A Statistical Distribution Function of Wide Applicability

By WALODDI WEIBULL,[1] STOCKHOLM, SWEDEN

This paper discusses the applicability of statistics to a wide field of problems. Examples of simple and complex distributions are given.

If a variable X is attributed to the individuals of a population, the distribution function (df) of X, denoted $F(x)$, may be defined as the number of all individuals having an $X \leq x$, divided by the total number of individuals. This function also gives the probability P of choosing at random an individual having a value of X equal to or less than x, and thus we have

$$P(X \leq x) = F(x) \ldots\ldots\ldots\ldots\ldots\ldots [1]$$

Any distribution function may be written in the form

$$F(x) = 1 - e^{-\varphi(x)} \ldots\ldots\ldots\ldots\ldots\ldots [2]$$

This seems to be a complication, but the advantage of this formal transformation depends on the relationship

$$(1 - P)^n = e^{-n\varphi(x)} \ldots\ldots\ldots\ldots\ldots\ldots [3]$$

The meri of this formula will be demonstrated on a simple problem.

Assume that we have a chain consisting of several links. If we have found, by testing, the probability of failure P at any load x applied to a "single" link, and if we want to find the probability of failure P_n of a chain consisting of n links, we have to base our deductions upon the proposition that the chain as a whole has failed, if any one of its parts has failed. Accordingly, the probability of nonfailure of the chain, $(1 - P_n)$, is equal to the probability of the simultaneous nonfailure of all the links. Thus we have $(1 - P_n) = (1 - P)^n$. If then the df of a single link takes the form Equation [2], we obtain

$$P_n = 1 - e^{-n\varphi(x)} \ldots\ldots\ldots\ldots\ldots\ldots [4]$$

Equation [4] gives the appropriate mathematical expression for the principle of the weakest link in the chain, or, more generally, for the size effect on failures in solids.

The same method of reasoning may be applied to the large group of problems, where the occurrence of an event in any part of an object may be said to have occurred in the object as a whole, e.g., the phenomena of yield limits, statical or dynamical strengths, electrical insulation breakdowns, life of electric bulbs, or even death of man, as the probability of surviving depends on the probability of not having died from many different causes.

Now we have to specify the function $\varphi(x)$. The only necessary general condition this function has to satisfy is to be a positive, nondecreasing function, vanishing at a value x_u, which is not of necessity equal to zero.

The most simple function satisfying this condition is

$$\frac{(x - x_u)^m}{x_0}$$

and thus we put

$$F(x) = 1 - e^{-\frac{(x - x_u)^m}{x_0}} \ldots\ldots\ldots\ldots\ldots\ldots [5]$$

The only merit of this df is to be found in the fact that it is the simplest mathematical expression of the appropriate form, Equation [2], which satisfies the necessary general conditions. Experience has shown that, in many cases, it fits the observations better than other known distribution functions.

The objection has been stated that this distribution function has no theoretical basis. But in so far as the author understands, there are—with very few exceptions—the same objections against all other df, applied to real populations from natural or biological fields, at least in so far as the theoretical basis has anything to do with the population in question. Furthermore, it is utterly hopeless to expect a theoretical basis for distribution functions of random variables such as strength properties of materials or of machine parts or particle sizes, the "particles" being fly ash, Cyrtoideae, or even adult males, born in the British Isles.

It is believed that in such cases the only practicable way of progressing is to choose a simple function, test it empirically, and stick to it as long as none better has been found. In accordance with this program the df Equation [5], has been applied not only to populations, for which it was originally intended, but also to populations from widely different fields, and, in many cases, with quite satisfactory results. The author has never been of the opinion that this function is always valid. On the contrary, he very much doubts the sense of speaking of the "correct" distribution function, just as there is no meaning in asking for the correct strength values of an SAE steel, depending as it does, not only on the material itself, but also upon the manufacturer and many other factors. In most cases, it is hoped that these factors will influence only the parameters. However, accidentally they may even affect the function itself.

The purpose of this paper has been to illustrate with a few examples the experience that the df, Equation [5], may sometimes render good service.

