REPORT:

Authors:
***Sigmund J. Amster** is a member of technical staff in the Quality Theory and Technology Department and **Jeffrey H. Hooper** is a supervisor in the Quality Theory and Technology Department at AT&T Bell Laboratories in Holmdel, New Jersey. Mr. Amster joined AT&T in 1962. He has a B.S. in business administration from the University of Pennsylvania, an M.S. in managerial statistics from Columbia University, and a Ph.D. in mathematical statistics from the University of North Carolina. Mr. Hooper joined AT&T in 1977. He has a B.S. in engineering, an M.Engr. in industrial engineering, and an M.S. and a Ph.D. in operations research, all from Cornell University.*

STATISTICAL METHODS FOR RELIABILITY IMPROVEMENT

Introduction

Statistical methods play an important role in the design and manufacture of reliable products. They are used to design, analyze, and present reliability improvement studies such as accelerated testing of critical parts, prototype testing, process design and qualification, first-office applications, and field-tracking studies.

This article defines statistical reliability methods, explains how a product-reliability program uses them, and presents an example of their use. It also explores available software tools, training materials, and future directions.

Product Reliability

Reliability does not just happen in a product. It must be planned for and built into the product at each stage of the product-realization process.

A product-reliability program helps to ensure that the delivered product meets or exceeds its reliability objectives, and does so economically. Such a program comprises a coordinated set of activities (Figure 1) that begin at the early stages of system design and continue through design and development, implementation, manufacturing, and sales and support. Many of these activities involve statistical methods for designing reliability studies and analyzing and presenting reliability data.

During system design, for example, we determine the system architecture that can best meet the overall product-reliability objectives, and then allocate these objectives down to the subsystem and circuit-pack level. The allocation process, called reliability budgeting, provides designers and developers with meaningful reliability goals.

As the next step in the reliability program, we identify those parts or subsystems that are critical to the system's reliability. (Small changes in their reliability cause large changes in system reliability.) Additional engineering effort or testing will be devoted to these parts to ensure that they meet their reliability objectives. However, because no hardware exists that can be tested, statistical reliability methods are generally not useful at this stage.

During the design and development phase, early hardware begins to be available and statistical reliability methods become useful. We do accelerated life tests on critical parts at this early stage to determine if these parts will meet their reliability objectives. When designing these tests, we determine what stress levels to use and how many parts to test at each stress level.

Accelerated life tests are also used to evaluate competing designs and determine how to modify a given design to improve its reliability. At this stage, we may test completed prototypes, too, to find out if the subsystem or system can meet its reliability objectives.

During the implementation phase, reliability data are often collected to evaluate, improve, and qualify process designs. This is particularly true in component manufacturing. We also collect reliability data to qualify vendor products as meeting their reliability requirements.

In the manufacturing phase, we must determine whether the product performs as intended in the field, and then ensure that it

