Financial Econometrics

UNC CHARLOTTE

BELK COLLEGE *of* BUSINESS

Lecture 1

Review of Statistics

1.1 Random Variables and Distributions

1.1.1 Distributions

A univariate distribution of a random variable x describes the probability of different values. If $f(x)$ is the probability density function, then the probability that x is between A and B is calculated as the area under the density function from A to B

$$\Pr(A \leq x < B) = \int_A^B f(x)dx. \tag{1.1}$$

See Figure 1.1 for illustrations of normal (gaussian) distributions.

Remark 1.1 *If $x \sim N(\mu, \sigma^2)$, then the probability density function is*

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

This is a bell-shaped curve centered on the mean μ and where the standard deviation σ determines the "width" of the curve.

Some Plots of Normal PDFs

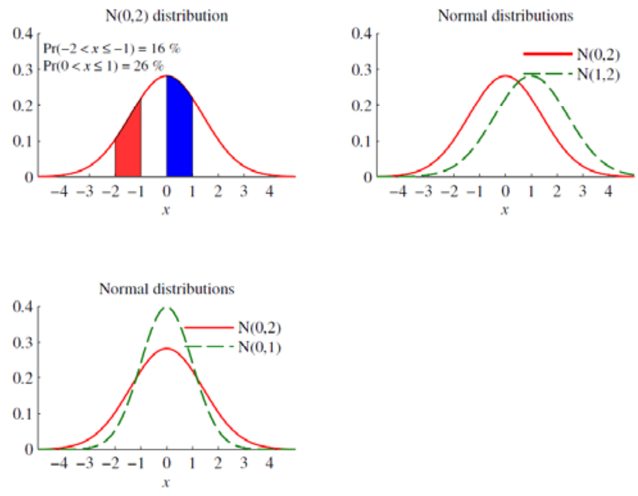


Figure 1.1: A few different normal distributions

Sampling from a Normal Distribution

- MATLAB allows us to create samples of pseudo random numbers from a wide range of distributions
- Simply use the random function
  - Inputs are the distribution, parameters of the distribution, and dimensions of the output vector (or matrix)
- Example
  - Normal distribution with mean 10 and standard deviation 5

```
rng(0);
y = random('Normal',10,5,1,T);
```

  - The first line sets the seed to zero
  - The second line generates a row vector of T variables

Bivariate Distributions

A bivariate distribution of the random variables x and y contains the same information as the two respective univariate distributions, but also information on how x and y are related. Let $h(x, y)$ be the joint density function, then the probability that x is between A and B and y is between C and D is calculated as the volume under the surface of the density function

$$\Pr(A \le x < B \text{ and } C \le x < D) = \int_A^B \int_C^D h(x, y) dx dy. \tag{1.2}$$

A joint normal distributions is completely described by the means and the covariance matrix

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}\right), \tag{1.3}$$

where μ_x and μ_y denote means of x and y, σ_x^2 and σ_y^2 denote the variances of x and y and σ_{xy} denotes their covariance. Some alternative notations are used: $\mathrm{E}\,x$ for the mean, $\mathrm{Std}(x)$ for the standard deviation, $\mathrm{Var}(x)$ for the variance and $\mathrm{Cov}(x, y)$ for the covariance.

Covariance, Correlation, and Independence

Clearly, if the covariance σ_{xy} is zero, then the variables are (linearly) unrelated to each other. Otherwise, information about x can help us to make a better guess of y. See Figure 1.2 for an example. The correlation of x and y is defined as

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \tag{1.4}$$

If two random variables happen to be independent of each other, then the joint density function is just the product of the two univariate densities (here denoted $f(x)$ and $k(y)$)

$$h(x, y) = f(x)\, k(y) \text{ if } x \text{ and } y \text{ are independent.} \tag{1.5}$$

This is useful in many cases, for instance, when we construct likelihood functions for maximum likelihood estimation.

