Distribution of Functions of a Random Variable

We have considered the expectation of a function u(X) of random variables. If we are concerned only about the moment of u(X), we do not need its distribution function. However, estimators of unknown parameters, or the test statistics (such as *t-test, F-test,* etc.) in econometrics, are functions of other random variables, and it is critical to know their distribution functions. The knowledge of the distribution function of u(X) may be necessary in other applications also. For example, one may wish to find out the probability of a positive profit in the earlier example of a competitive firm. We will first consider the distribution of a function of a single random variable. Let Y=u(X), where random variable X has cdf F(x) and pdf f(x). We wish to find the cdf G(y) and the pdf g(y) of Y.

Discrete Random Variable

When X is a discrete random variable, it is relatively straightforward to find the distribution function of u(X).

Example. Let X be a Bernoulli random variable with pdf $f(x)=\theta^x(1-\theta)^{1-x},\ x=0,1.$ Let $Y=X^2$. Then, Y takes a value 0 or 1, and P(Y=0)=P(X=0)=1-θ and P(Y=1)=P(X=1)=θ. Therefore, $g(y)=\theta^y(1-\theta)^{1-y},\ y=0,1.$ It is easy to see that $\sqrt{X}$ also has the same distribution as X in this example.

Example. Consider a discrete random variable X that takes values {-1, 0, 1} with probabilities θ_1, θ_2 and θ_3, where $\theta_1+\theta_2+\theta_3=1$. The pdf of $Y=X^2$ is g(0)=θ_2, g(1)=f(-1)+f(1)=$\theta_1+\theta_3$ and g(y)=0 for all other values of y.

Theorem. Let X be a discrete random variable that takes values $x_1,\ldots,x_n$ with pdf $f(x_i)$. The pdf of Y=u(X) is given by

$$g(y) = \sum_{\{i:\, y=u(x_i)\}} f(x_i)$$

Continuous Random Variable

When X is a continuous random variable, the procedure is more complicated. We will consider two alternative methods: *Moment Generating Function Method* and *Cumulative Distribution Function Method*. Which method to use depends on the problem under consideration. It is easier to use the mgf method for some problems, while the cdf method is more straightforward for some other problems.

Method of Moment Generating Function

Since the moment generating function is unique if it exists, this method finds the mgf of Y=u(X), and compare it with the mgf of a known distribution. Thus, the mgf method is used to prove that u(X) has a given distribution function.

Example. Let X be uniformly distributed over the unit interval (0,1), and let $Y=-\lambda^{-1}\ln X$. Since f(x)=1 for x∈(0,1), the mgf of Y is

$$E(e^{tY}) = E\left(e^{t\ln X^{-1/\lambda}}\right) = E(X^{-t/\lambda}) = \int_0^1 x^{-t/\lambda}dx = \frac{\lambda}{\lambda-t}$$

which is equal to the mgf of an exponential random variable with parameter λ. Hence, Y has an exponential density. ⏮

Theorem. Let X be a uniform random variate U(0,1). Then, $-\frac{1}{\lambda}\ln X$ is distributed as an exponential random variate with parameter λ.

Example. Let $X \sim N(\mu_x, \sigma_x^2)$ and Y=aX+b. We have already shown without the knowledge of the distribution function of Y that it has mean $\mu_y = a\mu_x + b$ and variance $\sigma_y^2 = a^2\sigma_x^2$. Now we wish to find the distribution function of Y by finding its mgf:

$$E(e^{tY}) = E(e^{t(aX+b)}) = e^{tb}E(e^{taX}) = e^{tb}e^{ta\mu_x + (ta)^2\sigma_x^2/2} = e^{t\mu_y + t^2\sigma_y^2/2}$$

We have shown that this is the mgf of a normal random variable with mean μ_y and variance σ_y^2, and hence, Y is distributed as $N(a\mu + b, a^2\sigma^2)$. In particular, if a=1/σ_x and b=-μ_x/σ_x, then Y is distributed as a standard normal variate $N(0,1)$. ⏮

Example. Let X ~ N(μ,1) and Y=X^2. The mgf of Y is

$$E(e^{tY}) = E(e^{tX^2}) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx^2} e^{-(x-\mu)^2/2}\,dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-Q}\,dx$$

Let $s^2 = (1-2t)^{-1}$. The exponent term Q can be rearranged for t<1/2 as

$$Q = \frac{(x-\mu)^2}{2} - tx^2 = \frac{x^2 - 2\mu s^2 x + \mu^2 s^2}{2s^2} = \frac{\left(x-\mu s^2\right)^2}{2s^2} + \frac{\mu^2(1-s^2)}{2} = \frac{\left(x-\mu s^2\right)^2}{2s^2} - t\mu^2 s^2$$

Therefore, for t<1/2,

$$E(e^{tY}) = s e^{ts^2\mu^2}\frac{1}{\sqrt{2\pi s^2}}\int_{-\infty}^{\infty} e^{-\left(x-\mu s^2\right)^2/2s^2}\,dx = s e^{ts^2\mu^2} = \frac{e^{t\mu^2/(1-2t)}}{\sqrt{1-2t}}$$

which is the mgf of a noncentral chi-square distribution $\chi^2(1,\delta)$ with one degree of freedom and noncentrality parameter $\delta = \sqrt{\mu^2}$. ⏮

Theorem. Let X be distributed as $N(\mu, \sigma^2)$. Then,

(a) aX+b ~ $N(a\mu + b,\ a^2\sigma^2)$, and $(X-\mu)/\sigma$ ~ N(0,1),

(b) $X^2/\sigma^2 \sim \chi^2(1, \delta)$, and $(X-\mu)^2/\sigma^2 \sim \chi^2(1, 0)$, where $\delta = \sqrt{\mu^2/\sigma^2}$.

Method of Cumulative Distribution Function.

This method finds the cdf of Y=u(X) by applying the definition of the cdf, G(y)=P(Y$\le$y)=P(u(X)$\le$y), and then derive the pdf by differentiating the cdf. This method is intuitively clear and leads to a simple formula for the pdf when the function u(X) is differentiable and strictly monotonic.

Monotonic Transformation

Example. Consider a linear function Y=aX+b, $a \neq 0$, and let F(x) and f(x) be the cdf and pdf of X. From the definition of the cdf, we can write the cdf of Y as

$$G(y) = P(Y \le y) = P(aX+b \le y) = \begin{cases} P(X \le (y-b)/a) = F\left(\frac{y-b}{a}\right) & \text{if } a > 0 \\ P(X \ge (y-b)/a) = 1 - P(X \le (y-b)/a) = 1 - F\left(\frac{y-b}{a}\right) & \text{if } a < 0 \end{cases}$$

Differentiating G(y) with respect to y we find its pdf

$$g(y) = \frac{\partial G(y)}{\partial y} = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

The pdf of Y is obtained by replacing x in f(x) with (y-b)/a, and then multiplying the *absolute value* of 1/a. Note that x=(y-b)/a is the unique *inverse function* of y=ax+b. The term 1/a is called the *Jacobian* of the transformation Y=aX+b.

To understand the role of the Jacobian, consider an interval $(x, x+\triangle_x]$ for X for a small positive number $\triangle_x$, and the corresponding interval $(y, y+\triangle_y]$ for Y, where y=ax+b and $\triangle_y$=a $\triangle_x$. The probability $P(x<X\leq x+\triangle_x)$ of X is transferred and spread over the interval $(y, y+\triangle_y]$ for Y, such that $P(y<Y\leq y+\triangle_y) = P(x<X\leq x+\triangle_x)$. When $\triangle_x$ is small, these probabilities can be approximated by $g(y)\triangle_y$ and $f(x)\triangle_x$, respectively. This gives the relationship between g(y) and f(x): $g(y)\triangle_y=f(x)\triangle_x$, or $g(y)=f(x)(\triangle_x/\triangle_y)$. When $\triangle_x$ converges to zero, we can write g(y)=f(x)|dx/dy|, where we take the absolute value of the derivative dx/dy because it represents the relative length of intervals. That is, the absolute value of the Jacobian represents the relative length of intervals for X and Y, over which the same amount of probability mass is spread. This procedure applies to any *monotonic differentiable* transformation, and presented below as a theorem, follwed by a few examples of differentiable monotonic transformation[1].

Theorem. Let f(x) be the pdf of a continuous random variable X, and let Y=u(X). If u(x) is *differentiable* and *strictly monotonic* over the interval of positive f(x), then the inverse transformation $u^{-1}(y)$ exists and

$$g(y) = f\left(u^{-1}(y)\right)|J|, \qquad J = \frac{\partial u^{-1}(y)}{\partial y}$$

where J is the *Jacobian* of the transformation Y=u(X). ◘

Examples of Differentiable Monotonic Transformation

(1) ***Linear Function***: Y=u(X)=aX+b, a≠0.

Since the inverse function is $u^{-1}(y)$=(y-b)/a and the Jacobian is J=1/a, the pdf of Y is given by

$$g(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right)$$

Example. X ~ U(0,1).

$$f(x) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \qquad g(y) = \begin{cases} \dfrac{1}{|a|} & \text{if } 0 < \dfrac{y-b}{a} < 1 \\ 0 & \text{otherwise} \end{cases}$$

The range of positive density for Y is the transformation of the range of positive density for X:

$$0 < \frac{y-b}{a} < 1 \quad \Leftrightarrow \quad \begin{cases} b < y < a+b & \text{if } a > 0 \\ a+b < y < b & \text{if } a < 0 \end{cases}$$

Note that the length of these intervals for Y is |a|, while that for X is 1. Therefore, the height of the density

[1] The procedure stated in the theorem below is often called a *Transformation (Change of Variable) Technique*. We will treat it as a result of the cumulative distribution function technique.

line is lowered if $|a|>1$, and raised up if $|a|<1$, by a factor of $1/|a|$. ⏮

Example. X~ N(μ, σ^2). We have shown by using the mgf technique that Y is distributed as a normal with mean $\mu_y = a\mu + b$ and variance $\sigma_y^2 = a^2\sigma^2$. We derive this result by using the theorem above. Since the normal pdf is

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}$$

we can write

$$g(y) = \frac{1}{|a|} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-[(y-b)/a - \mu_x]^2/2\sigma_x^2} = \frac{1}{\sqrt{2\pi a^2\sigma^2}} \exp^{-[y-(a\mu_x+b)]^2/2a^2\sigma_x^2} = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{-(y-\mu_y)^2/2\sigma_y^2}$$

which is a normal density with mean μ_y and variance σ_y^2. Recall that the linear transformation stretches (or shrinks) the finite support of the uniform distribution and the density curve is proportionally adjusted by lowering (or raising) it. Since the range of a normal r.v. is (-∞, ∞), the range remains unchanged even after we stretch it. However, if we consider any finite interval [α, β] of length (β-α), it gets stretched to [aα, aβ] of length a(β-α). Hence, we need to adjust the height of the pdf curve by the factor $1/|a|$. This is how we get the pdf g(y).

As a special case, consider a standard normal X~N(0,1) with pdf ϕ(x) and cdf Φ(x). Let Y=μ+σX. Then, the pdf and the cdf of Y are

$$g(y) = \frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma}\phi(x), \qquad G(y) = \Phi\left(\frac{y-\mu}{\sigma}\right) = \Phi(x)$$

These relationships between the standard normal and other normal random variables are useful in the econometric analysis of qualitative response models. ⏮

(2) ***Square Function***. Y=u(X)=X^2 and f(x)=0 for $x \le 0$.

Since the inverse function is $u^{-1}(y) = \sqrt{y}$ for y>0, and $J = 1/2\sqrt{y}$, the pdf of Y is

$$g(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y}) \qquad for\ y > 0$$

Example. X ~ U(0,1). Since f(x)=1 for x∈(0,1) and f(x)=0 otherwise,

$$g(y) = \frac{1}{2\sqrt{y}} f(\sqrt{y}) = \frac{1}{2\sqrt{y}} \qquad for\ 0 < \sqrt{y} < 1 \ \Leftrightarrow\ 0 < y < 1$$

The transformation Y=X^2 in this example does not change the entire length of the support (0,1). However, subintervals within the unit intervals are transformed differently. For example, an interval (0.1, 0.2) becomes (0.01, 0.04) of length 0.03, while (0.5, 0.6) becomes (0.25, 0.36) of length 0.11. X being a uniform distribution, the probability mass in the interval (0.1, 0.2) is the same as the probability mass in the interval (0.5, 0.6). This same probability mass will be spread over the intervals of different lengths for the random variable Y. Therefore, the density of Y in the interval (0.01, 0.04) will be higher than the density in the interval (0.25, 0.36). ⏮

(3) ***Exponential Function***: Y = u(X) = e^X. Since $u^{-1}(y) = lny$ for y>0, and J=1/y,

$$g(y) = \frac{1}{y}f(lny), \quad for\ y>0$$

Example. X ~ U(0,1).

$$g(y) = \frac{1}{|y|}f(\ln y) = \begin{cases} \frac{1}{y} & if\ \ 0<\ln y<1 \ \Leftrightarrow\ 1<y<e \\ 0 & otherwise \end{cases}$$

Example. X ~ N(μ, σ^2).

$$g(y) = \frac{1}{y\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(lny-\mu)^2}{2\sigma^2}\right), \quad for\ \ -\infty<\ln y<\infty \ \Leftrightarrow\ 0<y<\infty$$

which is the pdf of the *log-normal* random variable with parameters μ and σ^2. This relationship suggests an easy way to find the mean of a log-normal random variable from the mgf of a normal random variable. Since the mgf of a normal is $e^{t\mu+t^2\sigma^2/2}$ and E(Y)=E(e^X) is the mgf of a normal for t=1, $E(Y) = e^{\mu+\sigma^2/2}$.

(4) ***Logarithmic Function***: Y = u(X) = β *ln* (aX+b), aX+b > 0 , a≠0, β≠0.

Since the inverse function is $u^{-1}(y) = (e^{y/\beta}-b)/a$ and the Jacobian is $J = e^{y/\beta}/(a\beta)$, we have

$$g(y) = \frac{1}{|a||\beta|}e^{y/\beta}f\left(\frac{e^{y/\beta}-b}{a}\right)$$

Example: X ~ U(0,1).

(a) Y = *ln*X (i.e., β=1, a=1 and b=0). $\Rightarrow$ $g(y) = e^y f(e^y) = e^y$.
(b) Y = - (1/λ)*ln*X (i.e., β=-1/λ, a=1 and b=0). $\Rightarrow$ $g(y) = \lambda e^{-\lambda y}$
(c) Y = - (1/λ)*ln*(1-X) (i.e., β=-1/λ, a=-1 and b=1). $\Rightarrow$ $g(y) = \lambda e^{-\lambda y}$

The density $g(y) = \lambda e^{-\lambda y}$ is the pdf of an exponential random variable (also gamma random variable with parameters λ and k=1).

Example: Let X be a log-normal random variable with parameters μ and σ^2, and let Y=β*ln*(aX) for a>0. Since the log-normal density function is

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(lnx-\mu)^2}{2\sigma^2}\right) \quad for\ 0<x<\infty$$

we have

$$g(y) = \frac{1}{|a||\beta|}e^{y/\beta}f\left(\frac{e^{y/\beta}}{a}\right) = \frac{1}{|a||\beta|}e^{y/\beta}\,\frac{1}{\frac{e^{y/\beta}}{a}\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(ln\frac{e^{y/\beta}}{a}-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\beta^2\sigma^2}}\exp\left(-\frac{(y-\beta(\mu+\ln a))^2}{2\beta^2\sigma^2}\right)$$

which is the density function of a normal with mean $\beta(\mu+ln\,a)$ and variance $\beta^2\sigma^2$.

Example: *Exponential Gamma Distribution*. Let X be a Gamma random variable with parameters k and λ, and let Y=β*ln*(aX) for a>0. Since the Gamma density function is

$$f(x;k,\lambda) = \frac{\lambda^k}{\Gamma(k)} x^{k-1} e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0, \quad k > 0, \qquad \Gamma(k) = \int_0^\infty z^{k-1} e^{-z} dz$$

we have

$$g(y) = \frac{1}{|a||\beta|} e^{y/\beta} f\left(\frac{e^{y/\beta}}{a}\right) = \frac{1}{|a||\beta|} e^{y/\beta} \frac{\lambda^k}{\Gamma(k)} \left(\frac{e^{y/\beta}}{a}\right)^{k-1} e^{-(\lambda/a) e^{y/\beta}}, \qquad y \in \mathbb{R}, \quad \lambda > 0, \quad k > 0$$

This density function is called an *exponential Gamma* density, because e^Y has a Gamma density. If a=β=1, i.e., Y=*ln*(X), then the pdf of Y becomes

$$g(y) = \frac{\lambda^k}{\Gamma(k)} e^{ky} e^{-\lambda e^y}, \qquad y \in \mathbb{R}, \quad \lambda > 0, \quad k > 0$$

This distribution is used in conjunction with a Poisson distribution in the *Count Models*. We will discuss this later when the conditional and marginal distributions are introduced. ⏮

The results derived in these examples provide a convenient way to generate *random numbers* of a certain distribution function. Most computer softwares have a routine that generates standard uniform U(0, 1) (pseudo) random numbers that are in the interval (0,1), and standard normal (pseudo) random numbers. To convert the U(0,1) random numbers x to Y~U(a,b), simply compute y=a+(b-a)x. Random numbers of exponential distribution can be generated from the standard U(0,1) random numbers by computing y = -(1/λ) *ln*(1-x) or y = -(1/λ) *ln*x for any λ. To convert standard normal N(0,1) random numbers x to Y ~ N(μ, σ^2), compute y=μ+σx. Random numbers of log-normal distribution can be generated by transforming the random numbers of normal distribution by computing y=e^x. Some of the important results in examples above are summarized as a theorem.

Theorem.

(a) If X ~ N(μ, σ^2), then e^X is distributed as a log-normal with parameters μ and σ^2.
(b) If X is distributed as a log-normal with parameters μ and σ^2, then *ln*X is distributed as N(μ, σ^2).
(c) If X ~ U(0,1), then -(1/λ)*ln*X and -(1/λ)*ln*(1-X) are distributed as exponential with parameter λ.

(5) ***Box-Cox Transformation***.

$$Y = \begin{cases} \frac{X^\lambda - 1}{\lambda} & if\ \lambda \neq 0 \\ lnX & if\ \lambda = 0 \end{cases}$$

This transformation is used widely in econometrics when the functional form of a regression relationship is unknown. Consider the case of λ>0. Since the inverse function is $x=(\lambda y+1)^{1/\lambda}$,

$$g(y) = \begin{cases} f\left((\lambda y + 1)^{1/\lambda}\right)(\lambda y + 1)^{(1-\lambda)/\lambda} & in\ terms\ of\ variable\ y \\ f(x) x^{1-\lambda} & in\ terms\ of\ variable\ x \end{cases}$$

This can also be derived from the cdf of Y

$$G(y) = P\left(\frac{X^{\lambda}-1}{\lambda} \le y\right) = P\left(X \le (\lambda y+1)^{1/\lambda}\right) = F\left((\lambda y+1)^{1/\lambda}\right)$$

The expression of g(y) in terms of the original variable X is used in econometrics to specify the *log-likelihood* function. Note that the Box-Cox transformation applies for an arbitrary λ only if X is a positive random variable. In some applications in which X can be negative, the Box-Cox transformation is taken on a new random variable Z=X+c, where c is a positive constant large enough to ensure the positivity of Z.

(6) ***Probability Integral Transformation***. Let F(x) be the cdf of X, and let Y= F(X). (Note the change in notation from x to X.) Since Y is a function of X, it is a random variable and we may wish to find its cdf. Let the inverse function of F be $X=F^{-1}(Y)$. The Jacobian is

$$J = dx/dy = \frac{1}{dy/dx} = \frac{1}{f(x)} = \frac{1}{f(F^{-1}(y))}$$

Hence,

$$g(y) = f(F^{-1}(y))|J| = 1 \quad for \;\; 0 \le y \le 1$$

That is, Y has a uniform distribution over the unit interval. This can also be derived directly by using the cumulative distribution function technique:

$$G(y) = P(Y \le y) = P\left(F(X) \le y\right) = P\left(X \le F^{-1}(y)\right) = F\left(F^{-1}(y)\right) = y, \quad for \;\; 0 \le y \le 1$$

We showed earlier that $P(Y \le y)=y$ is the cdf of a uniform r.v. with the range [0,1]. Thus, Y is distributed as U[0,1]. This type of transformation is called *probability integral transformation.* This relationship also appears in some applications of IO and Macroeconomics.

Theorem. Let F(x) be a cdf of X and let Y=F(X). Then, Y ~ U(0, 1).
Theorem. Let X ~ U(0, 1). If G(y) is a continuous function that satisfies the conditions for a cdf as described before, then G is the cdf of $Y=G^{-1}(X)$.
Proof. Applying the cumulative distribution function technique, we have

$$P(Y \le y) = P\left(G^{-1}(X) \le y\right) = P\left(X \le G(y)\right) = G(y), \quad \because \; X \sim U(0,1)$$

Example. Let F(x) be a cdf of X and let Φ be the cdf of a standard normal random variable. Then, $Y = \Phi^{-1}\left(F(X)\right)$ is a standard normal random variable.

Non-Monotonic Transformation

When the transformation Y=u(X) is not strictly monotonic over the entire range of X, the inverse function is not unique. Consider a square function $Y=X^2$. If X can take both positive and negative values, the inverse function is $x=\sqrt{y}$ or $x=-\sqrt{y}$. The theorem above needs to be modified to take this non-uniqueness into account. We first consider the derivation of the cdf of Y:

$$G(y) = P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}) \quad for \;\; all \;\; y \ge 0$$

which gives the pdf

$$g(y) = \frac{\partial G(y)}{\partial y} = \frac{1}{2\sqrt{y}}\left[f(\sqrt{y}) + f(-\sqrt{y})\right] \qquad y > 0$$

If f(x) is symmetric around zero (i.e., f(x)=f(-x)), then this pdf is reduced to

$$g(y) = \frac{f(\sqrt{y})}{\sqrt{y}} \quad for\ all\ y > 0, \quad if\ f(x) = f(-x)$$

The pdf g(y) is a sum of two components. The first part comes from the positive inverse function $x=\sqrt{y}$ which has the Jacobian $1/(2\sqrt{y})$, and the second part from the negative inverse function $x=-\sqrt{y}$ which has the Jacobian $-1/(2\sqrt{y})$. That is, g(y) is derived by splitting the range of X into subintervals (-∞, 0] and (0, ∞), such that u(X) is strictly monotonic in each subinterval. Applying the earlier theorem to each subinterval, we can write

$$for\ x \in (-\infty, 0], \qquad u^{-1}(y) = -\sqrt{y} \quad and \quad J = -\frac{1}{2\sqrt{y}} \quad \Rightarrow \quad g(y) = f(-\sqrt{y})\frac{1}{2\sqrt{y}}$$
$$for\ x \in (0, \infty), \qquad u^{-1}(y) = \sqrt{y} \quad and \quad J = \frac{1}{2\sqrt{y}} \quad \Rightarrow \quad g(y) = f(\sqrt{y})\frac{1}{2\sqrt{y}}$$

and sum of the two parts gives the pdf g(y). These results are summarized in the following theorem.

Theorem. Let f(x) be the pdf of a continuous random variable X, and let Y=u(X). Partition the range of X into m subintervals such that u(x) is *differentiable* and *strictly monotonic* in each subintervals. Let $u_i^{-1}(y)$ denote the inverse transformation in subinterval i. Then,

$$g(y) = \sum_{i=1}^{m} f\left(u_i^{-1}(y)\right)|J_i|$$

where J_i is the *Jacobian* of the transformation Y=u(X) in subinterval i. ▣

Examples of Differentiable Non-Monotonic Transformation

(1) ***Square Function***. Y=u(X)=X^2.

$$\left.\begin{array}{lll} x \in (-\infty, 0]: & u^{-1}(y) = -\sqrt{y}, & J = -\dfrac{1}{2\sqrt{y}} \\ x \in (0, \infty): & u^{-1}(y) = \sqrt{y}, & J = \dfrac{1}{2\sqrt{y}} \end{array}\right\} \Rightarrow \quad g(y) = \frac{1}{2\sqrt{y}}\left[f(\sqrt{y}) + f(-\sqrt{y})\right] \quad y > 0$$

Example. X ~ U(-1, 1). Since f(x)=1/2 for x∈(-1, 1) and f(x)=0 otherwise, and f(x) is symmetric around zero,

$$g(y) = \frac{f(\sqrt{y})}{2\sqrt{y}} = \frac{1}{2\sqrt{y}} \qquad for\ 0 < y < 1$$

Example. X~N(0, 1). We have shown earlier by using the mgf technique that $X^2 \sim \chi^2(1,0)$. To apply the cdf technique, recall that the pdf $\phi(x)$ of X is symmetric around zero, and hence,

$$g(y) = \frac{\phi(\sqrt{y})}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}}y^{-1/2}e^{-y/2}, \qquad y > 0$$

Since $\Gamma(1/2)=\sqrt{\pi}$, this is the ***central chi-square*** density with 1 degree of freedom, $\chi^2(1,0)$.

(2) ***Absolute Value Function*** Y=|X|.

$$g(y) = \begin{cases} f(y) + f(-y) & if\ y \ge 0 \\ 0 & if\ y < 0 \end{cases}$$

Example. X ~ U(-1, 1) and Y=|X|. Since f(x)=1/2 for x∈(-1, 1), g(y)=f(y)+f(-y)=2f(y)=1 if y∈(0,1). Density g(y) is obtained by folding the probability mass of negative x over the probability mass of positive x.

Example. X ~ N(0, 1) and Y=|X|.

$$g(y) = \phi(y) + \phi(-y) = 2\phi(y) \qquad for \quad y \geq 0$$

Example. Let X~N(μ, σ^2) and Y= |X-μ|. Since X-μ is distributed as N(0, σ^2), which is symmetric around zero,

$$g(y) = 2f(y) = \frac{2}{\sqrt{2\pi\sigma^2}} e^{-y^2/2\sigma^2} \qquad for \ y \geq 0$$

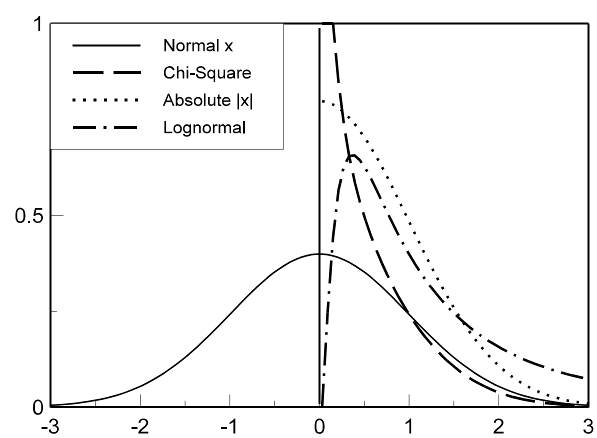
It is useful to consider the cdf of Y:

$$\begin{aligned} G(y) &= P(Y \leq y) = P(|X-\mu| \leq y) = P(-y \leq X-\mu \leq y) = P\left(-\frac{y}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{y}{\sigma}\right) \\ &= \Phi(y/\sigma) - \Phi(-y/\sigma) = \Phi(y/\sigma) - [1 - \Phi(y/\sigma)] = 2\Phi(y/\sigma) - 1 \end{aligned}$$

where we used the fact Φ(-c)=1-Φ(c) for any constant c. Applying this, we can find the probability

$$P(|X-\mu|<\sigma) = P(Y<\sigma) = G(\sigma) = 2\Phi(\sigma/\sigma) - 1 = 2\Phi(1) - 1 = 2*0.8413 - 1 = 0.6826$$

The figure below shows the pdf's of the functions of normal random variable.



Assignments #3.
(i) X ~ U(0,1). Find the pdf of $Y = \sqrt{X}$.
(ii) X is exponential with parameter λ. Find the pdf of Y=*ln* X.
(iii) X ~ U(0,1). Find the pdf of Y=-2*lnX.* Compare the result with the Gamma density.
(iv) X ~ U(0,1). Find the pdf of Y=$X^{-1/a}$ where a>0 . Compare the result with the Pareto density.
(v) Show that Y=1/X is a standard normal if the pdf of X is

$$f(x) = \frac{1}{\sqrt{2\pi}} x^{-2} e^{-1/(2x^2)}, \qquad x \in \mathbb{R}$$

(vi) X has a Weibull distribution. Show that $Y=(\alpha X)^{\beta}$ has an exponential distribution.

Multivariate Random Variables
Marginal Distribution, Joint Distribution, Conditional Distribution

When we discussed the probabilities of subsets A and B of the sample space Ω, we defined three probability concepts. The *marginal* probability is the probability of each event, P(A) and P(B), and the *joint* probability of A and B is the probability that both events will occur, $P(A \cap B)$. The *conditional* probability of A given B is the probability of A when the event B is known to have occurred, which is defined as $P(A|B)=P(A \cap B)/P(B)$ if $P(B) \neq 0$. We also introduced an important statistical property of events; Events A and B are *stochastically independent* if $P(A \cap B)=P(A)P(B)$. Stochastic independence of A and B means that the knowledge about event B does not add any new information about the probability of A, and vice versa. These definitions are carried over to the distribution of two random variables.

Let $A=\{\omega;\ X(\omega) \leq x\}$ and $B=\{\omega;\ Y(\omega) \leq y\}$ for random variables X and Y. Marginal probabilities P(A) and P(B) are nothing but the cdf's of X and Y: G(x)=P(A) and H(y)=P(B). When multiple random variables are under consideration, the cdf of each random variable is called the ***marginal*** *cumulative distribution function*. The joint probability of A and B is a function of both x and y, F(x,y), and it is called the ***joint*** *cumulative distribution function* of X and Y. The conditional probability of A given B, P(A|B)=F(x,y)/H(y), is the conditional distribution of X, given $Y \leq y$. The conditioning set in the conditional distribution needs not be of the form of set B. It is more common to consider the conditional distribution of X given a specific value of Y=y. This will be discussed in more detail after the joint probability density function is introduced. Finally, the independence of events A and B implies $P(A \cap B)=P(A)P(B)$, which is equivalent to writing F(x,y)=G(x)H(y).

Definition. ***Joint Cumulative Distribution Function***. The joint cdf of two random variables X and Y, denoted by F(x,y), is the joint probability of two events $A=\{\omega;\ X(\omega) \leq x\}$ and $B=\{\omega;\ Y(\omega) \leq y\}$:

$$F(x,y) = P\{\omega;X(\omega) \leq x,\ Y(\omega) \leq y\}$$

for any real numbers x and y.

Definition. ***Conditional Cumulative Distribution Function***. The conditional cdf of random variable Y given $X \leq x$, denoted by $F(y|X \leq x)$, is the conditional probability of event $B=\{\omega;\ Y(\omega) \leq y\}$, given $A=\{\omega;\ X(\omega) \leq x\}$:

$$F(y|X \leq x) = \frac{P\{\omega;X(\omega) \leq x,\ Y(\omega) \leq y\}}{P\{\omega;X(\omega) \leq x\}} = \frac{F(x,y)}{G(x)} \qquad if \quad G(x) \neq 0$$

for any real numbers x and y.

