KUNGL. TEKNISKA HÖGSKOLANS
HANDLINGAR

TRANSACTIONS OF THE ROYAL INSTITUTE OF TECHNOLOGY
STOCKHOLM, SWEDEN

Nr 27 1949

UDC 539. 431.

A STATISTICAL REPRESENTATION OF FATIGUE FAILURES IN SOLIDS

BY

W. WEIBULL

GÖTEBORG 1949
ELANDERS BOKTRYCKERI AKTIEBOLAG

Table of Contents

Introduction

When some ten years ago the Author (1, 2) published his statistical theory of the strength of materials, he was quite aware of the possibility of applying the theory to other problems than the ultimate strength of materials at statical loads. It appeared particularly promising to try the idea on fatigue problems for several reasons.

To begin with, the fatigue failures are much more frequent in the normal service of engineering products, and accordingly more important, than failures caused by statical overloads. Furthermore, fatigue failures occur almost always at stresses below the elastic limit of the material, thus allowing the exact calculation of the stress distribution on the fractured surface before the crack starts. This is not the case at a static load, where the fracture generally is preceded by large plastic strains combined with a complete change of the stress distribution in the neighbourhood of the crack. Finally, the scatter of an endurance test is usually much greater than that of a static test.

When it first became evident — which was not very long ago — that a small stress, sufficiently repeated, was able to cause a failure, the designer based his calculations on the simple rule that the stress should be well below the endurance limit, i. e. the stress below which it is known that an infinite number of stress cycles can be borne. It seemed too hazardous to allow stresses giving a finite life.

The increasing demand for more economical or less heavy constructions has actualized the question of the relation between the load and the life. It seems as though this problem was first introduced in the ball bearing design, where an infinite life would give economically prohibitive dimensions and later on in the airplane constructions, where the weight is of vital importance.

Now it is not sufficient to know the load, at which failure never will occur, but also the number of cycles endured at high loads. In this way, there has been a continuous transition from statical to dynamical theory of strength.

Furthermore, at all real, not idealized materials there is a dispersion of the properties of such a magnitude that statistical points of view have to be introduced.

The final question calling for an answer is accordingly: How great is the probability that a specimen at a given load will endure a certain number of load cycles? We have, thus, three factors, the relation between which must be known:

1) The probability of failure P, 2) The load S, defined by the range of stress applied during a cycle, and 3) The life N, given as the number of load cycles.

1. Statistical Aspects

If a progressively increasing load S is applied to a specimen, failure finally occurs at a load, denoted S_ν. If this procedure is repeated on n specimens of nominally the same material and dimensions, the ultimate loads S_ν, where $1 \leq \nu \leq n$, will differ, provided the load is determined with sufficient accuracy. This scatter of the values is caused by unknown factors, thus preventing the prediction of the value of some specified individual of the population. Then, S_ν is called a random variable.

The only way to describe such a variable is to give a statistical description of the properties of the population, i. e. to give the number of individuals with an ultimate strength equal to or less than any given load S. In many cases this knowledge is quite sufficient and may provide means to deduce statistical properties of other populations of interest.

In order to facilitate the understanding of the statistical procedure, we now let each one of the n specimens be represented by a card. On the front side of these n cards we write down the values S_ν, arrange the cards in the order of increasing magnitude of S_ν and number them from 1 to n. Thus, $S_\nu \leq S_{\nu+1}$. If some of the cards have the same value of S_ν the chosen order is arbitrary, but each card must have its individual number. On the back side of the cards we then write down the figure $P_\nu = \frac{\nu}{n}$. Thus, the card with the smallest load S_1 has on its back side the figure $P_1 = \frac{1}{n}$, and the card with the greatest load S_n has the figure $P_n = \frac{n}{n} = 1$. It is evident that $\frac{1}{n} \leq P_\nu \leq 1$. Obviously, P_ν is a step function.

At increasing n the steps of the variable P_ν will be less and less and it may be substituted with increasing precision by a continuous variable P, uniformly distributed over the half-open interval (0,1). Thus, $0 < P \leq 1$. The ultimate loads S_ν may also be replaced by a continuous variable S.

