

Technical Notes

A.1 BINOMIAL DISTRIBUTION

A coin toss is an example of a *Bernoulli trial*, a random experiment with two possible outcomes. The coin is not necessarily stipulated to be “fair.” The probability of heads can be equal to any $\pi \in [0, 1]$. If we assign the value $Y = 1$ to one of the outcomes of the Bernoulli trial and the value $Y = 0$ to the other, we say that Y follows a *Bernoulli distribution* with parameter π .

Suppose we repeat a Bernoulli trial n times and add up the resulting values of $Y_i, i = 1, \dots, n$. Successive trials are independent. The random variable $X = \sum_{i=1}^n Y_i$ is said to follow a *binomial distribution* with parameters π and n . We have

$$\begin{aligned} E[X] &= n\pi \\ \text{Var}[X] &= n\pi(1 - \pi) \end{aligned}$$

The Bernoulli and binomial distributions are both discrete distributions. But the binomial distribution converges to the normal distribution as the number of trials n grows larger. This convergence result is an application of the *central limit theorem*. Specifically, if we standardize a binomially distributed random variable X , we can get its probability distribution arbitrarily close to that of a standard normal variate by increasing n enough:

$$P \left[\frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \leq z \right] \rightarrow \Phi(z) \quad \text{as } n \rightarrow \infty$$

This book uses the binomial distribution in two applications, to analyze the probability distribution of the terminal state of a random walk (Chapter 2) and to characterize the distribution of credit losses in a portfolio of identical, uncorrelated credits (Chapter 8). In the case of the random walk,

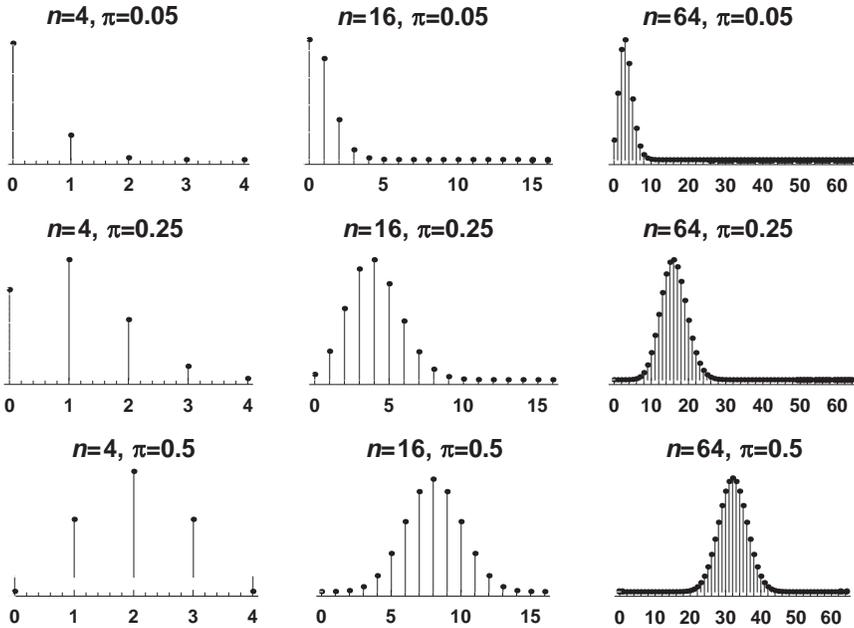


FIGURE A.1 Convergence of Binomial to Normal Distribution

The plots display the probability density of X_n , the terminal position of a random walk over an interval of length 1, with $n = 4, 16, 64$ time steps. The upper two rows of graphs set $\pi = 0.05$ and $\pi = \frac{1}{4}$. The distributions are not symmetrical, but converge to a symmetrical normal distribution. The lower row of graphs uses $\pi = \frac{1}{2}$, and all three distributions in the row are symmetrical.

we use convergence to the normal distribution to show that the discrete-step random walk converges to a continuous Brownian motion. Figure A.1 illustrates. Notice that if the probability of the Bernoulli trial is not equal to 0.5, the binomial distribution is skewed for small n . But as n grows larger, the skewness disappears. In the credit context, convergence is used to derive an approximation to the distribution of losses in a credit portfolio.

