Fire Research Note
No. 837

SOME POSSIBLE APPLICATIONS OF THE THEORY OF EXTREME VALUES FOR THE ANALYSIS OF FIRE LOSS DATA

by

G. RAMACHANDRAN

September 1970

FIRE
RESEARCH
STATION

Fire Research Station,
Borehamwood,
Herts.
Tel. 01-953-6177

F.R. Note No.837.

SOME POSSIBLE APPLICATIONS OF THE THEORY OF EXTREME VALUES FOR THE ANALYSIS OF FIRE LOSS DATA

by

G. Ramachandran

SUMMARY

This paper discusses the possibility of applying the statistical theory of extreme values to data on monetary losses due to large fires in buildings. The theory is surveyed in order to impart the necessary background picture. With the logarithm of loss as the variate, an initial distribution of the exponential type is assumed. Hence use of the first asymptotic distribution of largest values is illustrated. Extreme order statistics other than the largest are also discussed. Uses of these statistics are briefly outlined. Suggestions for further research are also made.

KEY WORDS: Large fires, loss, fire statistics.

MINISTRY OF TECHNOLOGY AND FIRE OFFICES' COMMITTEE
JOINT FIRE RESEARCH ORGANIZATION

SOME POSSIBLE APPLICATIONS OF THE THEORY OF EXTREME VALUES FOR THE ANALYSIS OF FIRE LOSS DATA

by

G. Ramachandran

INTRODUCTION

Large fire losses are of economic importance as their total cost is over 60 per cent of the total cost in all fires. However, they represent only one per cent (overall) of the total number of fires. But these values are at the upper tail of the fire loss distribution and hence, in the statistical sense, have to be regarded as extreme (largest) values.

The oldest problems connected with extreme values arise from floods. An inundation happens when water flows where it ought not to flow. Floods destroy life and property. Hence engineers are confronted with the need to forecast floods in order to tackle flood control problems. Floods are defined as the largest daily discharges of a stream during a year. The smallest daily discharges are droughts, a study of which is of essential value in treating problems arising from stream pollution, sewage disposal and water supply.

Although the theory is comparatively recent, it has attracted the attention of scientists working in different fields. Comparable meteorological phenomena are the extreme pressures, temperatures, rainfalls etc. Other examples occur in fracturing of metals, textiles and other materials (breaking strength) under applied force, which are problems arising from materials testing and quality control. Gusts are the largest strengths of wind and gust loads play an important role in aeronautics. An example in the actuarial field is 'oldest ages', i.e. the longest duration of life. The extreme span of human life is linked to the study of extremes. Recently Epstein[1] considered bacterial extinction time as an extreme value phenomenon. Once it is recognised that these phenomena are of a statistical nature, it is possible to obtain solutions to some of the problems connected with them by applying the asymptotic theory of extreme values. The most comprehensive treatise on the subject is the book by Gumbel[2]. This book also contains a discussion of applications and a bibliography of papers on the subject. The latest work on the subject appears to be the thesis of Harris[3].

Fire in a particular building is a rare event. However, in a group of identical risks fires are frequent. If a fire spreads beyond the material or room of origin it might cause widespread destruction. In the monetary sense a loss of £10,000 or more may be of economic importance. Such a phenomenon is similar to floods in rivers. The largest loss in a year may be regarded as equivalent to an annual 'flood'. The statistical problems being similar to those of flood control, the feasibility of applying the theory of extreme values to large fire losses is examined in this paper.

THE VARIABLE

The variable (x) considered is the direct monetary loss to a building and its contents due to fire. The values available are preliminary estimates made by the British Insurance Association. It is not possible to include consequential losses (due to loss of production etc) as estimates for such losses are not available.

The operational variable is the logarithm of x and is designated by z. If a minimum loss of one pound is assumed, the variable z is non negative. (One could always consider the variable $x^1 = x + 1$ which does not alter the picture materially especially when dealing with large values of x).

Theoretically the maximum loss that could occur in a building is the total value at risk in that building. However, there is a finite non-vanishing probability that fire could spread beyond the building of origin. Also, in a group of buildings, the value of the costliest building could be high. In these circumstances the variable z could assume values extending to infinity.

