WILEY

Reading 9: Common Probability Distributions

Learning Outcome Statements

- LOSs Covered
  - 9a, 9b, 9c, 9d, 9e, 9f, 9g, 9h, 9i, 9j, 9k, 9l, 9m, 9n, 9o, 9p, 9q
- LOSs Not Covered
  - None

Introduction

A random variable is a variable whose outcome cannot be predicted.

- A discrete random variable is one that can take on a countable number of values.
- The probability distribution of a random variable identifies the probability of each of the possible outcomes of a random variable.
- For a discrete random variable the probability of each possible outcome can be listed in the form of a probability function.
- A continuous random variable is one for which there are infinite possible outcomes, and therefore probabilities cannot be attached to specific outcomes.
- Therefore, we use a probability density function to interpret their probability structure.
- A cumulative distribution function expresses the probability that a random variable takes on a value less than or equal to a specific value.

Cumulative Distribution Function

The **cumulative distribution function** (or just **distribution function**) $F(x)$ gives the probability that X is less than or equal to a given value:

$$F(x) = P(X \leq x)$$

Example: Suppose that $p(x) = x^2/55,\ X = \{1, 2, 3, 4, 5\}$

$F(1)$	$P(X \leq 1)$	
$F(3)$	$P(X \leq 3)$	
$P(1 < X \leq 4)$	$F(4) - F(1)$	
$F(100)$	$P(X \leq 100)$	
$F(3.5)$	$P(X \leq 3.5)$	

Cumulative Distribution Function

The **cumulative distribution function** (or just **distribution function**) F(x) gives the probability that X is less than or equal to a given value:

$$F(x) = P(X \leq x)$$

Example: Suppose that $p(x) = x^2/55$, $X = \{1, 2, 3, 4, 5\}$

F(1)	$P(X \leq 1)$	1/55
F(3)	$P(X \leq 3)$	1/55 + 4/55 + 9/55 = 14/55
$P(1 < X \leq 4)$	F(4) – F(1)	4/55 + 9/55 + 16/55 = 29/55
F(100)	$P(X \leq 100)$	1
F(3.5)	$P(X \leq 3.5)$	14/55

Cumulative Distribution Function

Example (cont.): Suppose that $f(x) = x/12$, for X in [1, 5]; then

$$F(x) = \frac{x^2 - 1}{24}$$

$F(1)$	$P(X \leq 1)$	
$F(3)$	$P(X \leq 3)$	
$P(2 < X \leq 4)$	$F(4) - F(2)$	
$F(100)$	$P(X \leq 100)$	
$F(0)$	$P(X \leq 0)$	

Cumulative Distribution Function

Example (cont.): Suppose that $f(x) = x/12$, for X in [1, 5]; then

$$F(x) = \frac{x^2 - 1}{24}$$

$F(1)$	$P(X \le 1)$	$F(1) = \frac{1^2 - 1}{24} = 0$
$F(3)$	$P(X \le 3)$	$F(3) = \frac{3^2 - 1}{24} = \frac{1}{3}$
$P(2 < X \le 4)$	$F(4) - F(2)$	$P(2 < X \le 4) = F(4) - F(2) = \frac{4^2 - 1}{24} - \frac{2^2 - 1}{24} = \frac{1}{2}$
$F(100)$	$P(X \le 100)$	$F(100) = P(X \le 100) = 1$
$F(0)$	$P(X \le 0)$	$F(0) = P(X \le 0) = 0$

Common Discrete Distributions

A **discrete uniform distribution** has a finite number of possible outcomes, all equally likely.

Example: p(*x*) $= 0.25,\ X = \{4, 5, 6, 7\}$

A **Bernoulli random variable** has two outcomes, usually called "success" and "failure." Often, success is assigned a value of 1 and failure 0.

Example: p(success) = p(1) = 0.7, p(failure) = p(0) = 0.3

A **binomial random variable** measures the number of successes in *n* Bernoulli trials, where p(success) is constant across all of the trials.

