Probabilistic Availability Risk Assessments without Simulation

Jan B. Smith, PE, Reliability Engineering Consultant

Key Words: Probability, Risk Assessment, Availability, Capacity, Probabilistic Assessment, Probability Distributions

SUMMARY & CONCLUSIONS

Two methods for calculating availability and capacity probability distributions are discussed. One gives exact results, but has narrow application. The other is approximately exact but is a general method applicable to time varying data. Each method uses high level system data that is both available and accurate.

The two methods can be implemented using personal computers and spreadsheets. Templates with cell equations are provided so the user can build their own working spreadsheets. The first method presented is capacity outage probability tables (COPT) used routinely in the electric power industry. It calculates exact empirical instantaneous availability and capacity distributions for limited but practical applications. The second method is probability mass function (PMF) series, a general method for measuring accumulated system performance with a dense family of empirical distributions.

COPT is mathematically exact and PMF series is essentially so. PMF series is computer intensive with accuracy increasing with increasing computations. Excellent results are achievable on personal computers. In an example, COPT and PMF series results are shown to be comparable in the special case in which these totally different concepts can be compared.

Two non-conventional applications will be discussed as examples. One of these is personal injuries recorded by an industrial maintenance contractor. The probability and risk associated with any performance-based contract safety goal can be determined. If risk (measured in units of recorded injuries) is high, it can easily be identified and managed down by changing the "system" that generates injuries. Alternatively, new contract commitments can be negotiated, again with the new risk calculated.

The other nonconventional application is a business system that delivers a medical services product. The risk associated with management's goals for the next month or the next quarter can be assessed. Risk-based goals can be set, and once goals are established, unacceptable risk can be managed by changing the business system.

Computer simulation models or other traditional methods are not now being routinely used for these non-conventional applications. Likewise there are numerous risk assessment applications residing within engineered physical systems, such as manufacturing plants, that are underserved. Also, traditional applications, such as capacity assessment of manufacturing plants, can benefit due to no modeling requirement and the availability of high quality data.

1 CAPACITY OUTAGE PROBABILITY TABLES (COPT)

The electric power industry uses a system risk assessment method that is also applicable to other systems. The method is capacity outage probability tables (COPT) [1]. The forced outage rate and capacity of each individual generating unit are the input data. Only basic probability theory and simple arithmetic are used with this fundamental and available data for the system instantaneous capacity probability. The load demand curve is combined with the capacity distribution to obtain the loss of load probability (LOLP). The risk of inability of capacity to meet demand is the loss of load expectation (LOLE) in MW-hrs. Generating units are shutdown or started up as necessary to keep the risk within targeted values. This allows efficient operation of the system (minimum generating capacity on line) while controlling the expected value of the generating system capacity shortfall to an acceptable level.

COPT can be used for other applications. For example, when several steam generators in a refinery fail to produce a certain quantity of steam, refinery processes must be reduced in rate or shut down. The risk to the refinery is the product of probability of steam load being below demand (LOLP) and the economic impact when the steam supply is below demand.

Figure 1 is a fictitious system composed of three units that illustrates the COPT input and output data. The simple and readily obtainable unit input data of Figure 1 are placed in

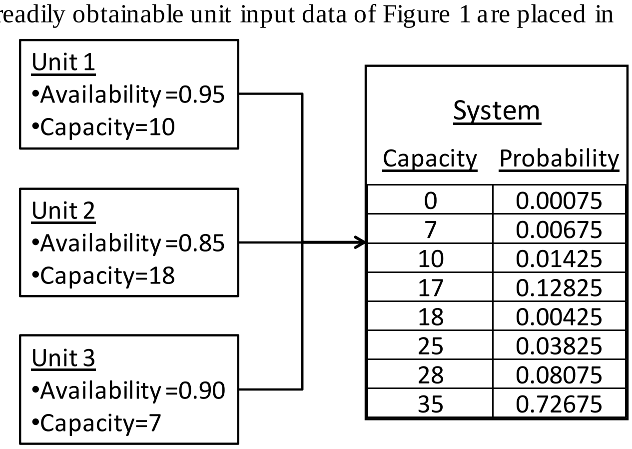


Figure 1

Illustration with a 3 unit system with unit availability and capacity used to calculate the system capacity probability distribution as shown in Table 1.

rows 2 and 3 of Table 1 to demonstrate how the COPT process is applied. The reader can use Table 1 as a template for building a larger spreadsheet to accommodate systems with more units. Only fundamental probability theory is used and the cell equations are shown. System operating state probability and the capacity at that state are calculated in columns H and L, respectively. Each operational state defines which units are operating or otherwise available, and which are under forced shutdown (unavailable). The number of operational states increases exponentially with the number of units in the system. For example, a system of ten units has 1,024 unique operational states. Units that are planned down are assigned an availability of zero during this time. Likewise, a spreadsheet for, say, 10 units can be used to assess a system of less than 10 units by assigning zero availability to the non-existing units.

