Key Concepts

LO 16.1

To compute the population mean, all the observed values in the population are summed and divided by the number of observations in the population.

Variance and standard deviation provide a measure of the extent of the dispersion in the values of the random variable around the mean.

LO 16.2

The mean of a population is expressed as:

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

Variance of a random variable is defined as:

$$\text{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$$
$$\text{where } \mu = E(X)$$

The square root of the variance is called the standard deviation.

LO 16.3

Expected value is the weighted average of the possible outcomes of a random variable, where the weights are the probabilities that the outcomes will occur. The expectation of a random variable *X* having possible values $x_1, \ldots, x_n$ is defined as:

$$E(X) = x_1 P(X = x_1) + \ldots + x_n P(X = x_n)$$

LO 16.4

Covariance measures the extent to which two random variables tend to be above and below their respective means for each joint realization. It can be calculated as:

$$\text{Cov}(A,B) = \sum_{i=1}^{N} P_i \left(A_i - \bar{A}\right)\left(B_i - \bar{B}\right)$$

Correlation is a standardized measure of association between two random variables; it ranges in value from –1 to +1 and is equal to:

$$\frac{\text{Cov}(A,B)}{\sigma_A \sigma_B}$$

LO 16.5

If *X* and *Y* are any random variables, then:

$$E(X + Y) = E(X) + E(Y)$$

If *X* and *Y* are independent random variables, then:

$$Var(X + Y) = Var(X) + Var(Y)$$

$$Var(X - Y) = Var(X) + Var(Y)$$

If *X* and *Y* are NOT independent, then:

$$Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X,Y)$$

$$Var(X - Y) = Var(X) + Var(Y) - 2 \times Cov(X,Y)$$

LO 16.6

The shape of a probability distribution is characterized by its raw moments and central moments. The first raw moment is the mean of the distribution. The second central moment is the variance. The third central moment divided by the cube of the standard deviation measures the skewness of the distribution, and the fourth central moment divided by the fourth power of the standard deviation measures the kurtosis of the distribution.

LO 16.7

Skewness describes the degree to which a distribution is nonsymmetric about its mean.

- A right-skewed distribution has positive skewness and a mean that is higher than the median that is higher than the mode.
- A left-skewed distribution has negative skewness and a mean that is lower than the median that is lower than the mode.

Kurtosis measures the peakedness of a distribution and the probability of extreme outcomes.

- Excess kurtosis is measured relative to a normal distribution, which has a kurtosis of three.
- Positive values of excess kurtosis indicate a distribution that is leptokurtic (fat tails, more peaked).
- Negative values of excess kurtosis indicate a platykurtic distribution (thin tails, less peaked).

Like mean and variance, we can generalize covariance to cross central moments. The third cross central moment is coskewness and the fourth cross central moment is cokurtosis.

LO 16.8

Desirable statistical properties of an estimator include unbiasedness (sign of estimation error is random), efficiency (lower sampling error than any other unbiased estimator), consistency (variance of sampling error decreases with sample size), and linearity (used as a linear function of sample data).

Concept Checkers

1. A distribution of returns that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean:
   A. is positively skewed.
   B. is a symmetric distribution.
   C. has positive excess kurtosis.
   D. has negative excess kurtosis.

2. The correlation of returns between Stocks A and B is 0.50. The covariance between these two securities is 0.0043, and the standard deviation of the return of Stock B is 26%. The variance of returns for Stock A is:
   A. 0.0331.
   B. 0.0011.
   C. 0.2656.
   D. 0.0112.

Use the following data to answer Questions 3 and 4.

Probability Matrix

Returns	R_B = 50%	R_B = 20%	R_B = –30%
R_A = –10%	40%	0%	0%
R_A = 10%	0%	30%	0%
R_A = 30%	0%	0%	30%

3. Given the probability matrix above, the standard deviation of Stock B is closest to:
   A. 0.11.
   B. 0.22.
   C. 0.33.
   D. 0.15.

4. Given the probability matrix above, the covariance between Stock A and B is closest to:
   A. –0.160.
   B. –0.055.
   C. 0.004.
   D. 0.020.

5. A discrete uniform distribution (each event has an equal probability of occurrence) has the following possible outcomes for X: [1, 2, 3, 4]. The variance of this distribution is closest to:
   A. 1.00.
   B. 1.25.
   C. 1.50.
   D. 2.00.

Concept Checker Answers

1. C A distribution that has a greater percentage of small deviations from the mean and a greater percentage of extremely large deviations from the mean will be leptokurtic and will exhibit excess kurtosis (positive). The distribution will be taller and have fatter tails than a normal distribution.

2. B $\text{Corr}(R_A,R_B) = \dfrac{\text{Cov}(R_A,R_B)}{[\sigma(R_A)][\sigma(R_B)]}$

$$\sigma^2(R_A) = \left[\frac{\text{Cov}(R_A,R_B)}{\sigma(R_B)\text{Corr}(R_A,R_B)}\right]^2 = \left[\frac{0.0043}{(0.26)(0.5)}\right]^2 = 0.0331^2 = 0.0011$$

3. C Expected return of Stock B = (0.4)(0.5) + (0.3)(0.2) + (0.3)(–0.3) = 0.17

   $\text{Var}(R_B) = 0.4(0.5 - 0.17)^2 + 0.3(0.2 - 0.17)^2 + 0.3(-0.3 - 0.17)^2 = 0.1101$

   Standard deviation = $\sqrt{0.1101} = 0.3318$

4. B $\text{Cov}(R_A,R_B) = 0.4(-0.1 - 0.08)(0.5 - 0.17) + 0.3(0.1 - 0.08)(0.2 - 0.17) + 0.3(0.3 - 0.08)(-0.3 - 0.17) = -0.0546$

5. B Expected value = (1/4)(1 + 2 + 3 + 4) = 2.5

   Variance = $(1/4)[(1 - 2.5)^2 + (2 - 2.5)^2 + (3 - 2.5)^2 + (4 - 2.5)^2] = 1.25$

   Note that since each observation is equally likely, each has 25% (1/4) chance of occurrence.

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Distributions

Topic 17

Exam Focus

This topic explores common probability distributions: uniform, Bernoulli, binomial, Poisson, normal, lognormal, chi-squared, Student's t, and F. You will learn the properties, parameters, and common occurrences of these distributions. Also discussed is the central limit theorem, which allows us to use sampling statistics to construct confidence intervals for point estimates of population means. For the exam, focus most of your attention on the binomial, normal, and Student's t distributions. Also, know how to standardize a normally distributed random variable, how to use a z-table, and how to construct confidence intervals.

Parametric and Nonparametric Distributions

Probability distributions are classified into two categories: parametric and nonparametric. **Parametric distributions**, such as a normal distribution, can be described by using a mathematical function. These types of distributions make it easier to draw conclusions about the data; however, they also make restrictive assumptions, which are not necessarily supported by real-world patterns. **Nonparametric distributions**, such as a historical distribution, cannot be described by using a mathematical function. Instead of making restrictive assumptions, these types of distributions fit the data perfectly; however, without generalizing the data, it can be difficult for a researcher to draw any conclusions.

LO 17.1: Distinguish the key properties among the following distributions: uniform distribution, Bernoulli distribution, Binomial distribution, Poisson distribution, normal distribution, lognormal distribution, Chi-squared distribution, Student's t, and F-distributions, and identify common occurrences of each distribution.

