Choosing Probability Distributions in Simulation

Probability Distributions may be selected on the basis of

- Data
- Theory
- Judgment
- a mix of the above...

General approach

1. Gather all known information about the uncertain variable
2. Match the information with a probability distribution

Gathering information about the uncertain variable

- Any relevant data available?
  If yes, consider using the “Fit...” option of Crystal Ball

- Is the variable discrete or continuous?
  Want discrete values out of a continuous distribution?
  Just round the values to integer =ROUND(Assumption Cell, 0)

- Do we know anything about the process underlying the variable?
The Uniform Distribution

*Almost Complete Ignorance*

- We know nothing except for a range (Min and Max)
- No reason to think that some values are more likely than others
- No reason to think that some values are less likely than others

Assessing the range of a variable:
- Think of "highest" lower bound, "lowest" upper bound
- Beware: *Overconfidence* ➔ range too narrow

The Triangular Distribution

*We know a “best guess” value*

- We know the most likely value (the Mode), the Min and the Max
- Min, Mode, Max are often based on judgment
- We believe: the larger a deviation from central value, the less likely
- May have a non symmetric shape
- Beware: most likely value ≠ expected value

Assessing the range of the variable:
- Think of bounds separately, not just as deviations from typical value
- Beware: *Anchoring* on Mode to judge Min and Max values ➔ range too narrow
The Normal Distribution

- We believe the variable results from the combination of many independent random factors, or a natural process.
- Deviations up or down from the mean are equally likely.
- We need to specify the Mean and Standard Deviation.
- Are any percentiles known?
  A distribution in Crystal Ball may be defined using percentiles, e.g. 10th and 90th.

![Normal Distribution Graph]

Stock return with 6% mean and 12% standard deviation.

Judgmentally, percentiles may be easier to assess than the standard deviation.

The Binomial Distribution

We know something about the process:

- We can think of the variable as a number of instances out of a specified number of trials.
  E.g.: Any web page viewer has a 10% chance of making a purchase. We’ll have 100 page viewers. How many sales could we have?
- We know the probability of occurrence on any trial (P)
  E.g. probability that any particular cell phone will ring during class.
- We know the total number of trials (N)
  E.g. total number of phones that are on ring mode during class.

![Binomial Distribution Graphs]
The Poisson Distribution

We know something about the process

- We can think of the variable as a number of instances over a specified time period, occurring randomly at a given rate.
  E.g.: Number of help requests arriving at the 4618 computer hotline per hour.
- We know the average rate of occurrence (e.g. 5 per hour)
- Average rate of occurrence is constant

Other Distributions

Other distributions may be used...

- Theoretical reasons
  Lognormal Distribution: the log of the random variable follows a Normal
  e.g. stock prices
  Beta Distribution: the random variable has natural bounds
  e.g. a proportion, probability, market share, etc.

- Empirical reasons: it has proved to fit well to past data
- Avoid using a distribution that you do not understand

Single event

- The =RAND() function returns a value uniformly distributed between 0 and 1. Can be used as part of other Excel formulas
  Ex: =(RAND()<0.35) will return 1 with prob. 0.35, 0 with prob. 0.65
Mixture Distributions

*Modeling “fat” tails with a mixture of normal distributions*

**Ex:** Mix a Normal with Mean=0, Stdev=1
with a Normal with Mean=0, Stdev=5

**Use:** =IF(RAND()>0.5, CB.Normal(0,1), CB.Normal(0,5))

---

Picking the right distribution = Art + Science

*All you know about the random variable’s Process, Range, Likely values, Percentiles, etc…*  
*Features and properties of common probability distributions*

... or make up a “Custom” distribution
Some useful short-cut functions provided by Crystal Ball

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Excel-style function (active when CB is loaded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>=CB.Uniform(Min, Max)</td>
</tr>
<tr>
<td>Normal</td>
<td>=CB.Normal(Mean, StdDev)</td>
</tr>
<tr>
<td>Triangular</td>
<td>=CB.Triangular(Min, Mode, Max)</td>
</tr>
<tr>
<td>Binomial</td>
<td>=CB.Binomial(Prob, Trials)</td>
</tr>
<tr>
<td>Poisson</td>
<td>=CB.Poisson(Rate)</td>
</tr>
<tr>
<td>Exponential</td>
<td>=CB.Exponential(Rate)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>=CB.Lognormal(Mean, StdDev)</td>
</tr>
<tr>
<td>Custom</td>
<td>=CB.Custom(Array of values/prob.)</td>
</tr>
</tbody>
</table>

Declaring random variables with “CB.functions”

Pros
• quick, easy to use
• provide great flexibility (e.g. can be used in ‘IF’ statements)
• can be combined with any other Excel function
• parameters can be dynamic

Cons
• will not appear as assumptions in the simulation Report
• will not be selected when you ‘Select Assumptions’ in your model
• will not be included in the Sensitivity estimations
• do not allow you to define explicit correlation with other variables
For next session...

Session 15

- Prepare SCOR-eSTORE.COM case
  The simulation model is already built
  Exploit the model to estimate the value of real options

- Individual case (will be accepted until Session 16)
  Turn in a hard copy in class as per instructions