Definition. ***Stochastic Independence***. Random variables X and Y are stochastically independent if F(x,y)=G(x)H(y) for every x and y, where F(x,y) is the joint cdf of X and Y, and G(x) and H(y) are the marginal cdf's of X and Y, respectively.

The pdf of a discrete random variable X is defined as $g(x)=P\{\omega;\ X(\omega)=x\}$, and the pdf of a continuous random variable is defined as an integrand g(x) whose integral is equal to the cdf, $G(x)=\int g(x)dx$, if such an integrand exists for all x. The pdf of each random variable is called the ***marginal*** *probability density function* to distinguish it from the joint probability distribution of multiple random variables. The joint pdf of random variables X and Y is the straightforward extension of the marginal pdf.

Definition. ***Joint Probability Density Function***. The joint pdf f(x,y) of *discrete* random variables X and Y is the joint probability of the two events A={ω; X(ω)=x} and B={ω; Y(ω)=y}:

$$f(x,y) \;=\; P\{\omega; X(\omega)=x,\; Y(\omega)=y\}$$

When X and Y are continuous random variables with a cdf F(x,y), their joint pdf is a *nonnegative* function f(x,y) such that

$$F(x,y) \;=\; \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\,du\,dv$$

if such a function exists for all x and y.

The conditional probability density function for discrete random variables is defined just as the definition of the conditional cumulative distribution. Let X and Y be discrete random variables with a joint pdf f(x,y) and the marginal pdf h(y) for Y. Let A={ω; X(ω)=x} and B={ω; Y(ω)=y}. Then, the conditional probability P(B|A) is the probability of (Y=y) given that (X=x). This conditional probability, denoted by f(y|X=x), or more compactly by f(y|x), is given by

$$f(y|x) \equiv P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{f(x,y)}{g(x)} \quad if \quad P(A) = g(x) \neq 0$$

This explanation of conditional pdf does not apply when X and Y are absolutely continuous random variables, because P(X=x)=0 for any specific value y for a continuous random variable. However, this form of conditional pdf is employed for continuous random variables also.

Definition. ***Conditional Probability Density Function***. The conditional pdf of Y, given X=x, is defined as

$$f(y|x) \;=\; \frac{f(x,y)}{g(x)} \qquad if\; g(x) \neq 0$$

where g(x) is the marginal pdf of X and f(x,y) is the joint pdf of X and Y.

The conditional pdf f(y|x) is in general a function of the fixed value x, which can be considered as a parameter of the distribution of Y. We have defined earlier the conditional cdf of Y for a given event A={ω; X(ω)≤x}. We can now define the conditional cdf of Y for a given specific value of x as the integral of the conditional pdf defined above.

Definition. ***Conditional Cumulative Distribution Function***. The conditional cdf of Y, given X=x, is defined as

$$F(y|x) = \begin{cases} \displaystyle\sum_{\{i:\, y_i \le y\}} f(y_i|x) & \textit{for discrete random variables} \\ \displaystyle\int_{-\infty}^{y} f(s|x)ds & \textit{for continuous random variables} \end{cases}$$

where f(y|x) is the conditional pdf of Y given X=x.

The stochastic independence of X and Y can also be defined in terms of the density functions. For discrete random variables, let A={ω; X(ω)=x} and B={ω; Y(ω)=y}. The independence of A and B, P(A∩B)=P(A)P(B), implies f(x,y)=g(x)h(y). For continuous random variables, independence of X and Y implies F(x,y)=G(x)H(y), and differentiation of this relationship gives the expression for the independence X and Y in terms of density functions:

$$f(x,y) = \frac{dF(x,y)}{dx\,dy} = \frac{d}{dx\,dy}\big[G(x)H(y)\big] = g(x)h(y)$$

Definition. ***Stochastic Independence***. Random variables X and Y are stochastically independent if and only if f(x,y)=g(x)h(y) for every x and y, where f(x,y) is the joint pdf of X and Y, and g(x) and h(y) are the marginal pdf's of X and Y, respectively.

Properties of Joint and Conditional Distributions

We have shown that a marginal cdf is a right continuous and nondecreasing function, and it takes a value in a closed unit interval, and that the marginal pdf is a nonnegative function and the area under the continuous pdf must be one. Similar properties hold for the joint cdf and the joint pdf.

Theorem. The joint cdf F(x,y) has the following properties:

(1) (a) $0 \le F(x,y) \le 1$, (b) F(x,y) = 0 if ***either*** $x \to -\infty$ ***or*** $y \to -\infty$, (c) F(x,y) = 1 if $x \to \infty$ ***and*** $y \to \infty$
(2) F(x,y) is monotonically nondecreasing in x and y.
(3) F(x,y) is right-continuous.

Proof. (1) (a) F(x,y) must take a value in the unit interval because it represents a probability. (b) This follows from the fact that the set A={ω; X(ω)≤x} becomes an empty set ∅ as x becomes a negative infinity, and hence A∩B=∅. (c) As both x and y become a positive infinity, both sets A and B become the sample space Ω, and hence, A∩B=Ω. The joint cdf takes a zero value if *either* x *or* y becomes negative infinite, but it takes a value of 1 if *both* x and y become positive infinite.
(2) Let A_i={ω; X(ω)≤x_i}, i=1,2, and $x_1 < x_2$. Then, $A_1 \subseteq A_2$ and hence (A_1∩B) is a subset of (A_2∩B), which implies $F(x_1,y) \le F(x_2,y)$ for all $x_1 < x_2$. A similar argument applies to y.
(3) This is a technical property that arises from the specification of the set by a weak inequality X(ω)≤x rather than a strict inequality.

The marginal pdf g(x) of a continuous random variable is represented by a curve that lies above the horizontal axis in a two dimensional space. The marginal cdf G(x) is the area under the density curve to the left of point x, and the area under the entire density curve is 1. The joint pdf f(x,y) of continuous random variables X and Y is represented by a surface in a three-dimensional space. The density surface lies above the floor: it can touch the floor, but can not cross it. The joint cdf F(x,y) is the volume under the surface that lies to the south-west of the point (x,y). The volume under the entire surface must be equal to 1. Similar interpretations hold for discrete random variables, and these properties are summarized in the following theorem.

Theorem. The joint pdf f(x,y) has the following relationships.

(1) $f(x,y) \geq 0 \quad \textit{for all } x \textit{ and } y$

$$(2) \quad F(x,y) = \begin{cases} \displaystyle\sum_{\{j;\, y_j \leq y\}} \sum_{\{i;\, x_i \leq x\}} f(x_i,y_j) & \textit{for discrete } X \textit{ and } Y \\ \displaystyle\int_{-\infty}^{y} \int_{-\infty}^{x} f(t,s)\,dt\,ds & \textit{for continuous } X \textit{ and } Y \end{cases}$$

$$(3) \quad \sum_{y_i} \sum_{x_i} f(x_i,y_i) = 1 \textit{ for discrete } X \textit{ and } Y, \qquad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1 \textit{ for continuous } X \textit{ and } Y$$

▣

These properties of the joint pdf and the joint cdf are parallel to those of the marginal pdf and cdf. Additional properties that arise from the multivariate distribution are the relationships between the marginal cdf and joint cdf, and between the marginal pdf and the joint pdf, as stated in the following theorem.

Theorem. Let f(x,y) and F(x,y) be the joint pdf and cdf of X and Y, and g(x) and G(x) be the marginal pdf and the marginal cdf of X, respectively. Then,

$$(1) \quad G(x) = \lim_{y \to \infty} F(x,y)$$

$$(2) \quad g(x) = \begin{cases} \displaystyle\sum_{y_j} f(x,y_j) & \textit{for discrete } X \textit{ and } Y \\ \displaystyle\int_{-\infty}^{\infty} f(x,y)\,dy & \textit{for continuous } X \textit{ and } Y \end{cases}$$

Proof. (1) Let A={ω; X(ω)≤x} and B={ω; Y(ω)≤y}. Note that $\lim_{y \to \infty} B = \{\omega\,; Y(\omega) \leq \infty\} = \Omega$. Hence,

$$\lim_{y \to \infty} F(x,y) = \lim_{y \to \infty} P\{\omega; X(\omega) \leq x,\, Y(\omega) \leq y\} = \lim_{y \to \infty} P\{A \cap B\} = P\{A \cap \Omega\} = P(A) = P\{\omega; X(\omega) \leq x\} = G(x)$$

(2) For discrete random variables, let A={ω; X(ω)=x} and B_j={ω; Y(ω)=y_j}, j=1,2,···n. Then, $A = \bigcup_{j=1}^{n} (A \cap B_j)$ and $(A \cap B_j)$ are mutually exclusive. Hence,

$$g(x) = P(A) = \sum_{j=1}^{n} P(A \cap B_j) = \sum_{j=1}^{n} f(x,y_j)$$

The marginal pdf g(x) of a discrete random variable is obtained by summing all probability for different values of y. For continuous random variables, we have from (1) above

$$G(x) = \lim_{y\to\infty} F(x,y) = \int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(s,y)dy\right)ds$$

Hence,

$$g(x) = \frac{dG(x)}{dx} = \frac{d}{dx}\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f(s,y)dy\right)ds = \int_{-\infty}^{\infty} f(x,y)dy \qquad \boxed{\bullet}$$

The joint pdf of continuous random variables X and Y is a surface in a three-dimensional space. If we push this surface against the x-axis wall, piling the surface up, the top profile of the pile on the x-axis wall is the marginal pdf g(x). If we cut the surface with a knife perpendicular to the Y-axis at a point Y=y, the conditional pdf is the upper line of the cut, blown up by the factor 1/h(y). We need to enlarge the upper line to make the area under the line to be equal to 1.

Theorem: Let f(x,y) and F(x,y) be the joint pdf and cdf of X and Y, f(y|x) and F(y|x) be the conditional pdf and cdf of Y given X=x, and g(x) be the marginal pdf of X, respectively. Then,

$$(a) \quad f(x,y) = f(y|x)g(x)$$

$$(b) \quad F(x,y) = \begin{cases} \displaystyle\sum_{\{i:\, x_i \le x\}} F(y|x_i)g(x_i) & \textit{for discrete X and Y} \\ \displaystyle\int_{-\infty}^{x} F(y|s)g(s)ds & \textit{for continuous X and Y} \end{cases}$$

Proof. (a) is obvious from the definition of the conditional pdf.
(b) This can be shown by using the relationships among the joint cdf, joint pdf and the conditional cdf. We will prove the case of continuous random variables.

$$\int_{-\infty}^{x} F(y|s)g(s)ds = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f(t|s)dt\right)g(s)ds = = \int_{-\infty}^{x}\int_{-\infty}^{y}\frac{f(s,t)}{g(s)}g(s)dtds = \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)dsdt = F(x,y)$$

where we used the definition of the conditional cdf and pdf. $\boxed{\bullet}$

Theorem. Let f(x,y) and F(x,y) be the joint pdf and cdf of X and Y. Then, for a<b and c<d,

$$P(a<X\le b,\ c<Y\le d) = F(b,d) - F(b,c) - F(a,d) + F(a,c) = \int_{c}^{d}\int_{a}^{b} f(x,y)dx\,dy$$

Proof. Let A={ω; X(ω)≤a}, B={ω; X(ω)≤b}, C={ω; Y(ω)≤c} and D={ω; Y(ω)≤d}. Then,

$$P\{\omega\,; a<X(\omega)\le b,\, c<Y(\omega)\le d\} = P\{(B-A)\cap(D-C)\} = P\{(B\cap D)-(B\cap C)-(A\cap D)+(A\cap C)\}$$
$$= P\{(B\cap D)\} - P\{(B\cap C)\} - P\{(A\cap D)\} + P\{(A\cap C)\}$$
$$= F(b,d) - F(b,c) - F(a,d) + F(a,c)$$

The second equality for continuous random variables is obvious. ■

Theorem. If X and Y are stochastically independent, then

(a) $f(x|y)=g(x)$ and $f(y|x)=h(y)$

(b) $P(a<X\le b,\ c<Y\le d) = P(a<X\le b)P(c<Y\le d)$.

(c) any measurable functions $Z=u(X)$ and $W=v(Y)$ are also independent.

Proof. (a) This is obvious from the definition of the condition pdf and independence.

(b) Since $F(x,y)=G(x)H(y)$ for independent random variables X and Y,

$$P(a<X\le b,\ c<Y\le d) = F(b,d) - F(b,c) - F(a,d) + F(a,c) = G(b)H(d) - G(b)H(c) - G(a)H(d) + G(a)H(c)$$
$$= [G(b)-G(a)][H(d)-H(c)] = P(a<X\le b)P(c<Y\le d)$$

(c) We need to show that, if X and Y are independent, $P(Z\le z,\ W\le w) = P(Z\le z)P(W\le w)$. We will consider only the monotonic functions, so that the inverse functions $u^{-1}(z)$ and $v^{-1}(w)$ exist.

$$P(Z\le z,\ W\le w) = P(X\le u^{-1}(z),\ Y\le v^{-1}(w)) = P(X\le u^{-1}(z))\,P(Y\le v^{-1}(w)) = P(Z\le z)\,P(W\le w) \quad ■$$

Example. Toss a pair of fair coins and let the sample space $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}$, where ω_1=(T,T), ω_2=(H,T), ω_3=(T,H), ω_4=(H,H), and $P(\omega_i)=1/4$, for all i. Define two random variables X and Y as follows. Random variable X takes 0 if the two coins show different faces and takes 1 if same faces. Random variable Y denotes the number of heads shown.

$$X(\omega) = \begin{cases} 0 & \textit{if } \omega=\omega_2 \textit{ or } \omega_3 \\ 1 & \textit{if } \omega=\omega_1 \textit{ or } \omega_4 \end{cases} \qquad Y(\omega) = \begin{cases} 0 & \textit{if } \omega=\omega_1 \\ 1 & \textit{if } \omega=\omega_2 \textit{ or } \omega_3 \\ 2 & \textit{if } \omega=\omega_4 \end{cases}$$

We wish to find the probability distributions of X and Y individually as well as jointly. The *marginal* pdf's are

$$g(x) = \begin{cases} 1/2 & \textit{if } x=0 \textit{ or } 1 \\ 0 & \textit{otherwise} \end{cases} \qquad h(y) = \begin{cases} 1/4 & \textit{if } y=0 \textit{ or } 2 \\ 1/2 & \textit{if } y=1 \\ 0 & \textit{otherwise} \end{cases}$$

The joint pdf of X and Y, $f(x,y) = P(X=x, Y=y)$, is computed for all combinations of the values of X and Y.

$f(0,0) = P(X=0, Y=0) = P(\{\omega_2,\omega_3\}\cap\{\omega_1\}) = P(\varnothing) = 0$

$f(0,1) = P(X=0, Y=1) = P(\{\omega_2,\omega_3\}\cap\{\omega_2,\omega_3\}) = P(\{\omega_2,\omega_3\})=1/2$

$f(0,2) = P(X=0, Y=2) = P(\varnothing) = 0$,

$f(1,0) = P(\omega_1) = 1/4$, $\quad f(1,1) = P(\varnothing) = 0$, $\quad f(1,2) = P(\omega_4) = 1/4$

$f(x,y) = 0$ for all other x, y.

Joint and Marginal pdf's

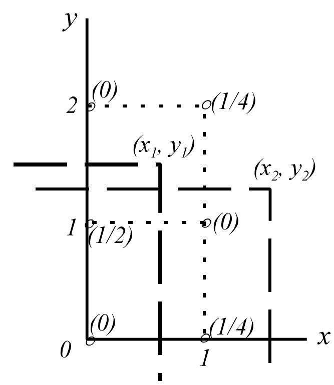
		Y			g(x)
		0	1	2	
X	0	0	1/2	0	1/2
	1	1/4	0	1/4	1/2
h(y)		1/4	1/2	1/4	

Conditional pdf f(y|x)

		Y		
		0	1	2
X	0	0	1	0
	1	1/2	0	1/2

The marginal pdf of Y is computed by summing all probabilities in the column for each value of Y. Similarly, the marginal pdf of X is computed by summing all probabilities in each row of X values. The conditional pdf f(y|x) is derived by dividing each cell of joint pdf with each marginal pdf of X. One readily sees that X and Y are not independent since the conditional pdf f(y|x) in the second table is not equal to the marginal pdf h(y). This can also be seen from the first table by noting that f(0,0)=0 is not equal to g(0)h(0)=1/8.

The cdf F(x, y) is the sum of the probabilities of all points that lie in the south-west area from the point (x, y). Thus, in the figure below, $F(x_1, y_1)=1/2$ and $F(x_2, y_2)=3/4$.



Example: *Bivariate Uniform Distribution*. Random variables X and Y have a joint uniform distribution if their joint pdf is given by

$$f(x,y) = \begin{cases} \dfrac{1}{(b-a)(d-c)} & \text{if} \quad x \in [a, b] \quad \text{and} \quad y \in [c, d] \\ 0 & \text{otherwise} \end{cases}$$

Jointly uniform random variables have a uniform marginal pdf, and they are stochastically independent, as shown below.

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{c}^{d} \frac{1}{(b-a)(d-c)} dy = \frac{1}{b-a}, \quad \text{if } x \in [a, b]$$

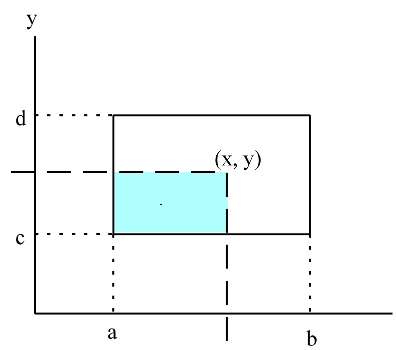
$$h(y) = \frac{1}{d-c}, \quad \text{if } y \in [c, d]$$

and hence, f(x,y)=g(x)h(y). Stochastic independence also implies that conditional pdf is equal to the marginal pdf;

$$f(y|x) = \frac{1}{d-c}, \quad \text{if } x \in [a,b] \text{ and } y \in [c,d]$$

which does not depend on the value of y.

The joint pdf is the top of a box with the base area (b-a)(d-c) and volume equal to 1. The cdf can be found by finding volume over the south-west area from the point (x,y) defined by [-∞, x] and [-∞, y]. Since the height of the box (pdf) is zero if either x<a or y<c, we have F(x, y)=0 if x<a or y<c. If x∈[a, b] and y∈[c, d], then there is a positive height over the area with base (x-a)(y-c), and hence the total volume is (x-a)(y-c)/[(b-a)(d-c)]. If x∈[a, b] and y>d, then F(x,y)=(x-a)/(b-a), and etc.



Example. ***Bivariate Normal Distribution***. Continuous random variables X and Y are said to have a jointly normal distribution if their joint pdf is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-Q/[2(1-\rho^2)]} \qquad x \in \Re, \quad y \in \Re, \quad -1 < \rho < 1$$

where

$$Q \equiv \tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2, \qquad \tilde{x} = (x-\mu_x)/\sigma_x, \qquad \tilde{y} = (y-\mu_y)/\sigma_y$$

and parameters μ_x, μ_y, σ_x, and σ_y are the means and standard deviations of X and Y, and ρ is called the *correlation coefficient* of X and Y. This joint pdf is a symmetrical bell-shaped surface with center (μ_x, μ_y) when ρ=0. The multivariate normal random variables play an extremely important role in theory as well as in practice. Their marginal distribution, conditional distribution and the condition for the stochastic independence are presented in the following theorem.

Theorem: Let X and Y be bivariate normal random variables with means μ_x and μ_y, standard deviations σ_x and σ_y, and correlation coefficient ρ. Then,

(a) The marginal distributions are normal: $X \sim N(\mu_x, \sigma_x^2), \quad Y \sim N(\mu_y, \sigma_y^2)$.
(b) The conditional distribution of Y given X is a normal with mean and variance

$$\mu_{y|x} = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x), \qquad \sigma^2_{y|x} = (1-\rho^2)\sigma^2_y$$

(c) X and Y are independent if and only if the correlation coefficient ρ=0.

Proof. (a) Rewrite Q by subtracting and adding term $\rho^2\tilde{x}^2$ to derive

$$\frac{Q}{2(1-\rho^2)} = \frac{(1-\rho^2)\tilde{x}^2 + \rho^2\tilde{x}^2 - 2\rho\tilde{x}\tilde{y} + \tilde{y}^2}{2(1-\rho^2)} = \frac{\tilde{x}^2}{2} + \frac{(\tilde{y}-\rho\tilde{x})^2}{2(1-\rho^2)} = \frac{(x-\mu_x)^2}{2\sigma^2_x} + \frac{(y-\mu_{y|x})^2}{2\sigma^2_{y|x}}$$

where $\mu_{y|x}$ and $\sigma^2_{y|x}$ are as defined above. The joint pdf can be written as

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-Q/[2(1-\rho^2)]} = \frac{1}{\sqrt{2\pi\sigma^2_x}}e^{-(x-\mu_x)^2/2\sigma^2_x}\cdot\frac{1}{\sqrt{2\pi\sigma^2_{y|x}}}e^{-(y-\mu_{y|x})^2/2\sigma^2_{y|x}}$$

The marginal density of X is given by

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2\pi\sigma^2_x}}e^{-(x-\mu_x)^2/2\sigma^2_x}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma^2_{y|x}}}e^{-(y-\mu_{y|x})^2/2\sigma^2_{y|x}}dy = \frac{1}{\sqrt{2\pi\sigma^2_x}}e^{-(x-\mu_x)^2/2\sigma^2_x}$$

which is a normal pdf with mean μ_x and variance σ^2_x. The integral in the third expression is equal to 1 because it is an integral of a form of a normal density with mean $\mu_{y|x}$ and variance $\sigma^2_{y|x}$. The normality of marginal pdf of Y can be shown in a similar way.

(b) From the expressions for f(x,y) and g(x), one can see

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{1}{\sqrt{2\pi\sigma^2_{y|x}}}e^{-(y-\mu_{y|x})^2/2\sigma^2_{y|x}}$$

which is a normal density with mean $\mu_{y|x}$ and variance $\sigma^2_{y|x}$.

(c) X and Y are independent if and only if f(x,y)=g(x)h(y) or f(y|x)=h(y). Since $Y \sim N(\mu_y, \sigma^2_y)$ and f(y|x) is a normal density, the equality f(y|x)=h(y) holds if and only if

$$\mu_{y|x} = \mu_y, \quad \sigma^2_{y|x} = \sigma^2_y$$

It is easy to verify that both equalities hold if and only if ρ=0. ▣

Example. ***Farlie-Morgenstern Family***. Let g(x) and G(x) be the pdf and cdf of X, and h(y) and H(y) be the pdf and cdf of Y. For any $\alpha\in$[-1, 1], let

$$F(x,y;\alpha) = G(x)H(y)\{1 + \alpha[1-G(x)][1-H(y)]\}$$

Then, F(x, y; α) is the joint cdf's of X and Y, each having the marginal cdf's G(x) and H(y), respectively. The joint pdf is given by

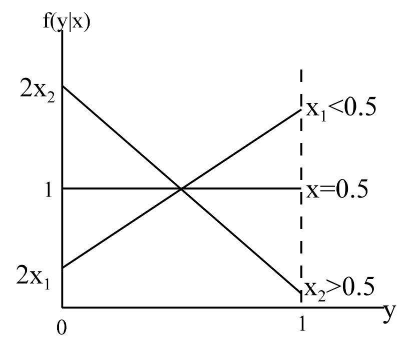
$$f(x,y;\alpha) = g(x)h(y)\{1 + \alpha[2G(x) - 1][2H(y) - 1]\}$$

This family of distributions provides a convenient way to construct various joint distribution from two given marginal distributions, if the closed forms of cdf are given.

Consider for an example the uniform random variables X~U(0,1) and Y~U(0,1). Since g(x)=1, G(x)=x for x∈(0,1) and h(y)=1, H(y)=y for y∈(0,1), the Farlie-Morgenstern family of density functions is

$$f(x,y) = \begin{cases} 1 + \alpha(2x-1)(2y-1), & x \in (0,1), \quad y \in (0,1) \\ 0 & otherwise \end{cases}$$

Note that if α=0, f(x,y) is the jointly uniform pdf. Since g(x)=1 and h(y)=1, the joint pdf is also the conditional pdf. If $\alpha \neq 0$, f(y|x) depends on the value of x as drawn in the figure for the case of α=-1, and X and Y are not independent.



Example. Consider a joint pdf

$$f(x,y) = \begin{cases} \lambda^2 e^{-\lambda y} & if \quad 0 \le x \le y \\ 0 & otherwise \end{cases}$$

where λ>0. The marginal and conditional pdf's are:

$$g(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{x}^{\infty} \lambda^2 e^{-\lambda y} dy = \lambda e^{-\lambda x} \quad for \ x \ge 0, \qquad h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{0}^{y} \lambda^2 e^{-\lambda y} dx = \lambda^2 y e^{-\lambda y} \quad for \ y \ge 0$$

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{1}{y} \quad for \ y \ge x \ge 0 \ and \ y > 0, \qquad f(y|x) = \frac{f(x,y)}{g(x)} = \lambda e^{-\lambda(y-x)} \quad for \ y \ge x \ge 0$$

⊡

Example. Consider a joint pdf

$$f(x,y) = \begin{cases} 2 & if \quad x+y \le 1, \ x \ge 0, \ y \ge 0 \\ 0 & otherwise \end{cases}$$

The marginal pdf of Y and the conditional pdf of X given Y=y are

$$h(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_{-\infty}^{0} f(x,y)dx + \int_{0}^{1-y} f(x,y)dx + \int_{1-y}^{\infty} f(x,y)dx = 0 + \int_{0}^{1-y} 2dx + 0 = 2(1-y)$$

$$f(x|y) = \begin{cases} \dfrac{f(x,y)}{h(y)} = \dfrac{2}{2(1-y)} = \dfrac{1}{1-y} & \text{if } 0 \le x \le 1-y, \quad 0 \le y < 1 \\ 0 & \text{otherwise} \end{cases}$$

◘

Example. ***Count Models***. A Poisson distribution is often used in *count models*. For example, suppose that the number of major derogatory reports Y on an individual's credit history in a given time period is distributed as a Poisson random variable with parameter α

$$f(y;\alpha) = \frac{\alpha^{y} e^{-\alpha}}{y!}, \qquad y = 0,1,2,\cdots, \qquad \alpha > 0$$

where α is the average number of major derogatory reports in a given time period. In economics applications, the mean α is assumed to be a function of individual's socio-economic characteristics, such as education level, income, age, etc. For example, for individual i with a characteristics x_i, one may assume a linear function $\alpha_i=\beta_0+\beta_1 x_i$, or a log-linear function $ln\alpha_i=\beta_0+\beta_1 x_i$. The latter is a more common specification in practice, because the linear function may lead to a negative value of α_i. The objective is to estimate β_0 and β_1 from a cross section data.

A drawback of using a Poisson distribution in this type of analysis is that its variance is equal to α, i.e., E(Y)=var(Y), which we have shown earlier. This is an undesirable restriction, because the sample variance of major derogatory reports across individuals exceeds the sample mean. To relax the restriction in the Poisson model, the mean α is treated as a random variable, and the Poisson pdf is interpreted as a conditional pdf of Y given α. For the log-linear model, we add a random variable ε_i: $ln\alpha_i=\beta_0+\beta_1 x_i+\varepsilon_i$. Droping the subscript, we can write

$$\alpha = e^{\beta_0+\beta_1 x} e^{\epsilon} = \delta e^{\epsilon}, \qquad \delta = e^{\beta_0+\beta_1 x}$$

If the conditional pdf of Y given α is Poisson as written above, what is the marginal pdf of Y? Let $g(\varepsilon)$ be the marginal pdf of ε. Then, the joint pdf of Y and ε is $f(y;\alpha)g(\alpha)$, and hence, the marginal pdf of Y is

$$h(y) = \int_{0}^{\infty} f(y;\alpha)g(\alpha)d\alpha$$

If ε is normal or if e^{ε} is lognormal, the closed form for h(y) can not be found. However, if ε is distributed as an *exponential Gamma*, i.e., if e^{ε} is distributed as a Gamma, then we can find a closed form for the marginal pdf h(y). Recall the pdf of a Gamma random variable $Z=e^{\varepsilon}$ with parameters λ and k:

$$f(z;k,\lambda) = \frac{\lambda^{k}}{\Gamma(k)} z^{k-1} e^{-\lambda z}, \quad z \ge 0, \quad \lambda > 0, \quad k > 0, \qquad \Gamma(k) = \int_{0}^{\infty} w^{k-1} e^{-w} dw$$

Since $\alpha=\delta Z$ is a linear function of Z and $\delta>0$, the pdf of α is

$$g(\alpha) = \frac{1}{\delta} f\left(\alpha/\delta\,; k,\lambda\right) = \frac{\lambda^k}{\delta\,\Gamma(k)} (\alpha/\delta)^{k-1} e^{-\lambda(\alpha/\delta)} = \frac{\lambda^k}{\delta^k\,\Gamma(k)} \alpha^{k-1} e^{-(\lambda/\delta)\alpha}$$

Substituting this into the equation for the marginal pdf of Y

$$h(y) = \int_0^\infty f(y;\alpha) g(\alpha) d\alpha = \int_0^\infty \frac{\alpha^y e^{-\alpha}}{y!} \frac{\lambda^k}{\delta^k\,\Gamma(k)} \alpha^{k-1} e^{-(\lambda/\delta)\alpha}\, d\alpha = \frac{\lambda^k}{y!\,\delta^k\,\Gamma(k)} \int_0^\infty \alpha^{y+k-1} e^{-\alpha(\delta+\lambda)/\delta}\, d\alpha$$

$$= \frac{\lambda^k}{y!\,\delta^k\,\Gamma(k)} \left(\frac{\delta}{\delta+\lambda}\right)^{y+k} \int_0^\infty w^{y+k-1} e^{-w}\, dw, \qquad \textit{by setting } w = \frac{\alpha(\delta+\lambda)}{\delta}$$

$$= \frac{\lambda^k}{y!\,\delta^k\,\Gamma(k)} \left(\frac{\delta}{\delta+\lambda}\right)^{y+k} \Gamma(y+k) = \frac{\Gamma(y+k)}{\Gamma(k)\,y!} \theta^k (1-\theta)^y, \qquad \theta = \frac{\lambda}{\delta+\lambda}$$

Since Γ(k)=(k-1)! for an integer k, this is a pdf of a *negative binomial* random variable, which represents the number of failures before *k*th success in repeated independent Bernoulli trials with success probability θ. The mean and variance of a negative binomial random variable are

$E(Y) = k(1-\theta)/\theta$, $\quad$ *var*$(Y) = k(1-\theta)/\theta^2$

and the variance is greater than the mean, because $\theta<1$. Substituting θ into these expressions, we have

$E(Y) = k\delta/\lambda$, $\quad$ *var*$(Y) = k\delta(\delta+\lambda)/\lambda^2$.

In practice, a restriction $k=\lambda$ is imposed in the choice of the Gamma pdf. This simplifies the unconditional mean and variance of Y to

$$E(Y) = \delta = e^{\beta_0+\beta x}, \qquad var(Y) = \delta + \delta^2/\lambda$$

Note that the mean is same as the mean of the Poisson model, while the variance is greater than the mean.

