Statistical Bulletin
Reliability & Variation Research

DETROIT RESEARCH INSTITUTE
P.O. BOX 36504 • GROSSE POINTE, MICHIGAN 48236 • (313) 886-7976

LEONARD G. JOHNSON
EDITOR

WANG H. YEE
DIRECTOR

Volume 11
Bulletin 3

July , 1981
Page 1

BASIC CONCEPTS OF PRODUCT ASSURANCE: ENTROPY AND EVIDENCE

TOPICS DISCUSSED

I HAZARD RATE AND ENTROPY

II SAMPLE SIZE

III CONFIDENCE BANDS

IV AVERAGE ENTROPY PER FAILURE TECHNIQUE

V SEQUENTIAL TESTING VS. FIXED SAMPLE SIZE

HAZARD RATE

When we speak of the hazard rate at a specific age x , we are referring only to items which have survived to age x . Then we ask, "What fraction of the survivors to age x will fail between age x and age(x + 1) ?"

This fraction is the hazard rate at age x . Thus , if x = age, then Hazard at x is given by

$$H(x) = \frac{\text{population portion failing between ages } x \ \& \ (x+1)}{\text{population portion surviving to age } x}$$

Now, from statistical definitions:

Population Portion Failing by Age x

= Cumulative Distribution Function of Population at Age x

= $F(x)$ (Capital letter F is used)

So, Population Portion Surviving to Age $x = 1 - F(x)$, and

Population Portion Failing Between Ages x and $(x + 1)$

$= F(x + 1) - F(x)$.

So,

$$H(x) = \frac{F(x+1) - F(x)}{1 - F(x)}$$

Now, let

$f(x)$ = Probability Density Function of the Life for the Population

small letter f = Derivative of $F(x)$ with respect to $x = \frac{dF(x)}{dx}$.

By definition:

$$\frac{dF(x)}{dx} = \lim_{\Delta \to 0} \frac{F(x+\Delta) - F(x)}{\Delta}$$

Now, take $\Delta = 1$.

Then $$\frac{dF(x)}{dx} \approx \frac{F(x+1) - F(x)}{1}$$

So,

$$H(x) = \frac{\frac{dF(x)}{dx}}{1 - F(x)} = \frac{f(x)}{1 - F(x)} \qquad (1)$$

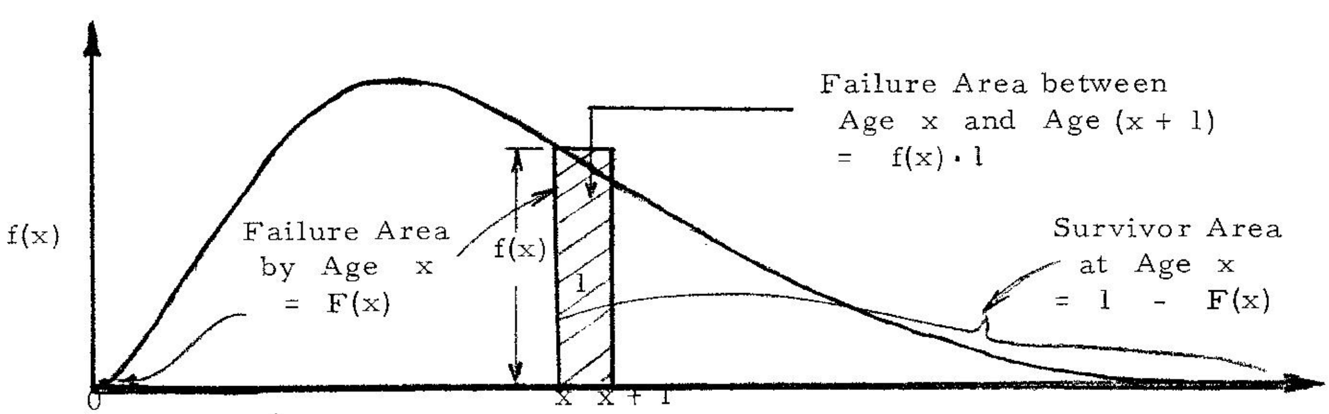


FIGURE 1

ENTROPY

Entropy at Age x is defined as the sum total of all Hazards between Age 0 and Age x.