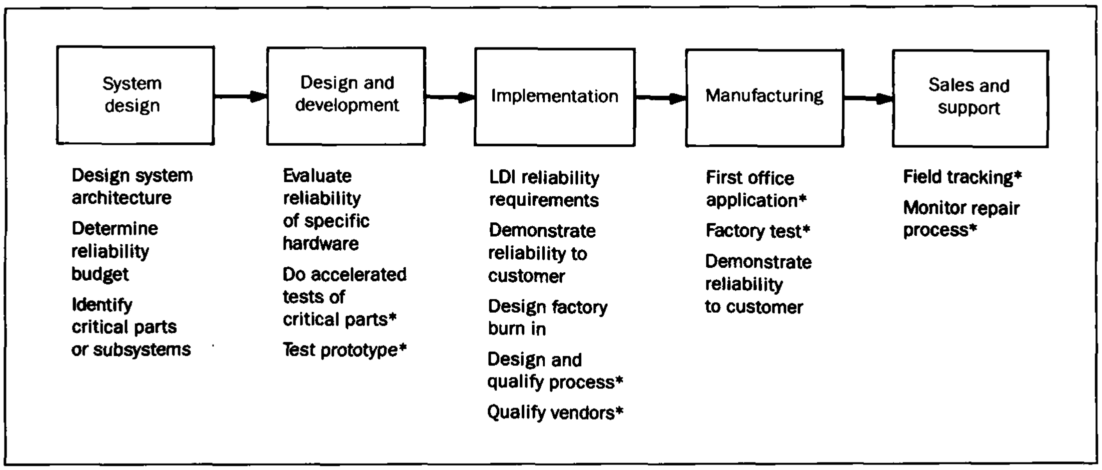


Figure 1. Product reliability program activities. Starred (*) items involve statistical reliability methods: the design of reliability studies, and the analysis and presentation of reliability data.

continues to do so. Therefore, we collect and analyze reliability data from a first-office application and from factory testing. We also analyze factory test data to prove product reliability to customers.

Reliable performance in the field is essential for important new-technology products. If we monitor this performance with a field-tracking study, any departure from the reliability objectives will trigger reliability-improvement activity.

While repair data can be messy, routinely extracting reliability information from these data can be important. For example, we can use information on the current replacement rate for circuit packs to provide customers with up-to-date information on product reliability, and determine current needs for spare circuit packs.

In addition, component-reliability information can sometimes be pulled out of repair data. We can feed this information into the reliability-information database to improve the reliability prediction for the next generation of the product.

Reliability Models

Many parametric reliability models have proven useful for analyzing reliability data. The models most frequently used are listed in Table I. All are special cases of the Generalized Gamma model, a fact we can use to determine which model is most appropriate for a given set of data.

It is possible to give a physical reason for using most of these models,[1] but the real justification is empirical. Because these models are flexible, we have found that they fit a wide variety of reliability data.

When several models fit a given set of data equally well, we can base our choice of model on conservative extrapolation, previous use in the same field, or other considerations. For example, the Weibull model eventually will predict a shorter probable failure time than the lognormal model.

The failure-rate function, $h(t)$, is frequently used to summarize a set of time-to-failure data. It describes how the instantaneous probability of failure changes with time. Because failure rate is a population quantity, a decreasing $h(t)$ may not imply that individual units are improving with age. The bathtub curve (Figure 2), or parts of it, characterizes many data sets.

When there is insufficient test time available to observe failures under use conditions, accelerated life testing is often employed. By increasing temperature, voltage, or other stress factors, it is possible to observe failures sooner. To extrapolate the data to use conditions, we have found two broad classes of models are effective: proportional-hazard (PH) models for T (time-to-failure) and accelerated-failure-time (AFT) models.[2,3]

The PH model is usually presented as

$$h(t/x) = h_0(t)\, e^{\sum_1^k \beta_i x_i}$$

Table I. Most Frequently Used Parameter Models

Model	Probability (Failure Time $>t$)	Failure Rate
Exponential	$e^{-\lambda t}$	Constant
Gamma	$\frac{1}{\Gamma(k)}\int_t^{\infty} u^{k-1}e^{-u}\,du$	Decreasing, or increasing and then decreasing
Lognormal	$\frac{1}{\sqrt{2\pi}}\int_{\left(\frac{\log t-\mu}{\sigma}\right)}^{\infty} e^{-(u^2/2)}\,du$	Increasing and then decreasing
Weibull	$e^{-(\lambda t)^{\beta}}$	Decreasing or constant or increasing

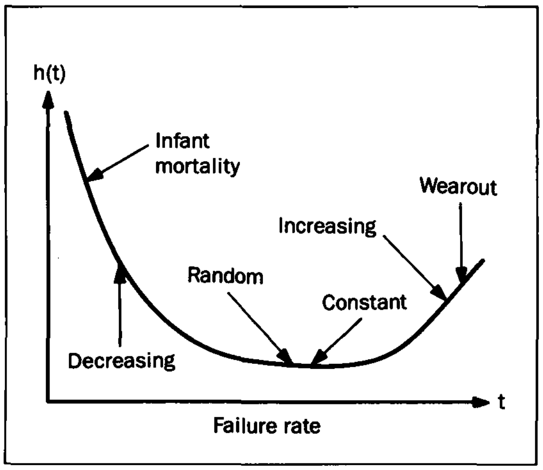


Figure 2. Plots of time-to-failure data usually produce the bathtub curve.