Example: *x* is the number of successes in 10 Bernoulli trials, where p(success) $= 0.7;\ X = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

Binomial Distribution

For a binomial distribution of *n* Bernoulli trials where p(success) = *p*, the probability of *x* successes is:

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

Example: For *n* = 5 and *p* = 0.7,

P(0)		
P(1)		
P(2)		
P(3)		
P(4)		
P(5)		

Binomial Distribution

For a binomial distribution of *n* Bernoulli trials where p(success) = *p*, the probability of *x* successes is:

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

Example: For $n = 5$ and $p = 0.7$,

P(0)	$= {}_5C_0(0.7^0)(0.3^5)$	= 0.0024
P(1)	$= {}_5C_1(0.7^1)(0.3^4)$	= 0.0284
P(2)	$= {}_5C_2(0.7^2)(0.3^3)$	= 0.1323
P(3)	$= {}_5C_3(0.7^3)(0.3^2)$	= 0.3087
P(4)	$= {}_5C_4(0.7^4)(0.3^1)$	= 0.3602
P(5)	$= {}_5C_5(0.7^5)(0.3^0)$	= 0.1681

Binomial Tree

At each period there are two possible price movements: up and down.

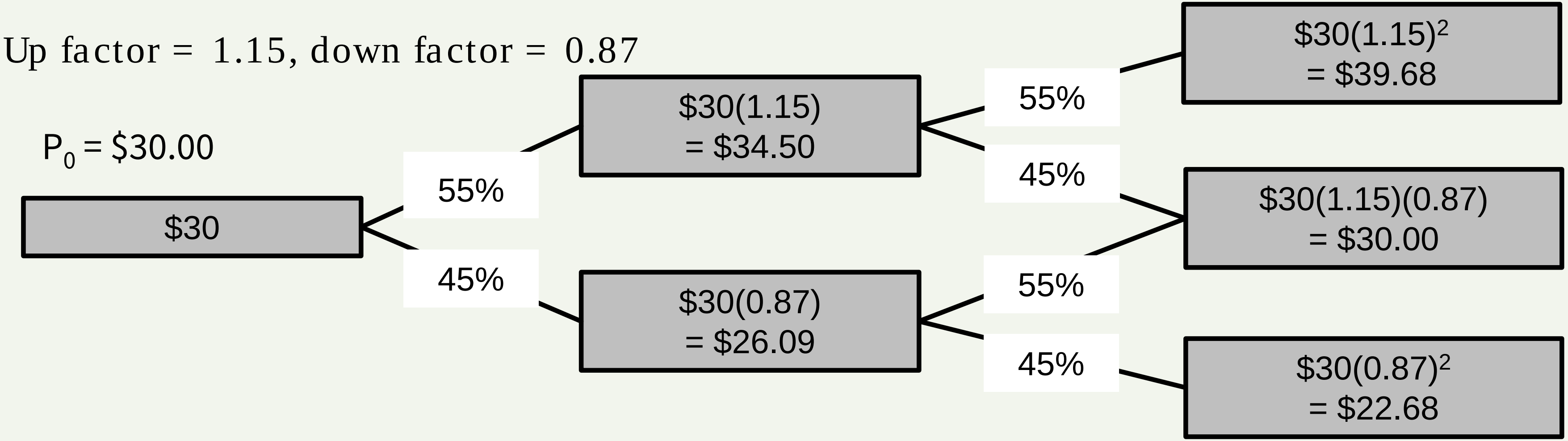
Example: P(up) = 55%, P(down) = 45%

$E(P_1)$ =

$E(P_2)$ =

Up factor = 1.15, down factor = 0.87

P_0 = $30.00



Binomial Tree

At each period there are two possible price movements: up and down.