The Uniform Distribution

The **continuous uniform distribution** is defined over a range that spans between some lower limit, *a*, and some upper limit, *b*, which serve as the parameters of the distribution. Outcomes can only occur between *a* and *b*, and since we are dealing with a continuous distribution, even if $a < x < b$, $P(X = x) = 0$. Formally, the properties of a continuous uniform distribution may be described as follows:

- For all $a \leq x_1 < x_2 \leq b$ (i.e., for all x_1 and x_2 between the boundaries *a* and *b*).
- $P(X < a \text{ or } X > b) = 0$ (i.e., the probability of *X* outside the boundaries is zero).
- $P(x_1 \leq X \leq x_2) = (x_2 - x_1)/(b - a)$. This defines the probability of outcomes between x_1 and x_2.

Don't miss how simple this is just because the notation is so mathematical. For a continuous uniform distribution, the probability of outcomes in a range that is one-half the whole range is 50%. The probability of outcomes in a range that is one-quarter as large as the whole possible range is 25%.

Example: Continuous uniform distribution

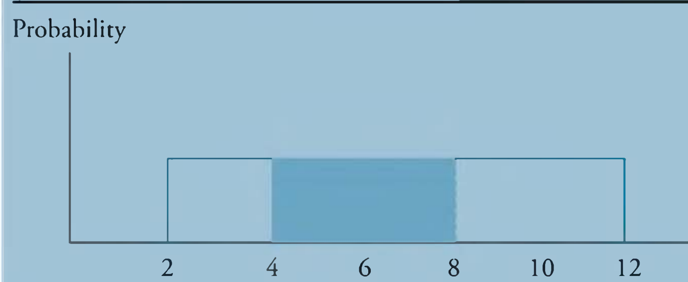
X is uniformly distributed between 2 and 12. **Calculate** the probability that *X* will be between 4 and 8.

Answer:

$$\frac{8-4}{12-2}=\frac{4}{10}=40\%$$

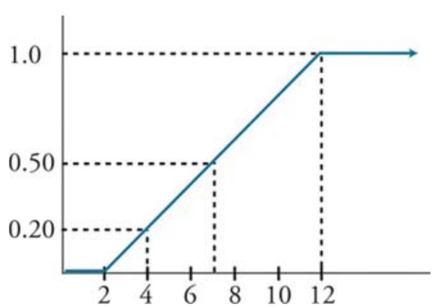
The figure below illustrates this continuous uniform distribution. Note that the area bounded by 4 and 8 is 40% of the total probability between 2 and 12 (which is 100%).

Continuous Uniform Distribution



Since outcomes are equal over equal-size possible intervals, the cumulative distribution function (cdf) is linear over the variable's range. The cdf for the distribution in the above example, Prob (X < x), is shown in Figure 1.

Figure 1: CDF for a Continuous Uniform Variable



The probability function for a continuous random variable is called the probability density function (pdf) and is denoted f(x). Symbolically, the probability density function for a continuous uniform distribution is expressed as:

$$f(x)=\frac{1}{b-a} \text{ for } a \leq x \leq b, \text{ else } f(x)=0$$

The mean and variance, respectively, of a uniform distribution are:

$$E(x) = \frac{a+b}{2}$$

$$Var(x) = \frac{(b-a)^2}{12}$$

The Bernoulli Distribution

A Bernoulli distributed random variable only has two possible outcomes. The outcomes can be defined as either a "success" or a "failure." The probability of success, *p*, may be denoted with the value "1" and the probability of failure, *1 – p*, may be denoted with the value "0." Bernoulli distributed random variables are commonly used for assessing whether or not a company defaults during a specified time period. In the default example, the random variable equals "1" in the event of default and "0" in the event of survival.

The Binomial Distribution

A **binomial random variable** may be defined as the number of "successes" in a given number of trials, whereby the outcome can be either "success" or "failure." The probability of success, *p*, is constant for each trial and the trials are independent. A binomial random variable for which the number of trials is 1 is called a Bernoulli random variable. Think of a trial as a mini-experiment (or Bernoulli trial). The final outcome is the number of successes in a series of *n* trials. Under these conditions, the binomial probability function defines the probability of *x* successes in *n* trials. It can be expressed using the following formula:

$$p(x) = P(X = x) = (\text{number of ways to choose } x \text{ from } n)p^x(1 - p)^{n-x}$$

where:

$$(\text{number of ways to choose x from n}) = \frac{n!}{(n-x)!x!}$$

p = the probability of "success" on each trial [don't confuse it with p(x)]

So the probability of exactly *x* successes in *n* trials is:

$$p(x) = \frac{n!}{(n-x)!x!}p^x(1-p)^{n-x}$$

Example: Binomial probability

Assuming a binomial distribution, **compute** the probability of drawing three black beans from a bowl of black and white beans if the probability of selecting a black bean in any given attempt is 0.6. You will draw five beans from the bowl.

Answer:

$$P(X=3)=p(3)=\frac{5!}{2!3!}(0.6)^3(0.4)^2=(120\ /\ 12)(0.216)(0.160)=0.3456$$

Some intuition about these results may help you remember the calculations. Consider that a (very large) bowl of black and white beans has 60% black beans and that each time you select a bean, you replace it in the bowl before drawing again. We want to know the probability of selecting exactly three black beans in five draws, as in the previous example.

One way this might happen is BBBWW. Since the draws are independent, the probability of this is easy to calculate. The probability of drawing a black bean is 60%, and the probability of drawing a white bean is 1 – 60% = 40%. Therefore, the probability of selecting BBBWW, in order, is 0.6 × 0.6 × 0.6 × 0.4 × 0.4 = 3.456%. This is the $p^3(1-p)^2$ from the formula and p is 60%, the probability of selecting a black bean on any single draw from the bowl. BBBWW is not, however, the only way to choose exactly three black beans in five trials. Another possibility is BBWWB, and a third is BWWBB. Each of these will have exactly the same probability of occurring as our initial outcome, BBBWW. That's why we need to answer the question of how many ways (different orders) there are for us to choose three black beans in five draws. Using the formula, there are $\frac{5!}{(5-3)!3!}=10$ ways; 10 × 3.456% = 34.56%, the answer we computed above.

Expected Value and Variance of a Binomial Random Variable

For a given series of *n* trials, the expected number of successes, or E(X), is given by the following formula:

$$\text{expected value of X} = E(X) = np$$

The intuition is straightforward; if we perform *n* trials and the probability of success on each trial is *p*, we expect *np* successes.

The variance of a binomial random variable is given by:

$$\text{variance of X} = np(1-p) = npq$$

Professor's Note: q = 1 – p is the probability that the event will fail to occur in a single trial (i.e., the probability of failure).

Example: Expected value of a binomial random variable

Based on empirical data, the probability that the Dow Jones Industrial Average (DJIA) will increase on any given day has been determined to equal 0.67. Assuming the only other outcome is that it decreases, we can state p(UP) = 0.67 and p(DOWN) = 0.33. Further, assume that movements in the DJIA are independent (i.e., an increase in one day is independent of what happened on another day).

Using the information provided, **compute** the expected value of the number of up days in a 5-day period.

Answer:

Using binomial terminology, we define success as UP, so $p = 0.67$. Note that the definition of success is critical to any binomial problem.

$$E(X \mid n = 5, p = 0.67) = (5)(0.67) = 3.35$$

Recall that the "|" symbol means *given*. Hence, the preceding statement is read as: the expected value of X given that n = 5, and the probability of success = 67% is 3.35.

Using the equation for the variance of a binomial distribution, we find the variance of X to be:

$$\text{Var}(X) = np(1 - p) = 5(0.67)(0.33) = 1.106$$

We should note that since the binomial distribution is a discrete distribution, the result $X = 3.35$ is not possible. However, if we were to record the results of many 5-day periods, the average number of up days (successes) would converge to 3.35.

Binomial distributions are used extensively in the investment world where outcomes are typically seen as successes or failures. In general, if the price of a security goes up, it is viewed as a success. If the price of a security goes down, it is a failure. In this context, binomial distributions are often used to create models to aid in the process of asset valuation.