Estimates of fire losses rounded to the nearest ten pounds (or even to the nearest thousand pounds in the case of large fires) would suffice for practical purposes. Theoretically, however, the damage could assume all values in terms of infinitesimal units. Hence the variate x or z is conceptually of a continuous nature.

THE INITIAL DISTRIBUTION

In order to apply extreme value theory it is essential to identify the nature of the initial or parent distribution. This is the distribution of the probabilities with which the cumulative monetary damage reaches various amounts in a single fire. Probability distributions which have been suggested are the pareto and the logarithmic normal. In a recent paper[4] the author discussed the possible relevance of distributions with an increasing failure rate and of mixtures of such distributions. For the variable x he suggested a distribution of the form

$$\phi(x) = \text{Prob}[X > x] = x^{-(\alpha + \beta \log x)}, \quad (\alpha, \beta > 0) \qquad \ldots\ldots (1)$$

the density function being

$$f(x) = (\alpha + 2\beta \log x)\, x^{-(\alpha + \beta \log x + 1)} \qquad \ldots\ldots (2)$$

The density function of z is of the form

$$f(z) = (\alpha' + 2\beta' z)\, e^{-z(\alpha' + \beta' z)}, \quad (\alpha', \beta' > 0) \qquad \ldots\ldots (3)$$

if z = k log x.

The function

$$\mu(z) = \frac{f(z)}{\phi(z)} = (\alpha' + 2\beta' z) \qquad \ldots\ldots (4)$$

(where $-\phi'(z) = f(z)$) is known as failure rate or hazard function. This is the conditional probability of extinction in the interval $(z, z+dz)$ given that the fire has survived till the value z has been reached. In the early stages of growth, fire has a greater tendency to spread so that $\mu(z)$ is decreasing. However, for large values of z, $\mu(z)$ is likely to be increasing due to the commencement of fire fighting and the exhaustion of natural forces contributing to fire spread. The structure of expression (4) could be generalised so that $\mu(z)$ is a polynomial in z or an exponential function.

Application of L'Hopital's rule gives the critical quotient $Q(z)$ as

$$Q(z) = \frac{\mu(z)}{-f'(z)/f(z)}$$

$$= \left[1 - \frac{\mu'(z)}{\mu^2(z)}\right]^{-1} = \left[1 - \frac{2\beta'}{(\alpha' + 2\beta' z)^2}\right]^{-1}$$

$$= 1 + \mathcal{E}(z)$$

with $\lim_{z \to \infty} \mathcal{E}(z) = 0$. Hence according to Gumbel's terminology $f(z)$ is a distribution of the exponential type and belongs to the 'first' class. The probability density converges towards zero for $z \to \infty$ faster than an exponential distribution. Therefore all moments exist for this distribution (expression (3)). For the application of extreme value theory it would suffice to note that the probability distribution of z is of exponential type; the values of the parameters α' and β' are not required. The well known distributions logistic, gamma, chisquare, normal and log normal all belong to this type, but pareto does not. The distribution (3) is for all risks taken together. However it is reasonable to assume that the distribution for any industry is of exponential type though with different values of the parameters.

ORDER STATISTICS

Consider the variable z. The cumulative distribution function $F(z)$ denotes the probability that the loss is less than or equal to z. It is given by $F(z) = 1 - \varphi(z)$ where it is assumed that

$$\varphi(z) = e^{-z(\alpha' + \beta' z)} \qquad \text{...... (7)}$$

the corresponding density function $f(z)$ being expression (3). The function $\varphi(z)$ is the survivor function giving the probability that the loss is greater than z. In reliability theory $\varphi(z)$ is known as the reliability function.

Now consider independent random samples $z_1, z_2, z_3, \ldots\ldots\ z_n$ drawn from the population with distribution function $F(z)$. Rearranging them in increasing order of their magnitudes we may write

$$z_{(1)} < z_{(2)} < z_{(3)} \ldots\ldots < z_{(n)} \qquad \text{...... (8)}$$

These are known as order statistics where $z_{(1)}$ is the smallest and $z_{(n)}$ the largest. A collection of useful papers on order statistics and their applications is contained in the book edited by Sarhan and Greenberg[5].

