Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

MODELS OF RANDOM MACHINE DOWNTIMES FOR SIMULATION

Averill M. Law

Simulation Modeling and Analysis Company
440 N. Alvernon Way, Suite 103
Tucson, Arizona 85711

ABSTRACT

A significant source of randomness for many simulation models of manufacturing systems is the unscheduled downtime of machines. However, there has been little discussion of this subject in the simulation or manufacturing literature. In this paper we discuss how to model machine downtimes both when system data are available and in the no-data case.

1. INTRODUCTION

The most important source of randomness for many manufacturing systems is that associated with machine breakdowns or unscheduled downtime. Random downtime results from such events as actual machine failures, part jams, and broken tools. The following example illustrates the importance of modeling machine downtimes correctly in simulation studies.

A company is going to buy a new machine tool from a vendor who claims that the machine will be down 10 percent of the time. However, the vendor has no data on how long the machine will operate before breaking down or on how long it will take to repair the machine. Some simulation analysts have accounted for random breakdowns by simply reducing the machine processing rate by 10 percent. We will see, however, that this can produce results that are quite inaccurate.

Suppose that the single-machine-tool system will actually operate in accordance with the following assumptions when installed by the purchasing company:

- Jobs arrive with exponential interarrival times with a mean of 1.25 minutes.
- Processing times for a job at the machine are a constant 1 minute.
- The machine operates for an exponential amount of time with mean 540 minutes (9 hours) before breaking down.
- The repair time for the machine has a gamma distribution (shape parameter equal to 2) with mean 60 minutes (1 hour).
- The machine is, thus, broken 10 percent of the time, since the mean length of the up-down cycle is 10 hours.

In column 1 of Table 1 are results from five independent simulation runs of length 160 hours (20 eight-hour days) for the above system; all times are in minutes. In column 2 of the table are results from five simulation runs of length 160 hours for the machine tool system with no breakdowns, but with the processing (cycle) rate reduced from 1 job per minute to 0.9 job per minute, as has sometimes been the approach of simulation practitioners.

Note first that the average weekly throughput is almost identical for the two simulations. (For a system with no capacity shortages that is simulated for a long period of time, the average throughput for a 40-hour week must be equal to the arrival rate for a 40-hour week, which is 1,920 here.) On the other hand, note that such measures of performance as average time in system for a job and maximum number of jobs in queue are vastly different for the two cases. Thus, the deterministic adjustment of the processing rate produces results that differ greatly from the correct results based on actual breakdowns of the machine.

Table 1. Simulation Results for Single-Machine-Tool System

Measure of performance	Breakdowns	No breakdowns
Average throughput per 40/hour week*	1908.8	1914.8
Average time in system*	35.1	5.6
Maximum time in system#	256.7	39.1
Average number in queue*	27.2	3.6
Maximum number in queue#	231.0	35.0

*Average over five runs. #Maximum over five runs.

Despite the importance of modeling machine breakdowns correctly, as demonstrated by the above example, there has been little discussion of this subject in the simulation or manufacturing literature. Thus, we now present an introduction to modeling random machine downtimes; see Law and Kelton [1] for a more detailed discussion. Deterministic downtimes such as breaks, shift changes, and scheduled maintenance are relatively easy to model and are not treated here.

2. MODELS FOR WHEN SYSTEM DATA ARE AVAILABLE

A machine goes through a sequence of cycles, with the ith cycle consisting of an up ("operating") segment of length U_i followed by a down segment of length D_i. During an up segment, a machine will process parts if any are available and if the machine is not blocked. The first two up-down cycles for a machine are shown in Figure 1. Let B_i and I_i be the amounts of time during U_i that the machine is busy processing parts and that the machine is idle (either starved for parts or blocked by the current finished part), respectively. Thus, $U_i = B_i + I_i$. Note that B_i and I_i may each correspond to a number of separated time segments and, thus, are not represented in Figure 1.

Let W_i be the waiting time from the ith "failure" of the machine until its subsequent repair begins, and let R_i be the length of this ith repair time. Thus, $D_i = W_i + R_i$, as shown in Figure 1.

We will assume for simplicity that cycles are independent of

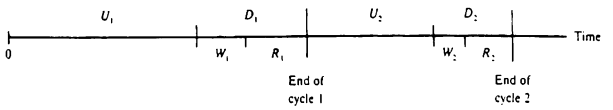

Figure 1. Up-Down Cycles for a Machine

each other and probabilistically identical, and also that U_i and D_i are independent for all i.