Professor's Note: We will examine binomial trees for stock option valuation in Book 4.

The Poisson Distribution

The Poisson distribution is another discrete probability distribution with a number of real-world applications. For example, the number of defects per batch in a production process or the number of calls per hour arriving at the 911 emergency switchboard are discrete random variables that follow a Poisson distribution.

While the Poisson random variable X refers to the *number of successes per unit*, the parameter lambda (λ) refers to the average or *expected number of successes per unit*. The mathematical expression for the Poisson distribution for obtaining X successes, given that λ successes are expected, is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

An interesting feature of the Poisson distribution is that both its mean and variance are equal to the parameter, λ.

Example: Using the Poisson distribution (1)

On average, the 911 emergency switchboards receive 0.1 incoming calls per second. What is the probability that in a given minute exactly 5.0 phone calls will be received, assuming the arrival of calls follows a Poisson distribution?

Answer:

We first need to convert the seconds into minutes. Note that λ, the expected number of calls per minute, is (0.1)(60) = 6.0. Hence:

$$P(X = 5) = \frac{6^5 e^{-6}}{5!} = 0.1606 = 16.06\%$$

This means that, given the average of 0.1 incoming calls per second, there is a 16.06% chance there will be five incoming phone calls in a minute.

Example: Using the Poisson distribution (2)

Assume there is a 0.01 probability of a patient experiencing severe weight loss as a side effect from taking a recently approved drug used to treat heart disease. What is the probability that out of 200 such procedures conducted on different patients, five patients will develop this complication? Assume that the number of patients developing the complication from the procedure is Poisson-distributed.

Answer:

Let X = expected number of patients developing the complication from the procedure
= np = (200)(0.01) = 2

$$P(X=5)=\frac{\lambda^{x}e^{-\lambda}}{x!}=\frac{2^{5}e^{-2}}{5!}=0.036=3.6\%$$

This means that given a complication rate of 0.01, there is a 3.6% probability that 5 out of every 200 patients will experience severe weight loss from taking the drug.

The Normal Distribution

The normal distribution is important for many reasons. Many of the random variables that are relevant to finance and other professional disciplines follow a normal distribution. In the area of investment and portfolio management, the normal distribution plays a central role in portfolio theory.

The probability density function for the normal distribution is:

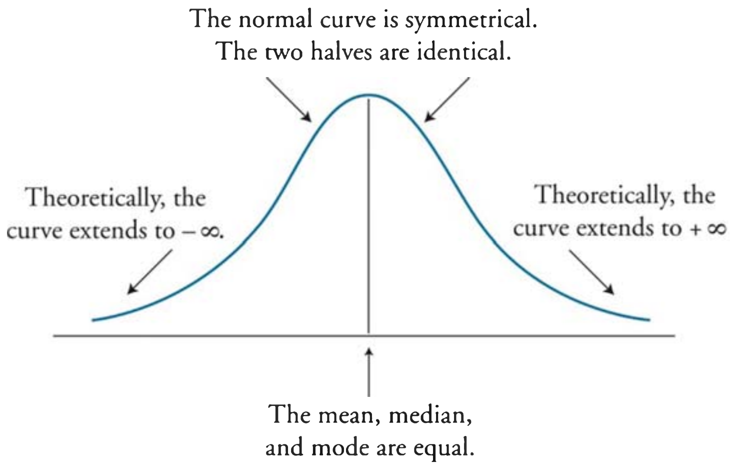
$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}$$

The normal distribution has the following key properties:

- It is completely described by its mean, μ, and variance, σ^2, stated as $X \sim N(\mu, \sigma^2)$. In words, this says that "*X* is normally distributed with mean μ and variance σ^2."
- Skewness = 0, meaning the normal distribution is symmetric about its mean, so that $P(X \leq \mu) = P(\mu \leq X) = 0.5$, and mean = median = mode.
- Kurtosis = 3; this is a measure of how flat the distribution is. Recall that excess kurtosis is measured relative to 3, the kurtosis of the normal distribution.
- A linear combination of normally distributed independent random variables is also normally distributed.
- The probabilities of outcomes further above and below the mean get smaller and smaller but do not go to zero (the tails get very thin but extend infinitely).

Many of these properties are evident from examining the graph of a normal distribution's probability density function as illustrated in Figure 2.

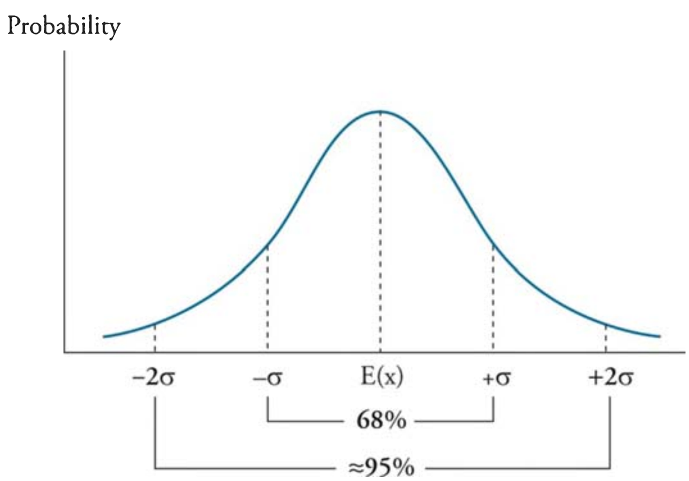
Figure 2: Normal Distribution Probability Density Function



A **confidence interval** is a range of values around the expected outcome within which we expect the actual outcome to be some specified percentage of the time. A 95% confidence interval is a range that we expect the random variable to be in 95% of the time. For a normal distribution, this interval is based on the expected value (sometimes called a point estimate) of the random variable and on its variability, which we measure with standard deviation.

Confidence intervals for a normal distribution are illustrated in Figure 3. For any normally distributed random variable, 68% of the outcomes are within one standard deviation of the expected value (mean), and approximately 95% of the outcomes are within two standard deviations of the expected value.

Figure 3: Confidence Intervals for a Normal Distribution



In practice, we will not know the actual values for the mean and standard deviation of the distribution, but will have estimated them as $\bar{X}$ and s. The three confidence intervals of most interest are given by:

- The 90% confidence interval for X is $\bar{X} - 1.65s$ to $\bar{X} + 1.65s$.
- The 95% confidence interval for X is $\bar{X} - 1.96s$ to $\bar{X} + 1.96s$.
- The 99% confidence interval for X is $\bar{X} - 2.58s$ to $\bar{X} + 2.58s$.

Example: Confidence intervals

The average return of a mutual fund is 10.5% per year and the standard deviation of annual returns is 18%. If returns are approximately normal, what is the 95% confidence interval for the mutual fund return next year?

Answer:

Here μ and σ are 10.5% and 18%, respectively. Thus, the 95% confidence interval for the return, R, is:

$$10.5 \pm 1.96(18) = -24.78\% \text{ to } 45.78\%$$

Symbolically, this result can be expressed as:

$$P(-24.78 < R < 45.78) = 0.95 \text{ or } 95\%$$

The interpretation is that the annual return is expected to be within this interval 95% of the time, or 95 out of 100 years.

The Standard Normal Distribution

A standard normal distribution (i.e., *z*-distribution) is a normal distribution that has been standardized so it has a mean of zero and a standard deviation of 1 [i.e., N~(0,1)]. To standardize an observation from a given normal distribution, the *z-value* of the observation must be calculated. The *z*-value represents the number of standard deviations a given observation is from the population mean. *Standardization* is the process of converting an observed value for a random variable to its *z*-value. The following formula is used to standardize a random variable:

$$z = \frac{\text{observation} - \text{population mean}}{\text{standard deviation}} = \frac{x - \mu}{\sigma}$$

Professor's Note: The term z-value will be used for a standardized observation in this topic. The terms z-score and z-statistic are also commonly used.

