Chapter 67

An introduction to random variables

Questions answered in this chapter:

- What is a random variable?
- What is a discrete random variable?
- What are the mean, variance, and standard deviation of a random variable?
- What is a continuous random variable?
- What is a probability density function?
- What are independent random variables?

In today's world, the only thing that's certain is uncertainty. In the next nine chapters, I'll give you some powerful techniques that you can use to incorporate uncertainty in business models. The key building block in modeling uncertainty is understanding how to use *random variables*.

Answers to this chapter's questions

What is a random variable?

Any situation whose outcome is uncertain is called an *experiment*. The value of a random variable emerges from the (uncertain) outcome of an experiment. For example, tossing a pair of dice is an experiment, and a random variable might be defined as the sum of the values shown on each die. In this case, the random variable could assume any of the values 2, 3, and so on, up to 12. As another example, consider the experiment of selling a new video-game console, for which a random variable might be defined as the market share for this new product.

What is a discrete random variable?

A random variable is *discrete* if it can assume a countable (but usually finite) number of possible values. Here are some examples of discrete random variables:

- Number of potential competitors for your product
- Number of aces drawn in a five-card poker hand
- Number of car accidents you have (hopefully zero!) in a year
- Number of dots showing on a die

- Number of free throws out of 12 that Kevin Durant makes during a basketball game

What are the mean, variance, and standard deviation of a random variable?

In Chapter 42, “Summarizing data by using descriptive statistics,” I discussed the *mean*, *variance*, and *standard deviation* for a data set. In essence, the mean of a random variable (often denoted by μ) is the average value of the random variable you would expect if you performed an experiment many times. The mean of a random variable is often referred to as the random variable’s *expected value*. The variance of a random variable (often denoted by s^2) is the average value of the squared deviation from the mean of a random variable that you would expect if you performed an experiment many times. The standard deviation of a random variable (often denoted by σ) is simply the square root of its variance. As with data sets, the mean of a random variable is a summary measure for a typical value of the random variable, whereas the variance and standard deviation measure the spread of the random variable about its mean.

As an example of how to compute the mean, variance, and standard deviation of a random variable, suppose you believe that the return on the stock market during the next year is governed by the following probabilities:

Probability	**Market return**
0.40	+20 percent
0.30	0 percent
0.30	-20 percent

Hand calculations show the following:

$\mu=0.40*(0.20)+0.30*(0.00)+0.30*(-0.20)=0.02$ or 2 percent $s^2=0.4*(0.20-0.02)^2+0.30*(0.0-0.02)^2+0.30*(-0.20-0.02)^2=0.0276$

Then σ=0.166 or 16.6 percent.

In the file Meanvariance.xlsx (shown in Figure 67-1), I verified these computations.

	B	C	D	E
1				
2				
3	Value	Probability	Squared deviation	
4	0.2	0.4	0.0324	=(B4-C9)^2
5	0	0.3	0.0004	=(B5-C9)^2
6	-0.2	0.3	0.0484	=(B6-C9)^2
7				
8				
9	Mean	0.02	=SUMPRODUCT(B4:B6,C4:C6)	
10	Variance	0.0276	=SUMPRODUCT(C4:C6,D4:D6)	
11	Standard deviation	0.166132477	=SQRT(C10)	

F67xx01

FIGURE 67-1 Computing the mean, standard deviation, and variance of a random variable.

F67xx01: This figure shows the calculation of the mean, variance, and standard deviation of a random variable.

I computed the mean of the market return in cell C9 with the formula SUMPRODUCT(B4:B6,C4:C6). This formula multiplies each value of the random variable by its probability, and then it sums the products.

To compute the variance of the market return, I determined the squared deviation of each value of the random variable from its mean by copying from D4 to D5:D6 the formula =(B4–C9)^2. Then, in cell C10, I computed the variance of the market return as the average squared deviation from the mean by using the formula =SUMPRODUCT(C4:C6,D4:D6). Finally, in cell C11, I computed the standard deviation of the market return with the formula =SQRT(C10).

