Definition **Joint moment generating function** The *joint moment generating function* of $(X_1, \ldots, X_k)$ is defined by

$$m_{X_1,\ldots,X_k}(t_1, \ldots, t_k) = \mathscr{E}\left[\exp \sum_{j=1}^{k} t_j X_j\right],$$

if the expectation exists for all values of $t_1, \ldots, t_k$ such that $-h < t_j < h$ for some $h > 0$, $j = 1, \ldots, k$. ////

Remark $m_X(t_1) = m_{X,Y}(t_1, 0) = \lim_{t_2 \to 0} m_{X,Y}(t_1, t_2)$, and $m_Y(t_2) = m_{X,Y}(0, t_2) = \lim_{t_1 \to 0} m_{X,Y}(t_1, t_2)$; that is, the marginal moment generating functions can be obtained from the joint moment generating function. ////

The rth moment of X_j may be obtained from $m_{X_1, \ldots, X_k}(t_1, \ldots, t_k)$ by differentiating it r times with respect to t_j and then taking the limit as all the t's approach 0. Also $\mathscr{E}[X_i^r X_j^s]$ can be obtained by differentiating the joint moment generating function r times with respect to t_i and s times with respect to t_j and then taking the limit as all the t's approach 0. Similarly other joint raw moments can be generated.

4.5 Independence and expectations

Theorem If X and Y are independent and $g_1(\cdot)$ and $g_2(\cdot)$ are two functions, each of a single argument, then

$$\mathscr{E}[g_1(X)g_2(Y)] = \mathscr{E}[g_1(X)] \cdot \mathscr{E}[g_2(Y)].$$

PROOF We will give the proof for jointly continuous random variables.

$$\begin{aligned}
\mathscr{E}[g_1(X)g_2(Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_{X,Y}(x, y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g_1(x)g_2(y)f_X(x)f_Y(y)\,dx\,dy \\
&= \int_{-\infty}^{\infty} g_1(x)f_X(x)\,dx \cdot \int_{-\infty}^{\infty} g_2(y)f_Y(y)\,dy \\
&= \mathscr{E}[g_1(X)] \cdot \mathscr{E}[g_2(Y)].
\end{aligned}$$

////

Corollary If X and Y are independent, then $\text{cov}[X, Y] = 0$.

PROOF Take $g_1(x) = x - \mu_X$ and $g_2(y) = y - \mu_Y$;

$$\begin{aligned}\text{cov}[X, Y] &= \mathscr{E}[(X - \mu_X)(Y - \mu_Y)] = \mathscr{E}[g_1(X)g_2(Y)] \\ &= \mathscr{E}[g_1(X)]\mathscr{E}[g_2(Y)] \\ &= \mathscr{E}[X - \mu_X] \cdot \mathscr{E}[Y - \mu_Y] = 0 \qquad \text{since } \mathscr{E}[X - \mu_X] = 0. \qquad ////\end{aligned}$$

Definition **Uncorrelated random variables** Random variables X and Y are defined to be *uncorrelated* if and only if $\text{cov}[X, Y] = 0$. ////

Remark The converse of the above corollary is not always true; that is, $\text{cov}[X, Y] = 0$ does not always imply that X and Y are independent.

EXAMPLE Let U be a random variable which is uniformly distributed over the interval (0, 1). Define $X = \sin 2\pi U$ and $Y = \cos 2\pi U$. X and Y are clearly not independent since if a value of X is known, then U is one of two values, and so Y is also one of two values; hence the conditional distribution of Y is not the same as the marginal distribution. $\mathscr{E}[Y] = \int_0^1 \cos 2\pi u \, du = 0$, and $\mathscr{E}[X] = \int_0^1 \sin 2\pi u \, du = 0$; so cov $[X, Y] = \mathscr{E}[XY] = \int_0^1 \sin 2\pi u \cos 2\pi u \, du = \frac{1}{2} \int_0^1 \sin 4\pi u \, du = 0$. ////

Theorem Two jointly distributed random variables X and Y are independent if and only if $m_{X,Y}(t_1, t_2) = m_X(t_1)m_Y(t_2)$ for all t_1, t_2 for which $-h < t_i < h$, $i = 1, 2$, for some $h > 0$.

PROOF [Recall that $m_X(t_1)$ is the moment generating function of X. Also note that $m_X(t_1) = m_{X,Y}(t_1, 0)$.] X and Y independent imply that the joint moment generating function factors into the product of the marginal moment generating functions by taking $g_1(x) = e^{t_1 x}$ and $g_2(y) = e^{t_2 y}$. The proof in the other direction will be omitted.

////

Cauchy-Schwarz inequality

Theorem 1 Cauchy-Schwarz inequality Let X and Y have finite second moments; then $(\mathcal{E}[XY])^2 = |\mathcal{E}[XY]|^2 \leq \mathcal{E}[X^2]\mathcal{E}[Y^2]$, with equality if and only if $P[Y = cX] = 1$ for some constant c.

PROOF The existence of expectations $\mathcal{E}[X]$, $\mathcal{E}[Y]$, and $\mathcal{E}[XY]$ follows from the existence of expectations $\mathcal{E}[X^2]$ and $\mathcal{E}[Y^2]$. Define $0 \leq h(t) = \mathcal{E}[(tX - Y)^2] = \mathcal{E}[X^2]t^2 - 2\mathcal{E}[XY]t + \mathcal{E}[Y^2]$. Now $h(t)$ is a quadratic function in t which is greater than or equal to 0. If $h(t) > 0$, then the roots of $h(t)$ are not real; so $4(\mathcal{E}[XY])^2 - 4\mathcal{E}[X^2]\mathcal{E}[Y^2] < 0$, or $(\mathcal{E}[XY])^2 < \mathcal{E}[X^2]\mathcal{E}[Y^2]$. If $h(t) = 0$ for some t, say t_0, then $\mathcal{E}[(t_0 X - Y)^2] = 0$, which implies $P[t_0 X = Y] = 1$. ////

Corollary $|\rho_{X,Y}| \leq 1$, with equality if and only if one random variable is a linear function of the other with probability 1.

PROOF Rewrite the Cauchy-Schwarz inequality as $|\mathscr{E}[UV]| \leq \sqrt{\mathscr{E}[U^2]\mathscr{E}[V^2]}$, and set $U = X - \mu_X$ and $V = Y - \mu_Y$. ////