Example: Standardizing a random variable (calculating *z*-values)

Assume the annual earnings per share (EPS) for a population of firms are normally distributed with a mean of \$6 and a standard deviation of \$2.

What are the *z*-values for EPS of \$2 and \$8?

Answer:

If EPS = x = \$8, then $z = (x - \mu) / \sigma = (\$8 - \$6) / \$2 = +1$

If EPS = x = \$2, then $z = (x - \mu) / \sigma = (\$2 - \$6) / \$2 = -2$

Here, $z = +1$ indicates that an EPS of \$8 is one standard deviation above the mean, and $z = -2$ means that an EPS of \$2 is two standard deviations below the mean.

Calculating Probabilities Using *z*-Values

Now we will show how to use standardized values (*z*-values) and a table of probabilities for *Z* to determine probabilities. A portion of a table of the cumulative distribution function for a standard normal distribution is shown in Figure 4. We will refer to this table as the *z*-table, as it contains values generated using the cumulative density function for a standard normal distribution, denoted by F(Z). Thus, the values in the *z*-table are the probabilities of observing a *z*-value that is less than a given value, *z* [i.e., $P(Z < z)$]. The numbers in the first column are *z*-values that have only one decimal place. The columns to the right supply probabilities for *z*-values with two decimal places.

Note that the *z*-table in Figure 4 only provides probabilities for positive *z*-values. This is not a problem because we know from the symmetry of the standard normal distribution that $F(-Z) = 1 - F(Z)$. The tables in the back of many texts actually provide probabilities for negative *z*-values, but we will work with only the positive portion of the table because this may be all you get on the exam. In Figure 4, we can find the probability that a standard normal random variable will be less than 1.66, for example. The table value is 95.15%. The probability that the random variable will be less than –1.66 is simply 1 – 0.9515 = 0.0485 = 4.85%, which is also the probability that the variable will be greater than +1.66.

Figure 4: Cumulative Probabilities for a Standard Normal Distribution

CDF Values for the Standard Normal Distribution: The z-Table										
z	**.00**	**.01**	**.02**	**.03**	**.04**	**.05**	**.06**	**.07**	**.08**	**.09**
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.5	.6915	Please note that several of the rows have been deleted to save space.*								
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

*A complete cumulative standard normal table is included in the Appendix.

Professor's Note: When you use the standard normal probabilities, you have formulated the problem in terms of standard deviations from the mean. Consider a security with returns that are approximately normal, an expected return of 10%, and standard deviation of returns of 12%. The probability of returns greater than 30% is calculated based on the number of standard deviations that 30% is above the expected return of 10%. 30% is 20% above the expected return of 10%, which is 20 / 12 = 1.67 standard deviations above the mean. We look up the probability of returns less than 1.67 standard deviations above the mean (0.9525 or 95.25% from Figure 4) and calculate the probability of returns more than 1.67 standard deviations above the mean as 1 – 0.9525 = 4.75%.

Example: Using the z-table (1)

Considering again EPS distributed with μ = \$6 and σ = \$2, what is the probability that EPS will be \$9.70 or more?

Answer:

Here we want to know P(EPS > \$9.70), which is the area under the curve to the right of the *z*-value corresponding to EPS = \$9.70 (see the distribution below).

The *z*-value for EPS = \$9.70 is:

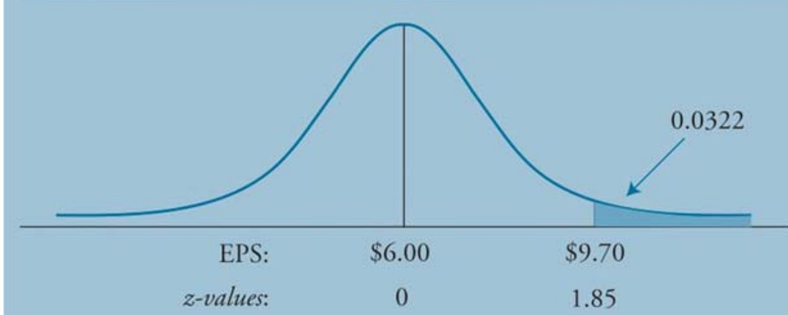
$$z = \frac{(x - \mu)}{\sigma} = \frac{(9.70 - 6)}{2} = 1.85$$

That is, \$9.70 is 1.85 standard deviations above the mean EPS value of \$6.

From the z-table we have F(1.85) = 0.9678, but this is P(EPS ≤ 9.70). We want P(EPS > 9.70), which is 1 – P(EPS ≤ 9.70).

P(EPS > 9.70) = 1 – 0.9678 = 0.0322, or 3.2%

P(EPS>$9.70)



Example: Using the z-table (2)

Using the distribution of EPS with μ = $6 and σ = $2 again, what percent of the observed EPS values are likely to be less than $4.10?

Answer:

As shown graphically in the distribution below, we want to know P(EPS < $4.10). This requires a 2-step approach like the one taken in the preceding example.

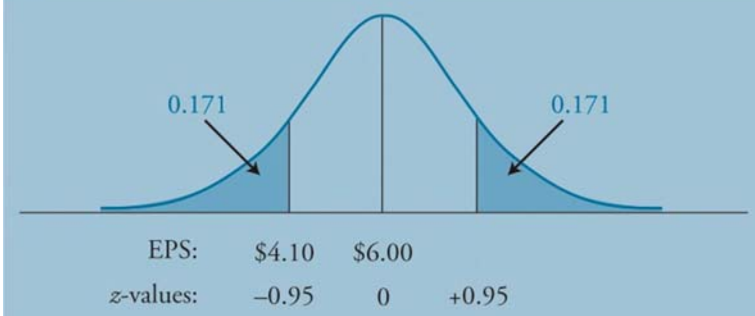
First, the corresponding z-value must be determined as follows:

$$z = \frac{(\$4.10 - \$6)}{2} = -0.95,$$

So $4.10 is 0.95 standard deviations below the mean of $6.00.

Now, from the z-table for negative values in the back of this book, we find that F(–0.95) = 0.1711, or 17.11%.

Finding a Left-Tail Probability



The *z*-table gives us the probability that the outcome will be more than 0.95 standard deviations below the mean.

The Lognormal Distribution

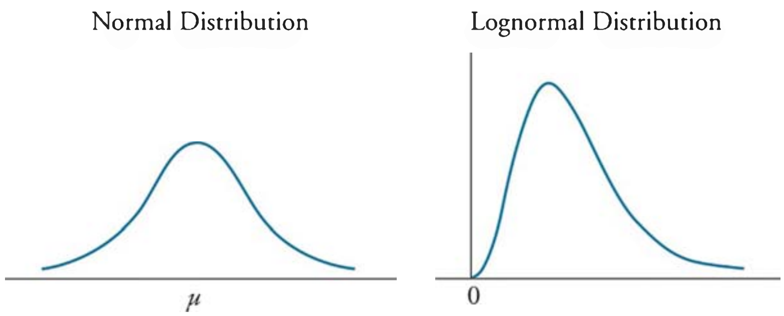
The lognormal distribution is generated by the function e^x, where x is normally distributed. Since the natural logarithm, ln, of e^x is x, the logarithms of lognormally distributed random variables are normally distributed, thus the name.

The probability density function for the lognormal distribution is:

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

Figure 5 illustrates the differences between a normal distribution and a lognormal distribution.

Figure 5: Normal vs. Lognormal Distributions



In Figure 5, we can see that:

- The lognormal distribution is skewed to the right.
- The lognormal distribution is bounded from below by zero so that it is useful for modeling asset prices which never take negative values.

