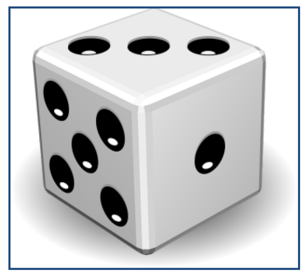
Those who can, do; those who can't – use Computer Simulation *



AMATYC S089 – Nov. 16, 2018
Jerry Tuttle, FCAS
University of Phoenix

* Apologies to GBS, 1902

Background

- Simulation is imitation to reproduce behavior of a real-world system.
- Useful for some problems when real-world is too expensive, too dangerous, too complicated.

- Other non-mathematical examples?

Other non-mathematical examples

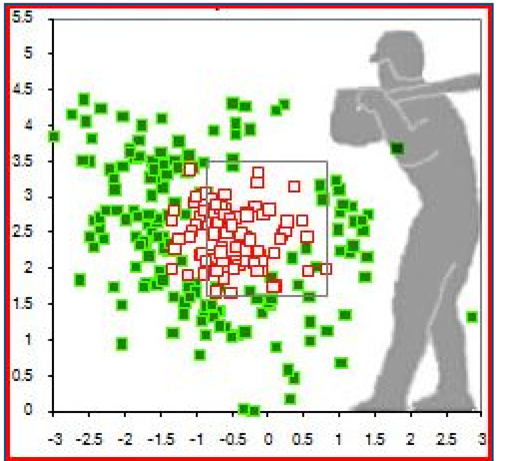
- Driving, flying training
- Disaster preparedness
- Science, engineering. Virus spreading
- Role playing
- Robotics. Biomechanics
- Games
- These are generally deterministic, not stochastic

Stochastic Simulation

- Stochastic simulation is sampling from a probability distribution to reproduce behavior of a system.
- Can help explain and illustrate difficult concepts.
- Can solve problems that are hard or impossible to solve directly.
- Can be fun.

Stochastic in Other Disciplines

- Modelling the macro economy – inflation, interest rates, govt spending, balance of trade, etc.
- Modelling the longevity of Social Security, and other insurance or business entities.
- Stock market, retirement simulations.
- Individual and team sports. Moneyball.
- Project management, queuing.



Simulation Process

- Construct a model.
- Design the experiment.
- Generate random #s from prob distribution(s).
- Analyze results.

Nice Online Simulations

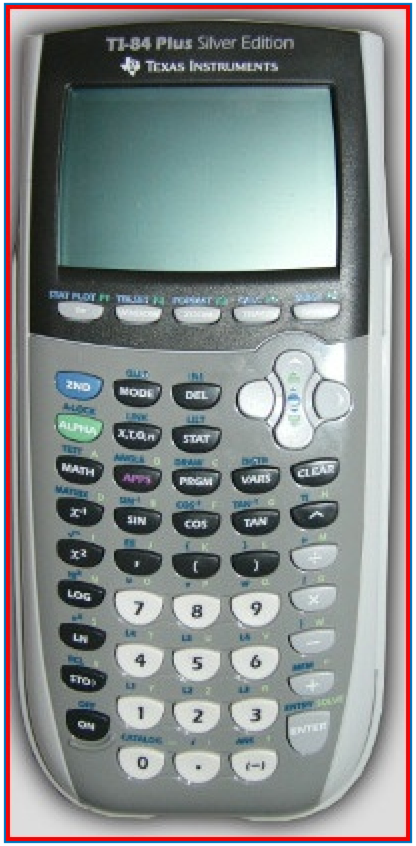
- Rice Univ: onlinestatbook.com/stat_sim/index.html
- Rossman/Chance: rossmanchance.com/applets/index.html
- Visualizing Statistics – UBC: www.zoology.ubc.ca/~whitlock/kingfisher/KFhomepage.htm
- DIGMATH – Farmingdale State Coll: www.farmingdale.edu/faculty/sheldon-gordon/dynamicstatistics.shtml

- But – let’s build our own in Excel !

Generating Uniform Random #s

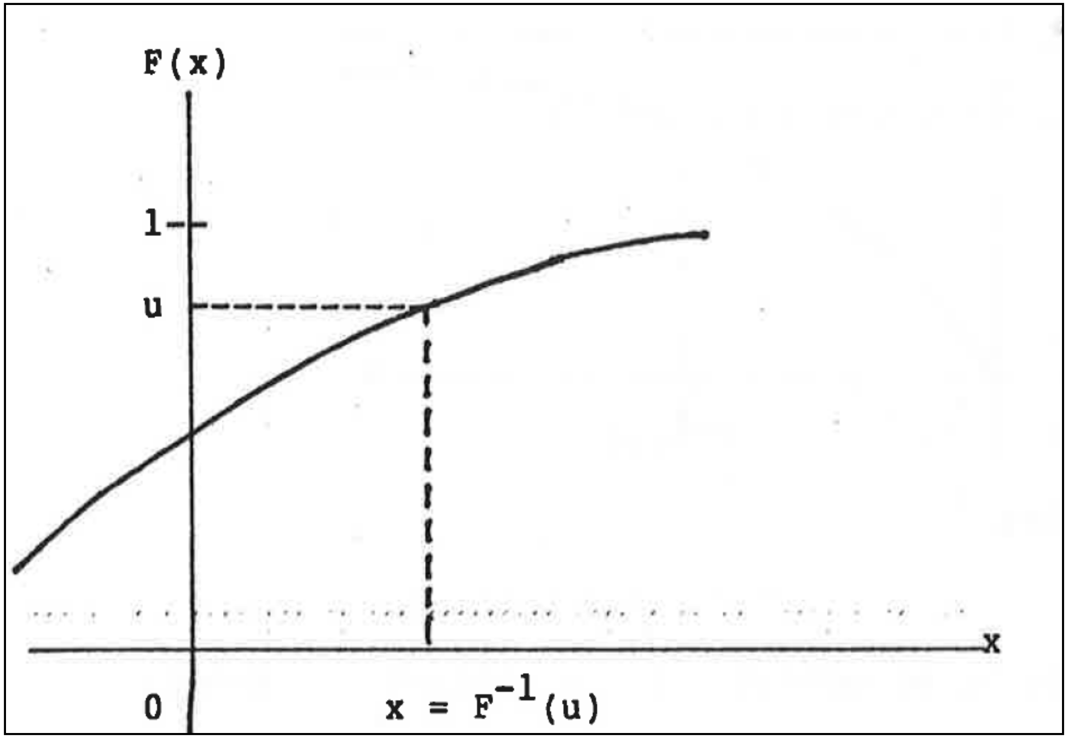
- Common algorithm to generate pseudorandom #s is $x_n = a*x_{n-1}$, mod m, for seed x_0 and large m.
- Here, x_n/m is approx a uniform U(0,1) variable.
- Excel has been criticized over insuff randomness, and is now using Mersenne Twister algoritm whose cycle is Mersenne prime.
- = rand() is random # U(0,1).
- =randbetween(min,max) : integer U(min,max)
- =(b-a)*rand() + a is random # U(a,b)

