Fire Research Note

No. 910

EXTREME VALUE THEORY AND FIRE LOSSES - FURTHER RESULTS

by

G. RAMACHANDRAN

January 1972

FIRE
RESEARCH
STATION

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

F.R. Note No.910
January 1972

EXTREME VALUE THEORY AND FIRE LOSSES - FURTHER RESULTS

by

G. Ramachandran

SUMMARY

In a previous paper the author illustrated the use of the theory of extreme values for analysing the largest losses due to fires in buildings. In this paper the theory is extended so as to cover the top 17 extreme losses in the textile industry. A few statistical problems concerning these extremes and their averages are discussed. Using the estimated values of the parameters of these extreme value distributions a method for assessing the total loss in smaller fires in a given year is also illustrated. This method could also be used to estimate the expected loss in a particular building of known value at risk. Problems for further research have also been suggested.

Conceptually, the intensity function of the probability distribution of fire loss is 'U' shaped. But, neglecting the infant and early stages of growth of fire this function increases exponentially. In 1967, there were about 105 fires in the textile industries with individual losses in the range between £55 and £10,000. The overall average loss in these fires was about £2,200. In the same year and in the same industry sprinklered buildings had an average loss of £1,600 for a comparable loss range. Hence in non-sprinklered buildings the average loss was about £2,800 indicating a saving of £1,200 per fire due to sprinklers in the range considered.

It is extremely unlikely that the total loss in all smaller fires (costing less than £10,000) in the textile industries in 1967 was more than £300,000.

KEY WORDS: Large fires, loss, fire statistics

DEPARTMENT OF THE ENVIRONMENT AND FIRE OFFICES' COMMITTEE

JOINT FIRE RESEARCH ORGANIZATION

F.R. Note No.910
January 1972

EXTREME VALUE THEORY AND FIRE LOSSES - FURTHER RESULTS

by

G. Ramachandran

INTRODUCTION

In many technological problems interest centres around a statistical assessment of the life characteristics of the material under investigation. One such problem is the phenomenon of the spread of fire, statistical treatment of which has been receiving attention only in recent times. A deeper understanding of the life of a fire under given conditions would provide a sound technical basis for planning fire fighting and fire protection strategy.

Essentially there are three ways in which the life of a fire could be expressed. First we could consider the physical extent of spread in terms of the area or volume destroyed. Secondly, the duration of burning in time units, which is useful in certain problems, especially those concerning fire brigade operations. Thirdly, the extent of financial damage, which plays a vital role in the economics of fire protection measures. The analysis in this paper relates only to the third aspect, viz. financial loss. Using suitably determined conversion factors it should be possible to translate the monetary values into equivalent values in time or physical damage, but this problem is one for later study.

A statistical assessment of the financial damage in a fire implies a study of the probability distribution of fire loss. This is the distribution of the probabilities with which the cumulative monetary damage in a fire reaches various amounts. The structure of this distribution is not likely to change from one set of conditions to another; but the values of the parameters would vary. The values of the parameters would thus serve as indices of fire risks.

If loss figures were available for all fires it would be possible to establish the structure and parametric values of the probability distribution fairly precisely. But at present loss data are available only for fires costing £10,000 or more. Hence the available observations are large or extreme values at the tail end of a given parent distribution. Repeated sets of extreme values are produced from year to year under different conditions. One

is therefore faced with something like the traditional experimental design and analysis situation with the difference that the data are known to be sets of extreme values from an unknown distribution. The relevant theory for tackling such data is the extreme value theory.

In a previous paper[1] the author reviewed the extreme value theory with special reference to financial losses due to fires in buildings. It has been assumed that the probability distribution in a fire, i.e. the parent distribution, is of the exponential type if the logarithm of loss is considered as the operational variable. Using the appropriate asymptotic distribution the largest losses in the textile industry in the United Kingdom during the period 1947 to 1967 have been analysed. In this paper the extension of the theory to cover the top 17 extremes is discussed.

THE DATA

The data used in this paper relate to large losses that occurred in the textile industry in the United Kingdom during the 21 year period from 1947 to 1967. The top 17 of these losses arranged in decreasing order of magnitude for each year are given in Table 1, Appendix 1. These were preliminary estimates furnished to the British Insurance Association and published in the 'Times'.

With the aid of the index numbers for retail prices a correction for inflation has been made to the observed loss figures and the corrected figures are shown in Table 2.

In the previous paper[1] it was explained in detail that the logarithm of loss follows a probability law of the exponential type. Hence in the actual analysis we have to consider the logarithms of corrected losses and these are given in Table 3.

THE REDUCED EXTREMES

Consider the variable Z which is the logarithm of the observed loss corrected for inflation. This has a probability distribution $F(Z)$ the structure of which is assumed to remain the same over the years. The fire losses in a given year constitute a sample of observations from $F(Z)$. Let these losses be $Z_1, Z_2 \ldots\ldots Z_n$ with n denoting the sample size, i.e. number of fires in the year. If these n observations are rearranged in decreasing order of magnitude, let $Z_{(1)}$ be the largest and $Z_{(n)}$ the smallest. Also let $Z_{(m)}$ be the mth observation from top.

Over the years, $Z_{(m)}$ has a probability distribution the structure of which depends upon the rank m, the parent distribution $F(Z)$ and the sample size n. But if $F(Z)$ is of exponential type, n is large and m is small in comparison with n. The asymptotic density function $\chi_m(y_m)$ of the mth extreme from top is

$$\chi_m(y_m) = \frac{m^m}{(m-1)!} e^{-m y_m - m e^{-y_m}} dy_m \tag{1}$$

where the reduced mth large value y_m is defined as

$$y_m = a_{mn}(Z_{(m)} - b_{mn}) \tag{2}$$

The parameters a_{mn} and b_{mn} in (2) are the solutions of

$$F(b_{mn}) = 1 - \frac{m}{n} \quad \text{and} \tag{3}$$

$$a_{mn} = \frac{n}{m} f(b_{mn}) \tag{4}$$

where $f(Z)\,(= F'(Z))$ is the density function of Z.

The distribution function corresponding to the density function (1) is

$$\Phi_m(y_m) = \int_{-\infty}^{y_m} \chi_m(y_m)$$

which could be rewritten as an incomplete gamma function

$$\Phi_m(u_m) = \int_{u_m}^{\infty} u^{m-1} e^{-u} du / \Gamma(m) \tag{5}$$

by introducing the transformation

$$u_m = m e^{-y_m} \tag{6}$$

In the application of the classical theory of extreme values it is assumed that the sample size n is constant. This assumption is not satisfied in the example considered as the frequency of fires, though large, has varied considerably from year to year[1]. This variation would be expected to affect the values of the extreme value parameters a_{mn} and b_{mn} as they are functions of n. This aspect of the problem has been investigated in detail in Appendix 2. It appears that the following model would suffice for all

practical purposes:

$$Z_{(m)j} = b_{mn_1} + \frac{y'_{mj}}{a_{mn_1}} \tag{7}$$

where

$$y'_{mj} = y_{mj} + \log(n_j/n_1) \tag{8}$$

n_1 = number of fires in 1947

n_j = number of fires in the jth year

b_{mn_1} = the characteristic m^{th} large value for samples of size n_1 as defined by expression (3) with $n = n_1$

a_{mn_1} = the failure rate of parent distribution (at b_{mn_1}) as defined by expression (4) with $n = n_1$

$Z_{(m)j}$ = the mth large value from top as observed in the jth year, and

y_{mj} = the reduced m^{th} large value corresponding to $Z_{(m)j}$ calculated on the assumption that the samples have the constant size of n_1 .

The values of $Z_{(m)j}$ for m = 1 to 17 are contained in Table 3, Appendix 1. If each of the 17 sets (of 21 values) of $Z_{(m)j}$ is arranged in increasing order of magnitude let R_{mj} be the rank of $Z_{(m)j}$. The ranks are shown in Table 4. The empirical value of the cumulative relative frequency of $Z_{(m)j}$ is

$$\Phi_m(Z_{(m)j}) = \frac{R_{mj}}{N+1} \tag{9}$$

where N = 21. Since cumulative frequencies are preserved under transformations

$$\Phi_m(Z_{(m)j}) = \Phi_m(y_{mj}) = \Phi_m(u_{mj})$$
$$= \int_{u_{mj}}^{\infty} u^{m-1} e^{-u} du / \Gamma(m) \tag{10}$$

where

$$u_{mj} = m e^{-y_{mj}} \tag{11}$$

The values of u_{mj} ($m = 1, \ldots 17$; $j = 1, \ldots 21$) corresponding to the cumulative frequencies (9) were calculated from tables of incomplete gamma functions. Then using (11) the values of y_{mj} were obtained and are given in Table 5.

The correction term $\log(n_j/n_1)$ for y_{mj} is independent of m. These correction terms are shown in Table 6. Using these values the corrected reduced variates y'_{mj} have been calculated using (8) and are shown in Table 7.

EXTREME VALUE PARAMETERS

The next step in the problem is to estimate the extreme value parameters a_{mn_1} and b_{mn_1}. This can be achieved by fitting the straight line (7) to the variables $Z_{(m)j}$ and y'_{mj} ($j = 1, \ldots 21$) using the method of least squares. The values of the parameters thus obtained are given in Table 8. Also given in the table are the mean $\bar{Z}_{(m)}$ and variance $s^2_{(m)Z}$ of $Z_{(m)j}$, and the mean $\bar{y}_m$ and variance s^2_{my} of y_{mj}.

Fitting the straight line (7) involves a residual error e_{mj}. The expected value of e_{mj} may be assumed to be zero while the variance of e_{mj} is equal to

$$RS^2_m = s^2_{mZ}\left(1 - r^2_m\right) \tag{12}$$

where r_m is the correlation coefficient between $Z_{(m)j}$ and y'_{mj}. The values of r_m are also given in Table 8. Since a high correlation has been observed in all the cases, RS^2_m is negligible in magnitude. The high correlation also strengthens the hypothesis that the logarithm of fire loss follows a probability law of the exponential type.

In the example under consideration the number of observations (years) N for each extreme is 21. For large values of N, the sample moments $\bar{y}_m$ and s^2_{my} of the reduced extreme y_{mj} tend to limiting or population values $\bar{Y}_m$ and σ^2_{my} respectively. The derivation of these limiting values is shown in Appendix 3 together with a tabulation for $m = 1$ to 40.

