STATISTICAL BULLETIN

Reliability & Variation Research

LEONARD G. JOHNSON
EDITOR

DETROIT RESEARCH INSTITUTE

P.O. Box 36504 • Grosse Pointe, MI 48236 • (313) 886-8435

WANG H. YEE
DIRECTOR

Volume 25
Bulletin 2

May, 1995
Page 1

THE SHOCKING TRUTH ABOUT THE WASTEFULNESS OF CLASSICAL TYPES OF SUCCESS RUN TESTING

INTRODUCTION

In this Statistical Bulletin of Detroit Research Institute we are summarizing the main points which were discussed at the Spring Meeting of the Industrial Mathematics Society on May 4, 1995. The topics discussed at the meeting had to do with the wastefully large sample sizes demanded by conventional Binomial Probability Theory for life tests in which items are run to some specific life (bogey) with all of the items being required to survive the bogey period of life. Such tests are called success runs. The big drawback in such tests (with successes only) is that they required such horrendously large sample sizes to demonstrate any acceptable reliability and confidence levels. In the Industrial Mathematics Society Meeting of May 4, 1995 our Editor and Consultant, Leonard Johnson presented some dramatic examples of the proper way to go about testing for product life durability by making proper use of past experience about minimum reliability levels which could be assumed, as well as the obvious facts that we can generate much more powerful life tests for reliability with much smaller sample sizes by testing a few items all the way to failure. This whole concept of life durability testing is of such importance to modern industry that we felt it necessary to publish our ideas in this special statistical bulletin in order that interested people would have it available as a reference and reminder that it's imperative that everyone engaged in life testing could improve his testing program by avoiding the wasteful practice of running success run tests based on outmoded classical binomial probabilities.

The Deprogramming of Success Run Joe

Success Run Joe was a reliability analyst at XYZ, Inc.. One day his supervisor (Mr. Cool) called Joe into his office, and the following conversation took place between the two of them:

Mr. Cool: "Our company has just been granted a new contract by a customer who wants us to supply him with our latest design of engines to power certain types of machines which he will sell under his own brand name. The customer wants our engines to be able to run for 2,000 hours with no more than one engine in 100 failing in a 2,000 hour run. You can see, therefore, that the customer wants 99% reliability to a **bogey** of 2,000 hours. Furthermore, he told me that he wants the 99% reliability to be at a confidence level of 90%. What I want you to do, Joe, is to come up with the proper sample size to be run for 2,000 hours without any engine failures."

Joe: "O.K., I'll see what I can come up with in the way of a success run testing program for our engines."

After about 30 minutes Joe went to Mr. Cool's office and told Mr. Cool "We must run 229 engines for 2,000 hours each in order to demonstrate 99% reliability with 90% confidence."

Mr. Cool: "Where did you ever get that sample size?"

Joe: "That comes from the **Success Run Theorem**."

Mr. Cool: "But, Joe, we don't have 229 engines available. I think we need only about 7 engines because our past experience indicates that our engine reliability couldn't ever be worse than 90%. Furthermore, we have found that the failure distribution function can be represented by a declining parabola with mode at 0% failed and upper limit at 10% failed."

Joe: "Oh, I see! I was assuming that the minimum reliability could be 0%, which means the product could possibly have 100% failed at the bogey. I'll go back to my office and figure the needed sample size for a success run under the assumptions you mention."

After about another hour Joe came back to talk to Mr. Cool about his findings. "You were right, Mr. Cool! I find that we only need 7 engines to survive 2,000 hours."

Mr. Cool: "O.K., Joe. Let us go ahead and run 7 engines for 2,000 hours."

Then 7 engines were successfully run in the company testing lab for 2,000 hours each. The customer was satisfied with the demonstration test. Thus, Success Run Joe had learned a very important lesson about prior assumptions in success run testing. Never again did Joe waste test specimens by using the standard classical sample sizes for such testing programs. He had been **deprogrammed** from such absurdities.

LET'S FACE IT!

If you are running a test with nothing but successes to a bogey, you are chasing a moving target.

5 successes	10 successes	15 successes	20 successes	
o	o	o	o	
o	o	o	o	
o	o	o	o	
o	o	o	o	
o	o	o	o	
Possibly	o	o	o	
1 bad in 6 →	o	o	o	
	o	o	o	
	o	o	o	etc. etc.
	o	o	o	
	Possibly	o	o	ouch!
	1 bad in 11 →	o	o	
		o	o	not there
		o	o	yet
		o	o	
		Possibly	o	When will
		1 bad in 16 →	o	it end?
			o	
			o	
			o	
			Possibly	
			1 bad in 21 →	
			Still chasing the target !	

