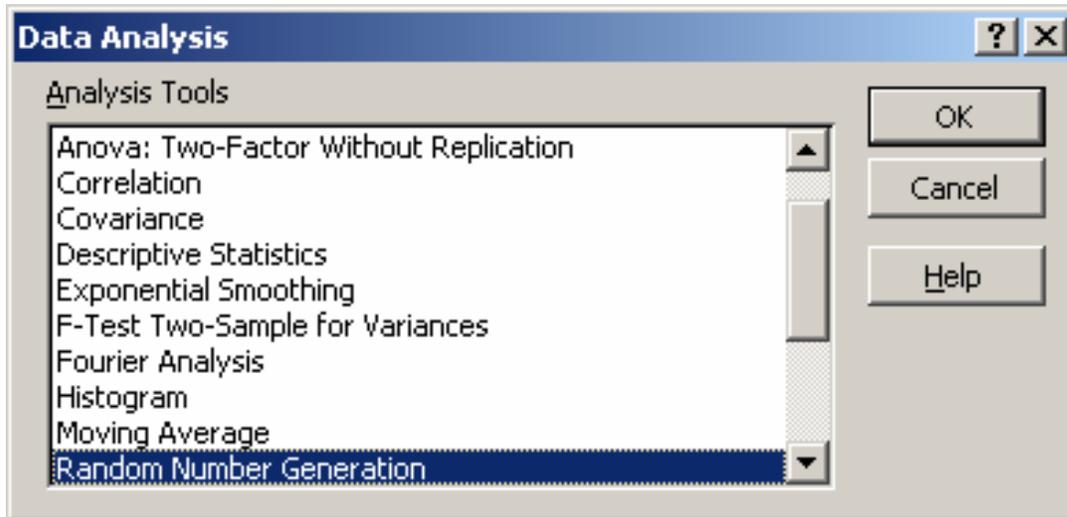


Using Excel to Simulate Probabilistic Events

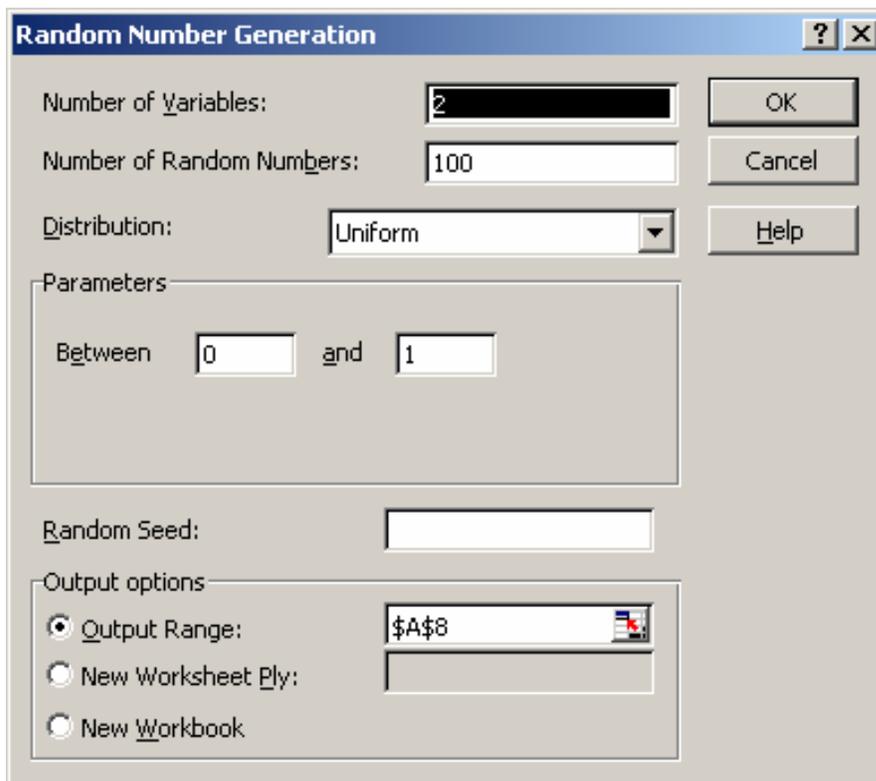
Generating Random Variables

1. First Method: Static Generation

The easiest way of generating random variables (but also the most restrictive way) is by choosing the Option Data Analysis ... Random Number Generation ...