A.2 QUANTILES AND QUANTILE TRANSFORMATIONS

Suppose we have a random variable X . A p -th *quantile* of X is a number $Q_p(X)$ such that $\mathbf{P}[X < Q_p(X)] \leq p$. We can define quantiles in terms of the cumulative distribution function $F(X)$. The p -th quantile is the value of

X such that, for any $x < Q_p(X)$,

$$F(x) \leq p$$

Every probability distribution has an *inverse probability distribution function* or *quantile function*. If X is a continuous random variable, with a cumulative distribution function (CDF) that is monotone increasing in X , we can define the quantile more simply. The p -th quantile is then a number $Q_p(X)$ such that $\mathbf{P}[X < Q_p(X)] = p$. It is equal to the value of the inverse function of the CDF, evaluated at p :

$$Q_p(X) = F^{-1}(p)$$

The most commonly mentioned quantile is the 0.5 quantile, or the *median*. If X is a random variable, there is a probability of $\frac{1}{2}$ of realizing a value of X less than or equal to the median. The “ p ” in “ p -th quantile” is commonly expressed as a percent, and the quantile is then called a *percentile*. A familiar example comes from the standard normal distribution. The median of the standard normal is zero, equal to its mean. The 0.01 quantile or first percentile of the standard normal is about 2.33.

We use quantile functions in a number of applications, including simulation, in applying copulas (Chapters 8 and 9), and in understanding risk neutral probability distributions (Chapter 10). The probability distribution tells us, for any value x of a random variable X , what the probability is of X having a realization that is less than or equal to x , that is, $\mathbf{P}[X \leq x]$. The quantile function tells us, for any $p \in [0, 1]$, what is the value of x such that $p = \mathbf{P}[X \leq x]$.

The most common approach to obtaining random variates with a desired probability distribution applies the *quantile transformation* or *inversion principle*. This principle exploits the fact that the cumulative probabilities $F(X)$ of a random variable X are uniform-[0, 1] distributed.

By the same token, values of the quantile function, applied to the uniform-[0, 1] variates, have the original distribution. Both the domain of a uniform random variable and the probabilities lie on [0, 1]; the domain of the uniform random variable happens to be the range of any cumulative distribution function. The inversion principle states: If we have a uniform distribution value, we can transform it into the value of another distribution $F(X)$ by finding the value for which that other distribution has a probability equal to the uniform distribution value. If U represents a uniform-[0, 1] variate, the random variable $F^{-1}(U)$ has the same distribution function as X , namely $F(X)$.

A.3 NORMAL AND LOGNORMAL DISTRIBUTIONS

A.3.1 Relationship between Asset Price Levels and Returns

The *lognormal distribution* is defined as the probability distribution of a random variable y with a logarithm that is normally distributed. If $\log(y) \sim N(a, b)$, then y has a lognormal distribution with parameters a and b . Those parameters, though, are not the mean and standard deviation of y .

We use the lognormal distribution to relate changes in the level of an asset price to its logarithmic returns. We want to find the probability distributions, at some future time $t + \tau$, of

- the asset price level $S_{t+\tau}$
- the change in price $S_t - S_{t+\tau}$
- the logarithmic return $\log\left(\frac{S_{t+\tau}}{S_t}\right)$

Typically, we have a reasonable assumption about the mean of the future price $E[S_{t+\tau}]$, based on the forward or futures price, or on a price forecast, and a reasonable assumption about the annualized volatility σ of the logarithmic return, based on historical or implied volatility.

How do we match these distributions and parameters? The expected discrete rate of return that “grows” $\log(S_t)$ to its time $t + \tau$ mean $E[S_{t+\tau}]$ in a straight line is the μ that satisfies

$$E[S_{t+\tau}] = S_t e^{\mu\tau}$$

But there is noise along the path; S_t has a volatility. The expected constant logarithmic rate of return that gets S_t to $E[S_{t+\tau}]$ is *not* μ , but $\mu - \frac{\sigma^2}{2}$:

$$\frac{1}{\tau} \log\left(\frac{E[S_{t+\tau}]}{S_t}\right) = \mu - \frac{\sigma^2}{2}$$