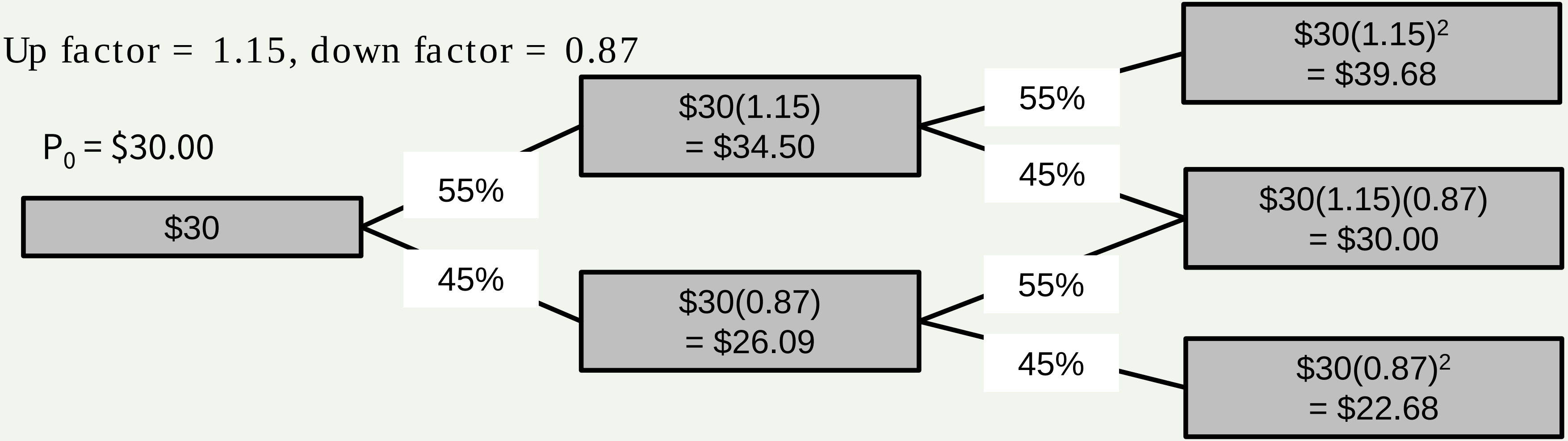
Example: P(up) = 55%, P(down) = 45%

$E(P_1) =$ 0.55($34.50) + 0.45($26.09) = $30.70

$E(P_2) =$ $31.45

Up factor = 1.15, down factor = 0.87

$P_0 = \$30.00$



Practice Question

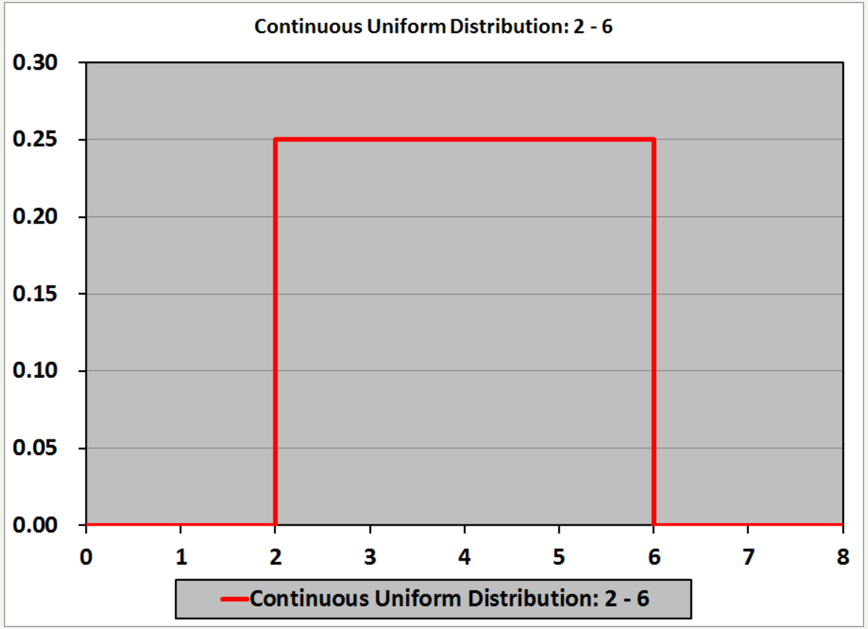
In a two-year binomial stock pricing tree, the current stock price is $20/share, there is a 70% chance of a price increase each year, and the expected return if a price increase occurs is 10%. The expected values of the stock prices one year and two years from today are *closest to*:

	One Year	Two Years
A.	$20.09	$20.24
B.	$20.09	$21.75
C.	$20.85	$21.75

Continuous Uniform Distribution

The **continuous uniform distribution** has a constant probability over the range of possible values.