The conditional pdf of α given y can also be found easily:

$$f(\alpha\,|y) = \frac{f(y;\alpha) g(\alpha)}{h(y)} = \frac{\tau^m}{\Gamma(m)} \alpha^{m-1} e^{-\tau\alpha}, \qquad m = y+k, \quad \tau = 1/(1-\theta)$$

which is a Gamma pdf with shape parameter m and scale parameter τ. Hence, the conditional mean and variance of α given y are

$$E(\alpha|y) = \frac{m}{\tau} = (1-\theta)(y+k) = \frac{\delta(y+k)}{\delta+\lambda}$$

$$var(\alpha|y) = \frac{m}{\tau^2} = (1-\theta)^2(y+k) = \frac{\delta^2(y+k)}{(\delta+\lambda)^2}$$

The analysis presented above is a Bayesian type analysis. The parameter α in the Poisson model is viewed as a random variable in Bayesian analysis (α varies from individual to individual in the population). The pdf g(α) is called the *prior density function* of α, and the conditional pdf of α given y is called the *posterior density function* of α. The Bayesian estimate of α is E(α|y), because this is the best predictor of α given Y=y.

Extensions to More than Two Random Variables

When there are n random variables, $X_1, X_2, \cdots, X_n$, their joint cdf is defined as

$$F(x_1,x_2,\ldots,x_n) = P\{\omega\,;X_1(\omega) \le x_1, X_2(\omega) \le x_2, \cdots, X_n(\omega) \le x_n\}$$

for any real numbers x_i, $i=1,2,\cdots,n$, and its properties are

(1) (a) $0 \le F(x_1,x_2,\cdots,x_n) \le 1$, (b) $F(x_1,x_2,\cdots,x_n) = 0$ if $x_i \to -\infty$ for ***any*** i,
(c) $F(x_1,x_2,\cdots,x_n) = 1$ if $x_i \to \infty$ for ***all*** i
(2) $F(x_1,x_2,\cdots,x_n)$ is monotonically nondecreasing in *each* x_i.
(3) $F(x_1,x_2,\cdots,x_n)$ is right-continuous.

The joint pdf is also defined analogously. For discrete random variables, it is

$$f(x_1,x_2,\ldots,x_n) = P\{\omega\,;X_1(\omega) = x_1, X_2(\omega) = x_2, \cdots, X_n(\omega) = x_n\}$$

and for continuous random variables it is the *nonnegative* integrand such that

$$F(x_1,x_2,\ldots,x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(s_1,s_2,\ldots,s_n)\, ds_1, ds_2, \ldots, ds_n$$

if such a function exists for all x_i.

The marginal pdf and cdf of a subset $(X_1,\cdots X_m)$ are derived from the joint distribution as

$$G(x_1,x_2,\ldots,x_m) = \lim_{\{x_j \to \infty\,;\, j \ge m+1\}} F(x_1,x_2,\ldots,x_m,\, x_{m+1}, x_{m+2}, \cdots, x_n)$$

$$g(x_1,x_2,\ldots,x_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1,x_2,\ldots,x_m,x_{m+1},x_{m+2},\cdots,x_n)\, dx_{m+1}, dx_{m+2}, \ldots, dx_n$$

Let $g(\cdot)$ and $h(\cdot)$ be the marginal distributions of subsets $(X_1,\cdots X_m)$ and $(X_{m+1},\cdots X_n)$. The conditional pdf of $(X_1,\cdots X_m)$ given $(X_{m+1},\cdots X_n)$ is defined as

$$f(x_1,x_2,\ldots,x_m \mid x_{m+1},x_{m+2},\cdots,x_n) = \frac{f(x_1,x_2,\ldots,x_m,x_{m+1},x_{m+2},\cdots,x_n)}{h(x_{m+1},x_{m+2},\cdots,x_n)}$$

and the two groups of random variables $(X_1,\cdots X_m)$ and $(X_{m+1},\cdots X_n)$are stochastically independent if and only if

$$f(x_1,x_2,\ldots,x_m,x_{m+1},x_{m+2},\cdots,x_n) = g(x_1,x_2,\ldots,x_m) \cdot h(x_{m+1},x_{m+2},\cdots,x_n)$$

or if and only if the conditional pdf is equal to the marginal pdf

$$f(x_1,x_2,...,x_m \,|\, x_{m+1},x_{m+2},\cdots,x_n) = g(x_1,x_2,...,x_m)$$

In particular, all random variables X_1, X_2,...,X_n are *mutually independent* if and only if

$$F(x_1,...,x_n) = \prod_{i=1}^{n} F_i(x_i) \quad \text{or} \quad f(x_1,...,x_n) = \prod_{i=1}^{n} f_i(x_i),$$

where $F_i(x_i)$ and $f_i(x_i)$ are the marginal cdf and pdf of X_i. It is important to note that the random variables are *pairwise* independent, but all of the n random variables are not independent (see an example below).

Multinomial Probability Distribution

The binomial distribution arises from n repeated ***independent*** tosses of a coin, where each trial can produce one of the two possible outcomes, head or tail. When the probability of head in each trial is θ, the number of heads in n trials, denoted by a random variable X, has the binomial pdf

$$f(x;n,\theta) = \begin{cases} \binom{n}{x} \theta^x (1-\theta)^{n-x} & \text{if } x=0,1,....,n \\ 0 & \text{otherwise} \end{cases}$$

and $E(X)=n\theta$ and $var(X)=n\theta(1-\theta)$.

Consider another experiment of n independent repeated trials where

(i) each trial can result in k possible outcomes (classes or cells), $k \geq 2$,

(ii) the probability that the outcome of each trial to fall into cell i is θ_i, $i=1,2,\cdots k$, and $\Sigma\theta_i = 1$

Let X_i be the number of trials whose outcome falls into cell i such that $\Sigma X_i = n$. The joint pdf for X_1, X_2,$\cdots X_k$ is given by

$$f(x_1,x_2,\cdots,x_k;n,\theta_1,\theta_2,\cdots\theta_k) = \begin{cases} \dfrac{n!}{x_1!x_2!\cdots x_k!}\theta_1^{x_1}\theta_2^{x_2}\cdots\theta_k^{x_k} & \text{if } x_i\text{'s are nonnegative integers such that } \sum_{i=1}^{k} x_i = n \\ 0 & \text{otherwise} \end{cases}$$

where $\sum_{i=1}^{k}\theta_i = 1$. This is the joint pdf of multinomial random variables.

The marginal pdf of X_i is binomial with parameters θ_i and n. This can be seen easily by treating all other sells as another cell beside the cell i. Therefore, the mean and variance of each random variable X_i are

$$E(X_i) = n\theta_i, \qquad var(X_i) = n\theta_i(1-\theta_i), \quad i=1,2,\cdots,k$$

The covariance between X_i and X_j is given by

$$cov(X_i, X_j) = -n\theta_i\theta_j, \qquad i \neq j = 1,2,\cdots,k$$

To prove this result, recall that the marginal distribution of X_i is a binomial with parameters θ_i and n. Similarly, treating cells i and j as one cell, one easily sees that X_i+X_j has a binomial distribution with parameters $(\theta_i+\theta_j)$ and n. Therefore,

$$var(X_i+X_j) = n(\theta_i+\theta_j)(1-\theta_i-\theta_j)$$

Using the relationship

$$var(X_i+X_j) = var(X_i) + var(X_j) + 2cov(X_i,X_j)$$

we can derive

$$\begin{aligned} cov(X_i,X_j) &= \frac{1}{2}\left[var(X_i+X_j) - var(X_i) - var(X_j)\right] \\ &= \frac{n}{2}\left[(\theta_i+\theta_j)(1-\theta_i-\theta_j) - \theta_i(1-\theta_i) - \theta_j(1-\theta_j)\right] \\ &= -n\theta_i\theta_j \end{aligned}$$

Example. Independent samples from multinomial distributions
Suppose we have two *independent* samples X_i and Y_i of size n and m from two multinomial populations. Consider the difference in frequencies in each cell between the two samples: $Z_i \equiv X_i/n - Y_i/m$. If the two populations have the same distribution function, we have

$$\begin{aligned} var(Z_i) \equiv var\left(\frac{X_i}{n} - \frac{Y_i}{m}\right) &= \frac{1}{n^2}var(X_i) + \frac{1}{m^2}var(Y_i) - 2\frac{1}{nm}cov(X_i, Y_i) \\ &= \frac{1}{n}\theta_i(1-\theta_i) + \frac{1}{m}\theta_i(1-\theta_i) - 0 \\ &= \frac{n+m}{nm}\theta_i(1-\theta_i) \end{aligned}$$

where $cov(X_i, Y_i) = 0$ because the two samples are independent samples. Similarly, $cov(Z_i, Z_j)$ can be derived by using the procedure employed above:

$$\begin{aligned} var(Z_i + Z_j) \equiv var\left(\frac{X_i + X_j}{n} - \frac{Y_i + Y_j}{m}\right) &= \frac{1}{n^2}var(X_i + X_j) + \frac{1}{m^2}var(Y_i + Y_j) - 2\frac{1}{nm}cov(X_i + X_j, Y_i + Y_j) \\ &= \frac{1}{n}(\theta_i+\theta_j)(1-\theta_i-\theta_j) + \frac{1}{m}(\theta_i+\theta_j)(1-\theta_i-\theta_j) - 0 \\ &= \frac{n+m}{nm}(\theta_i+\theta_j)(1-\theta_i-\theta_j) \end{aligned}$$

so that

$$\begin{aligned} cov(Z_i, Z_j) &= \frac{1}{2}\left[var(Z_i + Z_j) - var(Z_i) - var(Z_j)\right] \\ &= \frac{n+m}{2nm}\left[(\theta_i + \theta_j)(1 - \theta_i - \theta_j) - \theta_i(1-\theta_i) - \theta_j(1-\theta_j)\right] \\ &= -\frac{n+m}{nm}\theta_i\theta_j \end{aligned}$$

These results can be used in testing of the equality of two population distributions. See Gordon Anderson, "Nonparametric Tests of Stochastic Dominance in Income Distributions," Econometrica, Vol. 64, No. 5 (September, 1996), 1183-1193. See also C. R. Rao, section 6.b and 6.c.

Assignment #5

(1) For each of the following joint pdf's of X and Y, find the marginal pdf's g(x) and h(y), and conditional pdf's f(x|y) and f(y|x), and determine the independence of X and Y.

$(i)\quad f(x,y) = xe^{-x(y+1)}, \quad for\ 0 \le x < \infty, \quad 0 \le y < \infty, \quad and\ f(x,y) = 0,\ otherwise$

$(ii)\quad f(x,y) = 2 \quad for\ x > 0,\ y > 0,\ x + y < 1, \quad and\ f(x,y) = 0\ otherwise$

$(iii)\quad f(x,y) = (x+y)/30 \quad for\ x = 0,1,2,3, \quad y = 0,1,2, \quad and\ f(x,y) = 0\ otherwise$

(2) Show that X_1 and X_3, and X_2 and X_3, are *pairwise* independent, but all three r.v.'s are not independent, for the following joint pdf.

$$f(x_1,x_2,x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & for\ 0<x_1<1, \quad 0<x_2<1, \quad x_3>0 \\ 0 & otherwise \end{cases}$$

(3) Three fair coins are tossed. Let X denote the number of heads on the first two coins, and let Y denote the number of tails on the last two coins. (i) Find the joint pdf of X and Y, (ii) find marginal pdf's of X and Y, (iii) find the conditional pdf f(x|y), and (iv) determine whether X and Y are independent.

Expectations of Functions of Several Random Variables

We have defined the expected value of a function of a single random variable as the weighted average of the values that the function takes on, where each value is weighted by its probability. The expected value of a function of multiple random variables is also defined in an analogous way.

Definition. Let u(X,Y) be a Borel measurable function of two random variables X and Y, which have a joint pdf f(x,y). The expected value of u(X,Y) is defined as

$$E[u(X,Y)] = \begin{cases} \sum_{y_j} \sum_{x_i} u(x_i,y_j) f(x_i,y_j) & \text{for discrete random variables} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x,y) f(x,y)\,dx\,dy & \text{for continuous random variables} \end{cases}$$

and the variance of u(X,Y) is defined as

$$Var[u(X,Y)] = E[u(X,Y) - Eu(X,Y)]^2 = E[u(X,Y)]^2 - [Eu(X,Y)]^2$$

The *covariance* of u(X,Y) with another function v(X,Y) is defined as

$$Cov[u(X,Y), v(X,Y)] = E\{[u(X,Y) - Eu(X,Y)][v(X,Y) - Ev(X,Y)]\} = E[u(X,Y)v(X,Y)] - [Eu(X,Y)][Ev(X,Y)]$$

The mean and variance of X can be considered as a special case of u(X,Y)=X, where the function involves only one random variable. The mean and variance of a single random variable can of course be derived from the marginal distribution function. To show the equivalence of these two approaches consider the expected value of a continuous random variable X:

$$E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y)\,dy \right) dx = \int_{-\infty}^{\infty} x g(x)\,dx = E(X)$$

where we used the fact that the marginal pdf g(x) of X is the integral of the joint pdf with respect to y. When both random variables are included in the function, there are a few special functions whose expectations have special names.

Definition. Let X and Y be random variables with means μ_x and μ_y, and variances σ_x^2 and σ_y^2, respectively.

(a) The *mixed central moment* (or *product moment about the means*) is defined as $E[(X-\mu_x)^r(Y-\mu_y)^s]$ for $r \neq 0$ and $s \neq 0$. If r=s, it is called the *mixed r-th central moment*.

(b) The *covariance* of X and Y, denoted by σ_{xy} or *Cov*(X,Y), is defined as $E[(X-\mu_x)(Y-\mu_y)]$.

(c) The *correlation coefficient* of X and Y, denoted by ρ or ρ_{xy} or Corr(X,Y), is defined as the covariance of standardized random variables

$$\rho_{xy} = E\left[\frac{(X-\mu_x)}{\sigma_x} \cdot \frac{(Y-\mu_y)}{\sigma_y}\right] = \frac{Cov(X,Y)}{\sigma_x \sigma_y}$$

(d) The *joint moment generating function* of X and Y is defined as $m_{xy}(t,s) = E[e^{tX+sY}]$.

$u(X,Y)$	$E[u(X,Y)]$
$X,\ Y$	Means $\mu_x,\ \mu_y$
$(X-\mu_x)^2,\ \ (Y-\mu_y)^2$	Variances $\sigma_x^2,\ \ \sigma_y^2$
$(X-\mu_x)(Y-\mu_y)$	Covariance σ_{xy}
e^{tX+sY}	Joint mgf

The random variables X and Y are said to be *uncorrelated* if $\sigma_{xy}=0$, *positively correlated.* if $\sigma_{xy}>0$, and *negatively correlated* if $\sigma_{xy}<0$. Both covariance and correlation coefficient measure a *linear relationship* between X and Y in the sense that they tend to move in the same direction from their means with high probability if $\sigma_{xy}>0$, and vice versa. Absolute magnitude of the covariance does not mean much because it depends on the measurement units (and hence the variances). On the other hand, the correlation coefficient is a covariance between two standardized random variables X/σ_x and Y/σ_y, and hence does not depend on the measurement unit.

The covariance of X with itself, *Cov*(X,X), is the variance of X, and the correlation coefficient of X with itself, ρ_{xx}, is equal to 1. The perfect linear synchronization of the movements of two random variables is thus characterized by the unitary correlation coefficient, and the two random variables are said to be *perfectly correlated*. It is easy to verify that X is perfectly correlated with any linear function of X such as $Y=a+bX$, but not perfectly correlated with a nonlinear function such as $Y=a+bX^2$.

Properties of the mean, variance, covariance and correlation coefficient of simple linear functions that arise frequently in practice are presented in the theorem below.

Theorem. Let a, b, c and d be some constants. Then,

(a) $E(aX+bY) = aE(X) + bE(Y)$
(b) $Var(aX+bY+c) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y)$
(c) $Cov(X,Y) = E(XY) - E(X)E(Y)$
(d) $Cov(aX+c, bY+d) = ab\, Cov(X,Y)$
(e) $Corr(aX+c, bY+d) = sgn(ab)\, Corr(X,Y)$

Proof. (a) Let g(x) and h(y) be the marginal pdf's of X and Y, respectively. Then,

$$E(aX+bY) = \int_y\int_x (ax+by)f(x,y)dxdy = a\int_y\int_x xf(x,y)dxdy + b\int_y\int_x yf(x,y)dxdy$$
$$= a\int_x x\left(\int_y f(x,y)dy\right)dx + b\int_y y\left(\int_x f(x,y)dx\right)dy = a\int_x xg(x)dx + b\int_y yh(y)dy = aE(X) + bE(Y)$$

(b) Applying the definition of the variance, we can derive

$$Var(aX+bY+c) = E[(aX+bY+c) - E(aX+bY+c)]^2 = E[a(X-\mu_x) + b(Y-\mu_y)]^2$$
$$= a^2E(X-\mu_x)^2 + b^2E(Y-\mu_y)^2 + 2abE(X-\mu_x)(Y-\mu_y) = a^2\,Var(X) + b^2\,Var(Y) + 2ab\,Cov(X,Y)$$

(c) $Cov(X,Y) = E(X-\mu_x)(Y-\mu_y) = E(XY - \mu_xY - \mu_yX + \mu_x\mu_y) = E(XY) - \mu_xEY - \mu_yEX + \mu_x\mu_y = E(XY) - \mu_x\mu_y$

(d) Since $E(aX+c)=a\mu_x+c$ and $E(bY+d)=b\mu_y+d$,

$$Cov(aX+c, bY+d) = E[(aX+c)-(a\mu_x+c)][(bY+d)-(b\mu_y+d)] = E[a(X-\mu_x)b(Y-\mu_y)]$$
$$= abE[(X-\mu_x)(Y-\mu_y)] = ab\,Cov(X,Y)$$

(e) Since $Var(aX+c)=a^2\sigma_x^2$ and $Var(bY+d)=b^2\sigma_y^2$,

$$Corr(aX+c, bY+d) = \frac{Cov(aX+c, bY+d)}{\sqrt{Var(aX+c)}\sqrt{Var(bY+d)}} = \frac{ab\,Cov(X,Y)}{|a||b|\sigma_x\sigma_y} = sgn(ab)\,\rho_{xy}$$

where *sgn*(*ab*) denotes the sign of ab. ▣

There are in general no simple formulae for the mean and variance of a nonlinear function of the random variables. The moments of such functions may not even exist. A nonlinear function that arises often in econometrics is the ratio of random variables. For example, we estimate a linear regression model $y=\alpha+\beta X+\gamma Z+\varepsilon$. Standard computer software gives the estimates of coefficients $(\hat\alpha, \hat\beta, \hat\gamma)$ and their variances and covariances. If we wish to examine the statistical properties of $\hat\beta/\hat\gamma$, a Taylor approximation (***δ method***) of its moments is often used

Theorem. *Mean and variance of product and ratio*. Let μ_x and μ_y be the means of X and Y, σ_x and σ_y be their standard deviations, and σ_{xy} be their covariance.

(a) $E(XY) = \mu_x\mu_y + \sigma_{xy}$, $\quad Var(XY) = E(X^2Y^2) - [E(XY)]^2 = E(X^2Y^2) - (\mu_x\mu_y + \sigma_{xy})^2$

(b) $$E\left(\frac{Y}{X}\right) \approx \frac{\mu_y}{\mu_x} + \frac{\mu_y}{\mu_x^3}\sigma_x^2 - \frac{\sigma_{xy}}{\mu_x^2} = \frac{\mu_y}{\mu_x} + \frac{\sigma_x\mu_y}{\mu_x^2}\left(\frac{\sigma_x}{\mu_x} - \frac{\rho\sigma_y}{\mu_y}\right)$$

$$Var\left(\frac{Y}{X}\right) \approx \frac{\mu_y^2}{\mu_x^4}\sigma_x^2 + \frac{1}{\mu_x^2}\sigma_y^2 - \frac{2\mu_y}{\mu_x^3}\sigma_{xy} = \left(\frac{\mu_y}{\mu_x}\right)^2\left(\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x\mu_y}\right)$$

Proof. (a) E(XY) is obtained by rewriting Cov(X,Y), and Var(XY) is just the formula for the variance.
(b) This is proved by using the Taylor approximation (***δ method***), and the details are in the Appendix. ▣

Theorem. *Cauchy-Schwarz Inequality*. Let X and Y be the random variables with finite second moments. Then,

$$[E(XY)]^2 \le E(X^2)E(Y^2)$$

where the equality holds if and only if P(Y=cX)=1 for some constant c.
Proof. Let $q(t)=E[(tX-Y)^2]$, which is nonnegative because it is the expected value of a nonnegative function of random variables. Expanding the function, one sees that $q(t) = E(X^2)t^2 - 2E(XY)t + E(Y^2)$ is a quadratic function in t. If q(t)>0, then the roots of q(t) is not real, i.e., $[2E(XY)]^2 - 4E(X^2)E(Y^2) < 0$, which gives the result with the strict inequality. If q(t)=0 for some t=c, then $E[(cX-Y)^2] = 0$, which implies P[cX=Y] = 1. ▣

Corollary. The correlation coefficient ρ_{xy} of X and Y satisfies $-1 \le \rho_{xy} \le 1$.
Proof. Let $W=(X-\mu_x)/\sigma_x$ and $Z=(Y-\mu_y)/\sigma_y$ be the standardized random variables, so that Cov(W, Z)=E(WZ)=ρ_{xy} and $E(W^2)=E(Z^2)=1$. The Cauchy-Schwarz inequality for W and Z then indicates $\rho^2_{xy} \le 1$, which gives the desired result. One can prove this in an alternative way. Consider two variances of the sum and the difference of W and Z:

$$var(W+Z) = 1 + 1 + 2\rho_{xy} = 2(1+\rho_{xy})$$
$$var(W-Z) = 1 + 1 - 2\rho_{xy} = 2(1-\rho_{xy})$$

Since the variance is always nonnegative, $|\rho_{xy}| \le 1$ must hold. ▣

Conditional Expectation, Conditional Variance and Conditional Covariance

When the value of X is fixed at x, it ceases to be a random variable and the function u(X,Y) becomes a function of the random variable Y only. The proper weights for its expectation is then the conditional probability of Y given X=x. The expected value of u(X,Y) conditional on X=x is thus defined as the weighted average of the values that the function takes on, where each value is weighted by its conditional probability density f(y|x). Similar interpretations apply to the conditional variance.

Definition. The *conditional expectation* (*conditional mean*) and the *conditional variance* of a Borel measurable function u(X,Y) given X=x are defined as

$$\mu_{u|x} \equiv E[u(X,Y)|X=x] = \int_{-\infty}^{\infty} u(x,y)f(y|x)dy$$
$$\sigma^2_{u|x} \equiv Var[u(X,Y)|X=x] = E\left\{[u(X,Y) - \mu_{u|x}]^2|X=x\right\} = E[u(X,Y)^2|X=x] - \mu^2_{u|x}$$

where f(y|x) is the conditional pdf of Y given X=x. In particular, the *conditional mean* and the *conditional variance* of Y given X=x are

$$\mu_{y|x} \equiv E[Y|X=x] = \int_{-\infty}^{\infty} yf(y|x)\,dy$$

$$\sigma^2_{y|x} \equiv Var(Y|X{=}x) = E\left\{(Y-\mu_{y|x})^2|X{=}x\right\} = E(Y^2|X{=}x) - \mu^2_{y|x}$$

The expected value of the sum of random variables is the sum of expected values of each random variable. This result also holds for the conditional expectation. Furthermore, a function of X alone becomes a fixed parameter for any given x, and can be taken outside of the expectation operator. These are demonstrated in the following theorem.

Theorem. Let u and v be measurable functions of the random variables X and Y with finite moments. Then,

$$E[u(X,Y)+v(X,Y)|X{=}x] = E[u(x,Y)|X{=}x] + E[v(x,Y)|X{=}x]$$

$$E[u(X)v(X,Y)|X{=}x] = u(x)E[v(x,Y)|X{=}x]$$

Proof. The proofs are left to the readers as exercises.

The conditional expectation E[u(X,Y)|X=x] and the conditional variance Var[u(X,Y)|X=x] are, in general, the functions of x. When we wish to write them for all possible values of X, they are written as E[u(X,Y)|X] and Var[u(X,Y)|X] without specifying a specific value for X. Since they are the functions of random variable X, we can define the mean and variance of E[u(X,Y)|X], and the expected value of Var[u(X,Y)|X], with respect to the random variable X. The following theorem shows that the *expected conditional mean* is the *unconditional mean*, and that the *unconditional variance* is the sum of the *expected conditional variance* and the *variance of the conditional mean*. These results play important roles in a wide range of problems.

Theorem. Let Z=u(X,Y) be a real valued function of random variables X and Y with finite second moments. Then,

(a) $E(Z) = E[E(Z|X)]$

(b) $Var(Z) = E[Var(Z|X)] + Var[E(Z|X)]$

Proof. (a) Applying the definition of the expectation and the conditional expectation, we can write

$$E\{E(Z|X)\} = \int \{E(Z|X)\}f(x)dx = \int\left\{\int z\frac{f(x,y)}{f(x)}dy\right\}f(x)dx = \int\int zf(x,y)dydx = \int\int u(x,y)f(x,y)dydx = E[u(X,Y)]$$

(b) Using the relationship $Var(W) = E(W^2) - [E(W)]^2$ for any random variable W, we can write

$$E[Var(Z|X)] = E\{E(Z^2|X) - [E(Z|X)]^2\} = E(Z^2) - E\{[E(Z|X)]^2\}$$

$$Var[E(Z|X)] = E[E(Z|X)]^2 - \{E[E(Z|X)]\}^2 = E[E(Z|X)]^2 - [E(Z)]^2$$

which give

$$E[Var(Z|X)] + Var[E(Z|X)] = E(Z^2) - [E(Z)]^2 = Var(Z) = Var[u(X,Y)]$$

Alternatively, we can show

$$
\begin{aligned}
Var(Z) &= E(Z^2) - [E(Z)]^2 = E[E(Z^2|X)] - \{E[E(Z|X)]\}^2 \\
&= E\{[E(Z^2|X)] - [E(Z|X)]^2\} + E\{[E(Z|X)]^2\} - \{E[E(Z|X)]\}^2 \\
&= E\{Var(Z|X)\} + Var\{E(Z|X)\}
\end{aligned}
$$

Special cases of the theorem that arise frequently in application problems are listed in the following corollary.

Corollary. Let X and Y be the random variables with finite second moments. Then,

(a) $E(Y) = E[E(Y|X)]$
(b) $Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$
(c) $E[u(X)v(Y)] = E\{[u(X)E[v(Y)|X]\}$
(d) $E(XY) = E[XE(Y|X)]$
(e) $Cov(X,Y) = Cov[X, E(Y|X)]$

Proof. (a), (b) and (c) are special cases of the theorem above, and (d) is a special case of (c). To show (e), we use (d) and apply the relationship $Cov(W,Z) = E(WZ) - E(W)E(Z)$ for any random variables W and Z:

$$Cov(X, E(Y|X)) = E[XE(Y|X)] - E(X)E[E(Y|X)] = E(XY) - E(X)E(Y) = Cov(X,Y)$$

Corollary. Let u(X) and v(Y) be real valued functions of X and Y. Then, $u(X)$ and $\{v(Y)-E[v(Y)|X]\}$ are uncorrelated. In particular, X and {Y-E(Y|X)} are uncorrelated.
Proof. For notational simplicity, let $Z = v(Y)-E[v(Y)|X]$. Note that E(Z)=0 by (a) of the theorem above, and part (c) of the corollary above indicates that $E[u(X)Z]=0$. Therefore,

$$Cov\big(u(X), \{v(Y)-E[v(Y)|X]\}\big) = Cov\big(u(X), Z\big) = E[u(X)Z] - E[u(X)]E(Z) = 0$$

Regression Function

Suppose we wish to find a function r(X) that gives the ***best*** possible representation or estimation of variable Y, where the term "best" is used to mean the ***least mean squared error*** principle. That is, we wish to find a function r(X) that minimizes $E[Y-r(X)]^2$. The function r(X) that minimizes $E[Y-r(X)]^2$ is called the *minimum mean squared regression function of Y on X*, or simply, the *regression function of Y on X*. The following theorem provides the basis for finding the regression function Y on X.

Theorem. Let u(X) be a real valued function of a random variable X. Then,

$$E\{[Y - u(X)]^2\} = E[Var(Y|X)] + E\{[E(Y|X) - u(X)]^2\}$$

In particular,

$$E\{[Y - E(Y|X)]^2\} = E[Var(Y|X)]$$

Proof. We apply the result that the unconditional mean is the expected value of the conditional mean:

$$E\{[Y - u(X)]^2\} = E\big(E\{[Y - u(X)]^2 | X\}\big)$$

Now consider the conditional expectation in the parenthesis:

$$\begin{aligned} E\{[Y - u(X)]^2 | X\} &= E\{[Y - E(Y|X) + E(Y|X) - u(X)]^2 | X\} \\ &= E\{[Y - E(Y|X)]^2 | X\} + E\{[E(Y|X) - u(X)]^2 | X\} + 2E\{[Y - E(Y|X)][E(Y|X) - u(X)] | X\} \\ &= Var(Y|X) + [E(Y|X) - u(X)]^2 + 0 \end{aligned}$$

where the last equality follows because the first term is the definition of var(Y|X), and E(Y|X) in the second and third terms is a function of X and plays a constant as far as the conditional expectation is concerned. Taking the expectation on both sides gives the desired result. ◘

Corollary. The minimum of $E[Y-r(X)]^2$, among all possible functions r(X), is obtained by the function r(X)=E(Y|X], and the minimum of $E[Y-r(X)]^2$ is equal to $E[Var(Y|X)]$.
Proof. From the theorem above

$$E\{[Y - r(X)]^2\} = E[Var(Y|X)] + E\{[E(Y|X) - r(X)]^2\}$$

The first term does not involve the function r(X). The second term is nonnegative, because the expected value of a nonnegative random variable is nonnegative. Hence, it takes the smallest value when r(X)=E(Y|X), and the smallest value is the first term $E[Var(Y|X)]$. ◘

Theorem. Let $Z=Y-E(Y|X)$ be the *residual* of the regression of Y on X. Then,

(a) the mean and the variance of the residual are $E(Z)=0$ and $Var(Z) = E[Var(Y|X)] = Var(Y) - Var[E(Y|X)]$,

(b) the residual is not correlated with X, $Cov(X,Z)=0$,

(c) The conditional mean and the conditional variance of the residual are E(Z|X)=0 and Var(Z|X)=Var(Y|X).

Proof.

(a) $E(Z) = E(Y) - E[E(Y|X)] = E(Y) - E(Y) = 0$

$Var(Z) = E[Z - E(Z)]^2 = E(Z^2) = E[Y - E(Y|X)]^2 = E[Var(Y|X)]$

where the last equality follows from the theorem above.

(b) $Cov(X,Z) = E(XZ) - E(X)E(Z) = E(XZ) = E(XY) - E[XE(Y|X)] = E(XY) - E[E(XY|X)] = E(XY) - E(XY) = 0$

(c) $E(Z|X) = E(Y|X) - E[E(Y|X)|X] = E(Y|X) - E(Y|X) = 0$

$Var(Z|X) = E(Z^2|X) - [E(Z|X)]^2 = E(Z^2|X) = E\{[Y - E(Y|X)]^2 | X\} = Var(Y|X)$ ◘

Linear Regression Function

Instead of considering all possible functions r(X) to approximate (estimate, predict) Y, we may wish to consider only a family of linear functions: L(X)=$\alpha+\beta$X. Parameters α and β are called the linear regression coefficients.

