Randomness

by

W. A. Shewhart

Paper to be given before Mathematics
group of BTL at Murray Hill on
February 28, 1946

"A random variable is a measurable function defined on a measure space with total measure 1"

"The Foundations of Probability" by P. R. Halmos, The American Mathematical Monthly, Vol. LI, No. 9, November, 1944.

PART I

"We shall identify as Problem I the problem of setting up a formal calculus to deal with (probability) numbers. Within this discipline, once set up, the only problems will be mathematical. The concepts involved will be ordinary mathematical ones, constantly used in other fields. The words "probability", "independent", etc. will be given mathematical meanings, where they are used.

"We shall identify as Problem II the problem of finding a translation of the results of the formal calculus which makes them relevant to empirical practice. Using this translation, experiments may suggest new mathematical theorems. If so, the theorems must be stated in mathematical language, and their validity will be independent of the experiments which suggested them. (Of course, if a theorem, after translation into practical language, contradicts experience, the contradiction will mean that the probability calculus, or the translation, is inappropriate."

"Probability as Measure", by J. L. Doob, Annals of Mathematical Statistics, June, 1941.

'A man's reach should exceed his grasp
Or what's a Heaven for."

For twenty years I have been searching for a method of obtaining that "Happy Hunting Ground" known as a state of randomness.

Randomness is one of the most important concepts for the applied scientist. Randomness is a basic concept in any scientific theory of control that allows for the fact that the only observable constancy is that associated with a state of statistical control. It is a basic concept in the theory of control of all kinds of repetitive operations whether in the research laboratory or in the mass production process. It is a basic concept to any one who would apply probability theory.

A state of randomness is of the nature of a physical model. Twenty years ago, it was universally considered good practice to assume that any set of data taken under presumably the same essential conditions behaved in accord with such a model; even today I know of no textbook on the theory of errors that does not either implicitly or explicitly make such an assumption. However, work carried on by Bell System engineers within this period has gone a long way towards revealing that such sets of data seldom, if ever, behave in accord with such a model. This is true even for such elite measurements as those on the velocity of light, the gravitational constant, the charge on an electron, and Planck's constant.[1] In spite of this, physicists and chemists still persist in treating such sets of data as random whereas, as Will Feller recently pointed out in a review of "The Mathematics of Physics and Chemistry" by Margenau and Murphy: "They all show a complete lack of statistical control, and even the simplest methods of industrial quality control could be used for an improvement".[2]

- -

1. See, for example, Shewhart, Walter A., Statistical Method from the Viewpoint of Quality Control, Chapter II, published

by the Graduate School, Department of Agriculture, Washington, 1939.

2. Quarterly Journal of Applied Math. Vol. 2, page 91, 1944-45.

Whereas the validity of any mathematical theory of probability and statistics, like that of mathematical theory in any other field of science, does not depend on the existence of assumed model, nevertheless, for the empirical verification of the theory, it becomes a matter of fundamental importance to know if in a given case the physical situation is represented to a sufficient approximation by the assumed model.[1] It is interesting from the viewpoint of applied probability and statistics that the method developed for attaining statistical control in mass production is now becoming recognized by mathematical statisticians as essentially a practical empirical procedure for approximating a state of randomness.[2]

The mathematician makes no attempt to define randomness in observed sequences beyond a rough description in terms of the fluctuations found in sequences of repetitive measurements on operations performed under presumably "the same essential conditions". The mathematician then proceeds to develop a superstructure of formal mathematical theory based upon certain postulates that may or may not have a counterpart in nature. This procedure is, of course, the same as that followed in developing any mathematical theory of natural phenomena. If it were true that even the majority of sets of repetitive observations made under presumably the

- -

2. Wilks, S. S., Mathematical Statistics, p. 3, Princeton Univ. Press, 1943 and Doob, J. L., Annals Math. Stat., Vol.12, No. 2, June, 1941.
1. Cf. Cramer, Harold, Random Variables and Probability Distributions, p. 5, Cambridge Univ. Press, 1937.

