7. The Multivariate Normal Distribution

The Bivariate Normal Distribution

Definition

Suppose that U and V are independent random variables each, with the standard normal distribution. We will need the following five parameters: $\mu_1 \in \mathbb{R}$, $\mu_2 \in \mathbb{R}$, $\sigma_1 \in (0, \infty)$, $\sigma_2 \in (0, \infty)$, and $\rho \in [-1, 1]$. Now let X and Y be new random variables defined by

$$X = \mu_1 + \sigma_1\, U, \quad Y = \mu_2 + \sigma_2\, \rho\, U + \sigma_2 \sqrt{1 - \rho^2}\, V$$

The joint distribution of (X, Y) is called the **bivariate normal distribution** with parameters $(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$.

Basic Properties

For the following exercises, use properties of mean, variance, covariance, and the normal distribution.

1. Show that X is normally distributed with mean μ_1 and standard deviation σ_1.

2. Show that Y is normally distributed with mean μ_2 and standard deviation σ_2.

3. Show that $\text{cov}(X, Y) = \rho$

4. Show that X and Y are independent if and only if $\text{cov}(X, Y) = 0$.

Thus, for two random variables with a joint normal distribution, the random variables are independent if and only if they are uncorrelated.

5. In the bivariate normal experiment, change the standard deviations of X and Y with the scroll bars. Watch the change in the shape of the probability density functions. Now change the correlation with the scroll bar and note that the probability density functions do *not* change. For various values of the parameters, run the experiment 2000 times with an update frequency of 10. Observe the cloud of points in the scatterplot, and note the apparent convergence of the empirical density function to the probability density function.

The Probability Density Function

We will now use the change of variables formula to find the joint probability density function of (X, Y).

6. Show that inverse transformation is given by

$$u = \frac{x - \mu_1}{\sigma_1}, \quad v = \frac{x - \mu_1}{\sigma_1 \sqrt{1 - \rho^2}} - \rho \frac{y - \mu_2}{\sigma_2 \sqrt{1 - \rho^2}}$$

7. Show that the Jacobian of the transformation in the previous exercise is

$$\frac{\partial(u, v)}{\partial(x, y)} = \frac{1}{\sigma_1 \sigma_2 \sqrt{1 - \rho^2}}$$

Note that the Jacobian is a constant; this is because the transformation is linear.

8. Use the previous exercises, the independence of U and V, and the change of variables formula to show that the joint probability density function of (X, Y) is

$$f(x, y) = \frac{1}{2 \pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left(-\frac{1}{2(1 - \rho^2)} \left(\frac{(x - \mu_1)^2}{\sigma_1^2} - 2 \rho \frac{(x - \mu_1)(y - \mu_2)}{\sigma_1 \sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2}\right)\right),$$
$$(x, y) \in \mathbb{R}^2$$

If c is a constant, the set of points $\left\{(x, y) \in \mathbb{R}^2 : f(x, y) = c\right\}$ is called a **level curve** of f (these are curves of constant probability density).

9. Show that

a. The level curves of f are ellipses centered at (μ_1, μ_2).
b. The axes of these ellipses are parallel to the coordinate axes if and only if $\rho = 0$

10. In the bivariate normal experiment, run the experiment 2000 times with an update frequency of 10 for selected values of the parameters. Observe the cloud of points in the scatterplot and note the apparent convergence of the empirical density function to the probability density function.

Transformations

The following exercise shows that the bivariate normal distribution is preserved under affine transformations.

11. Define $W = a_1 X + b_1 Y + c_1$ and $Z = a_2 X + b_2 Y + c_2$. Use the change of variables formula to show that (W, Z) has a bivariate normal distribution. Identify the means, variances, and correlation.