In the same way, $Z_{(m)j}$ has a limiting mean $\hat{Z}_m$ and variance σ^2_{mZ}. Using the sample estimates of a_{mn_1} and b_{mn_1},

$$\hat{Z}_m = b_{mn_1} + \frac{\bar{Y}_m + \bar{P}_j}{a_{mn_1}} \quad \text{and} \tag{13}$$

$$\sigma^2_{mz} = \left(\frac{1}{a_{mn_1}}\right)^2 \left[\sigma^2_{my} + \sigma^2_p + 2\, Cov(y_{mj}, p_j) \right] \tag{14}$$

In (13), $\bar{P}_j$ is the mean (population) value of the correction factor

$$p_j = \log\left(n_j/n_1\right) \tag{15}$$

an estimate of which is furnished by the sample value

$$\bar{p}_j = \frac{1}{N} \sum_{j=1}^{N} p_j \tag{16}$$

Also in (14) sample estimates of σ^2_p and $Cov\,(y_{mj}, p_j)$ are furnished by

$$s^2_p = \frac{1}{N-1} \sum_{j=1}^{N} (p_j - \bar{p}_j)^2 \quad and \tag{17}$$

$$Cov\,(y_{mj}, p_j) = \frac{1}{N-1} \left[\sum_{j=1}^{N} y_{mj} \cdot p_j - N\, \bar{y}_m\, \bar{p}_j \right] \tag{18}$$

In (13) and (14) population moments are available only for y_{mj} i.e. $\bar{Y}_m$ and σ^2_{my} since the probability distribution of y_{mj} only is known. For the remaining parameters due to p_j only sample estimates can be inserted. By studying the trend in the frequency of fires it may be possible to obtain better estimates of $\bar{P}_j$, σ^2_p and $Cov\,(y_{mj}, p_j)$. However, using available estimates, the values of $\hat{Z}_m$ and σ^2_{mz} have been calculated and given in the last two columns of Table 8. These figures are improvements over the corresponding values $\bar{Z}_m$ and s^2_{mz} (Table 8) which are purely sample estimates.

AVERAGES OF EXTREMES

Consider the observed extremes $Z_{(m)j}$ ($m = 1, \ldots 17$) of the jth year. The expected value $\hat{Z}_m$ of $Z_{(m)j}$ is given by (13). If

$$\bar{Z}_{17,j} = \frac{1}{17} \sum_{m=1}^{17} Z_{(m)j} \tag{19}$$

the average $\bar{Z}_{17,j}$ for a year has a probability distribution for different values of j. Hence over the years, $\bar{Z}_{17,j}$ itself has the expected value

$$\begin{aligned}\bar{Z}_{17} &= E(\bar{Z}_{17,j}) \\ &= \frac{1}{17}\sum_{m=1}^{17} E(Z_{(m)j}) \\ &= \frac{1}{17}\sum_{m=1}^{17} \hat{Z}_m \end{aligned} \tag{20}$$

The variance of $\bar{Z}_{17,j}$ is given by

$$\begin{aligned}\bar{\sigma}^2_{17,Z} &= \frac{1}{17^2} Var \sum_{m=1}^{17} Z_{(m)j} \\ &= \frac{1}{289}\left[\sum_{m=1}^{17} Var\, Z_{(m)j} + 2\sum_{\substack{\ell=1 \\ \ell\neq m}}^{17}\sum_{m=1}^{17} Cov(Z_{(\ell)j}, Z_{(m)j})\right]\end{aligned} \tag{21}$$

The variance of $Z_{(m)j}$ is σ^2_{mZ} given by (14). For a given j (year) the covariance of two extreme order statistics $Z_{(\ell)j}$ and $Z_{(m)j}$, from (7), is given by

$$Cov(Z_{(\ell)j}, Z_{(m)j}) = Cov(y'_{\ell j}, y'_{mj}) \Big/ a_{\ell n_1} a_{m n_1}$$

It may be easily verified that

$$\begin{aligned}&Cov(y'_{\ell j}; y'_{mj}) \\ &= Cov(y_{\ell j}, y_{mj}) + Cov(y_{\ell j}, p_j) + Cov(y_{mj}, p_j) + \sigma^2_p\end{aligned} \tag{22}$$

In Appendix 4 it has been proved that if $m > \ell$, with the order counted from the top, the covariance of $y_{\ell j}$ and y_{mj} is equal to the variance σ^2_{my} of y_{mj} tabulated in Appendix 3.

Using the estimated values of $\hat{Z}_m$, σ^2_{mZ}, $a_{\ell n_1}$, $a_{m n_1}$ and the covariances it was found that

$$\begin{aligned}\bar{Z}_{17} &= 3.9904 \quad \text{and} \\ \bar{\sigma}^2_{17,Z} &= 0.1645\end{aligned} \tag{23}$$

Hence over the years, the averages of the top 17 losses in the textile industry were fluctuating around a value of 3.9904 (£54,100) with a standard error of 0.4056 (£1,500) both at 1947 prices.

There are two unsolved problems concerning an average of large order statistics. Firstly, we require the confidence limits for the average. These could be obtained either from the probability distribution of the average or by using the standard 't' value corrected for skewness and kurtosis as indicated by Gayen[2]. Both these methods are being attempted. Secondly, a test has to be devised for judging the significance of the difference btween two averages of extremes. This problem would arise when we wish to compare, say, the textile industry with another industry in terms of the averages. A solution to this problem is also being attempted.

RETURN PERIOD

In the previous paper[1] the usefulness of the 'return period' was illustrated with reference to the largest value ($m = 1$). This concept may be generalised with some modification. If the mth large observations follow each other in time, if the sample size is constant, and further if the distance between consecutive m^{th} large values is approximately constant (1 year in our case) then the function

$$T_m(z_{mrp}) = \left[1 - \Phi_m(z_{mrp})\right]^{-1} \tag{25}$$

defined as the 'return period for the mth large value', gives the average number of years necessary to obtain an mth value larger than z_{mrp}.

Now consider the 29 year period from 1947 to 1975. For $T = 29$ expression (25) gives $\Phi_m(z_{mrp}) = \frac{28}{29} = 0.966$. The reduced values y_{mrp} ($= 1, \ldots. 17$) corresponding to the above cumulative probability, obtained via incomplete gamma functions, are given in Table 9. If the sample size (number of fires) is constant at the value n_1 of the base year 1947, then only one of the mth values will exceed

$$z_{mrp} = b_{mn_1} + \frac{y_{mrp}}{a_{mn_1}} \tag{26}$$

in the course of 29 years from 1947. Since the sample size is increasing this will actually occur in a period much shorter than 29 years, or more than once in 29 years. But we are interested in a value of z_{mrp} which will be

exceeded only once before 1975. This has to be higher than the value given by (26). The following method of calculation may be adopted.

If it is assumed that the number of fires increases at a rate 'r' per annum the frequency n_j in the jth year is given approximately by

$$n_j = n_1 (1 + r)^{j-1} \tag{27}$$

where n_1 is the number of fires in the base year (1947). Fitting (27) to the data for the textile industries for the period 1947 to 1967 the value of r appears to be 0.038. Since n_1 = 465, about 1310 fires are likely to occur in the textile industries in the year 1975 (j = 29). From the straight line (7)

$$z'_{mrp} = b_{mn_1} + \frac{y_{mrp} + 1.036}{a_{mn_1}} \tag{28}$$

since log (n_{29}/n_1) has the value 1.036. The probability of log (n_j/n_1) being less than 1.036 is almost unity. Hence the probability of the observed loss being less than z'_{mrp} during the period 1947 to 1975 is 0.966. The probability of exceeding z'_{mrp} is 0.034 or one in 29, i.e. once before 1976. The values of z'_{mrp} are also given in Table 9.

If we take the mth loss from top in the textile industries every year before 1976 and correct for inflation (at 1947 values) only one of them is likely to exceed z'_{mrp} given in Table 9. The estimates in Table 9 are based on the current trend. During the course of the period up to 1970 none of the actual losses exceeded z'_{mrp} values, except the 4th and 5th extremes of 1969. Hence the excesses are likely to happen during the 5 year period 1971 to 1975. If they do not actually happen then it would appear that improved fire protection and fire fighting methods in textile industry fires were having an effect. Fire prevention activities would also have played some part in keeping the number of fires below the levels forecast in (27) thus reducing the values of log (n_j/n_1). On the other hand if more than one m^{th} loss were to exceed z'_{mrp} during 1971 to 1975 then there would be reason to doubt that fire fighting, fire prevention and fire protection methods are coping with the situation.

The forecast figures given in Table 9 are for the entire population of textile industry buildings. By doing further research it would be possible to

forecast z'_{mnp} values separately for, say, sprinklered and non-sprinklered buildings. From such an analysis it would be possible to estimate the saving due to sprinklers in extreme (very large) losses. The estimate could be used as a guide for assessing the economic value of sprinkler protection in buildings where such large losses are likely to occur.

THE PARENT DISTRIBUTION

Consider the parameter a_{mn_1} for the mth extreme from a sample of size n_1. From (3) and (4)

$$a_{mn_1} = f(b_{mn_1}) / 1 - F(b_{mn_1})$$
$$= h(b_{mn_1}) \qquad (29)$$

The function $h(u)$ gives the conditional probability of extinction of the fire in the interval $(u, u+du)$ given that it has survived till the value u has been reached. It denotes the ratio of the probability of extinction to the probability of survival or spread at the point u. This function is known as the 'intensity function' in extreme value theory, 'hazard' or 'failure rate' in reliability theory and 'force of mortality' in actuarial statistics. The parameter a_{mn_1} is the value of the intensity function of the parent distribution $F(z)$ at the characteristic large value b_{mn_1}.

A constant value of a_{mn_1} for varying m is an indication that the parent distribution is a simple exponential distribution, i.e.

$$f(z) = a e^{-az} \quad \text{or}$$
$$F(z) = 1 - e^{-az}$$

However, according to Table 8 a_{mn_1} decreases for increasing m or increases with increasing z, indicating an increasing failure rate. The increasing trend for m = 14 to 17 is likely to be due to random fluctuations. In earlier papers[1,3] the author discussed the possible relevance of distributions with an increasing failure rate for describing the probability distribution of fire loss. If a fire has been burning for a long time it is likely that fire fighting will have commenced meanwhile. Some items (e.g. oxygen, fuel) contributing to fire spread may also be getting exhausted. Given these factors the probability of extinction would increase at a rate higher than that of the probability of spread so that $h(u)$ (or $h(z)$) would increase for large values of u (or z).