same essential conditions behaved in accord with the formal mathematical theory, we would in a majority of cases be safe in applying the theory. Sad but true, however, such sets of observations almost never behave this way. How shall the practical man proceed under such conditions? In what follows, I shall briefly consider the light thrown on this question by Bell System engineers in their study of the theory of statistical control.

$$t = \frac{X - \bar{X}'}{\sigma} \qquad n = \text{sample size}$$

$$dy = \frac{\left(\frac{n-2}{2}\right)!}{\sqrt{\pi}\left(\frac{n-3}{2}\right)!}\left(1 + t^2\right)^{-n/2} dt$$

<u>Exact for any size of sample</u>

1. Before 1900. Large Samples.

1.1 Theory of errors. Gauss's work. Rayleigh's remark.

$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} dx$

1.1.1 Errors assumed to be distributed at random: normal law.

1.1.2 No concerted effort to check. Faith in large numbers.

1.2 Pearson and others in statistics.

1.2.1 Assume laws of probability.

$$dy = f(X, \alpha_1, \alpha_2, \ldots)dX.$$

1.2.2 Emphasis on discovery of laws.

Faith in $\lim_{n \to \infty} \bar{\alpha} = \alpha'$

1. Slide 15455 Statistical limit.

2. 1908. Small Samples.

2.1 Exact theory for small samples.
2.2 Still assume normality of $f(X)$.

2 Slide 22023 Birthplace of theory of small samples

3 Slide 16921 Confidence Limits

Beginning of work on small samples by Fisher and others.

Emphasis on exact theory of small samples.

Fiducial Probability

Concept of a unique sample.

$$\bar{X} \pm t_a \frac{\sqrt{(n+1)/n}}{\sigma}$$

<u>Size n of sample enters</u>

1924 - Quality Control Theory.

Errors of first and second kind.
4. Slide 19000 Testing hypotheses (statistical)
Emphasis on order.

First opportunity to see whether probability applies.

Since repetitive events do not happen "at random" and since it is desirable that something be done to make them happen that way, emphasis directed toward finding an

5 Slide 21879 Operation of Statistical Control

1928 - Concept of

6 Slide 16920 Tolerance limit prediction

Practical importance thereof.
Later work.

1924-46

7 Slide 23040 Science of statistical control

Hypotheses of Statistical Control

Hypothesis IIa: The maximum attainable degree of validity of prediction that an operation will give a value X lying within any previously specified tolerance limits is that based upon the prior knowledge that the probability of this event is q′ or more generally upon the prior knowledge of the law of chance underlying the operation.

Hypothesis IIb: The maximum degree of attainable control of the cause system underlying any repetitive operation in the physical world is that wherein the system of causes produce effects in accord with a law of probability.

Hypothesis IIc: It is assumed that some criterion or criteria may be found and methods developed for their application to the numbers obtained in a sequence of repetitions of any operation such that whenever a failure to meet the criterion or criteria is observed, an assignable cause of variability in the results given by the operation may be discovered and removed from the operation. It is further assumed that, by the removal of a comparatively small number of causes, a state of statistical control is approached where the results of repetitions of the operation behave in accord with a law of chance.

Experiment

Hypothesis shapes the experiment.

How choose a random sample.

Shuffle cards.

Shake well before taking.

Test of Control Hypothesis

Choice of criteria appropriate for testing errors of first and second kinds.

Note 1 - The choice of criteria depends upon type or types of assignable causes assumed present. Smallness of probability is not in itself indication of an assignable cause.

Note 2 - Test of statistical control hypothesis is not the same as a test of a statistical hypothesis. The latter depends on relevant information besides that given by a unique sample.

Part II

Repetitive Operation under the "same essential conditions"

1. Examples.

   1.1 Throw of an ace with a die.
   1.2 Draw of a chip from a bowl.
   1.3 Shooting at a target.
   1.4 Measurement of a physical constant.
   1.5 Routine analytical measurements.
   1.6 Mass production process.