When will we catch the true failure rate?

ANSWER

If you want to home in on the true failure rate (or reliability) to the bogey, you must run some items long enough to make them actual fail. In other words, go beyond the bogey until you get some failures. Failures can give us Weibull plots for life distributions. Successes just tell us that we haven't caught up with the true rating we should assign to the product.

If you want to evaluate a product's durability, then for economy's sake, get some failures. Then you can begin an evaluation of the product's true life and reliability. Successes alone only tell you that you aren't there yet as far as evaluating reliability is concerned.

LET'S USE SOME COMMON SENSE!

If I'm going to sell 100 engines, why should I test 229 engines (all successes) to a bogey just to be convinced of 99% reliability with 90% confidence? What a waste!

Even testing 100 of them is asking too much. I don't want to produce 200 (to allow 100 to be tested) when I'm going to sell only 100. It makes much more sense to test half a dozen all the way to failure. Then I'll have some idea how good the engines are.

What we need is a clear headed approach which really tests an item long enough to make it actually fail. Then we'll know how much it can take.

CONCLUSION

Success run testing without product verification by actual times to failure requires too large a sample size in order to arrive at its true and verified reliability. If we stop with nothing but successes to a bogey we still won't know how much better it might be.
Recommendation: Run some items to actual failure so that you can construct a Weibull plot, or an Entropy plot with confidence bands.

Success Run Testing for High Reliabilities is a Big Waste.

NOTE: In order to verify the validity of the sample which Mr. Cool suggested, two extra items were run to actual failure, as show on the next page. An Entropy plot then proved that the sample size employed was adequate.

RUNNING TWO EXTRA ITEMS TO FAILURE

RUNNING TIME	STATUS	
2000 Hrs.	7 Unfailed	Original Success Run
3271 Hrs.	1 Failed	(2 Extra Items Run to Failure)
4519 Hrs.	1 Failed	

TABULATION OF MEDIAN ENTROPY ANALYSIS

Time Interval	No. Active	No. Failed	Median Entropy
0 to 3271 Hrs.	2+7(2000/3271) = 6.28	1	(1-.3)/(6.28+.4) = 0.1048
3271 to 4519 Hrs.	1	1	0.1048+1(1+.4) = 0.8191

This last column is the Median Failures per Engine.

ANALYSIS OF THE DATA

Median Weibull Program (Using Median Entropies)