Theorem. The minimum mean squared *linear* regression function L(X)=$\alpha+\beta$X of Y on X is given by $\alpha = \mu_y - \beta\mu_x$ and $\beta = \rho\sigma_y/\sigma_x$, where, ρ is the correlation coefficient of X and Y. The minimum mean squared error is equal to $(1-\rho^2)\sigma_y^2$.

Proof. Let $\gamma = \alpha - \mu_y + \beta\mu_x$ for notational simplicity.

$$\begin{aligned} MSE(\alpha,\beta) \equiv E(Y-\alpha-\beta X)^2 &= E[(Y-\mu_y) - \beta(X-\mu_x) - (\alpha-\mu_y+\beta\mu_x)]^2 \\ &= E[(Y-\mu_y)^2] + \beta^2 E[(X-\mu_x)^2] - 2\beta E[(Y-\mu_y)(X-\mu_x)] + \gamma^2 - 2\gamma\,[E(Y-\mu_y) - \beta\,E(X-\mu_x)] \\ &= \sigma_y^2 + \beta^2\sigma_x^2 - 2\beta\sigma_{xy} + (\alpha-\mu_y+\beta\mu_x)^2 \end{aligned}$$

Setting the derivatives of the MSE with respect to α and β to zero, we derive the first order conditions

$$\partial MSE/\partial\alpha = 2(\hat{\alpha} - \mu_y + \hat{\beta}\mu_x) = 0$$

$$\partial MSE/\partial\beta = 2\hat{\beta}\sigma_x^2 - 2\rho\sigma_x\sigma_y + 2\mu_x(\hat{\alpha} - \mu_y + \hat{\beta}\mu_x) = 0$$

whose solutions are

$$\hat{\alpha} = \mu_y - \hat{\beta}\mu_x = \mu_y - \frac{\sigma_{xy}}{\sigma_x^2}\mu_x$$

$$\hat{\beta} = \frac{\rho\sigma_y}{\sigma_x} = \frac{\sigma_{xy}}{\sigma_x^2}$$

Substituting $\hat{\alpha}$ and $\hat{\beta}$,

$$E(Y-\alpha-\beta X)^2 = \sigma_y^2 + \hat{\beta}^2\sigma_x^2 - 2\hat{\beta}\sigma_{xY} = \sigma_y^2 + \rho^2\sigma_y^2 - 2\rho^2\sigma_y^2 = (1-\rho^2)\sigma_y^2$$

$\boxdot$

Theorem. The minimum mean squared error of linear regression function L(X) is greater than the minimum mean squared error of regression function E(Y|X), unless the latter is a linear function.
Proof. From the relationships

$$E[Y-L(X)]^2 = E\,[var(Y|X)] + E\left\{E(Y|X) - L(X)\right\}^2$$
$$E[Y-E(Y|X)]^2 = E\,[var(Y|X)]$$

we have $E[Y-L(X)]^2 - E[Y-E(Y|X)]^2 = E\left\{E(Y|X)-L(X)\right\}^2 \geq 0$. $\boxdot$

Projection of Y on X

Consider now a linear function without constant L(X)=βX. The minimum mean squared regression function L(X)=βX of Y on X is given by $\beta = E(XY)/E(X^2)$. This function is called the *projection of Y on X*:

$$proj(Y|X) = \frac{E(XY)}{E(X^2)}X$$

and we can write

$$Y = proj(Y|X) + \epsilon$$

The residual ϵ is orthogonal to X, i.e., $E(X\epsilon)=0$. This residual is called the *idiosyncratic risk* in finance literature.

Independent Random Variables

When the random variables are independent, the expectations of the functions of the random variables can be simplified as shown below.

Theorem. Let X and Y be independent random variables with means μ_x and μ_y, and standard deviations σ_x and σ_y, and let u(X) and v(Y) be measurable functions.

(a) $E[u(X)v(Y)] = E[u(X)]E[v(Y)]$
(b) $Cov[u(X), v(Y)] = Corr[u(X), v(Y)] = 0$
(c) $Var(XY) = (\mu_x\sigma_y)^2 + (\mu_y\sigma_x)^2 + (\sigma_x\sigma_y)^2$
(d) $E[v(Y)|X] = E[v(Y)]$

Proof. (a) Independence implies that the joint density is the multiplication of marginal pdf's: f(x,y)=g(x)h(y). Hence,

$$E[u(X)v(Y)] = \int_y\int_x u(x)v(y)f(x,y)dxdy = \int_y\int_x u(x)g(x)\, v(y)h(y)dxdy$$

$$= \int_x u(x)g(x)dx \int_y v(y)h(y)dy = E[u(X)]E[v(Y)]$$

(b) Since $E[u(X)v(Y)] = Eu(X)Ev(Y)$ by (a),

$$Cov[u(X), v(Y)] = E[u(X)v(Y)] - E[u(X)]E[v(Y)] = 0$$

(c) Applying the definition of the variance of a random variable

$$Var(XY) = E(X^2Y^2) - [E(XY)]^2 = E(X^2)E(Y^2) - \mu_x^2\mu_y^2 = [\sigma_x^2 + \mu_x^2][\sigma_y^2 + \mu_y^2] - \mu_x^2\mu_y^2$$

which gives the desired result.

(d) Since f(y|x)=f(x,y)/g(x)=h(y),

$$E[v(Y)|X=x] = \int v(y)f(y|x)dy = \int v(y)h(y)dy = E[v(Y)] \quad \boxed{\bullet}$$

It should be noted that, even if X and Y are independent, E[u(X)/v(Y)]$\neq$E[u(X)]/E[v(Y)] in general. The covariance of independent random variables is zero, but the converse does not always hold. That is, Cov(X,Y)=0 does not necessarily imply the stochastic independence of X and Y, as shown in the example below.

Extensions to More than Two Random Variables

When more than two random variables are involved in an analysis, the mean and variance of a function of the random variables can be defined in the same way as the case of two random variables. For example, the mean of a function u(X,Y,Z) of the three continuous random variables with the joint pdf f(x,y,z) is defined to be

$$E[u(X,Y,Z)] = \int\int\int u(x,y,z)f(x,y,z)dxdydz.$$

The covariance between two functions of the random variables such as u(X,Y) and v(W,Z) is defined as before. In particular, it is easy to verify the following relationship for the covariance between two linear functions:

$$Cov(aX+bY, cW+dZ) = ac\,Cov(X,W) + ad\,Cov(X,Z) + bc\,Cov(Y,W) + bd\,Cov(Y,Z)$$

The conditional mean of a function u(X,Y) given other random variables W=w and Z=z is defined to be

$$E[u(X,Y)\,|\,W=w,Z=z] = \int\int u(x,y)f(x,y\,|\,w,z)dxdy = \int\int u(x,y)\frac{f(x,y,w,z)}{h(w,z)}dxdy.$$

where f(x,y,w,z) is the joint pdf of X, Y, W and Z, and h(w,z) is the joint pdf of W and Z. In particular, the *conditional covariance* of X and Y conditional on Z=z is defined to be

$$\sigma_{xy|z} \equiv Cov(X,Y|Z=z) = E[(X-\mu_{x|z})(Y-\mu_{y|z})\,|\,Z=z] = E(XY|Z=z) - \mu_{x|z}\mu_{y|z}$$

where $\mu_{x|z}$ and $\mu_{y|z}$ are the conditional means of X and Y given Z=z. This can be extended to more than three variables in an obvious way. Some of the useful relationships are presented in the following theorems.

Theorem. Let X, Y and Z be random variables.

(a) $E(Y)=E\{E[E(Y|X)|Z]\}$

(b) $E(Y|X) = E[E(Y|X,Z)|X]$

(c) $Cov(X,Y)=Cov[X, E(Y|X)] = Cov\{X, E[E(Y|X,Z)|X]\}$

(d) $Var(Y) = E\{E[Var(Y|X)|Z]\} + E\{Var[E(Y|X)|Z]\} + Var\{E[E(Y|X)|Z]\}$

Proof.

(a) This is the extension of the previous result that the unconditional mean is the expectation of the conditional mean. Since E(Y|X) is a function of X and E[E(Y|X)|Z] is a function of Z, we will write u(X)=E(Y|X) and v(Z)=E[E(Y|X)|Z] for notational simplicity. Then, $E\{E[E(Y|X)|Z]\} = E\{v(Z)\} = E\{E[u(X)|Z]\}$, and $E[u(X)|Z]$ needs to be evaluated with the joint pdf f(x,z) of X and Z. Hence,

$$E\{E[E(Y|X)|Z]\} = E\{v(Z)\} = E\{E[u(X)|Z]\} = \int_z v(Z)f(z)dz = \int_z \left\{\int_x u(x)\frac{f(x,z)}{f(z)}dx\right\}f(z)dz$$

$$= \int_x u(x)\left\{\int_z f(x,z)\,dz\right\}dx = \int_x u(x)f(x)dx = E[E(Y|X)] = E(Y)$$

where the last equality has been shown earlier.

(b) Since $E(Y|X,Z)$ is a function of X and Z, let $u(X,Z) = E(Y|X,Z)$.

$$E[E(Y|X,Z)|X] = E[u(X,Z)|X] = \int_z u(x,z)\frac{f(x,z)}{f(x)}dz = \int_z \left(\int_y y\frac{f(x,y,z)}{f(x,z)}dy\right)\frac{f(x,z)}{f(x)}dz = \int_z\int_y y\frac{f(x,y,z)}{f(x)}dy\,dz$$

$$= \int_y \frac{y}{f(x)}\left(\int_z f(x,y,z)\,dz\right)dy = \int_y y\frac{f(x,y)}{f(x)}dy = E(Y|X)$$

(c) The first equality is proven earlier in Corollary (e) before the section on Regression Function, and the second equality follows from (b).

(d) The first right hand side term is $E\{E[Var(Y|X)|Z]\} = E[Var(Y|X)]$ because the mean of a conditional mean is the unconditional mean. Let $W = E(Y|X)$. Then, the last two terms is the decomposition of Var(W). Therefore, the right hand side becomes $E[Var(Y|X)] + Var[E(Y|X)]$, which is the decomposition of Var(Y).

▣

Theorem. Let X_i, i=1,2,...n, be random variables with mean μ_i, variance $\sigma_{ii} = \sigma_i^2$, and the covariance σ_{ij}. Let $Y = \sum_i a_i X_i$, where a_i's are finite constants. Then,

$$E(Y) = \sum_{i=1}^{n} a_i\mu_i \qquad\qquad Var(Y) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \sigma_{ij}$$

If they are *mutually independent* random variables, then $Var(Y) = \sum_i a_i^2\sigma_i^2$. ▣

Corollary. Let X_i, i=1,2,...,n, be independent random variables with a common mean μ and a common variance σ^2. Let $\bar{X} = \frac{1}{n}\sum_i X_i$. Then, $E(\bar{X}) = \mu$ and $Var(\bar{X}) = \sigma^2/n$. ▣

Example. Consider the example of the discrete random variables in the experiment of tossing two fair coins, where the random variable X is defined as X = 0 if different faces and X= 1 if same faces, and Y is the number of heads shown. The joint pdf f(x,y), the marginal pdf's g(x) and h(y), and the conditional pdf f(y|x) were shown earlier.

$E(X) = 1/2, \quad E(X^2) = 1/2, \quad \sigma_x^2 = 1/4$
$E(Y) = 1, \quad E(Y^2) = 3/2, \quad \sigma_y^2 = 1/2$
$E(XY) = 1/2, \quad E(X^2Y^2) = 1, \quad Var(XY) = 3/4$
$E(X/Y)$ *does not exist*
$Cov(X,Y) = E(XY) - E(X)E(Y) = 0 \quad \Rightarrow \quad \rho_{xy} = 0$

We have shown earlier that X and Y are not independent. But, their covariance is zero. To examine the conditional moments, first consider E(Y|X=x). Given X=0, Y takes values 0, 1 and 2 with probabilities 0, 1 and 0, respectively. Similarly, given X=1, Y takes values 0, 1 and 2 with probabilities 1/2, 0 and 1/2, respectively. Therefore,

$$E(Y|X=0) = 0 \cdot 0 + 1 \cdot 1 + 2 \cdot 0 = 1, \quad E(Y|X=1) = 0 \cdot (1/2) + 1 \cdot 0 + 2 \cdot (1/2) = 1$$

We can also find $E(Y^2|X=x)$ in a similar way. Given X=0, Y^2 takes values 0, 1 and 4 with probabilities 0, 1 and 0, respectively. Similarly, given X=1, Y^2 takes values 0, 1 and 4 with probabilities 1/2, 0 and 1/2, respectively. Therefore,

$$E(Y^2|X=0) = 0 \cdot 0 + 1 \cdot 1 + 4 \cdot 0 = 1, \quad E(Y^2|X=1) = 0 \cdot (1/2) + 1 \cdot 0 + 4 \cdot (1/2) = 2$$

Conditional variance var(Y|X=x) can then be computed from these results:

$$var(Y|X=0) = E(Y^2|X=0) - [E(Y|X=0)]^2 = 1 - 1 = 0$$
$$var(Y|X=1) = E(Y^2|X=1) - [E(Y|X=1)]^2 = 2 - 1 = 1$$

Other moments are computed as follows:

$$E[E(Y|X)] = E(Y|X=0) \cdot p(X=0) + E(Y|X=1) \cdot p(X=1) = 1 \cdot (1/2) + 1 \cdot (1/2) = 1 = E(Y)$$
$$E[Var(Y|X)] = Var(Y|X=0) \cdot p(X=0) + Var(Y|X=1) \cdot p(X=1) = 0 \cdot (1/2) + 1 \cdot (1/2) = 1/2$$
$$Var[E(Y|X)] = E\{[E(Y|X)]^2\} - \{E[E(Y|X)]\}^2 = E\{[E(Y|X)]^2\} - \{E(Y)\}^2 = 1 - 1 = 0$$
$$E[Var(Y|X)] + Var[E(Y|X)] = 1/2 = Var(Y)$$

$$E[XE(Y|X)] = (0 \cdot 1)p(X=0) + (1 \cdot 1)p(X=1) = 1/2 = E(XY)$$
$$Cov[X, E(Y|X)] = E[XE(Y|X)] - E(X)E[E(Y|X)] = 0 = Cov(X.Y)$$

Exercise: Find similar moments and relationships by using the conditional pdf f(x|y). ▣

Example. Random variables X and Y have a joint uniform distribution

$$f(x,y) = \begin{cases} \dfrac{1}{(b-a)(d-c)} & \textit{if} \quad x \in [a,b] \quad \textit{and} \quad y \in [c,d] \\ 0 & \textit{otherwise} \end{cases}$$

The marginal pdf's are U(a,b) for X and U(c,d) for Y. Since f(x,y)=g(x)h(y), X and Y are independent. We have shown earlier that E(X)=(a+b)/2 and Var(X)=$(b-a)^2/12$, and similarly for Y. Independence of X and Y implies

$Cov(X,Y) = 0$

$E(XY) = E(X)E(Y) = \frac{(a+b)(c+d)}{4}$

$var(XY) = \frac{(a+b)^2(d-c)^2}{48} + \frac{(c+d)^2(b-a)^2}{48} + \frac{(b-a)^2(d-c)^2}{144}$

and E(Y|X=x)=E(Y), Var(Y|X=x)=Var(Y), E[XE(Y|X)]=E(X)E(Y), and Cov[X, E(Y|X)]=0. ◙

Example. Let X and Y have a uniform distribution over a unit circle with the center at (0,0):

$$f(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } x^2+y^2 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

The marginal pdf's are

$$g(x) = \int f(x,y)dy = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi}dy = \frac{2\sqrt{1-x^2}}{\pi} \quad \text{for } -1 \le x \le 1$$

$$h(y) = \frac{2\sqrt{1-y^2}}{\pi} \quad \text{for } -1 \le y \le 1$$

And hence, f(x,y)≠g(x)h(y), and X and Y are not independent. However, Cov(X,Y)=0. This can be seen easily from the following facts: The marginal pdf's g(x) and h(y) are symmetric around zero and hence, E(X)=E(Y)=0. Or, this can be verified by

$$E(X) = \int_{-1}^{1} xg(x)dx = \frac{2}{\pi}\int_{-1}^{1} x\sqrt{1-x^2}dx = -\frac{2}{\pi}\frac{1}{3}(1-x^2)^{3/2}\Big|_{-1}^{1} = 0$$

And, similarly, the symmetric uniform joint pdf gives

$$E(XY) = \int\int xyf(x,y)dxdy = \frac{1}{\pi}\int_{-1}^{1}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} xydydx = \frac{1}{\pi}\int_{-1}^{1} x(1-x^2)dx = \frac{1}{\pi}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{-1}^{1} = 0$$

Therefore, Cov(X,Y)=E(XY)-E(X)E(Y)=0. ◙

Example. Let X and Y have the joint density

$$f(x,y) = \begin{cases} 2(2xy - x - y + 1) & \text{if } 0<x<1, \quad 0<y<1 \\ 0 & \text{otherwise} \end{cases}$$

This is a member of the Farlie-Morgenstern family based on two uniform distribution functions with parameter α=1. Hence, the marginal pdf's are uniform: $X \sim U(0,1)$, $Y \sim U(0,1)$, and

$E(X) = E(Y) = 1/2,\ var(X) = var(Y) = 1/12.$

$$E(XY) = \int_0^1\int_0^1 xyf(x,y)dxdy = \frac{5}{18} \quad \Rightarrow \quad cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{36}$$

Since the marginal density of X is g(x)=1 for 0<x<1, the conditional density f(y|x)=f(x,y)/g(x) is same as f(x,y). Therefore,

$$E(Y|X=x) = \int_0^1 yf(y|x)dy = \int_0^1 y(4xy-2x-2y+2)dy = \frac{4}{3}xy^3 - xy^2 - \frac{2}{3}y^3 + y^2\Big|_0^1 = \frac{1}{3} + \frac{1}{3}x$$

which is a linear function of x. To find the conditional variance

$$E(Y^2|X=x) = \int_0^1 y^2f(y|x)dy = \int_0^1 y^2(4xy-2x-2y+2)dy = xy^4 - \frac{2}{3}xy^3 - \frac{2}{4}y^4 + \frac{2}{3}y^3\Big|_0^1 = \frac{1}{6} + \frac{1}{3}x$$

$$Var(Y|X=x) = E(Y^2|X=x) - [E(Y|X=x)]^2 = \frac{1}{18} + \frac{x}{9} - \frac{x^2}{9}, \quad x\in(0,1)$$

which is the largest for x=1/2. To verify the relationship $Var(Y) = E[Var(Y|X)] + Var[E(Y|X)]$,

$$E[Var(Y|X)] = \int_0^1\left(\frac{1}{18} + \frac{x-x^2}{9}\right)g(x)dx = \frac{2}{27}$$

$$Var[E(Y|X)] = Var\left(\frac{1}{3} + \frac{X}{3}\right) = \frac{1}{9}Var(x) = \frac{1}{9}\frac{1}{12} = \frac{1}{108}$$

which give

$$Var(Y) = \frac{1}{12} = E[Var(Y|X)] + Var[E(Y|X)] \quad \boxed{\bullet}$$

Bivariate Normal Distribution

Let X_1 and X_2 be *bivariate normal* random variables with means μ_1 and μ_2, variances σ_{11} and σ_{22}, and covariance σ_{12}. We have shown that the conditional pdf $f(x_1|x_2)$ is a normal pdf with the conditional mean and variance

$$E(X_1|x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_{22}}(x_2 - \mu_2)$$

$$Var(X_1|x_2) = \sigma_{11} - \frac{\sigma_{12}^2}{\sigma_{22}}$$

Theorem. Let X_1 and X_2 be *bivariate normal* random variables with means μ_1 and μ_2, variances σ_{11} and σ_{22}, and covariance σ_{12}. Then, their joint moment generating function is

$$m(t_1, t_2) = E(e^{t_1X_1 + t_2X_2}) = \exp\left\{(t_1\mu_1 + t_2\mu_2) + \tfrac{1}{2}(t_1^2\sigma_{11} + 2t_1t_2\sigma_{12} + t_2^2\sigma_{22})\right\}$$

Proof. We will use the following results that have been proven earlier:

(a) The unconditional mean is the mean of a conditional mean
(b) The mgf of a normal random variable $N(\mu, \sigma^2)$ is $m(t) = E(e^{tX}) = e^{t\mu + t^2\sigma^2/2}$.
(c) The conditional pdf of X_1 given X_2 is normal $N(m, s^2)$, where $m = \mu_1 + \beta(x_2 - \mu_2)$, $\beta = \sigma_{12}/\sigma_{22}$ and $s^2 = \sigma_{11} - \beta\sigma_{12}$.
(d) The marginal pdf of X_i is normal $N(\mu_i, \sigma_{ii})$.

Using (a), we can write the joint mgf as

$$m(t_1, t_2) = E(e^{t_1X_1 + t_2X_2}) = E\left[E(e^{t_1X_1 + t_2X_2}|X_2)\right] = E\left[e^{t_2X_2}E(e^{t_1X_1}|X_2)\right]$$

From (b) and (c),

$$E(e^{t_1X_1}|X_2) = e^{t_1m + t_1^2s^2/2} = e^{t_1[\mu_1 + \beta(X_2 - \mu_2)] + t_1^2s^2/2} = e^{t_1\beta X_2}e^{t_1(\mu_1 - \beta\mu_2) + t_1^2s^2/2}$$

which leads to

$$m(t_1, t_2) = E\left[e^{t_2X_2}E(e^{t_1X_1}|X_2)\right] = E\left[e^{t_2X_2}e^{t_1\beta X_2}e^{t_1(\mu_1 - \beta\mu_2) + t_1^2s^2/2}\right] = \exp\left\{t_1(\mu_1 - \beta\mu_2) + t_1^2s^2/2\right\}E\left[e^{(t_1\beta + t_2)X_2}\right].$$

The last term is the mgf of X_2 with parameter $(t_1\beta+t_2)$ in the place of t. Since the marginal distribution of X_2 is normal $N(\mu_2, \sigma_{22})$, the last term is equal to

$$E\left[e^{(t_1\beta + t_2)X_2}\right] = \exp\left\{(t_1\beta + t_2)\mu_2 + (t_1\beta + t_2)^2\sigma_{22}/2\right\}.$$

Therefore,

$$\begin{aligned} m(t_1, t_2) &= \exp\left\{t_1(\mu_1 - \beta\mu_2) + t_1^2s^2/2\right\} \cdot \exp\left\{(t_1\beta + t_2)\mu_2 + (t_1\beta + t_2)^2\sigma_{22}/2\right\} \\ &= \exp\left\{(t_1\mu_1 + t_2\mu_2) + (t_1^2\sigma_{11}^2 + 2t_1t_2\sigma_{12} + t_2^2\sigma_{22}^2)/2\right\} \end{aligned}$$

▣

Stein's Lemma. Let X_1 and X_2 be *bivariate normal* random variables with means μ_1 and μ_2, variances σ_{11} and σ_{22}, and covariance σ_{12}. Let $u(X_1)$ be a differentiable function satisfying $E|u'(X_1)| < \infty$. Then,

$$cov[u(X_1), X_2] = E[u'(X_1)]\,cov(X_1, X_2)$$

Proof. We use the fact that the unconditional expectation is the expectation of conditional expectation:

$$cov[u(X_1), X_2] = E[u(X_1)X_2] - E[u(X_1)]E(X_2) = E[u(X_1)E(X_2|X_1)] - E[u(X_1)]E(X_2)$$

$$= E[u(X_1)\{\mu_2 + (\sigma_{12}/\sigma_{11})(X_1-\mu_1)\}] - \mu_2 E[u(X_1)]$$

$$= (\sigma_{12}/\sigma_{11})E[u(X_1)(X_1-\mu_1)] = (\sigma_{12}/\sigma_{11})\sigma_{11}E[u'(X_1)] = \sigma_{12}E[u'(X_1)]$$

where we used the result $E[u(X_1)(X_1-\mu_1)] = \sigma_{11}E[u'(X_1)]$ as shown below. $\boxdot$

Stein's Lemma. Let X be a *normal* random variable with mean μ and variance σ^2, and u(X) a *differentiable* function of X satisfying $E|u'(X)| < \infty$. Then,

$$E[u(X)(X-\mu)] = \sigma^2 E[u'(X)]$$

Proof. We apply the integration by parts to

$$E[u(X)(X-\mu)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} u(x)(x-\mu)e^{-(x-\mu)^2/2\sigma^2}dx$$

by setting u=u(x) and $dv = (x-\mu)e^{-(x-\mu)^2/2\sigma^2}dx$. Since $v = -\sigma^2 e^{-(x-\mu)^2/2\sigma^2}$, we can write

$$E[u(X)(X-\mu)] = \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} udv = \frac{1}{\sqrt{2\pi\sigma^2}} uv\Big|_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} vdu$$

$$= 0 + \frac{\sigma^2}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty} u'(x)e^{(x-\mu)^2/2\sigma^2}dx = \sigma^2 E[u'(X)]$$

where the first term is zero by the condition $E|u'(X)| < \infty$. $\boxdot$

Remark: This Lemma is used for the estimation of multivariate normal means. An extension of this lemma to the covariance of jointly normal random variables is also used in derivation of CAPM under normality assumption on risky returns.

Expectation of Truncated Distribution

Let f(x) and F(x) be the pdf and cdf of X. It was shown earlier that a distribution truncated from below by a constant c is a conditional distribution f(x|X>c):

$$f(x|X>c) = \begin{cases} \dfrac{f(x)}{1-F(c)} & if \quad x > c \\ 0 & otherwise \end{cases}$$

The mean and variance of this distribution are computed by the usual formulae

$$E(X|X>c) = \int_c^{\infty} x\frac{f(x)}{1-F(c)}dx, \qquad Var(X|X>c) = E(X^2|X>c) - [E(X|X>c)]^2$$

Consider for an example a normal random variable $X \sim N(\mu, \sigma^2)$. We know that $Z = (X-\mu)/\sigma$ is distributed as a standard normal $N(0, 1)$, and

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2} = \frac{1}{\sigma}\phi(z), \qquad z = \frac{x-\mu}{\sigma}$$

$$F(c) = P(X \le c) = P\left(\frac{X-\mu}{\sigma} \le \frac{c-\mu}{\sigma}\right) = P(Z \le b) = \Phi(b), \quad b = \frac{c-\mu}{\sigma}$$

Therefore,

$$E(X|X{>}c) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{1-F(c)} \int_c^\infty x e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{1-\Phi(b)} \int_b^\infty (\mu + \sigma z) e^{-z^2/2} (\sigma dz)$$

$$= \frac{\mu}{1-\Phi(b)}[1-\Phi(b)] + \frac{\sigma}{1-\Phi(b)} \frac{1}{\sqrt{2\pi}} e^{-b^2/2} = \mu + \sigma \frac{\phi(b)}{1-\Phi(b)} \equiv \mu + \sigma\lambda(b)$$

where λ(b) is the hazard function, which is also called the *inverse Mills ratio*. The conditional variance can be derived by first computing the conditional second moment

$$E(X^2|X{>}c) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{1-F(c)} \int_c^\infty x^2 e^{-(x-\mu)^2/2\sigma^2} dx = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{1-\Phi(b)} \int_b^\infty (\mu + \sigma z)^2 e^{-z^2/2} (\sigma dz)$$

$$= \frac{\mu^2}{1-\Phi(b)} \frac{1}{\sqrt{2\pi}} \int_b^\infty e^{-z^2/2} dz + \frac{2\sigma\mu}{1-\Phi(b)} \frac{1}{\sqrt{2\pi}} \int_b^\infty z e^{-z^2/2} dz + \frac{\sigma^2}{1-\Phi(b)} \frac{1}{\sqrt{2\pi}} \int_b^\infty z^2 e^{-z^2/2} dz$$

$$= \mu^2 + 2\sigma\mu\lambda(b) + b\sigma^2\lambda(b) + \sigma^2$$

where the last integral is obtained by the *integration by parts*, setting $u{=}z,\ dv{=}ze^{-z^2/2}dz$, and $v = -e^{-z^2/2}$:

$$\frac{1}{\sqrt{2\pi}} \int_b^\infty z^2 e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \int_b^\infty u\,dv = \frac{1}{\sqrt{2\pi}} uv \Big]_b^\infty - \frac{1}{\sqrt{2\pi}} \int_b^\infty v\,du$$

$$= b\phi(b) + [1-\Phi(b)]$$

Now the conditional variance is derived from

$$var(X|X{>}c) = E(X^2|X{>}c) - [E(X|X{>}c)]^2 = \left(\mu^2 + 2\sigma\mu\lambda(b) + b\sigma^2\lambda(b) + \sigma^2\right) - [\mu + \sigma\lambda(b)]^2$$
$$= \sigma^2\{1 - \lambda(b)[\lambda(b) - b]\}$$

Note that the conditional mean is greater than the unconditional mean, and the conditional variance is smaller than unconditional variance. Note also that $\{1 - \lambda(b)[\lambda(b) - b]\}$ must be positive for any finite b.

Exercise: Show that, for a normal random variable $X \sim (\mu, \sigma^2)$, and $b = (c-\mu)/\sigma$,

$$E(X|X<c) = \mu - \sigma\frac{\phi(b)}{\Phi(b)} \equiv \mu - \sigma\,\kappa(b)$$

$$var(X|X<c) = \sigma^2\{1 - \kappa(b)[\kappa(b) - b]\}$$

Exercise: Consider a logistic random variable X whose cdf is $F(x)=1/(1+e^{-z})$ and pdf $f(x)=F(x)F(-x)=F(x)[1-F(x)]$. Show

$$E(X|X<c) = \frac{cF(c) + ln(1-F(c))}{F(c)}, \qquad E(X|X>c) = -\frac{cF(c) + ln(1-F(c))}{1-F(c)}$$

Exercise: Suppose X~U(a, b). Find E(X|X>c) and var(X|X>c). ▣

Note: Moments of truncated standard normal distribution

$$f(x|X<b) = \frac{\phi(b)}{\Phi(b)}$$

$$m_r = E(X^r|X<b) = \frac{1}{\sqrt{2\pi}\Phi(b)} \int_{-\infty}^{b} x^r e^{-x^2/2} dx$$

$$= \frac{1}{\sqrt{2\pi}\Phi(b)} \int_{-\infty}^{b} x^{r-1} (x e^{-x^2/2} dx)$$

$$= \frac{1}{\sqrt{2\pi}\Phi(b)} \left\{ - x^{r-1} e^{-x^2/2} \Big|_{-\infty}^{b} + (r-1) \int_{-\infty}^{b} x^{r-2} e^{-x^2/2} dx \right\}$$

$$= -\frac{b^{r-1}\phi(b)}{\Phi(b)} + (r-1) m_{r-2}$$

where $m_0 = 1$

Example. Stabilization Policy
Suppose that the relationship between the GDP Y and a policy instrument G can be described by a linear function

$$Y = a + bG, \quad E(a) = \bar{a}, \quad var(a) = \sigma_a^2, \quad b = 1\ (normalization)$$

where ***a*** represents effects of all other factors. Variations of Y come from the variations of ***a*** and changes in G. If no stabilization policy is employed, the variation of Y, σ_y^2, is equal to σ_a^2. If the government attempts to stabilize Y by changing G,

$$\sigma_y^2 = \sigma_a^2 + \sigma_g^2 + 2\rho\sigma_a\sigma_g$$

where the variance σ_g^2 measures the magnitudes of the changes in G and the correlation coefficient ρ between ***a*** and G measures the timing of the policy changes. If ρ=-1, the policy changes are perfectly timed. If -1<ρ<0, the timing is correct more often than not, and if ρ>0, the timing is incorrect more often. Similarly, $\sigma_g = \sigma_a$ can be interpreted as a full (complete) stabilization policy, $\sigma_g < \sigma_a$ as a partial stabilization policy, and $\sigma_g > \sigma_a$ as an overreacting policy. To make the stabilization policy successful, ρ must be negative.