The following dialog box then appears:



The field *Number of Variables* will be the number of columns Excel will generate with this random variable; the field *Number of Random Numbers* is the number of rows Excel will fill with random numbers. The following distributions are available:

Uniform
Normal
Bernoulli
Binomial
Poisson
*Patterned*¹
Discrete

There is serious drawback when using this method. The random variables generated this way are *static*, i.e., they will never be recomputed, but the main drawback is that none of the parameters (e.g., the mean or standard deviation of a Normal distribution) can contain a reference to a cell, and certainly no formulae are allowed! In other words, this method is good for generating a set of simple random data, but is less appropriate for modeling more complex processes.

Example

Suppose we want to model the following simple gambling game. You can enter the “*Dice Game*” when you pay 1 dollar. Then the payoffs are as follows:

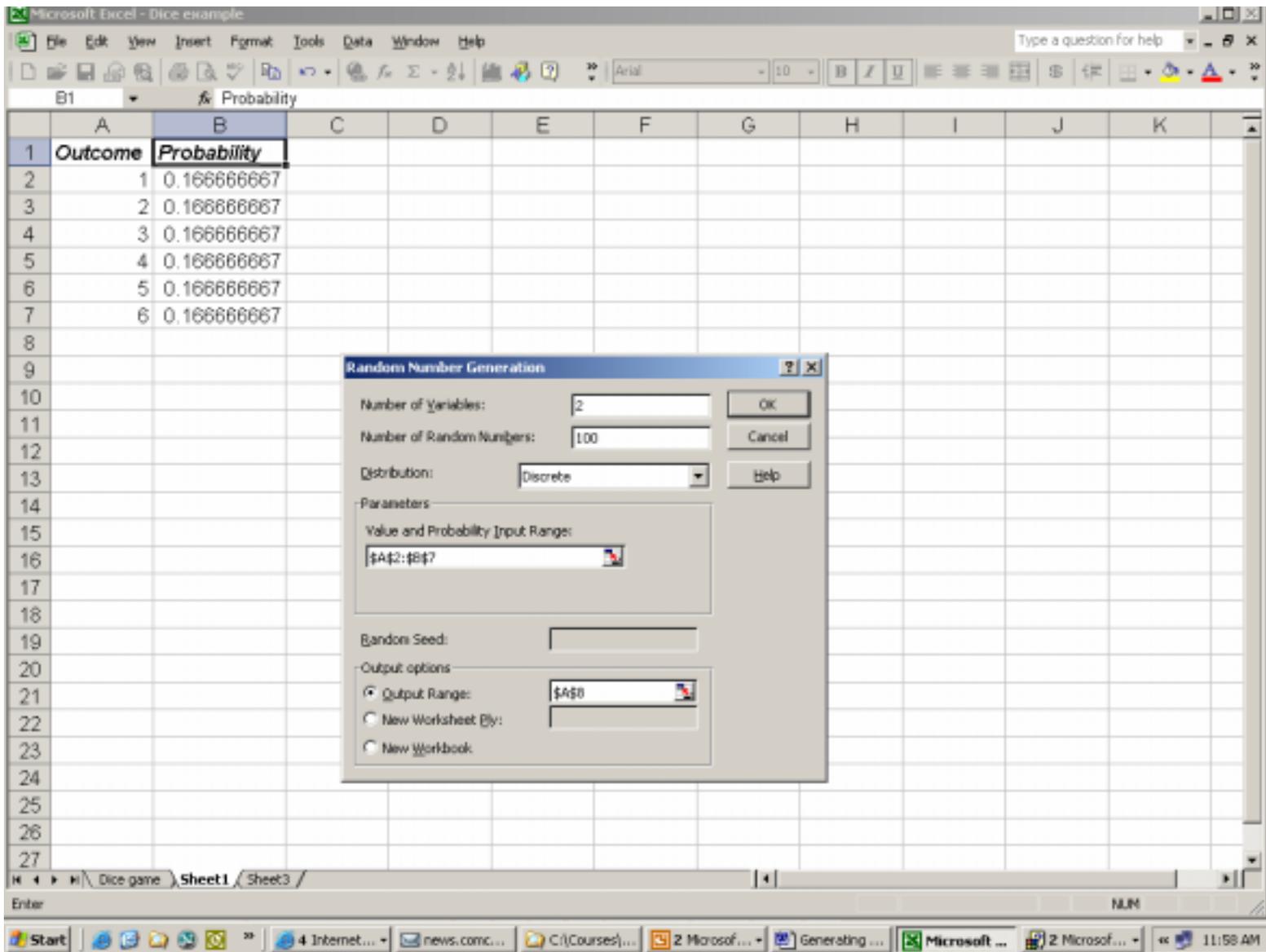
1. When the first roll of the die is a “two” or a “four”, you can roll the die again.
 - a. If the outcome of the second roll is less than or equal to the first roll (e.g., a “two”), you win the latter amount (i.e., 2 dollars).
 - b. If the outcome is greater than the first roll, you lose.
2. If the first roll is other than a two or four, you lose.