For a typical system, A Hazard Function plots into a so-called "Bathtub Curve" . (See Figure 2 below.)*

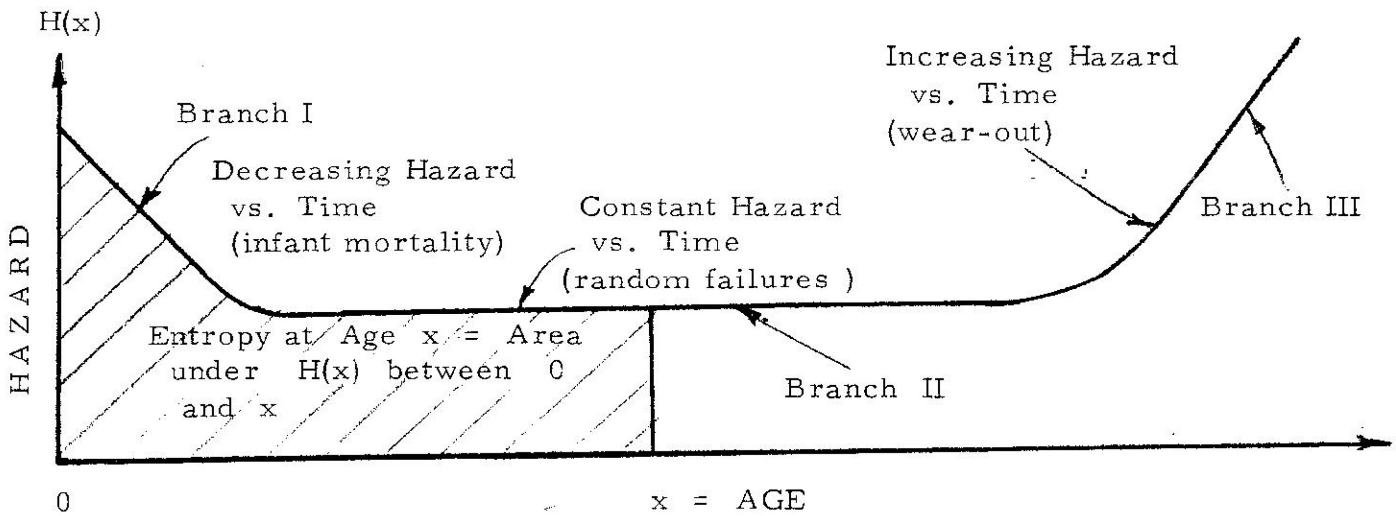


FIGURE 2

Thus , Entropy at x = Integral of H(x) from 0 to x

or $$\mathcal{E}(x) = \int_0^x H(x)\,dx \tag{2}$$

PERFORMING THE HAZARD INTEGRATION

The Entropy Function at Age x is

$$\mathcal{E}(x) = \int_0^x H(x)\,dx = \int_0^x \frac{f(x)dx}{1 - F(x)} = \ln \frac{1}{1 - F(x)} \tag{3}$$

Thus, the Entropy Function at Age x simplifies into the natural logarithm of the reciprocal of the fraction surviving to Age x .

Another name for the fraction of a population surviving to Age x is RELIABILITY TO AGE x .

* For certain systems, it may happen that not all of the three branches I, II, and III exist.

Thus, we can state that $\mathcal{E}(x) = \ln \frac{1}{R(x)}$

($R(x)$ = Reliability to Age x.)

THE CUMULATIVE DISTRIBUTION OF ENTROPY TO FAILURE

According to (3) :

$$\mathcal{E}(x) = \ln \frac{1}{1 - F(x)} \tag{3}$$

Where $\mathcal{E}(x)$ = Entropy at Age x

$F(x)$ = Fraction of Population Failed by Age x .

Solving (3) for $F(x)$ in terms of $\mathcal{E}(x)$,

we obtain

$$F(x) = 1 - e^{-\mathcal{E}(x)} \tag{4}$$

Thus, in terms of Entropy, the fraction failed becomes

$$\text{Fraction Failed} = 1 - e^{-\text{entropy}}$$

or

$$F = 1 - e^{-\mathcal{E}} , \tag{5}$$

which tells us that the ENTROPY TO FAILURE is EXPONENTIALLY DISTRIBUTED.

APPLICATION TO A WEIBULL DISTRIBUTION

I : TWO-PARAMETER WEIBULL

The CDF (cumulative distribution function) of Age at Failure (x) is

$$F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)^b} \tag{6}$$

Where x = Age at Failure

b = Weibull Slope

θ = Characteristic Life

$F(x)$ = Fraction of Population Failed by Age x .

Equation (6) is of the type $F(x) = 1 - e^{-\xi(x)}$ (see (4))

with $$\xi(x) = \left(\frac{x}{\theta}\right)^b \quad (7)$$

Thus, the two-parameter Weibull has a simple <u>Power Function</u> as its Entropy Function.

II : THREE-PARAMETER WEIBULL

The CDF of Age at failure is

$$F(x) = 1 - e^{-\left(\frac{x-\alpha}{\theta-\alpha}\right)^b} \quad (8)$$

Where x = Age at Failure

b = Weibull Slope

θ = Characteristic Life

α = Minimum Life

Equation (8) is of the type (4) with $\xi(x) = \left(\frac{x-\alpha}{\theta-\alpha}\right)^b$ (9)

Thus, in the case of the three-parameter Weibull Distribution, the Entropy Function is a Power Function of the Binomial $(x - \alpha)$.

