Agribusiness Analysis and Forecasting

Stochastic Simulation

Aleks Maisashvili

Texas A&M University

Stochastic Simulation

In economics we use simulation because we can not 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

- First step: be certain that the variable that you will directly scholastically draw is suitable
- Every stochastic draw you will make will for a variable will be independent of every other draw, even for the same variable in different time periods.
- The properties of the variable must be consistent with this simulation process
- In short, we need all draws for an individual variable to be *independently, identically distributed* (i.i.d.).
- We must therefore be certain that the variables we directly simulate have
    - Constant mean
    - Constant, finite variance
    - No autocorrelation

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)

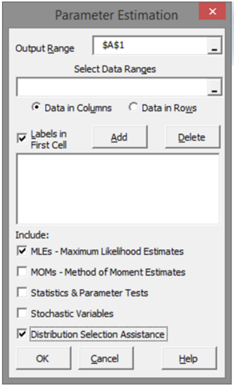
Steps for Parameter Estimation

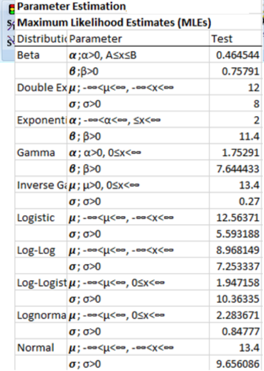
1. Again, be sure that you have removed any trend, cycle or structural pattern. Be sure that you have a constant mean and variance. i.i.d.!
2. Estimate parameters for several assumed distributions using historical data.
3. Simulate the data under different distributions.
4. Pick the best distribution 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.

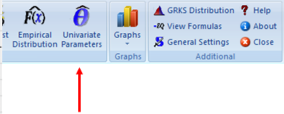




Parameter Estimation

Maximum Likelihood Estimates (MLEs)

Distributi	Parameter	Test
Beta	α ; $\alpha>0$, $A\leq x\leq B$	0.464544
	β ; $\beta>0$	0.75791
Double Ex	μ ; $-\infty<\mu<\infty$, $-\infty<x<\infty$	12
	σ ; $\sigma>0$	8
Exponenti	α ; $-\infty<\alpha<\infty$, $\leq x<\infty$	2
	β ; $\beta>0$	11.4
Gamma	α ; $\alpha>0$, $0\leq x<\infty$	1.75291
	β ; $\beta>0$	7.644433
Inverse G	μ ; $\mu>0$, $0\leq x<\infty$	13.4
	σ ; $\sigma>0$	0.27
Logistic	μ ; $-\infty<\mu<\infty$, $-\infty<x<\infty$	12.56371
	σ ; $\sigma>0$	5.593188
Log-Log	μ ; $-\infty<\mu<\infty$, $-\infty<x<\infty$	8.968149
	σ ; $\sigma>0$	7.253337
Log-Logist	μ ; $-\infty<\mu<\infty$, $0\leq x<\infty$	1.947158
	σ ; $\sigma>0$	10.36335
Lognorma	μ ; $-\infty<\mu<\infty$, $0\leq x<\infty$	2.283671
	σ ; $\sigma>0$	0.84777
Normal	μ ; $-\infty<\mu<\infty$, $-\infty<x<\infty$	13.4
	σ ; $\sigma>0$	9.656086



Uniform Distribution

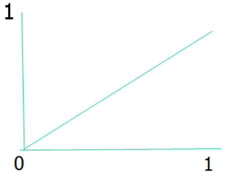
- Random variable where every interval has an equal probability of being observed (drawn).

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

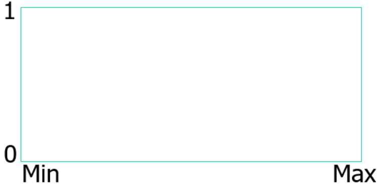
- Simulating Uniform in Simetar enter parameters as:
  - =UNIFORM(Minimum , Maximum)
  - =UNIFORM(0,1) which is the same as =UNIFORM() (this is **standard** uniform)
  - =UNIFORM(10,25), etc.
- A standard uniform RV is used to simulate all distributions. For example a normal distribution:
  - =norm(mean, standard deviation, U), where U is distributed standard uniform.

Standard Uniform Distribution

- CDF of the Standard Uniform Distribution.



- PDF of Standard Uniform Distribution.

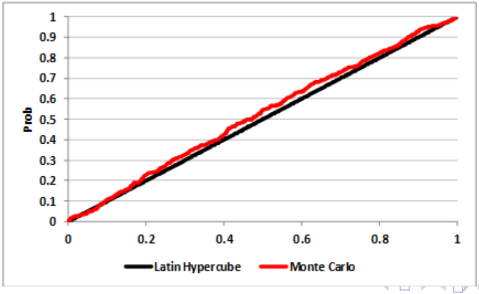


Basic Simulation Definitions

- 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.