If we used a normal distribution of returns to model asset prices over time, we would admit the possibility of returns less than –100%, which would admit the possibility of asset prices less than zero. Using a lognormal distribution to model *price relatives* avoids this problem. A price relative is just the end-of-period price of the asset divided by the beginning price (S_1/S_0) and is equal to (1 + the holding period return). To get the end-of-period asset price, we can simply multiply the price relative times the beginning-of-period asset price. Since a lognormal distribution takes a minimum value of zero, end-of-period asset prices cannot be less than zero. A price relative of zero corresponds to a holding period return of –100% (i.e., the asset price has gone to zero).

The Central Limit Theorem

LO 17.2: Describe the central limit theorem and the implications it has when combining independent and identically distributed (i.i.d.) random variables.

LO 17.3: Describe i.i.d. random variables and the implications of the i.i.d. assumption when combining random variables.

The central limit theorem states that for simple random samples of size *n* from a *population* with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean $\bar{x}$ approaches a normal probability distribution with mean μ and variance equal to $\frac{\sigma^2}{n}$ as the sample size becomes large. This is possible because, when the sample size is large, the sums of independent and identically distributed (i.i.d.) random variables (the individual items drawn for the sample) will be normally distributed.

The central limit theorem is extremely useful because the normal distribution is relatively easy to apply to hypothesis testing and to the construction of confidence intervals. Specific inferences about the population mean can be made from the sample mean, *regardless of the population's distribution*, as long as the sample size is "sufficiently large," which usually means $n \geq 30$.

Important properties of the central limit theorem include the following:

- If the sample size *n* is sufficiently large ($n \geq 30$), the sampling distribution of the sample means will be approximately normal. Remember what's going on here: random samples of size *n* are repeatedly being taken from an overall larger population. Each of these random samples has its own mean, which is itself a random variable, and this set of sample means has a distribution that is approximately normal.
- The mean of the population, μ, and the mean of the distribution of all possible sample means are equal.
- The variance of the distribution of sample means is $\frac{\sigma^2}{n}$, the population variance divided by the sample size.

Student's *t*-Distribution

Student's *t*-distribution, or simply the *t*-distribution, is a bell-shaped probability distribution that is symmetrical about its mean. It is the appropriate distribution to use when constructing confidence intervals based on *small samples* ($n < 30$) from populations with *unknown variance* and a normal, or approximately normal, distribution. It may also be appropriate to use the *t*-distribution when the population variance is unknown and the sample size is large enough that the central limit theorem will assure that the sampling distribution is approximately normal.

Student's *t*-distribution has the following properties:

- It is symmetrical.
- It is defined by a single parameter, the degrees of freedom (df), where the degrees of freedom are equal to the number of sample observations minus 1, n – 1, for sample means.
- It has more probability in the tails (fatter tails) than the normal distribution.
- As the degrees of freedom (the sample size) gets larger, the shape of the t-distribution more closely approaches a standard normal distribution.

When *compared to the normal distribution*, the *t*-distribution is flatter with more area under the tails (i.e., it has fatter tails). As the degrees of freedom for the *t*-distribution increase, however, its shape approaches that of the normal distribution.

The degrees of freedom for tests based on sample means are n – 1 because, given the mean, only n – 1 observations can be unique.

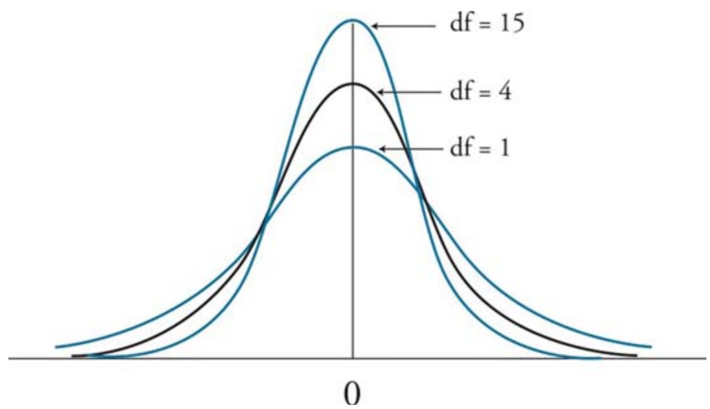
The table in Figure 6 contains one-tailed critical values for the t-distribution at the 0.05 and 0.025 levels of significance with various degrees of freedom (df). Note that, unlike the *z*-table, the *t*-values are contained within the table and the probabilities are located at the column headings. Also note that the level of significance of a *t*-test corresponds to the *one-tailed probabilities*, *p*, that head the columns in the *t*-table.

Figure 6: Table of Critical *t*-Values

	One-Tailed Probabilities, p	
df	*p = 0.05*	*p = 0.025*
5	2.015	2.571
10	1.812	2.228
15	1.753	2.131
20	1.725	2.086
25	1.708	2.060
30	1.697	2.042
40	1.684	2.021
50	1.676	2.009
60	1.671	2.000
70	1.667	1.994
80	1.664	1.990
90	1.662	1.987
100	1.660	1.984
120	1.658	1.980
∞	1.645	1.960

Figure 7 illustrates the different shapes of the *t*-distribution associated with different degrees of freedom. The tendency is for the *t*-distribution to look more and more like the normal distribution as the degrees of freedom increase. Practically speaking, the greater the degrees of freedom, the greater the percentage of observations near the center of the distribution and the lower the percentage of observations in the tails, which are thinner as degrees of freedom increase. This means that confidence intervals for a random variable that follows a *t*-distribution must be wider (narrower) when the degrees of freedom are less (more) for a given significance level.

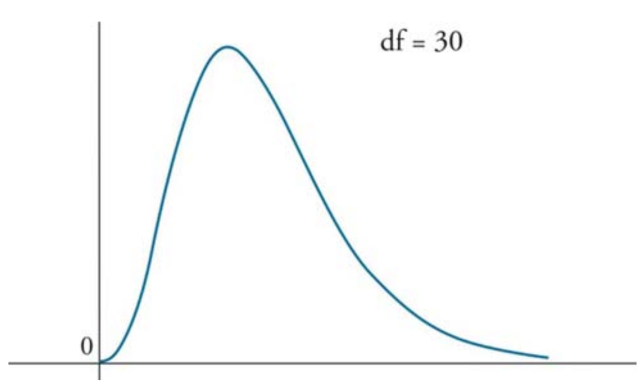
Figure 7: *t*-Distributions for Different Degrees of Freedom (df)



The Chi-Squared Distribution

As you will see in Topic 19, hypothesis testing of the population variance requires the use of a chi-squared distributed test statistic, denoted χ^2. The chi-square distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.

Figure 8: Chi-Squared Distribution



The chi-squared test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

$$\chi^2_{n-1} = \frac{(n-1)s^2}{\sigma_0^2}$$

where:
n = sample size
s^2 = sample variance
σ_0^2 = hypothesized value for the population variance

The chi-squared test compares the test statistic to a critical chi-squared value at a given level of significance to determine whether to reject or fail to reject a null hypothesis. Note that since the chi-squared distribution is bounded below by zero, chi-squared values cannot be negative.

The *F*-Distribution

As you will also see in Topic 19, the hypotheses concerned with the equality of the variances of two populations are tested with an *F*-distributed test statistic. Hypothesis testing using a test statistic that follows an *F*-distribution is referred to as the *F*-test. The *F*-test is used under the assumption that the populations from which samples are drawn are normally distributed and that the samples are independent.

The test statistic for the *F*-test is the ratio of the sample variances. The *F*-statistic is computed as:

$$F = \frac{s_1^2}{s_2^2}$$

where:
s_1^2 = variance of the sample of n_1 observations drawn from Population 1
s_2^2 = variance of the sample of n_2 observations drawn from Population 2

An *F*-distribution is presented in Figure 9. As indicated, the *F*-distribution is right-skewed and is truncated at zero on the left-hand side. The shape of the *F*-distribution is determined by *two separate degrees of freedom*, the numerator degrees of freedom, df_1, and the denominator degrees of freedom, df_2.