The ratio of the variances in Y with and without a stabilization policy is

$$\frac{\sigma_y^2}{\sigma_a^2} = 1 + \frac{\sigma_g^2}{\sigma_a^2} + 2\rho\frac{\sigma_g}{\sigma_a}$$

For a full stabilization policy to be successful in reducing the variation of Y (i.e., $\sigma_y^2/\sigma_a^2 < 1$), $\rho < -1/2$ must hold. The optimum stabilization policy that minimizes σ_y^2/σ_a^2 is given by $\sigma_g = -\rho\sigma_a$, at which $\sigma_y^2/\sigma_a^2 = 1 - \rho^2$. Thus, if ρ=-0.5, then the optimum size of stabilization policy reduces the income variation by only 25%.

Example: In macroeconomics we learn the effects of the fiscal (and monetary) policy variable G on the equilibrium GDP Y under certainty. However, the effect G is in general unknown and there are many other factors that make the GDP random. Suppose that the relationship between Y and G can be described by a linear function

$$Y = a + bG, \quad E(a) = \bar{a}, \quad Var(a) = \sigma_a^2, \quad E(b) = \bar{b} > 0, \quad Var(b) = \sigma_b^2, \quad Cov(a,b) = \rho\sigma_a\sigma_b$$

where ***a*** represents effects of all other factors. G is measured as a deviation from the current level so that G=0 means no change in the fiscal policy and the GDP is equal to ***a***. The government wishes to use G to bring the GDP to a target level Y^*.

If there is no uncertainty in the relationship, the optimum level of the policy variable is $G^*=(Y^*-a)/b$. If the GDP at the current level of G is below the target level (i.e., $a < Y^*$), then G^* is positive, which means an increase in G from the current level, and vice versa. The policy prescription under certainty thus requires a change of the policy *in the right direction* by the *full amount* to close the gap between the current Y and the target Y^*.

Under uncertainty, this cannot be accomplished. Instead, the government is assumed to set the level of

G to minimize the expected quadratic disutility function

$$\min_{G}\ E(Y - Y^*)^2$$

How does the randomness of ***a*** and ***b*** affect the policy decision compared to the case of certainty? To analyze this issue, rewrite the expected disutility as

$$E(Y - Y^*)^2 = E[(Y - \bar{Y}) + (\bar{Y} - Y^*)]^2 \ = \ \sigma_Y^2 + (\bar{Y} - Y^*)^2$$

where

$$\bar{Y} \equiv E(Y) = \bar{a} + \bar{b}G$$
$$\sigma_Y^2 \equiv Var(Y) = \sigma_a^2 + \sigma_b^2 G^2 + 2\rho\sigma_a\sigma_b G$$

Differentiation of the expected disutility with respect to G gives the first order condition

$$\frac{\partial E(Y - Y^*)}{\partial G} \ = \ 2\sigma_b^2 G + 2\rho\sigma_a\sigma_b + 2\bar{b}(\bar{a} + \bar{b}G - Y^*) \ = \ 0$$

and hence the optimum policy level

$$G^* \ = \ \frac{\bar{b}(Y^* - \bar{a}) - \rho\sigma_a\sigma_b}{\sigma_b^2 + \bar{b}^2}$$

Substituting G^* into the E(Y) equation, the expected GDP at the optimum policy level is

$$\bar{Y}^* \ = \ \frac{\bar{a}\sigma_b^2 + \bar{b}^2 Y^* - \bar{b}\rho\sigma_a\sigma_b}{\sigma_b^2 + \bar{b}^2}$$

We now consider a few special cases.
Case 1. $\sigma_a \neq 0, \quad \sigma_b = 0$ ... The effect of policy is certain.

$$G^* \ = \ \frac{Y^* - \bar{a}}{\bar{b}} \gtreqless 0 \quad as \quad Y^* \gtreqless \bar{a}$$

$$\bar{Y} \ = \ Y^*$$

Thus, the policy is in the *right direction* and attempts to close the gap by the full amount (*full policy*) on the average.

Case 2. $\sigma_a = 0, \ \ \sigma_b \neq 0$ ... *There is no other uncertainty in the rest of the economy.*

$$G^* = \frac{\bar{b}(Y^* - \bar{a})}{\sigma_b^2 + \bar{b}^2} \gtreqless 0 \quad as \quad Y^* \gtreqless \bar{a}$$

$$\bar{Y} = \frac{\bar{a}\sigma_b^2 + \bar{b}^2 Y^*}{\sigma_b^2 + \bar{b}^2} \lesseqgtr Y^* \quad as \quad Y^* \gtreqless \bar{a}$$

Thus, the policy is in the *right direction*, but attempts to close the gap by less than the full amount (*partial policy*) on the average.

Case 3. $\sigma_a \neq 0,\ \sigma_b \neq 0,\ \rho = 0$ ... *The two sources of uncertainty are uncorrelated.*

$$G^* = \frac{\bar{b}(Y^* - \bar{a})}{\sigma_b^2 + \bar{b}^2} \gtreqless 0 \quad as \quad Y^* \gtreqless \bar{a}$$

$$\bar{Y} = \frac{\bar{a}\sigma_b^2 + \bar{b}^2 Y^*}{\sigma_b^2 + \bar{b}^2} \lesseqgtr Y^* \quad as \quad Y^* \gtreqless \bar{a}$$

Thus, the policy is in the *right direction*, but attempts to close the gap by less than the full amount (*partial policy*) on the average. Note that this case is the same as the Case 2.

Case 4. $\sigma_a \neq 0,\ \sigma_b \neq 0,\ \rho \neq 0$ ... *The two sources of uncertainty are correlated.*

$$G^* = \frac{\bar{b}(Y^* - \bar{a}) - \rho\sigma_a\sigma_b}{\sigma_b^2 + \bar{b}^2} \gtreqless 0 \quad as \quad Y^* - \bar{a} \gtreqless \frac{\rho\sigma_a\sigma_b}{\bar{b}}$$

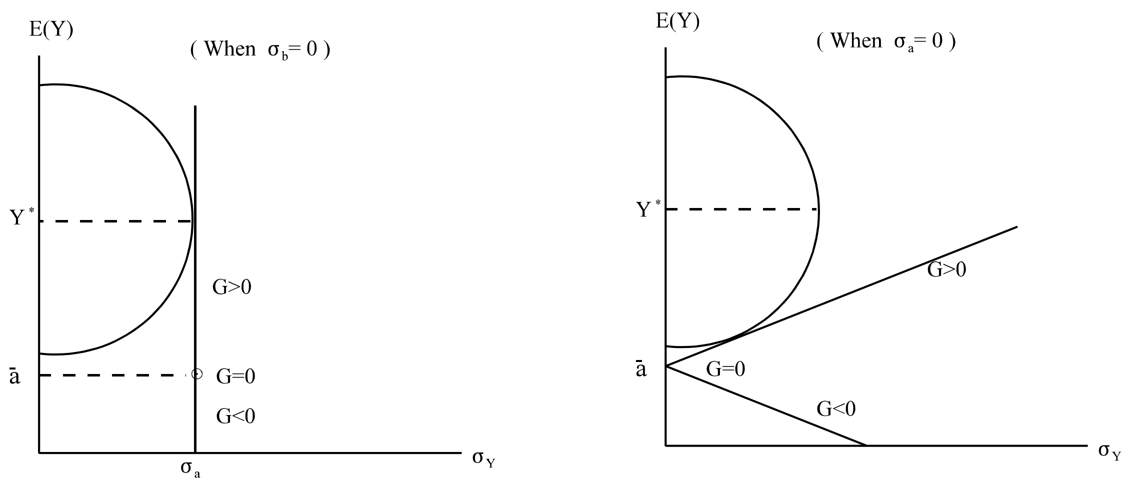
When $\rho>0$, the policy is in the *right direction* if the current expected GDP ($\bar{a}$) is higher than the target GDP (i.e., $Y^* - \bar{a} < 0$). However, if the current expected GDP ($\bar{a}$) is lower than the target GDP ($Y^* - \bar{a} > 0$), then the optimum policy may require a movement in the *wrong direction*. Similarly, when $\rho<0$, the optimum policy may be in the *wrong direction* if the current expected GDP ($\bar{a}$) is higher than the target GDP. ▣

Analysis by using the indifference curve and opportunity locus

An indifference curve represents all combinations of σ_Y and $\bar{Y}$ that give the same expected disutility $E(Y-Y^*)^2$. It is a half circle in a graph with σ_Y on the horizontal axis and $\bar{Y}$ on the vertical axis, and the center of the circle is $(0, Y^*)$. The indifference curve closer to this center gives a higher expected utility.

The opportunity locus can be found by solving the var(Y) equation for G and substituting the result into the E(Y) equation. When $\sigma_b=0$, for example, $\sigma_Y = \sigma_a$ and $\bar{Y} = \bar{a} + \bar{b}G$ does not depend on σ_Y. When $\sigma_a=0$ we have $\sigma_Y = \sigma_b|G|$, which implies

$$\bar{Y} = \begin{cases} \bar{a} + \dfrac{\bar{b}}{\sigma_b}\sigma_Y & \text{if} \quad G > 0 \\ \bar{a} & \text{if} \quad G = 0 \\ \bar{a} - \dfrac{\bar{b}}{\sigma_b}\sigma_Y & \text{if} \quad G < 0 \end{cases}$$



Example: Choice of Monetary Policy -- set M or r? (Poole, William: "Optimal Choice of Monetary Policy Instruments in a Simple Stochastic Macro Model," Quarterly Journal of Economics, 84 (1970), 197-216) The IS-LM model with a linear supply function consists of

$$\begin{aligned} &IS: \quad Y = a_0 - a_1 r + u \\ &LM: \quad M = b_0 + b_1 Y - b_2 r + b_3 P + v \\ &AS: \quad P = c_0 + c_1 Y + w \end{aligned}$$

where Y is real income, r real interest rate, M nominal money stock, P price level, and u, v and w are random disturbance terms. We assume that the random terms u, v and w have zero means and variances σ_u^2, σ_v^2 and σ_w^2, respectively. We also assume that u and v are correlated with correlation coefficient ρ, but w is uncorrelated with u nor with v.

The objective of the policy maker, who wishes to stabilize income around the target income Y^*, is to choose either the money stock or the interest rate that minimizes the expected quadratic disutility function DU=(Y-Y^*):

$$\min_{M \text{ or } r} E(DU) = E(Y - Y^*)^2 = E[(Y - EY) + (EY - Y^*)]^2 = Var(Y) + (EY - Y^*)^2$$

When the money stock is set by the authority, the random aggregate demand function is

$$Y_m^d = \frac{1}{a_1 b_1 + b_2}\left(a_0 b_2 - a_1 b_0 + a_1 M - a_1 b_3 P + b_2 u - a_1 v\right)$$

and the equilibrium income and price level are

$$Y_m^e = \frac{1}{a_1 b_1 + b_2 + a_1 b_3 c_1}\left(a_0 b_2 - a_1 b_0 - a_1 b_3 c_0 + a_1 M + b_2 u - a_1 v - a_1 b_3 w\right)$$
$$P_m^e = c_0 + c_1 Y_m^e + w$$

When the authority sets the interest rate r, the aggregate demand is

$$Y_r^d = a_0 - a_1 r + u$$

and the equilibrium income and price level are

$$Y_r^e = a_0 - a_1 r + u$$
$$P_r^e = c_0 + c_1 Y_r^e + w$$

Now consider the optimum supply of money that minimizes E(DU). The mean and variance of the equilibrium income are

$$E(Y_m^e) = \frac{1}{a_1 b_1 + b_2 + a_1 b_3 c_1}\left(a_0 b_2 - a_1 b_0 - a_1 b_3 c_0 + a_1 M\right)$$
$$Var(Y_m^e) = \frac{1}{(a_1 b_1 + b_2 + a_1 b_3 c_1)^2}\left(b_2^2 \sigma_u^2 + a_1^2 \sigma_v^2 + (a_1 b_3)^2 \sigma_w^2 - 2 a_1 b_2 \rho \sigma_u \sigma_v\right)$$

Since the $Var(Y_m^e)$ does not depend on the choice variable M, the expected disutility is minimized if $E(Y_m^e) = Y^*$, which gives the optimum money supply

$$M^* = \frac{1}{a_1}\left[(a_1 b_1 + b_2 + a_1 b_3 c_1) Y^* - (a_0 b_2 - a_1 b_0 - a_1 b_3 c_0)\right]$$

and the level of E[DU(M*)] is given by

$$E[DU(M^*)] = Var(Y_m^e) = \frac{b_2^2 \sigma_u^2 + a_1^2 \sigma_v^2 - 2 a_1 b_2 \rho \sigma_u \sigma_v + (a_1 b_3)^2 \sigma_w^2}{(a_1 b_1 + b_2 + a_1 b_3 c_1)^2}$$

When the interest rate is used as the policy instrument,

$$E(Y_r^e) = a_0 - a_1 r$$
$$Var(Y_r^e) = \sigma_u^2$$

and hence,

$$r^* = \frac{a_0 - Y^*}{a_1}, \qquad E[DU(r^*)] = Var(Y_r^e) = \sigma_u^2$$

Under certainty ($\sigma_u = \sigma_v = \sigma_w = 0$), DU(M*)=DU(r*)=0 and there is no difference between the two instruments. Under uncertainty, the money stock (M) instrument is better than the interest rate (r) instrument in the expected utility sense if and only if

$$\frac{E[DU(M^*)]}{E[DU(r^*)]} < 1 \quad \Leftrightarrow \quad \frac{(a_1\lambda_{vu} + b_2)^2 - 2a_1b_2\lambda_{vu}(1+\rho) + (a_1b_3)^2\lambda_{wu}^2}{(a_1b_1 + b_2 + a_1b_3c_1)^2} < 1, \qquad \lambda_{vu} = \frac{\sigma_v}{\sigma_u}, \quad \lambda_{wu} = \frac{\sigma_w}{\sigma_u}$$

Consider the Poole's case of a fixed price level: $c_1 = \sigma_w^2 = 0$. In this case, the M-instrument is better if and only if

$$\frac{E[DU(M^*)]}{E[DU(r^*)]} < 1 \quad \Leftrightarrow \quad \frac{(a_1\lambda_{vu} + b_2)^2 - 2a_1b_2\lambda_{vu}(1+\rho)}{(a_1b_1 + b_2)^2} < 1 \quad \Leftrightarrow \quad \frac{b_2^2\sigma_u^2 + a_1^2\sigma_v^2 - 2a_1b_2\rho\sigma_u\sigma_v}{(a_1b_1 + b_2)^2} < \sigma_u^2$$

where the left hand side in the last expression is the variance of aggregate demand given any price level under the M-instrument, and the right hand side is that under the r-instrument.

A *sufficient condition* for the superiority of the M-instrument in Poole's case is $\lambda_{vu} < b_1$. That is, if the monetary sector is sufficiently more stable than the real expenditure (investment in particular) sector, then M-instrument is better than the r-instrument. ⏮

Example. A quantity-setting monopoly firm faces a linear inverse demand function $P = \tilde{a} - (b/2)q$, and has a total cost function $C(q) = \alpha q + (\beta/2)q^2,\ \alpha \geq 0,\ \beta > 0$. The demand shift term $\tilde{\alpha}$ is unknown at the time of production decision, but is known to be distributed as $N(\mu, \sigma_a^2)$. Suppose that, before the production decision, the firm receives a noisy signal Z about the true value of $\tilde{a}$: $Z = \tilde{a} + \epsilon$, where the noise ϵ is a normal variate with mean zero and variance σ_ϵ^2, and $\tilde{a}$ and ϵ are independent. Find the production strategy as a function of signal Z, the strategy that maximizes the conditional expected profit E(π|Z).

Solution: The profit function is $\pi = (\tilde{a} - \alpha)q - \frac{b+\beta}{2}q^2$, and we found earlier the optimum quantity without information signal is $q^* = (\mu - \alpha)/(b + \beta)$. For the current problem we need to find E($\tilde{a}$|Z), for which we use the fact that $\tilde{a}$ and Z are jointly normal.

$$Z \sim N(\mu, \sigma_a^2 + \sigma_\epsilon^2),$$
$$cov(\tilde{a}, Z) = cov(\tilde{a}, \tilde{a} + \epsilon) = cov(\tilde{a}, \tilde{a}) + cov(\tilde{a}, \epsilon) = var(\tilde{a}) = \sigma_a^2$$

This gives

$$E(\tilde{a}|Z) = E(\tilde{a}) + \frac{cov(\tilde{a}, Z)}{var(Z)}\big(Z - E(Z)\big) = \mu + \delta(Z - \mu), \qquad \delta \equiv \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}$$

so that

$$E(\pi | Z) = \big[\mu - \alpha + \delta(Z - \mu)\big]q - \frac{b+\beta}{2}q^2.$$

Differentiation then gives

$$q^* = \frac{\mu - \alpha}{b + \beta} + \frac{\delta}{b + \beta}(Z - \mu).$$

If information simply confirms what was expected ($Z=\mu$), then the firm produces the quantity that would have been produced without information. If information indicates a demand higher than expected ($Z>\mu$), then the firm increases output proportional to the difference between the new information and prior information, and vice versa. The output adjustment is greater, the more precise information is (i.e., δ is greater) and the flatter the marginal revenue and/or the marginal cost curves are. ⏮

Example. Simultaneous Equation Bias
Let the demand and supply functions of a good are

$$\begin{array}{ll} \textit{Demand}: & Q = a - bP + cY + \varepsilon \\ \textit{Inverse Supply}: & P = \alpha + \beta Q + \gamma W + \eta \end{array}$$

where, a, b, c, α, β, and γ are constant parameters, P the price, Y income, and W weather. The conditional distribution of the disturbance terms ε and η, conditional on Y and W, are jointly normal with $E(\varepsilon)=E(\eta)=0$, $\mathrm{var}(\varepsilon)=\sigma_\varepsilon^2$, $\mathrm{var}(\eta)=\sigma_\eta^2$, and $\mathrm{cov}(\varepsilon,\eta)=\sigma_{\varepsilon\eta}$. Find E(Q|P,Y,W) for the demand equation. Under what condition(s), E(Q|P,Y,W)=a-bP+cY?
The equilibrium price is

$$P = \frac{\alpha + a\beta + c\beta Y + \gamma W}{1+b\beta} + v, \qquad v \equiv \frac{\beta\varepsilon + \eta}{1+b\beta}$$

which is distributed (conditional on Y and W) as a normal variate with

$$E(P|Y,W) = \frac{\alpha + a\beta + c\beta Y + \gamma W}{1+b\beta}$$

$$\sigma_P^2 \equiv var(P|Y,W) = var(v|Y,W) = \frac{\beta^2\sigma_\varepsilon^2 + \sigma_\eta^2 + 2\beta\sigma_{\varepsilon\eta}}{(1+b\beta)^2}$$

$$\sigma_{\varepsilon P} \equiv cov(P,\varepsilon\,|\,Y,W) = \frac{\beta\sigma_\varepsilon^2 + \sigma_{\varepsilon\eta}}{1+b\beta}$$

and hence,

$$\begin{aligned} E(Q|P,Y,W) &= a - bP + cY + E(\varepsilon\,|P,Y,W) = a - bP + cY + E(\varepsilon|Y,W) + \frac{\sigma_{\varepsilon P}}{\sigma_P^2}(P - E(P|Y,W)) \\ &= a - bP + cY + \frac{\sigma_{\varepsilon P}}{\sigma_P^2}\left(P - \frac{\alpha + a\beta + c\beta Y + \gamma W}{1+b\beta}\right) \end{aligned}$$

and E(Q|P,Y,W)=a-bP+cY if and only if $\sigma_{\varepsilon P}=0$, that is, if and only if $\beta=0$ *and* $\sigma_{\varepsilon\eta}=0$. When these conditions are satisfied, the simultaneous equation system is called recursive, and the ordinary least square estimator of the demand equation is an unbiased estimator. Otherwise, the ordinary least square estimator is biased, which is called the *simultaneous equation bias*. ⏮

***Example*. *Autoregressive Model*.** Consider a simple linear time series model

$$y_t = \alpha y_{t-1} + u_t, \qquad |\alpha| < 1$$

where u_t is an i.i.d. random disturbance term with zero mean and variance σ_u^2. By repeated substitution, this equation can be written as

$$\begin{aligned} y_t &= \alpha y_{t-1} + u_t = \alpha[\alpha y_{t-2} + u_{t-1}] + u_t = \alpha^2 y_{t-2} + u_t + \alpha u_{t-1} = \alpha^3 y_{t-3} + u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} = \cdots \\ &= \alpha^s y_{t-s} + \sum_{j=0}^{s-1} \alpha^j u_{t-j} \\ &= \sum_{j=0}^{\infty} \alpha^j u_{t-j} \end{aligned}$$

Note that u_t is independent of y_{t-1}, because y_{t-1} is a linear function of u_{t-j}, $j>0$, which are independent of u_t. Using these relationships one can show the moments of y_t as given below.

$$\begin{aligned} E(y_t) &= \sum_{s=0}^{\infty} \alpha^s E[u_{t-s}] = 0 \qquad \because\ E(u_t) = 0 \ \ for\ \ all\ \ t \\ var(y_t) &= var\left(\sum_{s=0}^{\infty} \alpha^s u_{t-s}\right) = \sum_{s=0}^{\infty} \alpha^{2s} var(u_{t-s}) = \sigma_u^2 \sum_{s=0}^{\infty} \alpha^{2s} = \frac{\sigma_u^2}{1-\alpha^2} \qquad \because\ cov(u_t, u_s) = 0 \ \ for\ \ t \neq s \end{aligned}$$

It was shown earlier that $cov(aX + bY, cZ) = ac\,cov(X,Z) + bc\,cov(Y,Z)$ for constants a, b and c. Using this relationship

$$\begin{aligned} cov(y_t, y_{t-1}) &= cov[(\alpha y_{t-1} + u_t), y_{t-1}] = \alpha\ cov(y_{t-1}, y_{t-1}) + cov(u_t, y_{t-1}) = \alpha\, var(y_{t-1}) = \frac{\alpha\sigma_u^2}{1-\alpha^2}, \qquad or \\ &= E(y_t y_{t-1}) - E(y_t)E(y_{t-1}) = \alpha E(y_{t-1}^2) + E(u_t y_{t-1}) = \alpha\, var(y_{t-1}) = \frac{\alpha\sigma_u^2}{1-\alpha^2}, \qquad or \\ &= cov\Big[(u_t + \alpha u_{t-1} + \alpha^2 u_{t-2} + \alpha^3 u_{t-3} + \alpha^4 u_{t-4} + \cdots),\ \ (u_{t-1} + \alpha u_{t-2} + \alpha^2 u_{t-3} + \alpha^3 u_{t-4} + \cdots)\Big] \\ &= \alpha\, var(u_{t-1}) + \alpha^3 var(u_{t-2}) + \alpha^5 var(u_{t-3}) + \cdots = \alpha\sigma_u^2(1 + \alpha^2 + \alpha^4 + \cdots) = \frac{\alpha\sigma_u^2}{1-\alpha^2} \end{aligned}$$

$$\begin{aligned} cov(y_t, y_{t-s}) &= cov\left(\alpha^s y_{t-s} + \sum_{j=0}^{s-1} \alpha^j u_{t-j},\ \ y_{t-s}\right) = \alpha^s var(y_{t-s}) = \frac{\alpha^s\sigma_u^2}{1-\alpha^2} \qquad \because\ y_{t-s}\ \ is\ \ independent\ \ of\ \sum_{j=0}^{s-1} \alpha^j u_{t-j} \\ cov(y_t, y_s) &= \frac{\alpha^{|t-s|}\sigma_u^2}{1-\alpha^2} \\ cov(u_t, y_{t-1}) &= cov\Big(u_t,\ (u_{t-1} + \alpha u_{t-2} + \alpha^2 u_{t-3} + \alpha^3 u_{t-4} + \cdots)\Big) = 0 \end{aligned}$$

The conditional moments are

$$\begin{aligned} E(y_t|y_{t-1}) &= \alpha y_{t-1} + E(u_t|y_{t-1}) = \alpha y_{t-1} + E(u_t) = \alpha y_{t-1} \qquad \because\ u_t\ \ is\ \ independent\ \ of\ \ y_{t-1} \\ var(y_t|y_{t-1}) &= var(u_t|y_{t-1}) = var(u_t) = \sigma_u^2 \end{aligned}$$

Note that, since u_t is independent of all y_{t-s}, $s \geq 1$, we have a more general results

$$E(y_t|y_{t-1}, y_{t-2}, \cdots) = \alpha y_{t-1}, \qquad var(y_t|y_{t-1}, y_{t-2}, \cdots) = \sigma_u^2 \quad \boxed{\bullet}$$

Example*. *Autoregressive model with autocorrelated disturbance term. Suppose that the disturbance term u_t in

$$y_t = \alpha y_{t-1} + u_t, \qquad |\alpha| < 1$$

is not i.i.d., but has the first order autocorrelation

$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad |\rho| < 1, \quad \varepsilon_t \ is \ i.i.d.(0, \sigma_\varepsilon^2)$$

From the previous example, we know that

$$E(u_t) = 0, \quad \sigma_u^2 = var(u_t) = \frac{\sigma_\varepsilon^2}{1 - \rho^2}, \quad cov(u_t, u_{t-s}) = \rho^s \sigma_u^2, \quad cov(u_t, u_s) = \rho^{|t-s|}\sigma_u^2$$

Therefore,

$$\begin{aligned}
cov(u_t, y_{t-1}) &= cov(u_t, \sum_{s=0}^{\infty} \alpha^s u_{t-1-s}) = cov\Big(u_t,\ (u_{t-1}+\alpha u_{t-2}+\alpha^2 u_{t-3}+\alpha^3 u_{t-4}+\cdots)\Big) \\
&= cov(u_t, u_{t-1}) + \alpha\, cov(u_t, u_{t-2}) + \alpha^2 cov(u_t, u_{t-3}) + \alpha^3 cov(u_t, u_{t-4}) + \cdots \\
&= \rho\sigma_u^2 + \alpha\rho^2\sigma_u^2 + \alpha^2\rho^3\sigma_u^2 + \alpha^3\rho^4\sigma_u^2 + \cdots = \rho\sigma_u^2(1 + \alpha\rho + \alpha^2\rho^2 + \alpha^3\rho^3 + \cdots) = \frac{\rho\sigma_u^2}{1 - \alpha\rho}
\end{aligned}$$

$$cov(u_t, y_t) = cov(u_t, \alpha y_{t-1} + u_t) = \alpha\, cov(u_t, y_{t-1}) + cov(u_t, u_t) = \frac{\alpha\rho\sigma_u^2}{1 - \alpha\rho} + \sigma_u^2 = \frac{\sigma_u^2}{1 - \alpha\rho}$$

Since $cov(u_t, u_s)$ depends only on |t-s|,

$$var(y_t) = var\left(\sum_{s=0}^{\infty} \alpha^s u_{t-s}\right) = var(y_{t-j}) = var\left(\sum_{s=0}^{\infty} \alpha^s u_{t-j-s}\right) \quad for \ all \ t \ and \ j$$

Using this result we can derive

$$var(y_t) = var(\alpha y_{t-1} + u_t) = \alpha^2 var(y_{t-1}) + 2\alpha\, cov(u_t, y_{t-1}) + var(u_t) = \alpha^2 var(y_t) + 2\alpha \frac{\rho\sigma_u^2}{1 - \alpha\rho} + \sigma_u^2$$

$$\Rightarrow \qquad var(y_t) = \frac{(1+\alpha\rho)\sigma_u^2}{(1-\alpha^2)(1-\alpha\rho)}$$

Now consider the conditional moments.

$$\begin{aligned}
&E(y_t|y_{t-1}) = \alpha y_{t-1} + E(u_t|y_{t-1}) \neq \alpha y_{t-1} \quad \because \ E(u_t|y_{t-1}) \neq 0 \\
&var(y_t|y_{t-1}) = var(u_t|y_{t-1}) \neq var(u_t)
\end{aligned}$$

In the case of normal variates,

$$E(u_t|y_{t-1}) = E(u_t) + \frac{cov(u_t,y_{t-1})}{var(y_{t-1})}[y_{t-1} - E(y_{t-1})] = \frac{\rho(1-\alpha^2)}{1+\alpha\rho}y_{t-1} \quad \Rightarrow \quad E(y_t|y_{t-1}) = \frac{\alpha+\rho}{1+\alpha\rho}y_{t-1}$$

$$var(y_t|y_{t-1}) = var(u_t|y_{t-1}) = var(u_t) - \frac{[cov(u_t,y_{t-1})]^2}{var(y_{t-1})} = \frac{(1-\rho^2)\sigma_u^2}{1-\alpha^2\rho^2}$$

Example. ***Autoregressive Conditional Heteroscedasticity (ARCH)***
Consider an infinite sequence of random variables $\{u_t, \varepsilon_t\}$, $t\in(-\infty, \infty)$, where ε_t's are i.i.d. random variables with zero mean and unit standard deviation ($E(\varepsilon_t) = 0,\ E(\varepsilon_t^2)=1,$ and $E(\varepsilon_t\varepsilon_s) = 0$ for $t\neq s$), and

$$u_t = \sigma_t\varepsilon_t, \quad \sigma_t^2 = \alpha + \gamma u_{t-1}^2, \quad 0<\alpha, \quad 0<\gamma<1$$

This model is called the ARCH(1) model. We are interested in the conditional and unconditional moments of u_t.

$$E(u_t|u_{t-1}) = \sigma_t E(\varepsilon_t|u_{t-1}) = 0$$
$$var(u_t|u_{t-1}) = \sigma_t^2 var(\varepsilon_t|u_{t-1}) = \sigma_t^2 = \alpha + \gamma u_{t-1}^2$$
$$cov[(u_t, u_{t-1})|u_{t-1}] = 0$$

Note that the conditional variance changes over time as the realized value of u_{t-1} varies. The unconditional moments are

$$E(u_t) = E_{t-1}[E(u_t|u_{t-1})] = 0$$
$$var(u_t) = E[var(u_t|u_{t-1})] + var[E(u_t|u_{t-1})] = E[\alpha + \gamma u_{t-1}^2] + 0 = \alpha + \gamma E(u_{t-1}^2) = \alpha + \gamma var(u_{t-1})$$
$$= \alpha + \gamma[\alpha + \gamma var(u_{t-2}) = \alpha(1+\gamma) + \gamma^2 var(u_{t-3}) = \cdots = \alpha(1+\gamma+\gamma^2+\gamma^3+\cdots) + \gamma^\infty var(u_{t-\infty}) = \frac{\alpha}{1-\gamma}$$
$$cov(u_t, u_{t-s}) = E(u_t u_{t-s}) - E(u_t)E(u_{t-s}) = E(\sigma_t\varepsilon_t u_{t-s}) - 0 = 0 \quad \because\ \varepsilon_t \textit{ is independent of } u_{t-s} \textit{ and } \sigma_t$$

Example. ***Rational Expectations Equilibrium***
Consider a simple supply-demand model with expectation

$$\textit{Demand}: \quad q_t = \alpha - \beta p_t$$
$$\textit{Supply}: \quad q_t = \delta + \gamma p_t^e + u_t$$

where p_t^e is the farmers' expected price and u_t is the random factor. Because of the production lag farmers can not respond their supply to the current price p_t. They have to make their production decisions based on the price they expect to prevail at the time of harvest. The issue is how the farmers formulate their expectations. A naive expectation is $p_t^e = p_{t-1}$, i.e., expects the last period price will prevail in the current period. This expectation leads to a cobweb model. An adaptive expectation model postulates that

$$p_t^e = p_{t-1}^e + \lambda(p_{t-1} - p_{t-1}^e) = \lambda p_{t-1} + (1-\lambda)p_{t-1}^e, \quad 0<\lambda<1$$
$$= \lambda\sum_{s=0}^{\infty}(1-\lambda)^s p_{t-1-s}$$

which is a weighted average of all past prices and the weights decline geometrically.