where $h_0(t)$ is the baseline hazard or failure-rate function (when all $x_i = 0$); the β_i are unknown parameters; and $x = (x_1, x_2, \ldots, x_k)$ is a vector of k explanatory variables.

We can estimate β_i either from a parametric model for $h_0(t)$ or in a model-free (nonparametric) fashion. For a given change in the value of x_i, a larger β_i means a larger change in $h(t/x)$, the failure-rate function.

For example, with $k = 1$, the ratio

$$\frac{h(t/x_{11})}{h(t/x_{12})} = e^{\beta_1(x_{11} - x_{12})}$$

is independent of t. In general, for the PH model, the failure rates for two different values of an explanatory variable are in constant proportion, independent of time.

The AFT model is usually presented as:

$$\log T = \mu(s) + \sigma e$$

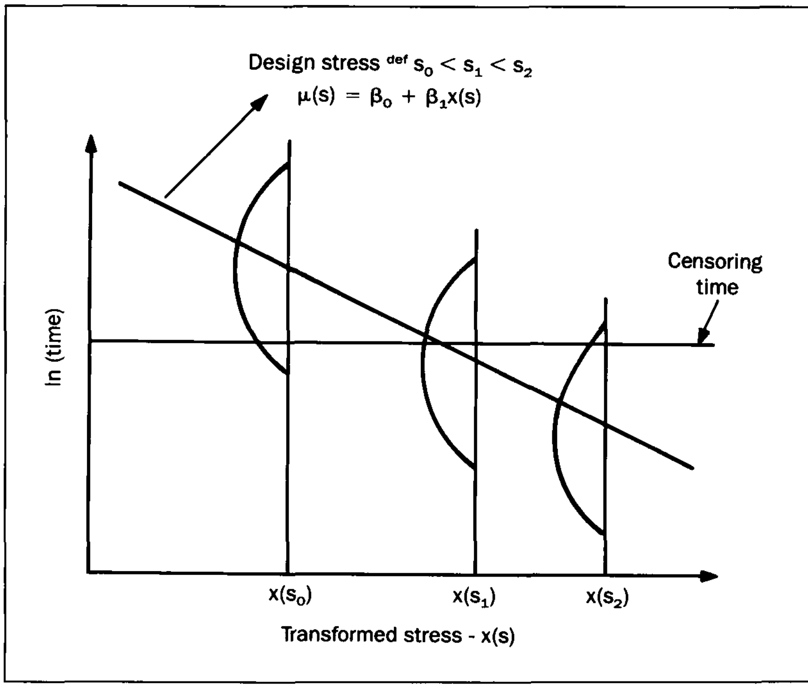


Figure 3. Accelerated-failure-time model. Assume that $\mu(s) = \beta_0 + \beta_1$ for $x(s)$, an increasing function of stress. Censoring time is the length of the experiment.

where T is time to failure, s is stress, $\mu(s)$ is the location parameter for the log T (log-failure-time) distribution, and σ is the distribution's scale parameter. That σ is independent of s is a consequence of the assumption that stress just speeds up time. We assume, as usual, that e follows a standard Gaussian (normal) model. When we can find a transformation of stress, $x(s)$, such that $\mu(s)$ is a linear function of stress, then the AFT model has the following highly intuitive interpretation.

As Figure 3 shows, if we increase the stress, the distribution moves down the straight line. Only in this case does the term *acceleration factor* mean simply a multiplier of time. An example of this is to use the Arrhenius relation for temperature stress, where β_1 is just a known constant times the activation energy.