The electric power industry uses input data of forced outage rate. Table 1 uses availability here defined as the complement of forced outage rate. Therefore, scheduled outages and other outages not related to forced downtime from unreliability, such as market dictated downtime, are appropriately excluded from the unit availability metric.

The binominal distribution should always be used to check the accuracy of the spreadsheet. When unit availabilities are equal, the spreadsheet system probability for each operating state will exactly equal the probability from the binominal distribution. If not identical, an operating state may have been missed or duplicated in building the spreadsheet.

This method gives exact results when unit operations are either up or down. Columns H and L produce the system instantaneous capacity discrete probability distribution. When unit capacities are normalized to a system maximum capacity of unity, columns H and L give the instantaneous availability discrete probability distribution.

The results of Table 1 are combined with a load or demand curve, as illustrated in Table 2. This spreadsheet determines the probability of failure to meet demand and the expected value of the capacity shortfall (equivalent to LOLP and LOLE). Again, Table 2 is designed to serve as a guide in setting up spreadsheets for larger systems and more complex demand curves. Cell equations are shown. The demand curve for Table 2 consist of three discrete loads, each applicable for some fraction of the total time. Demand curves can be as simple as a single value or as complex as a probability distribution. If demand is characterized by a continuous distribution, then the distribution should be approximated with discrete values. Whereas Table 2 has only three demand levels, an approximated demand distribution may have many levels as necessary for the appropriate accuracy. Except for approximation of the demand curves, the results of Table 2 are also mathematically exact.

The COPT process is completed with Tables 1 and 2 for our simple Figure 1 example. As Table 2 row 12 shows, there is a probability of a capacity shortfall of 21.38% (LOLP) and the expected value of the shortfall is 1.82 capacity units per time unit (LOLE). Units are the same as Table 1 row 3 units.

2 PROBABILITY MASS FUNCTION (PMF) SERIES

PMF series [2] is a new technology for probabilistic assessments. It is a fundamentally different concept from that of COPT or any other risk assessment method. It uses high level system data that is already on hand, such as in the case of engineered systems, time-between-failure (TBF) and time-to-restore (TTR). Where COPT produces a single instantaneous capacity distribution, PMF series produces a dense family of accumulated performance probability distributions.

The technology is not limited to physical systems. It is applicable to systems in the broadest sense of the word where the concepts of failure and availability do not apply and where the system can only be treated as a black box, as illustrated in

Cell equations shown in the table: E5 =IF(B5=“DOWN",1-B2,B2); H5 =E5*F5*G5; I5 =IF(B5=“DOWN",0,B3); L5 =SUM(I5:K5)

	A	B	C	D	E	F	G	H	I	J	K	L
1	Unit Number	1	2	3								
2	Unit Availability	0.95	0.85	0.9								
3	Unit Capacity	10	18	7								
4	System Operating State	Unit Operating State			Unit Probability at Operating State			System Probability at Operating State	Unit Capacity at Operating State			System Capacity at Operating State
5	1	UP	UP	UP	0.95	0.85	0.9	0.72675	10	18	7	35
6	2	UP	UP	DOWN	0.95	0.85	0.1	0.08075	10	18	0	28
7	3	UP	DOWN	UP	0.95	0.15	0.9	0.12825	10	0	7	17
8	4	DOWN	UP	UP	0.05	0.85	0.9	0.03825	0	18	7	25
9	5	UP	DOWN	DOWN	0.95	0.15	0.1	0.01425	10	0	0	10
10	6	DOWN	UP	DOWN	0.05	0.85	0.1	0.00425	0	18	0	18
11	7	DOWN	DOWN	UP	0.05	0.15	0.9	0.00675	0	0	7	7
12	8	DOWN	DOWN	DOWN	0.05	0.15	0.1	0.00075	0	0	0	0
13								Sum = 1.00000				

