Discrete Random Variables

Eco 2470: Economic Statistics

(Chapter 7)

Random Variables…

A ***random variable*** is a function or rule that assigns a number to all possible random outcomes of an expirement.

Alternatively, the ***value*** of a random variable is a numerical event.

Instead of talking about the coin flipping event as {heads, tails} think of it as

"the number of heads when flipping a coin"

{1, 0} ← *(numerical events)*

Random Variable Examples

- Flip a Coin X= 1 (if heads) X = 0 (if tails).
  (possible outcomes: heads or tails)
- Roll a die: X is the number on the face of the die.
  (possible outcomes: rolls of one through six)
- Roll two dice: X is the sum of the numbers on both dies.
  (possible outcomes: sums of 2,3,…,12)
- Roll two dice: X is the second roll minus the first roll
  (possible outcomes: differences of -5,-4,…,0,1,..5)

Two Types of Random Variables…

Discrete Random Variable

– one that takes on a ***countable*** number of values

– E.g. values on the roll of dice: 2, 3, 4, …, 12

Continuous Random Variable

– one whose values are ***not discrete***, not countable

– E.g. time (30.1 minutes? 30.10000001 minutes?)

Analogy:

Integers are Discrete, while Real Numbers are Continuous

Probability Distributions…

A ***probability distribution*** is a table, formula, or graph that describes the values of a random variable and the probability associated with these values.

Since we're describing a **random variable** (which can be discrete or continuous) we have two types of probability distributions:

- Discrete Probability Distribution, (this chapter) and
- Continuous Probability Distribution (Chapter 8)

Probability Notation…

An upper-case letter will represent the ***name*** of the random variable, usually **X**.

Its lower-case counterpart will represent the ***value*** of the random variable.

The probability that the random variable **X** will equal x is:

$$P(\mathbf{X} = x)$$

or more simply

$$P(x)$$

Discrete Probability Distributions…

The probabilities of the values of a ***discrete random variable*** may be derived by means of probability tools (e.g trees) so long as these two conditions apply:

$$1.\ 0 \le \mathrm{P}(\mathrm{x}) \le 1 \ \textit{for all } x$$

$$2.\ \sum_{all\ x_i} P(x) = 1$$

Discrete Probability Distribution: Examples

- Dummy Variable: Flip Coin: X = 1 (heads), X =0 (tails)

x	P(X=x)
0	0.5
1	0.5

$P(x) = \frac{1}{2}$ for $x = 0,1$

- Roll a single die: X=number on the face of the die

x	P(X=x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

$P(x) = 1/6$ for $x = 1,2,3,4,5,6$

Discrete Probability Distribution: Examples

- Roll 2 dice: X is the sum of the two faces.

x	P(x)=P(X=x)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

$P(x) = (6 - |7-x|)/36$ for $x = 2,3,\ldots, 12$

Relative frequency interpretation

- Discrete Probability Distributions can be interpreted as the relative frequency of the population
- Example: X = 1 (heads), 0 (tails)

  If we tossed a billion coins we would expect

outcome	relative frequency	x	P(x)
Tails	½	0	½
Heads	½	1	½

- Likewise if we tossed a billions dice X = face of die

outcome	relative frequency	x	P(x)
Roll a 1	1/6	1	1/6
Roll a 2	1/6	2	1/6 etc

Example 7.1 *Statistical Abstract of the U.S*

Number of Persons x	Number of Households (millions)	**P(x)**
1	31.1	31.1/116.0 = .268
2	38.6	38.6/116.0 = .333
3	18.8	18.8/116.0 = .162
4	16.2	16.2/116.0 = .140
5	7.2	7.2/116.0 = .062
6	2.7	2.7/116.0 = .023
7 or more	1.4	1.4/116.0 = .012
Total	116.0	1.000

Hidden assumption: Treat Statistical Abstract as population.