Some Plots of Bivariate Normal PDFs

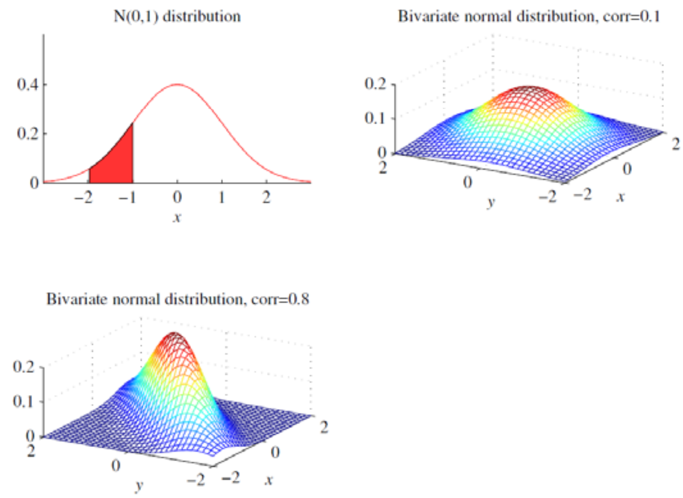
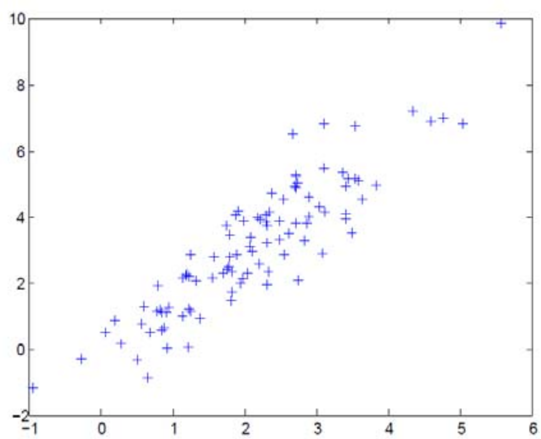


Figure 1.2: Density functions of univariate and bivariate normal distributions

Sampling from a Bivariate Normal Distribution

- Use the mvnrnd function
  - Inputs are the mean vector, covariance matrix, and number of rows in the output matrix
- Example

```
mu = [2 3];
SIGMA = [1 1.5; 1.5 3];
r = mvnrnd(mu,SIGMA,100);
plot(r(:,1),r(:,2),'+')
```



Conditional Distributions

If $h(x, y)$ is the joint density function and $f(x)$ the (marginal) density function of x, then the conditional density function is

$$g(y|x) = h(x, y)/f(x). \tag{1.6}$$

For the bivariate normal distribution (1.3) we have the distribution of y conditional on a given value of x as

$$y|x \sim N\left[\mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x), \sigma_y^2 - \frac{\sigma_{xy}\sigma_{xy}}{\sigma_x^2}\right]. \tag{1.7}$$

Notice that the conditional mean can be interpreted as the best guess of y given that we know x. Similarly, the conditional variance can be interpreted as the variance of the forecast error (using the conditional mean as the forecast). The conditional and marginal distribution coincide if y is uncorrelated with x. (This follows directly from combining (1.5) and (1.6)). Otherwise, the mean of the conditional distribution depends on x, and the variance is smaller than in the marginal distribution (we have more information). See Figure 1.3 for an illustration.

Some Plots of Conditional PDFs

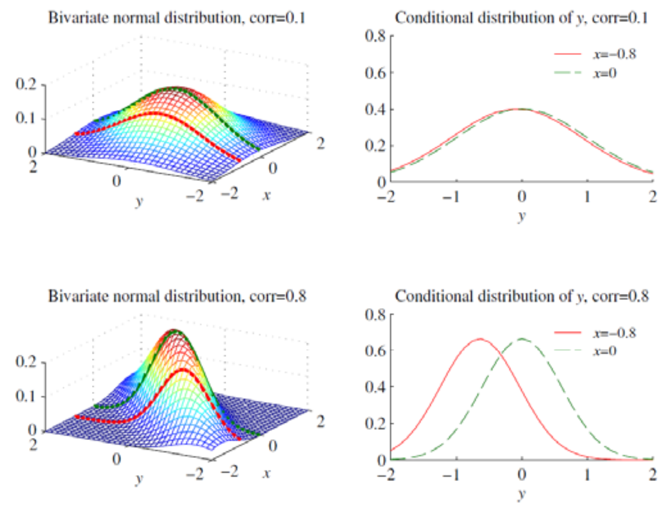


Figure 1.3: Density functions of normal distributions

Depicting Distributions Using Histograms

If we know that type of distribution (uniform, normal, etc) a variable has, then the best way of illustrating the distribution is to estimate its parameters (mean, variance and whatever more—see below) and then draw the density function.