The relation between these two quantities may be denoted

$$P = F(S) \qquad (1)$$

This function F is called the distribution function of S, and has the important property of giving the probability of failure at the load S, if we define this probability as the number of the specimens with an ultimate strength equal to or less than S divided by the total number of specimens of the population. From this definition it follows that $F(S)$ is a positive, non-decreasing function, which is everywhere continuous to the right.

Evidently,

$$1 - P \equiv 1 - F(S) \qquad (2)$$

is the probability that the load S will *not* cause failure.

If the population is finite, it is possible to have complete statistical information about the population by testing all the specimens, as the exact relation between S_ν and P_ν is then obtained.

If, on the contrary, the population is infinite, the only way is to choose at random a certain number of specimens n and to test them, i. e. to determine their values S_ν. From these values the corresponding values of P_ν have to be determined. This is exactly the same problem as to draw n cards from an infinitly large pack, read the values on the front sides and estimate the values on the back sides of the cards.

Before examining this problem in detail, it may be appropriate to point out that the preceding argumentation can easily be applied also to fatigue failures. If S denotes the repeated load defined by the range of stress applied during a cycle, the corresponding P is the number of specimens in percentage of the total number of the population which have failed after a number of stress cycles equal to or less than a given value N.

In this case, the distribution function is a function of two independent variables, S and N, and takes the form

$$P = F(S, N) \qquad (3)$$

This is the general expression, including also the statical case, as it coincides with (1) for $N = 1$.

Generally, there may be difficulties in finding an analytical closed expression for this two-dimensional probability, in which case one has to be satisfied with the relation between P and S at some given values of N. We introduce the following notations:

$$F_0(S), F_\infty(S), F_6(S) \text{ etc.}$$

for the distribution functions of the ultimate strengths, of the endurance limits, and of the loads giving a life based on 10^6 stress cycles etc.

In the classical theory of strength of material, the arithmetic mean $S_m = \frac{\Sigma S_\nu}{n}$ serves as a substitute for the distribution function $F(S)$, which is equivalent to the assumption that the distribution function has a discontinuity with a unity saltus at S_m. When the scatter of the S_ν is very small, as may be the case for the ultimate strength of ductile material, this assumption is quite acceptable, but for brittle materials such as cast iron, concrete, etc. and for fatigue failures in general, the deviations from the mean S_m are too great to be neglected, and a satisfactory theory demands the introduction of a distribution function.

We shall now return to the case of an infinite population. From an infinitely large pack, we draw n cards, arrange them in the order of increasing magnitude of S, and number them from 1 to n. It is easy to see that the unknown values $P_1 \ldots P_n$ are independent of the values S_ν on the front sides of the cards and accordingly, of the distribution function $F(S)$ in question with the only exception that if we arrange the cards in the order of increasing magnitude of S they will also be arranged in the order of increasing magnitude of P as $F(S)$ is a non-decreasing function. Thus, the distribution function of P_ν, denoted $F_{\nu/n}$, is a function of ν and n only and can be deduced as follows, even if $F(S)$ is unknown:

The probability that the value of P_ν falls between P and $P + dP$ is equal to the product of the following three probabilities:

As P is uniformly distributed over the interval (0,1),

1) the probability that *one* value lies inside the interval $(P, P + dP)$ is dP.

2) the probability that $\nu - 1$ values lie inside the interval (O, P) is $P^{\nu-1}$, and

3) the probability that $n - \nu$ values lie inside the interval $(P, 1)$ is $(1 - P)^{n-\nu}$

The probability that these three events, which are independent of each other, occur at the same time, is equal to the product $P_{\nu-1} (1 - P)^{n-\nu} \cdot dP$. Now, there are different possible combinations of the n values. If $\nu = 1$, anyone of the n values may be that one belonging to the interval dP, which makes n combinations. If $\nu = 2$ and one of the n values belongs to the interval dP, one of the $n - 1$ other values may belong to the interval (O, P), which makes $n\,(n - 1)$ combinations. Generally, the number of combinations is $\frac{n!}{(\nu - 1)!\,(n - \nu)!}$ and the total probability that P_ν lies inside the interval $(P, P + dP)$ is