Some early failures may already have been dead at the start of the life test. A useful model for dead-on-arrival failures is

$$P[T < t] = p + (1 - p)F(t)$$

where p is the probability of a dead-on-arrival and $F(t)$ is a parametric model (exponential, lognormal, etc.). We can use the data to estimate p, as well as the parameters of $F(t)$.

Reliability Analysis

Because reliability data are often either censored or truncated, many standard data-analysis techniques cannot be used. *Censoring* means that we know exact failure times for few, if any, units. For all other units, we know only that the failure time falls somewhere in an interval. *Truncation* means that we do not know how many units started at time zero, because only those that survived the first t_0 hours could be observed.

These special characteristics of reliability data make conventional statistical-analysis techniques inappropriate. For example, suppose we know some failure times, but we only know that three units live longer than 1000 hours. What is the average of the data?

A strategy that has proven useful in analyzing reliability data proceeds in three basic stages:

1. Nonparametric or model-free analysis to look at the data without restrictive model assumptions.
2. Graphical methods to choose the best parametric model.
3. Fitting the best parametric model and then using this fit to answer important reliability questions.

Without making any model assumptions, we can estimate cumulative-failure probability as a function of time. For example, the Kaplan-Meier estimate[2] can be used for data sets with only simple censoring, but the Turnbull estimate[4] is required when truncation or complex censoring is present.

We use the nonparametric estimate of cumulative-failure-probability to determine which parametric model, if any, best fits these data, without determining the model's specific parameters. Then, we take the best model and estimate the model parameters.

This gives us a single time-to-failure distribution that best fits the data. Usually we use the method of maximum likelihood. Then, we can use the resulting model to interpolate (smooth) the cumulative-failure-probability function over the range of the data and extrapolate this function beyond this point.

Future Directions

The latest methods for analyzing and presenting reliability data require both significant computation power and graphics capability.

Until recently, these methods were inaccessible to most engineers and technicians who need to analyze reliability data. Because they often used paper and pencils for analyzing studies, additional reliability information was not pulled out of the data.

Now, the latest reliability methods are readily available through STAR, a UNIX® system software tool for the analysis and presentation of reliability data. AT&T engineers and technicians can use STAR throughout a product-reliability program. Panel 1 describes one application.

STAR is easy to use and guides a user, from data entry, through data analysis, to the creation of presentation-quality tables and plots. Thus, it can quickly turn reliability data into useful information.

In a three-day statistical reliability workshop, we teach participants to use STAR effectively to improve product reliability, and help them understand the most important reliability models and their underlying assumptions. The workshop emphasizes how to use the STAR software package to fit these models to reliability data and assess the goodness of this fit.

In workshop exercises, participants gain hands-on experience with STAR to analyze reliability data. In particular, they are encouraged to analyze their own data.

Soon, stand-alone tools like STAR will be integrated with other reliability tools like SUPER to form a workbench that will provide comprehensive support for a product reliability program. (SUPER stands for: the system used for prediction and evaluation of reliability.)

Another important step is to integrate these reliability tools into computer-aided engineering, design, and manufacturing (CAE/CAD/CAM) systems. This will plug important reliability improvement methods directly into the product-realization process.

The next stage of integration will be to integrate reliability tools into factory and industrial automation systems. Thus, these tools will become part of the normal way of doing business.

The next generation of statistical reliability technology will focus on ways to build high reliability into products in a cost-effective way. There currently appear to be at least two fruitful directions for this technology.

The first involves a closer coupling between the physics of failure and reliability models, by developing a set of models based on the physics of failure. In this way, knowledge about an appropriate reliability model would provide insight into the physics of failure and vice versa.

The other fruitful direction for statistical reliability technology concerns measured-degradation data. With the advent of microprocessor-controlled life tests, it is now

High-K ceramic capacitors were to be used in a new system under a 25V operating voltage stress. To determine warranty expenses, the probability of failure in the first year of operation, 8760 hours, was needed. Because only a small proportion of capacitors were expected to fail within the first year, an extensive accelerated life test was conducted.