Table 1

Capacity outage probability table (COPT) for the Figure 1 system. This table is to be used as a template in building spreadsheets for larger systems. All possible operating states must be listed (B5:D12). For risk assessment, system capacity and probability results from columns H and L are combined with a load demand curve, as shown in Table 2.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Demand Level	30	25	20								
2	Time Fraction at	0.5	0.3	0.2								
3	System Capacity at Operating State	System Probability at Operating State	Loss of Load Probability at Operating State	Time Fraction Capacity < Demand	System Risk at Operating State	Total System Shortfall at Operating State	Capacity Shortfall at Demand Level			Time Fraction Capacity < Demand		
4	35	0.72675	0	0	0	0	0	0	0	0	0	0
5	28	0.08075	0.040375	0.5	0.08075	1	2	0	0	0.5	0	0
6	17	0.12825	0.12825	1	1.218375	9.5	13	8	3	0.5	0.3	0.2
7	25	0.03825	0.019125	0.5	0.095625	2.5	5	0	0	0.5	0	0
8	10	0.01425	0.01425	1	0.235125	16.5	20	15	10	0.5	0.3	0.2
9	18	0.00425	0.00425	1	0.036125	8.5	12	7	2	0.5	0.3	0.2
10	7	0.00675	0.00675	1	0.131625	19.5	23	18	13	0.5	0.3	0.2
11	0	0.00075	0.00075	1	0.019875	26.5	30	25	20	0.5	0.3	0.2
12			0.213750		1.817500							

- •A3 and B3 values from Table 1
- •C4 = B4*D4
- •D4=SUM(J4:L4)
- •E4=F4*B4
- •F4=G4*B2+H4*C2+I4*D2
- •G4=IF(A4>=B1,0,(B1-A4))
- •J4=IF(A4>=B1,0,B2)
- •C12=System Loss of Load Probability (in time period of A1 units)
- •E12=System Risk (measured in A1 units)

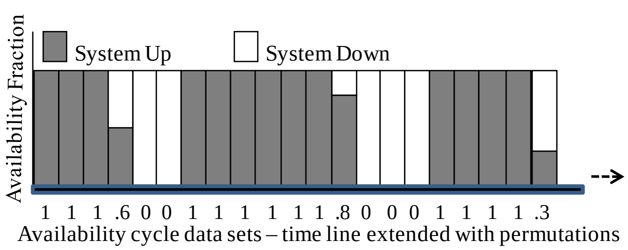
Table 2

Operating state capacity and probability are combined with a demand curve. The demand in this example is 3 discrete levels with the time fraction at each level. Cell equations are shown so the table can be used as a template to form spreadsheets for more complex systems and demand curves.

two examples to follow. It is also applied to reliability assessment of both repairable and non-repairable systems and time-to-first-failure of components and systems [2].

PMF series transforms conventional data into a new numbering system. Figure 2 summarizes the data conversion with a simple example of three failure events identified by TBF and TTR data [3]. The failure data can be historical, simulated, or surrogate (data from a similar system). It is not

	Conventional format		PMF Series format
Failure Number	TBF	TTR	Availability Cycles
1	3.6	2.4	1,1,1,.6,0,0
2	6.8	3.2	1,1,1,1,1,1,.8,0,0,0
3	4.3	0.7	1,1,1,1, 3



All permutations of the 3 independent failure events :
123, 132, 213, 231, 312, 321

Figure 2

Transformation of TBF and TTR data to availability cycle data sets consisting of values between 0 and 1for each small time interval is demonstrated with 3 failure events. These data sets form a time line that can be extended with permutations (arrangements) of the independent failure events.

necessary that raw data be in TBF and TTR form, but Figure 2 uses this form. Events can be cyclical or changing over time and dependent; however, Figure 3 is for a stable process and independent failure events. The TBF and TTR data are converted into availability cycles defined by a series of numbers ranging from 0 to 1. These cycles are laid end to end to form a time line of availability fractions for the smallest time interval of interest (STII) [3]. The STII is the unit of time for which we cease to be concerned for variation. For example, the STII for a manufacturing process with product storage is usually one day. For such a plant, there is typically no business purpose in knowing hour-to-hour capacity variation. The STII for a business system and a “system” that generates industrial injuries is one week in examples to follow. The availability cycle, the sum of TBF and TTR, must be made to sum to a whole number by rounding or changing units. The concept is computer intensive, and computations can be reduced by keeping the STII no smaller than necessary. The original data after transformation are seen graphically in the middle of Figure 2.