The rational expectations hypothesis postulates that p_t^e is the best predictor based on all available information, including the knowledge of the market structure:

$$p_t^e = E(p_t|\Omega_{t-1}) = E_{t-1}(p_t)$$

where Ω_{t-1} denotes the set of all information, such as the past realized prices, other private and public information about the random factor u_t, and the structure of the market, at the time of decision making. To find the rational expectation of the equilibrium price we first solve the model for the equilibrium price and then compute the conditional mean:

$$p_t = \frac{\alpha - \delta - \gamma p_t^e - u_t}{\beta} \quad \Rightarrow \quad p_t^e = E(p_t|\Omega_{t-1}) = \frac{\alpha - \delta - \gamma p_t^e}{\beta} - \frac{E(u_t|\Omega_{t-1})}{\beta} \quad \Rightarrow \quad p_t^e = \frac{\alpha - \delta}{\beta + \gamma} - \frac{E(u_t|\Omega_{t-1})}{\beta + \gamma}$$

Suppose that the information set $\Omega_{\tau-1}$ has just the past prices: $\Omega_{t-1} = \{p_{t-1}, p_{t-2}, \cdots\}$. Consider two cases.

(i) u_t is an i.i.d. random variable with mean zero and a finite variance σ^2. In this case, $E(u_t|\Omega_{t-1}) = 0$, and the rational expectation equilibrium price is the mean price $(\alpha-\delta)/(\beta+\gamma)$.

(ii) $u_t = \rho u_{t-1} + \epsilon_t, \quad \epsilon_t$ *is an i.i.d.*$(0, \sigma_\varepsilon^2)$

From the equilibrium price, we have $u_{t-1} = \alpha - \delta - \gamma p_{t-1}^e - \beta p_{t-1}$. Substituting this into the equilibrium price equation and using $E(\varepsilon_t|\Omega_{t-1}) = 0$, we can derive

$$\beta p_t = \alpha - \delta - \gamma p_t^e - \rho(\alpha - \delta - \gamma p_{t-1}^e - \beta p_{t-1}) - \varepsilon_t = (1 - \rho)(\alpha - \delta) - \gamma p_t^e + \rho\gamma p_{t-1}^e + \rho\beta p_{t-1} - \varepsilon_t$$

$$\Rightarrow \quad \beta p_t^e = \beta E(p_t|\Omega_{t-1}) = (1 - \rho)(\alpha - \delta) - \gamma p_t^e + \rho\gamma p_{t-1}^e + \rho\beta p_{t-1}$$

$$\Rightarrow \quad p_t^e = \frac{(1 - \rho)(\alpha - \delta)}{\beta + \gamma} + \frac{\rho\gamma}{\beta + \gamma} p_{t-1}^e + \frac{\rho\beta}{\beta + \gamma} p_{t-1} = \theta_0 + \theta_1 p_{t-1}^e + \theta_2 p_{t-1} \quad \Rightarrow \quad (1 - \theta_1 L)p_t^e = \theta_0 + \theta_2 p_{t-1}$$

$$\Rightarrow \quad p_t^e = \frac{\theta_0}{1 - \theta_1} + \theta_2 \sum_{s=0}^{\infty} \theta_1^s p_{t-1-s} = \frac{(1 - \rho)(\alpha - \delta)}{\beta + \gamma - \rho\gamma} + \frac{\rho\beta}{\beta + \gamma} \sum_{s=0}^{\infty} \left(\frac{\rho\gamma}{\beta + \gamma}\right)^s p_{t-1-s} = \frac{(1 - \rho)(\alpha - \delta)}{\beta + \gamma - \rho\gamma} + \frac{\beta}{\gamma} \sum_{s=1}^{\infty} \left(\frac{\rho\gamma}{\beta + \gamma}\right)^s p_{t-s}$$

Thus, the rational expectation of the equilibrium price is a linear function of all past prices in this case. Note that this is similar to the result of the adaptive expectation model. An important difference is that the geometric weights in the rational expectation are functions of the model parameters, while those in the adaptive expectations model are assumed to be exogenous parameters. Since the parameters of the model reflect decision makers' tastes, technology, etc., any change of them will also affect the decision makers' expectation formulation. This is not the case in the adaptive expectations model. ▣

Example. ***Effectiveness of monetary policy***
Consider a simple macro model

Aggregate Supply: $\quad y_t = \alpha(p_t - p_t^e) + \varepsilon_t$
Aggregate Demand: $\quad y_t = \theta(m_t - p_t) + \eta_t$

where y_t, m_t and p_t are log of real GNP, money supply and price level, respectively. The random shocks ε_t and η_t are assumed to be i.i.d. random variables with zero means and finite variances, and they are mutually independent. p_t^e is the public's expectation of the log of the price level at time t, formed at time t-1. The equilibrium y_t and p_t are

$$y_t = \frac{\alpha\theta}{\alpha+\theta}(m_t - p_t^e) + \frac{\alpha}{\alpha+\theta}\varepsilon_t + \frac{\theta}{\alpha+\theta}\eta_t$$
$$p_t = \frac{\theta}{\alpha+\theta}m_t + \frac{\alpha}{\alpha+\theta}p_t^e + \frac{1}{\alpha+\theta}(\varepsilon_t - \eta_t)$$

The monetary authority chooses a money supply to

$$\min_{m_t} E[(y_t - y^*)^2|\Omega_{t-1}] = var(y_t|\Omega_{t-1}) + [E(y_t|\Omega_{t-1}) - y^*]^2$$

where Ω_{t-1} is the information set available at the decision time and y^* is the target level of income. Consider two cases of expectations of price: (i) Naive expectations, $p_t^e = p_{t-1}$, and (ii) rational expectations, $p_t^e = E(p_t|\Omega_{t-1})$. In the case of naive expectations

$$p_t^e = p_{t-1} \quad\Rightarrow\quad E(y_t|\Omega_{t-1}) = \frac{\alpha\theta}{\alpha+\theta}(m_t - p_{t-1}), \qquad var(y_t|\Omega_{t-1}) = \frac{\alpha^2}{(\alpha+\theta)^2}\sigma_\varepsilon^2 + \frac{\theta^2}{(\alpha+\theta)^2}\sigma_\eta^2$$

hence, the optimum money supply is determined by

$$E(y_t|\Omega_{t-1}) = y^* \quad\Rightarrow\quad m_t^* = p_{t-1} + \frac{\alpha+\theta}{\alpha\theta}y^*, \qquad y_t = y^* + \frac{\alpha}{\alpha+\theta}\varepsilon_t + \frac{\theta}{\alpha+\theta}\eta_t$$

The equilibrium income under the optimum money supply fluctuates around the target income level y^*. The rational expectation is found from the equilibrium price equation

$$p_t^e = E(p_t|\Omega_{t-1}) = \frac{\theta}{\alpha+\theta}E(m_t|\Omega_{t-1}) + \frac{\alpha}{\alpha+\theta}p_t^e \quad\Rightarrow\quad p_t^e = E(m_t|\Omega_{t-1})$$

Therefore,

$$y_t = \frac{\alpha\theta}{\alpha+\theta}[m_t - E(m_t|\Omega_{t-1})] + \frac{\alpha}{\alpha+\theta}\varepsilon_t + \frac{\theta}{\alpha+\theta}\eta_t \quad\Rightarrow\quad E(y_t|\Omega_{t-1}) = \frac{\alpha\theta}{\alpha+\theta}[E(m_t|\Omega_{t-1}) - E(m_t|\Omega_{t-1})] = 0$$

There is no money supply to make the mean income to be the target income y^*. The monetary policy becomes ineffective any money supply will have the same outcome. ◙

Examples on regression function

Example. ***Autoregressive Model***. Consider a simple linear time series model

$$y_t = \alpha y_{t-1} + u_t, \qquad |\alpha| < 1$$

where u_t is an i.i.d. random disturbance term with zero mean and variance σ_u^2. We wish to find the *minimum mean squared error forecast* of y_t given past observations y_{t-s}, $s \geq 1$. It was shown earlier that $E(y_t|y_{t-1}, y_{t-2}, \cdots) = \alpha y_{t-1}$.

Example. ***Signal Extraction***. Let y_t be a random return of a stock, which is not observable ex ante. An expert's prediction, denoted by x_t, is not perfect and has a linear relationship $x_t = y_t + u_t$, where u_t is a random variable. We wish to estimate y_t based upon the observation on x_t. The minimum mean squared error signal extraction $r(x_t)$ is the conditional mean $E(y_t|x_t)$.

Example. ***Signal Extraction***. Duopolistic firms agreed to restrict their output levels to q_1 and q_2 to keep the market price high. The inverse market demand function is random: $P=\varepsilon-\alpha(q_1+q_2)$, where ε is a random variable and α is a known constant. The market price turned out to be unusually low even if firm 1 kept the agreement. It became suspicious of its rival breaking the agreement. It needs an estimate of q_2 given the observed price and its output level q_1.

Example. Consumption-based asset pricing model

A representative agent in the consumption-based asset pricing model receives income stream y_t at each time t, and allocates y_t between consumption c_t and savings (asset) s_t. The price of the asset is p_t and its gross payoff in the next period is x_{t+1}. The agent allocates the income to maximize the additive expected utility

$$EU = u(c_t) + \beta E_t[u(c_{t+1})].$$

subject to the budget constraint

$$c_t = y_t - p_t s_t, \qquad c_{t+1} = y_{t+1} + s_t x_{t+1}$$

where β is the time preference parameter (discount factor) and the subscript for the expectation operator indicates that the expectation is conditional expectation given all information available at time t. Quantities at time t+1 (y_{t+1}, x_{t+1}, c_{t+1}) are unknown at time t and they are random variables. The first order condition gives

$$p_t = E_t\left(\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1}\right) \equiv E_t(m_t x_{t+1}), \qquad m_t = \beta \frac{u'(c_{t+1})}{u'(c_t)}$$

The term m_t is called the stochastic discount factor. Instead of the price level, we may write this relationship in terms of the gross rate of return by dividing both sides with price:

$$1 = E_t\left(m_t \frac{x_{t+1}}{p_t}\right) = E_t(m_t R_{t+1})$$

Risk free rate R^f

A risk free gross rate has the current price p_t=1 and payoff $x_{t+1}=R^f$ with certainty. Hence,

$$p_t = E_t(m_t x_{t+1}) \quad \Rightarrow \quad 1 = E_t(m_t) R^f \quad \Rightarrow \quad R^f = \frac{1}{E_t(m_t)}$$

To evaluate $E(m_t)$ further, suppose that the utility function is a CRRA function

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}, \qquad u'(c) = c^{-\gamma}$$

For this utility function, the stochastic discount factor becomes

$$m_t = \beta\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma}$$

Suppose that the consumption growth (c_{t+1}/c_t) is distributed as a lognormal, i.e.,

$$z_t \equiv \ln\left(c_{t+1}/c_t\right) \sim N(\mu_t, \ \sigma_t^2)$$

Then, m_t is distributed as a lognormal

$$\ln(m_t) = \ln\beta - \gamma z_t \quad \sim N(\ln\beta - \gamma\mu_t, \quad \gamma^2\sigma_t^2)$$

Therefore,

$$E_t(m_t) = e^{\ln\beta - \gamma\mu_t + \gamma^2\sigma_t^2/2} = \beta\, e^{-\gamma\mu_t + \gamma^2\sigma_t^2/2}$$

and

$$R^f = \frac{1}{E_t(m_t)} = \frac{1}{\beta\, e^{-\gamma\mu_t + \gamma^2\sigma_t^2/2}} \qquad or \qquad \ln R^f = -\ln\beta + \gamma\mu_t - \gamma^2\sigma_t^2/2$$

Risk Correction and Beta Representation

The asset pricing equation can be rewritten as

$$p_t = E(m_t x_{t+1}) = cov(m_t, x_{t+1}) + E(m_t)E(x_{t+1}) = \frac{E(x_{t+1})}{R^f} + cov(m_t, x_{t+1})$$

where the subscript for expectation operator is dropped for notational simplicity. The first term represents the standard expected present value of payoffs, and the second term represents the risk adjustment. If the payoff is positively (negatively) correlated with the stochastic discount factor, the price of the asset will be higher (lower) than the standard expected present value of the payoff.

In terms of the rate of return this can be rewritten as

$$1 = E(m_t R^i_{t+1}) = cov(m_t, R^i_{t+1}) + E(m_t)E(R^i_{t+1}) = \frac{E(R^i_{t+1})}{R^f} + cov(m_t, R^i_{t+1})$$

or

$$E(R^i) = R^f - R^f cov(m, R^i) = R^f - \frac{cov(m, R^i)}{E(m)}$$

where the time subscript is omitted for notational simplicity and we used the relationship $R^f = 1/E(m)$ in the last equality. This relationship is also written as

$$E(R^i) = R^f + \frac{cov(m, R^i)}{var(m)}\left(-\frac{var(m)}{E(m)}\right) = R^f + \beta_{i,m}\lambda(m)$$

This is called the beta pricing model. $\beta_{i,m}$ is a linear regression coefficient of R^i on m, and it is called quantity of risk of asset i. λ(m) is common to all assets and it is called the price of risk.

Mean-Variance Frontier

The expression for the expected rate of return can also be written as

$$E(R^i) = R^f - \frac{cov(m, R^i)}{E(m)} = R^f - \rho_{i,m}\frac{\sigma_m}{E(m)}\sigma_i$$

where $\rho_{i,m}$ is the correlation coefficient between m and R^i, and σ_i and σ_m are the standard deviations of R^i and m, respectively. The expected return is a linear function of σ_i, given the correlation coefficient and the mean and variance of stochastic discount factor m. Since $|\rho_{i,m}| \le 1$, we have

$$\left|E(R^i) - R^f\right| \le \frac{\sigma_m}{E(m)}\sigma_i$$

The lines for $\rho_{i,m}=\pm1$ are called the mean-variance frontier. Means and standard deviations of all assets must be within the boundaries. All assets on the frontier are perfectly correlated with the stochastic discount factor m, and also perfectly correlated with each other.

Sharpe Ratio

The Sharpe ratio is defined as the standardized difference between the expected returns of risky and risk free assets

$$SR = \frac{E(R^i) - R^f}{\sigma_i}$$

The Sharpe ratio of all assets on the mean-variance frontier satisfies

$$|SR| = \frac{\sigma_m}{E(m)} = \sigma_m R^f$$

***Miscellaneous Topics*: Ellipse of Concentration, Concentration Ellipsoid**
This is an alternative measure of the dispersion of random variables around their means.

Case 1. Single random variable X with mean μ and SD σ. Suppose we wish to find another r.v. Y which is uniformly distributed over an interval (a,b) and has the same mean and SD as X. This interval (a,b) is interpreted as a geometrical representation of the concentration of the distribution of X about its center μ.

$E(Y)=(a+b)/2 = \mu$
$var(Y)=(b-a)^2/12 = \sigma^2$.

Solve for a and b: $a = \mu - \sigma\sqrt{3}$, $\quad b = \mu + \sigma\sqrt{3}$

Case 2. Two random variables X_i, i=1,2, with means μ_i, variance-covariance σ_{ij}, i=j=1,2. Let $X=(X_1,X_2)'$, $\mu=(\mu_1,\mu_2)'$ and $\Sigma=(\sigma_{ij})$. We wish to find a region over which another random vector $Y=(Y_1,Y_2)'$ has a uniform distribution with the same mean vector and variance-covariance matrix. The concentration ellipse is such an area. Some other form of area is also acceptable, but we restrict our search to an ellipse. (This ellipse also appears in the confidence region of regression estimators.)

For simplicity, assume zero means μ=0. And consider a general form of an ellipse in Y_i:

$h(y_1,y_2) = (y_1,y_2)(a_{ij})(y_1,y_2)' = c^2$, $a_{12}=a_{21}$.

If means are not zeros, replace y_i with y_i-μ_i. We wish to find a_{ij} and c such that Y vector has the same mean and var-cov matrix as X.

The area enclosed by this ellipse is $\pi c^2/|A|^{1/2}$, where $|A|=a_{11}a_{22}-a_{12}^2$. Thus, the pdf of Y is given by $f(y_1,y_2)=|A|^{1/2}/\pi c^2$ if $h(y_1,y_2)\leq c^2$, and zero otherwise. The concentration ellipse is given by in a general form $(y-\mu)'\Sigma^{-1}(y-\mu)=n+2$, where n is the number of element in vector y.

Appendix A: Proof of the mean and variance of the ratio Y/X

$$E\left(\frac{Y}{X}\right) \approx \frac{\mu_y}{\mu_x} + \frac{\mu_y}{\mu_x^3}\sigma_x^2 - \frac{\sigma_{xy}}{\mu_x^2} = \;= \frac{\mu_y}{\mu_x} + \frac{\sigma_x\mu_y}{\mu_x^2}\left(\frac{\sigma_x}{\mu_x} - \frac{\rho\sigma_y}{\mu_y}\right)$$

$$var\left(\frac{Y}{X}\right) \approx \frac{\mu_y^2}{\mu_x^4}\sigma_x^2 + \frac{1}{\mu_x^2}\sigma_y^2 - \frac{2\mu_y}{\mu_x^3}\sigma_{xy} = \left(\frac{\mu_y}{\mu_x}\right)^2\left(\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x\mu_y}\right)$$

Proof. This is based on the Taylor expansion of g(x,y) around a point (a,b):

$$g(x,y) = g(a,b) + g_x(a,b)(x-a) + g_y(a,b)(y-b) + \frac{1}{2!}\left(g_{xx}(a,b)(x-a)^2 + 2g_{xy}(a,b)(x-a)(y-b) + g_{yy}(a,b)(y-b)^2\right) + \;....$$

Applying this, the second order Taylor expansion of g(X,Y)=Y/X around the mean point ($\mu_{x,}$ μ_y) becomes

$$g_x = -\frac{Y}{X^2},\quad g_{xx} = \frac{2Y}{X^3},\quad g_y = \frac{1}{X},\quad g_{yy} = 0,\quad g_{xy} = -\frac{1}{X^2}$$

$$\frac{Y}{X} \approx \frac{\mu_y}{\mu_x} - \frac{\mu_y}{\mu_x^2}(X-\mu_x) + \frac{1}{\mu_x}(Y-\mu_y) + \frac{1}{2!}\left(\frac{2\mu_y}{\mu_x^3}(X-\mu_x)^2 - \frac{2}{\mu_x^2}(X-\mu_x)(Y-\mu_y) + 0\right)$$

For the E(Y/X), we use the second order approximation

$$E\left(\frac{Y}{X}\right) \approx \frac{\mu_y}{\mu_x} - 0 + 0 + \frac{1}{2!}\left(\frac{2\mu_y}{\mu_x^3}\sigma_x^2 - \frac{2}{\mu_x^2}\sigma_{xy}\right) = \frac{\mu_y}{\mu_x} + \frac{\mu_y}{\mu_x^3}\sigma_x^2 - \frac{\sigma_{xy}}{\mu_x^2} = \;= \frac{\mu_y}{\mu_x} + \frac{\sigma_x\mu_y}{\mu_x^2}\left(\frac{\sigma_x}{\mu_x} - \frac{\rho\sigma_y}{\mu_y}\right)$$

For the var(Y/X), we use only the first order approximation

$$var\left(\frac{Y}{X}\right) \approx var\left(\frac{\mu_y}{\mu_x} - \frac{\mu_y}{\mu_x^2}(X-\mu_x) + \frac{1}{\mu_x}(Y-\mu_y)\right) = \frac{\mu_y^2}{\mu_x^4}\sigma_x^2 + \frac{1}{\mu_x^2}\sigma_y^2 - \frac{2\mu_y}{\mu_x^3}\sigma_{xy} = \left(\frac{\mu_y}{\mu_x}\right)^2\left(\frac{\sigma_x^2}{\mu_x^2} + \frac{\sigma_y^2}{\mu_y^2} - \frac{2\sigma_{xy}}{\mu_x\mu_y}\right)$$

Distribution of Functions of Several Random Variables

Discrete Random Variables

It is relatively easy to find the pdf of Z=u(X,Y) when X and Y are discrete random variables. Consider the earlier two-coin toss example of discrete random variables X and Y with the joint pdf f(x,y):

		Y		
		0	1	2
X	0	0	1/2	0
	1	1/4	0	1/4

Consider a function Z=X+Y, which can take values {0, 1, 2, 3}. Their probabilities are:

$$\begin{aligned}
&P(Z=0) = P\{X=0,Y=0\} = 0,\\
&P(Z=1) = P\{X=0,Y=1\} + P\{X=1,Y=0\} = 1/2+1/4 = 3/4,\\
&P(Z=2) = P\{X=1,Y=1\} + P\{X=0,Y=2\} = 0+0 = 0,\\
&P(Z=3) = P\{X=1,Y=2\} = 1/4.
\end{aligned}$$

Therefore,

$$h(z) = \begin{cases} 3/4 & \textit{if} \quad z = 1 \\ 1/4 & \textit{if} \quad z = 3 \\ 0 & \textit{otherwise} \end{cases}$$

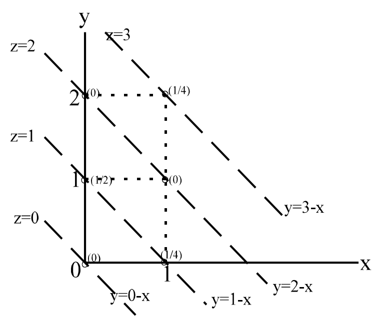
Similarly, the pdf of $Z=XY^2$, which can take values {0, 1, 4}, is P(Z=0)=3/4, P(Z=1)=0 and P(Z=4)=1/4. Note that, if the function is not well defined for the range of X and Y (eg. X/Y in this example), then its distribution is not computable. Formally, the pdf of Z=u(X,Y) for discrete X and Y can be written as

$$h(z) = \sum_{\{i,j\,;\, z = u(x_i,y_j)\}} f(x_i,y_j)$$

which is the sum of the joint pdf $f(x_i, y_j)$ for all x_i and y_j such that $z=u(x_i, y_j)$. In particular, the pdf of Z=X+Y can be written as

$$h(z) = \sum_i f(x_i, z-x_i) = \sum_j f(z-y_j, y_j)$$

Alternatively, we may use the cumulative distribution function technique and find $H(z) = P\{\omega; Z(\omega) \le z\}$. The figure below illustrates the case of Z=X+Y. Since $H(z)=P\{\{\omega;\ Y(\omega) \le z-X(\omega)\}$, we need to find the probability of all points that lie below and on the line y=z-x in the figure below.



If $z<1$,	$H(z) = P\{(0,0)\} = 0$
If $1 \le z < 2$,	$H(z) = P\{(0,0),(0,1),(1,0)\} = 0+1/2+1/4 = 3/4$
If $2 \le z < 3$,	$H(z) = P\{(0,0),(0,1),(1,0)\} + P\{(0,2),(1,1)\} = 3/4$
If $3 \le z$,	$H(z) = P\{(0,0),(0,1),(1,0)\} + P\{(0,2),(1,1)\} + P\{(1,2)\} = 1$

Therefore,

$$H(z) = \begin{cases} 0 & \text{if } \; z<1 \\ 3/4 & \text{if } \; 1 \le z < 3 \\ 1 & \text{if } \; z \ge 3 \end{cases}$$

which gives the same pdf as derived above.

Exercise: Find the pdf and cdf of Z=X-Y.

Continuous Random Variables

When X and Y are continuous random variables, we can use either the cumulative distribution function technique or the moment generating function technique.

[A] ***Moment Generating Function Technique***

As in the case of single random variable, this technique finds the mgf of u(X,Y), and then finds the distribution function that has the derived mgf. It is particularly useful in the proof of the distribution of the sum of reproducible independent random variables.

Example. ***Sum of Independent Exponential Variables***. We discussed earlier that if the waiting time X_i's are i.i.d. exponential, then their sum is distributed as a gamma distribution. This can be verified by using the mgf technique. The mgf of exponential distribution with parameter (arrival rate) λ is $\lambda/(\lambda$-t) for t<λ. Therefore,

$$E\left(e^{tY}\right) = E\left(e^{t\Sigma X_i}\right) = \prod_{i=1}^{n} E\left(e^{tX_i}\right) = \prod_{i=1}^{n} \frac{\lambda}{\lambda - t} = \left(\frac{\lambda}{\lambda - t}\right)^n$$

which is the mgf of a Gamma distribution with parameters λ and n. ◙

Example. ***Sum of Independent Normal Variables***. Let $X_i \sim N(\mu_i, \sigma_i^2)$, i=1,2,⋯,n, be *mutually independent* random variables, and let $Y=\sum X_i$. We know that $E(Y)=\mu=\sum\mu_i$ and $var(Y)=\sigma^2=\sum\sigma_i^2$ regardless of the distribution function of Y. The objective here is to show that Y is distributed as a normal random variable with mean μ and variance σ^2. The mgf of each normal random variable X_i is

$$E(e^{tX_i}) = e^{t\mu_i + t^2\sigma_i^2/2}$$

Therefore,

$$E(e^{tY}) = E(e^{t\sum X_i}) = E\left(\prod_{i=1}^{n} e^{tX_i}\right) = \prod_{i=1}^{n} E(e^{tX_i}) = \prod_{i=1}^{n} e^{t\mu_i + t^2\sigma_i^2/2} = e^{t\mu + t^2\sigma^2/2}$$

which is the mgf of a normal random variable $N(\mu, \sigma^2)$. This result also holds for any linear function $Y=\sum(a_iX_i + b_i)$ where a_i and b_i are constants. It was shown that $Z_i=a_iX_i+b_i$ is distributed as a normal with mean $a_i\mu_i + b_i$ and variance $a_i^2\sigma_i^2$. Hence $Y=\sum(a_iX_i + b_i)$ can be considered as a sum of independent normal random variables Z_i. ▣

Example. ***Joint Distribution of Linear Functions of Normal Variables***. Let X_1 and X_2 be distributed as *bivariate normal* with means μ_1 and μ_2, variances σ_{11} and σ_{22}, and covariance σ_{12}. Let $Y_i=a_{i1}X_1+a_{i2}X_2$, i=1,2, where a_{ij}'s are some constants. We wish to show that Y_1 and Y_2 have a *joint normal distribution* by using the mgf technique.

We have shown earlier that the joint mgf of jointly normal random variables X_1 and X_2 is

$$m(t_1, t_2) = E(e^{t_1X_1 + t_2X_2}) = \exp\left\{(t_1\mu_1 + t_2\mu_2) + \tfrac{1}{2}(t_1^2\sigma_{11} + 2t_1t_2\sigma_{12} + t_2^2\sigma_{22})\right\}$$

We will show that Y_1 and Y_2 have a joint mgf of same functional form.

$$\begin{aligned} m(t_1, t_2) &= E\left(e^{t_1Y_1 + t_2Y_2}\right) = E\left(e^{t_1(a_{11}X_1 + a_{12}X_2) + t_2(a_{21}X_1 + a_{22}X_2)}\right) = E\left(e^{(t_1a_{11} + t_2a_{21})X_1 + (t_1a_{12} + t_2a_{22})X_2}\right) \equiv E\left(e^{s_1X_1 + s_2X_2}\right) \\ &= \exp\left\{(s_1\mu_1 + s_2\mu_2) + \tfrac{1}{2}(s_1^2\sigma_{11} + 2s_1s_2\sigma_{12} + s_2^2\sigma_{22})\right\} \\ &= \exp\left\{(t_1\mu_{y_1} + t_2\mu_{y_2}) + \tfrac{1}{2}(t_1^2v_{11} + 2t_1t_2v_{12} + t_2^2v_{22})\right\} \end{aligned}$$

where

$$\begin{aligned} \mu_{y_i} &= a_{i1}\mu_1 + a_{i2}\mu_2, \qquad i=1,2 \\ v_{ii} &= a_{i1}^2\sigma_{11} + 2a_{i1}a_{i2}\sigma_{12} + a_{i2}^2\sigma_{22}, \qquad i=1,2 \\ v_{12} &= a_{11}a_{21}\sigma_{11} + (a_{11}a_{22} + a_{12}a_{21})\sigma_{12} + a_{12}a_{22}\sigma_{22} \end{aligned}$$

Since the joint mgf of Y_1 and Y_2 is the mgf of joint normal random variables, they have a joint normal distribution. Joint normality of Y_1 and Y_2 implies that the marginal distribution of Y_i is also normal $N(\mu_{y_i}, v_{ii})$. This result can be extended to linear functions of more than two jointly normal random variables as stated in the theorem below.

It is easy to verify that, if $a_{11}a_{22} = a_{12}a_{21}$, then $v_{22} = (a_{22}/a_{12})^2v_{11}$ and $v_{12} = (a_{22}/a_{12})v_{11}$. Therefore, the correlation coefficient between Y_1 and Y_2 is 1 if $a_{11}a_{22} = a_{12}a_{21}$. ▣

Example. ***Sum of Independent Chi-Square Variables***. Let X_i, i=1,2,⋯,n, be *mutually independent* non-central chi-square random variables: $\chi^2(\nu_i,\delta_i)$. Let $Y=\Sigma X_i$, $\nu=\Sigma\nu_i$, and $\delta=\Sigma\delta_i$. We have shown earlier that the mgf of each X_i is

$$E(e^{tX_i}) = \frac{e^{t\delta_i^2/(1-2t)}}{(1-2t)^{\nu_i/2}}$$

Therefore,

$$E\left(e^{tY}\right) = E\left(e^{t\Sigma X_i}\right) = \prod_{i=1}^{n} E\left(e^{tX_i}\right) = \prod_{i=1}^{n} \frac{e^{t\delta_i^2/(1-2t)}}{(1-2t)^{\nu_i/2}} = \frac{e^{t\delta^2/(1-2t)}}{(1-2t)^{\nu/2}}$$

which is the mgf of a non-central chi-square with ν degrees of freedom and non-centrality parameter δ. ▣

These important results are summarized in the theorems below.

