Agribusiness Analysis and Forecasting
Stochastic Simulation

Henry Bryant
Texas A&M University
In economics we use simulation because we cannot experiment on live subjects, a business, or the economy without injury.

In other fields they can create an experiment:

- Health sciences they feed (or treat) lots of lab rats on different chemicals to see the results.
- Animal science researchers feed multiple pens of steers, chickens, cows, etc. on different rations.
- Engineers run a motor under different controlled situations (temp, RPMs, lubricants, fuel mixes).
- Vets treat different pens of animals with different meds.
- Agronomists set up randomized block treatments for a particular seed variety with different fertilizer levels.
Probability Distributions

Parametric and Non-Parametric Distributions

- Parametric Dist. have known and well defined parameters that force their shapes to known patterns.
  - Normal Distribution - Mean and Standard Deviation.
  - Uniform - Minimum and Maximum
  - Bernoulli - Probability of true
  - Beta - Alpha, Beta, Minimum, Maximum

- Non-Parametric Distributions do not have pre-set shapes based on known parameters.
  - The parameters are estimated each time to make the shape of the distribution fit the data.
  - Empirical – Actual Observations and their Probabilities.
Typical Problem for Risk Analysis

- We have a stochastic variable that needs to be included in a business model. For example:
  - Price forecast has residuals we could not explain and they are the stochastic component we need to simulate.
  - Crop yield is forecasted by trend but it has residuals that are stochastic; risk caused by weather.

- Model will be solved (sampled) many times using alternative draws of random values for prices and yields.

- We have the data and a forecast model next we need to estimate parameters to define the stochastic variables.
  - NOTE: Parameters is the generic name for values that determine the location and shape of the distribution.
Steps for Simulating Random Variables

- For parametric distributions, we must make an assumption on a probability distribution for the random variables (e.g., Normal or Beta or Uniform...).

- Estimate/fit the parameters values to define the assumed distribution.

- Parameters for parametric distributions we will be using are:
  - Normal (Mean, Std Deviation)
  - Beta (Alpha, Beta, Min, Max)
  - Uniform (Min, Max)
  - Bernoulli (probability of true)
Steps for Parameter Estimation

1. Be sure that you have removed any trend, cycle or structural pattern. Be sure that you have a constant mean and variance.

2. Estimate parameters for several assumed distributions using historical data.

3. Simulate the data under different distributions.

4. Pick the best distribution based on.
   - Mean, Standard Deviation - use validation tests.
   - Minimum and Maximum.
   - Shape of the CDF vs. historical series.
   - Penalty function =CDFDEV() to quantify differences.
Parameter Estimator in Simetar

Use Theta Icon in Simetar
- Estimate parameters for up to 17 parametric distributions.
- Select MLE for parameter estimation.
- The tool provides ready-made cells simulating your variable under the various distributions.
Uniform Distribution

- Random variable where every interval has an equal probability of being observed (drawn).

  if $X$ is Uniform$(0, 1)$ then $P(0.1 < x < 0.2) = P(0.5 < x < 0.6)$

- Simulating Uniform in Simetar enter parameters as:
  - `=UNIFORM(Minimum, Maximum)`
  - `=UNIFORM(0,1)` which is the same as `=UNIFORM()` (this is standard uniform)
  - `=UNIFORM(10,25)`, etc.

- A standard uniform RV is used to simulate all distributions. For example a normal distribution:
  - `=norm(mean, standard deviation, U)`, where $U$ is distributed standard uniform.
**Standard Uniform Distribution**

- CDF of the Standard Uniform Distribution.

![CDF of the Standard Uniform Distribution](image)

- PDF of Standard Uniform Distribution.

![PDF of Standard Uniform Distribution](image)
Basic Simulation Definitions

- Stochastic Simulation Model - means the model has at least one random variable.
- Monte Carlo simulation model - same as a stochastic simulation model.
- Two ways to sample or simulate random values:
  1. Monte Carlo sampling - draw random values for the variables purely at random.
  2. Latin Hyper Cube sampling - draw random values using a systematic approach so we are certain that we sample ALL regions of the probability distribution.
- Monte Carlo sampling requires larger number of iterations to insure that model samples all regions of the probability distribution.
MC vs. LHC Sampling

- For a standard uniform random variable (uniform over the unit interval), the CDF is a 45-degree straight line.
- MC empirical CDF deviates from the 45-degree line.
- LHC empirical CDF is right on top of the population CDF.
- This is with 500 iterations.
- Simetar default is LHC.
When to Use the Normal Distribution

- Use the Normal distribution if you have lots of observations and have tested for normality.
- BUT watch for infeasible values from a Normal distribution (negative yields and prices).

![Normal Distribution Graph](Image)
How to Test for Normality

Simetar provides an easy to use procedure for testing Normality that includes:

- S-W (Shapiro-Wilk)
- A-D (Anderson-Darling)
- CvM (Cramer-Von Mises)
- K-S (Kolmogorov-Smirnov)
- Chi-Squared
Truncated Normal

- General formula for the Truncated Normal
  \[=\text{TNORM}(\text{Mean}, \text{Std Dev}, [\text{Min}], [\text{Max}],[\text{USD}])\]

- Truncated Downside only
  \[=\text{TNORM}(10, 3, 5)\]

- Truncated Upside only
  \[=\text{TNORM}(10, 3, , 15)\]

- Truncated Both ends
  \[=\text{TNORM}(10, 3, 5, 15)\]

- Truncated both ends with a USD in general form
  \[=\text{TNORM}(10, 3, 5, 15, [\text{USD}])\]
Parameter is \( p \) or the probability that the random variable is 1.0 or TRUE.

Simulate Bernoulli as:

\[
\text{=Bernoulli}(p) \\
\text{=Bernoulli}(0.25)
\]
### Bernoulli Distribution Application

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td><strong>Conditional Probability Distribution Example of Rain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>P(rain) in June</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Quantity of Rain IF it rains</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Min</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Max</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Use a Uniform distribution to simulate the amount of the rainfall</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Rainfall If it rained</td>
<td>3.728058</td>
<td>=UNIFORM(B16,B17)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Did it Rain?</td>
<td>1 =BERNOULLI(B14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>This is the value we want to simulate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>If It Rained the Amount</td>
<td>3.728058</td>
<td>=B21*B19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>How we could use the stochastic rainfall value in a simulation model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Assume a yield function for cotton that was Y = 400 + 15*Rainfall in June</td>
<td>455.9209 =400+15*B23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Simulated Yield is</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Press F9 several times to see the impact of random rainfall on yield</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Bernoulli Distribution Application

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>Simulate Machinery Repair Costs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Assume a 5% chance of a repair</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Repairs are $10,000, $20,000 or $30,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Bernoulli parameter</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Repairs costs range are:</td>
<td>10000</td>
<td>20000</td>
<td>30000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>If Repair is needed what is the stochastic repair cost?</td>
<td>30000 =DEMPRICAL(B36:D36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Repair?</td>
<td>1 =BERNOULLI(B35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>Simualted Repair Cost</td>
<td>30000 =B38*E37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>You must hit F9 about 22 times to get a value for simulated repair greater than zero.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>Think about it there is only a 5% chance of a repair or 1 in 20 chance.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Beta Distribution

- **Beta** is a continuous probability distribution.
- It is parametrized by two positive **shape parameters**, denoted by $\alpha$ and $\beta$.
- These two parameters define the shape of the distribution.
- Simulate *Beta* distribution using the function:
  $$=\text{beta.inv(USD, alpha, beta, minimum, maximum)}$$