Chapter 6.6 The Bivariate Normal Distribution

Jim Albert and Monika Hu

Chapter 6 Joint Probability Distributions

Introduction

- Suppose one collects multiple body measurements from a group of 30 students.
- One might collect the diameter of the wrist and the diameter of the ankle.
- If X and Y denote the two body measurements (measured in cm) for a student, then one might think that the density of X and the density of Y is each Normally distributed.
- Moreover, the two random variables would be positively correlated; if a student has a large wrist diameter, one would predict her to also have a large forearm length.

Bivariate Normal Density

- A convenient joint density function for two continuous measurements X and Y is the Bivariate Normal density with density given by

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho}} \exp\left[-\frac{1}{2(1-\rho^2)}(z_X^2 - 2\rho z_X z_Y + z_Y^2)\right],$$

where z_X and z_Y are the standardized scores

$$z_X = \frac{x-\mu_X}{\sigma_X}, \;\; z_Y = \frac{y-\mu_Y}{\sigma_Y},$$

and μ_X, μ_Y and σ_X, σ_Y are respectively the means and standard deviations of X and Y.

- The parameter ρ is the correlation of X and Y and measures the association between the two variables.

Contour plots of 4 Bivariate Normal Distributions

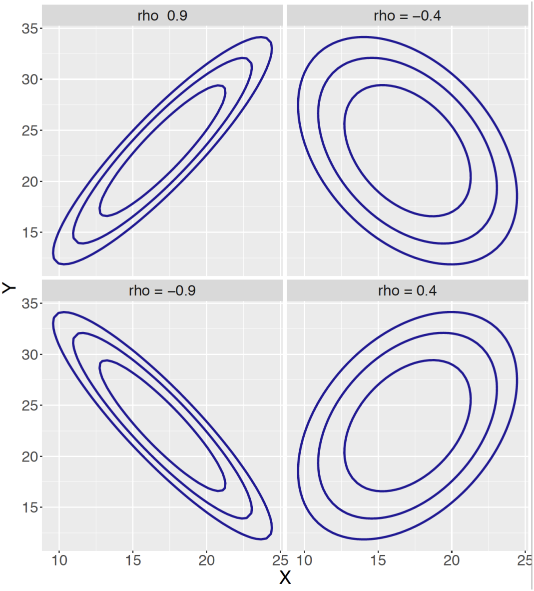

Figure 1: Contour graphs of four Bivariate Normal distributions with

Explanation

- The bottom right graph corresponds to the values $\mu_X = 17, \mu_Y = 23, \sigma_X = 2, \sigma_Y = 3$ and $\rho = 0.4$ where X and Y represent the wrist diameter and ankle diameter measurements of the student.
- The correlation value of $\rho = 0.4$ reflects the moderate positive correlation of the two body measurements.
- The other three graphs use the same means and standard deviations but different values of the ρ parameter.
- We see that the Bivariate Normal distribution is able to model a variety of association structures between two continuous measurements.

Properties

There are a number of attractive properties of the Bivariate Normal distribution.

1. **The marginal densities of X and Y are Normal.** So X has a Normal density with parameters μ_X and σ_X and likewise Y is $\mathrm{Norma}(\mu_Y, \sigma_Y)$.
2. **The conditional densities will also be Normal.** For example, if one is given that $Y = y$, then the conditional density of X given $Y = y$ is Normal where

$$E(X \mid Y = y) = \mu_X + \rho \frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

$$Var(X \mid Y = y) = \sigma_X^2(1 - \rho^2).$$

Properties (continued)

- Similarly, if one knows that $X = x$, then the conditional density of Y given $X = x$ is Normal with mean

$$\mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

and variance

$$\sigma_Y^2(1 - \rho^2)$$

.

Properties (continued)

3. **For a Bivariate Normal distribution, X and Y are independent if and only if the correlation $\rho = 0$.**

- In contrast, as the correlation parameter ρ approaches $+1$ and -1, then all of the probability mass will be concentrated on a line where $Y = aX + b$.

Bivariate Normal calculations

- Recall that X denotes the wrist diameter, Y represents the ankle diameter and we are assuming (X, Y) has a Bivariate Normal distribution with parameters $\mu_X = 17, \mu_Y = 23, \sigma_X = 2, \sigma_Y = 3$ and $\rho = 0.4$

Find the probability a student's wrist diameter exceeds 20 cm.

One wants the probability $P(X > 20)$. From the facts above, the marginal density for X will be Normal with mean $\mu_X = 17$ and standard deviation $\sigma_X = 2$. So this probability is computed using the function `pnorm()`:

```
1 - pnorm(20, 17, 2)
```

```
## [1] 0.0668072
```

Bivariate Normal calculations

Suppose one is told that the student's ankle diameter is 20 cm – find the conditional probability $P(X > 20 \mid Y = 20)$.

The distribution of X conditional on the value $Y = y$ is Normal with mean $\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$ and variance $\sigma_X^2(1 - \rho^2)$.

$$E(X \mid Y = 20) = \mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y) = 16.2.$$

$$SD(X \mid Y = 20) = \sqrt{\sigma_X^2(1 - \rho^2)} = 1.83.$$

So probability is

```r
1 - pnorm(20, 16.2, 1.83)
```

```
## [1] 0.01892374
```

Bivariate Normal calculations

Are X and Y independent variables?

- For a Bivariate Normal distribution, a necessary and sufficient condition for independence is that the correlation $\rho = 0$.
- Since the correlation between the two variables is not zero, the random variables X and Y can not be independent.

Find the probability a student's ankle diameter measurement is at 50 percent greater than her wrist diameter measurement, that is $P(Y > 1.5X)$.

- Simulation provides an attractive method of computing this probability.

Simulating Bivariate Normal measurements

- One simulates a large number, say 1000, draws from the Bivariate Normal distribution and then finds the fraction of simulated (x, y) pairs where $y > 1.5x$.
- Probability of interest is approximated by 0.242.