$$\frac{n!}{(\nu - 1)!\,(n - \nu)!} \cdot P^{\nu-1}\,(1 - P)^{n-\nu}\,dP$$

Hence, it follows that

$$F_{\nu/n} = \frac{n!}{(\nu - 1)!\,(n - \nu)!} \int_0^P P^{\nu-1}\,(1 - P)^{n-\nu} \cdot dP \quad \ldots\ldots (4)$$

and after partial integration

$$F_{\nu+1/n} = F_{\nu/n} - \frac{n!}{\nu!\,(n - \nu)!}\,(1 - P)^{n-\nu} \cdot P^{\nu} \quad \ldots\ldots. (5)$$

For the special case $\nu = 1$

$$F_{1/n} = 1 - (1 - P)^n \quad \ldots\ldots\ldots\ldots\ldots\ldots (6)$$

From (5) and (6) the other distribution functions can be calculated succesively.

In Fig. 1 the distribution functions $F_{1/1}$, $F_{1/2}$ and $F_{2/2}$ are shown, and in Fig. 2 the functions $F_{\nu/5}$. The latter curves have been verified experimentally by drawing 5 cards from a pack of 100 cards numbered from 00 to 99. This has been repeated 200 times and the values of $P_1 \ldots P_5$ have been noted. It thus happened, for example, 7 times that $P_1 = 0$, 9 times that $P_1 = 1$, and 12 times that $P_1 = 2$ etc. Accordingly, the observed values were

$$F_{1/5}(0) = \frac{7}{200};\; F_{1/5}(1) = \frac{7 + 9}{200};\; F_{1/5}(2) = \frac{7 + 9 + 12}{200} \text{ etc.}$$

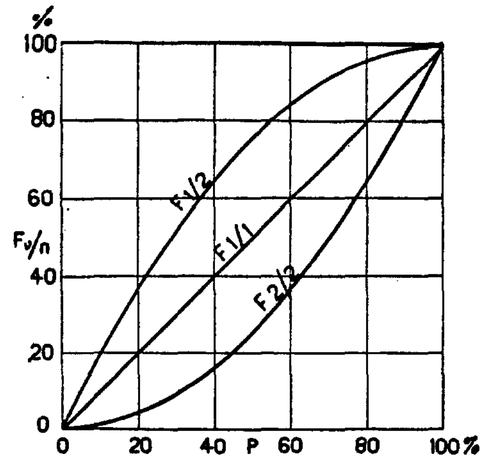

Fig. 1. The distribution functions $F_{1/1}$, $F_{1/2}$, and $F_{2/2}$.

These values are plotted in Fig. 2. As may be seen, observed points do not fall too far from the calculated curves, considering that the number of repetitions is not very large from a statistical point of view.

When the distribution functions are known, it is easy to calculate the arithmetic mean also called »the expected value» of P_ν according to

$$E\,(P_\nu) = \int_0^1 P \cdot dF_{\nu/n}\,(P) \quad \ldots\ldots\ldots\ldots\ldots \quad (7)$$

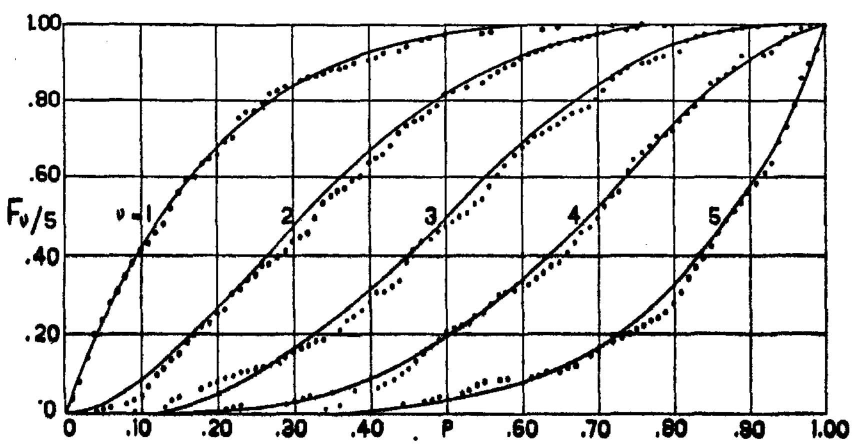

Fig. 2. The distribution functions $F_{\nu/5}$.