Note that $n_1 - 1$ and $n_2 - 1$ are the degrees of freedom used to identify the appropriate critical value from the *F*-table (provided in the Appendix).

Some additional properties of the *F*-distribution include the following:

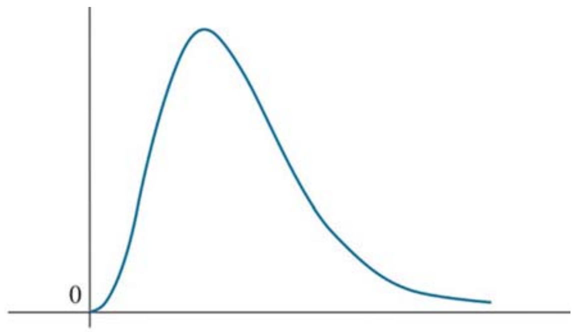
- The *F*-distribution approaches the normal distribution as the number of observations increases (just as with the *t*-distribution and chi-squared distribution).
- A random variable's *t*-value squared (t^2) with n – 1 degrees of freedom is *F*-distributed with 1 degree of freedom in the numerator and n – 1 degrees of freedom in the denominator.
- There exists a relationship between the *F*- and chi-squared distributions such that:

$$F = \frac{\chi^2}{\text{\# of observations in numerator}}$$

as the # of observations in denominator $\rightarrow \infty$

Figure 9: *F*-Distribution

numerator $df_1 = 10$, denominator $df_2 = 10$



Mixture Distributions

LO 17.4: Describe a mixture distribution and explain the creation and characteristics of mixture distributions.

The distributions discussed in this topic, as well as others, can be combined to create unique probability density functions. It may be helpful to create a new distribution if the underlying data you are working with does not currently fit a predetermined distribution. In this case, a newly created distribution may assist with explaining the relevant data.

To illustrate a mixture distribution, suppose that the returns of a stock follow a normal distribution with low volatility 75% of the time and high volatility 25% of the time. Here we have two normal distributions with the same mean, but different risk levels. To create a mixture distribution from these scenarios, we randomly choose either the low or high volatility distribution, placing a 75% probability on selecting the low volatility distribution. We then generate a random return from the selected distribution. By repeating this process several times, we will create a probability distribution that reflects both levels of volatility.

Mixture distributions contain elements of both parametric and nonparametric distributions. The distributions used as inputs (i.e., the component distributions) are parametric, while the weights of each distribution within the mixture are nonparametric. The more component distributions used as inputs, the more closely the mixture distribution will follow the actual data. However, more component distributions will make it difficult to draw conclusions given that the newly created distribution will be very specific to the data.

By mixing distributions, it is easy to see how we can alter skewness and kurtosis of the component distributions. Skewness can be changed by combining distributions with different means, and kurtosis can be changed by combining distributions with different variances. Also, by combining distributions that have significantly different means, we can create a mixture distribution with multiple modes (e.g., a bimodal distribution).

Creating a more robust distribution is clearly beneficial to risk managers. Different levels of skew and/or kurtosis can reveal extreme events that were previously difficult to identify. By creating these mixture distributions, we can improve risk models by incorporating the potential for low-frequency, high-severity events.

Key Concepts

LO 17.1

A continuous uniform distribution is one where the probability of *X* occurring in a possible range is the length of the range relative to the total of all possible values. Letting *a* and *b* be the lower and upper limit of the uniform distribution, respectively, then for: $a \le x_1 < x_2 \le b$,

$$P(x_1 \le X \le x_2) = \frac{(x_2 - x_1)}{(b - a)}$$

The binomial distribution is a discrete probability distribution for a random variable, *X*, that has one of two possible outcomes, success or failure. The probability of a specific number of successes in *n* independent binomial trials is:

$$p(x) = P(X = x) = \frac{n!}{(n - x)!x!} p^x (1 - p)^{n-x}$$

where p = the probability of success in a given trial

The Poisson random variable *X* refers to a specific number of successes per unit. The probability for obtaining *X* successes, given a Poisson distribution with parameter λ is:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The normal probability distribution has the following characteristics:

- The normal curve is symmetrical and bell-shaped with a single peak at the exact center of the distribution.
- Mean = median = mode, and all are in the exact center of the distribution.
- The normal distribution can be completely defined by its mean and standard deviation because the skew is always zero and kurtosis is always three.

A lognormal distribution exists for random variable *Y*, when $Y = e^X$, and *X* is normally distributed.

The *t*-distribution is similar, but not identical, to the normal distribution in shape—it is defined by the degrees of freedom, has a lower peak, and has fatter tails. The *t*-distribution is used to construct confidence intervals for the population mean when the population variance is not known.

Degrees of freedom for the *t*-distribution is equal to n – 1; Student's *t*-distribution is closer to the normal distribution when *df* is greater, and confidence intervals are narrower when df is greater.

The chi-squared distribution is asymmetrical, bounded below by zero, and approaches the normal distribution in shape as the degrees of freedom increase.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

The *F*-distribution is right-skewed and is truncated at zero on the left-hand side. The shape of the *F*-distribution is determined by two separate degrees of freedom.

LO 17.2

The central limit theorem states that for a population with a mean μ and a finite variance σ^2, the sampling distribution of the sample mean of all possible samples of size *n* will be approximately normally distributed with a mean equal to μ and a variance equal to σ^2/n.

LO 17.3

When a sample size is large, the sums of independent and identically distributed (i.i.d.) random variables will be normally distributed.

LO 17.4

Mixture distributions combine the concepts of parametric and nonparametric distributions. The component distributions used as inputs are parametric while the weights of each distribution within the mixture are based on historical data, which is nonparametric.

【梦轩考资网www.mxkaozi.com】QQ106454842 专业提供CFA FRM全程高清视频+讲义

Concept Checkers

1. Which of the following statements about the *F*-distribution and chi-squared distribution is least accurate? Both distributions:
   A. are asymmetrical.
   B. are bound by zero on the left.
   C. are defined by degrees of freedom.
   D. have means that are less than their standard deviations.

2. The probability that a standard normally distributed random variable will be more than two standard deviations above its mean is:
   A. 0.0217.
   B. 0.0228.
   C. 0.4772.
   D. 0.9772.

3. If 5% of the cars coming off the assembly line have some defect in them, what is the probability that out of three cars chosen at random, exactly one car will be defective? Assume that the number of defective cars has a Poisson distribution.
   A. 0.129.
   B. 0.135.
   C. 0.151.
   D. 0.174.

4. A recent study indicated that 60% of all businesses have a fax machine. Assuming a binomial probability distribution, what is the probability that exactly four businesses will have a fax machine in a random selection of six businesses?
   A. 0.138.
   B. 0.276.
   C. 0.311.
   D. 0.324.

5. What is the probability of an outcome being between 15 and 25 for a random variable that follows a continuous uniform distribution over the range of 12 to 28?
   A. 0.509.
   B. 0.625.
   C. 1.000.
   D. 1.600.

Concept Checker Answers

1. D There is no consistent relationship between the mean and standard deviation of the chi-squared distribution or *F*-distribution.