In case we are not sure about which distribution to use, the first step is typically to draw a histogram: it shows the relative frequencies for different bins (intervals). For instance, it could show the relative frequencies of a variable x_t being in each of the follow intervals: -0.5 to 0, 0 to 0.5 and 0.5 to 1.0. Clearly, the relative frequencies should sum to unity (or 100%), but they are sometimes normalized so the area under the histogram has an area of unity (as a distribution has).

See Figure 1.4 for an illustration.

Histograms for Monthly Stock Returns

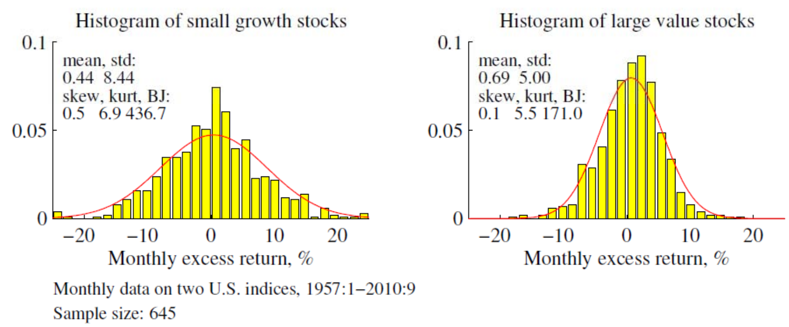


Figure 1.4: Histogram of returns, the curve is a normal distribution with the same mean and standard deviation as the return series

From Histograms to Kernel Densities

- Histograms approximate the PDFs of the returns
  - Constructed using the default number of bins
  - Could we approximate the PDFs better by using more bins?
    - No, as the number of bins increases, the variability of the height of each bin also increases, resulting in a jagged appearance.
    - To solve this problem we need to smooth across bins. This can be accomplished by using a kernel density smoother.

- Kernel density
  - In this context, the word "kernel" essentially refers to type of filter.
  - There are many different kernels that could be used
  - The "Gaussian" kernel typically works well for unimodal PDF's

Gaussian Kernel Density

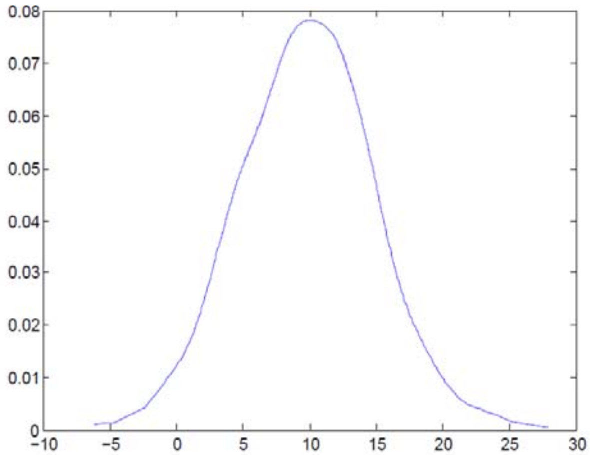
- The following function constructs the kernel density estimate

```
function f = kernel (x)     % kernel(x) returns the length(x)-point Gaussian
                            % kernel density estimate (h is the bandwidth)
n=length(x);
h=1.06*std(x)*n^(-1/5);
x=sort(x);
for i=1:n
  f(i)=1/(n*h*sqrt(2*pi))*sum(exp(-.5.*((x-x(i))./h).^2));
end
```

  - The plot is a "Gaussian" kernel density estimate because the density f used in the code is the Gaussian PDF.
  - The smoothing parameter h (the bandwidth) was set equal to a value that tends to work well in practice.