Reducing the rate of return by $\frac{\sigma^2}{2}$ just offsets the asymmetric, growth-increasing impact of volatility combined with compounding described in Chapter 2. The logarithmic rate of return is therefore distributed as

$$\log\left(\frac{S_{t+\tau}}{S_t}\right) \sim N\left[\left(\mu - \frac{\sigma^2}{2}\right)\tau, \sigma\sqrt{\tau}\right]$$

implying that

$$\frac{\log\left(\frac{S_{t+\tau}}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \sim N(0, 1)$$

By the definition of a lognormal random variable, we thereby establish that the distribution of $S_{t+\tau} - S_t$ is lognormal with parameters $\left(\mu - \frac{\sigma^2}{2}\right)\tau$ and $\sigma\sqrt{\tau}$. Shifting the mean of the logarithmic return by $\log(S_t)$, the distribution of $S_{t+\tau}$ is lognormal with parameters $\log(S_t) + \left(\mu - \frac{\sigma^2}{2}\right)\tau$ and $\sigma\sqrt{\tau}$.

In general, if y has a lognormal distribution with parameters a and b , then $E[y] = e^{a + \frac{b^2}{2}}$. The mean of the distribution of $S_{t+\tau} - S_t$ is therefore $e^{\mu\tau}$ and that of $S_{t+\tau}$ is $S_t e^{\mu\tau}$.

We can also match up the *quantiles* of these distributions. The p -th quantile of the lognormal distribution equals S_t times the exponential of the p -th quantile of the corresponding normal distribution, or

$$S_t e^{\left(\mu - \frac{\sigma^2}{2}\right)\tau + \sigma\sqrt{\tau}z} \quad \text{with} \quad \Phi(z) = p$$

The cumulative probability distribution function of $\log\left(\frac{S_{t+\tau}}{S_t}\right)$ is

$$\Phi\left[\frac{\log\left(\frac{S_{t+\tau}}{S_t}\right) - \left(\mu - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}}\right]$$

A.3.2 The Black-Scholes Distribution Function

We can use these relationships to interpret the Black-Scholes delta and other sensitivities under the risk-neutral probability distribution. In the Black-Scholes model, the future asset price follows a lognormal distribution, and the logarithmic return $\log\left(\frac{S_{t+\tau}}{S_t}\right)$ follows a normal distribution. What are the parameters of those risk-neutral distributions?

Suppose we have an asset with a current price of S_t , paying a risk-free, constant, and continuous dividend at a rate q . The constant and continuously compounded risk-free rate is r . The forward price $F_{t,\tau}$ is equal to the mean of the future price under the risk-neutral distribution:

$$\tilde{E}[S_{t+\tau}] = S_t e^{(r-q)\tau} = F_{t,\tau}$$

If σ is the implied volatility, $r - q - \frac{\sigma^2}{2}$ is the risk-neutral mean of the logarithmic return and $\sigma\sqrt{\tau}$ its risk-neutral standard deviation. This sets the parameters of the risk-neutral time $t + \tau$ asset price and return distributions, which we can express as

$$\frac{\log\left(\frac{S_{t+\tau}}{S_t}\right) - \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \sim N(0, 1)$$

With the derivations just above, we can summarize as follows:

- The risk-neutral probability distribution of the future asset price level $S_{t+\tau}$ is lognormal, with parameters $\log(S_t) + \left(r - q - \frac{\sigma^2}{2}\right)\tau$ and $\sigma\sqrt{\tau}$.
- The risk-neutral probability distribution of the change in price $S_t - S_{t+\tau}$ is lognormal, with parameters $\left(r - q - \frac{\sigma^2}{2}\right)\tau$ and $\sigma\sqrt{\tau}$.
- The risk-neutral probability distribution of the logarithmic return $\log\left(\frac{S_{t+\tau}}{S_t}\right)$ is normal, with mean $\left(r - q - \frac{\sigma^2}{2}\right)\tau$ and standard deviation $\sigma\sqrt{\tau}$.