The number of examples has, by space, been limited to the following:

1 Yield strength of a Bofors steel
2 Size distribution of fly ash
3 Fiber strength of Indian cotton
4 Length of Cyrtoideae
5 Fatigue life of a St-37 steel

In the Appendix:

6 Statures for adult males, born in the British Isles
7 Breadth of beans of Phaseolus Vulgaris

[1] Professor at the Royal Institute of Technology; Scientific Adviser to AB Bofors; mailing address Bofors, Sweden.

The correctness of fit has been checked by applying the chi-square method

Of those populations, Nos. 1–3 are distributed in good agreement with the df Equation [5], whereas the four remaining populations have to be split up into two components, before such an agreement is obtained. The first type will be called a "simple" and the second type a "complex" distribution.

The fundamental question now arises, whether this splitting-up is a purely formal operation, or whether it might unveil some hidden real causes. It may be said that any distribution may be represented by a sum of a sufficiently great number of simple distributions, just as any periodical function may be developed in a Fourier series. However, if the number of the components be small and the number of observations sufficiently large, the likelihood of real causes seems to increase. In any case, it is very easy to produce real complex distributions by syntheses.

It seems obvious that the components of examples 4 and 5 are due to real causes. In examples 6 and 7 it is impossible to decide whether the division is a formal one or a real one, but the fact itself may be a valuable stimulus to a closer examination of the observed material.

The specific data for the examples follow.

Yield Strength of a Bofors Steel

The observed values are obtained as routine tests of a Bofors steel, the quality of which was chosen at random for purposes of demonstration only. Fig. 1 gives the curve and Table 1 the values, observed and calculated. The parameters are $x_u = 38.57$ kg/mm², $x_0 = 7.74$ kg/mm², $m = 2.934$. Without pooling, the degrees of freedom (d of f) are $9 - 3 = 6$. Then $\chi^2 = 5.40$ gives $P = 0.49$. The agreement is thus very satisfactory.

TABLE 1 YIELD STRENGTH OF A BOFORS STEEL

(x = yield strength in 1.275 kg/mm²)

	x	Expected values n	Observed values n	Normal distribution n
1	32	10	10	8
2	33	36	33	28
3	34	84	81	71
4	35	150	161	141
5	36	224	224	225
6	37	291	289	301
7	38	340	336	351
8	39	369	369	376
9	40	383	383	386
10	42	389	389	388

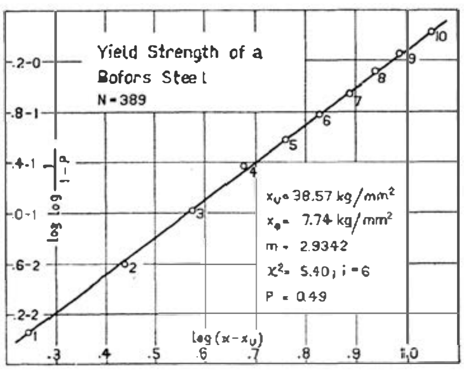

FIG. 1 YIELD STRENGTH OF A BOFORS STEEL

As a comparison, the values expected on the hypothesis of a normal distribution have been computed and are given in the last column of Table 1. If the classes 9–10 are pooled, the d of f are $8 - 2 = 6$. Then a $\chi^2 = 18.17$ gives a $P = 0.008$, which is not satisfactory at all.

Size Distribution of Fly Ash

The observed values are taken from J. M. Dalla Valle's work.[2] Fig. 2 gives the curve and Table 2 the values. The parameters are $x_u = 30\mu$, $x_0 = 128\,\mu$, $m = 2.288$. Without pooling, the d of f are $12 - 3 = 9$. Then $\chi^2 = 8.44$ gives a $P = 0.49$. If the classes 2–3 and 13–14 are pooled, the d of f are 7 and $\chi^2 = 8.44$ gives a $P = 0.29$.