Using the distribution to calculate probabilities

x	P(x)=P(X=x)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Roll 2 dice. What is the probability that they sum to less than 5?

$$P(X < 5) = P(X=2) + P(X=3) + P(X=4)$$
$$= 1/36 + 2/36 + 3/36 = 6/36$$

What is the probability of rolling more than 4, but less than 8.

$$P(4 < X < 8) = P(X=5) + P(X=6) + P(X=7)$$
$$= 4/36 + 5/36 + 6/36 = 15/36$$

Example 7.1

E.g. what is the probability there are 4 or more persons in any given household?

x	P(x)
1	.268
2	.333
3	.162
4	.140
5	.062
6	.023
7 or more	.012

$$P(\mathbf{X} \geq 4) = P(4) + P(5) + P(6) + P(\text{ 7 or more})$$
$$= .140 + .062 + .023 + .012 = .237$$

Cumulative Distribution Function (CDF)

- Probability Distribution: $P(X=x)$
  - the probability that X takes a certain value (x)
  - Interpret as relative frequency of the population
- Cumulative Distribution Function: $P(X \leq x)$
  - the probability that X is less than or equal to a certain value (x)
  - Interpret as cumulative relative frequency of the population

CDF - Examples

- Heads and Tails

x	P(X=x)	P(X ≤ x)
0	½	½
1	½	1

- Roll a Single Die

x	P(X=x)	P(X ≤ x)
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6

CDF of Example 7.1 (Statistical Abstract)

x	P(X=x)	P(X≤x)
1	0.268	0.268
2	0.333	0.601
3	0.162	0.763
4	0.140	0.903
5	0.062	0.965
6	0.023	0.988
7	0.012	1.000

- About 60 Percent of households have two or fewer people.
- About 90 percent of household have four or fewer people.

Example 7.1: Calculate probability from CDF

Example 7.1: Suppose we are only provided the CDF. Find the probability of a household with exactly 4 members.

x	P(X ≤ x)
1	0.268
2	0.601
3	0.763
4	0.903
5	0.965
6	0.988
7	1.000

Find $P(4) = P(X=4)$.

Solution (Work Backwards):

$$P(X=4) = P(X \leq 4) - P(X \leq 3)$$
$$= 0.903 - 0.763$$
$$= 0.140$$

Calculate Population Percentile from CDF of Ex. 7.1

- What is the 60th Population Percentile of Household size?

x	$P(X \leq x)$
1	0.268
2	0.601
3	0.763
4	0.903
5	0.965
6	0.988
7	1.000

Restate the problem mathematically:

Find x for which $P(X \leq x) = 0.60$

Answer: $P(X \leq 2) = 0.601$

So x =2 households is approximately 60th Percentile

The Population Mean

- Recall **Sample Mean**:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$ where **n** is the size of the sample

- **Population mean** (Denoted E[X] or Greek letter "mu")

E[X] = $$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$$ where **N** the population size

and the x_i are all the individuals in the population

Another Formula for Population Mean

Formula based on Probabilities:

$$E(X) = \mu = \sum_{all\ x} xP(x)$$

- Where the x's are all the values that X can take.
- Interpretation: A weighted average, where the probabilities are the weights.
- Which one is right? Both: they give the same answer as illustrated by the next example using Example 7.1 data.
- The intuition for why they are the same is that the probability weights can be interpreted as population relative frequencies.

Example: Both Population Mean Formulas Give Same Answer

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i = \text{Population Mean Household Size} =$$

$$\left[1(31.1) + 2(38.6) + 3(18.8) + 4(16.2) + 5(7.2) + 6(2.7) + 7(1.4)\right]/116$$

$$= 1(31.1/116) + 2(38.6/116) + 3(18.8/116) + 4(16.2/116) + 5(7.2/16) + 6(2.7/116) + (1.4/116)$$

$$= 1(0.268) + 2(0.333) + 3(0.162) + 4(0.140) + 5(0.062) + 6(0.023) + 7(0.012)$$

$$= 1P(X=1) + 2P(X=2) + 3P(X=3) + 4P(X=4) + 5P(X=5) + 6P(X=6) + 7P(X=7)$$

$$= \sum_{\text{all } x} xP(X=x)$$

x	Number of households (millions)	N	P(X=x)
1	31.1	116	0.268 = 31.1/116
2	38.6	116	0.333 = 38.6/116
3	18.8	116	0.162 = 18.8/116
4	16.2	116	0.140 = 16.2/116
5	7.2	116	0.062 = 7.2/116
6	2.7	116	0.023 = 2.7/116
7	1.4	116	0.012 = 1.4/116
Total	116.000		1.000

When to use which formula

- Use the formula based on the population mean if given information about all the individuals in the population

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

- Use the formula based on probabilities if given a list of possible outcomes and their probabilities.

$$E(X) = \mu = \sum_{all\ x} xP(x)$$

Examples: when to use which formula

Suppose that if the entire population of 34,930,000 Canadians were laid end to end it would create a human chain of 63,879,984 meters in length. Find the population mean Canadian height:

N= 34,930,000 $\sum_{i=1}^{N} x_i = 63,879,984$ meters

use pop. Mean version of formula

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = 63,879,984 / 34,930,000 = 1.8288$$

Examples (con't): which formula to use

- A fair six sided die will be rolled and the face of the die recorded. Find the population mean.