TI 84 & Random #s



- MATH → PRB
- rand(): U(0,1)
- randInt: U integer (min, max)
- randNORM: N(μ, σ); not σ^2.
- randBIN: Binomial(n, p)
- (b-a)*rand() + a: U(a,b)

Generating Non-Uniform Random #s



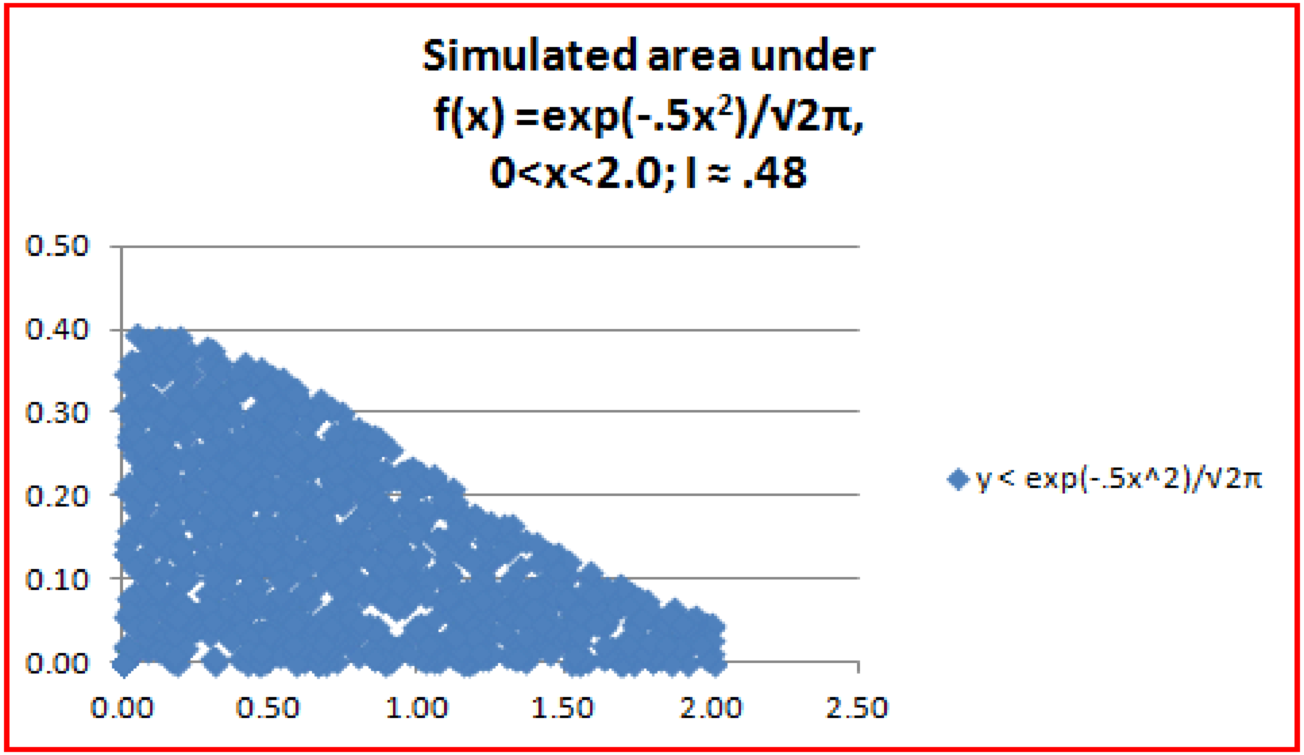
- A cumulative distribution function F(x) has inverse function $F^{-1}(z)$ in [0,1]. F can be discrete.
- If U is uniform random # in [0,1], then $x = F^{-1}(u)$ has cdf F.
- =NORM.INV(prob, μ, σ) Also GAMMA.INV, LOGNORM.INV

Some Excel Examples

- Definite integrals
- Coin flip – run of 6 consecutive heads or tails
- Birthday match; Monty Hall; Buffon's needle
- Sampling distribution
- Central Limit Theorem, with Normal & skewed populations
- Confidence intervals
- All Excel files available at professortuttle.com/amatyc

$$I = \int_0^2 \frac{exp(-\frac{x^2}{2})}{\sqrt{2\pi}} \, dx$$

- No simple anti-derivative exists from elem functions. (How do we know that?)
- Usually approximate such integrals with Riemann sums, Taylor series, others.
- If you can evaluate f(x) at any particular x value, you can approximate its definite integral with computer simulation.