During the period immediately following the ignition of the first material involved the failure rate is likely to be high. Such a phenomenon is known as 'infant mortality' in the analysis of life test data concerning, say, electric bulbs. It arises when the failure rate is relatively high in the early period of life. This phenomenon has also been observed in the case of a human life. It should be true in the case of fire. A high rate in the 'infant' stage may be attributed to the presence in the room of origin of materials which 'fail' to continue to burn after ignition or to which fire fails to spread. It may be of interest to note in this connection that in 1967, out of a total number of 982 fires attended by fire brigades in buildings concerned with textile manufacture, 524 fires were confined to exterior components, appliances and common service spaces[4].

In the early stages of growth after the infant stage, a fire has a greater tendency to spread so that $h(u)$ tends to decrease. Thereafter, after remaining constant for a short period, $h(u)$ will eventually increase till the fire become extinct. Of course, a fire cannot burn for ever. There is no possibility of checking the above assumptions regarding the infant and early stages of growth since data are not available for small losses.

Conceptually, therefore, $h(u)$ is a 'U' shaped curve. If, however, we disregard the infant mortality and early growth periods, for the remaining long range of the variable we may assume that $h(u)$ increases exponentially so that

$$h(u) = e^{\alpha + \beta u} \tag{30}$$

Consider now the year 1967. There were 982 fires (n_j) in textile industries in that year against n_1 (= 465) fires in 1947. Using the correction formulae

$$b_{rn_j} = b_{rn_1} + \frac{1}{a_{rn_1}} \log (n_j/n_1)$$

the values of b_{rn_j} for 1967 have been calculated and given in Table 10. These values are logarithms of characteristic values in units of one pound. A constant value 6.908 ($\log_e$ 1000) was added to b_{rn_1} as it was originally estimated in units of £1,000. Such a change implies only a change of origin in the straight line (7) without affecting the slope of the line, i.e. $1/a_{rn_1}$.

The estimated values of a_{rn_1} are also reproduced in Table 10.

Using the fact that $a_{rn_1} = h.(b_{rn_1})$, the exponential function in (30) was fitted to the data in Table 10. The following results were obtained.

$$\alpha = -4.0825 \quad \text{and}$$
$$\beta = 0.3839 \tag{31}$$

with a high correlation (0.9586). As explained earlier the function (30) could be expected to give reasonably good estimates of the failure rate for the major portion of the range of the fire loss variable Z excluding small values. In this connection it has to be mentioned that small claims are generally disallowed by an insurance firm as 'deductibles'; the insured himself having to bear a certain minimum loss. In 1967 the smallest loss in textile factories provided with sprinklers was £65. (For sprinklered buildings loss figures for fires costing less than £10,000 are available). Hence it is reasonable to assume a minimum value of £25 (x_0) or 3.219 (z_0) for **Z** at 1947 values. At 1967 values the minimum loss would be £55.

Acceptance of the values in (31) with z_0 = 3.219 provides a few interesting results. First consider the function

$$1 - F(z) = e^{-\int_{z_0}^{z} e^{\alpha+\beta u} du} \tag{32}$$

where $F(z)$ is the cumulative distribution function (expression 2, Appendix 2). The function (32) gives the probability of exceeding the value z given that the loss is greater than z_0 . A loss of £10,000 in 1967 corresponds to a loss of £4,500 at 1947 values. Hence if x_ℓ is 4,500 with the corresponding value z_ℓ = 8.4, the probability of loss exceeding z_ℓ as given by (32) is 0.385. In 1967 there were 65 large fires costing £10,000 or more. Hence in that year there were about 170 fires in textile industries each costing £55 or more, out of which about 105 were smaller fires costing less than £10,000 individually. According to the data furnished by the Fire Protection Association, there were 59 smaller fires in 1967 in textile buildings provided with sprinklers. An equal number of smaller fires in non-sprinklered buildings could have occurred since about 50 per cent of the textile industry is sprinklered. Also in 1967, out of a total of 982 fires

attended by fire brigades in this industry, only 194 fires spread beyond the room of origin[4].

The exponential function (30) with the values of the parameters given by (31) has another practical use. With the aid of this function it is possible to obtain an approximate estimate of the average loss in smaller fires in the range less than £10,000. (At present losses in smaller fires are not available for non-sprinklered buildings). The method of estimation is described in detail in Appendix 5. Accordingly the average loss in smaller fires in 1967 is given by

$$\bar{x}_e = (\beta\alpha')^{1/\beta} c_e (p_e - p_0)\, \Gamma\left(1 + \frac{1}{\beta}\right) \qquad (33)$$

where α and β are given by (31) and

$$\alpha' = e^{-\alpha} = 59.3$$

$$\xi_e = e^{\alpha+\beta z_e}/\beta = 1.104 \quad (z_e = 8.4)$$

$$\xi_0 = e^{\alpha+\beta z_0}/\beta = 0.1514 \quad (z_0 = 3.219)$$

$$c_e = \frac{1}{e^{-\xi_0} - e^{-\xi_e}} = 1.908$$

$$p_e = \frac{\int_0^{\xi_e} e^{-\xi}\xi^{1/\beta}\,d\xi}{\int_0^{\infty} e^{-\xi}\xi^{1/\beta}\,d\xi} = 0.046$$

$$p_0 = \frac{\int_0^{\xi_0} e^{-\xi}\xi^{1/\beta}\,d\xi}{\int_0^{\infty} e^{-\xi}\xi^{1/\beta}\,d\xi} = 0.0001 \text{ (negligible) and}$$

$$\Gamma\left(1 + \frac{1}{\beta}\right) = 3.32 \text{ approximately.}$$

Inserting the above values in (33) the estimated value of $\bar{x}_e$ (for all buildings) appears to be of the order of £1,000 at 1947 values or about £2,200 at 1967 values. In 1967 the average loss in smaller fires in sprinklered buildings engaged in the textile trade was about £1,600. Hence in non-sprinklered buildings the average loss was about £2,800 indicating a saving of about £1,200 per fire due to sprinklers in the range considered. Following the method described in Appendix 5 the standard deviation is about £1,100 at 1947 values or about £2,400 at 1967 values.

With an average loss of £2,200 the total loss in about 105 fires in the range £55 to £10,000 appears to be of the order of £231,000 with a standard error of about £24,000 (Appendix 5). Besides these there were about 812 smaller fires attended by fire brigades most of which were likely to have been confined to the room of origin, with an average loss of, say, £50. Also it was likely that a number of small fires extinguished by sprinklers and other means were neither attended by the fire brigades, nor reported to the organisation. The total losses in all these fires would have been only marginal. Adding all the above losses it is extremely unlikely that the total loss in all smaller fires in the textile industries in 1967 was more than £300,000.

DISCUSSION

As in other fields, a major task in fire protection economics is to evaluate the expected extent of damage in a given building or group of buildings. For this purpose it is necessary to find an expression defining the probability distribution of fire loss in the given risk. Estimation of the parametric values of this distribution would be reasonably easy if loss figures were available for the entire range. But at present, for a majority of fires, figures are available only for fires costing £10,000 or more.

Hence the precise structure of the parent probability distribution of fire loss is not known, though the logarithm of loss appears to belong to an exponential family. Available observations are located at the upper tail of this distribution. Therefore the treatment of the loss data has to rely on the techniques of extreme value theory. An application of this theory has been illustrated in this paper with the aid of data on the top 17 losses each year in the textile industry during the period 1947 to 1967. A few practical results have been obtained. Problems for further research have also been indicated.

The parts played by mean, standard deviation and standardised or reduced variable (i.e. $\frac{\text{variable-mean}}{\text{standard deviation}}$) in classical theory are taken up by b_m, a_m and y_m respectively in the theory concerning the mth extreme. In repeated sampling over, say, years, unlike normal theory, the expected value of y_m is not zero and its variance not unity. The moments of the error y_m depend upon the rank m from top. For different m in the same sample (year) the errors are not independent and hence have covariances. The errors are not normally distributed but tend to normality as m increases, i.e. as the centre of the parent distribution is approached. Considered individually

the extremes are not difficult to handle. But complications arise when they are to be used as a collection of extreme order statistics.

But extremes are useful. Their economic importance lies in the fact that more than 50 per cent of the total loss is in large fires. By studying their extreme value distributions over a period of years it is possible to get some idea of the parent distribution from which they arise. As seen in this paper the extreme value parameters a_{mn} are the values of the intensity function $h(u)$ at the characteristic large values b_{mn}. While a_{mn} could be assumed to be a constant (as a first approximation), b_{mn} increases with years. The unknown location parameter of the parent distribution is linked with the values of b_{mn} for varying m; hence, this parameter also increases over time due to inflation and the increasing number of fires. Such ideas of shifts in the parametric values denoting the changing trend in the parent distribution were expressed in a recent conference of the International Reinsurance Offices' Association[5].

The parameters a_{mn} and b_{mn} together describe the shape of the intensity curve $h(u)$ and hence, with sufficient accuracy, the parent distribution in the region of the extremes considered. Projection of this curve below the smallest extreme (largest m) considered is difficult. At this stage only conjectures are possible since data are not available for smaller losses.

Conceptually $h(u)$ will be roughly 'U' shaped due to infant mortality and decreasing failure rate for small values of u and increasing failure rate for large values. However, ignoring the infant and early stages which are not of economic importance, $h(u)$ will be an increasing function. Under this assumption and with the aid of the estimated values of the extreme value parameters, a method for estimating the expected loss in smaller fires has been described in detail in this paper, and applied to 1967 losses for purposes of illustration.

The above mentioned method could also be used to estimate the expected loss in a given building in the textile industry with a given monetary value 'v' at risk. Using expression (17), Appendix 5 the following results were obtained for a few magnitudes of v for the year 1967.

Value at risk (v) 1947 values (£)	Value at risk (v) 1967 values (£)	Expected loss ($E.L.$) 1947 values (£)	Expected loss ($E.L.$) 1967 values (£)	Expected loss ratio $\left(\frac{E.L.}{v}\right)$
1,000	2,200	320	700	0.320
5,000	11,000	1,100	2,420	0.220
10,000	22,000	1,830	4,030	0.183
50,000	110,000	5,480	12,060	0.110
100,000	220,000	7,840	17,250	0.078
500,000	1,100,000	12,270	26,990	0.025
1,000,000	2,200,000	12,920	28,420	0.013

From the above table it appears that the expected loss does not increase linearly with v. The proportionate damage (loss ratio) decreases perhaps exponentially, with increasing value at risk. As expressed in an earlier note[6], this apparent paradox may be due to the fact that a fire in a large building is more likely than one in a single room or a small building to be discovered and extinguished before involving the whole building. The proportion destroyed in a small building would therefore be expected to be greater than the proportion destroyed in a large building.