How do these distributions relate to the Black-Scholes formulas? The Black-Scholes model values for European puts and calls are:

$$\begin{aligned} v(S_t, \sigma, r, q) &= S_t e^{-q\tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] \\ &\quad - X e^{-r\tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] \\ w(S_t, \tau, X, \sigma, r, q) &= X e^{-r\tau} \Phi \left[-\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] \\ &\quad - S_t e^{-q\tau} \Phi \left[-\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] \end{aligned}$$

The Black-Scholes call delta is

$$\delta_{c,t} \equiv \frac{\partial}{\partial S_t} v(S_t, \tau, X, \sigma, r, q) = e^{-q\tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q + \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right]$$

The derivative of the Black-Scholes formula for the value of a European call with respect to the exercise price is less commonly used than the delta. We'll denote this "exercise price delta" by $\chi_{c,t}$:

$$\chi_{c,t} \equiv \frac{\partial}{\partial X} v(S_t, \tau, X, \sigma, r, q) = -e^{-r\tau} \Phi \left[\frac{\log\left(\frac{S_t}{X}\right) + \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right]$$

Multiplying the value at which the normal CDF is evaluated by -1 , we have

$$e^{r\tau} \chi_{c,t} = - \left\{ 1 - \Phi \left[\frac{\log\left(\frac{X}{S_t}\right) - \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right] \right\}$$

We showed in Chapter 10 that $\chi_{c,t}$ is related to the risk-neutral CDF by

$$\tilde{\Pi}(X) = 1 + e^{r\tau} \chi_{c,t}$$

Now we can see that in the specific case of the Black-Scholes model, the risk-neutral distribution of $S_{t+\tau}$ is lognormal, and that $\log\left(\frac{S_{t+\tau}}{S_t}\right)$ is a normally distributed random variable with CDF

$$\tilde{\Pi}(X) = \Phi \left[\frac{\log\left(\frac{X}{S_t}\right) - \left(r - q - \frac{\sigma^2}{2}\right)\tau}{\sigma\sqrt{\tau}} \right]$$

Example A.1 Let's take the dollar-euro exchange rate as an example. Suppose the current spot rate is 1.25 per euro, that the dollar and euro one-year funding rates are both 1 percent, and that the one-year implied volatility

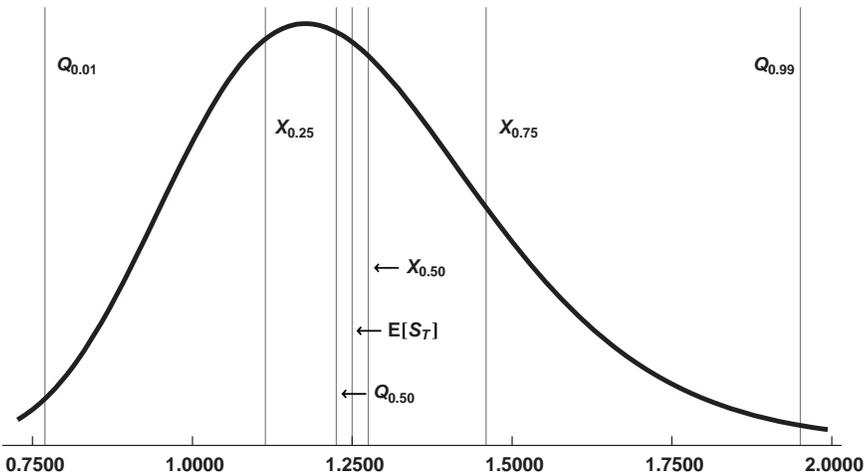


FIGURE A.2 The Black-Scholes Probability Density Function

The market-based parameters are as given in Example A.1. The grid lines are

$Q_{0.01}$	0.01 quantile of $S_{t+\tau}$	0.769414
$X_{0.25}$	Exercise price of 25- δ call	1.45942
$Q_{0.50}$	Median or 0.50 quantile of $S_{t+\tau}$	1.22525
$\tilde{E}[S_{t+\tau}]$	Expected value of $S_{t+\tau}$	1.25
$X_{0.50}$	Exercise price of 50- δ call, or expected value of $S_{t+\tau}$ conditional on exceeding 0.50 quantile	1.27525
$X_{0.75}$	Exercise price of 75- δ call	1.11432
$Q_{0.99}$	0.99 quantile of $S_{t+\tau}$	1.95114

is 20 percent. What is the risk-neutral distribution of the exchange rate in one year?

The one-year forward rate and the risk-neutral mean are both 1.25. The quantiles of the distribution are

Probability	Quantile
0.001	0.6604
0.01	0.7694
0.5	1.2253
0.99	1.9511
0.999	2.2732

Figure A.2 plots the exchange rate distribution in one year. Note that the median is below the mean. This is typical for continuous unimodal (“one-humped”) probability distributions. We’ve assumed a lognormal distribution, which is skewed to the right, so the mean is to the right of the median.