Theorem. Let X_j, j=1,2,⋯,n be jointly normal random variables with mean μ_j, variance σ_{jj}, and covariance σ_{jk}. Let

$$Y_i = b_i + \sum_{j=1}^{n} a_{ij}X_j, \qquad i = 1,2,\cdots,m$$

where a_{ij} and b_i are constants. Then, Y_i's are jointly normal and the marginal distribution of Y_i is normal also normal

$$Y_i \sim N(\mu_{y_i},\sigma_{y_i}^2), \qquad \mu_{y_i} = b_i + \sum_{j=1}^{n} a_{ij}\mu_j, \qquad \sigma_{y_i}^2 = \sum_{k=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ik}\sigma_{jk}$$

Theorem. Let X_i, i=1,2,⋯,n be mutually independent chi-square variates with ν_i degrees of freedom and noncentrality parameter δ_i. Then,

$$\sum_{i=1}^{n} X_i \sim \chi^2(\nu,\delta), \qquad \nu = \sum_{i=1}^{n}\nu_i, \quad \delta = \sum_{i=1}^{n}\delta_i$$

Corollary. Let X_i, i=1,2,⋯,n be mutually independent normal variates with mean μ_i and variance σ_i^2. Let c be a constant. Then,

$$(a) \quad \sum_{i=1}^{n} \frac{(X_i-\mu_i)^2}{\sigma_i^2} \sim \chi^2(n), \qquad (b) \quad \sum_{i=1}^{n} \frac{(X_i-c)^2}{\sigma_i^2} \sim \chi^2(n,\delta), \quad \delta = \sum_{i=1}^{n} |\mu_i - c|/\sigma_i$$

Proof.

(a) $$\frac{X_i-\mu_i}{\sigma_i} \sim N(0,1) \quad \Rightarrow \quad \left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi^2(1) \quad \Rightarrow \quad \sum_{i=1}^{n} \frac{(X_i-\mu_i)^2}{\sigma_i^2} \sim \chi^2(n)$$

(b) Let $m_i = (\mu_i - c)/\sigma_i$. Then,

$$\frac{X_i - c}{\sigma_i} \sim N(m_i, 1), \quad \Rightarrow \quad \left(\frac{X_i - c}{\sigma_i}\right)^2 \sim \chi^2(1, |m_i|) \quad \Rightarrow \quad \sum_{i=1}^{n} \frac{(X_i - c)^2}{\sigma_i^2} \sim \chi^2(n, \delta), \quad \delta = \sum_{i=1}^{n} |m_i|$$

q.e.d. ◘

Following theorems are useful and given without proof.

Theorem. Let $Y = \sum_{i=1}^{n} X_i^2$ where X_i's are independent *standard normal* random variables. If Y=Z+W and Z~χ^2(m), then W~χ^2(n-m).

Theorem. Let Y=Z+W, where Y~χ^2(n), Z~χ^2(m), and W is nonngeative. Then, W~χ^2(n-m).

[B] ***Cumulative Distribution Function Technique***

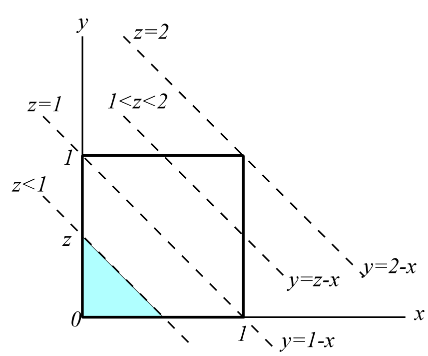
The method is same as the case of a function of a single random variable. We find the cdf of Z=u(X, Y) by applying the definition of the cdf, H(z)=P(u(X,Y)≤z), and then derive the pdf by differentiating the cdf.

Distribution of Sum of Two Random Variables

Consider a jointly uniform distribution:

$$f(x,y) = \begin{cases} 1 & \text{if} \quad x \in (0, 1), \quad y \in (0, 1) \\ 0 & \text{otherwise} \end{cases}$$

and we wish to find the pdf of Z=X+Y. The joint density is the top of a box of volume 1 on a unit square and the height of the box is 1. The cdf H(z)=P(Z≤z)=P(Y≤z-X) is the probability of all combination of x and y that satisfies y≤z-x for a given z. It is thus the volume of a section of the box that lies on and below the line y=z-x (see the figure below). Since the height of the density surface is 1, the volume is equal to the area of the box on and below the line y=z-x.



There are four cases to consider:

(i) If z<0, there is no box, and hence P(Z≤z)=0.

(ii) If 0≤z≤1, the base of the box below the line y=z-x is a triangle whose area is equal to $z^2/2$. Hence, P(Z≤z)=$z^2/2$.

(iii) If 1<z≤2, the base area can be computed by subtracting the area of a triangle at the upper-right corner

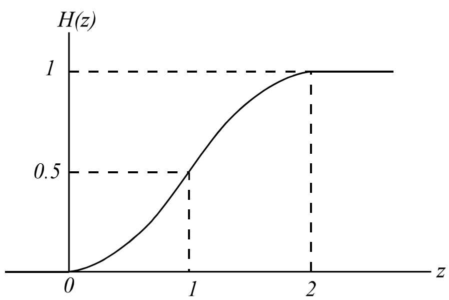
from the entire area 1. Since the triangle area is $(2-z)^2/2$, $P(Z \le z)=1-(2-z)^2/2$ for $1<z\le 2$.

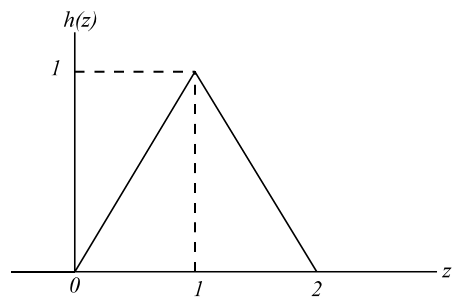
(iv) If $z>2$, the entire square is covered, and $P(Z\le z)=1$.

Therefore, the cdf and pdf of Z are

$$H(z) = \begin{cases} 0 & if \ \ z \le 0 \\ \dfrac{z^2}{2} & if \ \ 0 < z \le 1 \\ 1 - \dfrac{(2-z)^2}{2} & if \ \ 1 < z \le 2 \\ 1 & if \ \ 2 < z \end{cases} \qquad h(z) = \begin{cases} z & if \ \ 0 < z \le 1 \\ 2-z & if \ \ 1 < z \le 2 \\ 0 & otherwise \end{cases}$$

which are illustrated below.





Application of the cumulative distribution function technique to the sum of two random variables with joint pdf f(x,y) gives a very useful form as shown in the following theorem.

Theorem. Let f(x,y) be the joint pdf of continuous random variables X and Y, and let Z=X+Y. Then, the pdf of Z is given by

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

Proof. We find the cdf H(z) of Z, and take derivative of H(z) for the pdf of Z. The cdf of Z is

$$H(z) = P(Z \le z) = P(X+Y \le z) = \iint_{x+y\le z} f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-x} f(x,y)dy\right)dx = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f(x,v-x)dv\right)dx = \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f(x,v-x)dx\right)dv, \qquad or \ \ alternatively$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z-y} f(x,y)dx\right)dy = \int_{-\infty}^{\infty}\left(\int_{-\infty}^{z} f(u-y,y)du\right)dy = \int_{-\infty}^{z}\left(\int_{-\infty}^{\infty} f(u-y,y)dy\right)du$$

where the second expressions in the second and third lines are obtained by the change of variables (setting v=x+y, so that the upper limit z-x for y becomes the upper limit z for v, and y=v-x and dy=dv). The last expressions are simply exchanging the order of integrals. Differentiating the cdf with respect to z, we obtain the pdf of Z:

$$h(z) = \frac{\partial H(z)}{\partial z} = \int_{-\infty}^{\infty} f(x, z-x)dx, \qquad or, \qquad h(z) = \int_{-\infty}^{\infty} f(z-y, y)dy$$ ◘

Corollary. *Convolution*. Let $f_x(x)$ and $f_y(y)$ be the marginal pdf′s of *independent* random variables X and Y. Let Z=X+Y. Then, the pdf of Z is given by

$$h(z) = \int_{-\infty}^{\infty} f_x(x)f_y(z-x)dx = \int_{-\infty}^{\infty} f_x(z-y)f_y(y)dy$$

Proof. This result follows from the theorem above, because $f(x,y)=f_x(x)f_y(y)$ for independent random variables. ◘

Remark. The integral for h(z) for the independent random variables X and Y is called the ***convolution*** of functions f_x and f_y and frequently denoted by $f_x * f_y$. For discrete random variables, the convolution of f_x and f_y is

$$h(z) = \sum_i f_x(x_i)f_y(z-x_i) = \sum_j f_x(z-y_j)f_y(y_j)$$

which was shown earlier.

Corollary. Let Z=aX+bY, where a and b are finite constants. Then,

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{|b|} f\big(x, (z-ax)/b\big)dx, \qquad or, \qquad h(z) = \int_{-\infty}^{\infty} \frac{1}{|a|} f\big((z-by)/a,\, y\big)dy$$ ◘

Theorem. Let f(x,y) be the joint pdf of continuous random variables X and Y, and let W=X-Y. Then, the pdf of W is given by

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

If X and Y are independent,

$$g(w) = \int_{-\infty}^{\infty} f_x(x)f_y(x-w)dx = \int_{-\infty}^{\infty} f_x(w+y)f_y(y)dy$$ ◘

Example. Reconsider the example of the joint uniform distribution. Since X and Y are independent, we can use the convolution of f_x and f_y to derive the pdf of Z=X+Y. However, we have to be careful about the limits of x in the convolution. We know the marginal pdf's $f_x(x)=f_y(y)=1$ for $x\in(0,1)$ and $y\in(0,1)$. For a given value of z we need to find the range of x in which both densities $f_x(x)$ and $f_y(z\text{-}x)$ are positives. From the earlier graph, if 0<z<1, then both densities are positive for 0<x<z. If x<0, f_x=0, and if x>z, $f_y(z\text{-}x)$=0. Similarly, if 1<z<2, then both densities are positive only for z-1<x<1. The convolution for each range of z gives the pdf of Z as derived earlier.

Example. We have shown by using the mgf technique that the sum of i.i.d. exponential random variables is distributed as a gamma distribution. We derive it again here for two random variables by using the cumulative distribution function technique. Let X and Y be i.i.d. exponential random variables:

$$f_x(x) = \lambda e^{-\lambda x}, \qquad f_y(y) = \lambda e^{-\lambda y}, \qquad f(x,y) = f_x(x) f_y(y) = \lambda^2 e^{-\lambda(x+y)} \qquad for\ x>0,\ \ y>0$$

Let Z=X+Y. Then, applying the convolution of f_x and f_y,

$$h(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx$$

The lower and upper limits of the integral are 0 and z, respectively, because $f_x(x)=0$ if $x \le 0$ and $f_y(z\text{-}x)=0$ if $x \ge z$. Hence,

$$h(z) = \int_{-\infty}^{\infty} f_x(x) f_y(z-x) dx = \int_0^z f_x(x) f_y(z-x) dx = \int_0^z \lambda^2 e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z}, \qquad z>0$$

Noting that Γ(k)=(k-1)! for an integer k, it is easy to see that this is a gamma pdf with k=2 and λ.

Let W=X-Y. Then,

$$g(w) = \int_{-\infty}^{\infty} f_x(x) f_y(x-w) dx = \int_w^{\infty} f_x(x) f_y(x-w) dx = \int_w^{\infty} \lambda^2 e^{-\lambda(2x-w)} dx = \frac{\lambda e^{-\lambda w}}{2}, \qquad -\infty < w < \infty \quad \boxed{\bullet}$$

Exercise: Let X and Y be independent exponential random variables with parameters λ and θ, respectively. Find the pdf of Z=X+Y when $\lambda \neq \theta$. $\boxed{\bullet}$

Exercise. Let X and Y be independent normal random variables. Show by using the cdf technique that X+Y is distributed as a normal.

Distribution of Product and Quotient of Two Random Variables:

Theorem. Let Z=X/Y and W=XY, where X and Y have a joint pdf f(x,y). Then, the pdf's of Z and W are

(a) $$h(z) = \int_{-\infty}^{\infty} |y| f(zy,y) dy$$

(b) $$h(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, w/x)dx = \int_{-\infty}^{\infty} \frac{1}{|y|} f(w/y, y)dy$$

Proof. We will prove (a) here, and the proof of (b) is in the Appendix. Since P(X/Y$\le$z)=P(X$\le$zY) if Y>0, and P(X/Y$\le$z)=P(X$\ge$zY) if Y<0, we can split the integral as

$$H(z) = P(X/Y \le z) = \iint_{x/y \le z} f(x,y)dxdy = \int_{-\infty}^{0} \left(\int_{zy}^{\infty} f(x,y)dx \right) dy + \int_{0}^{\infty} \left(\int_{-\infty}^{zy} f(x,y)dx \right) dy$$

Change the variable by setting $v=x/y$, and substitute $x=yv$ for x, so that $dx=ydv$. The limits of integral in parentheses change as follows. The lower limit $x=zy$ in the first integral becomes $v=z$, and the upper limit x=∞ becomes $v=-\infty$ because y<0 for this integral. Similarly, the upper limit $x=zy$ of the integral in the second parenthesis becomes $v=z$ and the lower limit $x=-\infty$ becomes $v=-\infty$ because y>0 for this integral. Therefore,

$$H(z) = \int_{-\infty}^{0} \left(\int_{z}^{-\infty} yf(vy,y)dv \right) dy + \int_{0}^{\infty} \left(\int_{-\infty}^{z} yf(vy,y)dv \right) dy = \int_{-\infty}^{0} \left(\int_{-\infty}^{z} (-y)f(vy,y)dv \right) dy + \int_{0}^{\infty} \left(\int_{-\infty}^{z} yf(vy,y)dv \right) dy$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{0} (-y)f(vy,y)dy \right) dv + \int_{-\infty}^{z} \left(\int_{0}^{\infty} yf(vy,y)dy \right) dv = \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} |y| f(vy,y)dy \right) dv$$

Differentiation with respect to z gives the pdf h(z):

$$h(z) = \int_{-\infty}^{\infty} |y| f(zy,y)dy$$

q.e.d. ▣

Corollary. Let Z=X/(cY) where c is a positive constant. Then, the pdf of Z is

$$h(z) = \int_{-\infty}^{\infty} |cy| f(czy,y)dy$$

▣

Example: Jointly uniform X and Y over x∈(0,1) and y∈(0,1): $f_x(x)=1$ $f_y(y)=1$, f(x,y)=1 for x∈(0,1) and y∈(0,1). Note that X and Y are independent. Consider Z=XY. We find the limits of integral such that both $f_x(x)$ and $f_y(z/x)$ are positive. That is, 0<x<1 and 0<z/x<1, which give the limits 0<z<x<1. Therefore,

$$h(z) = \int_{z}^{1} \frac{1}{|x|} f_x(x) f_y(z/x) dx = \int_{z}^{1} \frac{1}{x} dx = -ln(z) \quad for\ z \in (0,1)$$

For the quotient Z=X/Y, the limits for the integral are (0, 1/z) if z>1 and (0,1) if z<1. Hence,

$$h(z) = \int_{0}^{1} y\,dy = \frac{1}{2} \qquad if\ z < 1, \qquad and \qquad h(z) = \int_{0}^{1/z} y\,dy = \frac{1}{2z^2} \qquad if\ z \ge 1$$

$\boxed{\bullet}$

Theorem: If X and Y are independent standard normal, then Z=X/Y has a *Cauchy* distribution.
Proof. Independence of X and Y implies $f(zy,y)=f_x(zy)f_y(y)$, and the pdf h(z) becomes

$$h(z) = \int_{-\infty}^{\infty} |y| f_x(zy) f_y(y) dy = \int_{-\infty}^{\infty} |y| \frac{1}{\sqrt{2\pi}} e^{-(zy)^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = = \frac{1}{2\pi} \int_{-\infty}^{\infty} |y| \; e^{-(1+z^2)y^2/2} dy$$
$$= \frac{1}{2\pi} \int_{-\infty}^{0} -y\, e^{-(1+z^2)y^2/2} dy + \frac{1}{2\pi} \int_{0}^{\infty} y\; e^{-(1+z^2)y^2/2} dy = \frac{1}{2\pi} \int_{0}^{\infty} y\, e^{-(1+z^2)y^2/2} dy + \frac{1}{2\pi} \int_{0}^{\infty} y\; e^{-(1+z^2)y^2/2} dy$$
$$= \frac{1}{\pi} \int_{0}^{\infty} y\; e^{-(1+z^2)y^2/2} dy = \frac{1}{\pi(1+z^2)} \int_{0}^{\infty} e^{-u} du \qquad \textit{by setting } u = (1+z^2)y^2/2$$
$$= \frac{1}{\pi(1+z^2)} \cdot \Gamma(1) = \frac{1}{\pi(1+z^2)}$$

which is a special case (α=0 and β=1) of a *Cauchy* pdf

$$f(x;\alpha,\beta) = \frac{1}{\pi\beta(1 + [(x-\alpha)/\beta]^2)}, \qquad -\infty < \alpha < \infty, \quad \beta > 0$$

This result cannot be proven by the mgf technique, because the mgf does not exist for a *Cauchy* random variable. $\boxed{\bullet}$

Following four theorems and a corollary play central roles in the test of hypothesis, and should be thoroughly studied. Their proofs are straightforward applications of the procedure described above, and they are presented in the Appendix.

Theorem. Let X~N(0,1), Y~χ^2(m), and X and Y be independent. Then, $\dfrac{X}{\sqrt{Y/m}} \sim t(m)$.

Theorem. Let X~χ^2(m), Y~χ^2(n), and X and Y be independent. Then, $\dfrac{X/m}{Y/n} \sim F(m,n)$.

Theorem. If X~N(μ,1), Y~χ^2(m), and X and Y are independent, $\dfrac{X}{\sqrt{Y/m}} \sim t(m;\mu)$.

Theorem. If X~χ^2(m, δ), Y~χ^2(n), and X and Y are independent, $\dfrac{X/m}{Y/n} \sim F(m,n,\delta)$.

Corollary. If X~N(μ,σ^2), Y~χ^2(m), and X and Y are independent,

$$\frac{(X-\mu)/\sigma}{\sqrt{Y/m}} \sim t(m), \qquad \frac{X/\sigma}{\sqrt{Y/m}} \sim t(m;\mu/\sigma)$$

$$\frac{(X-\mu)^2/\sigma^2}{Y/m} \sim F(1,m), \qquad \frac{X^2/\sigma^2}{Y/m} \sim F(1,m;|\mu|/\sigma)$$

Corollary. (a) If Z~t(m), then Z^2~F(1,m). (b) If Z~F(m,n), then Z^{-1}~F(n,m).

Change of Variable Approach

There is a common feature in the pdf's of the three functions we just considered:

$$Z = aX + bY: \qquad h(z) = \int_{-\infty}^{\infty} \frac{1}{|a|} f\big((z-by)/a,\, y\big)dy$$

$$Z = cXY: \qquad h(z) = \int_{-\infty}^{\infty} \frac{1}{|cy|} f\big(z/(cy),\, y\big)dy$$

$$Z = X/cY: \qquad h(z) = \int_{-\infty}^{\infty} |cy| f(czy,\, y)dy$$

All three are given as an integral, where the inegrands are a function of z and y and do not involve x. They appear to be joint densities of y and z, and the pdf of Z is derived as a marginal pdf. This is indeed the case. To see this, recall that in the case of a function of single variable Z=u(X), the pdf of Z is derived as $h(z) = f(u^{-1}(z))|J|$ if the inverse function exists. We can apply this technique in the case of the function of two variables Z=u(X,Y), if the inverse function exists between Z and Y holding X constant, or between Z and X holding Y constant. This approach leads to the derivation of joint pdf g(x,z) or g(y,z) from f(x,y), and then the pdf of Z is derived as the marginal pdf.

For instance, consider the pdf of Z=cXY. *Holding y constant*, x=z/cy and the Jacobian of the transformation between X and Z is $J=(cy)^{-1}$. Therefore, the joint pdf of x and z can be written as

$$g(y,z) = |J| f(z/(cy), y) = \frac{1}{|cy|} f\big(z/(cy), y\big)$$

which is the integrand of the pdf of z we derived earlier. If we hold x constant, a similar procedure gives a joint pdf of X and Z:

$$g(x,z) = \frac{1}{|cx|} f\big(x, z/(cx)\big)$$

In a general case of Z=u(X,Y), if the inverse function X=v(Y,Z) exists between X and Z, holding Y constant, then the joint pdf of Y and Z is given by

$$g(y,z) = |J| f\big(v(y,z), y\big), \qquad J = \frac{\partial v(y,z)}{\partial z}$$

and the pdf of Z is derived as the marginal pdf from g(y,z). This is a result of the following theorem.

Theorem. Let X and Y be continuous random variables with a joint pdf f(x,y). Let W=v(X,Y) and Z=u(X,Y) be partially differentiable functions over the range of X and Y where f(x,y)≠0. If these functions can be uniquely solved for X and Y in terms of W and Z such that X=φ(W,Z) and Y=ψ(W,Z), then the joint pdf g(w,z) of W and Z is given by

$$g(w,z) = |J| f\big(\varphi(w,z),\, \psi(w,z)\big)$$

where the Jacobian J is the determinant

$$J = \begin{vmatrix} \partial x/\partial w & \partial x/\partial z \\ \partial y/\partial w & \partial y/\partial z \end{vmatrix} = (\partial x/\partial w)(\partial y/\partial z) - (\partial y/\partial w)(\partial x/\partial z)$$

⊡

Example. Linear Functions. Let f(x,y) be the joint pdf of X and Y. Let $W=a_1X+b_1Y$ and $Z=a_2X+b_2Y$. We wish to find the joint pdf g(w,z) of W and Z. The inverse functions for X and Y are

$$X = \frac{b_2W - b_1Z}{d}, \qquad Y = \frac{a_1Z - a_2W}{d}, \qquad d \equiv a_1b_2 - a_2b_1$$

and the Jacobian of the transformation is

$$J = \begin{vmatrix} \partial x/\partial w & \partial x/\partial z \\ \partial y/\partial w & \partial y/\partial z \end{vmatrix} = \begin{vmatrix} b_2/d & -b_1/d \\ -a_2/d & a_1/d \end{vmatrix} = \frac{1}{d}$$

Hence,

$$g(w,z) = |J| f\big(\varphi(w,z),\, \psi(w,z)\big) = \frac{1}{|d|} f\Big((b_2w - b_1z)/d,\; (a_1z - a_2w)/d \Big)$$

Example. Let Z=u(X,Y)=X/(X+Y), and we wish to find the pdf of Z. Let W=v(X,Y)=X. Then,

$$\left.\begin{aligned} W &= v(X,Y) = X \\ Z &= u(X,Y) = \frac{X}{X+Y} \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} X &= \varphi(W,Z) = W \\ Y &= \psi(W,Z) = \frac{(1-Z)W}{Z} \end{aligned}\right\} \Rightarrow J = \begin{vmatrix} 1 & 0 \\ \dfrac{1}{1-z} & -\dfrac{w}{z^2} \end{vmatrix} = -\frac{w}{z^2}$$

Hence,

$$g(w,z) = |J| f\big(\varphi(w,z),\, \psi(w,z)\big) = |w/z^2| f\big(w,\, (1-z)w/z \big)$$

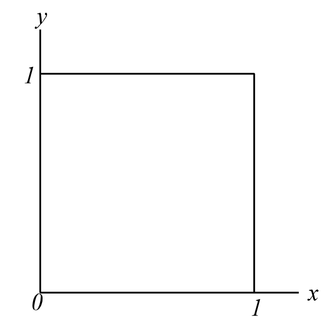
Since w=x, this can be written as a joint pdf of X and Z:

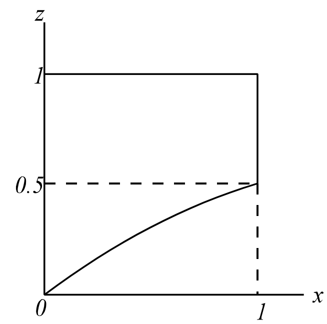
$$g(x,z) = |x/z^2| f\big(x,\, (1-z)x/z \big)$$

Suppose that X and Y are jointly uniform over the unit square; f(x,y)=1 for x∈(0,1) and y∈(0,1). Then, the joint pdf g(x,z) takes a positive value if

$$\begin{array}{l} 0 < x < 1 \\ 0 < y < 1 \end{array} \quad \Rightarrow \quad 0 < \frac{x(1-z)}{z} < 1 \quad \Rightarrow \quad \frac{x}{1+x} < z < 1$$

These regions are illustrated below.





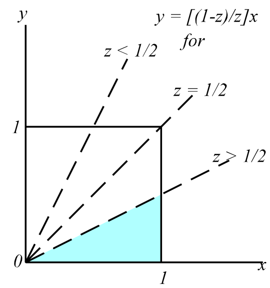
To find the marginal pdf of Z we integrate g(x,z) with respect to x. The range of x is where the density $f(x, x(1-z)/z)$ takes a positive value, i.e., 0 < x < min[1, z/(1-z)]. The upper limit of x is equal to z/(1-z) if $0<z\leq 1/2$, and equal to 1 if $1/2<z\leq 1$.Since f(x,y)=1 for this range of values, we have

$$h(z) = \begin{cases} \displaystyle\int_0^{z/(1-z)} \frac{x}{z^2} dx = \frac{1}{2(1-z)^2} & \text{if} \quad 0 < \dfrac{z}{1-z} \leq 1 \quad \Rightarrow \quad \text{if} \quad 0 < z \leq \frac{1}{2} \\ \displaystyle\int_0^{1} \frac{x}{z^2} dx = \frac{1}{2z^2} & \text{if} \quad \dfrac{z}{1-z} \geq 1 \quad \Rightarrow \quad \text{if} \quad \frac{1}{2} < z \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

It is instructive to work this out directly from the cdf of Z:

$$H(z) = P(Z \leq z) = P\left(\frac{X}{X+Y} \leq z\right) = P\left(\frac{1-z}{z} X \leq Y\right) = 1 - P\left(Y \leq \frac{1-z}{z} X\right)$$

The probability in the last term is the probability over a segment of the unit square that lies below the line y=[(1-z)/z]x. If $z\geq 1/2$, the area is a triangle. If $z\leq 1/2$, the area is a trapezoid, whose area can be computed by subtracting from 1 the area of triangle above the dashed line. Since the height of the box is 1 for this joint uniform distribution, the area itself is the probability.



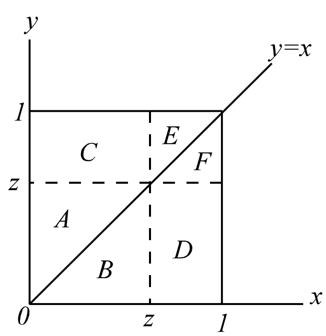
Thus,

$$H(z) = \begin{cases} 0 & \text{if } z \le 0 \\ \dfrac{z}{2(1-z)} & \text{if } 0 < z \le \frac{1}{2} \\ \dfrac{3z-1}{2z} & \text{if } \frac{1}{2} < z \le 1 \\ 1 & \text{if } 1 < z \end{cases} \qquad h(z) = \begin{cases} \dfrac{1}{2(1-z)^2} & \text{if } 0 < z \le \frac{1}{2} \\ \dfrac{1}{2z^2} & \text{if } \frac{1}{2} < z \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$\boxed{\bullet}$

Distribution of Extrema. max[X,Y] and min[X,Y].

Consider a uniform distribution: $f(x,y)=1$ if $x\in(0,1)$ and $y\in(0,1)$. In the figure below, $x\le y\le z$ in triangle area A, $y\le x\le z$ in triangle area B, $x\le z\le y$ in area C, $y\le z\le x$ in area D, $z\le x\le y$ in area E and $z\le y\le x$ in area F.



Thus, $\max[X,Y]\le z$ is satisfied only in areas A and B, and $P(\max[X,Y]\le z)=P(A)+P(B)$. These areas are characterized by the condition that *both* X and Y take a value less than or equal to z. Therefore,

$$H(z) = P(\max[X,Y] \le z) = P(X \le z, Y \le z) = F(z,z)$$

which is simply the volume of the box with base area z^2. Therefore,

$$H(z) = \begin{cases} 0 & if \;\; z \le 0 \\ z^2 & if \;\; 0 < z \le 1 \\ 1 & if \;\; 1 < z \end{cases} \qquad \Rightarrow \qquad h(z) = \begin{cases} 2z & if \;\; 0 \le z \le 1 \\ 0 & otherwise \end{cases}$$

Consider now P(min[X,Y]≤z). The inequality min[X,Y]≤z is clearly satisfied in areas A and B. In area C, min[X,Y]=X whose value is less than or equal to z. Similarly, min[X,Y]≤z in area D also. The only areas where the inequality is not satisfied are areas E and F. Therefore, P(min[X,Y]≤z)=P(A)+P(B)+P(C)+P(D). These areas are characterized by the condition that *either* X *or* Y, *or both* X and Y, take a value less than or equal to z. Instead of computing the probability by the sum of probabilities of A, B, C and D, we can also compute it by subtracting the probabilities of areas E and F from 1. Areas E and F are where *both* X and Y exceeds z. This gives

$$G(z) = P(\min[X,Y] \le z) = P(X \le z) + P(Y \le z) - P(X \le z, Y \le z) = 1 - P(X>z, Y>z)$$

For the uniform density under consideration, we have

$$G(z) = \begin{cases} 0 & if \;\; z \le 0 \\ 1 - (1-z)^2 & if \;\; 0 < z \le 1 \\ 1 & if \;\; 1 < z \end{cases} \qquad \Rightarrow \qquad g(z) = \begin{cases} 2(1-z) & if \;\; 0 \le z \le 1 \\ 0 & otherwise \end{cases}$$

These results are generalized for more than two random variables in the following theorem.