The m^{th} ($m = 1, 2, \ldots\ldots n$) among n observations taken in increasing order, viz. $z_{(m)}$ has the probability density function $f_n(z_m)$ which depends upon the initial density $f(z)$, the sample size n and the order m. The value of $f_n(z_m)$ is given by

$$f_n(z_m) = \frac{n!}{(n-m)!\,(m-1)!} \{F(z)\}^{m-1} \{\varphi(z)\}^{n-m} f(z) \qquad \text{...... (9)}$$

The generating function of the m^{th} order statistics can be written as an integral which will facilitate the calculation of mean, variance and higher moments of these statistics. However, for most of the initial distributions the integration is difficult. The book by Sarhan and Greenberg contains tables of expected values and standard deviations of order statistics from populations which are normal, exponential, gamma and rectangular. The tables give the values for a small number n of sample sizes. Calculation of moments of order statistics for the distribution $f_n(z_m)$ is equally difficult.

The median $\breve{z}_m$ of the m^{th} value is the solution of

$$F(\breve{z}_m) = \lambda(m, n)$$

where λ is given by

$$\frac{\int_0^\lambda F^{m-1}(1-F)^{n-m}\,dF}{\int_0^1 F^{m-1}(1-F)^{n-m}\,dF} = \frac{1}{2} \qquad \ldots\ldots (10)$$

and can be obtained from Pearson's tables of the Incomplete Beta Function. Evidently the median depends upon the initial distribution. The mode $\hat{z}_m$ is obtained from the logarithm of the distribution $f_n(z_m)$ as the solution of

$$\frac{m-1}{F}\cdot f - \frac{n-m}{1-F}\cdot f + \frac{f'}{f} = 0 \qquad \ldots\ldots (11)$$

DISTRIBUTION OF EXTREMES

In the previous section we have defined $z_{(m)}$ $(m = 1, 2, \ldots\ldots n)$ as order statistics given by an arrangement in increasing magnitudes of n independent observations. The first $z_{(1)}$ is the smallest and the last $z_{(n)}$ the largest one. Both are called extremes. In this paper we are concerned only with the largest $z_{(n)}$. If N such samples each of size n are drawn a distribution of largest values is obtained). The probability $\Phi_n(z_n)$ for z_n to be the largest value is

$$\Phi_n(z_n) = \{F(z_n)\}^n \qquad \ldots\ldots (12)$$

with the derivative

$$f_n(z_n) = n\{F(z_n)\}^{n-1} f(z_n) \qquad \ldots\ldots (13)$$

as the density function of the largest value. Expression (13) which could also be obtained from (9) by letting $m = n$, is known as the exact distribution of the extreme. The expected value and variance of z_n for a few initial distributions have already been studied and tabulated for different sample sizes n.

If n is large, expression (12) tends to an asymptotic probability $\Phi(z)$ of the largest value $z = z_n$. For intial distributions of the exponential type, as proved by Gumbel and others, $\Phi(z)$ is of the form

$$\Phi(z) = e^{-e^{-y}} \qquad \ldots\ldots (14)$$

where

$$y = a_n(z_n - b_n) \qquad \ldots\ldots (15)$$

is called the reduced variate. In (15), b_n is defined as the solution of

$$F(b_n) = 1 - \frac{1}{n} \qquad \ldots\ldots (16)$$

and is referred to as the 'characteristic largestvalue'. In n observations the expected number of values equal to or larger than b_n is unity. For $n = 2$ the characteristic largest value is equal to the initial median, for $n = 4$ to the upper quartile, for $n = 10$ to the upper decile of the initial distribution and so forth. For large samples the calculation may be simplified if an asymptotic expression for the probability $F(z)$ exists, as assumed in our case, to be given by

$$F(z) = 1 - e^{-z(\alpha' + \beta' z)} \qquad \ldots\ldots (17)$$

The value of b_n is obtained by solving

$$\begin{aligned} F(b_n) &= 1 - \frac{1}{n} \\ &= 1 - e^{-b_n(\alpha' + \beta' b_n)} \end{aligned} \qquad \ldots\ldots (18)$$

extracting the positive root of b_n.