- Simulate (x,y), count proportion of times that simulated y<f(x), multiply by rectangle area.
- Here, xЄU(0,2), yЄU(min f,max f) =U(0,1). Area ≈ (1196/5000)*2 = .48
- Similar problem approx π from circle.

SIMULATION FOR INTEGRATION

rand x	rand y	use x, if y < exp(-.5x^2)/√2π	use y, if y < exp(-.5x^2)/√2π
0.248285	0.19777	0.248285	0.197770
1.962702	0.101583	0.000000	0.000000
1.448921	0.733594	0.000000	0.000000
1.90526	0.940105	0.000000	0.000000
1.457725	0.139651	0.000000	0.000000
0.616879	0.511935	0.000000	0.000000
0.455571	0.60178	0.000000	0.000000
1.386689	0.367289	0.000000	0.000000
0.28517	0.661844	0.000000	0.000000
1.887179	0.661163	0.000000	0.000000
1.002119	0.782554	0.000000	0.000000
0.272148	0.642138	0.000000	0.000000
0.856317	0.819397	0.000000	0.000000
1.353089	0.044934	1.353089	0.044934
0.928526	0.920055	0.000000	0.000000
0.653388	0.293804	0.653388	0.293804
0.077484	0.956054	0.000000	0.000000
0.760549	0.777678	0.000000	0.000000
0.512969	0.142533	0.512969	0.142533

f(x) = (exp(.5*(-x^2)))/(√2π)		
Range for x	0	2
Range for y	0	1
Area of rectangle		2
Total points		5000
Points under curve		1209
Simulated area		0.48

COIN FLIP SIMULATION: RUN OF 6 HEADS OR 6 TAILS

N = 200	RAND #	H or T?	RUN of 6?
1	0.893733	T	
2	0.521595	T	
3	0.345944	H	
4	0.391363	H	
5	0.031342	H	
6	0.638856	T	0
7	0.606769	T	0
8	0.118357	H	0
9	0.802198	T	0
10	0.145905	H	0
11	0.976564	T	0
12	0.450911	H	0
13	0.280225	H	0
14	0.405565	H	0
15	0.353925	H	0
16	0.109392	H	0
17	0.421241	H	1

#RUNS	19
TOTAL	20
PERCENT	95%

DATA TABLE

SIM #	RUN
1	
2	RUN
3	RUN
4	RUN
5	RUN
6	RUN
7	RUN
8	RUN
9	RUN
10	RUN
11	RUN

Excel Data Table

	A	B	C	D	E	F	G	H	I
1				DATA TABLE					
2					SAMPLE				
3		n = 10		SIM #	MEAN				
4	n	SAMPLE			=B16				
5	1	97.49		1					
6	2	91.73		2			Highlight desired table range,		
7	3	84.15		3			incl the key formula.		
8	4	96.84		4			DATA > WHAT IF > DATA TABLE		
9	5	121.77		5			Ignore Row Input Cell		
10	6	73.74		6			Choose distant Column Input Cell		
11	7	85.92		7			OK		
12	8	86.91		8					
13	9	105.51		9					
14	10	89.82		10					
15				11					
16	xbar	93.39		12					
17				13					
18				14					
19				15					
20				16					
21				17					
22				18					
23				19					
24				20					

BIRTHDAY MATCH SIMULATION

N = 30	RAND #	DATE	MATCH?
1	21	21-Jan	
2	293	20-Oct	
3	260	17-Sep	
4	130	10-May	
5	344	10-Dec	
6	233	21-Aug	
7	138	18-May	
8	231	19-Aug	
9	292	19-Oct	
10	70	11-Mar	
11	98	8-Apr	
12	117	27-Apr	
13	86	27-Mar	
14	122	2-May	
15	115	25-Apr	
16	331	27-Nov	
17	49	18-Feb	
18	178	27-Jun	
19	101	11-Apr	
20	121	1-May	
21	31	31-Jan	
22	1	1-Jan	
23	3	3-Jan	
24	124	4-May	
25	17	17-Jan	**YES**
26	169	18-Jun	**YES**
27	17	17-Jan	**YES**
28	264	21-Sep	
29	169	18-Jun	**YES**
30	358	24-Dec	

#MATCH	15
TOTAL	20
PERCENT	75%
EXPECTED	70.6%

$$= 1 - {}_{365}P_k / 365^k$$

DATA TABLE

SIM #	MATCH
1	MATCH
2	
3	MATCH
4	MATCH
5	MATCH
6	
7	MATCH
8	MATCH
9	
10	MATCH
11	MATCH
12	MATCH
13	
14	MATCH
15	MATCH
16	MATCH
17	MATCH
18	
19	MATCH
20	MATCH