TABLE 2 SIZE DISTRIBUTION OF FLY ASH

(x = particle diameter in 20 microns)

x	Expected values n	Observed values $\bar{n}$
2	3	3
3	14	14
4	34	34
5	62	56
6	92	85
7	122	126
8	150	150
9	172	175
10	188	188
11	199	197
12	205	202
13	209	208
14	211	211

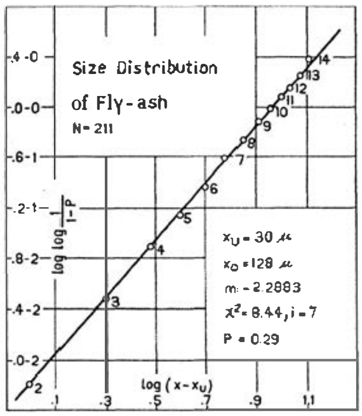

FIG. 2 SIZE DISTRIBUTION OF FLY ASH

Fiber Strength of Indian Cotton

The observed values are taken from R. S. Koshal and A. J. Turner.[3] Fig. 3 gives the curve, and Table 3 the values. The parameters are $x_u = 0.59$ gram, $x_0 = 3.73$ grams, $m = 1.456$. If the classes 14 to 16 are pooled, the d of f are $13 - 3 = 10$. Then $\chi^2 = 11.45$ gives a $P = 0.35$.

The authors[3] have pointed out that the most striking feature about the frequency curve is its asymmetry, showing a well-marked predominance of weak fibers. It was found—they say—that the observation curve would be well fitted by a theoretical curve of Pearson's Type 1, having the following equation

$$y = 599.3\left(1 + \frac{x}{18.777}\right)^{0.876716}\left(1 - \frac{x}{29.1947}\right)^{13.631284}$$

In this equation y represents the frequency of any strength x, expressed in grams.

The values computed from this not very handy equation are shown in the last column of Table 3. The d of f are $13 - 5 = 8$ (as there are 5 parameters). Then $\chi^2 = 14.43$ gives a $P = 0.07$.

[2] "Micromeritics," by J. M. Dalla Valle, Pitman Publishing Corporation, New York, N. Y., 1948, p. 57, Fig. 2.

[3] "Studies in the Sampling of Cotton for the Determination of Fiber Properties," by R. S. Koshal and A. J. Turner, Journal of the Textile Institute Transactions, vol. 21, 1930, pp. 325–370.

TABLE 3 FIBER STRENGTH OF INDIAN COTTON

(x = tensile strength in grams)

x	Expected values n	Observed values $\bar{n}$	Pearson Type 1 n
1	118	177	127
2	646	667	659
3	1232	1219	1255
4	1751	1729	1777
5	2161	2153	2184
6	2461	2465	2480
7	2667	2664	2683
8	2802	2813	2816
9	2886	2887	2899
10	2937	2933	2949
11	2966	2962	2978
12	2982	2985	2994
13	2991	2991	3003
14	2996	2995	3007
15	2999	2999	3009
16	3000	3000	3010

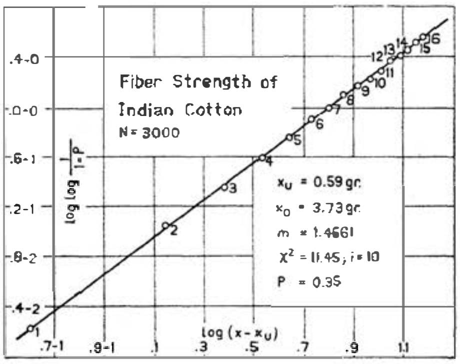

Fig. 3 Fiber Strength of Indian Cotton

TABLE 4 LENGTH OF CYRTOIDEAE

(x = length in microns)

	x	Expected values n_1	n_2	n_{1+2}	Observed values $\bar{n}_{1+2}$
1	10	1	..	1	0
2	20	5	..	5	5
3	30	13	..	13	12
4	40	23	..	23	24
5	50	35	..	35	38
6	60	47	..	47	45
7	70	58	..	58	58
8	80	67	..	67	69
9	90	74	..	74	70
10	100	79	..	79	80
11	110	82	..	82	82
12	120	85	..	85	84
13	130	86	..	86	86
14	140	86	4	90	90
15	150	86	7	93	93
16	160	86	9	95	95
17	170	86	11	97	97
18	180	86	12	98	98
19	190	86	13	99	99
20	200	86	14	100	100