Is it profitable to enter this game? In the long run, would you come out ahead, or lose money? This would require some (rather advanced) probabilistic computations, however, with Excel it is easy to simulate this game, say a hundred times, and see what happens.

In order to do this, we assume that every play of the game will involve two rolls of the die (when the outcome is other than a two or four, the second roll will just be ignored). Since the outcome of a roll of a die is clearly a discrete random variable, we start by simulating two rolls of a die.

I choose the option *Tools ... Data Analysis ... Random Number Generation* from the Excel menu and my worksheet and dialog look as follows:

¹ Strictly speaking, *patterned* is not a random number generator. That is, values are computed using a starting value and ending value using a given step size. In other words, it is very easy to compute a number once given the previous number, so there’s nothing “random” about this process.



with the dialog box:

Random Number Generation [?] [X]

Number of Variables:

Number of Random Numbers:

Distribution:

Parameters

Value and Probability Inter Range:

Random Seed:

Output options

Output Range:

New Worksheet Ply:

New Workbook

For a discrete random variable, Excel wants you to make a table with all outcomes and their probabilities (the sum of the latter should be one, of course). So, in cells A2 to B7, I listed the values of rolling a die, and (assuming we're working with a fair die!) every outcome is equally likely ($1/6$).

We will get two columns of variables and 100 hundred rows. After hitting the *OK* button, the worksheet looks like:

	A	B	C	D	E	F	G	H	I	J
1	Outcome	Probability								
2		1	0.166666667							
3		2	0.166666667							
4		3	0.166666667							
5		4	0.166666667							
6		5	0.166666667							
7		6	0.166666667							
8										
9										
10										
11	Roll 1	Roll 2	Revenue	Profit						
12	4	5								
13	1	1								
14	4	6								
15	2	1								
16	3	6								
17	5	4								
18	3	3								
19	4	1								
20	6	2								
21	3	6								
22	5	6								
23	1	3								
24	1	1								

For the first case, the first roll was a “four”, but the second roll was bigger than the first, so in this case the revenue should be computed to be 0. In the second case, the first roll is a “one” so that’s a losing game from the start. For the fourth play, the first roll is a “two”, the second is a “one” so the revenue is equal to 1 dollar, and after subtracting the one dollar payment to enter this game, the player just breaks even in this fourth play. Some trial-and-error with Excel functions, (esp. the IF() function), reveals that the correct formula to compute the **Revenue** is:

```
=IF(A12=2,
    IF(B12<=2,
        B12,
        0),
    IF(A12=4,
        IF(B12<=4,
            B12,
            0),
        0))
```

- * is first roll two?
- * is second roll two or less?
- * if yes, then revenue = second roll
- * if no, revenue = 0
- * is first roll four?
- * is second roll four or less?
- * if yes, then revenue = second roll
- * if no, revenue = 0
- * if first roll not two or four, revenue=0

The **Profit** column is computed as the Revenue minus one dollar, so the worksheet becomes:

The screenshot shows an Excel spreadsheet titled "Dice example" with the following data:

	A	B	C	D	E	F
1	Outcome	Probability				
2	1	0.166666667				
3	2	0.166666667				
4	3	0.166666667				
5	4	0.166666667				
6	5	0.166666667				
7	6	0.166666667				
8						
9						
10						
11	Roll 1	Roll 2	Revenue	Profit		
12	4	5	\$0	-\$1		
13	1	1	\$0	-\$1		
14	4	6	\$0	-\$1		
15	2	1	\$1	\$0		
16	3	6	\$0	-\$1		
17	5	4	\$0	-\$1		
18	3	3	\$0	-\$1		
19	4	1	\$1	\$0		
20	6	2	\$0	-\$1		
21	3	6	\$0	-\$1		

Depending on how many rows you generated and whether you input a value for “random Number Seed” in the Random Number Generation Dialog box, you average “winnings” in this game range between \$-0.4 and \$-0.65. So, in the long run you would lose money in this game.

2. Second Method: Dynamic Generation

Our second method allows to generate more complicated probabilistic events. For example, we can let the parameters be references to cells or to formulae; and (most importantly) we can generate random variables based on the outcome of other variables. For example, suppose we want to generate the weights for a population of 100 45-year olds consisting of males and females. It is known that weight is normally distributed weight mean 180lbs and standard deviation of 20lbs; whereas the weight for a 45-year old woman has a mean of 150lbs with the same standard deviation. Suppose we want to generate a male person with probability 0.6, a female with probability of 0.4.