We have been (implicitly) give the probability distribution:

x	P(X=x)
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Use probability version of formula

$$E(X) = \mu = \sum_{all\ x} xP(x)$$

=1(1/6) + 2(1/6) + 3(1/6) + 4(1/6)+ 5(1/6)+6(1/6)

= (1+2+3+4+5+6)/6 = 21/6 = 3.5

Forecast Interpretation of the Pop. Mean

- E[X] --- The "E" stands for **Expectation**
- X is a random variable – we do not know ahead of time what value it will take.
- But we can still form an expectation, which can be thought of us as our best guess or forecast.
- The population mean is a natural forecast because it falls in roughly in the middle of the all possible values that X can take.
- Of course, $E[X] \neq X$ in general. Forecasts are rarely that accurate.

Laws of Expected Value…

Let c be a constant and X and Y be random variables:

1. $E(c) = c$ e.g. $E[5] = 5$

   Our best guess of a constant is the constant – of course!

2. $E(X + c) = E(X) + c$

   If we add a constant to X, we add same constant to our forecast of X. Note, that we "pull" the constant out of the expected value.

3. $E(cX) = cE(X)$

   If we multiply X by constant, should multipy our forecast of X by same constant. Again we "pull" the constant out.

3. $E[X+Y] = E[X] + E[Y]$

   Our best guess of X+Y is our best guess of X plus our best guess of Y. <u>The expectation of the sum equals sum of the expectation.</u>

Example 7.4: Using rules of expectation

Monthly sales have a mean of $25,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of $6,000.

Find the mean monthly profit.

Profit = 0.30(sales) – 6,000

E(Profit) =E[.30(Sales) – 6,000]

=E[.30(Sales)] – 6,000 [by rule #2]

=.30E(Sales) – 6,000 [by rule #3]

=.30(25,000) – 6,000 = 1,500

Thus, the mean monthly profit is **$1,500**

Example: Using the rules of expectation

Example: ¾ of a client's retirement portfolio is invested in the TSX and ¼ is invested in the S&P 500. Your firm's analyst indicates an expected return of 7 percent for the TSX and 5 percent for the S&P 500 (in CAD). Provide the expected return of your client's portfolio.

Solution:

Let r_{TSX} = return on TSX = 7%

$r_{S\&P}$ = return on S&P500 =5%

r_P = return on clients porfolio

$$r_P = \tfrac{3}{4}\, r_{TSX} + \tfrac{1}{4}\, r_{S\&P}$$

$$E[\, r_P] = E[\tfrac{3}{4}\, r_{TSX} + \tfrac{1}{4}\, r_{S\&P}]$$

$$= E[\tfrac{3}{4}\, r_{TSX}\,] + E[\tfrac{1}{4}\, r_{S\&P}] \qquad \text{[by rule \#4]}$$

$$= \tfrac{3}{4}\, E[r_{TSX}\,] + \tfrac{1}{4}\, E[r_{S\&P}] \qquad \text{[by rule \#3]}$$

$$= \tfrac{3}{4}\,[7\%] + \tfrac{1}{4}\,[5\%] = 6.5\ \%$$

Expectations of Functions of Random Variables

Let X be a Random Variable and **g** be a function.

We are interested in E[g(X)]

•Interpret E[g(X)] population mean of g(X)

•Interpret E[g(X)] as forecast of g(X)

•Interpret E[g(X)] as weighted average of g(X), using the probabilities of X as population weights:

$$E[g(X)] = \Sigma_{\text{all } x}\, g(x)P(X=x)$$

•Unfortunately the "rules" don't help us unless the function g is linear.

$$E[g(X)] \neq g(E[X]) \text{ unless g is linear.}$$

Example: Expected Utility

- Suppose your utility from rolling one fair six-sided die is given by $U(X) = X^2-6$ where X is number on the face of the die. Find the expected utility from rolling the die.

