LA-UR- 98-2678

Title: AVAILABILITY GROWTH MODELING

CONF-980853--

Author(s): Joanne R. Wendelberger, TSA-1

Submitted to: American Statistical Association Meetings
Dallas, Texas
August 9-13, 1998

Los Alamos
NATIONAL LABORATORY

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

DRAFT

AVAILABILITY GROWTH MODELING

Joanne Wendelberger

Statistics Group, MS F600, Los Alamos National Laboratory, Los Alamos, NM 87544

May 27, 1998

KEY WORDS: cumulative availability, instantaneous availability, logistic model, empirical Bayes

ABSTRACT

In reliability modeling, the term availability is used to represent the fraction of time that a process is operating successfully. Several different definitions have been proposed for different types of availability. One commonly used measure of availability is cumulative availability, which is defined as the ratio of the amount of time that a system is up and running to the total elapsed time. During the startup phase of a process, cumulative availability may be treated as a growth process. A procedure for modeling cumulative availability as a function of time is proposed. Estimates of other measures of availability are derived from the estimated cumulative availability function. The use of empirical Bayes techniques to improve the resulting estimates is also discussed.

1. INTRODUCTION

The purpose of this paper is to describe procedures for estimating different types of availability for a system based on the history of uptimes and downtimes. Availability is often modeled using cumulative availability, the ratio of the amount of time that a system is up and running to the total elapsed time. First, we will discuss the estimation of cumulative availability as a function of time. We will then show how to derive estimates of other measures of availability from the estimated cumulative availability function.

The history of a system may be characterized by the periods of uptime when the system is operating successfully and the periods of downtime when the system is halted for repairs or other maintenance. The total system time at time t is divided up into irregularly spaced intervals corresponding to the periods of uptime and downtime. Let u_i denote the length of the ith interval of uptime, $i = 1, \ldots, m(t)$, where $m(t)$ denotes the number of intervals of uptime which occur prior to time t. Let d_j denote the length of the jth interval of downtime, $j = 1, \ldots, n(t)$, where $n(t)$ denotes the number of intervals of uptime which occur prior to time t. Then cumulative availability, $A(t)$, at time t, may be estimated by the sum of the uptimes divided by the sum of the uptimes plus the sum of the downtimes, for all uptimes and downtimes which occur prior to time t.

$$\widehat{A(t)} = \frac{\sum_{i=1}^{m(t)} u_i}{\sum_{i=1}^{m(t)} u_i + \sum_{j=1}^{n(t)} d_j}.$$

The numerator represents the time the system is operating successfully. The denominator is the total time, which includes both the uptimes and the downtimes.

Cumulative availability at time t includes all data from a given startup time up until time t. The use of cumulative availability to monitor process performance has the drawback that early data which may no longer represent current process behavior is included in estimates at subsequent time points. For example, a new process typically improves over time as the process is adjusted and fine-tuned. For this reason, other measures of availability besides cumulative availability are useful. In particular, a measure of availability which captures how well a process is doing at a specific time is desired.

Lie, Hwang, and Tillman (1977) describe several different classes and types of availability which have been discussed in the literature and provide numerous references. One classification, based on the time period considered, includes instantaneous, interval, and long-run availability. Interval availability refers to availability over a given time period. Cumulative availability, which measures availability from the starting time to a specified time t, is an example of interval availability. Long-run availability is the limiting value of the availability at arbitrarily long times. For example, if a processes stabilizes at a given level, the long-run availability measures the ultimate availability at which the process is capable of performing after it reaches a steady state. Instantaneous availability, $A_I(t)$, is the probability that a system is up at time t. Instantaneous availability measures the availability within an arbitrarily small interval, i.e, at a specific time, t. Instantaneous availability may be thought of as the limit of an interval availability, as the length of the interval gets smaller and smaller.

Following Lie et al, The relationships between these different types of availability may be described mathematically. Cumulative availability, $A(T)$, is the average of instantaneous availability from startup time to time T.

$$A(T) = \frac{1}{T} \int_0^T A_I(t)dt.$$

Long-run availability, A_{LR} is the limit of the cumulative availability as T grows large.

$$A_{LR} = \lim_{T \to \infty} A(T).$$

Lie et al. discuss several different approaches for modeling availability. The two main classes of models are Markovian models and models which use the ratio of uptime to total time. The Markovian approach typically involves the assumption of exponential distributions for failure times and repair times. The use of the Markovian approach leads to a definition of instantaneous availability based on the underlying distributional assumptions.

