The bivariate normal distribution is an extension of the univariate normal distribution to the case of two variables. It is the basis of many procedures including Pearson product-moment correlation analysis.

Definitions

A pair of continuous random variables X and Y will constitute a bivariate distribution if there is a joint density function $f(x, y)$ which defines probabilities P as follows

$$f(x, y) \geq 0$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$$

$$\iint_R f(x, y)\, dx\, dy = P(X, Y \in R)$$

$$\int_a^b \int_c^d f(x, y)\, dx\, dy = P(a \leq X \leq b \text{ and } c \leq Y \leq d)$$

where R is an arbitrary region in the X, Y plane, while $a \leq X \leq b$ together with $c \leq Y \leq d$ defines a rectangular region in the X, Y plane.

The marginal distributions are defined as

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

while the conditional distributions are

$$f_{X|Y}(x) = f(x, y) / f_Y(y)$$
$$f_{Y|X}(y) = f(x, y) / f_X(x).$$

Of particular interest are the definitions of variances σ_x^2 and σ_y^2, and the covariance $Cov(X, Y)$ between X and Y in terms of the expectations $\mu_x = E[X]$ and $\mu_y = E[Y]$

$$\sigma_x^2 = E[X - E(X)]^2$$
$$= E[X^2] - E[X]^2$$
$$\sigma_y^2 = E[Y - E(Y)]^2$$
$$= E[Y^2] - E[Y]^2$$
$$Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$$
$$= E[XY] - E[X]E[Y]$$

as this leads to the correlation coefficient ρ

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

where $-1 < \rho < 1$, and the fact that when the two variables are independent $\rho = 0$. However, the condition $\rho = 0$ is not a sufficient condition that two variables are independent as this requires the stronger condition

$$f(x, y) = g(x)h(y).$$

So, if variables X and Y are jointly distributed as a bivariate normal distribution, the density function is

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

$$\text{where } Q = \frac{1}{1 - \rho^2} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x - \mu_x)(y - \mu_y)}{\sigma_x\sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right)$$

with $\sigma_x > 0, \sigma_y > 0$, and $-1 < \rho < 1$.

Here the marginal and conditional densities for X and Y are normal, as will be clear from the following

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} \exp -\frac{1}{2}\left(\frac{x - \mu_x}{\sigma_x}\right)^2$$

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

with corresponding expressions for $f_Y(y)$ and $f_{Y|X}(y)$ by symmetry. At fixed probability levels, the quadratic form Q defines an ellipse in the X, Y plane which will have axes parallel to the X, Y axes if $\rho = 0$, but with rotated axes otherwise. Note that the marginal distribution for X has mean μ_x and variance σ_x^2, and the marginal distribution of Y has mean μ_y and variance σ_y^2, but also consider the expressions for the means in the conditional distributions, namely

$$m_{X|Y} = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$$

$$m_{Y|X} = \mu_y + \rho\frac{\sigma_y}{\sigma_x}(y - \mu_x)$$

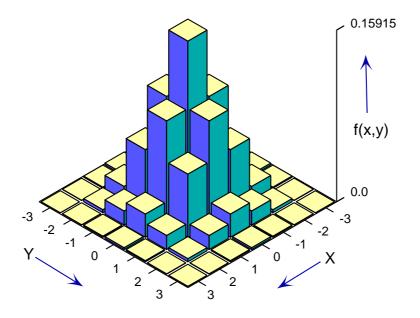which will be mentioned later.