PROPERTIES OF THE ENTROPY FUNCTION

<u>PROPERTY 1 :</u> At the beginning of a life distribution, the ENTROPY IS ZERO.

<u>PROPERTY 2 :</u> Entropy is a monotone increasing function of age.

<u>PROPERTY 3 :</u> With increasing age the entropy increases, and can reach any given large positive value, provided the Age is made large enough.

SAMPLE SIZE

QUESTION : What basic concept governs sample size decisions in testing programs and field data feedback ?

ANSWER : The concept of EVIDENCE.

A sample size must be of such a magnitude that it yields sufficient EVIDENCE in favor of or against the hypothesis under investigation. Then, and only then, are we wisely able either to ACEPT or REJECT the hypothesis.

THE MATHEMATICAL DEFINITION OF EVIDENCE

The Theory of Evidence is an exact mathematical science with an exact mathematical definition of the concept of evidence.

Let C = Confidence in some hypothesis, i.e., the confidence that the hypothesis is true . (confidence in favor)

Then 1 - C = Confidence against the hypothesis, i.e., the confidence that the hypothesis is false . (confidence against)

Then , Odds Ratio for the Hypothesis is:

$$W = \frac{C}{1 - C} = \frac{\text{Confidence in Favor}}{\text{Confidence Against}}$$

DEFINITION

Evidence in Favor of the Hypothesis $= E = \ln W = \ln\left(\frac{C}{1 - C}\right)$.

EVIDENCE = NATURAL LOGARITHM OF THE ODDS

QUESTION : What property of Evidence makes it so useful ?

ANSWER : Its additivity .

THE ADDITIVITY THEOREM

If Evidence is gathered from two separate and independent data gatherings on some hypothesis, then the total Evidence in favor of the hypothesis is the algebraic sum of the separate amounts of Evidence obtained from the two independent data sets .

Thus, if E_1 = Evidence from data set number 1 .

and E_2 = Evidence from data set number 2 .

Then, $\hat{E}$ = Total Evidence = $E_1 + E_2$

(This is the secret of the power of Sequential Analysis).

THE EXACT STATISTICAL DISTRIBUTION FUNCTION OF EVIDENCE

Since, in terms of Confidence C, the Evidence E is defined as

$$E = \ln\left(\frac{C}{1 - C}\right) ,$$

It follows that the mathematical formula for Confidence C (in favor of an hypothesis) in terms of Evidence E is the Inverse Function , written as follows :

$$C = \frac{1}{1 + e^{-E}}$$

This is called a LOGISTIC DISTRIBUTION FUNCTION .

NORMAL APPROXIMATION OF THE DISTRIBUTION OF EVIDENCE AS OBTAINED FROM THE LOGISTIC FUNCTION

From the Logistic Formula for Confidence C in terms of Evidence E , we can state the following rules :

The Evidence is approximately normally distributed with

$$\left\{ \begin{array}{l} \text{MEAN} = 0 \\ \text{STANDARD DEVIATION} = \pi/\sqrt{3} \end{array} \right\}$$

For an analytical discussion of the Logistic Distribution , see

E. J. Gumbel : Statistics of Extremes; Columbia University Press, 1958 page 126 -127

NON-PARAMETRIC SAMPLE SIZE AT QUANTILE LEVEL Q AS RELATED TO EVIDENCE AND CONFIDENCE BAND RANGE RATIO

I : POPULATION INFINITE

$$\text{SAMPLE SIZE} = N_\infty = \frac{12}{Q}\left(\frac{E}{\pi b \ln \rho}\right)^2$$

Q = Quantile Level (e.g., Q = .2 at B_{20} Life)

b = Weibull Slope of data plot

ρ = Range Ratio of $\left(\frac{\text{Right Hand Limit}}{\text{Left Hand Limit}}\right)$ of the Confidence Band .

E = Evidence corresponding to right hand one-sided confidence (e.g., in a 90% confidence band, we determing the Evidence corresponding th 95% confidence).