END PT. = 3271
NO. ACTIVE = 6.280037
ENTROPY INC. = .1047898
CUM. ENTROPY = .1047898

END PT. = 4519
NO. ACTIVE = 1
ENTROPY INC. = ..7142858
CUM. ENTROPY = .8190756

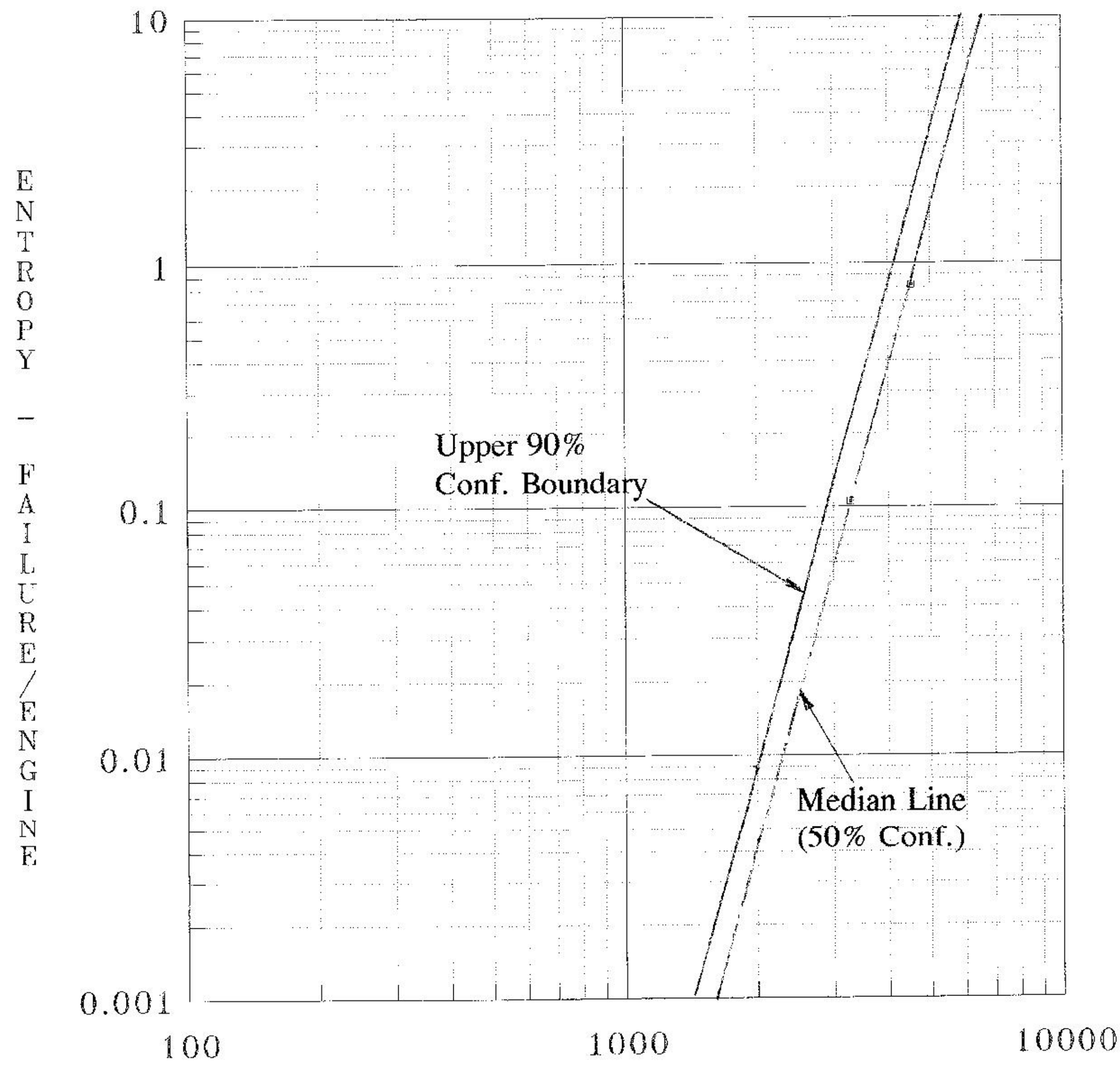
MIN. LIFE = 0
GOODNESS OF FIT = .9998256
WEIBULL SLOPE = 6.359923
CHARACTERISTIC LIFE = 4663.319
B1 LIFE = 2262.391
B10 LIFE = 3273.604
MEDIAN LIFE = 4402.176
B90 LIFE = 5316.771

TARGET = 2000
RELIABILITY TO TARGET WITH 50% CONF. = .9954219
DESIRED CONFIDENCE TO TARGET = .9
SAMPLE SIZE AT TARGET = .9
RELIABILITY WITH CONFIDENCE OF .9 = .991904

Here it is --- 99% Reliability with 90% Confidence to 2000 hours!

Mr. Cool was right! He didn't need 229 successes to 2,000 hours!

PROVING MR. COOL'S CONTENTION
BY RUNNING 2 EXTRA TO FAILURE



ENTROPY PLOT

SOMETHING TO THINK ABOUT

Observation: 100 Consecutive Successes to a Bogey.

This is not just a case of 50 - 50 pure luck (like tossing a coin). Since it was easily obtained on the first time we tested 100 in a row, we must conclude that the probability of getting such a set of 100 consecutive successes is quite high (say, a median chance of 90%). The question we must then answer is **"What Reliability does an individual item have in order that getting a set of 100 successes in a row would have a Median Probability of 90% of happening?"**

Deriving the Answer to the Question

Let R = Estimated probability (with 50% confidence) of 1 item surviving the bogey.

Then, R^{100} = Probability of 100 successes to the bogey.

We reasoned earlier that this probability should have a median value of .9 (90%), since having 100 items all succeeding in not a very likely event unless there is a good chance (like 90%) for such a remarkable thing to happen.

So, putting $R^{100} = .9$, gives us $R = .9^{1/100} = .99895$.

So, according to this type of reasoning, one individual item has a reliability estimated to be 99.895%. According to the **Classical Conservative** (i.e., pessimistic) **Theory**, the reliability of a single item would be estimated to be only 99%. **No wonder Classical Theory is so wasteful!**

Successes don't tell the total durability of a product. (They only tell us "So far, so good".) Only failures tell us how far a product can go. Just as in life insurance, we need mortality tables to pinpoint risks involved in survival.

CONCLUSION

To the honest observer it is quite obvious that the classical approach to success run testing to a bogey (and attribute testing in general) is plagued with a plethora of opinions resulting in much argumentation and confusion.

It is high time for modern industry to knuckle down to business and do some serious research about the question of permissible assumptions regarding the prior probabilities used in the reliability evaluations from attribute tests.