x	P(X=x)	$U(X)=x^2-6$
1	1/6	-5
2	1/6	-2
3	1/6	3
4	1/6	10
5	1/6	19
6	1/6	30
	Total	55
	Average	9.167

U() takes place of g() in formula.

$$E[U(X)] = \Sigma_{\text{all } x} U(x)P(X=x)$$

$$= 1/6(-5)+1/6(-2)+1/6(3) + 1/6(10) + 1/6(19) +1/6(30)$$

$$= 1/6\ (-5-2+3+10+19+30)$$

$$= 1/6\ (55)\ = 9.167$$

Population Variance

Suppose that we could observe the entire population:

$$x_1, x_2, \ldots, x_N$$

then the population variance could be calculated analogously to the sample variance as:

$$V(X) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

Population Variance as an Expectation

The population can equivalently be defined by the following expectation:

$$\sigma^2 = V(X) = E\left[(X - \mu)^2\right]$$

This an expectation of a function of X in the form:

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

We can express VAR(X) as a probability weighted sum:

$$\sigma^2 = V(X) = \sum_{all\ x} g(x)P(X = x) = \sum_{all\ x} (x - \mu)^2 P(X = x)$$

where we sum over all possible outcomes x.

Using the two definitions of variance

- Similar to the population mean, the variance has two definitions: one based the population and one based on probability weights applied to all possible outcomes
- Again, the two are equivalent. However, this time we will just accept this without working through the intuition.
- Which definition is useful depends on what information you are given.

Example Variance of Coin Toss

Example: One fair coin is tossed. X = 1 (heads) or 0 (tails). What are E(X) and V(X)?

Note: Given probability information (not a population). So use probability weights formulas:

	x	$P(x)$	$xP(x)$	μ	$x-\mu$	$(x-\mu)^2$	$(x-\mu)^2 P(x)$
Tails	0	0.5	0	0.5	-0.5	0.25	0.125
Heads	1	0.5	0.5	0.5	0.5	0.25	0.125
Total	$\mu = \sum_{all\ x} xP(x) =$		0.5	$V(X) = \sum_{all\ x} (x-\mu)^2 P(x) =$			0.25

Laws of Variance…

$V(c) = 0$

The variance of a constant (c) is zero.

$V(X + c) = V(X)$

Constants don't add any variation.

$V(cX) = c^2V(X)$

Don't forget to the square the constant when taking it out of variance.

Example 7.4…

Monthly sales have a mean of \$25,000 and a standard deviation of \$4,000. Profits are calculated by multiplying sales by 30% and subtracting fixed costs of \$6,000.

Find the standard deviation of monthly profits.

2) The ***variance*** of profit is = V(Profit)

=V[.30(Sales) – 6,000]

=V[.30(Sales)] [by rule #2]

$=(.30)^2$V(Sales) [by rule #3]

$=(.30)^2(16{,}000{,}000) = 1{,}440{,}000$

Again, ***standard deviation*** is the square root of ***variance***,

so standard deviation of Profit = $(1{,}440{,}000)^{1/2}$ = **\$1,200**

Bivariate Distributions…

Up to now, we have looked at ***univariate distributions***, i.e. probability distributions in **one** variable.

bivariate distributions are probabilities of combinations of **two** variables.

Bivariate probability distributions are also called ***joint probability***. A joint probability distribution of X and Y is a table or formula that lists the joint probabilities for all ***pairs*** of values x and y, and is denoted P(x,y).

$$P(x,y) = P(X=x \text{ and } Y=y)$$

Example 7.5…

Xavier and Yvette are real estate agents. Let X denote the number of houses that Xavier will sell in a month and let Y denote the number of houses Yvette will sell in a month. An analysis of their past monthly performances has the following joint probabilities (bivariate probability distribution).