Example of Kernel Density Estimate

```
rng(0);
y = random('Normal',10,5,1,1000);
f=kernel(y);
plot(x,f);
```



Confidence Bands

Confidence bands are typically only used for symmetric distributions. For instance, a 90% confidence band is constructed by finding a critical value c such that

$$\Pr(\mu - c \leq x < \mu + c) = 0.9. \tag{1.8}$$

Replace 0.9 by 0.95 to get a 95% confidence band—and similarly for other levels. In particular, if $x \sim N(\mu, \sigma^2)$, then

$$\Pr(\mu - 1.65\sigma \leq x < \mu + 1.65\sigma) = 0.9 \text{ and}$$
$$\Pr(\mu - 1.96\sigma \leq x < \mu + 1.96\sigma) = 0.95. \tag{1.9}$$

As an example, suppose x is not a data series but a regression coefficient (denoted $\hat{\beta}$)—and we know that the standard error equals some number σ. We could then construct a 90% *confidence band around the point estimate* as

$$[\hat{\beta} - 1.65\sigma, \hat{\beta} + 1.65\sigma]. \tag{1.10}$$

In case this band does not include zero, then we would be 90% that the (true) regression coefficient is different from zero.

Confidence Bands and t-Tests

Alternatively, suppose we instead construct the 90% confidence band around zero as

$$[0 - 1.65\sigma, 0 + 1.65\sigma]. \tag{1.11}$$

If this band does not include the point estimate ($\hat{\beta}$), then we are also 90% sure that the (true) regression coefficient is different from zero. This latter approach is virtually the same as doing a t-test, that, by checking if

$$\left|\frac{\hat{\beta} - 0}{\sigma}\right| > 1.65. \tag{1.12}$$

To see that, notice that if (1.12) holds, then

$$\hat{\beta} < -1.65\sigma \text{ or } \hat{\beta} > 1.65\sigma, \tag{1.13}$$

which is the same as $\hat{\beta}$ being outside the confidence band in (1.11).

1.2 Sample Moments

1.2.1 Mean and Standard Deviation

The mean and variance of a series are estimated as

$$\bar{x} = \sum_{t=1}^{T} x_t / T \text{ and } \hat{\sigma}^2 = \sum_{t=1}^{T} (x_t - \bar{x})^2 / T. \tag{1.14}$$

The standard deviation (here denoted $\mathrm{Std}(x_t)$), the square root of the variance, is the most common measure of volatility. (Sometimes we use $T-1$ in the denominator of the sample variance instead T.) See Figure 1.4 for an illustration.

A sample mean is normally distributed if x_t is normal distributed, $x_t \sim N(\mu, \sigma^2)$. The basic reason is that a linear combination of normally distributed variables is also normally distributed. However, a sample average is typically approximately normally distributed even if the variable is not (discussed below). If x_t is iid (independently and identically distributed), then the variance of a sample mean is

$$\mathrm{Var}(\bar{x}) = \sigma^2 / T, \text{ if } x_t \text{ is iid.} \tag{1.15}$$

Properties of the Sample Mean

A sample average is (typically) *unbiased*, that is, the expected value of the sample average equals the population mean, that is,

$$\mathrm{E}\,\bar{x} = \mathrm{E}\,x_t = \mu. \tag{1.16}$$

Since sample averages are typically normally distributed in large samples (according to the central limit theorem), we thus have

$$\bar{x} \sim N(\mu, \sigma^2/T), \tag{1.17}$$

so we can construct a *t-stat* as

$$t = \frac{\bar{x} - \mu}{\sigma/\sqrt{T}}, \tag{1.18}$$

which has an $N(0,1)$ distribution.

Higher-Order Sample Moments

1.2.2 Skewness and Kurtosis

The skewness, kurtosis and Bera-Jarque test for normality are useful diagnostic tools. They are

$$\begin{array}{lcll} & & \underline{\text{Test statistic}} & \underline{\text{Distribution}} \\ \text{skewness} & = & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t-\mu}{\sigma}\right)^3 & N\left(0, 6/T\right) \\ \text{kurtosis} & = & \frac{1}{T}\sum_{t=1}^{T}\left(\frac{x_t-\mu}{\sigma}\right)^4 & N\left(3, 24/T\right) \\ \text{Bera-Jarque} & = & \frac{T}{6}\text{skewness}^2 + \frac{T}{24}\left(\text{kurtosis}-3\right)^2 & \chi^2_2. \end{array} \tag{1.19}$$

This is implemented by using the estimated mean and standard deviation. The distributions stated on the right hand side of (1.19) are under the null hypothesis that x_t is iid $N(\mu, \sigma^2)$. The "excess kurtosis" is defined as the kurtosis minus 3. The test statistic for the normality test (Bera-Jarque) can be compared with 4.6 or 6.0, which are the 10% and 5% critical values of a χ^2_2 distribution.