Example: A continuous uniform distribution over the interval [2, 6] has this density function:



Continuous Uniform Distribution

For a continuous uniform distribution defined on [a, b], the width is

$$\text{width} = b - a$$

Because the height is constant, the probability of being in any range within [a, b] is just a ratio: the width of that range over the total width:

$$\mathrm{P}(c \leq X \leq d) = \frac{d - c}{b - a}$$

Example: Given a set of returns with a continuous uniform distribution on [–20%, 30%], what is the probability that a return is between 5% and 10%?

The probability is:

Continuous Uniform Distribution

For a continuous uniform distribution defined on [a, b], the width is

$$\text{width} = b - a$$

Because the height is constant, the probability of being in any range within [a, b] is just a ratio: the width of that range over the total width:

$$\mathrm{P}(c \leq X \leq d) = \frac{d - c}{b - a}$$

Example: Given a set of returns with a continuous uniform distribution on [–20%, 30%], what is the probability that a return is between 5% and 10%?

The probability is:

$$\frac{10\% - 5\%}{30\% - (-20\%)} = 0.10$$

Practice Question

An investment's annual returns follow a continuous uniform distribution from –10% to +40%. The probability that next year's return is between –5% and 15% is *closest to*:

A. 0.10

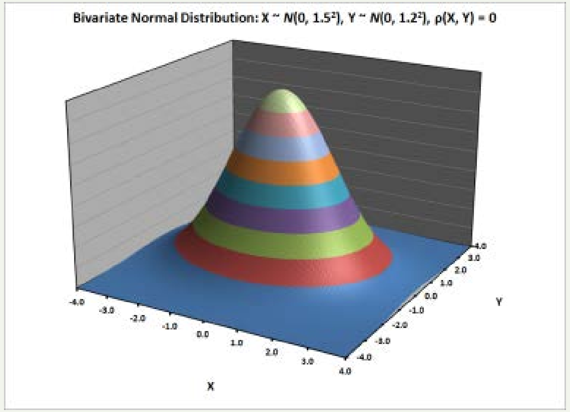
B. 0.20

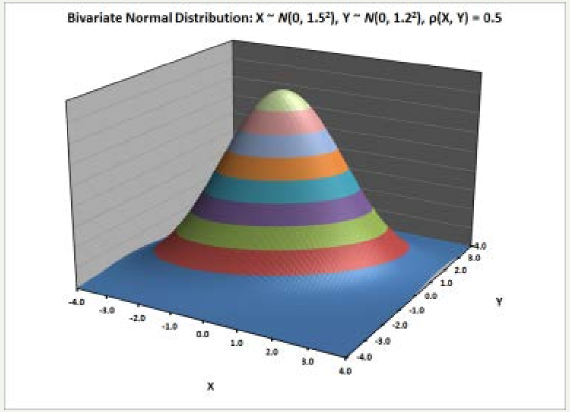
C. 0.40

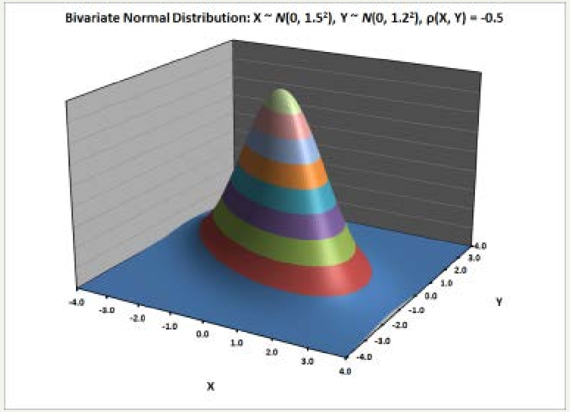
Properties of a Normal Distribution

- Completely described by its mean and variance (or mean and standard deviation): $X \sim N(\mu, \sigma^2)$
- Unimodal
- Skewness = 0: symmetric about its mean (so its mean, median, and mode are equal)
- Kurtosis = 3, so excess kurtosis = 0
- A linear combination of normal random variables is also normal; e.g., if X, Y, and Z are all normally distributed, then $W = 3X + 4Y - 6Z$ is also normally distributed
- Unbounded: defined for all values $(-\infty, +\infty)$
- Multivariate normal distribution
    - Means for each variable
    - Variances for each variable
    - Covariances for each distinct pair of variables