12. Show that the conditional distribution of Y given $X = x$ is normal with mean and variance given by

a. $\mathbb{E}(Y | X = x) = \mu_2 + \rho \sigma_2 \frac{x - \mu_1}{\sigma_1}$
b. $\text{var}(Y | X = x) = \sigma_2^2 (1 - \rho^2)$

13. Use the definition of X and Y in terms of the independent standard normal variables U and V to show

that

$$Y = \mu_2 + \rho\,\sigma_2 \frac{X - \mu_1}{\sigma_1} + \sigma_2 \sqrt{1 - \rho^2}\, V$$

Now give another proof of the result in Exercise 12 (note that X and V are independent).

14. In the bivariate normal experiment, set the standard deviation of X to 1.5, the standard deviation of Y to 0.5, and the correlation to 0.7.

a. Run the experiment 100 times, updating after each run.
b. For each run, compute $\mathbb{E}(Y|X = x)$ the predicted value of Y for the given the value of X.
c. Over all 100 runs, compute the square root of the average of the squared errors between the predicted value of Y and the true value of Y.

The following problem is a good exercise in using the change of variables formula and will be useful when we discuss the simulation of normal variables.

15. Recall that U and V are independent random variables each with the standard normal distribution. Define the polar coordinates $(R,\ \Theta)$ of $(U,\ V)$ by the equations $U = R\cos(\Theta)$, $V = R\sin(\Theta)$ where $R \geq 0$ and $0 \leq \Theta < 2\,\pi$. Show that

a. R has probability density function $g(r) = r\,e^{-\frac{1}{2}r^2},\ \ r \geq 0$.
b. Θ is uniformly distributed on $[0,\ 2\,\pi)$.
c. R and Θ are independent.

The distribution of R is known as the **Rayleigh distribution**, named for William Strutt, Lord Rayleigh. It is a member of the family of Weibull distributions, named in turn for Wallodi Weibull.

The General Multivariate Normal Distribution

The general multivariate normal distribution is a natural generalization of the bivariate normal distribution studied above. The exposition is very compact and elegant using expected value and covariance matrices, and would be horribly complex without these tools. Thus, this section requires some prerequisite knowledge of linear algebra. In particular, recall that $\boldsymbol{A}^T$ denotes the transpose of a matrix A.

The Standard Normal Distribution

Suppose that $\boldsymbol{Z} = (Z_1,\ Z_2,\ ...,\ Z_n)$ is a vector of independent random variables, each with the standard normal distribution. Then $\boldsymbol{Z}$ is said to have the n-dimensional **standard normal distribution**.

16. Show that $\mathbb{E}(\boldsymbol{Z}) = \boldsymbol{0}$ (the zero vector in $\mathbb{R}^n$).

17. Show that $\text{VC}(\boldsymbol{Z}) = \boldsymbol{I}$ (the $n \times n$ identity matrix).

18. Show that $\boldsymbol{Z}$ has probability density function

$$g(z) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z \cdot z\right), \quad z \in \mathbb{R}^n$$

19. Show that $\boldsymbol{Z}$ has moment generating function given by

$$\mathbb{E}(\exp(\boldsymbol{t} \cdot \boldsymbol{Z})) = \exp\left(\frac{1}{2} \boldsymbol{t} \cdot \boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^n$$

The General Normal Distribution

Now suppose that $\boldsymbol{Z}$ has the n-dimensional standard normal distribution. Suppose that $\boldsymbol{\mu} \in \mathbb{R}^n$ and that $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is invertible. The random vector $\boldsymbol{X} = \boldsymbol{\mu} + \boldsymbol{A}\,\boldsymbol{Z}$ is said to have an n-dimensional **normal distribution**.

20. Show that $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu}$.

21. Show that $\text{VC}(\boldsymbol{X}) = \boldsymbol{A}\,\boldsymbol{A}^T$, and that this matrix is invertible and positive definite.

