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 9 Bulletin 8 February, 1980

ENTROPY APPLICATIONS IN NORMAL ANALYSIS

INTRODUCTION

One of the most useful concepts for simplifying statistical decisions in all kinds of situations is the concept of ENTROPY. Mathematically speaking, entropy is represented by the absolute value of the logarithm of some type of probability. In reliability theory, RELIABILITY is a SURVIVAL PROBABILITY and, consequently, ENTROPY is the absolute value of the logarithm of the reliability. For this reason, in reliability studies we write

$$\text{ENTROPY} = \ln\left(\frac{1}{\text{RELIABILITY}}\right),$$

i.e., we use the reciprocal of the reliability in order to make the logarithm positive.

In other studies which make use of entropy, the probability under consideration is some kind of probability of success. Hence, in such studies,

$$\text{ENTROPY} = \ln\left(\frac{1}{\text{PROBABILITY OF SUCCESS}}\right),$$

where, again, the reciprocal of the probability is used in order to produce a POSITIVE ENTROPY.

In this bulletin we shall discuss various applications of entropy for problems whose probabilities are derived from NORMAL DISTRIBUTIONS.

DECISIONS ON NORMALLY DISTRIBUTED DESIGN GOALS

GENERAL SITUATION IN WHICH A NORMAL DISTRIBUTION IS USED:

Suppose the designed product is supposed to have a MEAN LIFE of M hours, with a STANDARD DEVIATION OF σ hours. Then, the DESIGN LIFE GOAL looks as follows (A Bell-Shaped Curve) :

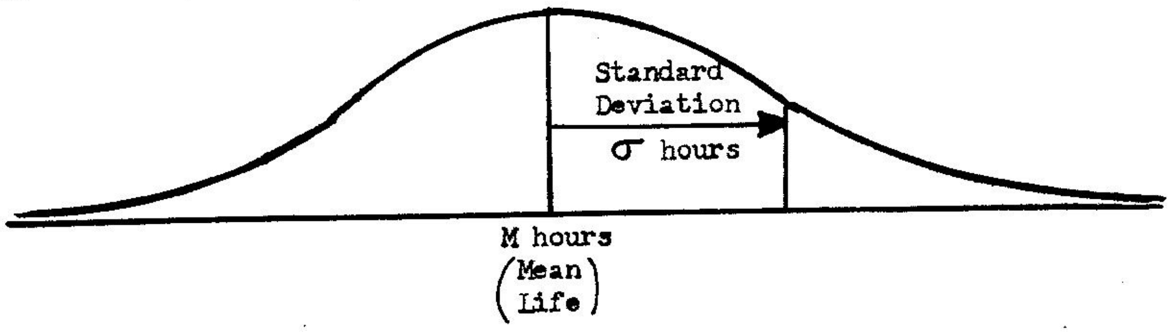

FIGURE 1

Now, suppose a test is run on a sample of N items of this particular design. In this test, all N items are run to failure, and the results are the following order statistics:

(TEST SAMPLE)

- X_1 Hrs. (Shortest Life of the N)
- X_2 Hrs.
- X_3 Hrs.
- .
- .
- .
- X_N Hrs. (Longest Life of the N)

QUESTION: To what extent (confidence) does the test sample demonstrate the product's ability to meet the design goal pictured in FIGURE 1 ?

DECISION MAKING BY THE USE OF ENTROPY

The test sample $(X_1, X_2, \ldots X_N)$, which is supposed to live up to a NORMAL LIFE DISTRIBUTION GOAL with $\left(\begin{array}{l}\text{MEAN LIFE} = M \\ \text{STANDARD DEVIATION} = \sigma\end{array}\right)$, can be judged by first calculating the ENTROPY represented by each of the lives $(X_1, X_2, \ldots X_N)$. Then these entropies are SUMMED UP and divided by N. The result is the AVERAGE ENTROPY PER FAILURE for the test sample. The decision concerning the extent to whish the test sample lives up to the design goal is then judged from the following statistical law:

THE STATISTICAL SAMPLING LAW OF ENTROPY:

Test samples of size N, which come from a SPECIFIED DESIGN GOAL DISTRIBUTION, will exhibit within that distribution

(a) MEAN of all AVERAGE ENTROPIES PER FAILURE = 1

(b) STANDARD DEVIATION of all AVERAGE ENTROPIES PER FAILURE = $\frac{1}{\sqrt{N}}$