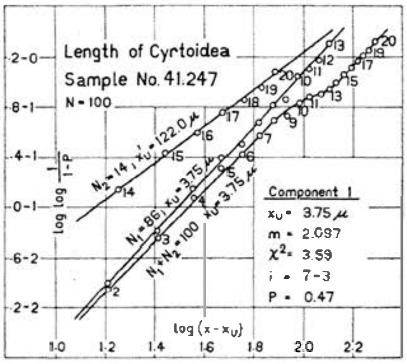

Fig. 4 Length of Cyrtoideae

In spite of the greater number of parameters, the fit of this distribution function is not as close as that of the first one.

Length of Cyrtoideae

This is the first example of a complex distribution. The observed values have been obtained from investigations by Dr. Gustaf Arrhenius, on submarine cores from the Swedish Deep-Sea Expedition With *Albatross.* The measurements were made by Dr. C. Jungk, taking samples from each 10 cm of the core, corresponding to an age interval of about 100,000 years. Some fifty populations have been analyzed statistically. About 20 per cent of the populations showed a simple distribution, as exemplifi in a previous paper.[4] The remaining samples showed a two-component distribution.

Fig. 4 gives the curves and Table 4 the values of one of the complex populations. The undivided sample gives the curve marked $N_1 + N_2$. It is easy to see that the distribution is a complex one, and that it is necessary to split up the population in two parts. By trial it was found that 86 of the individuals belonged to component No. 1, and 14 to component No. 2.

The parameters are: Component No. 1: $x_u = 3.75\ \mu$, $x_0 = 63.2\ \mu$, $m = 2.097$. Pooling the classes 2–3, 9–10, and 11–13 gives $\chi^2 = 3.59$. The d of f are $7 - 3 = 4$, and $P = 0.47$.

Component No. 2: $x_u = 122.0\ \mu$, $x_0 = 124.1\ \mu$, $m = 1.479$. The number of individuals is too small for the χ^2-test.

Fatigue Life of an St-37 Steel

The observed values are taken from Müller-Stock.[5] The frequency curve in Fig. 5[5] gives no impression of a complex distribution, which, on the other hand, may easily be seen when

[4] "A Statistical Analysis of the Size of Cyrtoideae in Albatross Cores From the East Pacific Ocean," by W. Weibull, *Nature*, vol. 164, 1949, p. 1047.

[5] "Der Einfluss dauernd und unterbrochen wirkender, schwingender Überbeanspruchung auf die Entwicklung des Dauerbruchs," by H. Müller-Stock, *Mitteilungen Kohle- und Eisenforschung*, (March, 1938), by measurements from his Fig. 13; reproduced in Fig. 5 of this paper.

TABLE 5 FATIGUE LIFE OF ST-37

(Rotating—beam test at ±32 kg/mm²)

	N 10^3	Expected values n_1	n_2	n_{1+2}	Observed values $\bar{n}_{1+2}$
1	17.5	4.6	...	4.6	4.6
2	22.5	47.4	...	47.4	47.4
3	27.5	125.1	...	125.1	125.1
4	32.5	161.2	8.1	169.3	169.2
5	37.5	164.9	28.0	192.9	192.7
6	42.5	165.0	41.9	206.9	207.3
7	47.5	165.0	51.0	216.0	215.9
8	52.5	165.0	57.0	222.2	222.2
9	57.5	165.0	61.0	226.0	225.9
10	62.5	165.0	63.7	228.7	228.7
11	67.5	165.0	65.6	230.6	230.5
12	72.5	165.0	66.9	231.9	231.9
13	77.5	165.0	67.9	232.9	232.9
14	82.5	165.0	68.6	233.6	233.5
15	87.5	165.0	69.1	234.1	233.9
16	92.5	165.0	70.0	235.0	235.0