During the test, groups of capacitors were subjected to a 50V, 100V or 150V accelerated voltage stress. First, the individual voltage-group data was examined to determine if the accelerated-failure-time model and the parametric time-to-failure model were adequate. Then, all the data were used to fit the AFT model.

As the first step in analyzing the data, a nonparametric estimate was obtained of the cumulative-failure probability as a function of time for each group. Then, probability plots were obtained for the parametric models that were considered candidates.

Because any model that fits the data well will plot as a nearly straight line, the plots were examined to determine which one was closest to linear. Simultaneous confidence bands furnish a quantitative measure of *goodness of fit* and usually provide evidence to eliminate models that fit poorly.

For the capacitor data, the lognormal distribution was selected and individual lognormal distributions were fit to the data at each voltage. The fitted distributions and nonparametric results, Figure a, give a visual measure of the equality of the slopes, which is an important assumption of the AFT model. (σ is the same at each voltage.)

Because voltage was the stress, the inverse-power-stress function was used in the AFT model. It was assumed that

$$e^{\mu(s)} = e^{\beta_0} s^{\beta_1}$$

or

$$\mu(s) = \beta_0 + \beta_1 \log s$$

and log $(T(s))$ has a lognormal distribution with parameters $\mu(s)$, σ.

Then, maximum likelihood was used to fit all the data to this model and obtain estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}$ of the unknown parameters. With these estimates, the original data was converted to estimated residuals. Every time, $T(s)$, was replaced with

$$\frac{\log(T(s)) - [\hat{\beta}_0 + \hat{\beta}_1 \log s]}{\hat{\sigma}}$$

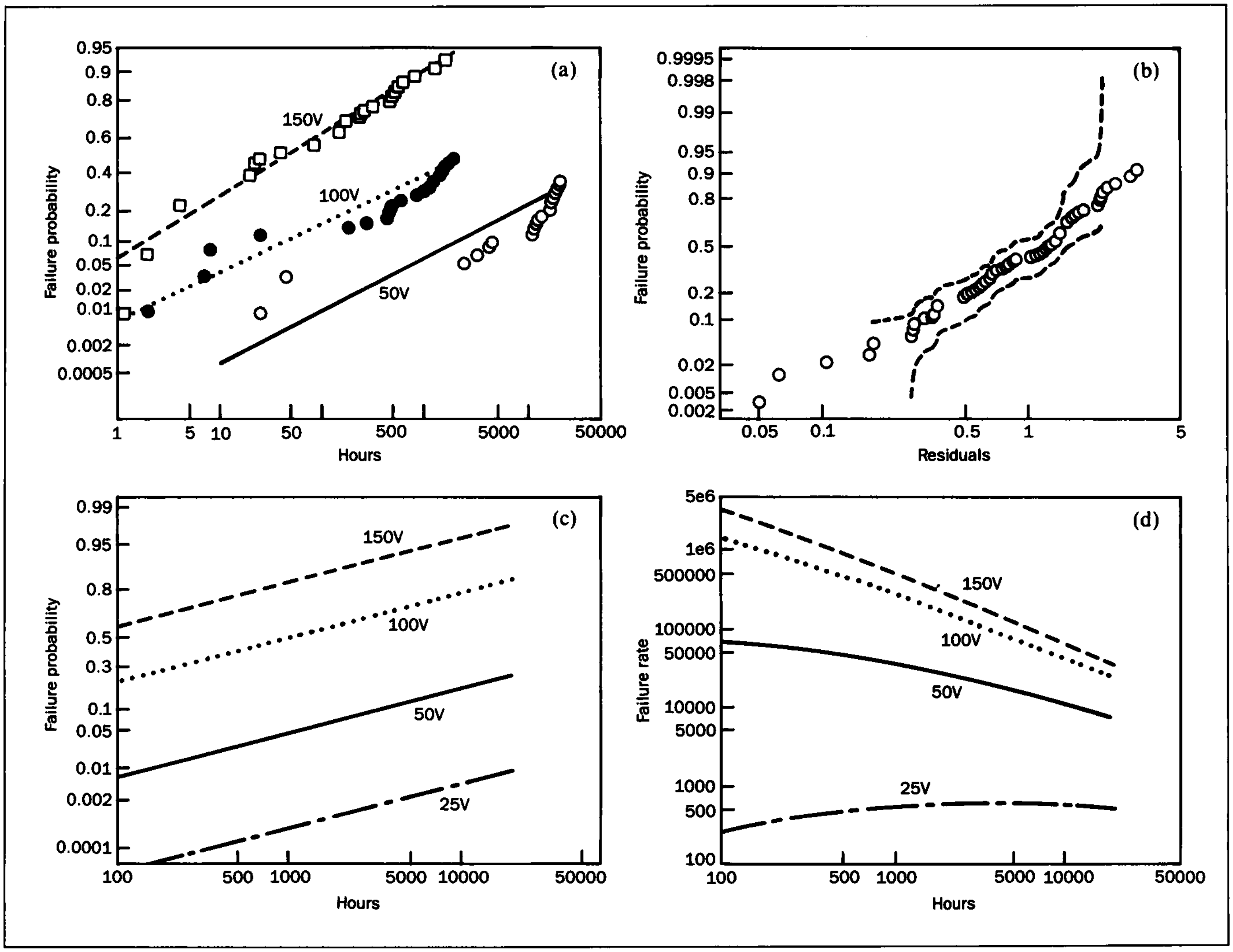
in Figure a and a normal probability plot was done with STAR, the statistical reliability analysis software.