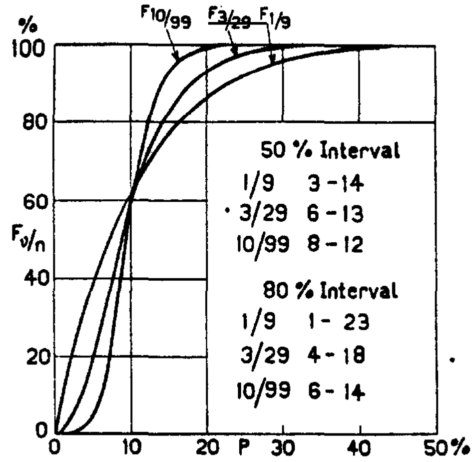

Fig. 3. The distribution functions $F_{1/9}$, $F_{3/29}$, and $F_{10/99}$.

This is the abscissa of the centre of gravity of the distribution of P_ν, which introducing (4) in (7) takes the simple form

$$E\,(P_\nu) = \frac{\nu}{n+1} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (8)$$

This value is to be expected if P_ν is repeatedly determined and the arithmetic mean of the observed values is calculated.

Applied to the above-mentioned experiment, the result was after 200 repetitions

$\nu =$	1	2	3	4	5
Theoretical values . . .	16.67	33.33	50.00	66.67	83.33
Observed values	16.45	34.02	51.02	66.18	83.46

It should be well observed that the distribution functions $F_{\nu/n}$ are calculated for an infinite population, i. e. an infinite pack of cards, and are thus only approximately valid for a pack of 100 cards.

The larger the value of n the more the values of P_ν cluster in the neighbourhood of the value $E\,(P_\nu)$. This is shown graphically in Fig. 3, where the distribution functions $F_{1/9}$, $F_{3/29}$ and $F_{10/99}$ are drawn. All of the curves have the same value $E\,(P_\nu) = 0.1$. The ranges of the intervals (0.25; 0.75) containing 50 % and (0.10; 0.90) containing 80 % of the values are also given.

From the distribution function, all information about the dispersion of the random variable may be obtained. A more vague measure of dispersion is known as the standard deviation, denoted D_ν.

We have

$$D_\nu = \sqrt{\frac{\nu\,(n+1-\nu)}{(n+1)^2\,(n+2)}} \quad \ldots\ldots\ldots\ldots\ldots \quad (9)$$

This gives, for the extreme values $\nu = 1$ and $\nu = n$

$$D_1 = D_n = \sqrt{\frac{n}{(n+1)^2\,(n+2)}} \quad \ldots\ldots\ldots\ldots \quad (10)$$

and for the median $\left(\nu = \frac{n+1}{2}\right)$

$$D_{\frac{n+1}{2}} = \frac{1}{2\sqrt{n+2}} \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad (11)$$

which for large n tends to

$$D_1 = D_n \rightarrow \frac{1}{n} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (12)$$

and

$$D_{\frac{n+1}{2}} \rightarrow \frac{1}{2\sqrt{n}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (13)$$

from which it follows that the extreme values will be determined with greater precision than the median.

If we now return to the drawing of the cards we may assume that the value of P_ν we expect to find on the back side of a card is calculated according to (8).

Thus, we put

$$P_\nu = \frac{\nu}{n+1} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (14)$$

and have to accept the inevitable deviations from this value as sampling errors.

There is no logical necessity of adopting the arithmetic mean as the value to be expected on the back side of the card. It would have been quite possible to choose, for instance, the median or the mode, but these values are not as simply calculated as the mean and there is no other reason to give them a preference. Besides, the differences between these three values tend to zero as $n \to \infty$.

If we now have calculated the values of P_ν from (14) it is very easy to obtain the corresponding value of S. As the function $F(S)$ is a non-decreasing function it follows from (1) that there is no change of the sequence if we arrange the cards in the order of increasing magnitude with regard to S instead of P. Then S_ν and P_ν are exactly corresponding values independent of sampling errors, though the exactness of the correspondence may be affected by measuring errors in S.