5 Highlight: The bivariate normal distribution

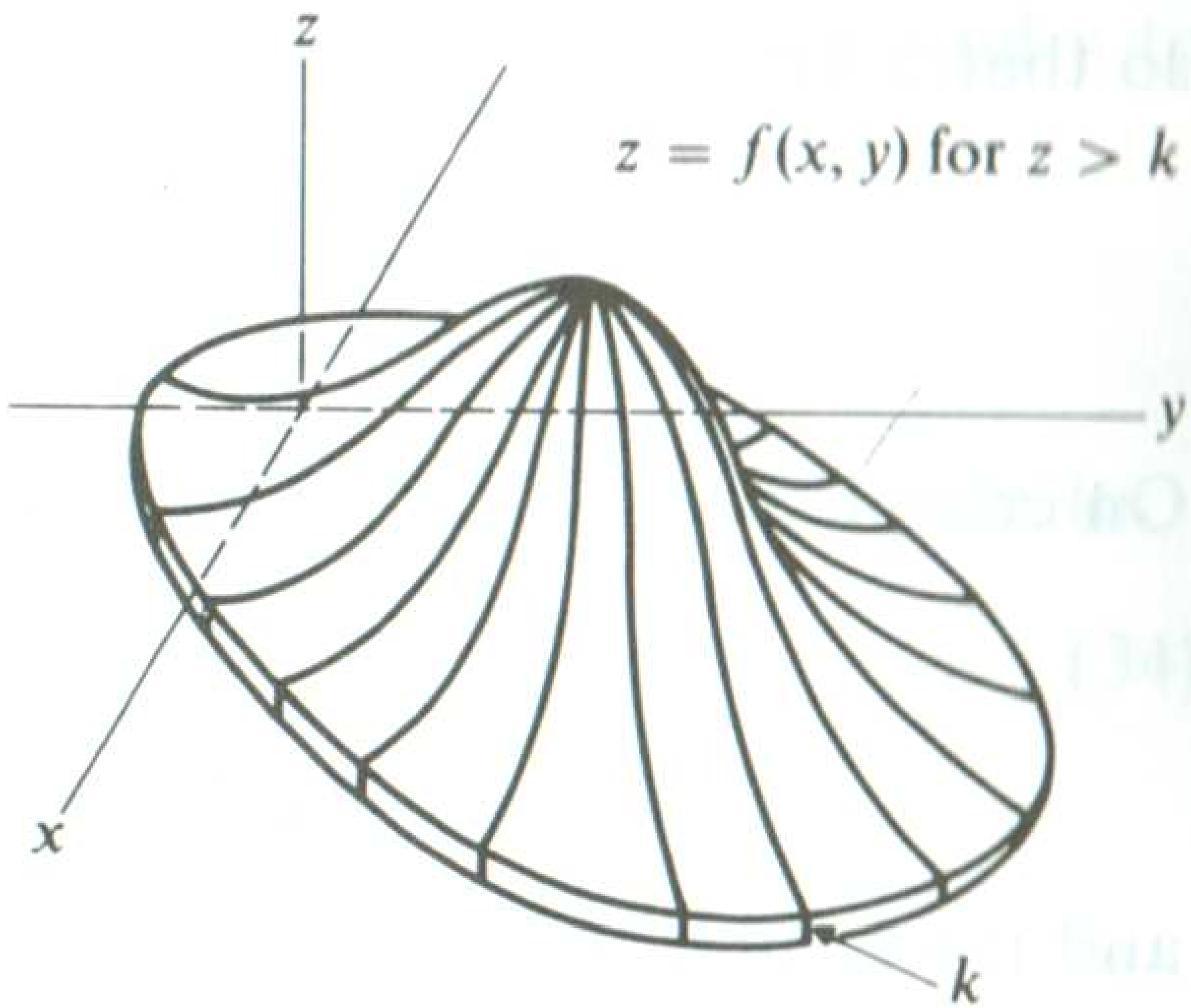
5.1 Density function

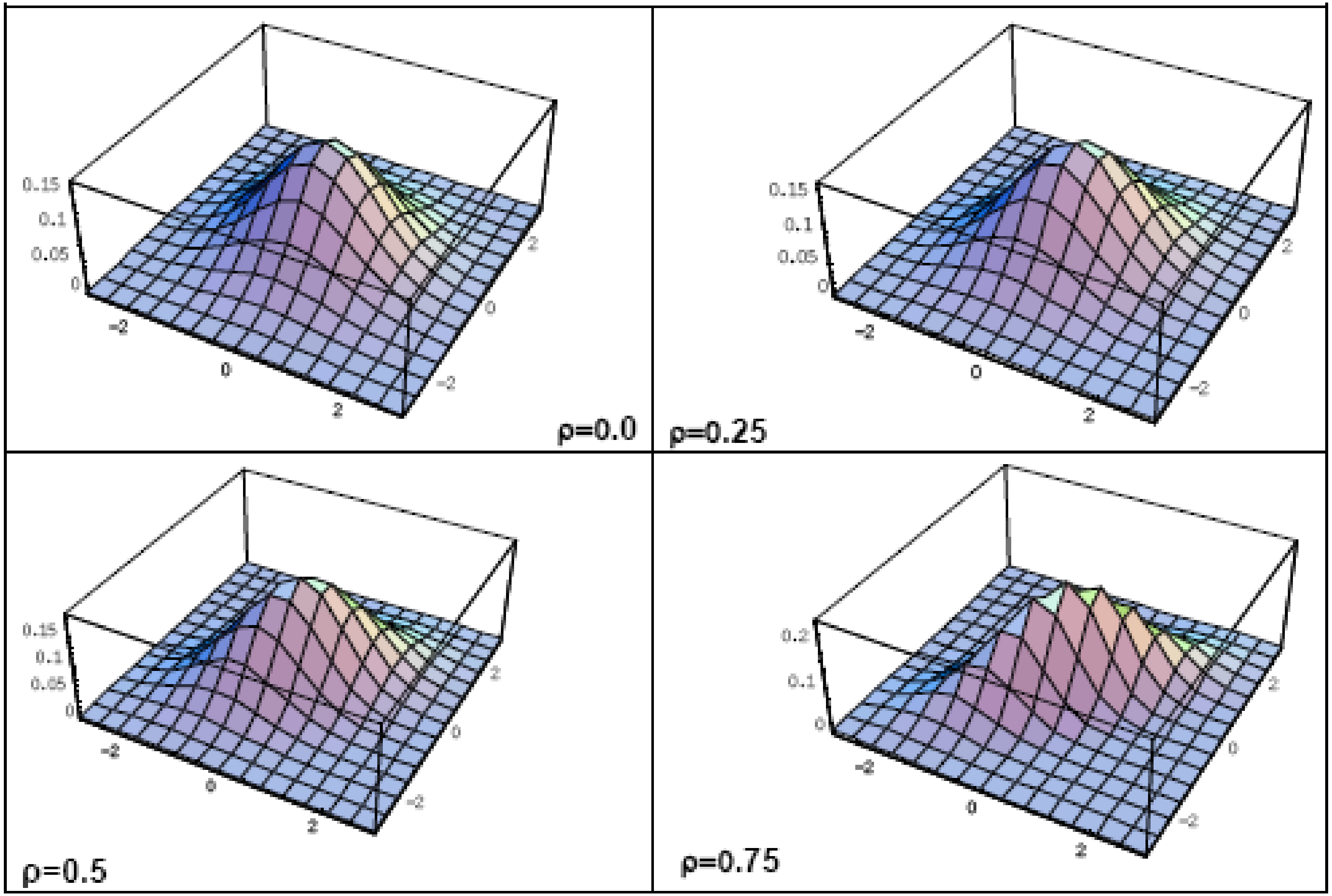
Definition Bivariate normal distribution Let the two-dimensional random variable (X, Y) have the joint probability density function

$$f_{X,Y}(x, y) = f(x, y) = \frac{1}{2\pi\sigma_X \sigma_Y \sqrt{1-\rho^2}}$$

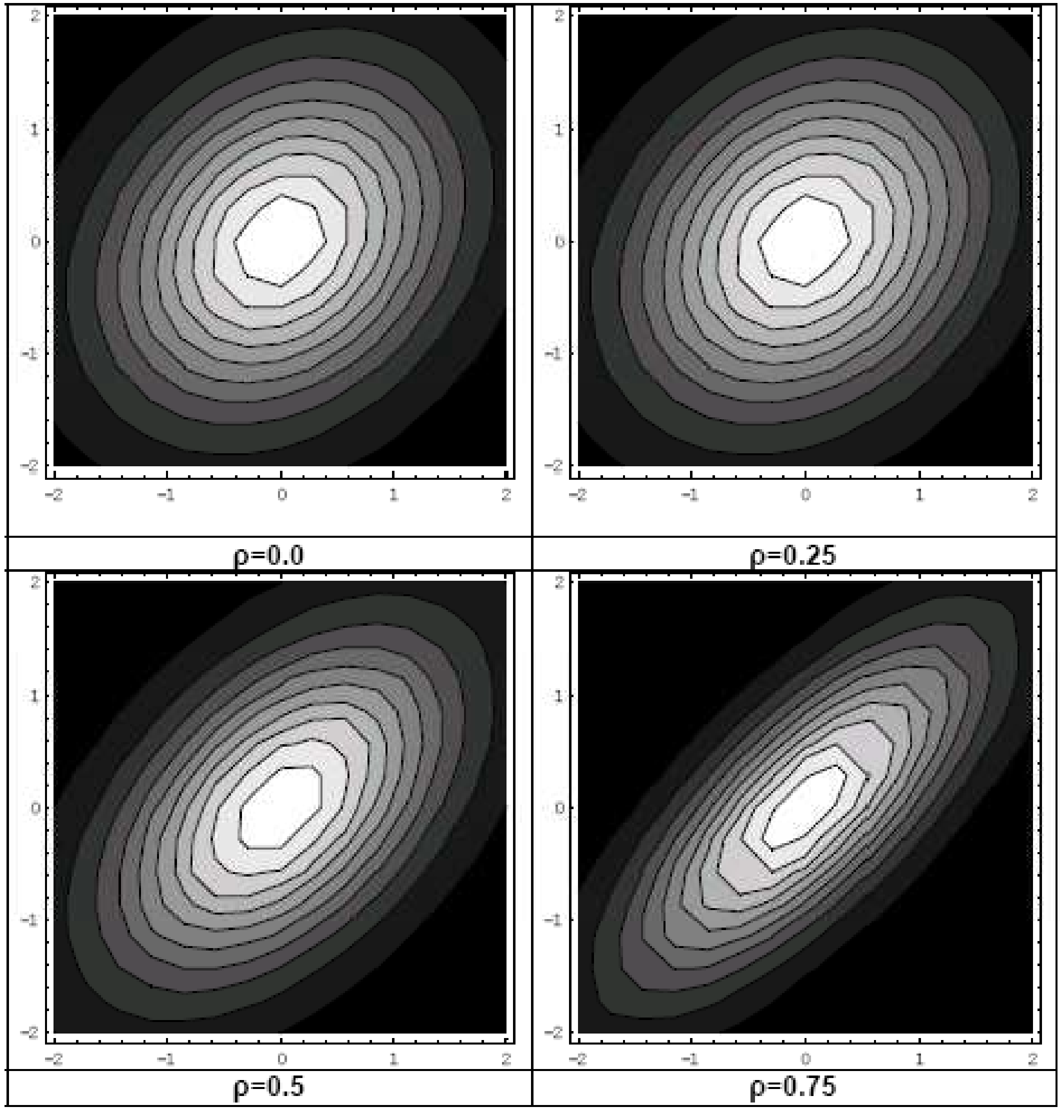
$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho \frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}$$