Models which use the ratio of uptime to total time approach availability more directly than the Markovian approaches. Typically, this approach involves the use of cumulative availability. An alternative, but closely related, approach expresses availability as a function of the mean time between failures and the mean time to repair.

$$A = \frac{MTBF}{MTBF + MTTR},$$

where MTBF is the mean time between failure, and MTTR is the mean time to repair.

Our approach is to use statistical estimation techniques to fit a cumulative availability model as a function of time based on uptime and downtime data. The fitted model is then used to estimate cumulative availability, interval availability, and long-run availability. Expressions for interval and instantaneous availability are derived from the cumulative availability function to allow estimation of availability at any point within the range of data, as well as extrapolation beyond the original data range, assuming the model is adequate within the region of interest.

Further improvements to the different availability estimates may be obtained by implementing an Empirical Bayes estimation procedure. The Empirical Bayes procedure can yield a substantial reduction in the variability of the estimated availabilities. The Empirical Bayes procedure provides a mechanism to incorporate historical information about similar systems into the availability estimates.

2. AVAILABILITY CONCEPTS AND RELATIONSHIPS

In this section, several different types of availability are defined, and relationships between different types of availability are developed. In the next section, procedures for estimating availability based on these concepts are proposed.

Let $u(t)$ represent an underlying uptime function which denotes the total amount of uptime which has occurred during the time period $[0, t]$. For simplicity, suppose that this true underlying function is continuously differentiable. In reality, the uptime function will be a step function. However, for sufficiently long time periods, the continuous approximation will be adequate.

2.1 Cumulative Availability

Cumulative availability, the fraction of time the system has been up at time t, is given by

$$A(t) = u(t)/t,$$

for $t > 0$.

2.2 Interval Availability

Cumulative availability is a special case of interval availability. Interval availability is the fraction of time the system is up during a specified interval $[t1, t2]$, where $t_1 < t2$. Let $A_{[t_1,t_2]}(t)$ denote the availability during the period of time from t_1 to t_2.

$$A_{[t_1,t_2]}(t) = \frac{u(t_2) - u(t_1)}{t_2 - t_1}.$$

Using this notation for interval availability, cumulative availability at time T is given by $A_{[0,T]}$.

Availability over a moving time window of width k, $A_k(t)$, may be obtained by specifying interval availability with starting point, $t - k$, and endpoint t, i.e.,

$$\begin{aligned} A_k(t) &= A_{[t-k,t]}(t) \\ &= \frac{u(t) - u(t-k)}{t - (t-k)} \\ &= \frac{u(t) - u(t-k)}{k}. \end{aligned}$$

Note that if $k = t$, then

$$
\begin{aligned}
A_k(t) &= A_t(t) \\
&= \frac{u(t) - u(t-t)}{t - (t-t)} \\
&= u(t)/t.
\end{aligned}
$$

At any time t, the moving window is calculated by dividing the change in uptime during the past time period of length k by the length of the time period, k.

2.3 Instantaneous Availability

The definition for moving window availability is now used to derive an expression for instantaneous availability. Instantaneous availability, $A_I(t)$, is defined as the limit of the moving window availability as the width of the window, k, goes to zero.

$$
\begin{aligned}
A_I(t) &= \lim_{k \to 0} A_k(t) \\
&= \lim_{k \to 0} \frac{u(t) - u(t-k)}{k} \\
&= u'(t),
\end{aligned}
$$

where $u'(t)$ is the derivative of the function $u(t)$ with respective to time. The derivative, $u'(t)$, is the rate of change of the uptime function $u(t)$. We could approximate the uptime function, $u(t)$, by an estimated uptime function, $\widehat{u(t)}$. Then the derivative function, $u'(t)$, could be evaluated either analytically or numerically. Instead, we note that $A(t) = u(t)/t \Longrightarrow u(t) = t \cdot A(t)$, for $t \neq 0$. It follows that we can get an expression for $A_I(t)$ by applying the rule for differentiation of a product to get

$$A_I(t) = u'(t) = t \cdot A'(t) + A(t),$$

where $A'(t)$ is the derivative of $A(t)$ with respect to t. The difference between instantaneous availability and cumulative availability is given by

$$A_I(t) - A(t) = t \cdot A'(t).$$

This difference is the amount by which the current availability is underestimated when cumulative availability is used instead of instantaneous availability. This difference may be explained by the presence of periods of lower availability early in the system's lifetime when a system's availability is increasing over time.

