers a and b with $a \neq 0$.

- 

$r \approx 0$ $\quad$ $r \approx -0.3$ $\quad$ $r \approx 0.9$

Relationship Between Correlation and Independence

- Independence leads to uncorrelatedness.

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

Relationship Between Correlation and Independence

- Independence leads to uncorrelatedness.

$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

- But not vice versa!

Relationship Between Correlation and Independence

- Independence leads to uncorrelatedness.

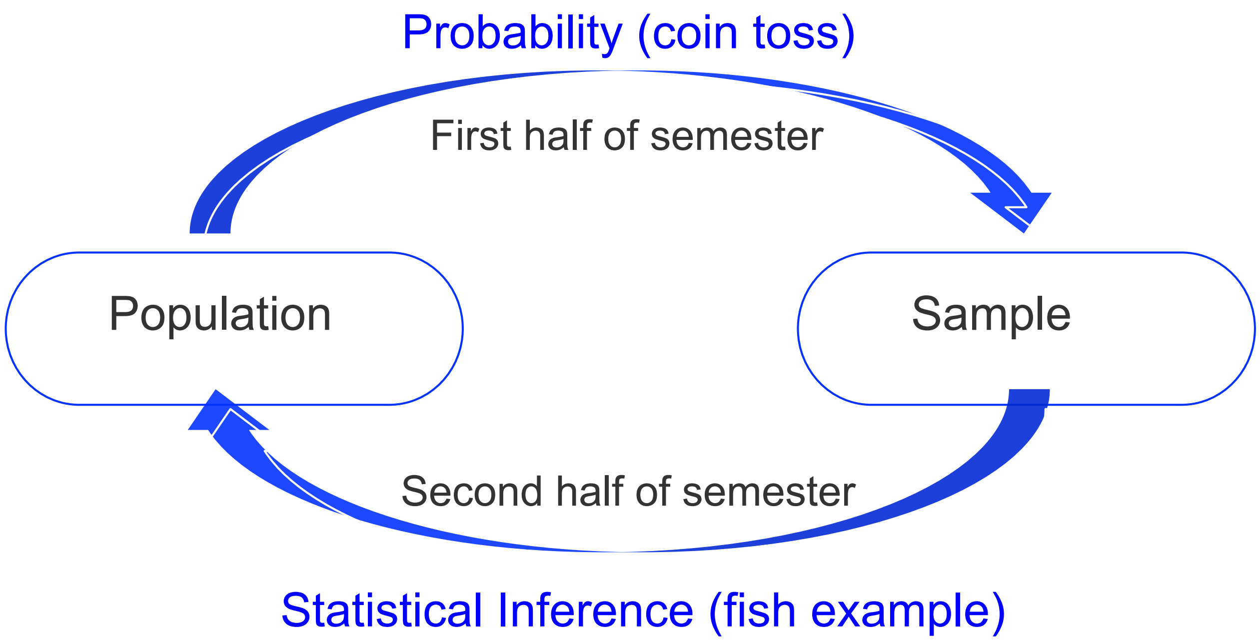
$$Cov(X, Y) = E(XY) - E(X)E(Y) = E(X)E(Y) - E(X)E(Y) = 0$$

- But not vice versa!
- We will talk about this more in regression.

Population and Sample

- We will start changing our discussion from probability to statistics, which means we need to think about samples and how they relate to the underlying population.
- Recall the relationship between population and sample (probability and inference) that we visualized in the first lecture.

Probability and Inference



Sample and Statistics

- A statistic is any quantity whose value can be calculated from sample data, such as Sample Mean and Sample Variance.
- Before obtaining data, a statistic is also a RV. The bulk of statistical inference is to find the distribution of the statistics, or the so-called *Sampling Distributions*.
- To make things easier, we often need to assume the observed data are *Simple Random Samples*, which means they are IID (Independently Identically Distributed).

Introduction to IID

- A sequence of random variables, X_1, X_2, …, X_n, is independent and identically distributed (i.i.d.) if each random variable has the same probability distribution as the others and all are mutually independent.

- In statistical analysis, we often assume the sampled data X_1, X_2, …, X_n, are i.i.d. from a common distribution $f(x)$. And usually, we end up analyzing a linear combination of the X_i's, that is

$$\mathrm{Y} = a_1\mathrm{X}_1 + \cdots + a_n\mathrm{X}_n = \sum_{i=1}^{n} a_i\mathrm{X}_i$$

Sample Mean***

- Let $X_1, X_2, \ldots, X_n$, be an i.i.d. sequence of rv's from a distribution with mean value μ and standard deviation σ.

- Notice that the sample mean or the sample total ($T = X_1+X_2+\ldots+X_n$) can also be viewed as a special case of linear combination of $X_1, X_2, \ldots, X_n$. In the i.i.d. case,

$$E(T) = E(X_1)+E(X_2)+\ldots+E(X_n) = n\mu$$

$$Var(T) = Var(X_1)+Var(X_2)+\ldots+Var(X_n) = n\sigma^2$$

- It is also easy to verify that for sample mean,

$$\mathrm{E}(\bar{\mathrm{X}}) = \mu_{\bar{X}} = \mu$$

$$\mathrm{Var}(\bar{\mathrm{X}}) = \sigma^2_{\bar{X}} = \sigma^2/n \Longrightarrow \sigma_{\bar{X}} = \sigma/\sqrt{n}$$

Invariance of Normal RV under Linear Transformation

- When $X_1, X_2, X_3, X_4, \ldots$ are normal random variables, then the linear combination of them

$$a_1 X_1 + a_2 X_2 + \ldots + a_n X_n = \sum_{i=1}^{n} a_i X_i$$

  is still a normal random variable.
- In particular, sample mean $\bar{X}$ is still a random variables.
- Remark: 1. No IID assumption is necessary; 2. This property is for Normal only.

Sample Mean of IID Normal

- If $X_1, X_2, \ldots, X_n$ IID $\sim N(\mu, \sigma^2)$, then what is the distribution of $\bar{X}$?

Sample Mean of IID Normal

- If $X_1, X_2, \ldots, X_n$ IID $\sim N(\mu, \sigma^2)$, then what is the distribution of $\bar{X}$?
- $$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

Sample Mean of IID Normal

- If $X_1, X_2, \ldots, X_n$ IID $\sim N(\mu, \sigma^2)$, then what is the distribution of $\bar{X}$?
- $$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$
- But how can we derive the distribution of Sample Mean when the sample are not normal? We need to use Central Limit Theorem.

CLT

- Theorem:

The Central Limit Theorem (CLT)

Let X_1, X_2, …, 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 $\sigma^2_{\bar{X}} = \sigma^2/n$;
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.