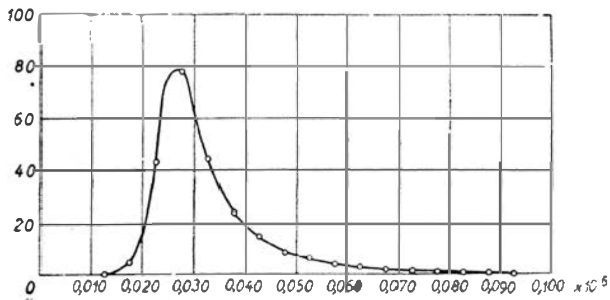

Fig. 5 Frequency Curve of Fatigue Life of St-37 Steel
(Number of specimens versus number of stress cycles.)

using the plottings in Fig. 6. The parameters are: Component No. 1: $x_u = 4.032$, $m = 5.956$; Component No. 2: $x_u = 4.484$, $m = 1.215$. Table 5 shows the close agreement between the observed and the calculated values.

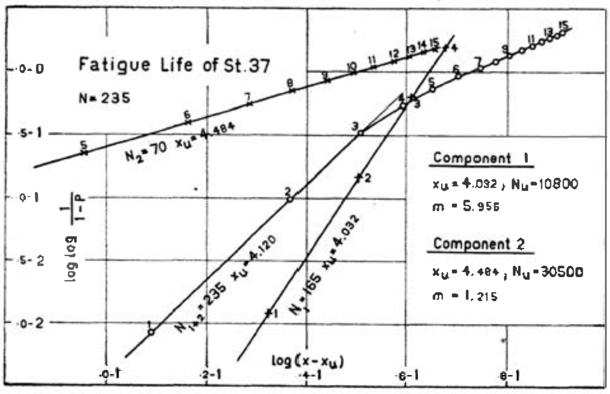

FIG. 6 FATIGUE LIFE OF ST-37 STEEL

It may be pointed out that the frequency curve in Fig. 5 seems to be the result of a smoothing operation on the cumulative frequency curve. Accordingly, the sampling errors of the observed values in Table 5 have been eliminated almost entirely (without affecting the function), which explains the really too good representation of the observed values.

The real causes of this splitting up in two components may be found by examining the frequency curve of the yield strength of the same material, Fig. 7. It is easy to see that the material, probably not being killed, is composed of two different kinds. If we suppose that all the specimens with a yield strength of less than 25 kg/mm² belong to Component No. 1, we obtain 14 specimens out of 20, making 70 per cent. Exactly the same proportion has been found by the statistical analysis, as $165/235 = 70$ per cent.

The reason why this partition is so easily seen in Fig. 7 and not at all in Fig. 5, depends, of course, upon the much larger scatter in fatigue life than in yield strength.

Appendix

The foregoing statistical methods have been applied to many problems outside the field of applied mechanics. It may perhaps be of interest to have examples of this kind, and for this reason, the following two are given with the tables only:

STATURES FOR ADULT MALES BORN IN THE BRITISH ISLES

The observed values are taken from Yule and Kendall.[6] This distribution is classified by the authors as being approximately of the symmetrical type, and there is no mention of its being composed of two parts.

By trial it was found that the population had to be split up into two parts: $N_1 = 6200$ and $N_2 = 2385$. The parameters are as follows:

Component No. 1: $x_u = 50.0$ in., $x_0 = 16.2$ in., $m = 9.6865$. If the classes 1–2 and 14–15 are pooled, we get $\chi^2 = 11.80$. As we have 7 parameters altogether, (one of them is the partition of the population), we take $3^1/_2$ to each of the components. The d of f are then $12 - 3^1/_2 = 8^1/_2$, which gives a $P = 0.20$.

Component No. 2: $x_u = 67.4$ in., $x_0 = 2.3$ in., $m = 1.4662$.

[6] "An Introduction to the Theory of Statistics," by G. U. Yule and M. G. Kendall, eleventh edition, J. B. Lippincott Company, Philadelphia, Pa., 1937, pp. 94 and 111.