for $-\infty < x < \infty$, $-\infty < y < \infty$, where σ_Y, σ_X, μ_X, μ_Y, and ρ are constants such that $-1 < \rho < 1$, $0 < \sigma_Y$, $0 < \sigma_X$, $-\infty < \mu_X < \infty$, and $-\infty < \mu_Y < \infty$. Then the random variable (X, Y) is defined to have a *bivariate normal distribution*. ////


z
z = f(x, y) for z > k
y
x
k



Standard bivariate normal plots for ρ = 0.0, 0.25, 0.5 and 0.75



Standard bivariate normal contours plots for ρ=0.0, 0.25, 0.5 & 0.75

5.2 Moment generating function and moments

To obtain the moments of X and Y, we shall find their joint moment generating function, which is given by

$$m_{X,Y}(t_1, t_2) = m(t_1, t_2) = \mathscr{E}[e^{t_1 X + t_2 Y}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1 x + t_2 y} f(x, y)\, dy\, dx.$$

Theorem 12 The moment generating function of the bivariate normal distribution is

$$m(t_1, t_2) = \exp[t_1 \mu_X + t_2 \mu_Y + \tfrac{1}{2}(t_1^2 \sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)]. \qquad (33)$$

PROOF Let us again substitute for x and y in terms of u and v to obtain

$$m(t_1, t_2)$$
$$= e^{t_1\mu_X + t_2\mu_Y} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1\sigma_X u + t_2\sigma_Y v} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-[\frac{1}{2}/(1-\rho^2)](u^2 - 2\rho uv + v^2)}\, dv\, du.$$

The combined exponents in the integrand may be written

$$-\frac{1}{2(1-\rho^2)}[u^2 - 2\rho uv + v^2 - 2(1-\rho^2)t_1\sigma_X u - 2(1-\rho^2)t_2\sigma_Y v],$$

and on completing the square first on u and then on v, we find this expression becomes

$$-\frac{1}{2(1-\rho^2)}\{[u - \rho v - (1-\rho^2)t_1\sigma_X]^2 + (1-\rho^2)(v - \rho t_1\sigma_X - t_2\sigma_Y)^2$$
$$- (1-\rho^2)(t_1^2\sigma_X^2 + 2\rho t_1 t_2\sigma_X\sigma_Y + t_2^2\sigma_Y^2)\},$$

which, if we substitute

$$w = \frac{u - \rho v - (1-\rho^2)t_1\sigma_X}{\sqrt{1-\rho^2}} \qquad \text{and} \qquad z = v - \rho t_1 \sigma_X - t_2 \sigma_Y,$$

becomes

$$-\tfrac{1}{2}w^2 - \tfrac{1}{2}z^2 + \tfrac{1}{2}(t_1^2\sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2),$$

and the integral may be written

$$m(t_1, t_2) = e^{t_1\mu_X + t_2\mu_Y} \exp[\tfrac{1}{2}(t_1^2\sigma_X^2 + 2\rho t_1 t_2 \sigma_X \sigma_Y + t_2^2 \sigma_Y^2)]$$
$$\times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-w^2/2 - z^2/2}\, dw\, dz$$
$$= \exp[t_1\mu_X + t_2\mu_Y + \tfrac{1}{2}(t_1^2\sigma_X^2 + 2\rho t_1 t_2 \sigma_X\sigma_Y + t_2^2 \sigma_Y^2)]$$

since the double integral is equal to unity. ////

Theorem If (X, Y) has bivariate normal distribution, then

$$\mathscr{E}[X] = \mu_X,$$
$$\mathscr{E}[Y] = \mu_Y,$$
$$\text{var}[X] = \sigma_X^2,$$
$$\text{var}[Y] = \sigma_Y^2,$$
$$\text{cov}[X, Y] = \rho\sigma_X\sigma_Y,$$

and

$$\rho_{X,Y} = \rho.$$

PROOF The moments may be obtained by evaluating the appropriate derivative of $m(t_1, t_2)$ at $t_1 = 0$, $t_2 = 0$. Thus,

$$\mathscr{E}[X] = \left.\frac{\partial m}{\partial t_1}\right|_{t_1, t_2 = 0} = \mu_X$$

$$\mathscr{E}[X^2] = \left.\frac{\partial^2 m}{\partial t_1^2}\right|_{t_1, t_2 = 0} = \mu_X^2 + \sigma_X^2 .$$

Hence the variance of X is

$$\mathscr{E}[(X - \mu_X)^2] = \mathscr{E}[X^2] - \mu_X^2 = \sigma_X^2 .$$

Similarly, on differentiating with respect to t_2, one finds the mean and variance of Y to be μ_Y and σ_Y^2. We can also obtain joint moments

$$\mathscr{E}[X^r Y^s]$$

by differentiating $m(t_1, t_2)$ r times with respect to t_1 and s times with respect to t_2 and then putting t_1 and t_2 equal to 0. The covariance of X and Y is

$$\begin{aligned}\mathscr{E}[(X-\mu_X)(Y-\mu_Y)] &= \mathscr{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= \mathscr{E}[XY] - \mu_X\mu_Y \\ &= \frac{\partial^2}{\partial t_1\,\partial t_2} m(t_1, t_2)\bigg|_{t_1=t_2=0} - \mu_X\mu_Y \\ &= \rho\sigma_X\sigma_Y .\end{aligned}$$

Hence, the parameter ρ is the correlation coefficient of X and Y. ////

Theorem 14 If (X, Y) has a bivariate normal distribution, then X and Y are independent if and only if X and Y are uncorrelated.

PROOF X and Y are uncorrelated if and only if $\text{cov}[X, Y] = 0$ or, equivalently, if and only if $\rho_{X,Y} = \rho = 0$. It can be observed that if $\rho = 0$, the joint density $f(x, y)$ becomes the product of two univariate normal distributions; so that $\rho = 0$ implies X and Y are independent. We know that, in general, independence of X and Y implies that X and Y are uncorrelated. ////

5.3 Marginal and conditional densities

Theorem If (X, Y) has a bivariate normal distribution, then the marginal distributions of X and Y are univariate normal distributions; that is, X is normally distributed with mean μ_X and variance σ_X^2, and Y is normally distributed with mean μ_Y and variance σ_Y^2.

PROOF The marginal density of one of the variables X, for example, is by definition

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

and again substituting

$$v = \frac{y - \mu_Y}{\sigma_Y}$$

and completing the square on v, one finds that

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_X\sqrt{1-\rho^2}} \times \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - \frac{1}{2(1-\rho^2)}\left(v - \rho\,\frac{x-\mu_X}{\sigma_X}\right)^2\right] dv.$$

Then the substitutions

$$w = \frac{v - \rho(x - \mu_X)/\sigma_X}{\sqrt{1-\rho^2}} \qquad \text{and} \qquad dw = \frac{dv}{\sqrt{1-\rho^2}}$$

show at once that

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right],$$

the univariate normal density. Similarly the marginal density of Y may be found to be

$$f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]. \qquad ////$$

Theorem If (X, Y) has a bivariate normal distribution, then the conditional distribution of X given $Y = y$ is normal with mean $\mu_X + (\rho\sigma_X/\sigma_Y)(y - \mu_Y)$ and variance $\sigma_X^2(1 - \rho^2)$. Also, the conditional distribution of Y given $X = x$ is normal with mean $\mu_Y + (\rho\sigma_Y/\sigma_X)(x - \mu_X)$ and variance $\sigma_Y^2(1 - \rho^2)$.