MONTY HALL SIMULATION

N = 100	RAND #	DOOR 1	DOOR 2	DOOR 3	MONTY SHOWS	SWITCH	NO SWITCH
1	1	CAR	GOAT	GOAT	3	GOAT	CAR
2	1	CAR	GOAT	GOAT	2	GOAT	CAR
3	2	GOAT	CAR	GOAT	3	CAR	GOAT
4	2	GOAT	CAR	GOAT	3	CAR	GOAT
5	1	CAR	GOAT	GOAT	2	GOAT	CAR
6	3	GOAT	GOAT	CAR	2	CAR	GOAT
7	2	GOAT	CAR	GOAT	3	CAR	GOAT
8	2	GOAT	CAR	GOAT	3	CAR	GOAT
9	2	GOAT	CAR	GOAT	3	CAR	GOAT
10	3	GOAT	GOAT	CAR	2	CAR	GOAT
11	1	CAR	GOAT	GOAT	2	GOAT	CAR
12	1	CAR	GOAT	GOAT	2	GOAT	CAR
13	2	GOAT	CAR	GOAT	3	CAR	GOAT
14	2	GOAT	CAR	GOAT	3	CAR	GOAT
15	1	CAR	GOAT	GOAT	2	GOAT	CAR
16	2	GOAT	CAR	GOAT	3	CAR	GOAT

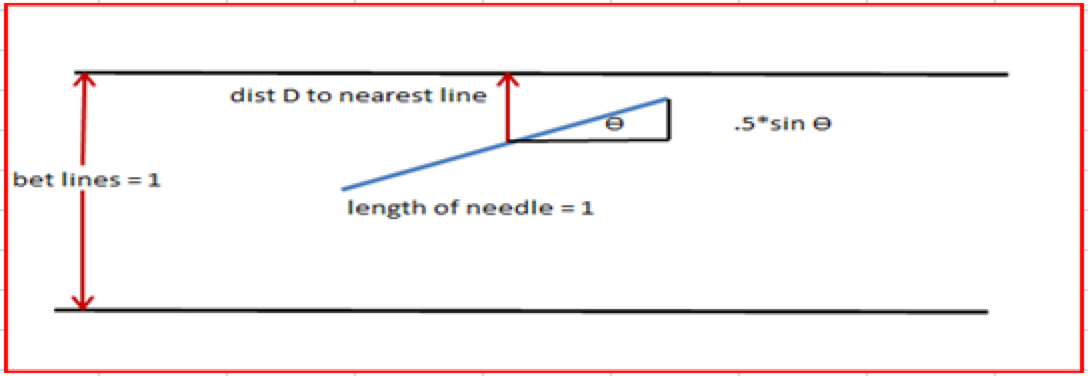
	SWITCH % CAR	NO SWITCH % CAR
SIMULATE	77%	23%
EXPECTED	67%	33%

RULES OF MONTY HALL GAME:

A car is randomly hidden behind one of three closed doors,
and goats are hidden behind the remaining two closed doors.
Contestant initially picks Door 1.
(The car is behind the random door of column B).
Monty opens a door not picked by contestant.
Monty opens a door that reveals a goat.
Monty offers contestant choice of switching to remaining closed
door, or not switching and keeping Door 1.

BUFFON'S NEEDLE SIMULATION (simplest case)
Length of needle = distance between lines = 1
Needle crosses a line if distance D from center of needle to closest line satisfies D ≤ .5*sinΘ

	D	Θ	HIT ?
N = 100	RAND #	RAND #	D ≤ .5*sinΘ ?
1	0.112365	100.8203	1
2	0.060878	47.90723	1
3	0.137762	1.938833	0
4	0.115839	88.01797	1
5	0.408763	40.69833	0
6	0.457648	55.44737	0
7	0.496481	131.3166	0
8	0.120049	100.3685	1
9	0.054349	24.39968	1
10	0.21657	159.0576	0
11	0.372209	149.8084	0
12	0.287214	77.13129	1
13	0.14769	52.82085	1
14	0.444877	138.1962	0
15	0.068408	163.7141	1
16	0.472185	137.1671	0
17	0.095532	94.16599	1



	% HITS	
SIMULATE	60%	
EXPECTED	64%	= 2 / π

CENTRAL LIMIT THEOREM SIMULATION (Normal Population)

N = 1000	RAND #	NORMAL POPULATION	RAND #	n = 30 SAMPLE
1	0.85788	182,084	938	(72,899)
2	0.126309	(39,401)	541	(17,603)
3	0.974762	270,590	679	185,304
4	0.491242	72,804	652	233,380
5	0.423971	55,825	469	279,159
6	0.447955	61,917	345	120,977
7	0.819979	166,528	753	146,721
8	0.117974	(43,517)	834	140,845
9	0.448779	62,125	680	147,824
10	0.622796	106,283	691	(11,049)
11	0.046078	(93,413)	359	206,375
12	0.722144	133,922	72	216,899
13	0.839353	174,180	39	223,517
14	0.994301	328,028	817	194,669
15	0.75403	143,723	273	(9,674)
16	0.884904	194,987	484	(66,185)
17	0.380568	44,601	464	135,248
18	0.716793	132,334	819	14,392
19	0.084864	(62,308)	886	159,094
20	0.210669	(5,410)	161	(67,573)