Theorem. Let X_i, $i=1,2,\cdots,n$, be the random variables with a joint cdf $F(x_1, x_2,...,x_n)$. Then, the cdf ′s of $\max[X_1, X_2,...,X_n]$ and $\min[X_1, X_2,...,X_n]$ are

$$H(z) = P(\max[X_1,X_2,\cdots,X_n] \le z) = P(X_1 \le z, X_2 \le z, \cdots, X_n \le z) = F(z,z,\cdots,z)$$

$$G(z) = P(\min[X_1,X_2,\cdots,X_n] \le z) = 1 - P(X_1>z, X_2>z, \cdots, X_n>z)$$

If X_i’s are *independent* and each X_i has the cdf $F_i(x)$, then the cdf′s become

$$H(z) = P(X_1 \le z)P(X_2 \le z) \cdots P(X_n \le z) = \prod_{i=1}^{n} P(X_i \le z) = \prod_{i=1}^{n} F_i(z)$$

$$G(z) = 1 - P(X_1>z)P(X_2>z) \cdots P(X_n>z) = 1 - \prod_{i=1}^{n} P(X_i>z) = 1 - \prod_{i=1}^{n} [1 - F_i(z)]$$

If X_i’s are *independent and identically distributed* (i.i.d.) with a common cdf F(x), then the cdf′s become

$$H(z) = \prod_{i=1}^{n} F(z) = [F(z)]^n$$

$$G(z) = 1 - \prod_{i=1}^{n} [1 - F(z)] = 1 - [1 - F(z)]^n$$

◘

Corollary. Let X_i, i=1,2,⋯,n, be the *i.i.d. random variables* with a common cdf F(x) and pdf f(x). Then, the pdf of max[$X_1, X_2,...,X_n$] is $h(z) = n[F(z)]^{n-1}f(z)$, and the pdf of min[$X_1, X_2,...,X_n$] is $g(z) = n[1 - F(z)]^{n-1}f(z)$. ◙

Theorem. Let X_i, i=1,2,⋯,n, be the i.i.d. *nonnegative* random variables with a common cdf F(x). Let Z=max[$X_1, X_2,...,X_n$] and W=min[$X_1, X_2,...,X_n$]. Then,

$$E(Z) = \int_0^\infty \left(1 - [F(z)]^n\right)dz, \qquad E(W) = \int_0^\infty [1 - F(w)]^n dw$$

Proof. We have shown earlier that, for a random variable Z with a cdf H(z),

$$E(Z) = \int_0^\infty (1 - H(z))dz - \int_{-\infty}^0 H(z)dz$$

Since the cdf of Z is $H(z) = [F(z)]^n$ and the cdf of W is $G(w) = 1 - [1 - F(w)]^n$, the results follow because of nonnegativity of the random variables. ◙

Theorem. Let X be a random variable with a cdf F(x) and α be a finite constant. Then,

$$E(\max[\alpha, X]) = \alpha + \int_\alpha^\infty (x - \alpha)\,dF(x)$$

Proof.

$$E(\max[\alpha, X]) = \alpha P(X \le \alpha) + \int_\alpha^\infty x\,dF(x) = \alpha \int_{-\infty}^\alpha dF(x) + \int_\alpha^\infty x\,dF(x)$$

$$= \alpha \int_{-\infty}^\infty dF(x) - \alpha \int_\alpha^\infty dF(x) + \int_\alpha^\infty x\,dF(x) = \alpha + \int_\alpha^\infty (x - \alpha)\,dF(x)$$

◙

Remark. This theorem is used in the theory of optimal stopping in sequential decision problems.

Example. Three people participate in a sealed-bid first price auction of a painting. Each person draws her bid price in dollars independently from a uniform distribution U(0,1). We do not consider in this example the strategic behavior of each agent in their choice of bid price.
(a) What is the probability of wining for a person who submits bid price 50 cents? 30 cents? 90 cents?
(b) What is the expected revenue for the seller?

Solution. Let X_i be the bid price of person i=1,2,3. Since X_i's are i.i.d. with a common cdf F(x)=x for $0 \le x \le 1$, the cdf of max[X_1, X_2, X_3] is

$$P(\max[X_1, X_2, X_3] \le z) = [F(z)]^3 = z^3 \quad \textit{if } 0 \le z \le 1$$

(a) A person with a bid price z wins the auction if max[X_1, X_2, X_3] $\le$z. Therefore, the probability of winning with a bid price 50 cents is 0.5^3=0.125; with a bid price 30 cents, 0.3^3=0.027; and with a bid price 90 cents, 0.9^3=0.729. Note that we ignored the possibility of a tie, because P(X_i=X_j)=0, i≠j, for continuous random variables.
(b) The expected value of Z=max[X_1, X_2, X_3] is

$$E(Z) = \int_0^\infty \left(1 - [F(z)]^3\right) dz = \int_0^1 (1 - z^3) dz = \frac{3}{4}$$ ■

Example. Fixed Sample Size Search. There are many stores that carry an identical good, and prices vary across the stores with a known cumulative distribution function F(x). A consumer wishes to determine how many stores to search for the lowest price. Once the number of stores (n) for search is determined, she randomly selects n stores, visits all n stores, and then buys from the store with a lowest price. Each trip to (or calling) the store costs \$c. The number of stores for search is determined to minimize the expected total costs of purchasing the item.

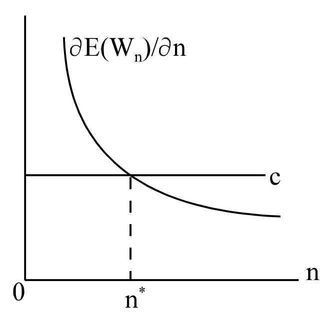
Solution. Let n be the number of stores for search, and X_i be the price of store i, i=1,2,...,n. Random selection of stores for search implies that X_i's are i.i.d. nonnegative random variables with common cdf F(x). Total cost of purchasing the item is the sum of total search costs \$nc, and the price she pays, $W_n=\min[X_1, X_2,...,X_n]$, which also depends on the decision variable n. We wish to minimize the expected total cost with respect to n:

$$\min_n TC_n \equiv \min_n [nc + E(W_n)] = \min_n \left\{ nc + \int_0^\infty [1 - F(w)]^n dw \right\}$$

Differentiating with respect to n, we get the first order condition (recall that $\frac{\partial a^x}{\partial x} = a^x lna$ for a constant a)

$$\frac{\partial TC_n}{\partial n} = c + \int_0^\infty [1 - F(w)]^n ln[1 - F(w)]\, dw = 0$$

Note that the first term "c" is the marginal cost of one more search, and the second term (the integral) is the marginal benefit of one more search in terms of a lower expected price. Note that $\partial E(W_n)/\partial n < 0$.



If X has an exponential distribution $F(x)=1-e^{-\lambda x}$ for instance, then

$$\int_0^\infty [1 - F(w)]^n ln[1 - F(w)]\, dw = \int_0^\infty [e^{-\lambda w}]^n ln(e^{-\lambda w}) dw = \int_0^\infty -\lambda w e^{-\lambda n w} dw = -\frac{1}{n}\int_0^\infty \lambda n w e^{-\lambda n w} dw$$

$$= -\frac{1}{n}\int_0^\infty \delta w e^{-\delta w} dw = -\frac{1}{n}\frac{1}{\delta} = -\frac{1}{\lambda n^2}$$

where the last integral can be found by integration by parts. Therefore, $n = 1/\sqrt{\lambda c}$, which is a decreasing function of the search cost, and an increasing function of the mean of the price distribution. Recall that the mean of the exponential distribution is $1/\lambda$. ▣

Example. Sequential Search with Recall

In the problem described above, the number of stores to be searched is fixed. The search strategy requires her to finish the search of all n stores even if she finds an unusually low price in the first store. A better search strategy is a sequential search. Let F(x) be the cdf of price distribution, and the marginal search cost is c. At any point of sequential search, she has two choices:

(1) Accept the lowest price w among all prices observed so far, and stop search.

(2) Search one more time, observing one more price X at an additional search cost c. She decides either to accept the lowest price so far, including the new price X, (i.e., same as choice (1)), or to continue one more search.

The question is when she should stop search.

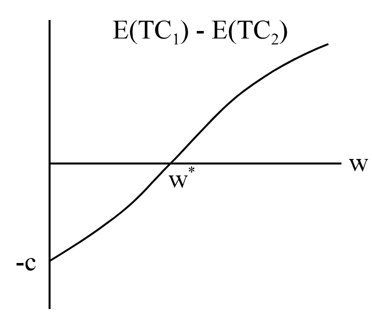
If she stops without one more search, case (1), the total cost TC_1 is the price she pays (w) plus all past search cost. Since past search cost is a sunk cost and does not affect the current decision of search one more or not, we will set past search cost to zero. If she searches one more time and then accepts the lowest price, the total cost will be $TC_2=c+X$ if $X<w$, and $TC_2=c+w$ if $X>w$. Before she makes one additional search, the total cost is unknown because X is random. Therefore, she needs to compare TC_1 with the expected TC_2, assuming that she is risk neutral. The expected total cost $E(TC_2)$ is

$$\begin{aligned} E(TC_2) &= P(X>w)\cdot(c+w) + P(X<w)\cdot E(c+X\,|\,X<w) \\ &= [1-F(w)]\cdot(c+w) + F(w)\cdot\int_0^w (c+x)\frac{f(x)}{F(w)}dx = [1-F(w)]\cdot(c+w) + F(w)\cdot c + \int_0^w xf(x)dx \\ &= c + w[1-F(w)] + \int_0^w xf(x)dx = c + w[1-F(w)] + wF(w) - \int_0^w F(x)dx = c + w - \int_0^w F(x)dx \end{aligned}$$

Therefore,

$$E(TC_1) - E(TC_2) = \int_0^w F(x)dx - c$$

At any current lowest price w, she will keep search if $E(TC_1)-E(TC_2)>0$. At what value of w should she stop search? It is easy to see that $E(TC_1)-E(TC_2)$ is an increasing function of w, because the cdf F(x) is an increasing function. This is illustrated below.



As long as the current lowest price is higher than w^*, the total cost is expected to be lower when she searches one more store. When she observes a price lower than w^* at any point of her search, she will accept that price and stops searching. This critical value w^* is called the *reservation price*. ▣

Studentized maximum modulus distribution

Let $Z_1, Z_2, \cdots, Z_n$ be the n *independent* standard normal random variables, and let Y be a central chi-square variate: $Y \sim \chi^2(m)$. Then,

$$X \equiv \frac{\max_i |Z_i|}{\sqrt{Y/m}}$$

has the studentized maximum modulus distribution. The closed form of the pdf is not available. Its cdf has two parameters m and n, and is given by

$$F(x;m,n) = \int_0^\infty [2\Phi(xw) - 1]^n g(w;m)dw, \quad x \geq 0$$

where $\Phi(z)$ is the cdf of a standard normal variate, and $g(w;m)$ is the pdf of $W = \sqrt{Y/m}$:

$$g(w;m) = \frac{m^{m/2}}{\Gamma(m/2)2^{(m-2)/2}} w^{m-1} e^{-mw^2/2}, \quad w \geq 0$$

This distribution arises in a simultaneous statistical inference. See Rupert G. Miller, Jr, *Simultaneous Statistical Inference*, pp.70-75. See also Gordon Anderson, "Nonparametric Tests of Stochastic Dominance in Income Distribution," Econometrica, Vol. 64, No. 5 (September, 1996), 1183-1193.

Assignment #6

(1) Let f(x,y) be the joint pdf of continuous random variables X and Y. Show that the pdf of Z=X-Y is

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

(2) Let X and Y have the joint pdf

$$f(x,y) = \frac{xy}{36} \quad \textit{for } x = 1,2,3, \quad \textit{and} \quad y = 1,2,3.$$

Find the pdf's for Z=X+Y, W=XY and V=X/Y.

(3) Let X and Y be i.i.d. U(0,1). Find the pdf's of Z=X+Y and W=X-Y, and draw them.

(4) Let X and Y be independent random variables having exponential densities with parameters λ and θ. Find the pdf of Z=X+Y when (i) $\lambda \neq \theta$, and (ii) $\lambda = \theta$.

(5) Let X and Y be i.i.d. exponential random variables with parameters λ. Find the pdf of W=X/Y. Show that W is independent of Z=X+Y.

(6) Let X and Y be i.i.d. U(0,1). Find the pdf's of Z=max[X,Y] and W=min[X,Y] and draw them.

(7) Let X be a random variable with a cdf F(x) and α be a finite constant. Find the expression for $E(\min[\alpha, X])$

APPENDIX

Theorem. Let Z=XY. Then, the pdf of Z is

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, z/x)dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{|x|} f_x(x) f_y(z/x)dx \quad \textit{if X and Y are independent}$$

Proof. Since P(XY≤z)=P(Y≤z/X) if X>0, and P(XY≤z)=P(Y≥z/X) if X<0, we can split the integral as

$$H(z) = P(Z \le z) = P(XY \le z) = \iint_{xy \le z} f(x,y)dxdy = \int_{-\infty}^{0} \left(\int_{z/x}^{\infty} f(x,y)dy \right) dx + \int_{0}^{\infty} \left(\int_{-\infty}^{z/x} f(x,y)dy \right) dx$$

Let $v=xy$, and substitute $y=v/x$ for y, so that $dy=x^{-1}dv$. The limits of integral in parentheses change as follows. The lower limit $y=z/x$ in the first integral becomes $v=z$, and the upper limit y=∞ becomes $v=-\infty$ because x<0 for this integral. Similarly, the upper limit $y=z/x$ of the integral in the second parenthesis becomes $v=z$ and the lower limit $y=-\infty$ becomes $v=-\infty$ because x>0 for this integral. Therefore,

$$H(z) = \int_{-\infty}^{0} \left(\int_{z}^{-\infty} x^{-1} f(x, v/x)\, dv \right) dx + \int_{0}^{\infty} \left(\int_{-\infty}^{z} x^{-1} f(x, v/x)\, dv \right) dx$$

$$= \int_{-\infty}^{0} \left(\int_{-\infty}^{z} (-x^{-1}) f(x, v/x)\, dv \right) dx + \int_{0}^{\infty} \left(\int_{-\infty}^{z} x^{-1} f(x, v/x)\, dv \right) dx$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{0} (-x^{-1}) f(x, v/x)\, dx \right) dv + \int_{-\infty}^{z} \left(\int_{0}^{\infty} x^{-1} f(x, v/x)\, dx \right) dv$$

$$= \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} |x|^{-1} f(x, v/x)\, dx \right) dv$$

Differentiating with respect to z we derive the pdf h(z):

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, z/x)dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{|x|} f_x(x) f_y(z/x)dx \quad \textit{if X and Y are independent}$$

If we substitute x=v/y instead, we will have equivalent results

$$H(z) = \int_{-\infty}^{z} \left(\int_{-\infty}^{\infty} |y|^{-1} f(v/y, y)\, dy \right) dv$$

$$h(z) = \int_{-\infty}^{\infty} \frac{1}{|y|} f(z/y, y) dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{|y|} f_x(z/y) f_y(y) dy \quad \textit{if X and Y are independent}$$

q.e.d. ▣

Theorem. Let X~N(0,1), Y~χ^2(m), and X and Y be independent. Then,

$$Z = \frac{X}{\sqrt{Y/m}} \sim t(m)$$

i.e., a central t with m degrees of freedom.

Proof. The pdf of a central χ^2(m) is

$$f_y(y; m) = \frac{1}{\Gamma(m/2) 2^{m/2}} y^{(m-2)/2} e^{-y/2}, \quad y \geq 0$$

Let $W = \sqrt{Y/m}$. This transformation has the inverse function $Y = mW^2$ and the Jacobian J=2mW. Therefore, from the previous derivation of the pdf of a square root function, the pdf of W can be written as

$$f_w(w) = |J| f_y(mw^2) = \frac{1}{\Gamma(m/2) 2^{m/2}} |2mw| (mw^2)^{(m-2)/2} e^{-mw^2/2} = \frac{m^{m/2}}{\Gamma(m/2) 2^{(m-2)/2}} w^{m-1} e^{-mw^2/2}, \quad w > 0$$

Since Z=X/W, the density of Z is derived by

$$h(z) = \int_{-\infty}^{\infty} |w| f_x(zw) f_w(w) dw = \int_0^{\infty} w \frac{1}{\sqrt{2\pi}} e^{-z^2 w^2/2} \cdot \frac{m^{m/2}}{\Gamma(m/2) 2^{(m-2)/2}} w^{m-1} e^{-mw^2/2} dw$$

$$= \frac{m^{m/2}}{\sqrt{2\pi}\, \Gamma(m/2) 2^{(m-2)/2}} \int_0^{\infty} e^{-(m+z^2) w^2/2} w^m dw$$

To find the integral in the last expression, let $u = (m+z^2) w^2/2$. Then,

$$w^m dw = \frac{2^{(m-1)/2}}{(m+z^2)^{(m+1)/2}} u^{(m-1)/2} du$$

so that

$$h(z) = \frac{m^{m/2}}{\sqrt{2\pi}\,\Gamma(m/2)2^{(m-2)/2}}\,\frac{2^{(m-1)/2}}{(m+z^2)^{(m+1)/2}}\int_0^\infty e^{-u}u^{(m-1)/2}du = \frac{1}{\Gamma(m/2)\sqrt{\pi m}\left(1+z^2/m\right)^{(m+1)/2}}\int_0^\infty e^{-u}u^{(m-1)/2}du$$
$$= \frac{\Gamma\left((m+1)/2\right)}{\Gamma(m/2)\sqrt{\pi m}\left(1+z^2/m\right)^{(m+1)/2}}$$

which is the pdf of a central t-distribution with m degrees of freedom. The last equality follows from the definition of the Gamma function

$$\Gamma(n) = \int_0^\infty e^{-x}x^{n-1}dx$$

q.e.d. ▣

Theorem. Let X~χ^2(m), Y~χ^2(n), and X and Y be independent. Then,

$$Z = \frac{X/m}{Y/n} \sim F(m,n)$$

Proof. Let c=m/n. Then, Z=X/(cY) where c is a positive constant. Hence, from the earlier derivation of the pdf of Z

$$h(z) = \int_{-\infty}^{\infty} |cy| f_x(czy)f_y(y)dy$$
$$= \int_0^\infty |cy| \frac{1}{\Gamma(m/2)2^{m/2}}(czy)^{(m-2)/2}e^{-czy/2}\cdot\frac{1}{\Gamma(n/2)2^{n/2}}y^{(n-2)/2}e^{-y/2}\,dy$$
$$= \frac{c^{m/2}z^{(m-2)/2}}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}}\int_0^\infty y^{(m+n-2)/2}e^{-(1+cz)y/2}\,dy$$
$$= \frac{c^{m/2}z^{(m-2)/2}}{\Gamma(m/2)\Gamma(n/2)2^{(m+n)/2}}\int_0^\infty \left(\frac{2u}{1+cz}\right)^{(m+n-2)/2}e^{-u}\left(\frac{2}{1+cz}\right)du \qquad where \qquad u=\frac{(1+cz)y}{2}$$
$$= \frac{c^{m/2}\,z^{(m-2)/2}}{\Gamma(m/2)\Gamma(n/2)\,(1+cz)^{(m+n)/2}}\int_0^\infty u^{(m+n-2)/2}e^{-u}du = \frac{c^{m/2}\,z^{(m-2)/2}}{\Gamma(m/2)\Gamma(n/2)\,(1+cz)^{(m+n)/2}}\Gamma((m+n)/2)$$

which is the pdf of F-distribution with degrees of freedom m and n: F(m,n). q.e.d. ▣

Theorem. If X~χ^2(m, δ), Y~χ^2(n), and X and Y are independent,

$$\frac{X/m}{Y/n} \sim F(m,n,\delta)$$

Proof. The noncentral chi-square pdf for X is

$$f(x;m,\delta) = \sum_{j=0}^{\infty}\left(\frac{e^{-\alpha}\alpha^j}{j!}\right)\frac{x^{k_j-1}e^{-x/2}}{\Gamma(k_j)\,2^{k_j}} \qquad x>0,\quad \delta\ge 0,\quad \alpha = \frac{\delta^2}{2},\quad k_j = j+\frac{m}{2}$$

and the central chi-square pdf for Y is

$$f(y;n) = \frac{1}{\Gamma(n/2)2^{n/2}} y^{(n-2)/2} e^{-y/2}, \quad y \geq 0$$

Let c=m/n. Since X and Y are independent,

$$\begin{aligned}
h(z) &= \int_{-\infty}^{\infty} |cy| f(czy, y) dy = \int_{-\infty}^{\infty} |cy| f_x(czy) f_y(y) dy \\
&= \int_{0}^{\infty} |cy| \sum_{j=0}^{\infty} \left(\frac{e^{-\alpha} \alpha^j}{j!} \right) \frac{(czy)^{k_j - 1} e^{-czy/2}}{\Gamma(k_j) 2^{k_j}} \cdot \frac{1}{\Gamma(n/2) 2^{n/2}} y^{(n-2)/2} e^{-y/2} dy \\
&= \sum_{j=0}^{\infty} \left(\frac{e^{-\alpha} \alpha^j}{j!} \right) \frac{c^{k_j} z^{k_j - 1}}{\Gamma(k_j) \Gamma(n/2) 2^{k_j + (n/2)}} \int_{0}^{\infty} y^{k_j + (n/2) - 1} e^{-(1+cz)y/2} dy
\end{aligned}$$

Let u=(1+cz)y/2. Then, the integral term can be written as

$$\int_{0}^{\infty} y^{k_j + (n/2) - 1} e^{-(1+cz)y/2} dy = \left(\frac{2}{1+cz} \right)^{\lambda_j} \int_{0}^{\infty} u^{\lambda_j - 1} e^{-u} du = \left(\frac{2}{1+cz} \right)^{\lambda_j} \Gamma(\lambda_j), \qquad \lambda_j = k_j + \frac{n}{2}$$

Substitution of this result into h(z) then gives the desired result:

$$h(z) = \sum_{j=0}^{\infty} \left(\frac{e^{-\alpha} \alpha^j}{j!} \right) \left(\frac{m}{n} \right)^{k_j} \frac{\Gamma(\lambda_j)}{\Gamma(k_j)\Gamma(n/2)} \frac{z^{k_j - 1}}{[1 + (m/n)z]^{\lambda_j}}, \qquad x > 0$$

which is the pdf of a noncentral F with degrees of freedom m and n, and noncentrality parameter δ. q.e.d. ◙

Theorem. Let X~N(μ,1), Y~χ^2(m), and X and Y be independent. Then,

$$Z = \frac{X}{\sqrt{Y/m}} \sim t(m, \mu)$$

i.e., a non-central t with m degrees of freedom and noncentrality parameter μ..

Proof. The pdf of a central χ^2(m) for Y is

$$f_y(y;m) = \frac{1}{\Gamma(m/2)2^{m/2}} y^{(m-2)/2} e^{-y/2}, \quad y \geq 0$$

Let $W = \sqrt{Y/m}$. This transformation has the inverse function $Y = mW^2$ and the Jacobian J=2mW. Therefore, from the previous derivation of the pdf of a square root function, the pdf of W can be written as

$$f_w(w) = |J| f_y(mw^2) = \frac{1}{\Gamma(m/2)2^{m/2}} |2mw| (mw^2)^{(m-2)/2} e^{-mw^2/2} = \frac{m^{m/2}}{\Gamma(m/2)2^{(m-2)/2}} w^{m-1} e^{-mw^2/2}, \qquad w > 0$$

Since Z=X/W, the density of Z is derived by

$$h(z) = \int_{-\infty}^{\infty} |w| f_x(zw) f_w(w) dw = \int_0^{\infty} w \frac{1}{\sqrt{2\pi}} e^{-(zw-\mu)^2/2} \cdot \frac{m^{m/2}}{\Gamma(m/2) 2^{(m-2)/2}} w^{m-1} e^{-mw^2/2} dw$$

$$= \frac{m^{m/2}}{\sqrt{2\pi}\,\Gamma(m/2) 2^{(m-2)/2}} \int_0^{\infty} w^m e^{-[(zw-\mu)^2 + mw^2]/2} dw$$

To find the integral in the last expression, rewrite the exponential function as

$$e^{-[(zw-\mu)^2 + mw^2]/2} = e^{-\mu^2/2} e^{-(m+z^2)w^2/2} e^{\mu zw}$$

Taylor expansion of the last exponential function with respect to w around w=0 gives

$$e^{\mu zw} = \sum_{j=0}^{\infty} \frac{(\mu z)^j w^j}{j!}$$

Therefore,

$$\int_0^{\infty} w^m e^{-[(zw-\mu)^2 + mw^2]/2} dw = \int_0^{\infty} w^m e^{-\mu^2/2} e^{-(m+z^2)w^2/2} \sum_{j=0}^{\infty} \frac{(\mu z)^j w^j}{j!} dw$$

$$= e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu z)^j}{j!} \int_0^{\infty} w^{m+j} e^{-(m+z^2)w^2/2} dw$$

Let $u = (m+z^2)w^2/2$. Then,

$$w^{m+j} dw = \frac{2^{(m+j-1)/2}}{(m+z^2)^{(m+j+1)/2}} u^{(m+j-1)/2} du$$

so that

$$\int_0^{\infty} w^{m+j} e^{-(m+z^2)w^2/2} dw = \frac{2^{(m+j-1)/2}}{(m+z^2)^{(m+j+1)/2}} \int_0^{\infty} u^{(m+j-1)/2} e^{-u} du = \frac{2^{(m+j-1)/2}}{(m+z^2)^{(m+j+1)/2}} \Gamma\left(\frac{m+j+1}{2}\right)$$

The last equality follows from the definition of the Gamma function

$$\Gamma(n) = \int_0^{\infty} e^{-x} x^{n-1} dx$$

Substituting these results into the equation for h(z),

$$h(z) = \frac{m^{m/2}}{\sqrt{2\pi}\,\Gamma(m/2) 2^{(m-2)/2}} \int_0^{\infty} w^m e^{-[(zw-\mu)^2 + mw^2]/2} dw$$

$$= \frac{m^{m/2}}{\sqrt{2\pi}\,\Gamma(m/2) 2^{(m-2)/2}} e^{-\mu^2/2} \sum_{j=0}^{\infty} \frac{(\mu z)^j}{j!} \frac{2^{(m+j-1)/2}}{(m+z^2)^{(m+j+1)/2}} \Gamma\left(\frac{m+j+1}{2}\right)$$

$$= \frac{m^{m/2}}{\sqrt{2\pi}\,\Gamma(m/2) 2^{(m-2)/2}} e^{-\mu^2/2} \frac{2^{(m-1)/2}}{(m+z^2)^{(m+1)/2}} \sum_{j=0}^{\infty} \frac{(\mu z)^j}{j!} \frac{2^{j/2}}{(m+z^2)^{j/2}} \Gamma\left(\frac{m+j+1}{2}\right)$$

$$= \frac{e^{-\mu^2/2}}{\sqrt{\pi m}\,\Gamma(m/2)\,(1+z^2/m)^{(m+1)/2}} \sum_{j=0}^{\infty} \left(\frac{\mu^j}{j!}\right) \left(\frac{2z^2}{m+z^2}\right)^{j/2} \Gamma\left(\frac{m+j+1}{2}\right)$$

which is the pdf of a non-central t-distribution with m degrees of freedom and noncentrality parameter μ. ■

Theorem. Let $X=(X_1,X_2,\cdots,X_n)'$ be a vector of n normal random variables with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\Sigma=(\sigma_{ij})$. Let Y=AX, where A is a nxn nonsingular matrix of constants. Then, Y has a multivariate normal distribution with mean vector $\mathbf{A}\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{A\Sigma A'}$.
Proof. The joint pdf of the random vector X is

$$f(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}$$

where $|\Sigma|$ is the determinant of Σ. Now we use the method of change of variables in finding the pdf of a function of variables. Since $X=A^{-1}Y$, the Jacobian of this linear transformation is

$$|\partial x_i/\partial y_j| = |A^{-1}| = |A|^{-1}$$

Therefore, the joint pdf of random vector Y is

$$\begin{aligned} g(y) &= \frac{1}{(2\pi)^{n/2}|A||\Sigma|^{1/2}}\exp\left\{-\frac{1}{2}(A^{-1}y-\mu)'\Sigma^{-1}(A^{-1}y-\mu)\right\} \\ &= \frac{1}{(2\pi)^{n/2}|A|^{1/2}|\Sigma|^{1/2}|A'|^{1/2}}\exp\left\{-\frac{1}{2}(y-A\mu)'A'^{-1}\Sigma^{-1}A^{-1}(y-A\mu)\right\} \\ &= \frac{1}{(2\pi)^{n/2}|A\Sigma A'|^{1/2}}\exp\left\{-\frac{1}{2}(y-A\mu)'(A\Sigma A')^{-1}(y-A\mu)\right\} \end{aligned}$$

which is a joint pdf of multivariate normal with mean vector $A\mu$ and variance-covariance matrix $A\Sigma A'$.

Corollary. Let $X=(X_1,X_2,\cdots,X_n)'$ be a vector of n normal random variables with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\Sigma=(\sigma_{ij})$. Let Y=BX, where B is a mxn matrix of constants with rank m. Then, Y has a multivariate normal distribution with mean vector $\mathbf{B}\boldsymbol{\mu}$ and variance-covariance matrix $\mathbf{B\Sigma B'}$.
Proof. This follows from the fact that marginal distribution of multivariate normal random variables is normal.

Copula Function

Let X and Y have a joint distribution function F(x,y) with marginals G(X) and H(Y). Let the corresponding densities be denoted by f(x,y), g(x) and h(y), respectively. We wish to find a joint distribution function of transformed random variables U=G(X) and V=H(Y). Note that U and V are standard uniform random variables. Since $X = G^{-1}(U)$ and $Y = H^{-1}(V)$,

$$\frac{\partial X}{\partial U} = \left(\frac{\partial U}{\partial X}\right)^{-1} = \left(\frac{dG(X)}{dX}\right)^{-1} = \frac{1}{g(X)}$$

and similarly for $\partial Y/\partial V = 1/h(Y)$, and $\partial X/\partial V = \partial Y/\partial U = 0$. Therefore, the Jacobian of the transformation is

$$J = \begin{vmatrix} \partial X/\partial U & \partial X/\partial V \\ \partial Y/\partial U & \partial Y/\partial V \end{vmatrix} = \frac{1}{g(X)h(Y)}$$

and the joint density $c(u,v)$ of U and V is given by

$$c(u,v) = f\left(G^{-1}(u), F^{-1}(u)\right)\frac{1}{g\left(G^{-1}(u)\right)h\left(H^{-1}(v)\right)}$$

which can be written as

$$f\left(G^{-1}(u), F^{-1}(u)\right) = g\left(G^{-1}(u)\right)h\left(H^{-1}(v)\right) \times c(u,v)$$

or, in terms of variables x and y as

$$f(x,y) = g(x)\,h(y) \times c\left(G(x), H(y)\right)$$

The density function c(u,v) is called the *copula density* of X and Y.

Alternatively, we can derive the copula function $C(u, v)$ (a cdf of copula density) by applying the cumulative distribution function technique as follows.

$$\begin{aligned} C(u,v) &= P\{U \le u, V \le v\} = P\{G(X) \le u, H(Y) \le v\} = P\{X \le G^{-1}(u), Y \le H^{-1}(v)\} \\ &= F\left(G^{-1}(u), H^{-1}(v)\right) \end{aligned}$$

which can be rewritten as

$$F(x,y) = C\left(G(x), H(y)\right)$$

The joint density can be derived by differentiation

$$f(x,y) = \frac{d^2F(x,y)}{dx\,dy} = \frac{d^2}{dx\,dy}C\left(G(x), H(y)\right) = g(x)\,h(y)\,c(u,v), \qquad c(u,v) = \frac{d^2C(u,v)}{dudv}$$

Note that, if X and Y are independent, then $f(x,y) = g(x)h(y)$, which implies $c(u,v) = 1$. This indicates that the copula function represents the dependence between X and Y. This property is very useful when we wish to construct a joint distribution function from specific marginal distributions. The choice of the copula function will determine the dependence between X and Y.

The copula function is useful to construct a joint distribution given marginal distributions.