System failure events are typically independent. In a petrochemical plant, for example, within a database period of a few years the same equipment with the same failure mode will usually not be repeated. Reliability engineers work hard to assure this is so. As soon as a failure occurs, action is taken to prevent it from repeating. Dependent events, systems undergoing improvement or deterioration and cyclical performance cannot be reordered. These require a modification to the methodology not discussed in this paper. The six (3! or 3 factorial) ways the three failure events can be reordered are seen at the bottom of Figure 2. The time line is extended by using all data permutations, though not shown in Figure 2 due to space. Greater detail has been published [3].

Table 3 shows the extended time line formed in Figure 2 placed in a spreadsheet – column B. Only the first permutation (original data) and part of the second permutation are shown, but the reader can easily build their own spreadsheet from the pattern established. Permutation and failure event numbers are seen in column A with the availability cycle data forming the time line in column B. For some applications, all series over a long interval of time is needed. For example, the risk of not meeting a daily shipping schedule requires evaluating risk one day into the future, two days ahead, three days ahead, etc. This may continue for 365 days. In Table 3 we limit the series for

	A	B	C	D	E	F	K
1	Permutation/ Failure Number	1-Time Unit	2-Time Units	3-Time Units	4-Time Units	5-Time Units	10-Time Units
2	1/1	1.0000					
3		1.0000	1.0000				
4		1.0000	1.0000	1.0000			
5		0.6000	0.8000	0.8667	0.9000		
6		[illegible]	0.3000	0.5333	0.6500	0.7200	
7		0.0000	0.0000	0.2000	0.4000	0.5200	
8	1/2	1.0000	0.5000	0.3333	0.4000	0.5200	
9		1.0000	1.0000	[illegible]	[illegible]	[illegible]	
10		1.0000	1.0000	1.0000	0.7500	0.6000	
11		1.0000	1.0000	1.0000	1.0000	0.8000	0.7600
12		1.0000	1.0000	1.0000	1.0000	1.0000	0.7600
13		1.0000	1.0000	[illegible]	[illegible]	[illegible]	0.7600
14		0.8000	0.9000	0.9333	0.9500	0.9600	0.7400
15		0.0000	0.4000	0.6000	0.7000	0.7600	0.6800
16		0.0000	0.0000	0.2667	0.4500	0.5600	0.6800
17		0.0000	0.0000	0.0000	0.2000	0.3600	0.6800
18	1/3	1.0000	0.5000	0.3333	0.2500	0.3600	0.6800
19		1.0000	1.0000	0.6667	0.5000	0.4000	0.6800
20		1.0000	1.0000	1.0000	0.7500	0.6000	0.6800
21		1.0000	1.0000	1.0000	1.0000	0.8000	0.6800
22		0.3000	0.6500	0.7667	0.8250	0.8600	0.6100
23	2/1	1.0000	0.6500	0.7667	0.8250	0.8600	0.6100
24		1.0000	1.0000	0.7667	0.8250	0.8600	0.6300
25		1.0000	1.0000	1.0000	0.8250	0.8600	0.7300
26		0.6000	0.8000	0.8667	0.9000	0.7800	0.7900
27		0.0000	0.3000	0.5333	0.6500	0.7200	0.7900
28		0.0000	0.0000	0.2000	0.4000	0.5200	0.6900
29	2/3	1.0000	0.5000	0.3333	0.4000	0.5200	0.6900
30		1.0000	1.0000	0.6667	0.5000	0.5200	0.6900
31		1.0000	1.0000	1.0000	0.7500	0.6000	0.6900
32		1.0000	1.0000	1.0000	1.0000	0.8000	0.7600
33		0.3000	0.6500	0.7667	0.8250	0.8600	0.6900

=AVERAGE(B2:B3)
=AVERAGE(B3:B4)
=AVERAGE(B2:B11)
=AVERAGE(B5:B14)

Table 3

The time line of Figure 2 is placed in column B with the use of permutations to extend the line. Only the first permutation and part of the second are shown. The cell equations will allow the user set up their own spreadsheet. Columns B through K random variable values form distributions for accumulated availability for the time period shown in row 1.

simplicity of illustration. Every number in Table 3 columns B through K is a random variable value for that particular series accumulated availability distribution. The random variable values are grouped into a histogram and displayed as seen in Figure 3 for the fifth and tenth series. Figure 3 is accumulated availability in cumulative form. For example, if our production goal for the next 10 days is 65% of maximum capacity, i.e. without any scheduled downtime we need at least 65% availability, we can read from Figure 3 that there is 20% chance of not making the goal. Full risk assessment with probability and consequence is explained in detail in [3].