As explained in Appendix 5 the standard error of the total loss in fires in a group of independent buildings is equal to the standard deviation of the individual loss for the group multiplied by $\sqrt{n}$ where n is the number of fires in the group and period considered. The greater the value of n the better will be the prediction for the total loss. That is to say that the estimated value of the total loss for a large group will have a small coefficient of variation and hence be more reliable than the estimated total loss for a smaller group with a larger coefficient of variation. The number of fires in a group depends upon the number of buildings in the group. It is not claimed that studies of this nature could help an insurance firm to decide the number of policies it could accept from a particular range of values at risk or sums assured. The studies, as such, are also not likely to be useful for tackling reinsurance problems. However, it should be mentioned that attempts to apply extreme value theory for reinsurance strategies have been made by Beard[7], Hooge[8] and Jung[9].

The method of estimating the expected loss and standard error for a given range described in this paper may require refinement in the light of further

research, and the availability of additional information. Confidence limits are needed for the expected loss. At any rate extreme value theory appears to have practical applications in the field of fire loss.

CONCLUSIONS

Extreme value distributions of exponential type parents fitted well with the top 17 observed fire losses in the textile industry during the period from 1947 to 1967. A high correlation between observed and theoretical values was obtained for each extreme. This strengthens the assumption that the parent probability distribution of fire loss belongs to the exponential family if the logarithm of loss is considered as the operational variable after applying the necessary correction for inflation.

In the textile industry the top 17 losses over a period of years had an expected value of 3.9904 (£54,100) with a standard error of 0.4056 (£1,500) both at 1947 prices. In the calculation of these estimates the non-normality of the extremes and the dependence between them have been taken into consideration.

The frequency of fires in the textile industry increases at a rate of 3.8 per cent per annum. About 1,310 fires are likely to occur in this industry in 1975.

A 'return period' analysis yielded certain forecast values (Table 9) at 1947 prices for the top 17 extremes. If we take the mth losses from top that actually occur in the textile industry every year before 1976 and correct them for inflation (to 1947 values) only one is likely to exceed the corresponding forecast value. The forecast figures are based on the current trend; only drastic changes in fire fighting and fire protection methods or the industrial processes would be likely to alter this picture for better or worse.

A 'U' shaped model for the intensity function of the parent probability distribution of fire loss appears to be physically relevant. This function appears to increase exponentially in the range excluding the infant and early stages of fire growth.

With the aid of the estimated values of the parameters of the extreme value distributions it is possible to estimate the values of two parameters describing the parent distribution. These results indicate that, in 1967, there were about 170 fires in the textile industries each costing £55 or more of which about 105 were smaller fires costing less than £10,000 individually.

A method for estimating the expected loss in smaller fires has been described in this paper. Following this method it appears that, in 1967, at 1967 values, the average loss in the textile industry in the range £55 to £10,000 was about £2,200. In the same year and industry and for the same range the known average loss in sprinklered buildings was £1,600. These figures gave an average loss of about £2,800 in non-sprinklered buildings indicating an average saving of about £1,200 per fire due to sprinklers in the given range of smaller losses.

It appears that, in 1967, the total loss in all smaller fires in the textile industries was not more than £300,000. In the same year and industry, 65 large fires each costing £10,000 or more caused a total loss of £4.55 million. Thus large fires appear to have accounted for nearly 94 per cent of the total loss in the industry.

The expected losses due to fire have been estimated for a few values at risk in buildings engaged in the textile trade. It appears that the proportionate damage (loss ratio) decreases with increasing value at risk.

For a given or acceptable level of the coefficient of variation of the total loss for a group of buildings it is possible to determine the corresponding number of fires (and buildings) for a particular range of value at risk. These studies, as such, are not likely to help an insurance firm to decide the maximum number of policies it could accept from a particular range of sums assured. It is also not claimed at this stage that the studies would be useful for tackling reinsurance problems.

REFERENCES

1. RAMACHANDRAN, G. Some possible applications of the theory of extreme values for the analysis of fire loss data. Ministry of Technology and Fire Offices' Committee Joint Fire Research Organization F.R. Note No.837.

2. GAYEN, A. K. The distribution of 'student's' t in random samples of any size drawn from non-normal universes. Biometrika (1949), 36, 353-369.

3. RAMACHANDRAN, G. The Poisson Process and fire loss distribution. Thirty-seventh session of the International Statistical Institute, London, Sept. 1969.

4. United Kingdom Fire Statistics, 1967. Her Majesty's Stationery Office.

5. Inflation above the retention line. Paper read at the annual meeting of the International Reinsurfance Offices' Association, London, June 1971. Policy Holder Insurance Journal, 18th June 1971.

6. RAMACHANDRAN, G. Fire Loss Indexes. Ministry of Technology and Fire Offices' Committee Joint Fire Research Organization F.R. Note No.839.

7. BEARD, R. E. (1963). Some notes on the statistical theory of extreme values. Astin Bull. Vol.III, Pt.I, 6-12.

8. HOOGE, L. D. (1965). Theorie des valeurs extremes et la tarification de "l'excess of loss". Astin Bull. Vol.III, Pt.II, 163-177.

9. JUNG, J. (1965). On the use of extreme values to estimate the premium for an excess of loss reinsurance. Astin Bull. VOL.III, Pt.II, 178-184.

APPENDIX 1

Table 1

Fire losses in the textile industry

(£'000)

Extremes	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1	460	350	210	350	550	1000	460	150	320	250	400	340	570	269	310	532	493	392	1912	445	1033
2	270	200	150	300	350	450	320	150	230	180	350	200	363	140	200	250	450	325	1002	400	300
3	198	191	140	173	130	250	285	90	200	150	125	200	250	110	175	165	450	300	635	309	290
4	190	143	135	115	75	150	275	75	190	145	100	150	241	80	142	155	286	290	502	275	286
5	135	105	120	110	70	125	176	65	160	125	80	140	200	55	100	110	175	245	445	257	280
6	75	100	100	100	70	90	150	50	110	100	75	120	188	50	97	110	167	225	370	230	268
7	45	100	86	100	65	80	86	25	100	90	45	112	170	50	92	82	165	191	290	205	203
8	30	65	60	80	56	65	85	20	100	90	35	110	120	48	79	77	126	180	275	172	192
9	27	46	55	75	50	60	80	20	100	75	32	75	100	45	75	72	126	170	200	143	114
10	20	32	35	75	50	60	60	15	100	75	30	72	80	44	52	65	115	151	200	142	112
11	17	31	25	74	49	59	58	11	80	74	29	69	75	40	50	64	90	144	199	110	109
12	15	22	24	65	40	50	50	10	70	60	25	53	71	34	46	63	64	129	185	108	95
13	14	18	21	60	35	40	49	9*	50	50	24	50	60	31	35	60	60	120	180	100	90
14	13	16	20	59	34	35	45	9*	49	49	20	49	55	28	33	45	52	88	120	90	85
15	13	13	19	50	30	30	45	9*	49	42	20	46	46	28	30	45	50	87	105	77	85
16	11	11	15	40	25	28	40	9*	45	40	20	44	43	25	23	44	50	81	82	75	82
17	10	10	15	36	25	28	35	9*	40	35	20	43	38	25	22	40	40	67	71	75	75

Table 2

Fire losses corrected for inflation (1947 values)

(£'000)

Extremes	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1	460	324	189	307	440	735	329	105	215	160	247	204	339	159	177	291	265	204	951	212	470
2	270	185	135	263	280	331	229	105	154	115	216	120	216	83	114	137	242	169	499	190	136
3	198	177	126	152	104	184	204	63	134	96	77	120	149	65	100	90	242	156	316	147	132
4	190	132	122	101	60	110	196	52	128	93	62	90	143	47	81	85	154	151	250	131	130
5	135	97	108	96	56	92	126	45	107	80	49	84	119	33	57	60	94	128	221	122	127
6	75	93	90	88	56	66	107	35	74	64	46	72	112	30	55	60	90	117	184	110	122
7	45	93	77	88	52	59	61	17	67	58	28	67	101	30	53	45	89	99	144	98	92
8	30	60	54	70	45	48	61	14	67	58	22	66	71	28	45	42	68	94	137	82	87
9	27	43	50	66	40	44	57	14	67	48	20	45	60	27	43	39	68	89	100	68	52
10	20	30	32	66	40	44	43	10	67	48	19	43	48	26	30	36	62	79	100	68	51
11	17	29	23	65	39	43	41	8	54	47	18	41	45	24	29	35	48	75	99	52	51
12	15	20	22	57	32	37	36	7	47	38	15	32	42	20	26	34	34	67	92	51	43
13	14	17	19	53	28	29	35	6	34	32	15	30	36	18	20	33	32	63	90	48	41
14	13	15	18	52	27	26	32	6	33	31	12	30	33	17	19	25	28	46	60	43	39
15	13	12	17	44	24	22	32	6	33	27	12	28	27	17	17	25	27	45	52	37	39
16	11	10	14	35	20	20	29	6	30	26	12	26	26	15	13	24	27	42	41	36	37
17	10	9	14	32	20	20	25	6	27	22	12	26	23	15	13	22	22	35	35	36	34

Table 3

Logarithms of extremes

Extremes	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957
1	6.131	5.781	5.242	5.727	6.087	6.600	5.796	4.654	5.371	5.075	5.509
2	5.598	5.220	4.905	5.572	5.635	5.802	5.434	4.654	5.037	4.745	5.380
3	5.288	5.176	4.836	5.024	4.644	5.215	5.318	4.143	4.898	4.564	4.344
4	5.247	4.883	4.804	4.615	4.094	4.700	5.278	3.951	4.852	4.533	4.127
5	4.905	4.575	4.682	4.564	4.025	4.522	4.836	3.807	4.673	4.382	3.892
6	4.317	4.533	4.500	4.477	4.025	4.190	4.673	3.555	4.304	4.159	3.829
7	3.807	4.533	4.344	4.477	3.951	4.078	4.111	2.833	4.205	4.060	3.332
8	3.401	4.094	3.989	4.249	3.829	3.871	4.111	2.639	4.205	4.060	3.091
9	3.332	3.761	3.912	4.190	3.689	3.784	4.043	2.639	4.205	3.871	2.996
10	2.996	3.434	3.466	4.190	3.689	3.807	3.761	2.303	4.205	3.871	2.944
11	2.833	3.401	3.135	4.174	3.664	3.761	3.738	2.079	3.989	3.850	2.890
12	2.708	2.996	3.091	4.043	3.497	3.611	3.584	1.946	3.850	3.638	2.773
13	2.639	2.833	2.944	3.970	3.332	3.367	3.555	1.792	3.526	3.466	2.708
14	2.565	2.708	2.890	3.951	3.296	3.258	3.466	1.792	3.497	3.434	2.485
15	2.565	2.485	2.833	3.784	3.178	3.091	3.466	1.792	3.497	3.296	2.485
16	2.398	2.303	2.639	3.555	2.996	2.996	3.367	1.792	3.401	3.258	2.485
17	2.303	2.197	2.639	3.466	2.996	3.045	3.219	1.792	3.296	3.091	2.485

Cont'd .....

