The Bivariate Normal Distribution

Most of the following discussion is taken from Wilks, *Statistical Methods in the Atmospheric Sciences*, section 4.5.

First, lets define the bivariate normal distribution for two related, normally distributed variables $x \sim \mathcal{N}(\mu_x, \sigma_x^2)$, and $x \sim \mathcal{N}(\mu_y, \sigma_y^2)$. Then, the bivariate normal distribution is defined by the following probability density function:

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

The bivariate normal PDF difines a surface in the x-y plane (see Figure 1). Like its one dimensional counterpart, the bivariate normal distribution has the following properties:

$$\int_{\mathcal{Y}} \int_{x} f(x, y) dx dy = 1 \tag{2}$$

$$f(x,y) \ge 0 \tag{3}$$

As might be inferred, the probability of observing a value x between $x_0 and x_1$, and y between y_0



Bivariate PDF based on CTI and SOI

Figure 1: Bivariate Normal PDF calculated for parameters based on the Cold Tongue Index (x axis) and the Southern Oscillation Index (y-axis).



Figure 2: Scatterplot of the standardized Cold Tongue Index (x axis) and the standardized Southern Oscillation Index (y-axis).

and y_1 (or, $P\{y_0 \le y \le y_1 \cap x_0 \le x \le x_1\}$) is just the volume integral under the surface f(x, y):

$$P\{y_0 \le y \le y_1 \cap x_0 \le x \le x_1\} = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y) dx dy$$
(4)

This integration is typically done numerically.

For normalized variables $z_x = (x - \mu_x)/\sigma_x$ and $z_y = (y - \mu_y)/\sigma_y$, the bivariate normal PDF becomes:

$$f(z_x, z_y) = \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left[-\frac{z_x^2 + z_y^2 - 2\rho z_x z_y}{2(1-\rho^2)}\right]$$
(5)

The bivariate standard normal distribution has a maximum at the origin. Note that the only parameter in the bivariate standard normal distribution is the correlation ρ between x and y. If xand y are independent ($\rho = 0$) then the surfaces of constant f(x, y) are concentric circles around the origin. As ρ increases, the distribution is stretched diagonally, forming elliptical isopleths with positive sloped major axes. For negative ρ , the major axes have a negative slope. Figure 1 depicts the bivariate standard normal distribution based on a correlation $\rho = -0.8$ between the Cold Tongue Index (CTI; SST averaged along the equator) and the Southern Oscillation Index (SOI; SLP at Tahiti minus SLP at Darwin, or equivalent). Figure 2 depicts the scatterplot of the CTI and the SOI, which graphically serves the same purpose as a histogram would for a univariate distribution.

A powerful feature of the bivariate normal distribution is that the conditional probability distribution function for one of the variables, given a known value for the other variable, is normally distributed (or, $f(x|y = y_0) \sim \mathcal{N}(\mu_{x|y=y_0}, \sigma_{x|y=y_0}^2)$). The conditional mean and variance of x, given that $y = y_0$ is:

$$\mu_{x|y=y_0} = \mu_x + \rho \sigma_x \frac{(y - \mu_y)}{\sigma_y} \tag{6}$$

$$\sigma_{x|y=y_0} = \sigma_x \sqrt{1-\rho^2} \tag{7}$$

From these parameters, we can determine the probability that x will fall in a given range $x_0 \le x \le x_1$ from the normal distribution. This point is illustrated in Figure 3.



Figure 3: Conditional Probability from the Bivariate Normal Distribution. Top: bivariate normal distribution based on a correlation of 0.4. Bottom: conditional distribution for variable x, given that variable y = 1.5. The shading indicates the probability that x will exceed 1.5 standard deviations if nothing is known about y (dark shading for lower curve), and if it is known that y = 1.5 (light shading for upper curve). Note the large difference in probabilities, though the correlation is not exceptionally high.

There are a couple of additional features to note here. First, the major axis of the ellipse defined by the bivariate normal distribution may be obtained by linear least squares regression. In this case, the least squares regression is taken to minimize the perpindicular distance between the data and the regression line. This can be accomplished with Empirical Orthogonal Function analysis, in that the regression analysis simply rotates the coordinate system. The leading eigenvector from the EOF analysis points in the direction of the major axis of the ellipse, and the distance from the maximum of the distribution (at μ_x, μ_y) to an isopleth of the distribution is given by $\chi^2_K(F) \cdot \sqrt{\lambda_1}$, where K = 2 (for a bivariate distribution), and F is the cumulative probability (the volume under the area bounded by the ellipse). The minor axis is defined in the direction of the second eigenvector (always perpindicular to the first), and has similarly defined distances to the isopleths of the distribution function. See Wilks, Example 9.4 (pp 372–373).