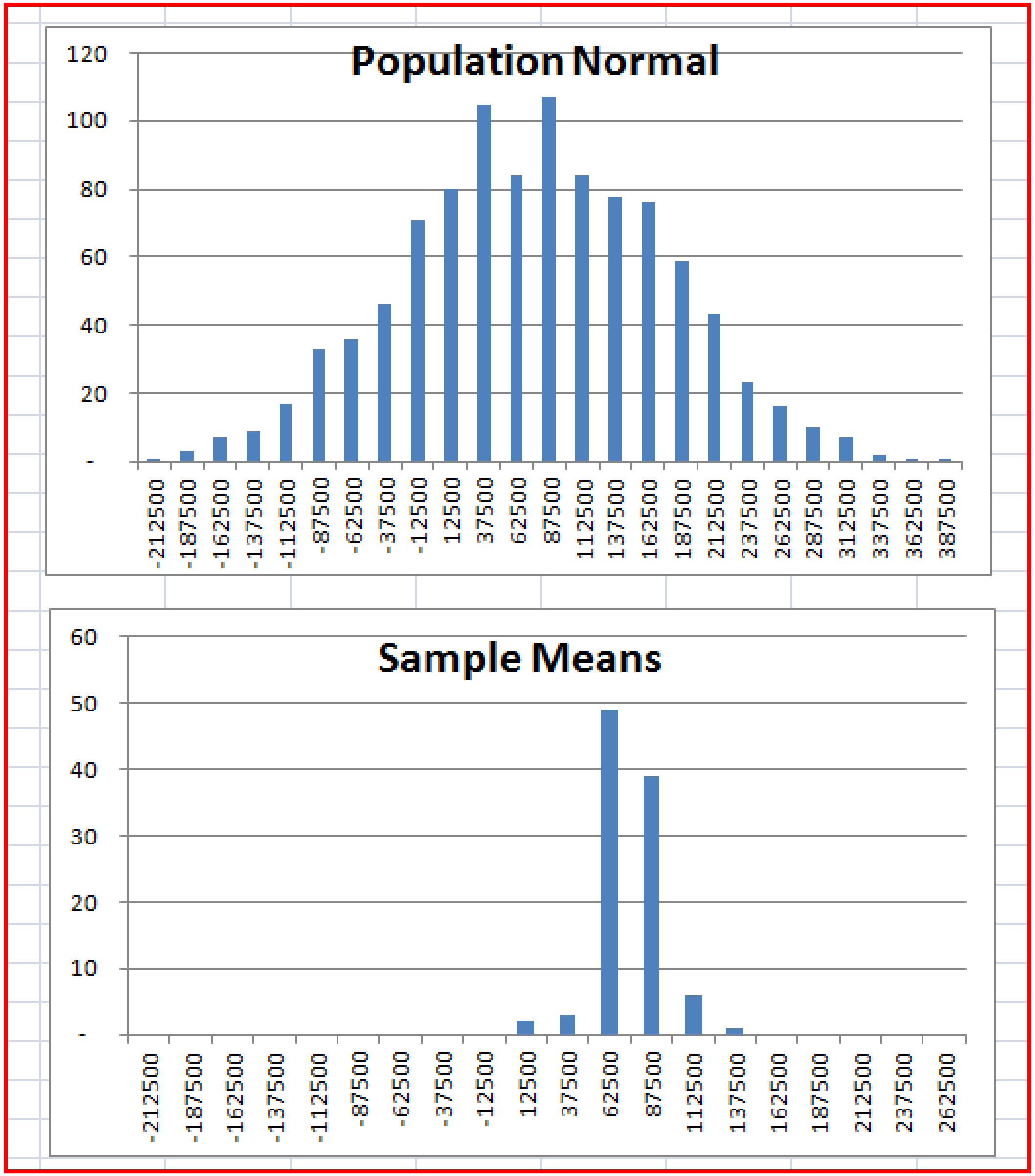
DATA TABLE

SIM #	SAMPLE MEAN
	87,985
1	92,248
2	21,684
3	126,677
4	81,560
5	101,715
6	51,858
7	89,292
8	79,039
9	67,346
10	56,896
11	55,205
12	75,005
13	63,372
14	74,686
15	76,620
16	62,466
17	92,018
18	78,989
19	62,635
20	83,377

Pop μ	75,406
Pop σ	99,140

Averages of samples:

xbar	74,026
s_xbar	17,819
σ / √30	18,100


Population Normal
120
100
80
60
40
20
-
-212500
-187500
-162500
-137500
-112500
-87500
-62500
-37500
-12500
12500
37500
62500
87500
112500
137500
162500
187500
212500
237500
262500
287500
312500
337500
362500
387500
Sample Means
60
50
40
30
20
10
-
-212500
-187500
-162500
-137500
-112500
-87500
-62500
-37500
-12500
12500
37500
62500
87500
112500
137500
162500
187500
212500
237500
262500

CENTRAL LIMIT THEOREM SIMULATION (Pareto Population)

N = 5000	RAND #	PARETO POPULATION	RAND #	n = 50 SAMPLE
1	0.225261	67,528	3,738	36,258
2	0.741881	30,506	3,222	60,633
3	0.522803	38,523	3,210	34,193
4	0.292647	56,717	3,835	35,496
5	0.637494	33,751	2,595	34,800
6	0.147259	89,650	2,742	43,769
7	0.571046	36,321	1,193	310,221
8	0.37536	48,044	4,506	112,600
9	0.65285	33,220	4,018	674,400
10	0.187097	76,424	2,461	46,627
11	0.706413	31,519	2,957	53,212
12	0.104879	112,412	3,596	45,416
13	0.585674	35,714	3,136	32,183
14	0.554942	37,020	3,233	35,179
15	0.989813	25,171	400	27,282
16	0.074924	140,668	3,403	28,317
17	0.661308	32,936	2,331	45,595
18	0.085989	128,325	3,414	30,255
19	0.598073	35,218	2,117	31,010
20	0.239673	64,793	3,780	81,273