A.4 HYPOTHESIS TESTING

In the standard approach to statistical hypothesis testing, we start with a statement about the parameters of a statistical distribution, called the *null hypothesis*, denoted \mathfrak{H}_0 . The null hypothesis is then tested against the *alternative hypothesis*, which can be the statement that the null hypothesis is not true, or a more specific hypothesis. The null or the alternative hypothesis is called a *simple* or *point hypothesis* if it states that the parameter is equal to a particular value.

For example, we might stipulate that the physical heights of a human population are normally distributed with a standard deviation of 5 inches, and formulate the simple null hypothesis that the mean height is 72 inches. The next step is then to devise a test procedure for determining whether we will accept the null hypothesis. Typically, using a sample of data from the population, we can calculate a *test statistic*. The test procedure is framed so that, if the null hypothesis is true, we can determine the probability distribution of the test statistic. Given the distribution of the test statistic, we can determine the probability with which the test statistic takes on any particular value. If the sample leads to a value of the test statistic that is highly unlikely, we will be inclined to reject the null hypothesis.

Continuing our physical height example, we can calculate a sample mean $N^{-1} \sum_{n=1}^N b_n$ by measuring the heights b_n of N randomly chosen individuals. Since we have stipulated that the population's distributional family is normal and its standard deviation is $\sigma = 5$ inches, the test statistic

$$\frac{\sqrt{N} N^{-1} \sum_{n=1}^N (b_n - \mu)}{\sigma}$$

where μ is the hypothesized population mean, is a standard normal variate. Note that the distributional family and standard deviation are *maintained hypotheses*, which we are *not* testing. The maintained hypotheses drive the probability distribution of the test statistic, while the data drive the value it takes on in our particular case. The z -test, which uses a standard normal as a test statistic, is appropriate here; if we used the sample rather than population standard deviation, and had a relatively small sample, we would prefer to use a t -test. The null hypothesis is

$$\mathfrak{H}_0 : \mu = 72$$

We'll test the null against the alternative hypothesis that the mean is

$$\mathfrak{H}_1 : \mu \neq 72$$

Suppose our sample consists of $N = 10$ individuals, with a sample mean of 70 inches. The value of the test statistic is then

$$\frac{\sqrt{10}(70 - 72)}{5} = -1.265$$

To complete the test procedure, we need to establish criteria for accepting or rejecting the null hypothesis. This is not a matter of mathematics or science, but rather of the relative penalty we want to place on two types of error, mistakenly accepting the null even though it is false, or mistakenly rejecting the null even though it is true, known respectively as *Type I* and *Type II errors*. There is an unavoidable trade-off: If we reduce the probability of a Type I error, we increase the probability of a Type II error, and vice versa.

The test criteria can be expressed by dividing the range of possible values of the test statistic we are testing into two regions. If the test statistic falls in the *critical region*, we reject the null hypothesis, and if the test statistic falls in the *acceptance region*, we do not reject the null hypothesis. In our example, we are carrying out a *two-sided test*, since the alternative hypothesis is an inequality. At a conventional significance level of, say, 5 percent, we would reject the null if the test statistic fell outside the range $(-1.96, +1.96)$. Our test statistic value of -1.265 is comfortably within the range, so we do not reject the null.

A.5 MONTE CARLO SIMULATION

Once we have a probability distribution, we could in principle use it to describe random outcomes algebraically. This is not, however, the way it is actually done in practice. More often than not, quantitative results are derived by simulation.

Why? Often, the random variables involved, say, asset returns or default times, are complicated functions of the underlying random drivers that we are confident we can model. The algebra needed to describe the distribution we are interested in may be too complicated.

Simulation is in a way a brute-force approach that cuts through the algebra and goes straight from what we think we know about the underlying drivers to the distribution of the random outcomes. It is often easier or more accurate. In this appendix, we outline the basics of how this is done. In Chapters 3 through 6, 8, and 9, we apply simulation to concrete market and credit risk measurement problems.

