2001 Joint Statistical Meetings

Atlanta, GA

An Excel Add-In for Capturing Simulation Statistics

David C. Trindade

david.trindade@sun.com

Sun Microsystems, Inc.

Cupertino, CA

David Meade

david.meade@amd.com

Advanced Micro Devices, Inc.

Austin, TX

Uses of Computer Simulation

Powerful tool to recreate possible outcomes of random processes

Allows checking models and analysis procedures without extensive experimentation

Where theory is limited, may be the only avenue for solving complex problems

Source of data for training

Outline

- Illustrate the capability of computer simulation using spreadsheets for solving many different types of simple and complex problems in design, manufacturing, analysis, etc.
- Describe a general procedure for simulating data for any distribution
- Apply to example for estimating parameters of Weibull distribution
- Provide a macro to capture distribution statistics from specified simulation runs
- Present example results to show power of spreadsheet simulation for analyzing and modeling random processes

Simulating a Distribution

We want to **produce failure times** from a **specific distribution**.

To simulate **any distribution**, we will use a special property of the cumulative distribution function (**CDF)**.

*Let **U** be a uniform random variable in the interval (0,1).*

*If **F(t) is any CDF** with an **inverse t** = F^{-1}, then substituting **U** for **F** in the inverse expression generates a **random variable T** distributed according to **F**.*

Weibull Distribution Example

The CDF for a Weibull distribution with characteristic life ***c*** and shape parameter ***m*** is

$$F(t) = 1 - e^{-(t/c)^m}$$

F(t) is the probability of failure by time ***t***.

The inverse expression is :

$$t = F^{-1} = c\left[-\ln(1-F)\right]^{1/m}$$

Simulating Weibull Times to Failure

Substitute the unit uniform random variable *U* in the inverse expression to get :

$$T = F^{-1}(U) = c\left[-\ln(1-U)\right]^{1/m}$$

Since both 1-*U* and *U* are a uniformly distributed variate in the interval (0,1), simplify the inverse expression as :

$$T = c\left[-\ln U\right]^{1/m}$$

Each *U* value produces a **time to failure** *T* from a Weibull distribution with parameters ***c*** and ***m***.

Weibull Simulation Example in Spreadsheet

- Estimating the parameters ***c*** and ***m*** of a Weibull distribution is not simple, even for complete data. For MLE's, iterative numerical methods are required.

- Let's investigate the suitability of the following simple procedure for estimating the parameters.

- Write the 25th and 75th percentiles for a Weibull distribution as:

$$t_{25} = c\left[-\ln(1-0.25)\right]^{1/m} = c\left[\ln(4/3)\right]^{1/m}$$

$$t_{75} = c\left[-\ln(1-0.75)\right]^{1/m} = c\left[\ln(4)\right]^{1/m}$$

Parameter Estimation for Weibull Distribution

Solve the second equation for ***c*** and substitute into the first to get

$$t_{25} = t_{75} \frac{\left[\ln(4/3)\right]^{1/m}}{\left[\ln(4)\right]^{1/m}} = \left[0.207519\right]^{1/m}$$

or

$$m \ln\left(\frac{t_{25}}{t_{75}}\right) = \ln(0.207519) = -1.572534$$

Solve for ***m*** directly in terms of the ratio of the two percentiles.

$$\hat{m} = \frac{-1.572534}{\ln\left(\frac{t_{25}}{t_{75}}\right)}$$

Parameter Estimation for Weibull Distribution

With ***m***, we can easily solve for ***c*** using any convenient percentile such as the median, that is,

$$\hat{c} = \frac{t_{50}}{[\ln(2)]^{1/\hat{m}}}$$

How well does this procedure work for fitting a Weibull model to data?

To get an answer, let’s try simulation in a spreadsheet.

Weibull Distribution Simulation in Spreadsheet

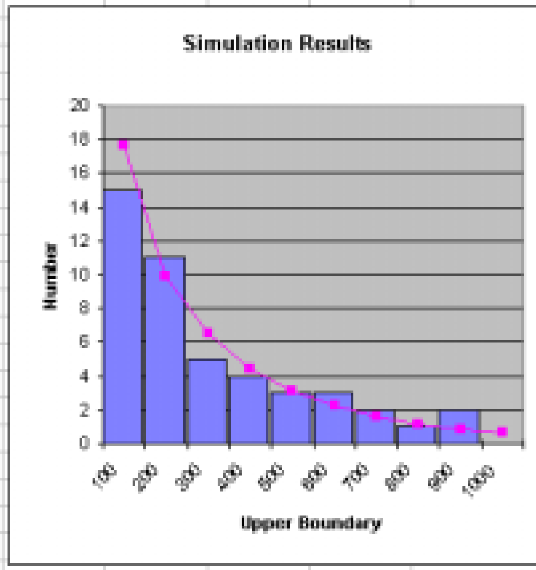
We generate 50 pseudo-random numbers ***U*** in a column using the spreadsheet function =**rand()**.