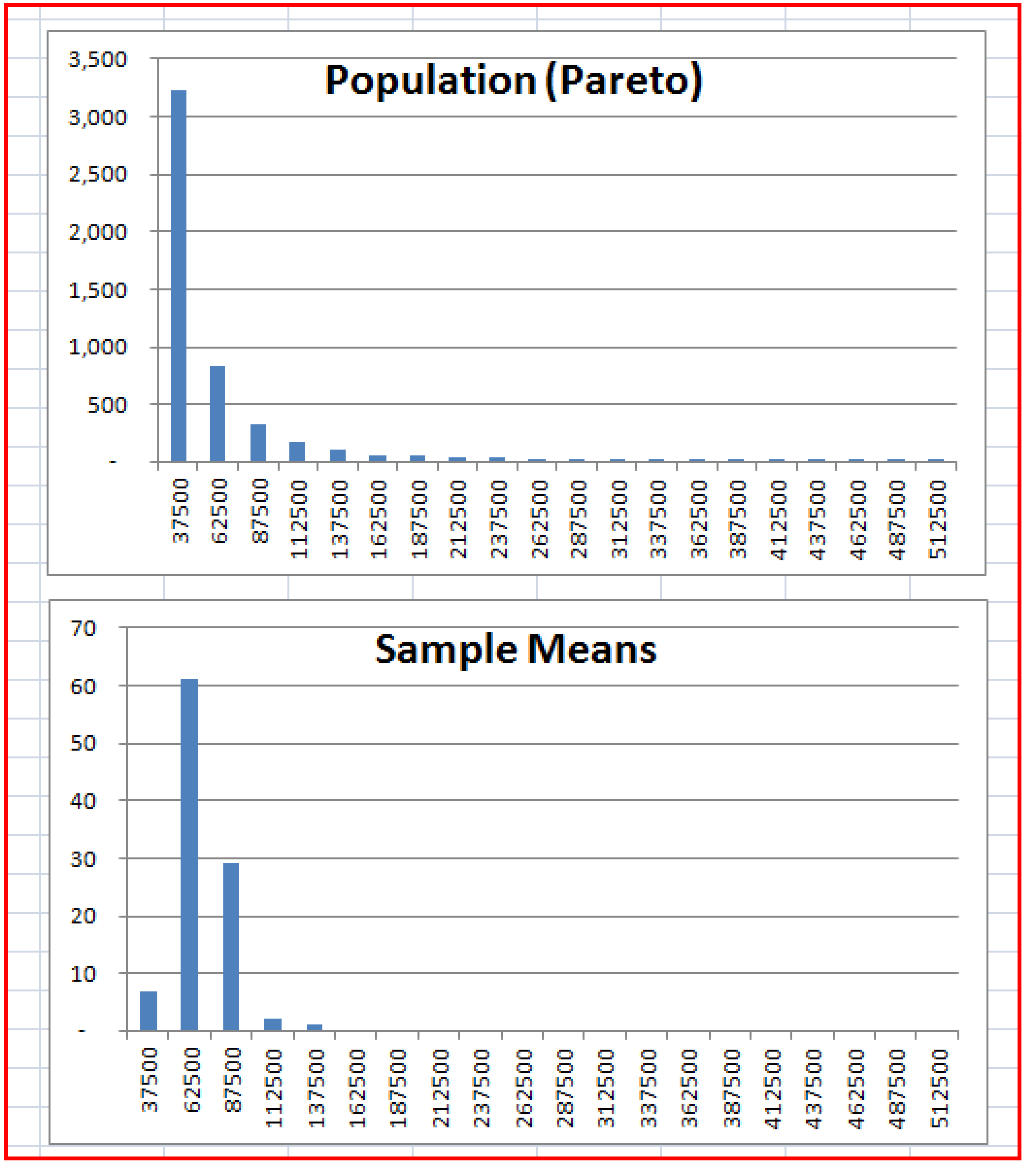
DATA TABLE

SIM #	SAMPLE MEAN
	74,654
1	76,453
2	61,993
3	49,159
4	83,947
5	75,929
6	79,708
7	54,160
8	61,553
9	56,426
10	111,522
11	78,826
12	58,895
13	90,812
14	68,255
15	68,513
16	55,763
17	57,219
18	59,265
19	81,743
20	66,800

Pop µ	66,674
Pop σ	93,964

Averages of samples:

xbar	68,467
s_xbar	15,119
σ / √50	13,289


Population (Pareto)
3,500
3,000
2,500
2,000
1,500
1,000
500
-
37500
62500
87500
112500
137500
162500
187500
212500
237500
262500
287500
312500
337500
362500
387500
412500
437500
462500
487500
512500
Sample Means
70
60
50
40
30
20
10
-
37500
62500
87500
112500
137500
162500
187500
212500
237500
262500
287500
312500
337500
362500
387500
412500
437500
462500
487500
512500

CONFIDENCE INTERVAL SIMULATION (Normal Population)

N = 30	RAND #	N(100,15) SAMPLE
1	0.729505	109.17
2	0.427737	97.27
3	0.838603	114.83
4	0.095694	80.40
5	0.069778	77.84
6	0.90101	119.31
7	0.069013	77.75
8	0.538799	101.46
9	0.709483	108.28
10	0.129084	83.04
11	0.435723	97.57
12	0.486013	99.47
13	0.009452	64.79
14	0.167338	85.53
15	0.536354	101.37
16	0.861084	116.28
17	0.237989	89.31
18	0.777641	111.46
19	0.252453	90.00
20	0.853521	115.77
21	0.378229	95.35