Monte Carlo simulation is one approach to estimating distributions. It comprises three stages:

1. We *generate random variates* according to the distributional hypotheses that we believe govern the terminal value of the portfolio.
2. The *model* transforms the raw random data in some way. In market risk applications, we may have a pricing model. In credit risk applications, we may have a model that values the outcomes of credit events.
3. The *results* of the simulation are a set of Monte Carlo *realizations* (also called *replications* or *threads*). The realizations have a distribution that can be described by its sample statistics, such as mean and variance. In risk measurement, we are usually interested in low quantiles such as the first percentile or the 0.01 quantile.

A.5.1 Fooled by Nonrandomness: Random Variable Generation

The first step in generating random variates is to generate uniformly distributed random numbers, that is, equiprobable real numbers on $[0, 1]$. In a later step, we transform these into random variates that follow other distributions.

There are two ways to generate uniform randoms. The first is to use a physical process, such as radioactive decay or flipping a coin. It is unusual to use physical techniques, because we often need many random numbers and it is both expensive and time-consuming to accurately generate and record physically generated ones.

The second, more typical, approach is to use *pseudo random numbers*. As the name indicates, pseudo random numbers are not truly random, the way the results of physical random processes can be. Rather, they are deterministic, since they are generated by algorithms. The main advantage of pseudo random numbers is that, since computing power is now so cheap, they are cheap to produce.

An example of a pseudo random number generator is

$$I_{j+1} = \frac{1}{2^{31} - 1} [7^5 I_j \bmod (2^{31} - 1)]$$

where $x \bmod (y)$ denotes the remainder obtained on dividing x by y . For example, $3 \bmod (3) = 0$ and $4 \bmod (3) = 1$. This type of pseudo random number generator is called a *linear congruential generator*.

The initial value I_0 is provided by the user (or automatically by the computer or the high-level application, for example, by setting it equal to the computer's internal clock time) and is called the *random seed*. Because the pseudo random number generator is deterministic, for a given seed, a sequence of N random numbers will always be identical. This is another

advantage of pseudo random numbers: The results can always be replicated exactly, realization by realization, by reusing the seed.

Pseudo random number generators can be evaluated by testing their outputs for whether they are close to truly random: for example, to see if there is repetition or if the realizations are serially correlated.

A.5.2 Generating Nonuniform Random Variates

Once we have a set of uniform random numbers, we turn them into random numbers drawn from the distribution we are really interested in. The inversion or transformation principle cited above tells us that the range of a distribution function follows a uniform distribution. Therefore, if we have a uniform distribution value, we can transform it into the value of another distribution by finding its quantile, that is, the value for which that other distribution has a probability equal to the uniform distribution value. In this way, once we have generated a set of uniform-[0, 1] variates, we can use the results to generate a random sample from any other distribution.

Figure A.3 illustrates this approach for the standard normal distribution. It plots the quantile function $z = \Phi^{-1}(u)$, which maps real numbers on [0, 1] to a real number on $(-\infty, \infty)$ such that $u = \Phi(z)$. The graph also displays 200 pseudo random uniform-[0, 1] variates. As the number of simulations grows, the simulations trace out the target, normal, distribution as precisely as desired.

A.6 HOMOGENEOUS FUNCTIONS

A function $f(x_1, \dots, x_N)$ is *homogeneous of degree p* if $\forall t \geq 0$,

$$f(tx_1, \dots, tx_N) = t^p f(x_1, \dots, x_N) \quad (\text{A.1})$$

The partial derivatives of functions homogeneous of degree p are homogeneous of degree $p - 1$:

$$\frac{\partial f(x_1, \dots, x_N)}{\partial x_n} = t^{p-1} \frac{\partial f(x_1, \dots, x_N)}{\partial x_n}, \quad n = 1, \dots, N$$

To see this, differentiate (A.1) w.r.t. any of the x_n :

$$\frac{\partial f(tx_1, \dots, tx_N)}{\partial x_n} = t \frac{\partial f(tx_1, \dots, tx_N)}{\partial x_n} = t^p \frac{\partial f(x_1, \dots, x_N)}{\partial x_n} \quad n = 1, \dots, N$$