PROOF The conditional distributions are obtained from the joint and marginal distributions. Thus, the conditional density of X for fixed values of Y is

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)},$$

and, after substituting, the expression may be put in the form

$$f_{X|Y}(x|y) = \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x - \mu_X - \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y)\right]^2\right\},$$

which is a univariate normal density with mean $\mu_X + (\rho\sigma_X/\sigma_Y)(y - \mu_Y)$ and with variance $\sigma_X^2(1 - \rho^2)$. The conditional distribution of Y may be obtained by interchanging x and y to get

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left[y - \mu_Y - \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)\right]^2\right\}.$$

////

Link with regression analysis

As we already noted, the mean value of a random variable in a conditional distribution is called a *regression curve* when regarded as a function of the fixed variable in the conditional distribution. Thus the regression for X on $Y = y$ in Eq.) is $\mu_X + (\rho\sigma_X/\sigma_Y)(y - \mu_Y)$, which is a linear function of y in the present case. For bivariate distributions in general, the mean of X in the conditional density of X given $Y = y$ will be some function of y, say $g(\cdot)$, and the equation

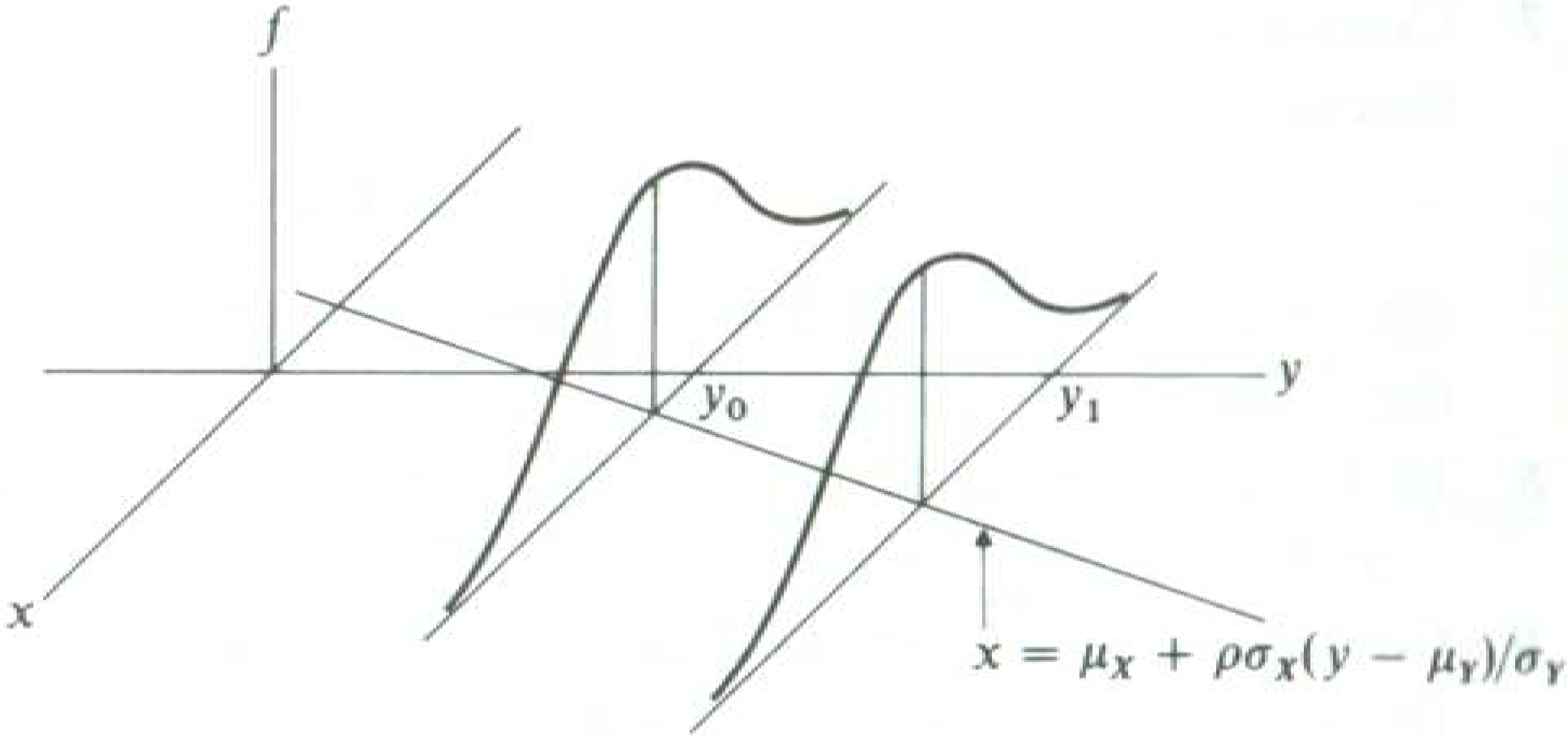
$$x = g(y)$$

when plotted in the xy plane gives the regression curve for X. It is simply a curve which gives the location of the mean of X for various values of Y in the conditional density of X given $Y = y$.

For the bivariate normal distribution, the regression curve is the straight line obtained by plotting

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

as shown in Fig The conditional density of X given $Y = y$, $f_{X|Y}(x|y)$, is also plotted in Fig for two particular values y_0 and y_1 of Y.


f
y
y_0
y_1
x
$x = \mu_X + \rho\sigma_X(y - \mu_Y)/\sigma_Y$

PART 2: Distributions for functions of random variables

1 Introduction

- Real-life examples often present themselves with far more complex density functions that the one described so far.
- In many cases the random variable of interest is a function of one that we know better, or for which we are better able to describe its density or distributional properties
- For this reason, we devote an entire part on densities of “functions of random variables”

2 Expectations of functions of random variables

2.1 Expectation two ways

An expectation of a function of a set of random variables can be obtained two different ways. To illustrate, consider a function of just one random variable, say X. Let $g(\cdot)$ be the function, and set $Y = g(X)$. Since Y is a random variable, $\mathscr{E}[Y]$ is defined (if it exists), and $\mathscr{E}[g(X)]$ is defined (if it exists). For instance, if X and $Y = g(X)$ are continuous random variables, then by definition

$$\mathscr{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy, \qquad (1)$$

and

$$\mathscr{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx; \qquad (2)$$

but $Y = g(X)$, so it seems reasonable that $\mathscr{E}[Y] = \mathscr{E}[g(X)]$. This can, in fact, be proved; although we will not bother to do it.

Thus we have two ways of calculating the expectation of $Y = g(X)$; one is to average Y with respect to the density of Y, and the other is to average $g(X)$ with respect to the density of X.

In general, for given random variables $X_1, \ldots, X_n$, let $Y = g(X_1, \ldots, X_n)$; then $\mathscr{E}[Y] = \mathscr{E}[g(X_1, \ldots, X_n)]$, where (for jointly continuous random variables)

$$\mathscr{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy \tag{3}$$

and

$$\mathscr{E}[g(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n. \tag{4}$$

EXAMPLE Let X be a standard normal random variable, and let $g(x) = x^2$. For $Y = g(X) = X^2$,

$$\mathscr{E}[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy,$$

and

$$\mathscr{E}[g(X)] = \mathscr{E}[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\, dx.$$

Now

$$\mathscr{E}[X^2] = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\, dx = 1$$

$$\mathscr{E}[Y] = \int_{0}^{\infty} y \frac{1}{\Gamma(1/2)} (1/2)^{\frac{1}{2}} y^{-\frac{1}{2}} e^{-\frac{1}{2}y}\, dy = 1,$$

using the fact that Y has a gamma distribution with parameters $r = \frac{1}{2}$ and $\lambda = \frac{1}{2}$.

////

2.2 Sums of random variables

Theorem For random variables $X_1, \ldots, X_n$

$$\mathscr{E}\left[\sum_1^n X_i\right] = \sum_1^n \mathscr{E}[X_i], \tag{5}$$

and

$$\operatorname{var}\left[\sum_1^n X_i\right] = \sum_1^n \operatorname{var}[X_i] + 2\sum_{i<j}\sum \operatorname{cov}[X_i, X_j]. \tag{6}$$

PROOF That $\mathscr{E}\left[\sum_1^n X_i\right] = \sum_1^n \mathscr{E}[X_i]$ follows from a property of expectation

$$\operatorname{var}\left[\sum_1^n X_i\right] = \mathscr{E}\left[\left(\sum_1^n X_i - \mathscr{E}\left[\sum_1^n X_i\right]\right)^2\right] = \mathscr{E}\left[\left(\sum_1^n (X_i - \mathscr{E}[X_i])\right)^2\right]$$

$$= \mathscr{E}\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - \mathscr{E}[X_i])(X_j - \mathscr{E}[X_j])\right]$$

$$= \sum_{i=1}^n \sum_{j=1}^n \mathscr{E}[(X_i - \mathscr{E}[X_i])(X_j - \mathscr{E}[X_j])]$$

$$= \sum_{i=1}^n \operatorname{var}[X_i] + 2 \sum_{i<j}\sum \operatorname{cov}[X_i, X_j].$$