SIM #	xbar	s	E	xbar-E	xbar+E	Does CI contain μ ?
	99.02	15.93	5.70	93.32	104.72	
1	99.30	13.57	4.86	94.44	104.15	YES
2	99.87	16.15	5.78	94.09	105.65	YES
3	102.44	17.95	6.42	96.02	108.87	YES
4	98.57	12.75	4.56	94.00	103.13	YES
5	104.94	15.71	5.62	99.31	110.56	YES
6	99.22	13.90	4.97	94.25	104.20	YES
7	98.52	15.79	5.65	92.87	104.17	YES
8	95.81	13.14	4.70	91.11	100.51	YES
9	102.62	17.02	6.09	96.53	108.72	YES
10	103.22	17.36	6.21	97.01	109.44	YES
11	104.03	13.52	4.84	99.19	108.87	YES
12	99.23	14.84	5.31	93.92	104.54	YES
13	103.58	14.16	5.07	98.52	108.65	YES
14	102.52	17.64	6.31	96.21	108.84	YES
15	101.08	14.41	5.16	95.92	106.23	YES
16	102.68	13.83	4.95	97.73	107.63	YES
17	99.65	12.46	4.46	95.19	104.10	YES
18	100.21	12.76	4.57	95.65	104.78	YES
19	94.77	12.00	4.29	90.47	99.06	**NO**
20	95.23	15.34	5.49	89.74	100.72	YES
21	102.62	14.93	5.34	97.28	107.96	YES

Does CI contain μ = 100 ?

YES	94	94%
NO	6	6%
	100	

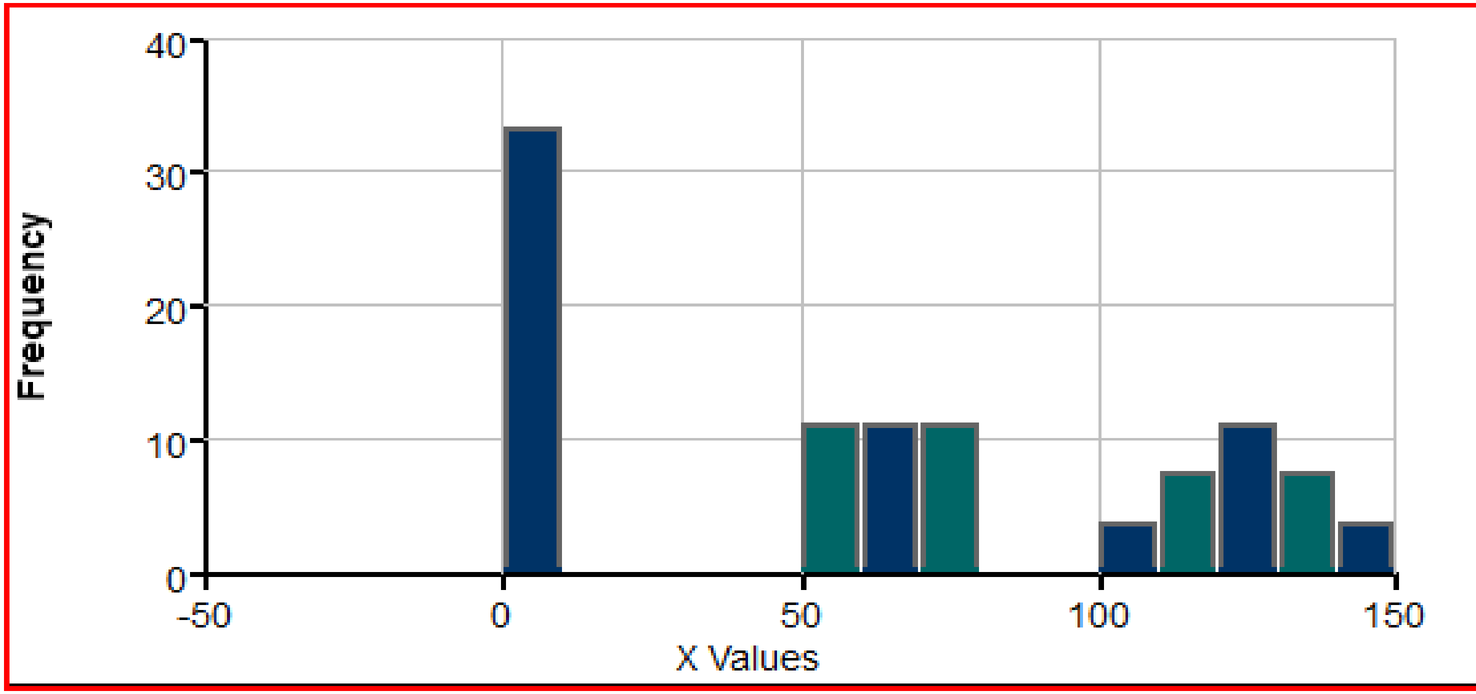
Two Actuarial Problems

- Suppose I insure 50 insurance policies.
- Each has avg frequency f = 1 claim/yr,
  and avg claim size s = $60,000.
- 1. What is 95th percentile of all claims?
- 2. How much should I pay to lay off $1M in excess of $3.5M to another insurer (reinsurer)?

Convolutions

- Problems 1 & 2 require we consider all possible #s of claims & claim sizes, not just expecteds.
- Ex: a. Suppose each policy may have either 0,1, or 2 claims; b. each claim is 50K, 60K, or 70K; and c. everything is equally likely.
- So the possibilities are 0 claims, 1 $50K claim, a $50K claim and a $60K claim, etc.
- List (convolute) all these possibilities and their probabilities. This is called aggregate claim distribution.