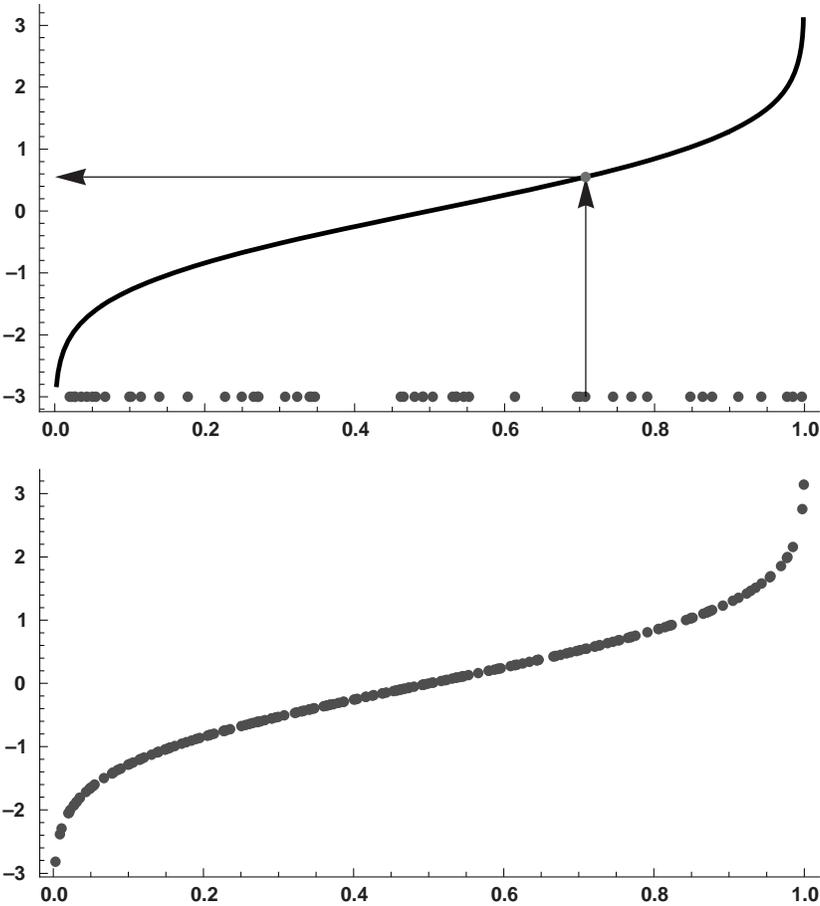


FIGURE A.3 Transforming Uniform into Normal Variates
 The plot in the upper panel represents the inverse standard normal distribution. The points represent 50 simulated values of the uniform distribution. The lower panel plots the results of generating 200 standard normal variates. Each point has coordinates $u_i, \Phi^{-1}(u_i), i = i, \dots, 200$.

The first equality differentiates the left-hand side of Equation (A.1) using the chain rule, while the expression following the second equality is the derivative of the right-hand side. Dividing both sides by t gives the result.

Homogeneous functions of degree p have the *Euler property*:

$$\sum_{n=1}^N x_n \frac{\partial f(x_1, \dots, x_N)}{\partial x_n} = pf(x_1, \dots, x_N)$$

For example, the weighted sum of the partial first derivatives of a linearly homogeneous function, with the weights set equal to the value of the function's arguments, is equal to the value of the function itself.

To prove this, differentiate (A.1) w.r.t. t . Using the chain rule to differentiate w.r.t. the tx_n , we have

$$\frac{\partial f(tx_1, \dots, tx_N)}{\partial t} = \frac{\partial f(tx_1, \dots, tx_N)}{\partial tx_n} \frac{\partial (tx_n)}{\partial t} = x_n \frac{\partial f(tx_1, \dots, tx_N)}{\partial x_n} \quad n = 1, \dots, N$$

Therefore, differentiating both sides of (A.1) w.r.t. t gives

$$\sum_{n=1}^N x_n \frac{\partial f(tx_1, \dots, tx_N)}{\partial x_n} = pt^{p-1} f(x_1, \dots, x_N)$$

Since this is true for all t , it is also true for $t = 1$, giving the result.

FURTHER READING

Most of the material in this Appendix is reviewed in intermediate textbooks on probability and statistics. Pfeiffer (1990) and Wasserman (2004) stand out for clarity of presentation. The lognormal distribution is covered in option textbooks such as Hull (2000). The explication in Jarrow and Rudd (1983) is particularly lucid. Numerical techniques and simulation are covered in Stoer and Bulirsch (1993).