Stat 451
3-5-20

Review from Tuesday:

(1)

$\sigma_{XY} = E[XY] - E[X]E[Y]$ Covariance

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$ Correlation

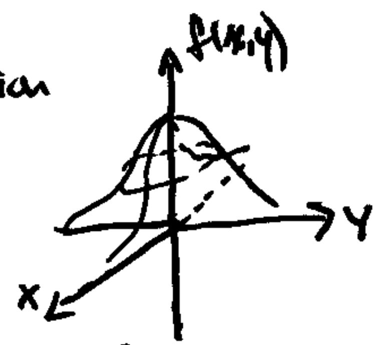
$-1 \leq \rho_{XY} \leq 1$

Independence $\Rightarrow \rho_{XY} = 0$ But $\rho_{XY} = 0 \not\Rightarrow$ independence

$\rho_{XY} = \pm 1 \Rightarrow X$ is a linear function of Y

(2)

Bivariate normal distribution



$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

5 parameters: $\mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho_{XY}$

What does this joint distribution look like when $\rho = 0$?

(3)

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \cdot \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$$

$$= f_X(x) \cdot f_Y(y)$$

$\Rightarrow$ X & Y are independent

Big result: In the bivariate normal distribution,

$$\rho = 0 \Leftrightarrow \text{independence}$$

(4)

Example of finding a correlation for a pair of discrete random variables

Defn: The Multinomial distribution

Run n independent trials

Each trial has r possible outcomes

The probabilities remain the same on each trial

The joint p.m.f. is

$$p(x_1,\dots,x_r) = \frac{n!}{x_1!\cdots x_r!} p_1^{x_1}\cdots p_r^{x_r}$$

(5)

Note: $x_1 + x_2 + \dots + x_r = n$

$p_1 + p_2 + \dots + p_r = 1$

Pick X_i and X_j + find their correlation

Step 1: Find the marginal distributions of X_i & X_j.

Note: $X_i \sim \text{Bino}(n, p_i)$

and $X_j \sim \text{Bino}(n, p_j)$

Step 2: $E[X_i] = np_i \quad V[X_i] = np_i \underbrace{q_i}_{1-p_i}$ and

$E[X_j] = np_j \quad V[X_j] = np_j q_j$

(6)

Step 3: Find $E[X_i X_j]$

$X_i + X_j \sim \text{Bino}(n, p_i + p_j)$

So $V[X_i + X_j] = n(p_i + p_j)(1 - p_i - p_j)$

$\qquad = V[X_i] + V[X_j] + 2\,\text{Cov}(X_i, X_j)$

$$\text{Cov}(X_i, X_j) = \frac{n(p_i+p_j)(1-p_i-p_j) - np_i \overset{(1-p_i)}{q_i} - np_j \overset{(1-p_j)}{q_j}}{2}$$

$$= \frac{n\left(\cancel{p_i} - \cancel{p_i^2} - p_i p_j + \cancel{p_j} - p_i p_j - \cancel{p_j^2} - \cancel{p_i} + \cancel{p_i^2} - \cancel{p_j} + \cancel{p_j^2}\right)}{2}$$ 7

$$Cov(X_i, X_j) = -n p_i p_j$$

$$Corr(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{V[X_i]V[X_j]}}$$

$$= \frac{-n p_i p_j}{\sqrt{n p_i q_i \, n p_j q_j}} = -\sqrt{\frac{p_i p_j}{q_i q_j}}$$

In a previous class, 8

$X_1 = $ # wins $\quad p_1 = .5$

$X_2 = $ # losses $\quad p_2 = .4 \qquad n = 3$

$X_3 = $ # ties $\quad p_3 = .1$

$$Corr(X_1, X_2) = -\sqrt{\frac{(.5)(.4)}{(.5)(.6)}} = -\sqrt{\frac{2}{3}}$$

$$= -.816$$

(9)

Suppose that $X_1, X_2, \ldots, X_n$ are <u>independent</u> random variables, each having the same mean, μ and the same variance, σ^2.

Let $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. This is a new random variable.

Find its mean and variance.

(10)

$$\mu_{\bar{X}} = E[\bar{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu$$

$$= \frac{1}{n} n\mu = \mu$$

$$\sigma^2_{\bar{X}} = V[\bar{X}] = V\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n^2}\left(\sum_{i=1}^{n} V[X_i] + \cancel{\text{covariance terms}}^{\,0 \text{ by indep.}}\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

HW #9 due 3/12

p. 237 # 51, 52, 56

51. The time taken by a randomly selected applicant for a mortgage to fill out a certain form has a normal distribution with mean value 10 min and standard deviation 2 min. If five individuals fill out a form on one day and six on another, what is the probability that the sample average amount of time taken on each day is at most 11 min?

52. The lifetime of a certain type of battery is normally distributed with mean value 10 hours and standard deviation 1 hour. There are four batteries in a package. What lifetime value is such that the total lifetime of all batteries in a package exceeds that value for only 5% of all packages?

56. A binary communication channel transmits a sequence of "bits" (0s and 1s). Suppose that for any particular bit transmitted, there is a 10% chance of a transmission error (a 0 becoming a 1 or a 1 becoming a 0). Assume that bit errors occur independently of one another.

a. Consider transmitting 1000 bits. What is the approximate probability that at most 125 transmission errors occur?

b. Suppose the same 1000-bit message is sent two different times independently of one another. What is the approximate probability that the number of errors in the first transmission is within 50 of the number of errors in the second?