When the modeling assumptions are true, if we subtract an estimate of $\mu(s)$ and divide by an estimate of σ, the data will reduce to a common $\mu = 0$ and $\sigma = 1$. In fact, the linearity of the residual probability plot indicates that both the AFT model and the inverse-power-stress function fit the data. The plot in Figure b should be linear within the 95-percent confidence bands, if all the assumptions hold.

Because this diagnostic check appears to be satisfied, the same $\mu(s)$ and σ are used to estimate $\mu(s)$ for 25V and get the result in Figure c. Here, the parallel lines are a consequence of the equal σ assumption.

Figure d shows the failure rates (in units of 10^9 hours) of the fitted distributions at the four voltages: 25V, 50V, 100V, and 150V. All lines would show that the failure rate initially increases for sufficiently small times. But, despite the apparent decreasing failure rates exhibited by *all* the test data, we expect capacitors operated at 25V to exhibit a slightly increasing failure rate over the first 10,000 hours.

From a STAR analysis, the probability that a high-K ceramic capacitor operating at 25V will fail in the first year is approximately 0.0063, with a 95-percent confidence interval of (0.00064, 0.059).



Lognormal failure probability for high-K ceramic capacitors. (a) shows the fitted distributions and the circles, dots, and squares give the nonparametric results. (b), a lognormal probability plot with 95-percent simultaneous confidence bands, shows the residuals. (c) shows the probability of failure for four voltages, and (d) shows the failure rates (in units of 10^9 hours) of the fitted distributions for these voltages.

economical to measure periodically the change in degradation of product parameter values. (Previously we simply checked the time when these parameter values exceed a threshold.)

Thus, much useful data is collected even on products that do not fail. This is especially important for some new high-reliability products where few units would fail during a life test.

In addition, if we know how a parameter value will change, we could compensate for this change in the product design, thus greatly improving product reliability.

References

1. N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, Vols. 1 and 2, Houghton Mifflin, Boston, 1970.
2. D. R. Cox, "Regression Models and Life Tables" (with discussion), *Journal of the Royal Statistical Society, Series B. Methodological*, Vol. 34, 1972, pp. 187-220.
3. J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, John Wiley and Sons, New York, 1982.
4. B. W. Turnbull, "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society, Series B. Methodological*, Vol. 38, 1976.

(Manuscript received November 22, 1985)