The parameter a_n in (15) is defined as

$$\begin{aligned} a_n &= n \cdot f(b_n) \\ &= \frac{f(b_n)}{1 - F(b_n)} = \frac{f(b_n)}{\Phi(b_n)} \\ &= \mu(b_n) \text{ from (4)} \end{aligned} \qquad \ldots\ldots (19)$$

As already mentioned $\mu(z)$ is the failure or hazard rate. The function $\mu(z)$ is also called 'the force of mortality' in actuarial statistics and 'intensity function' in extreme value theory. The parameter a_n is the value of the intensity function at the characteristic largest value b_n.

APPLICATION OF THE LARGEST VALUE

Extreme value theory is concerned with the analysis of largest (or smallest) values from samples of a large number n of observations. If a month is chosen as the period of observation the sample size n would be small as only a few fires happen in a month in a particular risk category. Hence a year is preferable. We also require a large number N of samples of n values from which we can obtain the N extreme values. In other words we require data on extreme values (largest losses) for a number of years. Loss figures are available for a number of years for fires costing £20,000 or more.

Another assumption in the theory is that the sample size n is constant. We might perhaps assume constancy in cases where the variation in the value of n is negligible. But even this assumption is difficult to make in our case as the frequency of fires increases considerably over a period of time. Due to this, the analysis may later require slight modification. But this aspect will be examined in a subsequent study.

In the table below, the number of fires in buildings engaged in the manufacture of textiles during the period 1947 to 1967 is shown under col.(2). In col.(3) the largest of the losses that occurred during each year is given. The loss figures corrected for inflation (with 1947 as the base year) are given in col.(4). Such a correction is necessary as the theory requires that the influence of time on the parameters has been taken into account or eliminated.

Table 1

Number of fires and the largest losses in the textile industry

Year	Number of fires	Observed largest loss (x_n) (£'000)	Corrected largest loss (x'_n) (£'000)
(1)	(2)	(3)	(4)
1947	465	460	460
1948	478	350	324
1949	512	210	189
1950	574	350	307
1951	728	550	440
1952	568	1000	735
1953	725	460	329
1954	662	150	105
1955	740	320	215
1956	716	250	160
1957	645	400	247
1958	560	340	205
1959	872	570	339
1960	760	269	159
1961	696	310	177
1962	724	532	291
1963	790	493	265
1964	998	392	204
1965	964	1912	951
1966	1050	445	212
1967	982	1033	470

The first step in the analysis is to test the goodness of fit of the extreme value distribution given by (14) to the logarithms of the corrected losses given in Table 1. In order to do this we have to estimate the parameters a_n and b_n so that we could form the reduced values y using expression (15). This is possible if the initial distribution and its parameters are known so that the parametric values a_n and b_n could be obtained from their definitions in expressions (19) and (16). This is not possible at present as figures are not available for fires costing less than £10,000.

However, since we have reason to believe that the initial distribution is of the exponential type, we may estimate the parameters a_n and b_n from the N (= 21) observed largest values alone. Gumbel has given a method based on order statistics. Kimball has done considerable work on the estimation of these parameters by the method of maximum likelihood. Gumbel has also suggested a general method which is discussed below.

The parameters a_n and b_n acquire a special significance if only the observed extremes are used to estimate them. The parameter b_n becomes the modal largest value and $1/a_n$ which is proportional to the standard deviation of extremes becomes asymptotically the rate of increase of the most probable largest value with the natural logarithm of the number of samples N.

Now consider the N observed (corrected) extreme values $z_{(m)}$ ($m = 1, 2, \ldots\ldots N$) ordered in increasing magnitude. The empirical value of the cumulative relative frequency of $z_{(m)}$ is

$$\Phi(z_{(m)}) \doteq \frac{m}{N+1} \qquad \ldots\ldots (20)$$

The reason for using $N + 1$ instead of N as the denominator in (20) has been explained by Gumbel. We may observe from (14) that

$$y = -\log_e\left(-\log_e \frac{m}{N+1}\right) \qquad \ldots\ldots (21)$$

The values of y for different cumulative relative frequencies $\frac{m}{N+1}$ are given in Table 2 of 'Probability Tables for the analysis of extreme value data' published by the National Bureau of Standards[6]. These values of the reduced variate y are reproduced in col.5 of Table 2 together with the rank, the corresponding observations $z_{(m)}$ and the cumulative relative frequencies.