////

Corollary If $X_1, \ldots, X_n$ are uncorrelated random variables, then

$$\text{var}\left[\sum_1^n X_i\right] = \sum_1^n \text{var}[X_i]. \qquad ////$$

Theorem Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be two sets of random variables, and let $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$ be two sets of constants; then

$$\text{cov}\left[\sum_1^n a_i X_i, \sum_1^m b_j Y_j\right] = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}[X_i, Y_j]. \tag{7}$$

////

Corollary If $X_1, \ldots, X_n$ are random variables and $a_1, \ldots, a_n$ are constants, then

$$\operatorname{var}\left[\sum_1^n a_i X_i\right] = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \operatorname{cov}[X_i, X_j] \qquad (8)$$

$$= \sum_{i=1}^n a_i^2 \operatorname{var}[X_i] + \sum_{i \neq j}\sum a_i a_j \operatorname{cov}[X_i, X_j].$$

In particular, if $X_1, \ldots, X_n$ are independent and identically distributed random variables with mean μ_X and variance σ_X^2 and if $\bar{X}_n = (1/n) \sum_1^n X_i$, then

$$\mathscr{E}[\bar{X}_n] = \mu_X, \qquad \text{and} \qquad \operatorname{var}[\bar{X}_n] = \frac{\sigma_X^2}{n}. \qquad (9)$$

PROOF Let $m = n$, $Y_i = X_i$, and $b_i = a_i$, $i = 1, \ldots, n$ in the above theorem; then

$$\operatorname{var}\left[\sum_1^n a_i X_i\right] = \operatorname{cov}\left[\sum_1^n a_i X_i, \sum_1^m b_j Y_j\right],$$

and Eq. (8) follows from Eq. (7). To obtain the variance part of Eq. (9) from Eq. (8), set $a_i = 1/n$ and $\sigma_X^2 = \operatorname{var}[X_i]$. The mean part of Eq. (9) is routinely derived as

$$\mathscr{E}[\overline{X}_n] = \frac{1}{n}\mathscr{E}\left[\sum_1^n X_i\right] = \frac{1}{n}\sum_1^n \mathscr{E}[X_i] = \frac{1}{n}\sum_1^n \mu_X = \mu_X. \qquad ////$$

2.3 Product and quotient

Theorem 3 Let X and Y be two random variables for which var $[XY]$ exists; then

$$\mathscr{E}[XY] = \mu_X \mu_Y + \text{cov}[X, Y], \tag{12}$$

and

$$\begin{aligned}\text{var}[XY] &= \mu_Y^2 \,\text{var}[X] + \mu_X^2 \,\text{var}[Y] + 2\mu_X \mu_Y \,\text{cov}[X, Y] \\ &\quad - (\text{cov}[X, Y])^2 + \mathscr{E}[(X - \mu_X)^2 (Y - \mu_Y)^2] + \\ &\quad 2\mu_Y \mathscr{E}[(X - \mu_X)^2 (Y - \mu_Y)] + 2\mu_X \mathscr{E}[(X - \mu_X)(Y - \mu_Y)^2].\end{aligned} \tag{13}$$

PROOF

$$XY = \mu_X \mu_Y + (X - \mu_X)\mu_Y + (Y - \mu_Y)\mu_X + (X - \mu_X)(Y - \mu_Y).$$

Calculate $\mathscr{E}[XY]$ and $\mathscr{E}[(XY)^2]$ to get the desired results. ////

Corollary If X and Y are independent, $\mathscr{E}[XY] = \mu_X \mu_Y$, and var $[XY] = \mu_Y^2$ var $[X] + \mu_X^2$ var $[Y]$ + var $[X]$ var $[Y]$.

PROOF If X and Y are independent,

$$\mathscr{E}[(X - \mu_X)^2(Y - \mu_Y)^2] = \mathscr{E}[(X - \mu_X)^2]\mathscr{E}[(Y - \mu_Y)^2]$$
$$= \text{var}[X] \text{ var}[Y],$$
$$\mathscr{E}[(X - \mu_X)^2(Y - \mu_Y)] = \mathscr{E}[(X - \mu_X)^2]\mathscr{E}[Y - \mu_Y] = 0,$$

and

$$\mathscr{E}[(X - \mu_X)(Y - \mu_Y)^2] = 0.$$ ////

Theorem 4

$$\mathscr{E}\left[\frac{X}{Y}\right] \approx \frac{\mu_X}{\mu_Y} - \frac{1}{\mu_Y^2}\operatorname{cov}[X, Y] + \frac{\mu_X}{\mu_Y^3}\operatorname{var}[Y], \tag{14}$$

and

$$\operatorname{var}\left[\frac{X}{Y}\right] \approx \left(\frac{\mu_X}{\mu_Y}\right)^2 \left(\frac{\operatorname{var}[X]}{\mu_X^2} + \frac{\operatorname{var}[Y]}{\mu_Y^2} - \frac{2\operatorname{cov}[X, Y]}{\mu_X \mu_Y}\right). \tag{15}$$

PROOF To find the approximate formula for $\mathscr{E}[X/Y]$, consider the Taylor series expansion of x/y expanded about (μ_X, μ_Y); drop all terms of order higher than 2, and then take the expectation of both sides. The approximate formula for var $[X/Y]$ is similarly obtained by expanding in a Taylor series and retaining only second-order terms. ////

Important consequence

the method of proof of Theorem 4 can be used to find approximate formulas for the mean and variance of functions of X and Y other than the quotient. For example,

$$\mathscr{E}[g(X, Y)] \approx g(\mu_X, \mu_Y) + \frac{1}{2}\operatorname{var}[X]\frac{\partial^2}{\partial x^2}g(x, y)\bigg|_{\mu_X, \mu_Y} + \frac{1}{2}\operatorname{var}[Y]\frac{\partial^2}{\partial y^2}g(x, y)\bigg|_{\mu_X, \mu_Y} + \operatorname{cov}[X, Y]\frac{\partial^2}{\partial y\,\partial x}g(x, y)\bigg|_{\mu_X, \mu_Y}, \tag{16}$$

and

$$\operatorname{var}[g(X, Y)] \approx \operatorname{var}[X]\left\{\frac{\partial}{\partial x}g(x, y)\bigg|_{\mu_X, \mu_Y}\right\}^2 + \operatorname{var}[Y]\left\{\frac{\partial}{\partial y}g(x, y)\bigg|_{\mu_X, \mu_Y}\right\}^2 + 2\operatorname{cov}[X, Y]\left\{\frac{\partial}{\partial x}g(x, y)\bigg|_{\mu_X, \mu_Y}\cdot\frac{\partial}{\partial y}g(x, y)\bigg|_{\mu_X, \mu_Y}\right\}. \tag{17}$$

3 Cumulative-distribution-function technique

3.1 Description of the technique

If the joint distribution of random variables $X_1, \ldots, X_n$ is given, then, theoretically, the joint distribution of random variables of $Y_1, \ldots, Y_k$ can be determined, where $Y_j = g_j(X_1, \ldots, X_n)$, $j = 1, \ldots, k$ for given functions $g_1(\cdot, \ldots, \cdot), \ldots, g_k(\cdot, \ldots, \cdot)$. By definition, the joint cumulative distribution function of $Y_1, \ldots, Y_k$ is $F_{Y_1, \ldots, Y_k}(y_1, \ldots, y_k) = P[Y_1 \le y_1; \ldots; Y_k \le y_k]$. But for each $y_1, \ldots, y_k$ the event $\{Y_1 \le y_1; \ldots; Y_k \le y_k\} \equiv \{g_1(X_1, \ldots, X_n) \le y_1; \ldots; g_k(X_1, \ldots, X_n) \le y_k\}$. This latter event is an event described in terms of the given functions $g_1(\cdot, \ldots, \cdot), \ldots, g_k(\cdot, \ldots, \cdot)$ and the given random variables $X_1, \ldots, X_n$. Since the joint distribution of $X_1, \ldots, X_n$ is assumed given, presumably the probability of event $\{g_1(X_1, \ldots, X_n) \le y_1; \ldots; g_k(X_1, \ldots, X_n) \le y_k\}$ can be calculated and consequently $F_{Y_1, \ldots, Y_k}(\cdot, \ldots, \cdot)$ determined. The above described technique for deriving the joint distribution of $Y_1, \ldots, Y_k$ will be called the *cumulative-distribution-function technique*.