We now discuss how to model machine-up segments in a simulation model assuming that "appropriate" breakdown data are available. The following two methods are widely used:

Calendar time. Assume that the uptime data U_1, U_2, ... are available and that we can fit a standard probability distribution (e.g., exponential) F_U to these data using techniques discussed in Law and Kelton [1]. Alternatively, if no distribution provides a good fit, assume that an empirical distribution (see [1]) is used to model the U_i's. (Standard and empirical distributions are available in most simulation software.) Then, starting at time 0, we generate a random value u_1 from F_U and $0 + u_1 = u_1$ is the time of the first failure of the machine in the simulation. When the machine actually fails at time u_1, note that it may either be busy or idle. Suppose that d_1 is determined to be the first downtime (to be discussed below) for the machine. Then the machine goes back up at time $u_1 + d_1$. (If the machine was processing a part when it failed at time u_1, then it is often assumed that the machine finishes this part's remaining processing time starting at time $u_1 + d_1$.) At time $u_1 + d_1$ another value u_2 is randomly generated from F_U and the machine is up during the time interval $[u_1 + d_1, u_1 + d_1 + u_2)$, etc.

There are two potential drawbacks of the calendar-time approach. First, it allows the machine to break down when it is idle, which may not be realistic. Also, assume that the machine in question is part of a larger system and has machines both upstream and downstream from it. Then, for two different versions of the larger system, the machine could fail at the same points in time, but have significantly different amounts of actual busy time.

Busy time. Assume that the busy-time data B_1, B_2, ... are available and that we can fit a distribution F_B to these data. (Alternatively, an empirical distribution can be used.) Then, starting at time 0, we generate a random value b_1 from F_B. The machine is up until its total accumulated busy (processing) time reaches a value of b_1, at which point the busy machine fails. (For example, suppose that b_1 is equal to 60.7 minutes and each processing time is a constant 1 minute. Then the machine fails while processing its 61st part.) If f_1 is the simulated time when the machine fails for the first time ($f_1 \geq b_1$) and d_1 is the first downtime, then the machine goes back up at time $f_1 + d_1$, etc.

In general, the busy-time approach is more natural than the calendar-time approach. We would expect the next time of failure of a machine to depend more on total busy time since the last repair than on calendar time since the last repair. However, in practice, the busy-time approach may not be feasible, since uptime data (U_1, U_2, ...) may be available but not busy-time data (B_1, B_2, ...). In many factories, only the times that the machine fails and the times that the machine goes back up (completes repair) are recorded. Thus, the uptimes U_1, U_2, ... may be easily computed, but the actual busy times B_1, B_2, ... may be unknown. (In computing the U_i's, time intervals where the machine is off, e.g., idle shifts, should probably be subtracted out.)

We now discuss how to model machine-down segments assuming that factory data are available. Assume first that the waiting time to repair, W_i, for the ith cycle is zero or negligible relative to the repair time R_i (for i = 1, 2, ...). Then we fit a distribution (e.g., gamma) F_D to the observed downtime data D_1, D_2, ... Each time the machine fails, we generate a new random value from F_D and use it as the subsequent downtime (repair time).

Suppose that the W_i's may sometimes be "large," due to waiting for a repairman to arrive. If only D_i's are available (and not the W_i's and R_i's separately) as is often the case in practice, then fit a distribution F_D to the D_i's and randomly sample from F_D each time a downtime is needed in the simulation model. The reader should be aware, however, that F_D is a valid downtime distribution for only the current number of repairmen and the maintenance requirements of the system from which the D_i's were collected.

Finally, assume that the W_i's may be significant and that the W_i's and R_i's are individually available. Then one approach is to model the waiting time for a repairman as a maintenance resource with a finite number of units and to fit a distribution F_R to the R_i's. If a repairman is available when the machine fails, the waiting time is zero unless there is a travel time, and the repair time is generated from F_R. If a repairman is not available, the broken machine joins a queue of machines waiting for a repairman, etc.

Several simulation software packages have an option that allow a user to specify the distribution of $C_i = U_i + D_i$ (the total length of an up-down cycle or the time between failures) and the distribution of D_i. We do not believe that this is a good approach, in general, since it theoretically allows the generated value of D_i to be larger than the generated value of C_i, which should be impossible. Also, this approach makes D_i and U_i negatively correlated (i.e., D_i large makes U_i small, and vice versa), since $U_i = C_i - D_i$.

3. A MODEL FOR THE NO-DATA CASE

Suppose now that factory data are not available to support either the calendar-time or busy-time breakdown models previously discussed. This often occurs when simulating a proposed manufacturing facility, but may also be the case for an existing plant when there is inadequate time for data collection and analysis. We now present a tentative model for this no-data case, which is likely to be more accurate than many of the approaches used in practice (see the above example).

We will first assume that the amount of machine busy time, B, before a failure has a gamma distribution with shape parameter $\alpha_B = 0.7$ and scale parameter β_B to be specified. Note that the exponential distribution (gamma distribution with $\alpha_B = 1.0$) does not appear, in general, to be a good model for machine busy times (see Law and Kelton [1]), even though it is often used in simulation models for this purpose.