Table 3 (cont'd)

Extremes	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1	5.318	5.826	5.069	5.176	5.673	5.580	5.323	6.858	5.357	6.153
2	4.788	5.375	4.419	4.736	4.920	5.489	5.130	6.213	5.247	4.913
3	4.788	5.004	4.174	4.605	4.500	5.489	5.050	5.756	4.990	4.883
4	4.500	4.963	3.850	4.394	4.443	5.037	5.017	5.521	4.875	4.868
5	4.431	4.779	3.497	4.043	4.094	4.543	4.852	5.398	4.804	4.844
6	4.277	4.719	3.401	4.007	4.094	4.511	4.762	5.215	4.700	4.804
7	4.220	4.615	3.401	3.970	3.829	4.489	4.595	4.970	4.585	4.522
8	4.190	4.263	3.332	3.807	3.738	4.220	4.543	4.920	4.407	4.466
9	3.807	4.094	3.296	3.738	3.664	4.220	4.489	4.605	4.234	3.951
10	3.784	3.892	3.258	3.401	3.584	4.127	4.382	4.605	4.220	3.932
11	3.714	3.807	3.178	3.367	3.555	3.871	4.317	4.595	3.951	3.932
12	3.466	3.738	3.045	3.258	3.555	3.526	4.205	4.522	3.951	3.761
13	3.401	3.584	2.890	2.996	3.497	3.434	4.143	4.500	3.871	3.714
14	3.401	3.526	2.833	2.944	3.219	3.332	3.829	4.094	3.761	3.664
15	3.332	3.296	2.833	2.833	3.219	3.296	3.807	3.951	3.611	3.664
16	3.258	3.258	2.708	2.565	3.178	3.296	3.738	3.714	3.584	3.611
17	3.258	3.135	2.708	2.565	3.091	3.091	3.555	3.555	3.584	3.526

Table 4

Ranks of extremes

Extremes	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1	18	14	5	13	17	20	15	1	9	3	10	6	16	2	4	12	11	7	21	8	19
2	18	11	6	17	19	20	15	2	9	4	14	5	13	1	3	8	16	10	21	12	7
3	18	16	9	14	7	17	19	1	11	5	3	8	13	2	6	4	20	15	21	12	10
4	19	15	11	9	3	10	20	2	12	8	4	7	16	1	5	6	18	17	21	14	13
5	20	12	14	11	4	9	17	2	13	7	3	8	15	1	5	6	10	19	21	16	18
6	11	15	13	12	5	8	16	2	10	7	3	9	18	1	4	6	14	19	21	17	20
7	4	17	13	14	6	9	10	1	11	8	2	12	20	3	7	5	15	19	21	18	16
8	4	11	9	16	7	8	12	1	14	10	2	13	17	3	6	5	15	20	21	18	19
9	4	8	12	16	6	9	14	1	17	11	2	10	15	3	7	5	18	20	21	19	13
10	3	6	7	17	9	12	10	1	18	13	2	11	14	4	5	8	16	20	21	19	15
11	2	7	4	19	9	12	11	1	18	14	3	10	13	5	6	8	15	20	21	17	16
12	2	4	6	19	9	13	12	1	17	14	3	8	15	5	7	11	10	20	21	18	16
13	2	4	6	19	8	9	15	1	14	12	3	10	16	5	7	13	11	20	21	18	17
14	3	4	6	20	10	9	14	1	15	13	2	12	16	5	7	18	11	19	21	18	17
15	4	3	7	19	9	8	15	1	16	11	2	14	13	5	6	10	12	20	21	17	18
16	3	2	6	17	9	8	15	1	16	11	4	13	12	7	5	10	14	21	20	18	19
17	3	2	6	17	8	9	14	1	16	10	4	15	13	7	5	1	12	20	19	21	18

Table 5

Reduced extremes (y_{mj}) - uncorrected

Rank (j)	Cumulative frequency ($j/N+1$)	Extremes (m)								
		1	2	3	4	5	6	7	8	9
1	0.0455	-1.125	-0.884	-0.760	-0.681	-0.619	-0.574	-0.538	-0.509	-0.483
2	0.0909	-0.874	-0.695	-0.598	-0.537	-0.489	-0.454	-0.426	-0.402	-0.383
3	0.1364	-0.691	-0.559	-0.485	-0.438	-0.398	-0.370	-0.347	-0.329	-0.312
4	0.1818	-0.533	-0.445	-0.389	-0.351	-0.324	-0.300	-0.283	-0.267	-0.254
5	0.2273	-0.394	-0.345	-0.307	-0.281	-0.256	-0.241	-0.227	-0.216	-0.205
6	0.2727	-0.261	-0.253	-0.229	-0.213	-0.197	-0.185	-0.175	-0.167	-0.160
7	0.3182	-0.136	-0.165	-0.158	-0.152	-0.143	-0.135	-0.129	-0.123	-0.118
8	0.3636	-0.011	-0.078	-0.087	-0.091	-0.089	-0.085	-0.082	-0.080	-0.078
9	0.4091	0.112	0.006	-0.020	-0.031	-0.036	-0.038	-0.039	-0.039	-0.039
10	0.4545	0.239	0.092	0.047	0.028	0.016	0.010	0.005	0.002	-0.001
11	0.5000	0.367	0.177	0.115	0.086	0.068	0.057	0.048	0.042	0.037
12	0.5455	0.502	0.244	0.185	0.146	0.122	0.105	0.094	0.084	0.077
13	0.5909	0.643	0.356	0.258	0.207	0.176	0.154	0.138	0.126	0.116
14	0.6364	0.793	0.451	0.333	0.271	0.233	0.205	0.185	0.170	0.157
15	0.6818	0.960	0.556	0.415	0.340	0.294	0.260	0.236	0.217	0.202
16	0.7273	1.143	0.669	0.503	0.415	0.360	0.318	0.290	0.267	0.249
17	0.7727	1.357	0.799	0.603	0.500	0.434	0.385	0.351	0.324	0.302
18	0.8182	1.605	0.962	0.716	0.594	0.518	0.461	0.419	0.387	0.359
19	0.8636	1.923	1.132	0.859	0.708	0.619	0.553	0.503	0.470	0.433
20	0.9091	2.350	1.372	1.043	0.863	0.756	0.668	0.610	0.564	0.523
21	0.9545	3.078	1.774	1.334	1.102	0.972	0.856	0.781	0.715	0.671

Cont'd .....

Table 5 (cont'd)

Rank (j)	Cumulative frequency ($j/N+1$)	Extremes (m)							
		10	11	12	13	14	15	16	17
1	0.0455	-0.462	-0.437	-0.428	-0.412	-0.399	-0.387	-0.377	-0.367
2	0.0909	-0.365	-0.351	-0.338	-0.327	-0.316	-0.307	-0.297	-0.290
3	0.1364	-0.299	-0.288	-0.277	-0.268	-0.259	-0.251	-0.243	-0.238
4	0.1818	-0.244	-0.234	-0.226	-0.218	-0.212	-0.206	-0.201	-0.195
5	0.2273	-0.197	-0.190	-0.183	-0.177	-0.172	-0.167	-0.163	-0.158
6	0.2727	-0.153	-0.150	-0.144	-0.139	-0.135	-0.131	-0.128	-0.127
7	0.3182	-0.114	-0.111	-0.108	-0.104	-0.101	-0.099	-0.097	-0.094
8	0.3636	-0.076	-0.074	-0.072	-0.070	-0.072	-0.067	-0.066	-0.064
9	0.4091	-0.039	-0.039	-0.039	-0.038	-0.038	-0.037	-0.037	-0.036
10	0.4545	-0.002	-0.004	-0.005	-0.006	-0.006	-0.007	-0.007	-0.008
11	0.5000	0.033	0.030	0.028	0.026	0.024	0.022	0.021	0.020
12	0.5455	0.071	0.066	0.062	0.058	0.055	0.053	0.051	0.048
13	0.5909	0.108	0.102	0.096	0.091	0.087	0.083	0.080	0.077
14	0.6364	0.148	0.139	0.131	0.125	0.120	0.114	0.110	0.106
15	0.6818	0.190	0.179	0.169	0.161	0.154	0.148	0.143	0.138
16	0.7273	0.233	0.221	0.209	0.200	0.191	0.184	0.177	0.171
17	0.7727	0.283	0.269	0.255	0.244	0.233	0.225	0.216	0.209
18	0.8182	0.339	0.321	0.305	0.289	0.280	0.268	0.259	0.250
19	0.8636	0.407	0.386	0.367	0.350	0.336	0.323	0.312	0.302
20	0.9091	0.494	0.468	0.445	0.426	0.407	0.391	0.378	0.368
21	0.9545	0.626	0.598	0.568	0.541	0.520	0.501	0.482	0.465

Table 6

Correction due to fire frequency

Serial No. (f)	Year	Number of fires (n_f)	Correction log (n_f/n_1)
1	1947	465	0.000
2	1948	478	0.028
3	1949	512	0.097
4	1950	574	0.211
5	1951	728	0.449
6	1952	568	0.201
7	1953	725	0.445
8	1954	662	0.354
9	1955	740	0.465
10	1956	716	0.432
11	1957	645	0.328
12	1958	560	0.186
13	1959	872	0.629
14	1960	760	0.492
15	1961	696	0.404
16	1962	724	0.443
17	1963	790	0.530
18	1964	998	0.764
19	1965	964	0.730
20	1966	1050	0.815
21	1967	982	0.748