EXAMPLE Let there be only one given random variable, say X, which has a standard normal distribution. Suppose the distribution of $Y = g(X) = X^2$ is desired.

$$F_Y(y)$$

$$= P[Y \leq y] = P[X^2 \leq y] = P[-\sqrt{y} \leq X \leq \sqrt{y}] = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

$$= 2\int_0^{\sqrt{y}} \phi(u)\,du = 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,du$$

$$= \frac{2}{\sqrt{2\pi}} \int_0^y \frac{1}{2\sqrt{z}} e^{-\frac{1}{2}z}\,dz = \int_0^y \frac{1}{\Gamma(\frac{1}{2})} \frac{1}{\sqrt{2z}} e^{-\frac{1}{2}z}\,dz, \quad \text{for } y > 0,$$

which can be recognized as the cumulative distribution function of a gamma distribution with parameters $r = \frac{1}{2}$ and $\lambda = \frac{1}{2}$. ////

3.2 Distribution of minimum and maximum

Let $X_1, \ldots, X_n$ be n given random variables. Define $Y_1 = \min[X_1, \ldots, X_n]$ and $Y_n = \max[X_1, \ldots, X_n]$. To be certain to understand the meaning of $Y_n = \max[X_1, \ldots, X_n]$, recall that each X_i is a function with domain Ω, the sample space of a random experiment. For each $\omega \in \Omega$, $X_i(\omega)$ is some real number. Now Y_n is to be a random variable; that is, for each ω, $Y_n(\omega)$ is to be some real number. As defined, $Y_n(\omega) = \max[X_1(\omega), \ldots, X_n(\omega)]$; that is, for a given ω, $Y_n(\omega)$ is the largest of the real numbers $X_1(\omega), \ldots, X_n(\omega)$.

The distributions of Y_1 and Y_n are desired.

Theorem If $X_1, \ldots, X_n$ are independent random variables and $Y_n = \max[X_1, \ldots, X_n]$, then

$$F_{Y_n}(y) = \prod_{i=1}^{n} F_{X_i}(y). \qquad (18)$$

If $X_1, \ldots, X_n$ are independent and identically distributed with common cumulative distribution function $F_X(\cdot)$, then

$$F_{Y_n}(y) = [F_X(y)]^n. \qquad (19)$$

////

The distributions of Y_1 and Y_n are desired. $F_{Y_n}(y) = P[Y_n \leq y] = P[X_1 \leq y; \ldots; X_n \leq y]$ since the largest of the X_i's is less than or equal to y if and only if all the X_i's are less than or equal to y. Now, if the X_i's are assumed independent, then

$$P[X_1 \leq y; \ldots; X_n \leq y] = \prod_{i=1}^{n} P[X_i \leq y] = \prod_{i=1}^{n} F_{X_i}(y);$$

so the distribution of $Y_n = \max[X_1, \ldots, X_n]$ can be expressed in terms of the marginal distributions of $X_1, \ldots, X_n$. If in addition it is assumed that all the $X_1, \ldots, X_n$ have the same cumulative distribution, say $F_X(\cdot)$, then

$$\prod_{i=1}^{n} F_{X_i}(y) = [F_X(y)]^n.$$

Corollary If $X_1, \ldots, X_n$ are independent identically distributed continuous random variables with common probability density function $f_X(\cdot)$ and cumulative distribution function $F_X(\cdot)$, then

$$f_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y). \tag{20}$$

PROOF

$$f_{Y_n}(y) = \frac{d}{dy} F_{Y_n}(y) = n[F_X(y)]^{n-1} f_X(y). \qquad ////$$

Theorem If $X_1, \ldots, X_n$ are independent random variables and $Y_1 = \min[X_1, \ldots, X_n]$, then

$$F_{Y_1}(y) = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(y)]. \tag{21}$$

And if $X_1, \ldots, X_n$ are independent and identically distributed with common cumulative distribution function $F_X(\cdot)$, then

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n. \tag{22}$$

////

Corollary If $X_1, \ldots, X_n$ are independent identically distributed continuous random variables with common probability density $f_X(\cdot)$ and cumulative distribution $F_X(\cdot)$, then

$$f_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y). \qquad (23)$$

PROOF

$$f_{Y_1}(y) = \frac{d}{dy} F_{Y_1}(y) = n[1 - F_X(y)]^{n-1} f_X(y). \qquad ////$$

3.3 Distribution of sum and difference of two random variables

Theorem Let X and Y be jointly distributed continuous random variables with density $f_{X,Y}(x, y)$, and let $Z = X + Y$ and $V = X - Y$. Then,

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx = \int_{-\infty}^{\infty} f_{X,Y}(z - y, y)\, dy, \qquad (24)$$

and

$$f_V(v) = \int_{-\infty}^{\infty} f_{X,Y}(x, x - v)\, dx = \int_{-\infty}^{\infty} f_{X,Y}(v + y, y)\, dy. \qquad (25)$$

PROOF We will prove only the first part of Eq. (24); the others are proved in an analogous manner.

$$F_Z(z) = P[Z \le z] = P[X + Y \le z] = \iint\limits_{x+y \le z} f_{X,Y}(x, y)\, dx\, dy$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_{X,Y}(x, y)\, dy \right] dx$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z} f_{X,Y}(x, u - x)\, du \right] dx$$

by making the substitution $y = u - x$.

Now

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \frac{d}{dz} \left\{ \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} f_{X,Y}(x, u - x)\, dx \right] du \right\}$$

$$= \int_{-\infty}^{\infty} f_{X,Y}(x, z - x)\, dx. \qquad ////$$

Corollary If X and Y are independent continuous random variables and $Z = X + Y$, then

$$f_Z(z) = f_{X+Y}(z) = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x)\, dx = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy. \qquad (26)$$

Remark The formula given in Eq. (26) is often called the *convolution* formula. In mathematical analysis, the function $f_Z(\cdot)$ is called the *convolution* of the functions $f_Y(\cdot)$ and $f_X(\cdot)$. ////

3.4 Distribution of product and quotient

Theorem Let X and Y be jointly distributed continuous random variables with density $f_{X,Y}(x, y)$, and let $Z = XY$ and $U = X/Y$; then

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{z}{y}, y\right) dy, \tag{27}$$

and

$$f_U(u) = \int_{-\infty}^{\infty} |y| f_{X,Y}(uy, y)\, dy. \tag{28}$$

EXAMPLE Suppose X and Y are independent random variables, each uniformly distributed over the interval (0, 1). Let $Z = XY$ and $U = X/Y$.

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{|x|} I_{(0,1)}(x) I_{(0,1)}\left(\frac{z}{x}\right) dx$$

$$= I_{(0,1)}(z) \int_0^1 \frac{1}{x} I_{(z,1)}(x)\, dx$$

$$= I_{(0,1)}(z) \int_z^1 \frac{1}{x}\, dx = -\log z\, I_{(0,1)}(z).$$

$$
\begin{aligned}
f_U(u) &= \int_{-\infty}^{\infty} |y| f_{X,Y}(uy, y)\, dy \\
&= \int_{-\infty}^{\infty} |y| I_{(0,1)}(uy) I_{(0,1)}(y)\, dy \\
&= \int_{-\infty}^{\infty} |y| \{I_{(0,1)}(u) I_{(0,1)}(y) + I_{[1,\infty)}(u) I_{(0,1/u)}(y)\}\, dy \\
&= I_{(0,1)}(u) \int_0^1 y\, dy + I_{[1,\infty)}(u) \int_0^{1/u} y\, dy \\
&= \frac{1}{2} I_{(0,1)}(u) + \frac{1}{2}\left(\frac{1}{u}\right)^2 I_{[1,\infty)}(u).
\end{aligned}
$$

Note that $\mathscr{E}[X/Y] = \mathscr{E}[U] = \frac{1}{2}\int_0^1 u\, du + \frac{1}{2}\int_1^{\infty}(1/u)\, du = \infty$, quite different from $\mathscr{E}[X]/\mathscr{E}[Y] = 1$. ////

4 Moment-generating-function technique

4.1 Description of the technique

There is another method of determining the distribution of functions of random variables which we shall find to be particularly useful in certain instances. This method is built around the concept of the moment generating function and will be called the *moment-generating-function technique*.

This method is quite powerful in connection with certain techniques of advanced mathematics (the theory of transforms) which, in many instances, enable one to determine the distribution associated with the derived moment generating function.