(In general , for a band of width W (confidence-wise), we determine the Evidence for the one-sided confidence level of (1 + W/2))

II : FINITE POPULATION OF SIZE T .

$$\text{NON-PARAMETRIC SAMPLE SIZE} = N_T = \frac{1}{\frac{1}{T} + \frac{1}{N_\infty}}$$

T = Population Size

N_∞ = Required Sample Size from an infinite population

$$= \frac{12}{Q}\left[\frac{\ln\left(\frac{1+W}{1-W}\right)}{\pi b \ln \rho}\right]^2$$

W = Confidence Band Width (Confidence-Wise)

ρ = $\left(\frac{\text{Right Limit}}{\text{Left Limit}}\right)$ of Confidence Band

b = Slope of Weibull plot

Q = Quantile Level

PARAMETRIC SAMPLE SIZE FORMULA

If we know the population Weibull slope (for sure), then we can omit the Q in the denominator of the Sample Size Formula for an infinite population , and write

$$N_{\infty} = 12 \left[\frac{\ln\left(\frac{1 + W}{1 - W}\right)}{\pi \; b \; \ln \rho} \right]^2 \tag{1}$$

(Parametric Sample Size from an Infinite Population.)

For a Finite Population of Size T ,

$$N_T = \frac{1}{\frac{1}{T} + \frac{1}{N_{\infty}}} = \text{PARAMETRIC SAMPLE SIZE FROM A FINITE POPULATION}$$

N_{∞} given by Formula (1) above .

CONFIDENCE BAND WIDTHS

Range Ratio ρ for 90% Bands:

(For Weibull Slope b and Sample Size N_{∞})

NON - PARAMETRIC : $\rho = 19^{\left(\frac{1}{\pi b \sqrt{\frac{N_{\infty} Q}{12}}}\right)}$

PARAMETRIC : $\rho = 19^{\left(\frac{1}{\pi b \sqrt{\frac{N_{\infty}}{12}}}\right)}$

N_{∞} = Sample Size from infinite population

NOTE : If we are given a sample N_{actual} from a population of size T,

then $$N_{\infty} = \frac{N_{actual}}{1 - \frac{N_{actual}}{T}}$$

Range Ratio ρ for a W Band :

(For Weibull slope b and Sample Size N_∞ (from an infinite population) at Quantile Level Q)

NON-PARAMETRIC :

$$\rho = \left(\frac{1+w}{1-w}\right)^{\left(\frac{1}{\pi b \sqrt{\frac{N_\infty Q}{12}}}\right)}$$

PARAMETRIC :

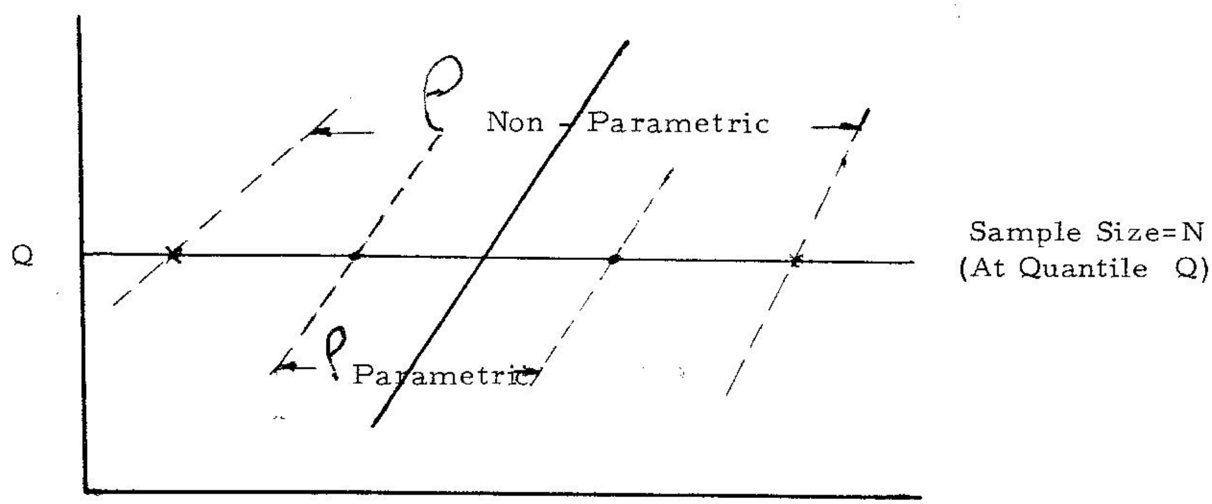
$$\rho = \left(\frac{1+w}{1-w}\right)^{\left(\frac{1}{\pi b \sqrt{\frac{N_\infty}{12}}}\right)}$$

(N_∞ = Sample Size from an infinite population)

NOTE: If we are given a sample size N_{actual} from a finite population of size T , then

$$N_\infty = \frac{N_{actual}}{1 - \frac{N_{actual}}{T}}$$

IMPORTANT FUNDAMENTAL RELATIONSHIP



For the same Sample Size N at Quantile Level Q, the relationship between the Non-Parametric Range Ration and the Parametric Range Ratio for corresponding Confidence Bands is

$$\rho_{\text{NON-PARAMETRIC}} = \rho_{\text{PARAMETRIC}}^{\frac{1}{\sqrt{Q}}}$$