Table 2

Largest values and reduced variates

Rank (m)	Largest loss (corrected) £'000 (x')	Transformed value ($z = \log_e x'$)	Cumulative relative frequency ($m/N+1$)	Reduced variate (y)
(1)	(2)	(3)	(4)	(5)
1	105	4.654	0.0455	-1.125
2	159	5.069	0.0909	-0.874
3	160	5.075	0.1364	-0.691
4	177	5.176	0.1818	-0.533
5	189	5.242	0.2273	-0.394
6	204	5.318	0.2727	-0.261
7	205	5.323	0.3182	-0.136
8	212	5.357	0.3636	-0.011
9	215	5.371	0.4091	0.112
10	247	5.509	0.4545	0.239
11	265	5.580	0.5000	0.367
12	291	5.673	0.5455	0.502
13	307	5.727	0.5909	0.643
14	324	5.781	0.6364	0.793
15	329	5.796	0.6818	0.960
16	339	5.826	0.7273	1.143
17	440	6.087	0.7727	1.357
18	460	6.131	0.8182	1.605
19	470	6.153	0.8636	1.923
20	735	6.600	0.9091	2.350
21	951	6.858	0.9545	3.078

If the theory is true the observed values z should lie scattered about the straight line

$$z = b_n + \left(\frac{1}{a_n}\right) y \qquad \ldots\ldots (22)$$

which is merely another form of expression (15). The correlation coefficient between z and y is as high as 0.99. The method of least squares gave the best theoretical straight line as

$$z = 5.384 + 0.476 y \qquad \ldots\ldots (23)$$

Hence $b_n = 5.384$ or £219,000 is the modal (annual) largest loss at 1947 prices. The value of $^1/a_n$ is 0.476. It will be a useful study to compare the estimates obtained by the least square method with those by other methods like maximum likelihood.

If the observations follow each other in time, if the sample size is constant, and further if the distance between consecutive large values is approximately constant (1 year in our case) then the function

$$T(z) = \left[1 - \Phi(z)\right]^{-1} \qquad \ldots\ldots (24)$$

called 'the return period' gives the average number of observations necessary to obtain the value equal to or larger than z. For large values of z the return periods converge towards e^y. The straight line (23) may be used to determine the most probable largest loss corresponding to a given return period.

For example, for T = 26, expression (24) gives $\Phi(z) = \frac{25}{26} = 0.96$ and we have the theoretical value y

$$y = -\log_e(-\log_e 0.96) = 3.199$$

from which using (23) we obtain the corresponding value of z as 6.907. This is the size of the largest loss in logarithmic terms or one million pounds (at 1947 prices) we may expect in the textile industry within the next (26 - 21) or 5 years (before 1973). This estimate is based on the current trend. Only drastic changes in fire protection and fire fighting methods could alter this picture.

Inversely we may estimate the return period of a largest loss of given size against which we require fire protection. For example, if we want to protect a building (engaged in the textile trade) against a loss of £1,500,000 (at 1947 prices), z being 7.314 the return period could be estimated as 59 years, so that this degree of protection would suffice for the next (59-21) or 38 years.

The loss mentioned in the first example may serve as a guide for the allocation of financial resources for any activity requiring planning for a relatively short period. In the same way, the loss mentioned in the second example may be used for defining a level of justifiable expenditure on fire protection measures in a factory building to be constructed or additional measures required for a building already in existence. A word of caution is necessary before the figures quoted above are used. The figures are applicable only to buildings with values at risk at least as high as the figures themselves. If adjustments were made for the regression of loss on value density and other factors, the appropriate loss figures could be estimated for a building of particular characteristics.

THE m^{th} EXTREME

It is true that one has to be prepared for the worst viz. the largest loss. But the cost involved in providing such a high degree of protection would be enormous. The question therefore arises as to whether it would not be better to protect against the loss next to the largest. Extending this argument further, we may like to think in terms of a few top large values. Let the observations be arranged in descending order z_m $(m = 1, \dots r)$ where $m = 1$ is the largest. These observations correspond to the observations $z_{(n)}, z_{(n-1)}, \dots z_{(n-r+1)}$ in the ascending order arrangement in expression (8). The exact distribution of the m^{th} extreme from top [or $(n - m + 1)^{th}$ from bottom], from (9) is

$$f_n(z_m) = \frac{n!}{(m-1)!\,(n-m)!}\{F(z)\}^{n-m}\{\varphi(z)\}^{m-1} f(z) \qquad \dots\dots (25)$$