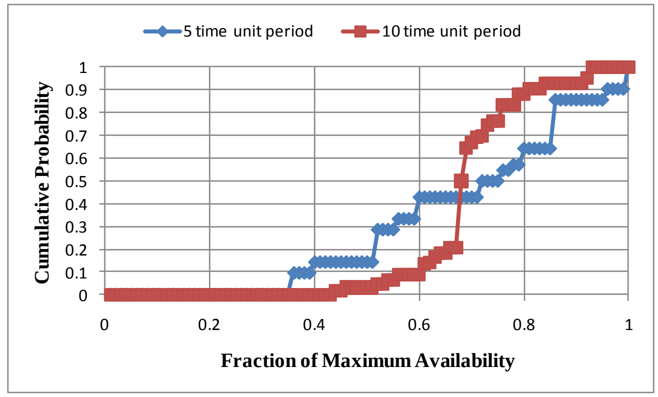


Figure 3

5 and 10 time unit accumulated availability distributions from histogram data of Table 3 columns F and K in cumulative form. Note useable distributions from only 3 failures.

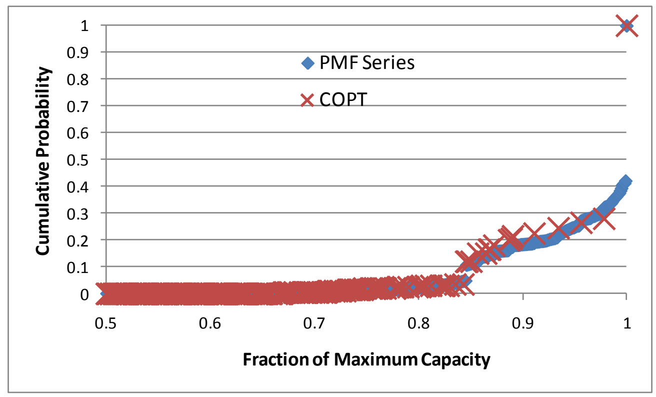


Figure 4

A 10 unit system with PMF series using TBF and TTR data and COPT using availability from the same data. Results are comparable although concepts are totally different.

The accumulated availability values in column C through K in Table 3 are not entirely accurate because the running average does not give full weight to the first few and last few numbers in the time line of column B. With adjustment for the weighing, the distributions from Table 3 are mathematically exact. In practice, the time line is so long that this is insignificant and does not require correction.

COPT and PMF series can be compared side by side in special circumstances. Though different concepts, they produce comparable results when the 1-day PMF series (such as in column B Table 3) and COPT unit capacities (such as in row 3 Table 1) are capacity units per day – they have common time units. A demonstrative 10 unit system with a range of unit capacities were created from TBF and TTR data from a number of power and petrochemical plants. Unit availabilities calculated from this data were used for COPT input data. PMF series used permutations of the data to form individual unit time lines that were combined to form a system time line. The methodology to combine unit time lines into a system time line is similar to that described for time-to-first- failure [2].

For this system, Figure 4 shows that the the first of the PMF series, such as from column B of Table 3, gives results that are comparable to COPT.

The prior simple illustration of Figure 2 and Table 3 with three failure events is smaller than normal. Realistic databases have more failure events, and with permutations the analyzed data quickly become astronomically large. For example, a petrochemical plant had 66 failures over a 1,924 day time line, with scheduled outages excluded. The total time line consists of 66! permutations and is over 1E96 values long. These large numbers may be incomprehensible without some sort of reference. If these values were printed with 500 values per 8x11 inch page and neatly stacked with 250 pages per inch height, the volume of paper would fill the universe 26 million times, given a spherical universe of 125 billion light-year diameter.