The adjacent column, based on the inverse CDF calculation, contains the random times distributed according to a Weibull with specified parameters ***c*** and ***m***.

After estimating the parameters according to the described procedure, we produce a histogram of the data along with a model Weibull fit.

Simulation Results After One Run

Weibull Parameters	c =	300	m =	1	
U	T		Bins	Freq.	Model Fit
0.733	93		100	15	17.6
0.595	156		200	11	9.9
0.806	65		300	5	6.5
0.838	53		400	4	4.5
0.825	58		500	3	3.1
0.064	824		600	3	2.2
0.645	132		700	2	1.6
0.768	79		800	1	1.2
0.473	225		900	2	0.8
0.639	135		1000	0	0.6
0.114	651		**Sum**	46	
0.947	16				
0.755	84		**25th Pcntl**	80	
0.067	812		**Median**	170	
0.233	437		**75th Pcntl**	479	
0.449	240				
0.092	715		**m est**	0.88	
0.171	530		**c est**	257	
0.680	116				
0.554	177				
0.027	1084				
0.375	294				

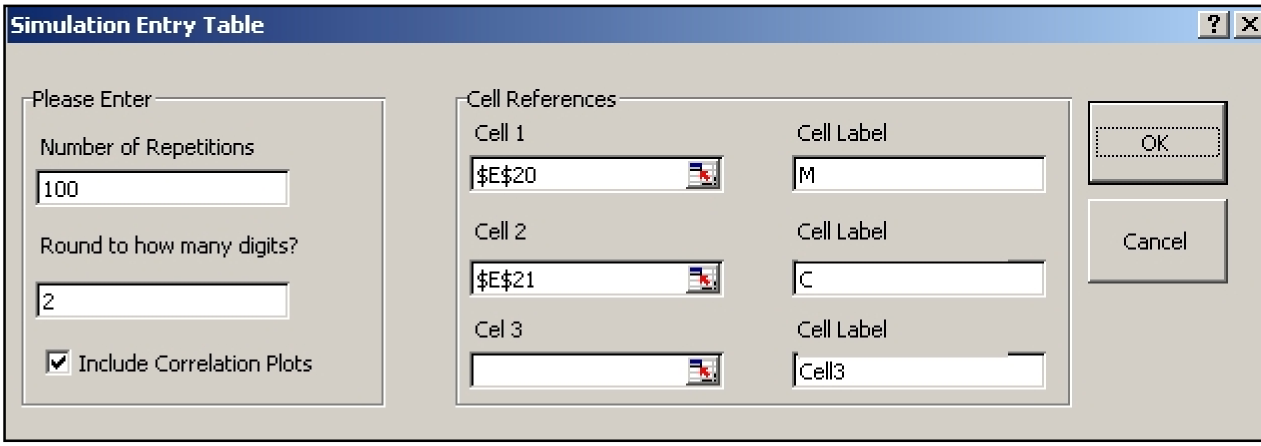


Adding Simulation Runs

By making additional simulation runs, and collecting parameter estimates from each simulation, we can determine many facets of the problem, such as:

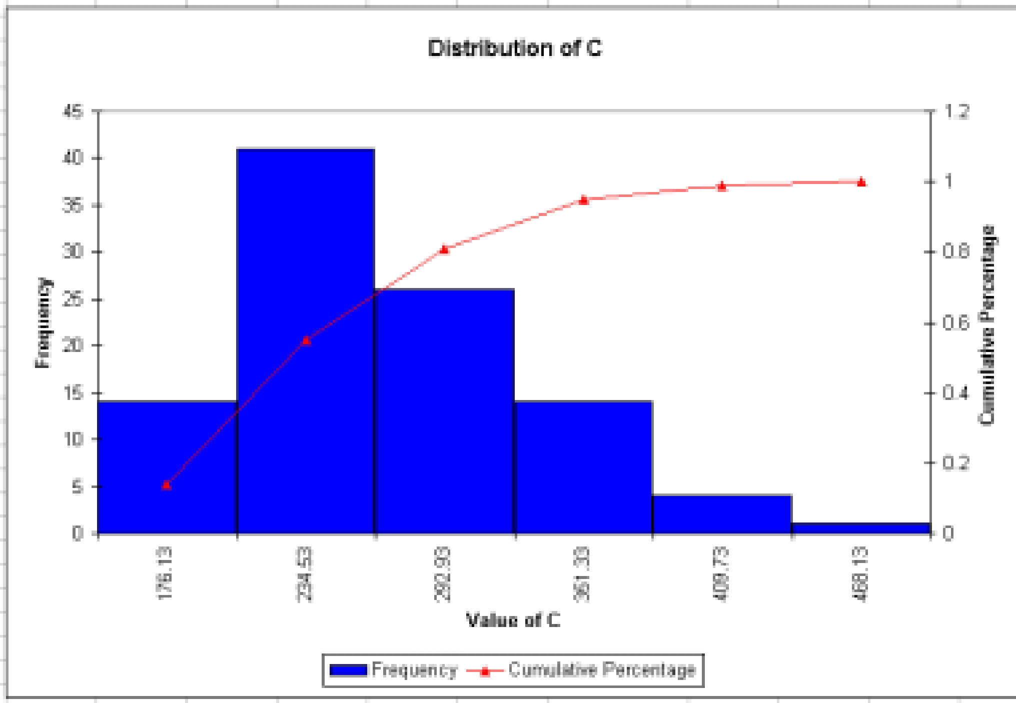
- the sampling distribution of the estimates
- confidence intervals on the estimates
- the suitability of the Weibull model based on these estimates

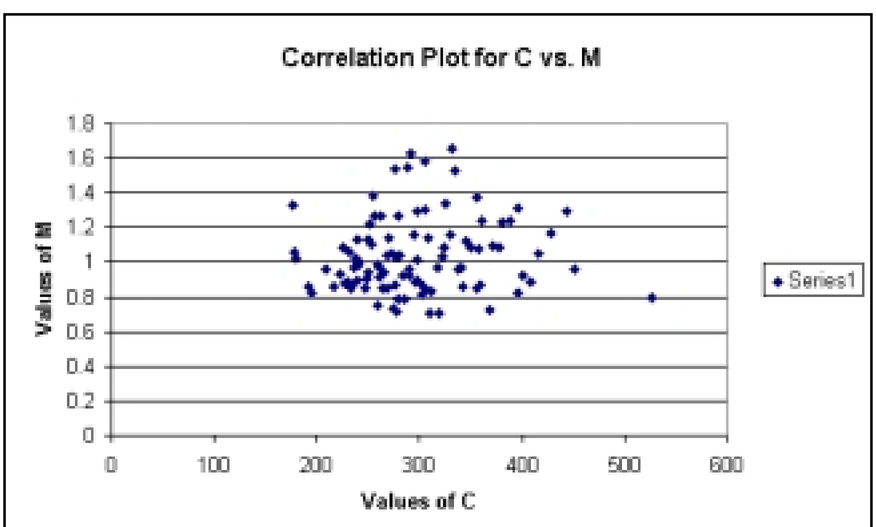
EXCEL Macro Add-In



Macro Output for One Parameter

Value of C	Bin	Frequency	Cumulative Percentage		
276.5156	176.13	14	0.14	Mean	295.6787
239.9777	234.53	41	0.55	Median	284.62
273.8751	292.93	26	0.81	Stdev	64.07935
302.4419	351.33	14	0.95	Variance	4106.163
269.3063	409.73	4	0.99	Max	526.5397
246.5857	468.13	1	1	Min	176.1325
277.8623				Count	100
238.6206				5th Percentile	208.0297
238.4521				95th Percentile	408.068
262.6684					
296.9746					
331.809					
259.0179					
356.709					
261.4644					
280.6267					
249.5432					
334.283					
330.6841					
227.0141					
309.4868					
284.3952					
323.0608					
275.4762					
308.3676					
427.5994					
263.5015					
349.5224					
288.797					
176.1325					
526.5397					
229.8454					
318.5078					
267.7796					
275.4182					
379.9728					
177.5735					
237.81					





Benefits of Simulation Studies

Check adequacy and variation in parameter estimation formulas

Determine confidence intervals along with model estimates

Try different representations of displaying results

Experiment with new estimation formulas

Show range of variability in possible system behavior

Summary

Simulation is a powerful tool for understanding and analyzing repairable system processes.