If we define two parameters a_m and b_m as the solutions of

$$\left.\begin{aligned} F_n(b_m) &= 1 - \frac{m}{n} \text{ and} \\ a_m &= \frac{n}{m} f_n(b_m) \end{aligned}\right\} \qquad \dots\dots (26)$$

the asymptotic density function $\chi_m(z)$ of the m^{th} extreme value $z = z_m$ has been proved to be

$$\chi_m(z) = \frac{m^m}{(m-1)!} a_m e^{-m y_m - m e^{-y_m}} \qquad \dots\dots (27)$$

where the reduced m^{th} largest value y_m is defined as

$$y_m = a_m (z_m - b_m) \qquad \dots\dots (28)$$

For the largest value $m = 1$, $\chi_1(z)$ is given by

$$\chi_1(z) = a_1 e^{-y_1 - e^{-y_1}} \qquad \ldots\ldots (29)$$

which is the derivative of the asymptotic distribution function in expression (14). The distribution function (probability) corresponding to (27) is

$$\Phi_m(z) = \int_{-\infty}^{z} \chi_m(z)\, dz \qquad \ldots\ldots (30)$$

The asymptotic probability in expression (30) could be rewritten as

$$\Phi_m(y_m) = \int_{-\infty}^{y_m} (m e^{-y})^{m-1} \exp(-m e^{-y})\, m e^{-y} dy \,/\, \Gamma(m) \qquad \ldots\ldots (31)$$

If we introduce

$$m e^{-y_m} = u$$

the probability leads to the incomplete Gamma function

$$\Phi_m(y_m) = \int_{m e^{-y_m}}^{\infty} u^{m-1} e^{-u} du \,/\, \Gamma(m) \qquad \ldots\ldots (32)$$

The successive probabilities for the m^{th} extreme can be expressed by the probability of the largest value. Integration of (32) by parts leads, after reversing the order of summation to

$$\Phi_m(y_m) = \Phi_1^m(y) \sum_{\nu=0}^{m-1} \frac{m^{\nu} e^{-\nu y_m}}{\nu!} \qquad \ldots\ldots (33)$$

Reduced values y_m obtained from (32) are given in Table 4 of the Probability Tables of the National Bureau of Standards[6] for a few selected probability points and for the top 15 extremes. But the spacing of the probabilities is wide and varying making interpolation for the required values a difficult task. Once we obtain the values of the reduced variate y_m for the probability points under col.(4) of Table 2, we could fit straight lines similar to (23) for the largest value. Thus we could estimate the parameters a_m and b_m for, say, the top ten values. If these straight lines fit the data well, we could combine the information on the top ten largest values and improve our decisions.

DISCUSSION

Solutions to problems in fire protection economics depend upon the structure of the probability distribution of fire loss. Available data suggest a distribution of the form in expression (2) or its counterpart in expression (3). If sufficient data become available it will be possible to estimate the values of the parameters α and β for each group of buildings with similar fire risks. These parametric values could serve as indices of fire risks for purposes of planning fire protection and fire fighting strategies on a national scale.

The values will also be useful in forecasting the expected total loss in a particular area during a given planning period. Assume, for example, that the expected number of fires could be estimated for a period in the given area. If the losses (on a log scale) are $z_{(m)}$ $(m = 1, 2, \ldots n)$, using expression (9) it is possible to estimate the expected value and standard deviation of the sum $S_n = \sum_{m=1}^{n} Z_m$. It may be preferable to obtain the parametric values for an individual area based on past experience.

At present, loss data are available only for fires costing £10,000 or more, though for a minority group, viz. sprinklered buildings, the information is available for the lower ranges.

However it is an accepted fact that (in repeated sampling) large or extreme values have their own distributions with three types of asymptotic forms depending on three types of initial distributions. The asymptotic distribution, for initial distributions of the exponential type (as assumed in our case), is known as the first asymptotic distribution of largest (smallest) values. The asymptotic distributions of extreme values are highly skewed and non-normal. In his book Gumbel has published a table giving the expected value, standard deviation and coefficients of skewness and kurtosis for the top ten reduced extremes. (It is possible to extend this table to the top 30 or 40 largest values).