22. Let $\boldsymbol{V} = \text{VC}(\boldsymbol{X}) = \boldsymbol{A}\,\boldsymbol{A}^T$. Use the multivariate change of variables theorem to show that $\boldsymbol{X}$ has probability density function

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det(\boldsymbol{V})}} \exp\left(-\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu}) \cdot \boldsymbol{V}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^n$$

23. Show that $\boldsymbol{X}$ has moment generating function given by

$$\mathbb{E}(\exp(\boldsymbol{t} \cdot \boldsymbol{X})) = \exp\left(\boldsymbol{\mu} \cdot \boldsymbol{t} + \frac{1}{2} \boldsymbol{t} \cdot \boldsymbol{V}\,\boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^n$$

Note that the matrix $\boldsymbol{A}$ that occurs in the transformation is not unique, but of course the variance-covariance matrix $\boldsymbol{V}$ is unique. In general, for a given positive definite matrix $\boldsymbol{V}$, there are many invertible matrices $\boldsymbol{A}$ such that $\boldsymbol{A}\,\boldsymbol{A}^T = \boldsymbol{V}$. A theorem in matrix theory states that there is a unique lower triangular matrix $\boldsymbol{L}$ with this property.

24. Identify the lower triangular matrix $\boldsymbol{L}$ for the bivariate normal distribution.

Transformations

The multivariate normal distribution is invariant under two basic types of transformations: affine transformation with an invertible matrix, and the formation of subsequences.

25. Suppose that $\boldsymbol{X}$ has an n-dimensional normal distribution. Suppose also that $\boldsymbol{a} \in \mathbb{R}^n$ and that $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is invertible. Show that $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\,\boldsymbol{X}$ has a multivariate normal distribution. Identify the mean

vector and the variance-covariance matrix of $\boldsymbol{Y}$.

26. Suppose that $\boldsymbol{X}$ has an n-dimensional normal distribution. Show that any permutation of the coordinates of $\boldsymbol{X}$ also has an n-dimensional normal distribution. Identify the mean vector and the variance-covariance matrix. *Hint*: Permuting the coordinates of $\boldsymbol{X}$ corresponds to multiplication of $\boldsymbol{X}$ by a **permutation matrix**--a matrix of 0's and 1's in which each row and column has a single 1.

27. Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ has an n-dimensional normal distribution. Show that if $k < n$, $\boldsymbol{W} = (X_1, X_2, ..., X_k)$ has a k-dimensional normal distribution. Identify the mean vector and the variance-covariance matrix.

28. Use the results of Exercise 26 and Exercise 27 to show that if $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ has an n-dimensional normal distribution and if $(i_1, i_2, ..., i_k)$ is a sequence of distinct indices, then $\boldsymbol{W} = \left(X_{i_1}, X_{i_2}, ..., X_{i_k}\right)$ has a k-dimensional normal distribution.

29. Suppose that $\boldsymbol{X}$ has an n-dimensional normal distribution, $\boldsymbol{a} \in \mathbb{R}^m$, and that $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ has linearly independent rows (thus, $m \le n$). Show that $\boldsymbol{Y} = \boldsymbol{a} + \boldsymbol{B}\,\boldsymbol{X}$ has an m-dimensional normal distribution. *Hint*: There exists an invertible $n \times n$ matrix $\boldsymbol{C}$ for which the first m rows are the rows of $\boldsymbol{B}$. Now use Exercise 25 and Exercise 27.

Note that the results in Exercises 25, 26, 27, and 28 are special cases of the result in Exercise 29.

30. Suppose that $\boldsymbol{X}$ has an n-dimensional normal distribution, $\boldsymbol{Y}$ has an m-dimensional normal distribution, and that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. Show that $(\boldsymbol{X}, \boldsymbol{Y})$ has an $m + n$ dimensional normal distribution. Identify the mean vector and the variance-covariance matrix.

31. Suppose that $\boldsymbol{X}$ is a random vector in $\mathbb{R}^m$, $\boldsymbol{Y}$ is a random vector in $\mathbb{R}^n$, and that $(\boldsymbol{X}, \boldsymbol{Y})$ has an $m + n$-dimensional normal distribution. Show that $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if and only if $\text{cov}(\boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{0}$ (the $m \times n$ zero matrix).