2. B 1 – F(2) = 1 – 0.9772 = 0.0228

3. A The probability of a defective car (*p*) is 0.05; hence, the probability of a non-defective car (q) = 1 – 0.05 = 0.95. Assuming a Poisson distribution:

   $\lambda = np = (3)(0.05) = 0.15$

   Then,

   $$P(X=1)=\frac{\lambda^x e^{-\lambda}}{x!}=\frac{(0.15)^1 e^{-0.15}}{1!}=0.129106$$

4. C Success = having a fax machine:

   $[6!\,/\,4!(6-4)!](0.6)^4(0.4)^{6-4} = 15(0.1296)(0.16) = 0.311$

5. B Since a = 12 and b = 28:

   $$P(15\leq X\leq 25)=\frac{(25-15)}{(28-12)}=\frac{10}{16}=0.625$$

The following is a review of the Quantitative Analysis principles designed to address the learning objectives set forth by GARP®. This topic is also covered in:

Bayesian Analysis

Topic 18

Exam Focus

Bayes' theorem is used to update a given set of prior probabilities for a given event in response to the arrival of new information. Updating a prior probability of an event requires knowledge of both conditional and unconditional probabilities. For the exam, be prepared to calculate updated probabilities when applying Bayesian analysis based on the probability of conditional and unconditional events occurring. Also, be prepared to contrast the Bayesian approach with the frequentist approach.

Bayes' Theorem

LO 18.1: Describe Bayes' theorem and apply this theorem in the calculation of conditional probabilities.

Bayesian analysis is applied in numerous disciplines and is growing in interest in finance and risk management. The foundation of Bayesian analysis is **Bayes' theorem**. Bayes' theorem for two random variables *A* and *B* is defined as follows:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

For this topic, it is helpful to recall the notation and definitions of conditional, unconditional, and joint probabilities. The notation for a **conditional probability** is shown on the left-hand side of the equation, P(A | B). The conditional probability is read as the probability of event *A* occurring, given that event *B* has already occurred. The **unconditional probability** of event *A* occurring is noted as P(A). This is an overall probability of event *A* occurring regardless of the outcome of other events.

The numerator of the above equation [P(B | A) × P(A)] is the joint probability of events *A* and *B*. The joint probability of two events occurring at the same time can also be stated as P(AB). Therefore, another way of expressing Bayes' theorem based on the joint probability of both events occurring is shown as follows:

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

The joint probability of both events *A* and *B* occurring can be determined by the following two equations. Notice that it does not matter which event occurred first. The first equation is used if event *B* occurred first and the second equation is used if event *A* occurred first.

$$P(AB) = P(A \mid B) \times P(B)$$
$$P(AB) = P(B \mid A) \times P(A)$$

Regardless of which unconditional event occurred first, the joint probability of both occurring is the same. Thus, these two equations can be combined. Notice that if we divide each side of this equation by P(B), we have the first derivation of Bayes' theorem introduced in this topic.

$$P(A \mid B) \times P(B) = P(B \mid A) \times P(A)$$

Bayes' theorem provides a framework for determining the probability of one random event occurring given that another random event has already occurred. This is known as a conditional probability. The following example illustrates how to determine the probability of one bond defaulting given that another bond has already defaulted.

Suppose a bond manager is interested in knowing the probability of Bond A defaulting given that Bond B is already in default. Figure 1 provides a probability matrix defining two events for both bonds, default and no default. Bonds A and B each have a 12% probability of default and an 88% probability of not defaulting. The bottom row of Figure 1 sums the total probabilities for Bond A for no default and default as 88% and 12%, respectively. Likewise, the last column of Figure 1 sums the total of no default and default for Bond B as 88% and 12%, respectively. The joint probability of both bonds defaulting is 4% in this example. Similarly, the joint probability of no defaults for either bond is 80%.

Figure 1: Probability Matrix for Bond A and Bond B

		Bond A		
		No Default	Default	
Bond B	No Default	80%	8%	88%
	Default	8%	4%	12%
		88%	12%	100%

Professor's Note: The two events for each bond must sum to 100% (88% + 12% = 100%). Each bond will either be in a state of default or no default.

The recent financial crisis beginning in 2007 illustrated that bond defaults are highly correlated. If the probabilities of bond defaults were independent, then the probability of both bonds defaulting would be calculated as 1.44% (i.e., 12% × 12%). However, the actual joint probability of both bonds defaulting is much higher at 4%. In addition, the joint probability that both bonds do not default is 80%. This probability is higher than the probability for two independent events each with an 88% probability of occurring (i.e., 88% × 88% = 77.44%).

As mentioned, an unconditional probability is a random event that is not contingent on any additional information or events occurring. The unconditional probability of Bond A defaulting is the overall probability of Bond A default given in the example as 12%. In other words, there is a 12% probability of Bond A defaulting regardless of the state of Bond B.

The conditional probability of Bond A defaulting given that Bond B is already in default is defined by: $P(A \mid B) = P(AB) / P(B)$. The numerator is the joint probability of both defaulting, $P(AB) = 4\%$. The denominator is the unconditional probability of Bond B defaulting, $P(B)$. Thus, the conditional probability can be computed as:

$$P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{4\%}{12\%} = \frac{1}{3} \text{ or } 33.3333\%$$

Professor's Note: If two events are highly correlated, the conditional probability of the event occurring (e.g., Bond A defaults given that Bond B is in default) is always higher than the unconditional probability of the event occurring.

Now we will look at another example that does not have everything neatly presented in a probability matrix.

Example: Bayes' theorem (1)

Suppose you are an equity analyst for ABC Insurance Company. You manage an equity fund of funds and use historical data to categorize the managers as excellent or average. Excellent managers are expected to outperform the market 70% of the time. Average managers are expected to outperform the market only 50% of the time. Assume that the probabilities of managers outperforming the market for any given year is independent of their performance in prior years. ABC Insurance Company has found that only 20% of all fund managers are excellent managers and the remaining 80% are average managers.

A new fund manager to the portfolio started three years ago and outperformed the market all three years. What is the probability that the new manager was an excellent manager when she first started managing portfolios three years ago?

Answer:

The last probabilities stated in the problem are the probabilities that a random fund manager is either an excellent manager [P(E) = 20%] or an average manager [P(A) = 80%].

The unconditional probability will answer the question related to the new manager (a random event occurring given no other information). There was a 20% probability that the new manager was an excellent manager when she first joined three years ago.

Bayesian analysis requires updating prior beliefs based on new information. In the prior example, we have new information that the manager outperformed the market three years in a row. Therefore, this information will change our prior beliefs regarding the probabilities that the manager is either excellent or average. This next example illustrates how Bayesian analysis updates prior beliefs based on new information.

Example: Bayes' theorem (2)

Using the same information given in the previous example, what are the probabilities that the new manager is an excellent or average manager today?

Answer:

To solve this problem, we first summarize the conditional probabilities related to the probability of outperforming the market given that the fund manager is either excellent or average.

- The probability of an excellent manager outperforming the market is 70% [P(O | E) = 70%]. The notation is read as the probability that a manager outperforms the market given she is an excellent manager equals 70%.
- The probability of an average manager outperforming the market is 50% [P(O | A) = 50%].

Next, we need to use Bayes' theorem to determine the probability that the new manager is excellent given that the manager outperformed the market three years in a row.

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)}$$

The numerator of Bayes' theorem is the probability that an excellent manager outperforms the market three years in a row [P(O | E) × P(E)]. In other words, it is a joint probability of a manager being excellent and outperforming the market three years in a row. The manager's performance each year is independent of the performance in prior years. The probability of an excellent manager outperforming the market in any given year was given as 70%. Thus, the probability of an excellent manager outperforming the market three years in a row is 70% to the third power or 34.3% [P(O | E) = 0.7^3 = 0.343].

The denominator of Bayes' theorem is the unconditional probability of outperforming the market for three years in a row [P(O)]. This is calculated by finding the weighted average probability of both manager types outperforming the market three years in a row. If there is a 20% probability that a manager is excellent, then there is an 80% probability that a manager is average. The probabilities of the manager being excellent or average are used as the weights of 20% and 80%, respectively.