2.4 Long-Run Availability

Once a system reaches a steady state, the availability levels off. In this case, the difference between the cumulative availability and the instantaneous availability decreases to zero, and both the instantaneous and cumulative availabilities will converge to a long-run availability, A_{LR}.

$$\lim_{t\to\infty} A_I(t) = \lim_{t\to\infty} u'(t) = \lim_{t\to\infty} t \cdot A'(t) + A(t) = \lim_{t\to\infty} (t \cdot 0 + A(t)) = A_{LR},$$

provided $A'(t)$ goes to zero faster than t goes to infinity.

3. AVAILABILITY MODELING

Many papers in the existing literature on availability assume that processes arise from specified probability distributions. For a given distribution, the availability may be predicted based on the underlying distributional parameters. An alternative approach is to estimate the availability based on the observed data on uptimes and downtimes. This second approach allows for changes in the underlying process and provides a more direct measure of the resulting availability.

Modeling the availability based on observed uptimes and downtimes allows for changes in the underlying system behavior over time. For example, reliability growth models have been proposed to describe the increasing reliability of a system as it is adjusted and improved over time. Tobias and Trindade (1995) discuss reliability growth models in their recent book on applied reliability.

Growth curve models, widely used in diverse statistical applications, are useful candidates for empirical availability models. (See, for example Draper and Smith (1981).) In growth curve modeling, an analytical function f is used to model the relationship between availability and time.

$$A(t) = f(\theta, t) + \epsilon,$$

where θ represents a vector of growth curve parameters, and ϵ represents an additive error term.

The parameters of the function θ may be estimated by least squares. The resulting estimated cumulative availability is given by

$$A(t) = f(\hat{\theta}, t).$$

Once an adequate fit is obtained, the formulas from Section 4 may be applied to the estimated cumulative availability function to estimate interval availability, instantaneous availability, and long-run availability.

4. EXAMPLE

As an example, consider a logistic growth curve function model for cumulative availability. The availability model using the logistic function is given by

$$A(t) = \frac{\theta_1}{1 + \theta_2 e^{-\theta_3 t}}.$$

The derivative of the logistic cumulative availability model is given by

$$A'(t) = \frac{\theta_1 \theta_2 \theta_3 e^{-\theta_3 t}}{(1 + \theta_2 e^{-\theta_3 t})^2}.$$

The instantaneous availability is then computed using the forumula from Section 3.3,

$$A_I(t) = u'(t) = t \cdot A'(t) + A(t)$$

5. EMPIRICAL BAYES ESTIMATION

Empirical Bayes techniques may be used to improve the precision of the estimated availability quantities. The classical results obtained from fitting logistic models to individual systems can be improved upon by incorporating prior information from similar systems.

6. SUMMARY

The approach described here models availability as a function of time using data on uptime and downtime history. The concept of instantaneous availability, which has been used in the literature on Markov models, is applied to the estimated availability function. This procedure is motivated by experience with data which exhibits growth in availability. Instead of deriving instantaneous availability from an underlying probability structure which remains stable over time, a more direct approach which models availability from uptime and downtime data is suggested which allows for changes in the underlying process.

REFERENCES

1. Barlow, R. E. and Proschan, F. (1981), *Statistical Theory of Reliability and Life Testing*, reprint of 1975 edition with corrections, To Begin With, Silver Spring, MD.
2. Draper, N. R. and Smith, H. (1981), *Applied Regression Analysis*, 2nd ed., John Wiley & Sons, Inc., 505-513.

3. Duane, J. T. (1964), "Learning Curve Approach to Reliability Monitoring," *IEEE Transactions on Aerospace*, **2**, 563-566.

4. Lie, C. H., Hwang, C. L., and Tillman, F. A. (1977), "Availability of Maintained Systems: A State-of-the-Art Survey," AIIE Transactions, **9**, 247-259.

5. Nicola, V., Bobbio, A. and Trivedi, K. (1992), "A Unified Performance Reliability Analysis of a System With a Cumulative Down Time Constraint," Microelectronics and Reliability, **32**, 49-65.

6. Tillman, F. A., Kuo, W., Nassar, R. F., and Hwang, C. L. (1983), "Numerical Evaluation of Instantaneous Availability," *IEEE Transactions*, **R-32**, 1, 119-123.

7. Tobias, P. A. and Trindade, D. C. (1995), *Applied Reliability*, 2nd ed., Van Nostrand Reinhold, NY, 372-385.