EXAMPLE Suppose X has a normal distribution with mean 0 and variance 1. Let $Y = X^2$, and find the distribution of Y.

$$m_Y(t) = \mathscr{E}[e^{tY}] = \int_{-\infty}^{\infty} e^{tx^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2(1-2t)}\,dx$$

$$= \frac{1}{\sqrt{2\pi}} \cdot \frac{(1-2t)^{-\frac{1}{2}}}{(1-2t)^{-\frac{1}{2}}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2(1-2t)}\,dx$$

$$= (1-2t)^{-\frac{1}{2}} = \left(\frac{\frac{1}{2}}{\frac{1}{2}-t}\right)^{\frac{1}{2}} \qquad \text{for} \quad t < \frac{1}{2},$$

which we recognize as the moment generating function of a gamma with parameters $r = \frac{1}{2}$ and $\lambda = \frac{1}{2}$. It is also called a chi-square distribution with one degree of freedom. ////

4.2 Distribution of sums of independent random variables

Theorem If $X_1, \ldots, X_n$ are independent random variables and the moment generating function of each exists for all $-h < t < h$ for some $h > 0$, let $Y = \sum_1^n X_i$; then

$$m_Y(t) = \mathscr{E}[\exp \sum X_i t] = \prod_{i=1}^{n} m_{X_i}(t) \qquad \text{for} \quad -h < t < h.$$

PROOF

$$m_Y(t) = \mathscr{E}[\exp \sum X_i t] = \mathscr{E}\left[\prod_{i=1}^{n} e^{X_i t}\right]$$

$$= \prod_{i=1}^{n} \mathscr{E}[e^{X_i t}] = \prod_{i=1}^{n} m_{X_i}(t)$$

////

EXAMPLE Suppose that $X_1, \ldots, X_n$ are independent Bernoulli random variables; that is, $P[X_i = 1] = p$, and $P[X_i = 0] = 1 - p$. Now

$$m_{X_i}(t) = pe^t + q.$$

So

$$m_{\sum X_i}(t) = \prod_{i=1}^{n} m_{X_i}(t) = (pe^t + q)^n,$$

the moment generating function of a binomial random variable; hence $\sum_1^n X_i$ has a binomial distribution with parameters n and p. ////

The central limit theorem

- One of the most important theorems of probability theory is the central limit theorem. It gives an approximate distribution of an average.

Theorem **Central-limit theorem** If for each positive integer n, $X_1, \ldots, X_n$ are independent and identically distributed random variables with mean μ_X and variance σ_X^2, then for each z

$$F_{Z_n}(z) \text{ converges to } \Phi(z) \text{ as } n \text{ approaches } \infty, \tag{31}$$

where

$$Z_n = \frac{(\overline{X}_n - \mathscr{E}[\overline{X}_n])}{\sqrt{\text{var}\,[\overline{X}_n]}} = \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}}. \qquad ////$$

Corollary If $X_1, \ldots, X_n$ are independent and identically distributed random variables with common mean μ_X and variance σ_X^2, then

$$P\left[a < \frac{\overline{X}_n - \mu_X}{\sigma_X/\sqrt{n}} < b\right] \approx \Phi(b) - \Phi(a), \tag{32}$$

$$P[c < \overline{X}_n < d] \approx \Phi\left(\frac{d - \mu_X}{\sigma_X/\sqrt{n}}\right) - \Phi\left(\frac{c - \mu_X}{\sigma_X/\sqrt{n}}\right), \tag{33}$$

or

$$P\left[r < \sum_1^n X_i < s\right] \approx \Phi\left(\frac{s - n\mu_X}{\sqrt{n}\sigma_X}\right) - \Phi\left(\frac{r - n\mu_X}{\sqrt{n}\sigma_X}\right). \tag{34}$$

////

5 The transformation *Y=g(X)*

The last of our three techniques for finding the distribution of functions of given random variables is the *transformation technique*. It is discussed in this section for the special case of finding the distribution of a function of a uni-dimensional random variable. That is, for a given random variable X we seek the distribution of $Y = g(X)$ for some function $g(\cdot)$. Discussion of the general

5.1 Distribution of *Y=g(X)*

A random variable X may be transformed by some function $g(\cdot)$ to define a new random variable Y. The density of Y, $f_Y(y)$, will be determined by the transformation $g(\cdot)$ together with the density $f_X(x)$ of X.

Theorem Suppose X is a continuous random variable with probability density function $f_X(\cdot)$. Set $\mathfrak{X} = \{x: f_X(x) > 0\}$. Assume that:

(i) $y = g(x)$ defines a one-to-one transformation of $\mathfrak{X}$ onto $\mathfrak{Y}$.

(ii) The derivative of $x = g^{-1}(y)$ with respect to y is continuous and nonzero for $y \in \mathfrak{Y}$, where $g^{-1}(y)$ is the inverse function of $g(x)$; that is, $g^{-1}(y)$ is that x for which $g(x) = y$.

Then $Y = g(X)$ is a continuous random variable with density

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) I_{\mathfrak{Y}}(y).$$

PROOF The above is a standard theorem from calculus on the change of variable in a definite integral; so we will only sketch the proof. Consider the case when $\mathfrak{X}$ is an interval. Let us suppose that $g(x)$ is a monotone increasing function over $\mathfrak{X}$; that is, $g'(x) > 0$, which is true if and only if $(d/dy)g^{-1}(y) > 0$ over $\mathfrak{Y}$. For $y \in \mathfrak{Y}$, $F_Y(y) = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y))$, and hence $f_Y(y) = (d/dy)F_Y(y) = [(d/dy)g^{-1}(y)] f_X(g^{-1}(y))$ by chain rule of differentiation. On the other hand, if $g(x)$ is a monotone decreasing function over $\mathfrak{X}$, so that $g'(x) < 0$ and $(d/dy)g^{-1}(y) < 0$, then $F_Y(y) = P[g(X) \leq y] = P[X \geq g^{-1}(y)] = 1 - F_X(g^{-1}(y))$, and therefore $f_Y(y) = -[(d/dy)g^{-1}(y)]f_X(g^{-1}(y)) = |(d/dy)g^{-1}(y)| f_X(g^{-1}(y))$ for $y \in \mathfrak{Y}$. ////

EXAMPLE Let X be a random variable with uniform distribution over the interval (0, 1) and let $Y = g(X) = X^2$. The density of Y is desired. Now

$$F_Y(y) = P[Y \leq y] = P[X^2 \leq y] = \int_{\{x:\, x^2 \leq y\}} f_X(x)\, dx = \int_0^{\sqrt{y}} dx = \sqrt{y}$$

for $0 < y < 1$; so

$$F_Y(y) = \sqrt{y} I_{(0,1)}(y) + I_{[1,\infty)}(y),$$

and therefore

$$f_Y(y) = \frac{1}{2} \frac{1}{\sqrt{y}} I_{(0,1)}(y). \qquad ////$$

Application of the cumulative-distribution-function technique to find the density of $Y = g(X)$, as in the above example, produces the transformation technique,

EXAMPLE Suppose X has a beta distribution. What is the distribution of $Y = -\log_e X$? $\mathfrak{X} = \{x: f_X(x) > 0\} = \{x: 0 < x < 1\}$. $y = g(x) = -\log_e x$ defines a one-to-one transformation of $\mathfrak{X}$ onto $\mathfrak{Y} = \{y: \ y > 0\}$. $x = g^{-1}(y) = e^{-y}$, so $(d/dy)g^{-1}(y) = -e^{-y}$, which is continuous and nonzero for $y \in \mathfrak{Y}$. By Theorem 11,

$$f_Y(y) = \left|\frac{d}{dy} g^{-1}(y)\right| f_X(g^{-1}(y)) I_{\mathfrak{Y}}(y)$$

$$= e^{-y} \frac{1}{\mathrm{B}(a, b)} (e^{-y})^{a-1} (1 - e^{-y})^{b-1} I_{(0, \infty)}(y)$$

$$= \frac{1}{\mathrm{B}(a, b)} e^{-ay} (1 - e^{-y})^{b-1} I_{(0, \infty)}(y).$$

In particular, if $b = 1$, then $\mathrm{B}(a, b) = 1/a$; so $f_Y(y) = ae^{-ay} I_{(0, \infty)}(y)$, an exponential distribution with parameter a. ////

5.2 Probability integral transform

The transformation $Y = F_X(X)$ is called the *probability integral transformation*. It plays an important role in the theory of distribution-free statistics and goodness-of-fit tests.

Theorem If X is a random variable with continuous cumulative distribution function $F_X(x)$, then $U = F_X(X)$ is uniformly distributed over the interval (0, 1). Conversely, if U is uniformly distributed over the interval (0, 1), then $X = F_X^{-1}(U)$ has cumulative distribution function $F_X(\cdot)$.

PROOF $P[U \leq u] = P[F_X(X) \leq u] = P[X \leq F_X^{-1}(u)] = F_X(F_X^{-1}(u)) = u$ for $0 < u < 1$. Conversely, $P[X \leq x] = P[F_X^{-1}(U) \leq x] = P[U \leq F_X(x)] = F_X(x)$. ////

Why is this an important result?

In various statistical applications, particularly in simulation studies, it is often desired to generate values of some random variable X. To generate a value of a random variable X having continuous cumulative distribution function $F_X(\cdot)$, it suffices to generate a value of a random variable U that is uniformly distributed over the interval (0, 1). This follows from Theorem since if U is a random variable with a uniform distribution over the interval (0, 1), then $X = F_X^{-1}(U)$ is a random variable having distribution $F_X(\cdot)$. So to get a value, say x, of a random variable X, obtain a value, say u, of a random variable U, compute $F_X^{-1}(u)$, and set it equal to x. A value u of a random variable U is called a *random number*. Many computer-oriented random-number generators are available.