The extent to which the ACTUAL TEST SAMPLE shows an AVERAGE ENTROPY PER FAILURE above UNITY will determine the CONFIDENCE that the sample comes from a population SUPERIOR to the DESIGN GOAL POPULATION. In fact, the desired confidence is given by the formula:

CONFIDENCE OF BEATING THE DESIGN GOAL = NORMAL AREA to a Z-SCORE of $\sqrt{N}(\mathcal{E}_{ave.}-1)$,

where $\mathcal{E}_{ave.}$ = Average Entropy per Failure for the test sample.

HOW TO CALCULATE THE AVERAGE ENTROPY PER FAILURE

In order to determine the Average Entropy per Failure for the test sample $(X_1, X_2, \ldots X_N)$, proceed as follows:

$$\text{ENTROPY OF } X_1 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } X_1 \text{ IN THE DESIGN GOAL DISTRIBUTION}}\right)$$

$$= \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_1 = \frac{X_1 - M}{\sigma}}\right) \quad (1)$$

Likewise,

$$\text{ENTROPY OF } X_2 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_2 = \frac{X_2 - M}{\sigma}}\right) \quad (2)$$

$$\vdots$$

$$\text{ENTROPY OF } X_N = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_N = \frac{X_N - M}{\sigma}}\right) \quad (N)$$

TOTAL ENTROPY OF TEST SAMPLE $(X_1, X_2, \ldots X_N)$

$$= \sum_{i=1}^{N} \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_i = \frac{X_i - M}{\sigma}}\right)$$

Dividing this summation of entropies by N gives

$$\text{AVERAGE ENTROPY PER FAILURE} = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_i = \frac{X_i - M}{\sigma}}\right)$$

A NUMERICAL EXAMPLE IN LIFE TESTING

Suppose a product has as its design life goal a NORMAL DISTRIBUTION with $\left\{\begin{array}{l}\text{MEAN LIFE} = 2000 \text{ hrs.} \\ \text{STANDARD DEVIATION} = 400 \text{ hrs.}\end{array}\right\}$.

This design goal is pictured by the following bell-shaped curve:

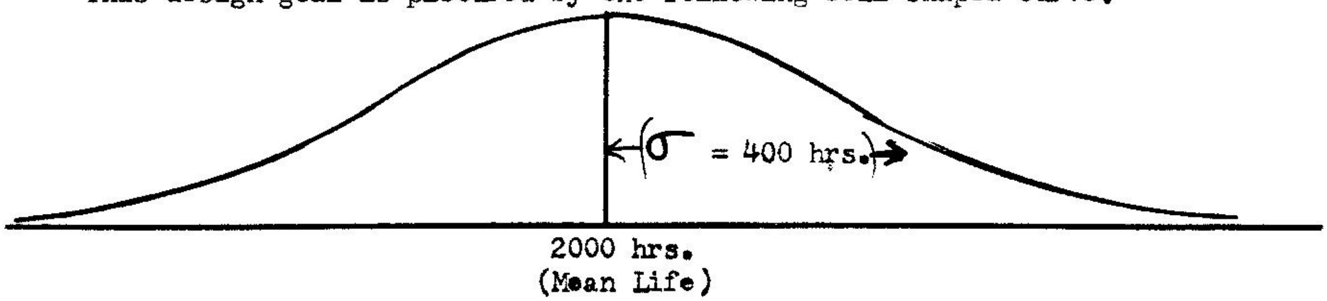

FIGURE 2

TEST RESULTS: Suppose 10 items of this design are tested to failure, with the results listed below:

X_1 = 1750 Hrs.

X_2 = 1996 Hrs.

X_3 = 2076 Hrs.

X_4 = 2280 Hrs.

X_5 = 2410 Hrs.

X_6 = 2501 Hrs.

X_7 = 2550 Hrs.

X_8 = 2625 Hrs.

X_9 = 2708 Hrs.

X_{10} = 2915 Hrs.