Multivariate Normal Distribution

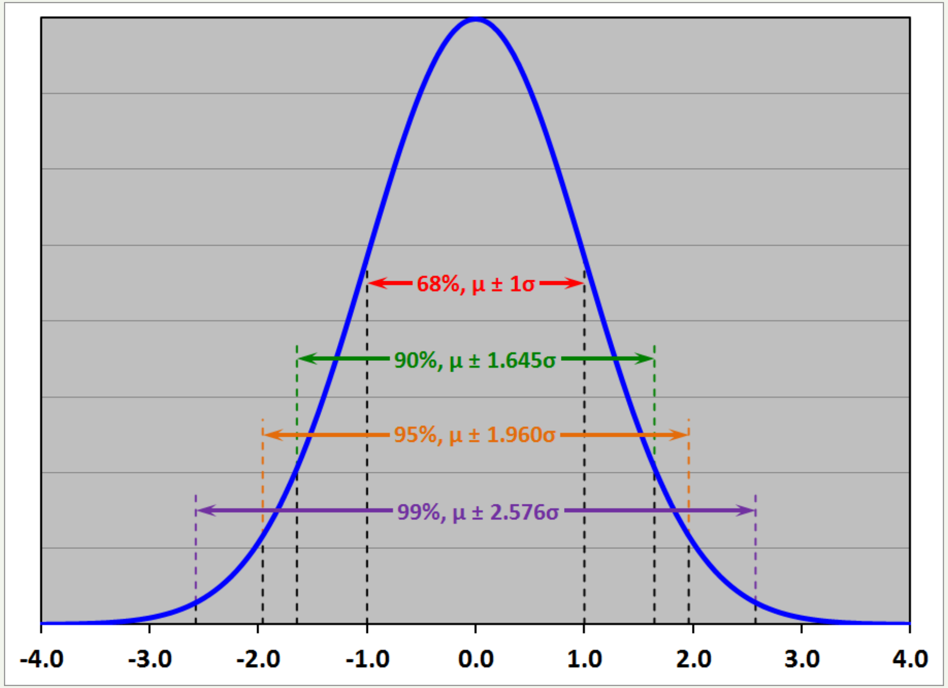






Normal Distribution: Confidence Intervals

For a given $k \geq 0$, ***all*** normal distributions have <u>the same percentage of their values</u> in the range $(\mu - k\sigma,\ \mu + k\sigma)$:



Confidence Intervals

Example

- Suppose that the monthly returns on a bond portfolio are normally distributed with a mean of 0.4% and a standard deviation of 0.3%. Construct a 95% confidence interval for next month's return.

Solution

- A 95% confidence interval encompasses the range:

Confidence Intervals

Example

- Suppose that the monthly returns on a bond portfolio are normally distributed with a mean of 0.4% and a standard deviation of 0.3%. Construct a 95% confidence interval for next month's return.

Solution

- A 95% confidence interval encompasses the range:

$$(\mu - 1.96\sigma,\ \mu + 1.96\sigma)$$

$$[0.4\% - 1.96(0.3\%),\ 0.4\% + 1.96(0.3\%)]$$

$$(-0.188\%,\ 0.988\%)$$

Standardizing a Normal Distribution

A standard normal distribution has mean = 0, standard deviation = 1.

To standardize any other normal distribution, subtract the mean (to move it to 0) and divide by the standard deviation (to change it to 1): If $X \sim N(\mu, \sigma^2)$, and

$$Z = \frac{X - \mu}{\sigma}$$

then $Z \sim N(0, 1)$.