Excel’s RAND() function generates a *uniformly distributed* random number between 0 and 1. So, we can use this function to determine whether we have a “male” or “female”. If the RAND() function generates a number less than or equal to 0.6, we have a male, otherwise we have a female. Or, combining, we can use the Excel formula for the cell “**Sex**”:

=IF(rand()<=0.6,"M","F")

Based upon this value, we can determine whether the **Mean Weight** should be 180lbs (for the males), or 160 lbs (for the females). This is done by using the IF function again:

=IF(A2="M",180,160)

Copying this formula 100 hundred times down the first column, our worksheet so far will look like:

The screenshot shows the Microsoft Excel interface with the following data in the spreadsheet:

	A	B	C	D	E	F	G
1	Sex	Mean Weight	Random Weight				
2	M	180					
3	F	160					
4	F	160					
5	M	180					
6	M	180					
7	M	180					
8	F	160					
9	F	160					
10	F	160					
11	F	160					
12	M	180					
13	M	180					
14	F	160					
15	M	180					
16	M	180					
17	M	180					
18	M	180					
19	M	180					
20	M	180					
21	F	160					
22	M	180					

The formula bar at the top shows the formula: `=IF(RAND()<=0.6,"M","F")`. The spreadsheet has three sheets: Sheet1, Sheet2, and Sheet3. The status bar at the bottom shows 'Ready' and 'NUM'.

Now, the next thing we need to do is generate a Normal distributed random variable **weight**. For this, we need the Excel function NORMINV. Generating a normal random variable with this function is done like this

=NORMINV(RAND(),mean, standard_deviation)

where “mean” and “standard_deviation” should be substituted with numbers, or cell references or a formula. In our case, the “mean” is given in the second column (a cell reference) and standard_deviation can just be typed in as 20. So, the formula becomes:

=NORMINV(RAND(),A2, 20)

So, the worksheet with our 100 individuals becomes:

	A	B	C	D	E	F	G
1	Sex	Mean Weight	Random Weight				
2	M	180	156.42				
3	M	180	170.32				
4	M	180	178.77				
5	F	160	181.75				
6	M	180	189.87				
7	M	180	156.74				
8	M	180	195.34				
9	F	160	140.81				
10	F	160	134.28				
11	M	180	149.97				
12	F	160	170.37				
13	F	160	190.97				
14	M	180	185.96				
15	F	160	158.44				
16	M	180	205.53				

Other distributions can be generated with the INV functions as well. Following is a complete list of functions available in Excel ²:

² We won't be needing most of them in practice, but it's good to know they're all there!

List of functions ³:

CHIINV(RAND(),degrees_of_freedom)

Generates a Chi-square distribution with *degrees_of_freedom*

FINV(RAND(),degrees_of_freedom1, degrees_of_freedom2)

Generates an *F*-distribution with *degrees_of_freedom1* in the numerator and *degrees_of_freedom2* in the denominator

GAMMAINV(RAND(),alpha,beta)

Generates a Gamma-distribution with parameters *alpha* and *beta*.

LOGINV(RAND(),mean,standard_deviation)

Generates a lognormal distribution with specified *mean* and *standard_deviation*

NORMINV(RAND(),mean,standard_deviation)

Generates a normal distribution with specified *mean* and *standard_deviation*

NORMSINV(RAND())

Generates a standard normal distribution (i.e., *mean=0* and *standard_deviation=1*)

TINV(RAND(),degrees_of_freedom)

Generates a *T*-distribution with *degrees_of_freedom*

Unfortunately, there does not exist an “INV” function to generate Uniformly distributed random variables. However, such function is easy to construct. For a UNIFORM distribution, use the formula:

$$=a+RAND()*(b-a)$$

where *a* is the *lowest value* the random variable can take on, and *b* is its *maximum value*.

The second method for generating random variables is the more cumbersome one, but allows for more complicated probabilistic models. Another advantage is that in this method, it is easy to “repeat” the simulation with new random numbers. Pressing the **F9** key recomputes the worksheet, thereby generating new values for the random variables. For the first method, if the simulation were to be repeated with new values for the random variables one would have to go through the menu steps *Tools ... Data Analysis ... Random Number Generation* again, and in order to guarantee new values for the random variables, put in some new number in the “Random Number Seed” field.

³ Words in upper case must be typed in unchanged, lower case terms need to be substituted with their proper values (numbers, cell references or formulae).