Table 7

Reduced extremes - corrected

Extremes	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957
1	1.605	0.821	-0.297	0.854	1.806	2.551	1.405	-0.771	0.577	-0.259	0.567
2	0.962	0.205	-0.156	1.010	1.581	1.573	1.001	-0.341	0.471	-0.013	1.080
3	0.716	0.531	0.077	0.544	0.291	0.804	1.304	-0.406	0.580	0.125	-0.157
4	0.708	0.368	0.183	0.180	0.011	0.229	1.308	-0.183	0.611	0.341	-0.023
5	0.756	0.150	0.330	0.279	0.125	0.165	0.879	-0.135	0.641	0.289	-0.070
6	0.057	0.288	0.251	0.316	0.208	0.116	0.763	-0.100	0.475	0.297	-0.042
7	-0.283	0.379	0.235	0.396	0.274	0.162	0.450	-0.184	0.234	0.350	-0.098
8	-0.267	0.070	0.058	0.478	0.326	0.121	0.529	-0.155	0.635	0.434	-0.074
9	-0.254	-0.050	0.174	0.460	0.289	0.162	0.602	-0.129	0.767	0.469	-0.055
10	-0.299	0.251	-0.017	0.494	0.410	0.272	0.443	-0.108	0.804	0.540	-0.037
11	-0.351	-0.083	-0.137	0.597	0.410	0.267	0.216	-0.083	0.186	0.571	0.040
12	-0.338	-0.198	-0.047	0.578	0.147	0.297	0.507	-0.074	0.720	0.563	0.051
13	-0.327	-0.190	-0.042	0.561	0.379	0.163	0.606	-0.058	0.590	0.490	0.060
14	-0.259	-0.184	-0.038	0.618	0.443	0.163	0.565	-0.045	0.619	0.519	0.012
15	-0.206	-0.223	0.305	0.534	0.412	0.134	0.593	-0.033	0.649	0.552	0.021
16	-0.243	-0.269	-0.031	0.427	0.164	0.383	0.588	-0.023	0.642	0.453	0.127
17	-0.238	-0.262	-0.030	0.420	0.385	0.165	0.551	-0.013	0.636	0.424	0.133

Table 7 (Cont'd)

Reduced extremes - corrected

Extreme	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967
1	0.503	1.772	-0.382	-0.129	0.945	0.897	0.050	3.808	0.804	2.671
2	-0.159	0.684	-0.392	-0.155	0.365	1.199	0.856	2.504	1.059	0.583
3	0.099	0.887	-0.106	0.175	0.054	1.573	1.179	2.064	1.000	0.795
4	0.034	1.044	-0.189	0.123	0.230	1.124	1.264	1.832	1.086	0.955
5	0.097	0.923	-0.127	0.148	0.246	0.546	1.383	1.702	1.175	1.266
6	0.148	1.090	-0.082	0.104	0.258	0.735	1.317	1.586	1.200	1.416
7	0.559	1.239	0.145	0.275	0.216	0.766	1.267	1.511	1.234	1.038
8	0.312	0.953	0.163	0.237	0.227	0.747	1.328	1.445	1.202	1.218
9	0.185	0.831	0.180	0.286	0.238	0.889	1.287	1.401	1.248	0.864
10	0.219	0.777	0.248	-0.169	0.367	0.763	1.258	1.356	1.222	0.938
11	0.441	0.731	0.302	0.254	0.369	0.709	1.232	1.328	1.084	0.969
12	0.377	0.798	0.309	0.296	0.558	0.438	1.209	1.298	1.120	0.957
13	0.180	0.829	0.315	0.300	0.534	0.556	1.190	1.271	1.104	0.992
14	0.241	0.820	0.320	0.303	1.095	0.554	1.100	1.250	1.095	0.981
15	0.300	0.515	-0.070	0.361	0.436	0.682	1.155	1.231	1.040	1.016
16	0.709	0.237	0.395	0.241	0.436	0.640	1.246	1.108	1.074	1.060
17	0.324	0.706	0.398	0.246	-0.013	0.578	1.132	1.032	1.280	0.998

Table 8

Results

Extreme (m)	a_{mn_1}	b_{mn_1}	$\bar{z}_m$	s^2_{mz}	$\bar{y}_m$	s^2_{my}	$\bar{y}'_m$	$s^2_{my'}$	$\hat{z}_m$	σ^2_{mz}	r_m
1	2.247	5.214	5.634	0.278	0.526	1.203	0.943	1.300	5.656	0.345	0.961
2	1.785	4.829	5.201	0.193	0.246	0.489	0.663	0.543	5.214	0.221	0.945
3	1.626	4.534	4.890	0.174	0.161	0.305	0.578	0.383	4.899	0.179	0.912
4	1.460	4.327	4.693	0.196	0.118	0.221	0.535	0.324	4.702	0.181	0.880
5	1.387	4.113	4.483	0.202	0.096	0.175	0.513	0.282	4.488	0.171	0.853
6	1.424	3.988	4.336	0.187	0.079	0.142	0.496	0.280	4.341	0.157	0.857
7	1.239	3.749	4.145	0.255	0.067	0.121	0.484	0.255	4.145	0.188	0.807
8	1.163	3.564	3.975	0.271	0.059	0.105	0.476	0.253	3.977	0.208	0.830
9	1.212	3.448	3.859	0.224	0.052	0.093	0.469	0.230	3.839	0.174	0.837
10	1.034	3.259	3.728	0.291	0.047	0.083	0.464	0.226	3.711	0.232	0.853
11	0.973	3.137	3.610	0.318	0.043	0.075	0.460	0.211	3.613	0.244	0.838
12	0.925	2.972	3.465	0.328	0.039	0.069	0.456	0.206	3.468	0.262	0.857
13	0.886	2.832	3.341	0.351	0.036	0.064	0.453	0.201	3.347	0.277	0.853
14	0.924	2.749	3.235	0.304	0.033	0.059	0.450	0.187	3.239	0.236	0.847
15	0.937	2.680	3.158	0.283	0.031	0.055	0.448	0.182	3.161	0.223	0.856
16	0.950	2.583	3.052	0.282	0.029	0.052	0.446	0.186	3.055	0.221	0.857
17	1.002	2.537	2.981	0.247	0.027	0.049	0.444	0.180	2.983	0.191	0.851

Table 9.

Return period analysis

| Extreme (m) | | |

Table 10

Characteristic large values for 1967
(At 1947 values)

Extreme (r)	Characteristic large values (b_{rn_j})	a_{rn_j}
1	12.455	2.247
2	12.156	1.785
3	11.902	1.626
4	11.747	1.460
5	11.560	1.387
6	11.421	1.424
7	11.261	1.239
8	11.115	1.163
9	10.973	1.212
10	10.890	1.034
11	10.814	0.973
12	10.689	0.925
13	10.584	0.886
14	10.467	0.924
15	10.386	0.937
16	10.278	0.950
17	10.192	1.002

APPENDIX 2

CORRECTION FOR VARIATION IN SAMPLE SIZE

INTRODUCTION

Extreme value theory is concerned with the distributions of extreme order statistics in repeated samples from a given parent distribution. The classical theory assumes that the sample sizes are maintained at a constant value. This is to ensure that the parametric values of the distributions of extremes remain constant during the process of sampling. However, there may be practical situations where the sizes of the samples vary considerably. In the case of fire losses, for example, the frequency of fires in a year increases significantly over a period of time. In such cases the classical theory needs to be modified. This aspect of the problem is examined in the succeeding sections.

THE FAILURE RATE FUNCTION

By definition

$$F(b_{rn}) = 1 - \frac{r}{n} \tag{1}$$

where $F(z)$ is the parent distribution function and b_{rn} the characteristic r^{th} large value from top in samples of size n from $F(z)$. Also by definition

$$F(z) = 1 - e^{-\int_0^z h(u)\,du} \tag{2}$$

where $h(u)$ is the failure rate function. From (1) and (2),

$$\begin{aligned} H(b_{rn}) &= \int_0^{b_{rn}} h(u)\,du \\ &= \log(n/r) \end{aligned} \tag{3}$$

Also according to fundamental results, if z_{rnj} is the observed r^{th} large value in the j sample of size n, we have

$$F(z_{rnj}) = 1 - \frac{r}{n} e^{-a_{rn}(z_{rnj} - b_{rn})} \tag{4}$$

where

$$a_{rn} = h(b_{rn}) \tag{5}$$

Approximations (1) and (4) which are true for exponential type distributions have been obtained under the assumption that L'Hopital's rule is applicable for large values of 2. According to this rule the critical quotient $Q(z)$ given by

$$Q(z) = \frac{h(z)}{-f'(z)/f(z)} \tag{6}$$

$Q(z)$ tends to unity for large z. For large z the density of probability $f(z)$ becomes very small and the same holds for the probability $\{1 - F(z)\}$ of a value exceeding z. If the variate is unlimited the derivative $f'(z)$ also converges towards zero. From (6) we may write

$$h(z) = \frac{f(z)}{1 - F(z)} \sim -f'(z)/f(z) \tag{7}$$

In (7), $f'(z)$ is the derivative of $f(z)$ and $-f(z)$ the derivative of $\{1 - F(z)\}$

By taking further derivatives we may extend (7) to write that, for large Z,

$$h(z) \sim -\frac{f'(z)}{f(z)} \sim -\frac{f''(z)}{f'(z)} \sim -\frac{f'''(z)}{f''(z)} \cdots\cdots \tag{8}$$

We have, from (7),

$$h'(z) = \frac{f'(z)}{1 - F(z)} - \frac{f(z)}{\{1 - F(z)\}^2}\{-F'(z)\}$$

$$= \frac{f'(z)}{1 - F(z)} + \frac{(f(z))^2}{\{1 - F(z)\}^2}$$

But from (8),

$$f'(z) \sim -\frac{(f(z))^2}{1 - F(z)}$$

Hence $h'(z)$ tends to zero for $z \to \infty$ In the same way

$$h''(z) = \frac{d}{dz}\left\{\frac{f'(z)}{1 - F(z)} + \left(\frac{f(z)}{1 - F(z)}\right)^2\right\}$$

$$= \frac{d}{dz}\left\{\frac{f'(z)\,h(z)}{f(z)} + (h(z))^2\right\}$$

$$= \frac{f'(z)}{f(z)} h'(z) + h(z) \left\{ \frac{f''(z)}{f(z)} - \left(\frac{f'(z)}{f(z)}\right)^2 \right\}$$

$$+ 2\, h(z) . h'(z)$$

The above expression for $h''(z)$ tends to zero since $h'(z)$ does so and from (8)

$$\frac{f''(z)}{f(z)} \sim \left(\frac{f'(z)}{f(z)}\right)^2$$

In the same way it can be shown that all the derivatives of $h(z)$ tend to zero for large z The asymptotic distribution of extreme values for the exponential family has also been derived under the condition that

$$\lim_{z \to \infty} \left[\frac{d}{dz} \left(\frac{1 - F(z)}{f(z)} \right) \right]$$

$$= \lim_{z \to \infty} \frac{d}{dz} \left(\frac{1}{h(z)} \right) = 0$$

The above mentioned property of the derivatives of $h(z)$ is implied in the asymptotic probability of z_{rn} given by (4) with the density function given by expression (1) in the text.