Example: If $X \sim N(10, 49)$, what is the standard normal z-value corresponding to $x = 24$?

Standardizing a Normal Distribution

A standard normal distribution has mean = 0, standard deviation = 1.

To standardize any other normal distribution, subtract the mean (to move it to 0) and divide by the standard deviation (to change it to 1): If $X \sim N(\mu, \sigma^2)$, and

$$Z = \frac{X - \mu}{\sigma}$$

then $Z \sim N(0, 1)$.

Example: If $X \sim N(10, 49)$, what is the standard normal z-value corresponding to $x = 24$?

$$z = \frac{x - \mu}{\sigma} = \frac{24 - 10}{7} = 2$$

Normal Distribution: Probabilities

Because any normal distribution can be standardized, probabilities for any normal distribution can be computed using a table for a standard normal distribution.

Example: If $X \sim N(10, 49)$, what is the probability that $3 \leqslant X \leq 24$?

Normal Distribution: Probabilities

Because any normal distribution can be standardized, probabilities for any normal distribution can be computed using a table for a standard normal distribution.

Example: If $X \sim N(10, 49)$, what is the probability that $3 \leqslant X \leq 24$?

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{3 - 10}{7} = -1$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{24 - 10}{7} = 2$$

$P(-1 \leq Z \leq 2)$ $= ?$

$P(Z \leq 2)$ $= 0.9772$

$P(Z < -1)$ $= 0.1587$

$P(-1 \leq Z \leq 2)$ $= 0.9772 - 0.1587 = 81.85\%$

Z	0.00	0.01
–0.9	0.1841	0.1814
–1.0	0.1587	0.1562
–1.1	0.1357	0.1335
1.9	0.9713	0.9719
2.0	0.9772	0.9778
2.1	0.9821	0.9826

Practice Question

Given that $X \sim N(5, 10)$, the standard normal range corresponding to the range $0 \leqslant X \leq 15$ is *closest to*:

A. $-4.472 \leq Z \leq 2.236$

B. $-1.581 \leq Z \leq 3.162$

C. $-0.500 \leq Z \leq 1.000$

Shortfall Risk

Shortfall risk is the risk that the value of a portfolio will fall below an acceptable level over a given time horizon.

The Safety-First ratio is a measure of shortfall risk:

$$\text{SFRatio} = \frac{\mathrm{E}(R_P) - R_L}{\sigma_P}$$

where R_L is the threshold (minimum acceptable) return.

Shortfall Risk

Example: If the expected return is 5%, the minimum acceptable return is 4%, and the standard deviation of returns is 3%, what is the Safety-First ratio?

Shortfall Risk

Example: If the expected return is 5%, the minimum acceptable return is 4%, and the standard deviation of returns is 3%, what is the Safety-First ratio?

$$SFRatio = \frac{\mathrm{E}(R_P) - R_L}{\sigma_P} = \frac{5\% - 4\%}{3\%} = 0.333$$

Shortfall Risk

Example: Suppose that you have three investments: ABC, DEF, GHI. The details are:

Investment	$E(R_P)$	σ_P	
ABC	5%	4%	
DEF	8%	6%	
GHI	10%	9%	

If the minimum acceptable return is 2%, which investment is best, according to the Safety-First criterion?

Shortfall Risk

Example: Suppose that you have three investments: ABC, DEF, GHI. The details are:

Investment	$E(R_P)$	σ_P	$\text{SFRatio} = \dfrac{E(R_P) - R_L}{\sigma_P}$
ABC	5%	4%	
DEF	8%	6%	
GHI	10%	9%	

If the minimum acceptable return is 2%, which investment is best, according to the Safety-First criterion?