5.3 The transformation $Y_1 = g_1(X_1, ..., X_n), ..., Y_k = g_k(X_1, ..., X_n)$

Suppose that the discrete density function $f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)$ of the n-dimensional discrete random variable $(X_1, \ldots, X_n)$ is given. Let $\mathfrak{X}$ denote the mass points of $(X_1, \ldots, X_n)$; that is,

$$\mathfrak{X} = \{(x_1, \ldots, x_n): f_{X_1, \ldots, X_n}(x_1, \ldots, x_n) > 0\}.$$

Suppose that the joint density of $Y_1 = g_1(X_1, \ldots, X_n), \ldots, Y_k = g_k(X_1, \ldots, X_n)$ is desired. It can be observed that $Y_1, \ldots, Y_k$ are jointly discrete and $P[Y_1 = y_1; \ldots; Y_k = y_k] = f_{Y_1, \ldots, Y_k}(y_1, \ldots, y_k) = \sum f_{X_1, \ldots, X_n}(x_1, \ldots, x_n)$, where the summation is over those $(x_1, \ldots, x_n)$ belonging to $\mathfrak{X}$ for which $(y_1, \ldots, y_k) = (g_1(x_1, \ldots, x_n), \ldots, g_k(x_1, \ldots, x_n))$.

Theorem Let X_1 and X_2 be jointly continuous random variables with density function $f_{X_1, X_2}(x_1, x_2)$. Set $\mathfrak{X} = \{(x_1, x_2): f_{X_1, X_2}(x_1, x_2) > 0\}$. Assume that:

(i) $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ defines a one-to-one transformation of $\mathfrak{X}$ onto $\mathfrak{Y}$.

(ii) The first partial derivatives of $x_1 = g_1^{-1}(y_1, y_2)$ and $x_2 = g_2^{-1}(y_1, y_2)$ are continuous over $\mathfrak{Y}$.

(iii) The Jacobian of the transformation is nonzero for $(y_1, y_2) \in \mathfrak{Y}$.

Then the joint density of $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ is given by

$$f_{Y_1, Y_2}(y_1, y_2) = |J| f_{X_1, X_2}(g_1^{-1}(y_1, y_2), g_2^{-1}(y_1, y_2)) I_{\mathfrak{Y}}(y_1, y_2). \tag{40}$$

EXAMPLE Let X_1 and X_2 be two independent standard normal random variables. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1/X_2$. Then

$$x_1 = g_1^{-1}(y_1, y_2) = \frac{y_1 y_2}{1 + y_2} \quad \text{and} \quad x_2 = g_2^{-1}(y_1, y_2) = \frac{y_1}{1 + y_2}.$$

$$J = \begin{vmatrix} \frac{y_2}{1 + y_2} & \frac{y_1}{(1 + y_2)^2} \\ \frac{1}{1 + y_2} & -\frac{y_1}{(1 + y_2)^2} \end{vmatrix} = -\frac{y_1(y_2 + 1)}{(1 + y_2)^3} = -\frac{y_1}{(1 + y_2)^2}.$$

$$f_{Y_1, Y_2}(y_1, y_2)$$

$$= \frac{|y_1|}{(1 + y_2)^2} \frac{1}{2\pi} \exp\left\{ -\frac{1}{2} \left[\frac{(y_1 y_2)^2}{(1 + y_2)^2} + \frac{y_1^2}{(1 + y_2)^2} \right] \right\}$$

$$= \frac{1}{2\pi} \frac{|y_1|}{(1 + y_2)^2} \exp\left[-\frac{1}{2} \frac{(1 + y_2^2) y_1^2}{(1 + y_2)^2} \right].$$

To find the marginal distribution of, say, Y_2, we must integrate out y_1; that is

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty} f_{Y_1, Y_2}(y_1, y_2)\, dy_1$$

$$= \frac{1}{2\pi} \frac{1}{(1 + y_2)^2} \int_{-\infty}^{\infty} |y_1| \exp\left[-\frac{1}{2} \frac{(1 + y_2^2) y_1^2}{(1 + y_2)^2} \right] dy_1.$$

Let

$$u = \frac{1}{2}\frac{(1 + y_2^2)}{(1 + y_2)^2} y_1^2;$$

then

$$du = \frac{(1 + y_2^2)}{(1 + y_2)^2} y_1 \, dy_1$$

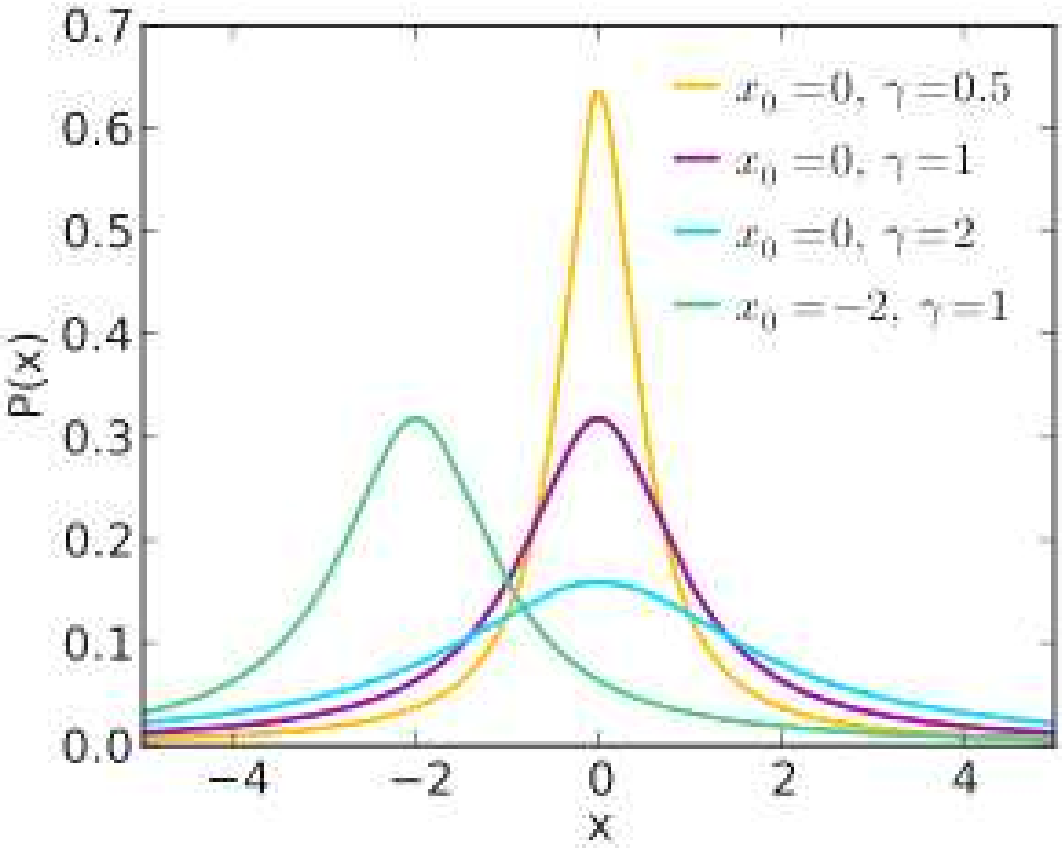
and so

$$f_{Y_2}(y_2) = \frac{1}{2\pi} \cdot \frac{1}{(1 + y_2)^2} \cdot \frac{(1 + y_2)^2}{1 + y_2^2} (2) \int_0^\infty e^{-u} \, du = \frac{1}{\pi} \cdot \frac{1}{1 + y_2^2},$$

a Cauchy density. That is, the *ratio of two independent standard normal random variables has a Cauchy distribution.* ////

A note on the Cauchy (-Lorentz) distribution

- The Cauchy distribution is important as an example of a pathological case. Cauchy distributions look similar to a normal distribution. However, they have much heavier tails.
- When studying hypothesis tests that assume normality, seeing how the tests perform on data from a Cauchy distribution is a good indicator of how sensitive the tests are to heavy-tail departures from normality.
- Likewise, it is a good check for robust techniques that are designed to work well under a wide variety of distributional assumptions.
- Finally, the mean and standard deviation of the Cauchy distribution are undefined. The practical meaning of this is that collecting 1,000 data points gives no more accurate an estimate of the mean and standard deviation than does a single point.



where x_0 is the location parameter, specifying the location of the peak of the Cauchy distribution, and *γ* is the scale parameter which specifies the half-width at half-maximum