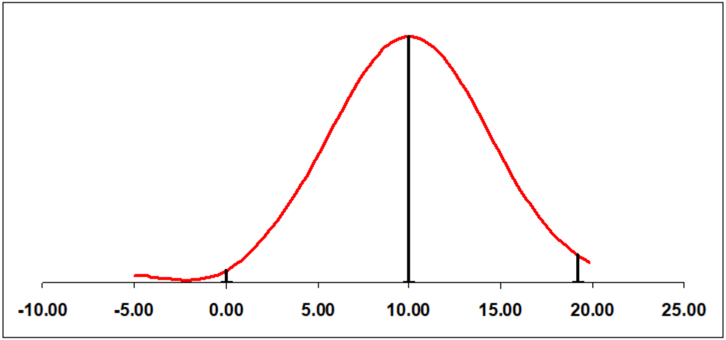
MC vs. LHC Sampling

- 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

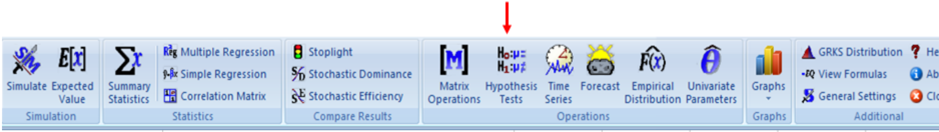
- Use the Normal distribution if you have lots of observations and have tested for normality.
- BUT watch for infeasible values from a *Normal* distribution (negative yields and prices).



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

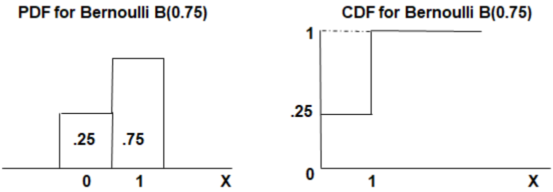


Test for Normality of Distribution for SOYBEAN			
Confidence Level		95.00%	
Procedure	Test Value	p-Value	
S-W	0.932020794	0.151105044	*Fail to Reject the Ho that the Distribution is Normally Distributed**
A-D	0.572620313	0.119703423	*Fail to Reject the Ho that the Distribution is Normally Distributed**
CvM	0.098783019	0.108520615	*Fail to Reject the Ho that the Distribution is Normally Distributed**
K-S	0.173747255	NA	*Consult Critical Value Table*
Chi-Sqared	12.52380952	0.862032304	*Fail to Reject the Ho that the Distribution is Normally Distributed**
			**Based on approximate p-values*

Truncated Normal

- General formula for the Truncated Normal
  =TNORM(Mean, Std Dev, [Min], [Max],[USD])
- Truncated Downside only
  =TNORM(10, 3, 5)
- Truncated Upside only
  =TNORM(10, 3, , 15)
- Truncated Both ends
  =TNORM(10, 3, 5, 15)
- Truncated both ends with a USD in general form
  =TNORM(10, 3, 5, 15, [USD])

Bernoulli Distribution



PDF and CDF for a Bernoulli Distribution.

- Parameter is p or the probability that the random variable is 1.0 or TRUE.
- Simulate Bernoulli as:
  =Bernoulli(p)
  =Bernoulli(0.25)

Bernoulli Distribution Application

	A	B	C	D	E
13	Conditional Probability Distribution Example of Rain				
14	P(rain) in June	0.2			
15	Quantity of Rain IF it rains				
16	Min	2			
17	Max	5			
18	Use a Uniform distribution to simulate the amount of the rainfall				
19	Rainfall If it rained	3.728058	=UNIFORM(B16,B17)		
20					
21	Did it Rain?	1	=BERNOULLI(B14)		
22	This is the value we want to simualt				
23	If It Rained the Amount	3.728058	=B21*B19		
24	How we could use the stochastic rainfall value in a simulation model				
25	Assume a yield function for cotton that was Y = 400 + 15*Rainfall in June				
26					
27	Simulated Yield is	455.9209	=400+15*B23		
28	Press F9 several times to see the impact of random rainfall on yield				

Bernoulli Distribution Application

	A	B	C	D	E	F	G
32	Simulate Machinery Repair Costs						
33	Assume a 5% chance of a repair						
34	Repairs are $10,000, $20,000 or $30,000						
35	Bernoulli parameter	0.05					
36	Repairs costs range are:	10000	20000	30000			
37	If Repair is needed what is the stochastic repiar cost?				30000	=DEMPIRICAL(B36:D36)	
38	Repair?	1	=BERNOULLI(B35)				
39							
40	Simualted Repair Cost	30000	=B38*E37				
41	You must hit F9 about 22 times to get a vlue for simulated repair greater than zero.						
42	Think about it there is only a 5% chance of a reapir or 1 in 20 chance.						

Beta Distribution

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