Implementing Tests for Normality

- Option 1: Use the built-in jbtest function
- Option 2: Program your own function

```
function [chiSqStat, pVal] = bjTest (x)
  n=length(x);
  chiSqStat=(n/6)*skewness(x)^2+(n/24)*(kurtosis(x)-3)^2;
  pVal=1-chi2cdf(chiSqStat,2);
end

>> rng(123)
>> y = random('Normal',10,5,1,1000);
>> [stat,pval]=bjTest(y)

stat =
   1.4735
pval =
   0.4787
```

Joint Sample Moments

1.2.3 Covariance and Correlation

The covariance of two variables (here x and y) is typically estimated as

$$\hat{\sigma}_{xy} = \sum_{t=1}^{T} (x_t - \bar{x})(y_t - \bar{y})/T. \tag{1.21}$$

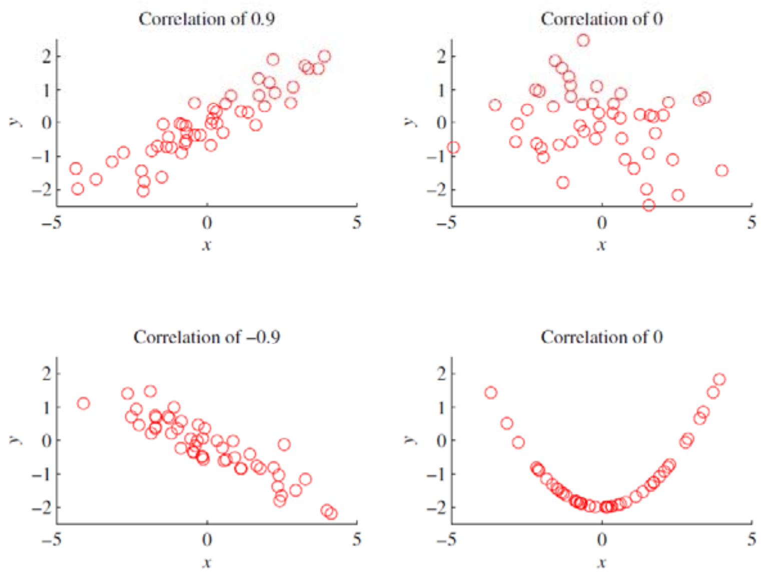
(Sometimes we use $T-1$ in the denominator of the sample covariance instead of T.)

The correlation of two variables is then estimated as

$$\hat{\rho}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y}, \tag{1.22}$$

where $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are the estimated standard deviations. A correlation must be between -1 and 1. Note that covariance and correlation measure the degree of *linear* relation only. This is illustrated in Figure 1.5.

Correlation Measures Linear Dependence



Distributions Commonly Used in Tests

1.3.1 Standard Normal Distribution, $N(0, 1)$

Suppose the random variable x has a $N(\mu, \sigma^2)$ distribution. Then, the test statistic has a standard normal distribution

$$z = \frac{x - \mu}{\sigma} \sim N(0, 1). \tag{1.25}$$

To see this, notice that $x - \mu$ has a mean of zero and that x/σ has a standard deviation of unity.

Student-t Distribution

1.3.2 t-distribution

If we instead need to estimate σ to use in (1.25), then the test statistic has t_{df}-distribution

$$t = \frac{x - \mu}{\hat{\sigma}} \sim t_n, \tag{1.26}$$

where n denotes the "degrees of freedom," that is the number of observations minus the number of estimated parameters. For instance, if we have a sample with T data points and only estimate the mean, then $n = T - 1$.

The t-distribution has more probability mass in the tails: gives a more "conservative" test (harder to reject the null hypothesis), but the difference vanishes as the degrees of freedom (sample size) increases. See Figure 1.7 for a comparison and Table A.1 for critical values.

Chi Square Distribution

1.3.3 Chi-square Distribution

If $z \sim N(0, 1)$, then $z^2 \sim \chi_1^2$, that is, z^2 has a chi-square distribution with one degree of freedom. This can be generalized in several ways. For instance, if $x \sim N(\mu_x, \sigma_{xx})$ and $y \sim N(\mu_y, \sigma_{yy})$ and they are uncorrelated, then $[(x - \mu_x)/\sigma_x]^2 + [(y - \mu_y)/\sigma_y]^2 \sim \chi_2^2$.