With such a large quantity of data, the ability to obtain a sufficient sample of the permutations may be questionable. However, distributions quickly converge with increasing sample size within the capability of personal computers. Figure 5 is the 30-day distribution of accumulated availability/capacity in cumulative form of the petrochemical plant example. Only when a small area of Figure 5 is enlarged, as in Figure 6, can differences in resolution (number of permutations sampled and length of the time line) be seen. The solid lines are formed from three different time lines of about one million rows in a spreadsheet. They are closer together than the dashed lines formed with three different time lines of about 65,000 rows. Figures 5 and 6 show that the distribution

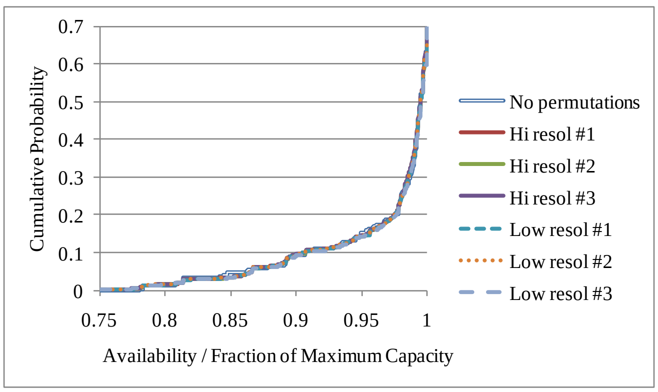


Figure 5

30-day accumulated availability for a petrochemical plant with 66 failures. Distributions calculated using different number of permutations (different resolution) show similar results. See Figure 6.

formed with no permutations approaches the others. With large quantities of events, such as the 66 petrochemical plant failures, reasonably good distributions can be formed without permutations and the necessity for the events to be independent.

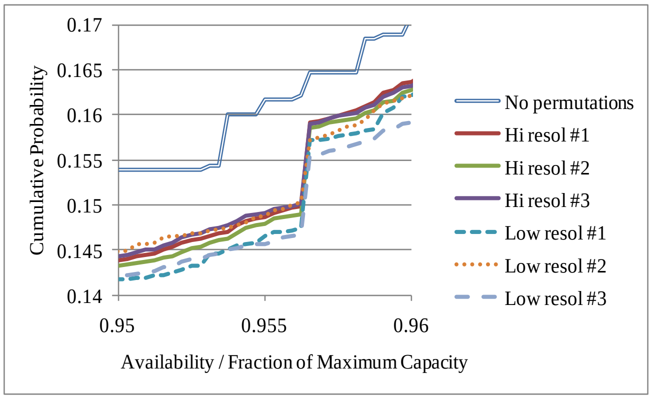


Figure 6

The distributions of Figure 5 show better convergence as the number of permutations is increased. For large quantities of raw data, no permutations (upper line) provides useable results

Some systems do not have distinguishable failure events, and use of permutations is not possible. Polypropylene plants are an example. Failure events are not easily defined, or impact on production easily allocated. Process problems overlap equipment problems such that all we can measure accurately is total unreliability impact for the day. Here the daily production quantities and the reliability related production losses are used to form the availability fraction for each day. Permutations are not allowed because the daily values are not independent. The running average errors such as seen in the first few and last few values of Table 3 are driven to insignificance by conceptually looping the time line. In practice, this is accomplished by extended the time line as previously discussed, but the order of events are maintained.

PMF series is a general approach applicable to a wide range of non-conventional applications. Two will be presented. A business process delivers a particular medical service. Sales data for 13 weeks were used to forecast the system capacity for the next quarter. The business system in the near future is expected to reflect the recent past. Sampling of the 13! permutations and developing the time line as in column B of Figure 3 produces the 3-month accumulated capacity distribution of Figure 7. The probability and risk associated with any management objective can be assessed. Management adjustments can be made if risk is unacceptable.

Another non-conventional application is for industrial safety. A large maintenance service company's recordable injuries over a 15 month period were analyzed. The numbers of injuries by week were normalized by dividing all values by the largest weekly value. This converts all values to a value between zero and one, convenient for PMF series. The normalization also serves to code the real data. For confidential or sensitive data, the analyst can work with only normalized data. The risk assessment is done with a maximum value of 1 (or any lower positive number) and real values are revealed only by a multiplier which can be confidential. Results are seen in Figures 8 and 9.