		x = 0	x = 1	x = 2	
y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.00

Marginal Probabilities…

As before, we can calculate the ***marginal probabilities*** by summing across rows and down columns to determine the probabilities of X and Y individually:

		x = 0	x = 1	x = 2	
y	0	0.12	0.42	0.06	0.6
	1	0.21	0.06	0.03	0.3
	2	0.07	0.02	0.01	0.1
		0.4	0.5	0.1	1.00

x	P(x)
0	0.4
1	0.5
2	0.1

y	P(y)
0	0.6
1	0.3
2	0.1

E.g the probability that Xavier sells 1 house = P(X=1) =0.50

Describing the Bivariate Distribution…

We can describe the mean, variance, and standard deviation of ***each variable*** in a bivariate distribution by working with the ***marginal probabilities***…

x	P(x)
0	0.4
1	0.5
2	0.1

E(X) =	**0.7**
V(X) =	**0.41**
σ_x =	**0.64**

y	P(y)
0	0.6
1	0.3
2	0.1

E(Y) =	**0.5**
V(Y) =	**0.45**
σ_y =	**0.67**

same formulae as for univariate distributions…

Expectations for Bivariate Distributions

- Recall that:

$$E\left[g(x)\right] = \sum_{all\ \text{x}} g(x)P(x)$$

- What if g is a function of both x and y: g(x,y)?

$$E\left[g(x,y)\right] = \sum_{all\ \text{x}} \sum_{all\ \text{y}} g(x,y)P(x,y)$$

where $P(X,Y) = P(X = x \text{ and } Y = y)$ the joint probability

- We can use this to define the population covariance…

Population Covariance:

The ***covariance*** of two discrete variables is defined as:

$$COV(X,Y) = E\left[(X-\mu_x)(Y-\mu_y)\right] = \sum_{all\ x}\sum_{all\ y}(x-\mu_x)(y-\mu_y)P(x,y)$$

or alternatively using this shortcut method:

$$COV(X,Y) = E[XY] - \mu_x\mu_y = \sum_{\text{all } x}\sum_{\text{all } y} xyP(x,y) - \mu_x\mu_y$$

where $P(x,y) = P(X = x \text{ and } Y = y)$ the joint probability

Population Covariance:

- There is also a version based on observation of the entire population:

$$x_1, x_2, \ldots, x_N$$

given by:

$$COV(X,Y) = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \mu_x\right)\left(y_i - \mu_y\right)$$

Recap: Two Formulas For Pop Covariance

- Note: Just as for the population mean and variance, we have two formulas for the population covariance:
  - –A version based on the individuals in the population, which looks similar to the sample version (see chapter 4)
  - –A version based on a probability weighted average or expectation (see chapter 7)
- Although they look different the two formulas again lead to the same answer.
- Which we use is a matter of convenience:
  - –If you have the population or information about it, use the population based version.
  - –If you have information on the probabilities – use the probability based version.

Example: Pay & Job Satisfaction I

- A company is interested in the relationship between its employees' annual wage (X) and their job satisfaction (Y). They survey the entire population of all their employees and report:

$$\frac{1}{N}\sum_{i=1}^{N} x_i = 50 \text{ (in thousands)} \qquad \frac{1}{N}\sum_{i=1}^{N} y_i = 7 \text{ (on scale of 1 to 10)}$$

$$\sum_{i=1}^{N}(x_i - 50)(y_i - 7) = 15{,}000 \qquad N = 10{,}000$$

Find the population covariance between wage and job satisfaction.

Sol: This population info, not prob info. So use population version of formula to obtain.

$$COV(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y) = \frac{1}{10{,}000}\sum_{i=1}^{N}(x_i - 50)(y_i - 7) = \frac{15{,}000}{10{,}000} = 1.5$$

Example: Pay and Job Satisfaction II

An employee is to be drawn at random. Let X=0 if the employee is poorly paid and X=1 if she is well paid. Let Y= 0 if her job satisfaction low and Y=1 if her job satisfaction is high. Using following means & joint distribution, find COV(X,Y)

Y

X

?	0	1
0	0.3	0.2
1	0.2	0.3

$$m_x = 0.5$$

$$m_y = 0.5$$

Solution: Next slide

Example, Continued

	y = 0	y = 1
x = 0	0.3	0.2
x = 1	0.2	0.3

Solution: Use probability version of the formula

x	y	$x - \mu_x$	$y - \mu_y$	$P(x,y)$	$(x - \mu_x)(y - \mu_x)P(x,y)$
0	0	-0.5	-0.5	0.3	0.075
0	1	-0.5	0.5	0.2	-0.05
1	0	0.5	-0.5	0.2	-0.05
1	1	0.5	0.5	0.3	0.075

$$COV(X,Y) = \sum_{all\ x} \sum_{\text{all } y} (x - \mu_x)(y - \mu_x)P(x,y)$$

$$= 0.075 - 0.05 - 0.05 + 0.075 = 0.05$$

Population Coefficient of Correlation…

The coefficient of correlation is calculated in the same way as described earlier…

$$\rho = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

Laws…

We can derive laws of expected value and variance for the sum of two variables as follows…