We chose the gamma distribution because of its flexibility (i.e., its density can assume a wide variety of shapes) and because it has the general shape of many busy-time histograms when $\alpha_B \leq 1$. (The Weibull distribution could also have been used, but its mean is harder to compute.) The particular shape parameter $\alpha_B = 0.7$ for the gamma distribution was determined by fitting a gamma distribution to four different sets of busy-time data, with 0.7 being the average shape parameter obtained. In none of the four cases was the estimated shape parameter close to 1.0 (the exponential distribution). The density function for a gamma distribution with shape and scale parameters 0.7 and 1.0, respectively, is shown in Figure 2(a).

We will assume that machine downtime (or repair time) has a gamma distribution with shape parameter $\alpha_D = 1.4$ and scale

parameter β_D to be specified. This particular shape parameter was determined by fitting a gamma distribution to six different sets of downtime data, with 1.4 being the average shape parameter obtained. The density function for a gamma distribution with shape and scale parameters 1.4 and 1.0, respectively, is shown in Figure 2(b). This density function has the same general shape as downtime histograms typically experienced in practice.

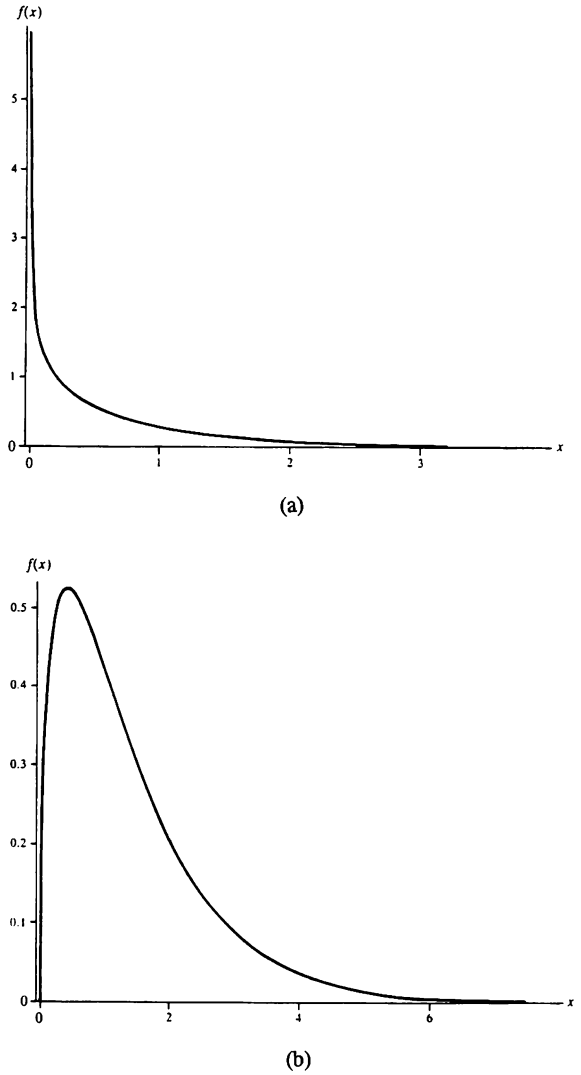

Figure 2. (a) Gamma(0.7,1.0) Distribution
(b) Gamma(1.4,1.0) Distribution

In order to complete our model of machine downtimes in the absence of data, we need to specify the scale parameters β_B and β_D. This can be done by soliciting two pieces of information from system "experts" (e.g., engineers or vendors). We have found it convenient and typically feasible to obtain an estimate of mean downtime μ_D = E(D) and an estimate of machine efficiency e, which we now define. The efficiency e is defined to be the long-run proportion of potential processing time (i.e., parts present and machine not blocked) that the machine is actually processing parts, and is given by

$$e = \mu_B/(\mu_B + \mu_D)$$

where μ_B = E(B) is the mean amount of machine busy time before a failure. If the machine is never starved or blocked, then $\mu_B = \mu_U$ = E(U) and e is the long-run proportion of time during which the machine is processing parts. Using the values of μ_D and e (and also the fact that the mean of a gamma distribution is the product of its shape and scale parameters), it is easy to show that the required scale parameters are given by

$$\beta_B = e\ \mu_D/0.7(1 - e)$$

and

$$\beta_D = \mu_D/1.4$$

Thus, our model for machine downtimes when no data are available has been completely specified.

4. SUMMARY

We have discussed above basic models for the breaking down and repair of machines. However, in practice there are a number of additional complications that often occur, such as multiple independent causes of machine failure.

REFERENCE

Law, A.M. and W.D. Kelton (1990), *Simulation Modeling and Analysis*, Second Edition, McGraw-Hill, New York, NY.