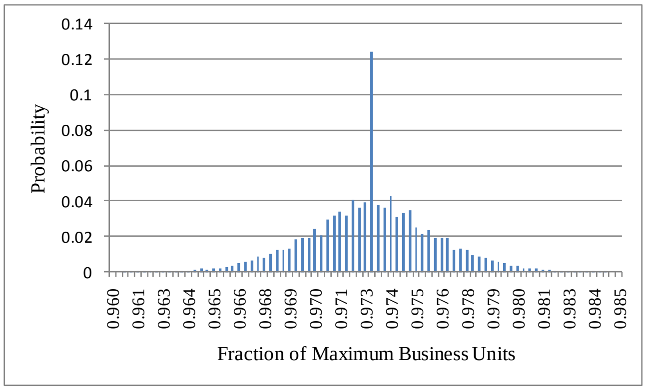


Figure 7

A medical services business system capacity for a 3-month period. Business commitments can be assessed for risk and management adjustments made if risk is unacceptable.

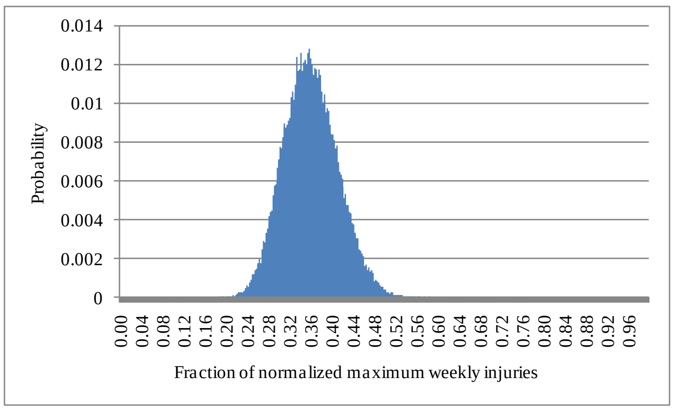


Figure 8

The 13-week PMF series maintenance injury distribution. Assessment reveals the risk associated with any performance-based contract commitment. High risk can be managed down by changing the "system" that generates injury.

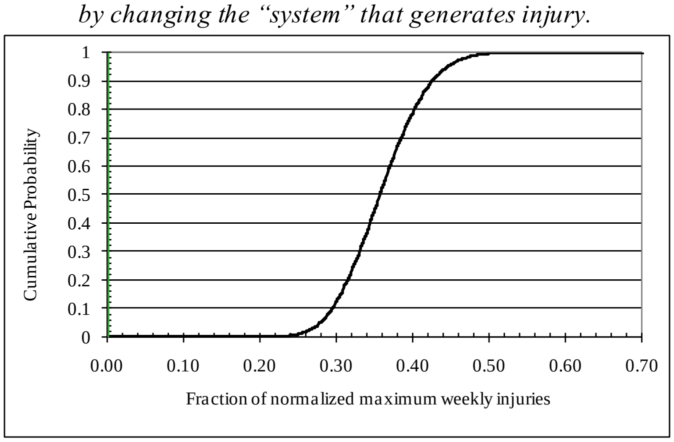


Figure 9

The injury distribution of Figure 8 in cumulative form. There is a 10% chance that the number of injuries will exceed 42.7% of the maximum weekly number that is confidential. Data such as this helps drive safety improvements.

REFERENCES

1. R Billington, R. N. Allan, *Reliability Evaluation of Power Systems, 2nd edition,* New York, Plenum Press, 1996.

2. J. B. Smith, "PMF Series for Availability and Reliability Probabilistic Assessments" *Proc. Ann. Reliability & Maintainability Symp.*, (Jan.) 2009.

3. J. B. Smith, W. B. Smith, "Probabilistic Assessment of Availability from System Performance Data," *Proc. Ann. Reliability & Maintainability Symp.*, (Jan.) 2002.

BIOGRAPHY

Jan B. Smith, P.E.
Reliability Engineering Consultant
2907 Forest Garden Dr.
Kingwood, TX 77345
E-mail: janb.smith@cebridge.net

Jan B. Smith, P.E., BSME, is a Registered Professional Engineer with a 40 year career devoted to process plant reliability. His experience includes forming and managing reliability organizations for major companies at the plant and corporate level as well as independent reliability consulting companies. His principle work includes reliability-centered maintenance, root cause failure analysis; finite element analysis; statistical and probabilistic analysis; vibration analysis; and capacity and availability forecasting. He has chaired conferences on plant reliability, authored several technical papers on the subject, and has organized and taught seminars on reliability issues. He holds three patents on business methods using new probabilistic methods.