Hence the failure rate function of the parent distribution could be regarded as a constant in the vicinity of any characteristic large value b_{rn} provided r is small compared with large n and z. Small deviations in the value of z around b_{rn} do not appear to produce any significant changes in the value of a_{rn}. However, for $r = 1, 2, \ldots$ the sequence a_{rn} may assume any pattern depending upon the parent since the characteristic large values need not be sufficiently close to each other. In the case of a parent of simple exponential form (ie with density $\mu e^{-\mu t}$ where μ is a constant) the failure rate a_{rn} is equal to the constant μ for all r.

MODIFIED MODEL

Let b_{rnj} be the characteristic r^{th} large value from top in samples of size n_j from $F(z)$. In the neighbourhood of b_{rn} we have

$$H(b_{rnj}) - H(b_{rn})$$
$$= (b_{rnj} - b_{rn})\, H'(b_{rn}) + \frac{(b_{rnj} - b_{rn})^2}{2} H''(b_{rn}) + \cdots\cdots$$
$$= (b_{rnj} - b_{rn})\, h(b_{rn}) + \frac{(b_{rnj} - b_{rn})^2}{2} h'(b_{rn}) + \cdots\cdots$$

From (3), the left hand side of the above equation is equal to log (n_j/n). It has also been proved in the previous section that the derivatives of $h(z)$ are of negligible magnitudes in the vicinity of b_{rn} Hence

$$b_{rnj} = b_{rn} + \frac{1}{a_{rn}} \log(n_j/n) \qquad (9)$$

In an investigation to be undertaken separately it is hoped to evaluate by numerical methods the errors in adopting the first approximation given by (9) for different distributions of exponential type.

We have,

$$Z_{rnj} = b_{rnj} + \frac{y_{rnj}}{a_{rnj}}$$

where y is the reduced value. Hence using (9)

$$Z_{rnj} = b_{rn} + \frac{1}{a_{rn}} \log(n_j/n) + \frac{y_{rnj}}{a_{rnj}}$$
$$= b_{rn} + \frac{y_{rnj} + \log(n_j/n)}{a_{rn}} \qquad (10)$$

since a_{rnj} is equal to the constant value a_{rn} for values of b_{rnj} in the neighbourhood of b_{rn}. The random variable y_{rnj} is independent of the sample size n_j provided it is large. Its value corresponding to z_{rnj} may be obtained by treating n as constant for the samples.

APPENDIX 3

POPULATION VALUES OF REDUCED EXTREMES

The reduced i^{th} large order statistic from top is

$$y_i = \alpha_i (x_i - u_i) \tag{1}$$

where x_i is the observed i^{th} large order statistic, u_i the characteristic i^{th} extreme and α_i the value at u_i of the intensity function of the parent distribution. It is known that the moment generating function $G_i(t)$ of y_i is given by

$$G_i(t) = i^t \Gamma(i-t) / \Gamma(i) \tag{2}$$

Using the Weistrass form of $\Gamma(x)$ we can write

$$\log G_i(t) = t\{\nu + \log i - S'_{1,i}\} + \sum_{k=2}^{\infty} \frac{t^k}{k}(S_k - S'_{k,i}) \tag{3}$$

where ν is Euler's constant and

$$S_k = \sum_{r=1}^{\infty} (1/r^k) \quad \text{and} \quad S'_{k,i} = \sum_{r=1}^{i-1} (1/r^k) \tag{4}$$

with $S'_{k,1} = 0$ Using Bernoulli numbers the approximate value of S_2 is 1.6449. From (3) and (4) we have

$$E(y_i) = \nu + \log i - \sum_{r=1}^{i-1} (1/r)$$

and

$$\sigma_i^2 = \text{variance of } y_i \tag{5}$$

$$= 1.6449 - \sum_{r=1}^{i-1} (1/r^2)$$

With the aid of the expression (5), the expected value and variance of the top 40 large order statistics have been obtained and tabulated below

Order (i)	Extremes (y_i)	
	Expected value	Variance (σ_i^2)
1	0.5772	1.6449
2	0.2704	0.6449
3	0.1758	0.3949
4	0.1302	0.2838
5	0.1033	0.2213
6	0.0857	0.1813
7	0.0731	0.1535
8	0.0637	0.1331
9	0.0565	0.1175
10	0.0508	0.1051
11	0.0461	0.0951
12	0.0422	0.0869
13	0.0390	0.0799
14	0.0362	0.0740
15	0.0337	0.0689
16	0.0316	0.0645
17	0.0297	0.0606
18	0.0280	0.0571
19	0.0265	0.0540
20	0.0252	0.0512
21	0.0240	0.0487
22	0.0228	0.0465
23	0.0219	0.0444
24	0.0210	0.0425
25	0.0201	0.0408
26	0.0193	0.0392
27	0.0186	0.0377
28	0.0179	0.0363
29	0.0173	0.0351
30	0.0167	0.0339
31	0.0162	0.0327
32	0.0156	0.0317
33	0.0151	0.0307
34	0.0147	0.0298
35	0.0143	0.0289
36	0.0138	0.0281
37	0.0134	0.0273
38	0.0131	0.0266
39	0.0128	0.0259
40	0.0125	0.0252

APPENDIX 4

COVARIANCE OF LARGE ORDER STATISTICS FROM EXPONENTIAL TYPE DISTRIBUTIONS

It is necessary to recall first the notation and some of the fundamental results already obtained. We may assume, for the sake of simplicity, that the population under consideration has the distribution function $F(x)$ with density function $f(x)$ which is continuous. Consider the elements of a random sample of size n drawn from the population which are mutually random variables. Rearranging them in decreasing order of their magnitudes we may write

$$X_1 > X_2 > X_3 \cdots\cdots\cdots > X_n \tag{1}$$

If N such samples of size n are drawn from the same parent distribution $F(x)$, the ith order statistics X_i will have a probability distribution with a density function given by

$$\chi_i(x_i)\,dx_i = \frac{n!}{(i-1)!\,(n-i)!}\left[F(x_i)\right]^{n-i}\left[1-F(x_i)\right]^{i-1} f(x_i)\,dx_i \tag{2}$$

We can now define two parameters α_i and u_i with reference to the ith extreme from top as the solutions of

$$F_n(u_i) = 1 - \left(\frac{i}{n}\right) \text{ and} \tag{3}$$

$$\alpha_i = \left(n/i\right) f_n(u_i) \tag{4}$$

Following Gumbel[1] we may expand $F(x)$ for large values of x about the characteristic i^{th} largest value u_i. If $F(x)$ is of the exponential type we may write approximately, for large x.

$$F(x_i) = 1 - \left(\frac{i}{n}\right) e^{-y_i} \text{ and} \tag{5}$$

$$f(x_i) = \left(\frac{i}{n}\right) \alpha_i e^{-y_i} \tag{6}$$

where

$$y_i = \alpha_i (x_i - u_i) \tag{7}$$

With these values, for large n, the density function $\psi_i(x_i)$ tends to

$$\psi_i(x_i) = \{ i^i/(i-1)! \}\, d_i \exp\{-i y_i - i \exp(-y_i)\}\, dx_i \text{ or}$$

$$\psi_i(y_i) = \{ i^i/(i-1)! \} \exp\{-i y_i - i \exp(-y_i)\}\, dy_i \tag{8}$$

for $-\infty \leq y_i \leq \infty$. For the largest value $i = 1$ we obtain the density

$$\psi_1(y_1) = \exp\{-y_1 - \exp(-y_1)\}\, dy_1 \tag{9}$$

with the distribution function

$$\Phi_1(y_1) = \exp\{-\exp(-y_1)\} \tag{10}$$

3. COVARIANCE

If the sample observations are arranged in decreasing order of magnitude as in (1) the joint distribution of the ith and j^{th} order statistics with $i > j$ could be written as

$$\frac{n!}{(n-i)!\,(i-j-1)!\,(j-1)!} \left[F(x_i)\right]^{n-i} \left[F(x_j) - F(x_i)\right]^{i-j-1} \left[1 - F(x_j)\right]^{j-1}$$

$$\cdot\, f(x_i) \cdot f(x_j)\, dx_i\, dx_j \tag{11}$$

which is true for the domain $x_i < x_j$. With the aid of the values in (5) and (6) we may write the asymptotic form of (11) as

$$\frac{i\, j^j}{(i-j-1)!\,(j-1)!} \exp\{-y_i - j y_j - i \exp(-y_i)\} \left(i e^{-y_i} - j e^{-y_j}\right)^{i-j-1} dy_i\, dy_j \tag{12}$$

for large values of n, x_i and x_j. By writing

$$\xi_i = i e^{-y_i} \quad \text{and} \quad \xi_j = j e^{-y_j} \tag{13}$$

we can rewrite (12) more elegantly as

$$\psi_{i,j} = \frac{1}{(i-j-1)!(j-1)!} e^{-\xi_i} (\xi_i - \xi_j)^{i-j-1} \xi_j^{j-1} \, d\xi_i \, d\xi_j \qquad (14)$$

which is true for the domain $0 \leq \xi_j \leq \xi_i \leq \infty$
since $i > j$ and $F(x_i) < F(x_j)$ expression (5)

Denoting the expected value by the letter E the covariance σ_{ij} of y_i and y_j is given by

$$\sigma_{ij} = E(y_i y_j) - E(y_i) E(y_j) \qquad (15)$$

From (14),

$$E(y_i y_j)$$

$$= \frac{1}{(i-j-1)!(j-1)!} \int_0^\infty e^{-\xi_i} \log(i/\xi_i) \, d\xi_i \int_0^{\xi_i} (\xi_i - \xi_j)^{i-j-1} \xi_j^{j-1} \log(j/\xi_j) \, d\xi_j \qquad (16)$$

The evaluation of the integral in (16) is shown in appendix 4(a)
We have:-

$$E(y_i y_j) = E(y_i^2) + E(y_i) \left\{ \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) - \log(i/j) \right\} \qquad (17)$$

But from (5) of Appendix 3

$$E(y_i) = \nu + \log i - \sum_{r=1}^{i-1} \left(\frac{1}{r}\right) \text{ and}$$

$$E(y_j) = \nu + \log j - \sum_{r=1}^{j-1} \left(\frac{1}{r}\right)$$

Hence

$$E(y_j) - E(y_i) = \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) - \log(i/j) \qquad (18)$$

From (15), (17) and (18) we obtain the covariance

$$\sigma_{ij} = E(y_i^2) - \{E(y_i)\}^2$$

$$= \text{variance of } y_i$$

Therefore, the covariance of the extreme order statistics y_i and y_j is the same as the variance of y_i where $i > j$. This result is analogous to the covariance for order statistics from the simple exponential distribution obtained by Sarhan[2]. Greenberg and Sarhan have also tabulated the expected values and the variances and covariances of the order statistics in samples of size ≤ 10 from the exponential distribution[3].