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

$V(X + Y) = V(X) + V(Y) + 2COV(X, Y)$

IF X and Y are independent, $COV(X, Y) = 0$ and thus

$V(X + Y) = V(X) + V(Y)$ for X and Y independent

More Laws…

We also have laws of covariance…

- $COV(X,X) = V(X)$
- $COV(X+c, Y+d) = COV(X,Y)$
- $COV(cX,dY) = cdCOV(X,Y)$

Again, If X and Y are independent, $COV(X, Y) = 0$

Example: Portfolio Diversification and Asset Allocation

Consider an investor who forms a portfolio, consisting of only two stocks, by investing $4,000 in one stock and $6,000 in a second stock. Suppose that the results after 1 year are:

One-Year Results

Stock	Initial Investment	Value of Investment After One Year	Rate of Return on Investment
1	$4,000	$5,000	R_1 = .25 (25%)
2	$6,000	$5,400	R_2 =-.10 (-10%)
Total	$10,000	$10,400	R_p = .04 (4%)

OR

$$R_p = w_1R_1 + w_2R_2 = (.4)(.25) + (.6)(-.10) = .04$$

Portfolio Diversification and Asset Allocation

Mean and Variance of a Portfolio of Two Stocks

$$E(R_p) = w_1\,E(R_1) + w_2\,E(R_2)$$

$$V(R_p) = w_1^2\,V(R_1) + w_2^2\,V(R_2) + 2w_1w_2\,COV(R_1, R_2)$$

$$= w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2 r s_1 s_2$$

where w_1 and w_2 are the proportions or weights of investments 1 and 2, $E(R_1)$ and $E(R_2)$ are their expected values, σ_1 and σ_2 are their standard deviations, and r is the coefficient of correlation

Example 7.8

An investor forms a portfolio by putting 25% of his money into McDonald's stock and 75% into Cisco Systems stock. Suppose that the expected returns will be 8% and 15%, respectively, with standard deviations 12% and 22%, respectively.

a Find the expected return on the portfolio.

b Compute the standard deviation of the returns on the portfolio assuming that

(i) the two stocks' returns are perfectly positively correlated

(ii) the coefficient of correlation is .5

(iii) the two stocks' returns are uncorrelated

Example 7.8 Solution

a The expected values of the two stocks are

$E(R_1) = .08$ and $E(R_2) = .15$

The weights are $w_1 = .25$ and $w_2 = .75$.

Thus,

$$
\begin{aligned}
E(R_2) &= w_1 E(R_1) + w_2 E(R_2) \\
&= .25(.08) + .75(.15) \\
&= .1325
\end{aligned}
$$

Example 7.8 Solution

The standard deviations are $\sigma_1 = .12$ and $\sigma_2 = .22$. Thus,

$$V(Rp) = w_1^2 s_1^2 + w_2^2 s_2^2 + 2 w_1 w_2 r s_1 s_2$$

$$= (.25^2)(.12^2) + (.75^2)(.22^2) + 2(.25)(.75) r (.12)(.22)$$

$$= .0281 + .0099 r$$

When $r = 1$

$$V(R_p) = .0281 + .0099(1) = .0380$$

When $r = 0.5$

$$V(R_p) = .0281 + .0099(.5) = .0331$$

When $r = 0$

$$V(R_p) = .0281 + .0099(0) = .0281$$

Portfolio Diversification in Practice

The formulas introduced in this section require that we know the expected values, variances, and covariance (or coefficient of correlation) of the investments we're interested in.

The question arises: How do we determine these parameters?

The most common procedure is to estimate the parameters from historical data, using sample statistics.

Bernoulli Random Variable

- Has two outcomes:
- X = 1 if "success"

  0 if "failure"
- P(X=1) = p (probability of "success")
- E.g. Fair coin: p = 0.5
- Recall the Indicator or Dummy Variable
- This is a Random Indicator Variable
- Try at home: Find E[X], VAR(X)

Binomial Distribution…

The ***binomial distribution*** is the probability distribution that results from doing a "***binomial experiment***". Binomial experiments have the following properties:

Fixed number of trials, represented as **n**.

Each trial has two possible outcomes, a "success" and a "failure".