Aggregate claim distribution

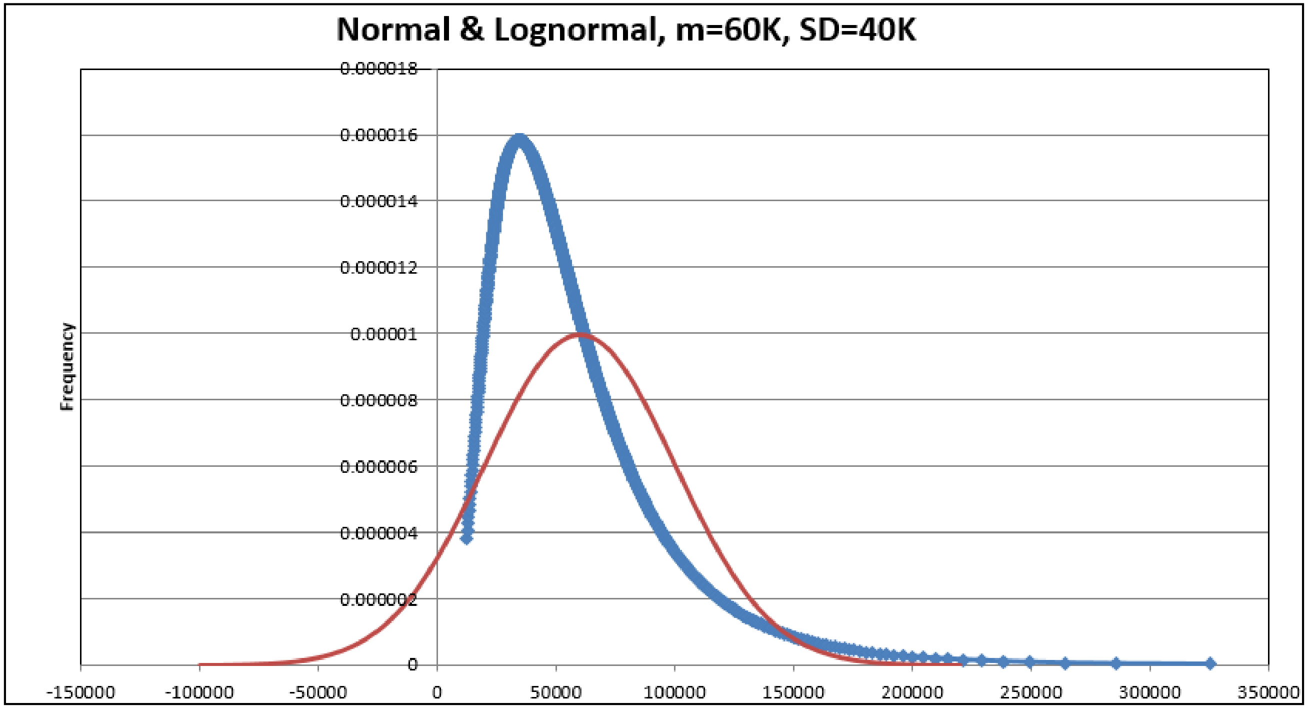


- As number of possibilities increases, listing them becomes unwieldy.
- Would be nice to have formula pdfs f(x) for number of claim distribution & g(x) for claim size distributiion.

Few convolution formulas

- There are many common pdfs, e.g., Poisson & Negative binomial for number of claims, Gamma, Lognormal, Pareto for claim sizes. While some convolutions have nice formulas, e.g., Poisson with Gamma, most do not.
- Claim size pdfs in insurance need to be positively skewed (most claims are small, i.e., fender-benders vs. total loss).
- Now we are getting to the simulation.

Claim Size Simulation



- =NORM.INV(rand(),60K,40K)
- =LOGNORM.INV(rand(),μ,σ), where here $\mu = LN(60K) - .5LN[(40/60)^2 + 1] = 3.91$ & $\sigma = \sqrt{LN[(40/60)^2 + 1]} = .61$

Simulation software result

- Simulate count N, then simulate N claims. Repeat 40K times, & sort aggregate claims.
- Neg binomial freq M=50, SD=10; & lognormal severity M=60K, SD=40K; 40K simulations:

Percentile	Agg Loss, in K
EL	3,000
50.0%	2,959
80.0%	3,546
90.0%	3,877
95.0%	4,161
98.0%	4,482
99.9%	5,409
Max	6,201
	Expected Loss
1000 xs 3500	90

Notes

- Gordon, S.P. and Gordon F.S. (Jul/Aug 2009).Visualizing and understanding probability and statistics: graphical simulations using Excel. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*,*19* (4) 346-369.
- Herzog, T.N. (1986). *Intro to stochastic simulation*. Society of Actuaries.
- Melard, G. (Oct., 2014.) On the accuracy of statistical procedures in Microsoft Excel 2010. Computational Statistics. 29(5). 1095-1128. http://homepages.ulb.ac.be/~gmelard/rech/gmelard_csda23.pdf
- Ross, S.M. (2013). *Simulation*. San Diego: Academic Press.
- Verschuuren, G.M. (2016). *100 Excel VBA Simulations*.

Images

- Die graphic from https://pixabay.com/en/dice-games-play-1294902/
- CPR graphic from https://pixabay.com/en/cpr-dummy-medical-training-course-1255746/
- Driver simulator graphic from http://uniquenovelinterfaces.blogspot.com/2014/02/car-driving-simulator.html
- Simulation graphic from https://www.hybris.com/medias/sys_master/root/h04/h0d/8814979711006/banner-factsheet-pricing-simulation.svg
- Baseball graphic from http://drspikecook.com/files/2012/11/sabermetrics-1xtkzx1.png
- Tree graphic from http://www.clipartkid.com/if-you-have-any-questions-with-the-calendar-please-email-6qv4gI-clipart/
- Excel graphic from https://play.google.com/store/apps/details?id=com.microsoft.office.excel
- Calculator graphic from https://images-na.ssl-images-amazon.com/images/I/71RH7Vj8FUL._SY355_.jpg
- Inverse function graphic from Herzog, T.N. (1986). *Intro to stochastic simulation*. Society of Actuaries.
- CAS graphic from http://www.casact.org/about/index.cfm?fa=contact