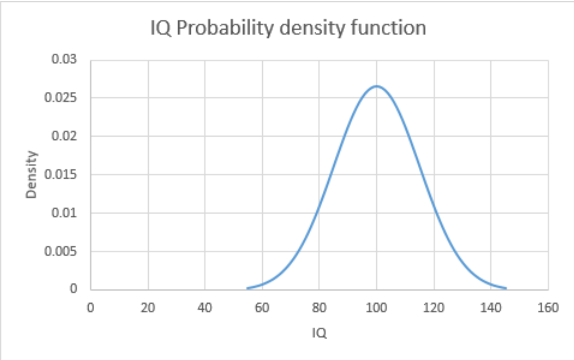
What is a continuous random variable?

A *continuous random variable* is a random variable that can assume a very large number or, to all intents and purposes, an infinite number of values, including all values on some interval. The following are some examples of continuous random variables:

- Price of Microsoft stock one year from now
- Market share for a new product
- Market size for a new product
- Cost of developing a new product
- Newborn baby's weight
- Person's IQ
- Dirk Nowitzki's three-point shooting percentage during next season

What is a probability density function?

A *discrete random variable* can be specified by a list of values and the probability of occurrence for each value of the random variable. Because a continuous random variable can assume an infinite number of values, you can't list the probability of occurrence for each value of a continuous random variable. A continuous random variable is completely described by its *probability density function*. For example, the probability density function for a randomly chosen person's IQ is shown in Figure 67-2.



F67xx02

FIGURE 67-2 The probability density function for IQs.

F67xx02: This figure shows the probability density function for IQs.

A probability density function (PDF) has the following properties:

- The value of the PDF is always greater than or equal to 0.
- The area under the PDF equals 1.
- The height of the PDF for a value *x* of a random variable is proportional to the likelihood that the random variable assumes a value near *x*. For example, the height of the density for an IQ of 83 is roughly half the height of the density for an IQ of 100. This tells you that IQs near 83 are approximately half as likely as IQs around 100. Also, because the density peaks at 100, IQs around 100 are most likely.
- The probability that a continuous random variable assumes a range of values equals the corresponding area under the PDF. For example, the fraction of people having IQs from 80 through 100 is simply the area under the density from 80 through 100.
- Note that a discrete random variable that assumes many values is often modeled as a *continuous random variable*. (See Chapter 70, “The normal random variable and Z-scores.”) For example, while the number of half gallons of milk sold in a single day by a small grocery store is discrete, it proves more convenient to model this discrete random variable as a continuous random variable.

What are independent random variables?

A set of random variables is *independent* if knowledge of the value of any of their subsets tells you nothing about the values of the other random variables. For example, the number of games won by the Indiana University football team during a year is independent of the percentage return on Microsoft during the same year. Knowing that Indiana did very well would not change your view of how Microsoft stock did during the year.

On the other hand, the returns on Microsoft stock and Intel stock are not independent. If you are told that Microsoft stock had a high return in one year, in all likelihood, computer sales were high, which tells you that Intel probably had a good year as well.

Problems

1. Identify the following random variables as discrete or continuous:
   - Number of games the Seattle Seahawks win next season
   - Number that comes up when spinning a roulette wheel
   - Unit sales of tablet PCs next year
   - Length of time that a light bulb lasts before it burns out
2. Compute the mean, variance, and standard deviation of the number of dots showing when a die is tossed.
3. Determine whether the following random variables are independent:
   - Daily temperature and sales at an ice-cream store
   - Suit and number of a card drawn from a deck of playing cards
   - Inflation and return on the stock market
   - Price charged for each and the number of units sold of a car
4. The current price of a company's stock is $20. The company is a takeover target. If the takeover is successful, the company's stock price will increase to $30. If the takeover is unsuccessful, the stock price will drop to $12. Determine the range of values for the probability of a successful takeover that would make it worthwhile to purchase the stock today. Assume your goal is to maximize your expected profit. Hint: Use the Goal Seek command, which is discussed in detail in Chapter 18, "The Goal Seek command."
5. When a roulette wheel is spun, the possible outcomes are 0, 00, 1, 2, …, 36. If you bet on a number coming up, you win $35 if your number comes up and you lose $1 otherwise. What is the mean and standard deviation of your winnings on a single play of the game?
6. A stock currently sells for $40. In the next month, there is a 60 percent chance the stock price will double and a 40 percent chance the stock will drop 50 percent. In a month, you will sell the stock. Find the mean and standard deviation of your profit (in dollars).
7. Suppose you bet on an odd number coming up in roulette. If an odd number comes up, you win $1, and if an odd number does not come up, you lose $1. Find the mean and standard deviation of your profit.