P(success)=p (and thus: P(failure)=1–p), for all trials.

The trials are ***independent***, which means that the outcome of one trial does not affect the outcomes of any other trials.

Binomial Random Variable…

The binomial random variable ***counts*** the number of successes in **n** trials of the binomial experiment. It can take on values from 0, 1, 2, …, **n**. Thus, its a ***discrete*** random variable.

To calculate the probability associated with each value we use combintorics:

$$P(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{for } x=0, 1, 2, \ldots, n$$

Simple Example: 2 Coin tosses

- Please try this example at home:
- Toss Fair Coin Twice
- X = number of Heads
- What is the probability distribution of X
- Solve once using intuition
- Solve again using formula

Pat Statsdud…

Pat Statsdud is a (not good) student taking a statistics course. Pat's exam strategy is to rely on luck for the next quiz. The quiz consists of 10 multiple-choice questions. Each question has five possible answers, only one of which is correct. Pat plans to guess the answer to each question.

What is the probability that Pat gets no answers correct?

What is the probability that Pat gets two answers correct?

Pat Statsdud…

n=10, and **P(success) = .20**

What is the probability that Pat gets ***no answers*** correct?

I.e. # success, **x**, = 0; hence we want to know P(x=0)

$$P(0) = \frac{10!}{0!(10-0)!}(.2)^0(1-.2)^{10-0}$$

$$= 1(1)(.8)^{10} = .1074$$

Pat has about an 11% chance of getting no answers correct using the guessing strategy.

Pat Statsdud…

n=10, and **P(success) = .20**

What is the probability that Pat gets ***two answers*** correct?

I.e. # success, **x**, = 2; hence we want to know P(x=2)

$$P(2) = \frac{10!}{2!(10-2)!}(.2)^2(1-.2)^{10-2}$$

$$= 45(.04)(.1678) = .3020$$

Pat has about a 30% chance of getting exactly two answers correct using the guessing strategy.

Binomial Distribution…

As you might expect, statisticians have developed general formulas for the mean, variance, and standard deviation of a binomial random variable. They are:

$$\mu = np$$

$$\sigma^2 = np(1-p)$$

$$\sigma = \sqrt{np(1-p)}$$

Poisson Distribution…

Named for Simeon Poisson, the ***Poisson distribution*** is a discrete probability distribution and refers to the number of events (a.k.a. successes) within a specific time period or region of space. For example:

- The number of cars arriving at a service station in 1 hour. (The interval of time is 1 hour.)
- The number of flaws in a bolt of cloth. (The specific region is a bolt of cloth.)
- The number of accidents in 1 day on a particular stretch of highway. (The interval is defined by both time, 1 day, and space, the particular stretch of highway.)

The Poisson Experiment…

Like a binomial experiment, a ***Poisson experiment*** has four defining characteristic properties:

The number of successes that occur in any interval is independent of the number of successes that occur in any other interval.

The probability of a success in an interval is the same for all equal-size intervals

The probability of a success is proportional to the size of the interval.

The probability of more than one success in an interval approaches 0 as the interval becomes smaller.

Poisson Distribution…

The ***Poisson random variable*** is the number of successes that occur in a period of time or an interval of space in a Poisson experiment.

successes

E.g. On average, 96 trucks arrive at a border crossing every hour.

time period

E.g. The number of typographic errors in a new textbook edition averages 1.5 per 100 pages.

successes (?!)

interval

Poisson Probability Distribution…

The probability that a Poisson random variable assumes a value of **x** is given by:

$$P(x) = \frac{e^{-\mu}\mu^{x}}{x!} \quad for \quad x = 0,\ 1,\ 2,\ldots$$

where μ is the mean number of successes in the interval

and **e** is the natural logarithm base.

FYI: $E(X) = V(X) = \mu$

Example 7.12…

A statistics instructor has observed that the number of typographical errors in new editions of textbooks varies considerably from book to book. After some analysis he concludes that the number of errors is Poisson distributed with a mean of 1.5 per 100 pages. The instructor randomly selects 100 pages of a new book. What is the probability that there are no typos?

That is, what is P(X=0) given that $\mu = 1.5$?

$$P(0) = \frac{e^{-\mu}\mu^{x}}{x!} = \frac{e^{-1.5}1.5^{0}}{0!} = .2231$$

"There is about a 22% chance of finding zero errors"