TABLE 6 STATURES FOR ADULT MALES BORN IN THE BRITISH ISLES
(x = height in inches)

	Expected values			Observed values
x	n_1	n_2	n_{1+2}	$\bar{n}_{1+2}$
57	2	...	2	2
58	6	...	6	6
59	20	...	20	20
60	56	...	56	61
61	143	...	143	144
62	333	...	333	313
63	702	...	702	707
64	1350	...	1350	1376
65	2351	...	2351	2366
66	3641	...	3641	3589
67	4917	...	4917	4918
68	5787	339	6126	6148
69	6134	1079	7213	7211
70	6197	1671	7868	7857
71	6200	2039	8239	8249
72	6200	2233	8433	8451
73	6200	2324	8524	8530
74	6200	2363	8563	8562
75	6200	2378	8578	8578
76	6200	2383	8583	8583
77	6200	2385	8585	8585

TABLE 7 PHASEOLUS VULGARIS
(Breadth of beans = $0.25\,x + 6.70$ mm)

	Expected values			Observed values
x	n_1	n_2	n_{1+2}	$\bar{n}_{1+2}$
1	32	...	32	32
2	130	...	130	135
3	400	...	400	374
4	1011	...	1011	998
5	2145	...	2145	2185
6	3832	...	3832	3835
7	5718	...	5718	5718
8	7140	486	7626	7648
9	7761	1525	9286	9286
10	7890	2510	10400	10416
11	7900	3229	11129	11153
12	7900	3671	11571	11580
13	7900	3908	11808	11801
14	7900	4022	11922	11911
15	7900	4071	11971	11968
16	7900	4091	11991	11992
17	7900	4098	11998	11998
18	7900	4100	12000	12000

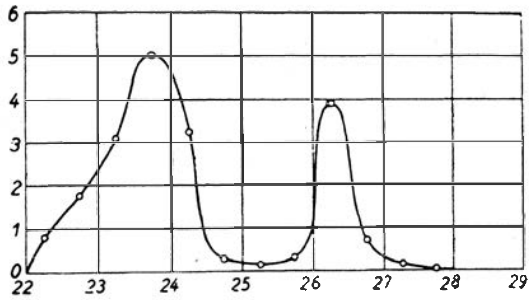

FIG. 7 FREQUENCY CURVE OF YIELD STRENGTH OF ST-37 STEEL
(Number of specimens versus yield strength in kg/mm².)

If the classes 20–21 are pooled, $\chi^2 = 5.11$. The d of f are $8 - 3^1/_2 = 4^1/_2$, which gives a $P = 0.35$.

BREADTH OF BEANS OF PHASEOLUS VULGARIS

This is a classical example, quoted from Charlier[7] to exemplify the expansion in Edgeworth's series.

If the population is divided into two parts, $N_1 = 7900$ and $N_2 = 4100$, each of them may be very well fitted to a simple distribution function with the following parameters:

Component No. 1: $x_u = -3.0$ (= 5.95 mm), $m = 6.2805$. Without pooling we have the $\chi^2 = 7.70$, and the d of f $10 - 3^1/_2 = 6^1/_2$ giving a $P = 0.29$.

Component No. 2: $x_u = +7.2$ (= 8.50 mm), $m = 1.6098$.

[7] "Die Grundlagen der Mathematischen Statistik," by C. V. L. Charlier, second edition, 1920, p. 73, quoted by Harald Cramér in his book: "Mathematical Methods of Statistics," Princeton University Press, Princeton, N. J., 1945, p. 440.

If the classes 17–18 are pooled, the value of $\chi^2 = 4.50$, and the d of f $9 - 3^1/_2 = 5^1/_2$ give a $P = 0.56$.

It may be of interest to compare this result with those of Charlier and Cramér.

Charlier says that, at the first look, the agreement with the normal distribution seems very satisfactory, but that a closer examination shows a small negative skewness and a small positive kurtosis.

Cramér has calculated the values of χ^2 on the hypotheses of normal distribution and asymptotic expansions from it. The result was as follows:

Normal distribution	$\chi^2 = 196.5$ d of f 13	$P < 0.001$
First approximation	$\chi^2 = 34.3$ d of f 12	$P < 0.001$
Second approximation	$\chi^2 = 14.9$ d of f 11	$P = 0.19$

The agreement is satisfactory in the third case only, requiring four terms of the series. This operation is certainly of a purely formal character.