REFERENCES

1. GUMBEL, E.J. (1958) Statistics of Extremes. Columbia University Press, New York.
2. SARHAN, A.E. (1954). Estimation of the mean and standard deviation by order statistics. Ann. Math. Statist, 25, 317-328.
3. SARHAN, A.E. and GREENBERG, B.G. (1958) Estimation problems in the exponential distribution using order statistics. Proceedings of the Statistical Techniques in Missile Evaluation Symposium, Blacksburg, Va, 123-175.

APPENDIX 4(a)

Consider

$$I_1 = \int_0^{\xi_i} (\xi_i - \xi_j)^{i-j-1} \xi_j^{\,j-1} \log(j/\xi_j)\, d\xi_j$$

Making the substitution $\xi_j = \xi_i \eta_j$

$$I_1 = \xi_i^{i-1} \Big\{ \log(j/\xi_i) \int_0^1 \eta_j^{j-1} (1-\eta_j)^{i-j-1} d\eta_j$$
$$- \int_0^1 \eta_j^{j-1} (1-\eta_j)^{i-j-1} \log \eta_j \, d\eta_j \Big\}$$
$$= \xi_i^{i-1} \Big\{ \log(j/\xi_i)\, B(j, i-j) - I_2 \Big\}$$

where B denotes the beta function and

$$I_2 = \int_0^1 \eta_j^{j-1} (1-\eta_j)^{i-j-1} \log \eta_j \, d\eta_j$$
$$= \sum_{r=0}^{m} {}_m C_r (-1)^r \int_0^1 \eta_j^{j+r-1} \log \eta_j \, d\eta_j$$

where $m = i-j-1$. Hence

$$I_2 = - \sum_{r=0}^{m} {}_m C_r (-1)^r / (j+r)^2 \qquad (1)$$

It can be proved numerically and otherwise that the series on the right hand side of (1) is equivalent to

$$- B(j, i-j) \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) \qquad (2)$$

Hence

$$I_1 = B(j, i-j)\, \xi_i^{i-1} \left\{ \log(j/\xi_i) + \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) \right\}$$

Now

$$E(y_i y_j) = \frac{1}{B(j, i-j)\,\Gamma(i)} \int_0^\infty e^{-\xi_i} \log(i/\xi_i)\, I_1\, d\xi_i$$

$$= \frac{1}{\Gamma(i)} \left[\int_0^\infty e^{-\xi_i} \xi_i^{i-1} \log(j/\xi_i) \log(i/\xi_i)\, d\xi_i + \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) \int_0^\infty e^{-\xi_i} \xi_i^{i-1} \log(i/\xi_i)\, d\xi_i \right]$$

$$= \frac{1}{\Gamma(i)} \left[I_3 + \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) I_4 \right] \qquad (3)$$

Now

$$I_3 = \int_0^\infty e^{-\xi_i} \xi_i^{i-1} \log(j/\xi_i) \log(i/\xi_i)\, d\xi_i$$

$$= \log i \log j\, \Gamma(i) - (\log i + \log j) \int_0^\infty e^{-\xi_i} \xi_i^{i-1} \log \xi_i\, d\xi_i + \int_0^\infty e^{-\xi_i} \xi_i^{i-1} (\log \xi_i)^2\, d\xi_i$$

But

$$\int_0^\infty e^{-\xi_i} \xi_i^{i-1} \log \xi_i\, d\xi_i$$

$$= \log i \int_{-\infty}^\infty i^i e^{-iy_i - ie^{-y_i}}\, dy_i - \int_{-\infty}^\infty i^i e^{-iy_i - ie^{-y_i}}\, y_i\, dy_i$$

$$= \Gamma(i) \left\{ \log i - E(y_i) \right\} \qquad (4)$$

Expression (4) follows from (8) and (13) in Appendix 4. Similarly it can be shown that

$$\int_0^\infty e^{-\xi_i} \xi_i^{i-1} (\log \xi_i)^2 \, d\xi_i$$
$$= \Gamma(i) \left\{ (\log i)^2 - 2 \log i \, E(y_i) + E(y_i^2) \right\}$$

Hence, after simplifying

$$I_3 = \Gamma(i) \left\{ \log (j/i) E(y_i) + E(y_i^2) \right\} \tag{5}$$

and

$$I_4 = \Gamma(i) \left\{ \log i - \log i + E(y_i) \right\}$$
$$= \Gamma(i) \, E(y_i) \tag{6}$$

It follows from (3), (5) and (6) that

$$E(y_i y_j) = E(y_i^2) + E(y_i) \left\{ \sum_{r=j}^{i-1} \left(\frac{1}{r}\right) - \log(i/j) \right\}$$

APPENDIX 5

EXPECTED LOSS IN SMALLER FIRES

From expression (2), appendix 2, the density function of the parent distribution is

$$f(z) = h(z)\, e^{-\int_0^z h(u)\,du}\, dz \qquad (1)$$

where $h(u)$ is the failure rate or intensity function. If $h(u)$ is of the form

$$h(u) = e^{\alpha+\beta u} \qquad (2)$$

it is easily seen that

$$f(z) = K e^{\alpha+\beta z - \frac{e^{\alpha+\beta z}}{\beta}}\, dz \qquad (3)$$

where

$$K = e^{\frac{e^{\alpha}}{\beta}} \qquad (4)$$

In (3), put

$$e^{\alpha+\beta z} = \beta \xi \qquad (5)$$

so that

$$f(\xi) = K e^{-\xi}\, d\xi \qquad (6)$$

Since $0 \leq z \leq \infty$, we have $\frac{e^{\alpha}}{\beta} \leq \xi \leq \infty$.

It may be verified that the integral of (6) over the prescribed range is unity so that $f(\xi)$ can represent a density function. If $e^{\alpha/\beta}$ is almost zero K may be assumed to be equal to unity.

From text it may be observed that Z is the logarithm of the loss x. Hence from (5)

$$\log \beta + \log \xi - \alpha = \beta \log x \quad \text{or}$$

$$x = (\beta \xi \alpha')^{1/\beta} \tag{7}$$

where

$$\alpha' = e^{-\alpha} \tag{8}$$

We are neglecting values of x less than x_0 as due to infant mortality. Also let x_e correspond to the amount £10,000 corrected for inflation. The expected value of x is required for the range $x_0 \leq x \leq x_e$. If $z_e = \log x_e$ and $z_0 = \log x_0$, the corresponding upper and lower limits for ξ are

$$\xi_e = e^{\alpha + \beta z_e} / \beta \quad \text{and} \tag{9}$$

$$\xi_0 = e^{\alpha + \beta z_0} / \beta \tag{10}$$

The density function in the range $\xi_0 \leq \xi \leq \xi_e$ is given by

$$f_e(\xi) = C_e e^{-\xi} d\xi \tag{11}$$

The integral of (11) over the given range ought to be unity so that

$$C_e = \frac{1}{e^{-\xi_0} - e^{-\xi_e}} \tag{12}$$

If $\bar{x}_e$ is the expected value in the range $x_0 \leq x \leq x_e$ from (7),

$$\bar{x}_e = (\beta\alpha')^{\frac{1}{\beta}} \underset{e}{E}(\xi^{1/\beta}) \tag{13}$$

But from (11),

$$\begin{aligned} \underset{e}{E}(\xi^{1/\beta}) &= C_e \int_{\xi_0}^{\xi_e} e^{-\xi} \xi^{1/\beta} d\xi \\ &= C_e \left[\int_0^{\xi_e} e^{-\xi} \xi^{1/\beta} d\xi - \int_0^{\xi_0} e^{-\xi} \xi^{1/\beta} d\xi \right] \end{aligned} \tag{14}$$

The ratios

$$p_e = \frac{\int_0^{\xi_e} e^{-\xi} \xi^{1/\beta} d\xi}{\int_0^{\infty} e^{-\xi} \xi^{1/\beta} d\xi}$$

and

$$p_0 = \frac{\int_0^{\xi_0} e^{-\xi} \xi^{1/\beta} d\xi}{\int_0^{\infty} e^{-\xi} \xi^{1/\beta} d\xi}$$

could be obtained from Tables of Incomplete Gama Functions. Thus

$$\int_0^{\xi_e} e^{-\xi} \xi^{1/\beta} d\xi = p_e \Gamma\left(1 + \frac{1}{\beta}\right) \tag{15}$$

and

$$\int_0^{\xi_0} e^{-\xi} \xi^{1/\beta} d\xi = p_0 \Gamma\left(1+\frac{1}{\beta}\right) \tag{16}$$

From (13), (14), (15) and (16)

$$\bar{x}_e = (p_e - p_0)\, c_e (\beta\alpha')^{1/\beta} \Gamma\left(1+\frac{1}{\beta}\right) \tag{17}$$

From (7),

$$\underset{e}{E}(x^2) = (\beta\alpha')^{2/\beta} \underset{e}{E}\left(\xi^{2/\beta}\right)$$

Following the algebra in (14)

$$\underset{e}{E}(x^2) = (q_e - q_0)\, c_e (\beta\alpha')^{2/\beta} \Gamma\left(1+\frac{2}{\beta}\right) \tag{18}$$

where

$$q_e = \frac{\int_0^{\xi_e} e^{-\xi} \xi^{2/\beta} d\xi}{\int_0^{\infty} e^{-\xi} \xi^{2/\beta} d\xi}$$

and

$$q_0 = \frac{\int_0^{\xi_0} e^{-\xi} \xi^{2/\beta} d\xi}{\int_0^{\infty} e^{-\xi} \xi^{2/\beta} d\xi}$$

The variance of x in $x_0 \leq x \leq x_e$ is given by

$$\sigma_e^2 = E_e(x^2) - [E_e(x)]^2$$
$$= E_e(x^2) - \bar{x}_e^2$$
$$= \text{expression (18)} - (\text{Expression (17)})^2 \qquad (19)$$

If we consider n fires within the range $x_0 \leq x \leq x_e$ the total T_n of the losses in these n fires would have the expected value of $n\bar{x}_e$ and standard error $\sqrt{n} \cdot \sigma_e$. For the variance of T_n is

$$Var\, T_n = Var(x_1 + \cdots x_n)$$
$$= n\, Var(x)$$
$$= n \cdot \sigma_e^2$$