More generally, we have

$$v' \Sigma^{-1} v \sim \chi_n^2, \text{ if the } n \times 1 \text{ vector } v \sim N(0, \Sigma). \tag{1.27}$$

See Figure 1.7 for an illustration and Table A.2 for critical values.

F Distribution

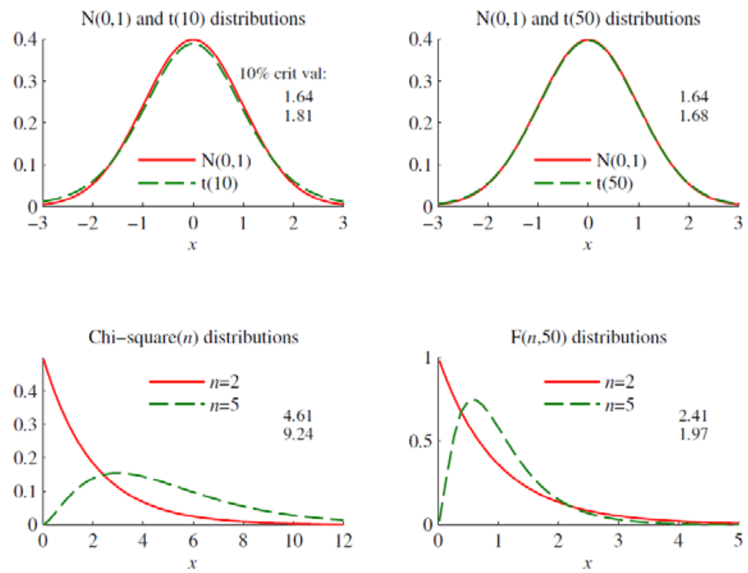
1.3.4 *F*-distribution

If we instead need to estimate Σ in (1.27) and let n_1 be the number of elements in v (previously called just n), then

$$v' \hat{\Sigma}^{-1} v / n_1 \sim F_{n_1, n_2} \tag{1.28}$$

where F_{n_1,n_2} denotes an F-distribution with (n_1, n_2) degrees of freedom. Similar to the t-distribution, n_2 is the number of observations minus the number of estimated parameters. See Figure 1.7 for an illustration and Tables A.3–A.4 for critical values.

Shapes of Commonly Used Distributions

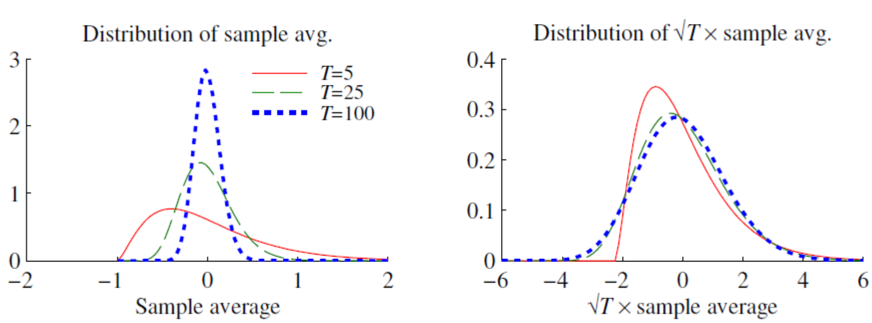


1.4 Asymptotic Theory

The *law of large numbers* (LLN) says that the sample mean converges to the true population mean as the sample size goes to infinity. This holds for a very large class of random variables, but there are exceptions. A sufficient (but not necessary) condition for this convergence is that the sample average is unbiased (as in (1.16)) and that the variance goes to zero as the sample size goes to infinity (as in (1.15)). (This is also called convergence in mean square.) To see the LLN in action, see Figure 1.8.

The *central limit theorem* (CLT) says that $\sqrt{T}\bar{x}$ converges in distribution to a normal distribution as the sample size increases. See Figure 1.8 for an illustration. This also holds for a large class of random variables—and it is a very useful result since it allows us to test hypothesis. Most estimators (including least squares and other methods) are effectively some kind of sample average, so the CLT can be applied.

The Central Limit Theorem in Action



Sample average of $z_t - 1$ where z_t has a $\chi^2(1)$ distribution