370 Fundamental Problem of prediction } Formalist route / Control route

Definitions of Probability

1. Three Kinds of Probability

   1.1 Formal - A mathematical theory of arrangement.

   Ratio P_m of the number m of members of a subclass to the number n of members of the whole class.

   1.2 Relative frequency p of an event in the long run.

   1.3 Degree of belief p_b.

2. Discussion.

   2.1 Formal - Example: Probability P_m of an ace for a die is 1/6.

   2.2 Relative Frequency.

A. If an event can happen in a certain number of ways, all of which are equally likely, and if a certain number of these be called favorable, then the ratio

(over)

Chance of throwing an ace with a die is 1/6 judged upon the symmetry of the die. In general, however, calling the event equally likely is begging the question. Furthermore, in most repetitive operations, including those of mass production, we have no way of determining conditions that are equally likely.

The symmetry of the die is not the whole influence in the repetitive operation because the operation includes the act of throwing.

2 If an event which can happen in two different ways be repeated a great number of times <u>under the same essential conditions</u>, the ratio of the number of times that it happens in one way, to the total number of trials, will approach a definite limit, as the latter number increases indefinitely.

When are conditions essentially the same?

<u>Example</u> - Let us consider n throws of an ace with a die in which m of the n throws give an ace. The classical approach is to assume that the limit

$$\lim_{n \to \infty} m/n = P_m$$

is a mathematical probability. One such series of throws gave

00000000000010000000010000001110001001011
00001000001010000010000000000000000000001
101000000010000000

for 102 throws where 1 represents the occurrence of an ace and 0 any other number.

If, however, we got a sequence in which 17 one's were followed by 85 zeros, or a sequence like

1000001000001000001000001000001000001000001

where each 1 was followed by five 0's, or some other similar series, would you in such a case use 1/6 as a probability. In other words, it is assumed that there are certain sequences to which probability does not apply. To get around this difficulty, von Mises proposed that in any infinite sequence if we choose, without looking at the sequence, by some rule a subsequence in which we observe m_1 successes in n_1 trials,

$$\lim_{n_1 \to \infty} m_1/n_1$$

should be P_m.

Certain formal difficulties, partially removed by the work of Wald, Copeland, Church, and others. Practical difficulty - no way of getting an infinite sequence.

2.3 Degree of belief

Original data as Evidence E ———— Prediction P

Degree of belief p_b in prediction P based on evidence E

Importance of p_b in throw of die

" " " in mass production

Statistical Control Approach to a Practical Problem

Slide 21614 - 144 Observations of thickness of inlay (in arbitrary units).

The question that the control theorist asks is not: Do these vary at random along the strip?, because this is more or less meaningless. Instead he asks: Are there any assignable or findable causes of variation such as must be removed before probability theory applies rigorously?

Slide 21617 - Observed order and an order drawn from a bowl.

Slide 25405 - Four arrangements of the possible 144 different arrangements.

Probably no one here would call in a statistician before looking for assignable cause.

Slide 25407 - Three other arrangements of the 144 numbers.

You might think of calling in a statistician even on the fifth. The 6th was drawn from a bowl with thorough mixing; the 7th is as they occurred.

Three Types of Assignable Causes

Continuous: Effect of rolls, end effects, correlation.
Discontinuous: Nonhomogeneity in crystal structure.
Mixture of above.

In general,

$$ {}_{L}R_{N} = \frac{{}_{L}C_{N}}{V_{N}} = \frac{X_1 X_{L+1} + X_2 X_{L+2} + \cdots + X_N X_L - (\Sigma X_i)^2/N}{\Sigma X_i^2 - (\Sigma X_i)^2/N} $$

Let

$$ \delta^2 = \frac{\sum_{i=1}^{n-1} (X_{i+1} - X_i)^2}{n - 1} \qquad s^2 = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n} $$

$$ \eta = \delta^2 / s^2 $$

the statistical control engineer postulates certain causes and then looks to see if there is evidence of the existence of these causes in the data. He needs a test or statistical criterion for each effect in order that he may not look for trouble more than an economic percent of the time when it is not there.