Using program MAKDAT to simulate f(x,y)

From the SIMFIT main menu select [A/Z], then run program **makdat** choosing the bivariate normal distribution. Using the parameters

$$\mu_x = \mu_y = 0, \ \sigma_x = \sigma_y = 1, \ \rho = 0, \ -3 \le x \le 3, \ -3 \le y \le 3, \ n_x = n_y = 7$$

allows the creation of the following three dimensional skyscraper plot for a bivariate normal distribution, where the volume of the individual bars indicates the probability of a x, y pair of random variables occurring in the area of the X, Y plane at the base of the bars. Increasing the number of divisions to $n_x = n_y = 50$ as in the subsequent plot indicates how the three dimensional bar chart converges to a smooth surface as the number of divisions increases.
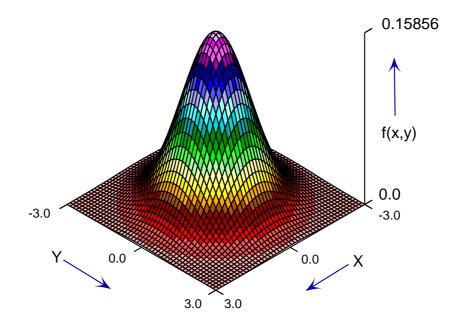
Bivariate Normal Distribution
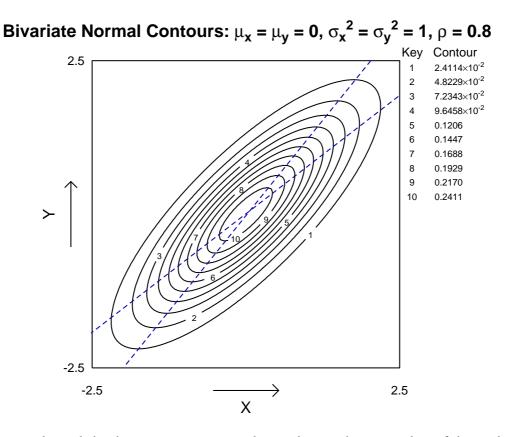## $\mu_x = \mu_y = 0, \sigma_x = \sigma_y = 1, \rho = 0$

Bivariate Normal Distribution
$\mu_x = \mu_y = 0, \sigma_x = \sigma_y = 1, \rho = 0$

3

Plotting bivariate normal contour diagrams

Contour diagrams are at the basis of checking scatter diagrams when using Pearson product-moment analysis, and especially the practise of plotting lines on such plots to indicate linear correlation. From SimF$_I$T program **makdat** simply change the the current value of $\rho = 0$ which models the absence of correlation to a value such $\rho = 0.8$ for positive correlation then create the contour diagram below.

Bivariate Normal Contours: $\mu_x = \mu_y = 0$, $\sigma_x^2 = \sigma_y^2 = 1$, $\rho = 0.8$

Key	Contour
1	2.4114×10^{-2}
2	4.8229×10^{-2}
3	7.2343×10^{-2}
4	9.6458×10^{-2}
5	0.1206
6	0.1447
7	0.1688
8	0.1929
9	0.2170
10	0.2411

This diagram also includes the two mean regression lines indicating the mean values of the conditional distributions as functions of their arguments. These lines show the limiting positions that would be obtained with a swarm of random points following a bivariate normal distribution when plotting best-fit lines for Y as a function of X and X as a function of Y. The lines are

$$\text{For } m_{Y|X}: \quad y = \rho \frac{\sigma_x}{\sigma_y} x + \left\{ \mu_y - \rho \frac{\sigma_y}{\sigma_x} \mu_x \right\}$$

$$\text{For } m_{X|Y}: \quad y = \frac{1}{\rho} \frac{\sigma_y}{\sigma_x} x + \left\{ \mu_y - \frac{1}{\rho} \frac{\sigma_x}{\sigma_y} \mu_x \right\}$$

where it is understood that $\rho^2 < 1$ and $\rho \neq 0$.

Given two lines with slopes α and β then the angle between them (subject to the usual sign conventions) is given by $\tan \theta = (\alpha - \beta)/(1 + \alpha\beta)$. So, for the above conditional mean regression lines, we have

$$\tan \theta = \frac{1}{2} \left(\rho \frac{\sigma_x}{\sigma_y} - \frac{1}{\rho} \frac{\sigma_y}{\sigma_x} \right)$$

which, apart from the singular case with θ as a right angle when $\rho = 0$, defines $|\theta|$ as a decreasing function of ρ^2 thereafter.