It is apparent that the averages of extremes are also non-normal. Hence the usual tests of significance based on the normal theory are not applicable to such averages. The same difficulty arises in the case of analysis of variance or regression problems with extreme value data perhaps introducing heteroscedasticity in the error variance. Statistical inference based on extreme values thus requires the application of the theory of non-normal distributions which unfortunately appears to be still in its early stages of development.

The extreme value theory may also be helpful in constructing approximate estimates of the parameters in expression (3) which would otherwise require the collection of data on smaller fires. For, if we are able to obtain the reduced values corresponding to probability points not widely spaced for the, say, top 20 or 30 extremes, we could estimate the parameters a_m and b_m for these extremes as in the case of the largest value. With these values it may be possible to estimate the parameters α and β in the initial density function $f(z)$. Incidentally, the hypothesis regarding $f(z)$ i.e. increasing failure rate function, could be proved or rejected for a particular group of buildings. An investigation on the above lines appears to be worthwhile and useful.

CONCLUSION

At present for a majority of cases figures for monetary damages are available only for fires costing £10,000 or more. For answering questions in the fields of fire protection economics, one is forced to work with only these figures which are inadequate in the statistical sense but of considerable economic importance. It may be some time before figures for smaller losses become available to this Organization.

In these circumstances special tools are necessary for analysing the data on large fires. The obvious technique to be used is the statistical theory of extreme values. In this paper the possibility of applying this theory has been investigated.

From data available the probability distribution of fire loss appears to be of the exponential type. Large losses are extreme observations from this distribution. The theory was applied to the largest losses in the textile industry during the 21 years from 1947 to 1967. The logarithm of these figures after correcting for inflation fitted well with the first asymptotic distribution of largest values. Based on this relationship it appears that the modal largest loss in the textile industry was of the order of £219,000. Also the expected largest loss in the next five years is about one million pounds. Lastly, if it is planned to protect the building against a loss of £1.5 million this should be sufficient for the next 38 years. This means that, during the next 38 years, only one fire would be expected to have a loss equal to or larger than £1.5M. When the theory is sufficiently developed it should be possible to show that the maximum expected loss could be held to a lower level, say, £0.5M if certain fire protection measures are adopted. We could then combine the expected reduction in loss of £1M with fire frequency and compare the expected benefit with the expenditure on fire protection. The theory thus enables us to define an acceptable level of loss, i.e. to ensure that the probability of any loss in excess of this level being acceptably small. Only for purposes of illustration certain levels and the associated periods have been used in this paper. The actual period of planning would depend upon the expected life of the building, the changes which might take place in the occupancy of the premises and the expected useful life of the protection itself. The life of an industrial building has been estimated to be about 40 years.

The figures quoted above are all at 1947 prices and are based on current trends. But drastic changes in the fire protection measures or in the trade carried on in the building could alter the picture. The figures are also applicable only to fairly large buildings with values at risk at least equal to the expected large losses. For smaller buildings with lesser values at risk adjustments are necessary. The regression between loss and value at risk would yield such adjustment factors.

Planning fire protection measures on the basis of the largest value may not be economically feasible. Hence we may have to use the mean, median or other suitable statistics of a few extremes at the top. But this involves research into statistical methods specially suited to handle extremes. Extremes are not normally distributed; neither therefore are statistics derived from them. Research in this direction is likely to yield results useful for practical purposes.

REFERENCES

1. EPSTEIN, B. Bacterial extinction time as an extreme value phenomenon. Biometrics (1967), 23, 835-839.

2. GUMBEL, E. J. Statistics of extremes. Columbia University Press, New York, 1958.

3. HARRIS, R. Contributions to the theory of extreme values. Ph.D. Thesis, University of California, 1968.

4. RAMACHANDRAN, G. The Poisson Process and fire loss distribution. Paper read at the 37th Session of the International Statistical Institute, London, Sept.1969.

5. SARHAN, A. E., GREENBERG, B. G. Contributions to order statistics. John Wiley & Sons, New York, 1962.

6. Probability Tables for the analysis of extreme value data. U.S. Department of Commerce, National Bureau of Standards. Applied Mathematics Series 22, 1953.