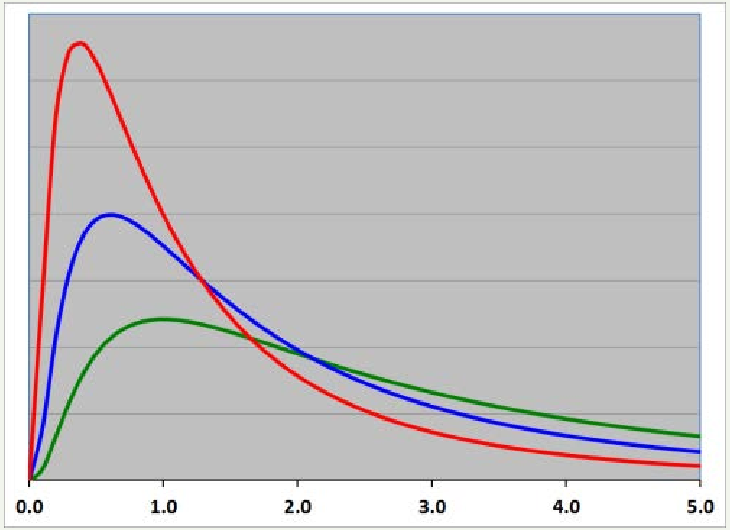
Investment DEF is the best.

Lognormal Distribution

X follows a **lognormal distribution** if ln(*X*) follows a normal distribution; *the log of X is normal*.

A lognormal distribution is bounded below by zero, and is skewed to the right, as shown:

If the continuously compounded return of an investment is normally distributed, the investment's price is lognormally distributed.



Continuous Compounding

If r_{nom} is the nominal annual interest rate compounded m times per year, then the effective annual rate (EAR) satisfies:

$$1 + \text{EAR} = \left(1 + \frac{r_{nom}}{m}\right)^m$$

As the number of compounding periods increases without bound (**continuous compounding**),

$$1 + \text{EAR} = e^{r_{nom}}$$

For continuous compounding, the nominal rate is called the continuously compounded rate. Taking natural logarithms:

$$r_{nom} = \ln(1 + EAR)$$

Continuous Compounding

Examples:

- An effective annual rate of 5% is equivalent to a continuously compounded (annual) rate of 4.879% because $\ln(1.05) = 0.04879$, or $e^{0.04879} = 1.05$.

- A \$100 investment grows to \$110 in one year. Its EAY is $(\$110/\$100) - 1 = 10\%$. Its continuously compounded (annual) rate of return is $\ln(\$110/\$100) = 9.531\%$.

- An investment that has a continuously compounded return of 7% has a holding period yield of 7.251% because $e^{0.07} - 1 = 0.07251$.

Monte Carlo Simulation

A **Monte Carlo simulation** is a computer model used to generate probability distributions that are too difficult to generate analytically.

Uses: Calculate derivative prices, VAR, NPV, etc.

Modus operandi:

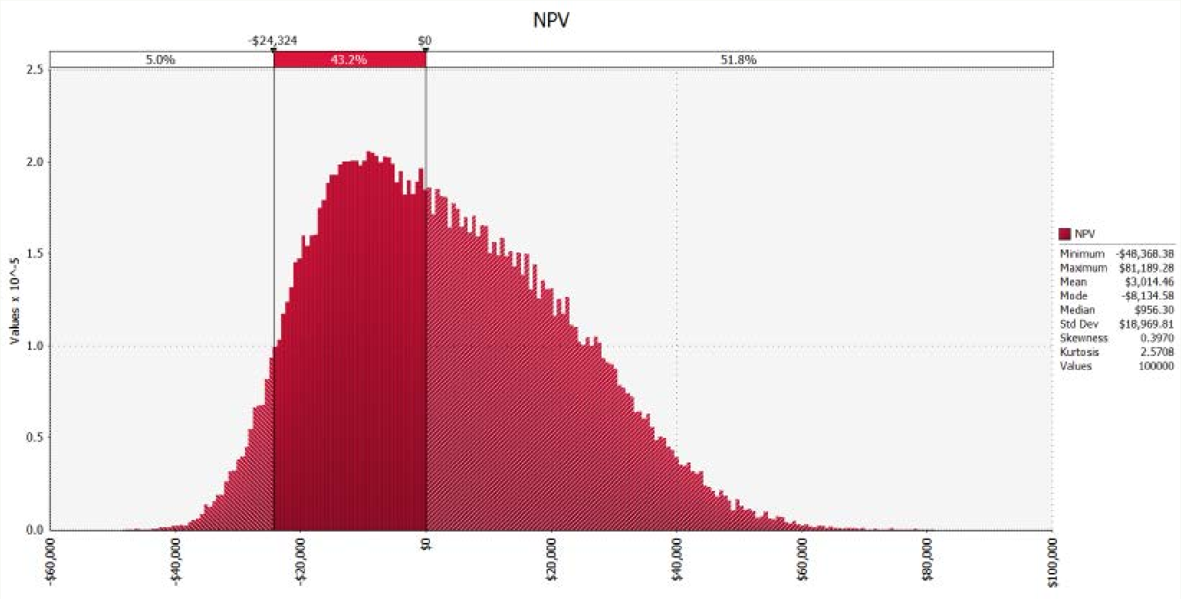
- Create a model for the desired value (e.g., NPV as a function of future cash flows and discount rate).
- Assign probability distributions to the input variables in the model (e.g., normal distributions for the cash flows, uniform distribution for the discount rate); optionally, specify correlations of input variables.
- Select random values for the input variables according to their probability distributions and calculate the desired value.
- Repeat the selection/calculation process hundreds or thousands of times and tabulate the results of each iteration.
- Compute statistics (e.g., mean, median, mode, standard deviation, skewness, kurtosis, and so on) for the results of the simulation.