Statistical Criteria

Distribution of averages.
" " standard deviations.
" " ranges
" " lambda
" " max and min
" " p
" " length of runs up and down
" " length of runs above and below median) BTL
" " length of runs above and below percentile) BTL
" " auto-correlation
" " eta (Von Neumann)
" " variance.
t, z.

Further Study of 144 Observations

Control Chart

Slide 21616 - Points go outside control limits 4 times; in commercial work this indicates that an assignable cause is present.

Serial Correlation - Correlogram for different lags; series not long enough to show cycles if they exist.

→ DTA Fig

<u>Runs up and down</u> - 6 longest run.

<u>Runs above and below median</u>

Expected no. of runs	73	sig. diff.
Observed " " "	55	

One run of 13 above median - In a sequence of 144 a run of thirteen might be expected to occur less than once in every one hundred such sequences.

Three runs of 7 and one run of 8.

<u>Conclusion</u> - Even though there is some serial correlation, there is no definite evidence of pseudo-cycles. In other words, there is definite evidence of discontinuities; certainly not in a state of statistical control along the strip.

Birge Data

Residuals for 500 Measurements of a Spectrum Line

Slide 26563 - Data grouped together and normal law test applied: typical of what was done at the turn of the century.

Slide 25376 - Autocorrelation

n = 500
Lag 1   r = .103   1% level = .102

Runs above and below median

OK for 500
Samples of 4: Long runs in both averages and standard deviations.
Samples of 20: Long runs in average
Run of 8   P = .007

t-test

Significant differences between groups of 100.

P for t-test between last 200 and first 300 considerably less than .01.

Simple BTL test reveals assignable causes.

Other Problems

Under water devices
Bombing devices
Secret message devices
Radar
Radiotrons
Condenser paper
Contact phenomena
Rubber
Navy big rifles

Slide 25408 - Butting type relays

Runs above and below median, lambda test, and maximum range test useful in checking against the types of causes that may be present.

Slide 25375 - Resistance measurements, panel bank contacts.

Criterion of control indicates certain contacts are good and others are bad.

Slide 25406 - Noise measurements on column of 100 contacts - sliding brush.

indicates kind of trouble.

Analytical Control

B. L. Clarke program.

9 Slide 26562

0 Slide 26561

Differences large enough that they do not need statistics.

Such information vital if we are to control raw materials in order to set most effective specification requirements in testing materials to insure quality of product that will meet the post-war conditions of competition.

Accuracy-precision chart - Simple device for using pairs of observations that the chemist likes so much.

Outside of circle means assignable cause present; outside limit lines means change in standard deviation and possibly change in expected value. Too many points above diagonal indicates correlation. Points lying within the limits but outside the circle indicates change in expected value. All of this obtained by simple plotting of the points.

Conclusion

Simpler aspects of statistical control have been introduced into inspection and production throughout the United States, Great Britain, Australia, and Canada. Some of the simpler procedures used by both Army and Navy. Savings during the war running into millions reported from many different concerns.

Committee under the National Research Council, with liaison members from the Army and Navy, the National Physical Laboratory in Great Britain, the Council for Scientific and Industrial Research in Australia, and the National Research Council of Canada, aims to further research in the development of statistical control procedures.

A national society has recently been formed with G. D. Edwards of our Laboratories as president, for the promulgation and study of the applications in inspection and production

My own feeling is that we have only scratched the surface of potential contributions in applied research not only in the field of engineering but in the production of foods, drugs, and everything consumed by the public. Some of the fields of particular importance in our own industry are the following:

Statistical control of

1. Method of measurement (chemical
(physical
2. New (materials
(processes
(types of apparatus
3. Specifications
4. Field trials
5. Cost accounting practices
6. Consumer standards

Slides no. 15455
22023
16921
19000
21879
16920
23040
22370
21614
21617
25405
25407
21616
26563
25376
25408
25375
25406
26562
26561

20 slides - 1-3/4 hours with interruptions.