The ten ENTROPIES from these ten values are calculated as follows:

$$\mathcal{E}(X_1) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_1 = \frac{1750 - 2000}{400} = -.625}\right) = .30923$$

$$\mathcal{E}(X_2) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_2 = \frac{1996 - 2000}{400} = -.01}\right) = .68520$$

$$\mathcal{E}(X_3) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_3 = \frac{2070 - 2000}{400} = +.175}\right) = .84274$$

$$\mathcal{E}(X_4) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_4 = \frac{2280 - 2000}{400} = +.7}\right) = 1.41898$$

$$\mathcal{E}(X_5) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_5 = \frac{2410 - 2000}{400} = +1.025}\right) = 1.87948$$

$$\mathcal{E}(X_6) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_6 = \frac{2501 - 2000}{400} = +1.2525}\right) = 2.25199$$

$$\mathcal{E}(X_7) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_7 = \frac{2550 - 2000}{400} = +1.375}\right) = 2.47018$$

$$\mathcal{E}(X_8) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_8 = \frac{2625 - 2000}{400} = +1.5625}\right) = 2.82869$$

$$\mathcal{E}(X_9) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_9 = \frac{2708 - 2000}{400} = +1.77}\right) = 3.26074$$

$$\mathcal{E}(X_{10}) = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_{10} = \frac{2915 - 2000}{400} = +2.2875}\right) = 4.50352$$

TOTAL ENTROPY = 20.45075

$$\mathcal{E}_{\text{ave.}} = \frac{\text{TOTAL ENTROPY}}{10} = \text{AVERAGE ENTROPY PER FAILURE} = 2.04508$$

To calculate the CONFIDENCE that this sample indicates that it comes from a population AT LEAST AS GOOD as the DESIGN GOAL, we evaluate

$$\text{Z-score} = \sqrt{N}\,(\mathcal{E}_{ave.} - 1) = \sqrt{10}\,(2.04508 - 1)$$
$$= 3.30483$$

Then,

$$\text{CONFIDENCE} = \text{NORMAL AREA TO LEFT OF } Z = 3.30483$$
$$= .99951 \quad \text{(ans.)}$$

Thus, the test data in this example indicate that we can be 99.951% confifident that the items tested come from a population at least as good as the design goal having a mean life of 2000 hours, with a standard deviation of 400 hours.

APPLICATION TO EMISSION COMPLIANCE TESTS

A similar procedure could be used on data which consist of emission rates of vehicles as compared to a desired design goal.

For example, the DESIGN GOAL for HYDROCARBON EMISSIONS might be a NORMAL DISTRIBUTION with a MEAN of .26 g./mi. and σ = .05 g./mi.

Now, suppose 5 vehicles are tested, with the resulting HC emission rates (in numerical order) being

$$\left\{ \begin{array}{l} .201 \text{ g./mi.} \\ .220 \text{ g./mi.} \\ .251 \text{ g./mi.} \\ .265 \text{ g./mi.} \\ .271 \text{ g./mi.} \end{array} \right\}.$$

The ENTROPIES for these 5 values in the design goal are as follows:

$$\mathcal{E}_1 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_1 = -1.18}\right) = .12670$$

$$\mathcal{E}_2 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_2 = -.8}\right) = .23807$$

$$\mathcal{E}_3 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_3 = -.18}\right) = .55963$$

$$\mathcal{E}_4 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_4 = +.1}\right) = .77618$$

$$\mathcal{E}_5 = \ln\left(\frac{1}{\text{NORMAL AREA TO RIGHT OF } Z_5 = +.22}\right) = .88448$$

ENTROPY SUM = 2.58506

$$\mathcal{E}_{\text{ave.}} = \frac{\text{ENTROPY SUM}}{5} = \frac{2.58506}{5} = .517012$$

Now, to calculate the CONFIDENCE of being at least as good as the design goal (emission-wise) , we MODIFY the life testing confidence formula, and state that

$$\text{CONFIDENCE} = \text{NORMAL AREA TO } \underline{\text{RIGHT}} \text{ of } Z = \sqrt{N}(\mathcal{E}_{ave.} - 1)$$

$$= \text{NORMAL AREA TO RIGHT OF } Z = \sqrt{5}(.517012 - 1)$$

$$= \text{NORMAL AREA TO RIGHT OF } Z = -1.08$$

$$= .86 \text{ (ans.)}$$

Thus, the 5 vehicles tested yield us 86% confidence of being able to meet the desired goal. In other words, we are 86% confident that the 5 vehicles tested come from a population which is at least as good (emission-wise) as the design goal.