Monte Carlo Simulation

Example: NPV

Without Monte Carlo simulation, NPV = +$1,527: Do the project.

With Monte Carlo simulation, P(NPV > 0) = 51.8%: Maybe not high enough to justify the project.



Historical Simulation

Historical simulation is similar to Monte Carlo simulation:

- Create a model for the value sought.
- Select random values for the input variables according to their probability distributions and calculate the quantity of interest.
- Repeat the selection/calculation process hundreds or thousands of times and tabulate the results of each iteration.
- Compute statistics (e.g., mean, median, mode, standard deviation, skewness, kurtosis, and so on) for the results of the simulation.

Historical simulation differs from Monte Carlo simulation in that it uses probability distributions based on historical numbers, not those chosen by the analyst.

Advantage: Easier to determine the probability distributions to use

Disadvantages: Less flexible, cannot do “what if” analysis, historical data may not include possible extreme values

WILEY

Practice Questions with Solutions

Practice Question

In a two-year binomial stock pricing tree, the current stock price is $20/share, there is a 70% chance of a price increase each year, and the expected return if a price increase occurs is 10%. The expected values of the stock prices one year and two years from today are *closest to*:

	One Year	Two Years
A.	$20.09	$20.24
B.	$20.09	$21.75
C.	$20.85	$21.75

Correct answer: C. $20.85 $21.75

Up factor = 1.10, down factor = 1/1.10 = 0.909

Expected return = [0.70(1.10) + 0.30(0.909)] – 1 = 4.27%

$20(1.0427) = $20.85, $20.85(1.0427) = $21.75

Practice Question

An investment's annual returns follow a continuous uniform distribution from –10% to +40%. The probability that next year's return is between –5% and 15% is *closest to*:

A. 0.10

B. 0.20

C. 0.40

Correct answer: C. 0.40

$$P(-5\% \leq \text{return} \leq 15\%)$$

$$= [15\% - (-5\%)]/[40\% - (-10\%)]$$

$$= 20\%/50\% = 0.40$$

Practice Question

Given that $X \sim N(5, 10)$, the standard normal range corresponding to the range $0 \leqslant X \leq 15$ is *closest to*:

A. $-4.472 \leq Z \leq 2.236$

B. $-1.581 \leq Z \leq 3.162$

C. $-0.500 \leq Z \leq 1.000$

Correct answer: B. $-1.581 \leq Z \leq 3.162$

$$z_1 = (x_1 - \mu)/\sigma = (0 - 5)/\sqrt{10} = -1.581$$

$$z_2 = (x_2 - \mu)/\sigma = (15 - 5)/\sqrt{10} = 6.162$$