We are given that excellent managers are expected to outperform the market 70% of the time and we just determined that the probability of an excellent manager outperforming three years in a row is 34.3%. Similarly, the probability of an average manager outperforming the market three years in a row is determined by taking the 50% probability to the third power: (0.5^3 = 0.125).

With this information, we can solve for the unconditional probability of a random manager outperforming the market for three years in a row. This is computed as a weighted average of the probabilities of outperforming three years in a row for each type of manager:

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.7^3 \times 0.2) + (0.5^3 \times 0.8) \\ &= 0.0686 + 0.1 \\ &= 0.1686 \end{aligned}$$

We can now answer the question, "What is the probability that the new manager is excellent or average after outperforming the market three years in a row?" by incorporating the information required for Bayes' theorem.

Probability for excellent manager:

$$P(E \mid O) = \frac{P(O \mid E) \times P(E)}{P(O)} = \frac{0.343 \times 0.2}{0.1686} = 0.4069 = 40.7\%$$

Probability for average manager:

$$P(A \mid O) = \frac{P(O \mid A) \times P(A)}{P(O)} = \frac{0.125 \times 0.8}{0.1686} = 0.5931 = 59.3\%$$

The fact that the new manager outperformed the market three years in a row increases the probability that the new manager is an excellent manager from 20% to 40.7%. The probability that the new manager is an average manager goes down from 80% to 59.3%.

Professor's Note: The denominator is the same for both calculations as it is the unconditional probability of a random manager outperforming the market for three years in a row. In addition, the sum of the updated probabilities must still equal 100% (i.e., 40.7% + 59.3%), because the manager must be excellent or average.

Example: Bayes' theorem (3)

Using the same information given in the previous two examples, what is the probability that the new manager will beat the market next year, given that the new manager outperformed the market the last three years?

Answer:

This question is determined by finding the unconditional probability of the new manager outperforming the market. However, now we will use 40.7% as the weight for the probability that the manager is excellent and 59.3% as the weight for the probability that the manager is average:

$$\begin{aligned} P(O) &= P(O \mid E) \times P(E) + P(O \mid A) \times P(A) \\ &= (0.7 \times 0.407) + (0.5 \times 0.593) \\ &= 0.2849 + 0.2965 \\ &= 0.5814 \end{aligned}$$

Thus, the probability that the new manager will outperform the market next year is 58.14%.

Bayesian Approach vs. Frequentist Approach

LO 18.2: Compare the Bayesian approach to the frequentist approach.

The **frequentist approach** involves drawing conclusions from sample data based on the frequency of that data. For example, the approach suggests that the probability of a positive event will be 100% if the sample data consists of only observations that are positive events. The primary difference between the Bayesian approach and the frequentist approach is that the Bayesian approach is instead based on a prior belief regarding the probability of an event occurring.

In the previous examples, we began under the assumptions that excellent managers outperform the market 70% of the time, average managers outperform the market only 50% of the time, and only 20% of all managers are excellent. The Bayesian approach was used to update the probabilities that the new manager is either an excellent manager (updated from 20% to 40.7%) or an average manager (updated from 80% to 59.3%). These updated probabilities were based on the new information that the manager outperformed the market three years in a row. Next, under the Bayesian approach, the updated probabilities were used to determine the probability that the new manager outperforms the market next year. The Bayesian approach determined that there is a 58.14% probability that the new manager will outperform the market next year.

Conversely, under the frequentist approach there is a 100% probability that the new manager outperforms the market next year. There was a sample of three years with the manager outperforming the market each year (i.e., 3 out of 3 = 100%). The frequentist approach is simply based on the observed frequency of positive events occurring.

Obviously, the frequentist approach is questionable with a small sample size. It is difficult to believe that there is no way the new manager can underperform the market next year. However, individuals who apply the frequentist approach point out the weakness in relying on prior beliefs in the Bayesian approach. The Bayesian approach requires a beginning assumption regarding probabilities. In the prior examples, we assumed specific probabilities for a manager being excellent or average and specific probabilities related to the probability of outperforming the market for each type of manager. These prior assumptions are often based on a frequentist approach (i.e., number of events occurring during a sample period) or some other subjective analysis.

With small sample sizes, such as three years of historical performance, the Bayesian approach is often used in practice. With larger sample sizes, most analysts tend to use the frequentist approach. The frequentist approach is also often used because it is easier to implement and understand than the Bayesian approach.

Bayes' Theorem with Multiple States

LO 18.3: Apply Bayes' theorem to scenarios with more than two possible outcomes and calculate posterior probabilities.

In prior examples, we assumed there were only two possible outcomes where either a manager was excellent or average. Suppose now that we add another possible outcome where a manager is below average. The prior belief regarding the probabilities of a manager outperforming the market are 80% for an excellent manager, 50% for an average manager, and 20% for a below average manager. Furthermore, there is a 15% probability that a manager is excellent, a 55% probability that a manager is average, and a 30% probability that a manager is below average. These probabilities of manager performance are noted as follows:

$P(p = 0.8) = 15\%$
$P(p = 0.5) = 55\%$
$P(p = 0.2) = 30\%$

Example: Bayes' theorem with three outcomes

Suppose a new fund manager outperforms the market two years in a row. Given the manager performance probabilities above, how is Bayesian analysis applied to update prior expectations regarding the new manager's ability?

Answer:

The first step is to calculate the probability of each type of manager outperforming the market two years in a row, assuming the probability of outperforming the market is independent for each year. The probability that an excellent manager outperforms the market two years in a row is calculated by multiplying 80% by 80%. Thus, the probability that an excellent manager outperforms the market two years in a row is 64%.

$$P(O \mid p = 0.8) = 0.8^2 = 0.64$$

The probability that an average manager outperforms the market two years in a row is 25%.

$$P(O \mid p = 0.5) = 0.5^2 = 0.25$$

The probability that a below average manager outperforms the market two years in a row is 4%.

$$P(O \mid p = 0.2) = 0.2^2 = 0.04$$

Next, we calculate the unconditional probability of a random manager outperforming the market two years in a row. Previously, with two possible outcomes, we used a weighted average of probabilities to calculate unconditional probabilities. This weighted average is now updated to include a third possible outcome for below average managers. The weights are based on prior beliefs regarding the probabilities that a manager is excellent (15%), average (55%), or below average (30%). The following calculation determines the unconditional probability that a manager outperforms the market two years in a row.

$$P(O) = (15\% \times 64\%) + (55\% \times 25\%) + (30\% \times 4\%) = 0.096 + 0.1375 + 0.012 = 0.2455$$

We now use Bayes' theorem to update our beliefs that the manager is excellent, average, or below average by calculating the following **posterior probabilities:**

$$P(p = 0.8 \mid O) = \frac{P(O \mid p = 0.8) \times P(p = 0.8)}{P(O)} = \frac{0.64 \times 0.15}{0.2455} = 39.1\%$$

$$P(p = 0.5 \mid O) = \frac{P(O \mid p = 0.5) \times P(p = 0.5)}{P(O)} = \frac{0.25 \times 0.55}{0.2455} = 56.01\%$$

$$P(p = 0.2 \mid O) = \frac{P(O \mid p = 0.2) \times P(p = 0.2)}{P(O)} = \frac{0.04 \times 0.3}{0.2455} = 4.89\%$$

Notice that after the new manager outperforms the market for two consecutive years, the probability that the manager is an excellent manager more than doubles from 15% to 39.1%. In this example, the 15% is known as a prior belief, which is set *before* seeing the manager outperform the market two years in a row. The 39.1% is known as a posterior belief, which is set *after* seeing the manager outperform the market two years in a row. The updated probability that the manager is average goes up slightly from 55% to 56.01%, and the updated probability that the manager is below average goes down significantly from 30% to 4.89%. Notice that the updated probabilities still sum to 100% (= 39.1% + 56.01% + 4.89%).