The Probability Distribution of Fire Loss Amount

Author(s): David C. Shpilberg
Source: *The Journal of Risk and Insurance*, Vol. 44, No. 1 (Mar., 1977), pp. 103-115
Published by: American Risk and Insurance Association
Stable URL: http://www.jstor.org/stable/251860
Accessed: 23-05-2017 11:11 UTC

The Probability Distribution of Fire Loss Amount

David C. Shpilberg

ABSTRACT

Theoretical distributions frequently used to model fire loss amount are discussed. The problem of selecting models solely on the basis of statistics is addressed. Use of probabilistic arguments applied in Reliability theory to infer the type of probability distribution, is explored. The concept of failure rate of a fire is discussed and used to explore implications of the Pareto and Lognormal models as to the fire growth phenomenon. It is concluded that probabilistic arguments, regarding the nature of the fire growth process can aid analysts in their choice of an appropriate model for the probability distribution of fire loss amount.

It is a basic assumption in all actuarial research and risk theory studies that there is a probability distribution of loss amount underlying the risk process. In other words, if a loss occurs, there is the probability $S(x)$ that the loss will be for an amount less than or equal to x. In theoretical studies this distribution often is presented as continuous, having a derivative $S'(x) = s(x)$, which is called the probability density function of fire loss amount.[1]

At a certain point in time, the results of the theoretical work have to be applied to practical situations. For example, the distribution of actual losses experienced by an insurer is then considered as a sample from an underlying model without defining the corresponding distribution, the characteristics of which are taken to agree with the corresponding statistics of the observed distribution. Many results can be obtained simply by using these sample statistics. However, it is often more desirable to work with analytically defined loss distributions, and the statistics are then used to establish suitable values of the parameters involved in the analytical distributions. When working with these analytical distributions, the researcher must make use of properties of the distributions other than those covered by the statistics observed.

There are two main aspects of general insurance in which a knowledge

David Shpilberg, Ph.D., is Associate Professor of Operations Research and Insurance at the Instituto de Estudios Superiores de Administración (IESA), a Graduate School of Management in Caracas, Venezuela. The research for this paper was partly financed by a grant of the Factory Mutual Research Corporation. The paper was presented at the 1976 annual meeting of The American Risk and Insurance Association. Dr. Shpilberg received the 1975 *Journal of Risk and Insurance* Award for the best paper published in the 1975 issues.

of the structure of the elements of risk variation is needed:[2] first, in the rate making process; and second, in dealing with the question of financial stability (monetary risk). Traditional methods of rate making are based only on an estimate of the mean expected loss. Financial stability studies (e.g., studies addressing the probability of ruin of an insurer or evaluating the risk of unbearable monetary loss for a corporation which chooses not to insure its property) usually are based on an estimate of the variance of the possible loss. However, in the area of industrial fire losses, the probability distributions involved are markedly skewed in character (very small probabilities of a very large loss).

Knowledge of its higher moments (in essence, the shape of the tail of the distribution) becomes essential if meaningful quantitative estimates of risk are to be made. Most often, this step involves assumptions regarding the behavior of losses larger than those observed in the sample of available loss experience. Thus, unless there is some theoretical support (not merely observed statistics) for an inference that a particular type of probability distribution is a more reasonable model for the distribution of fire loss amounts as a function of size, inferences derived for any region of the distribution outside the available data will be no better than a straight extrapolation on the data.

This paper presents a sumary review of work in the area of modeling the probability distribution of fire loss amount, and attempts to illustrate how probabilistic arguments relating to the physical nature of the phenomenon (an approach extensively used in life testing of material failure and in reliability analysis of systems' components) can effectively aid in the choice of an appropriate model.

Fire Loss As a Stochastic Process

The total amount of fire losses in a given period can be modeled as a risk process characterized by two stochastic variables: the number of fires and the amount of the losses. If,

$P_r(t)$ = Probability of r losses in the observed period, t
$S(x)$ = Probability that, given a fire loss, its amount is $\leq$ x
$S^{*r}(x)$ = r^{th} convolution of the distribution function of fire loss amount, $S(x)$,

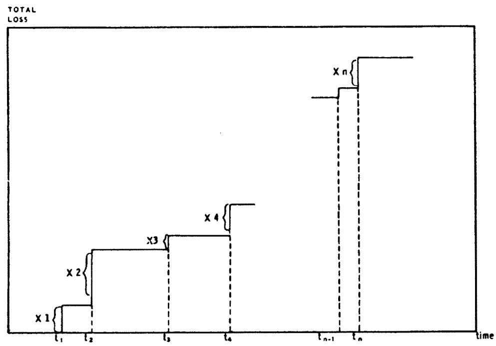
then the probability (see Figure 1) that total loss in a period of length t, is $\leq$ x can be expressed as

$$F(x,t) = \sum_{r=0}^{\infty} P_r(t) S^{*r}(x) \tag{1}$$

a compound distribution whose mean and variance are given by:

$$\begin{aligned} m_F &= m_p m_s \\ \sigma^2{}_F &= m_p \sigma_s^2 + m_s{}^2 \sigma^2{}_p \end{aligned} \tag{2}$$

FIGURE 1
STOCHASTIC PROCESS OF LOSS ACCUMULATION IN TIME



t_i ($i = 1,2,\ldots\ldots n$) = Independent random occurrence of random size X_i. (fires)
x_i ($i = 1,2,\ldots\ldots n$) = Independent identically distributed random variables of random length distributed according to S (x) (loss amounts per fire)

$$F(x,t) = \sum_{r=0}^{\infty} P_r(t)\, S^{*r}(x)$$

where m_p and m_s are the mean number of losses and the mean loss per fire, and σ_p^2 and σ_s^2 are the corresponding variances.[3]

While the total loss distribution, F(x, t) is of great importance for insurers in their task of determining appropriate premiums, reinsurance policies, credibility measures and contingency retention levels, it is the probability distribution of an individual fire loss amount, S(x), (Figure 2) that is relevant when the decisions to purchase insurance and/or invest in fire protection are to be made by a property owner or when deductible schedules are to be designed by an insurer. This paper is concerned with the choice of an appropriate model for this distribution*.

Prior Work

Several theoretical distributions have been postulated by researchers in the field as models for the fire loss amount distribution, S(x). They are all relatively successful in modeling the mid-range (the area around the median) of the distribution, but their adequacy to model the region of the larger values of x (tail) cannot be determined with confidence from observed data. Paradoxically, the least known part of S(x) has the greatest effect on the numerical results of many practical problems such as optimal investment decisions in insurance and/or fire protection.

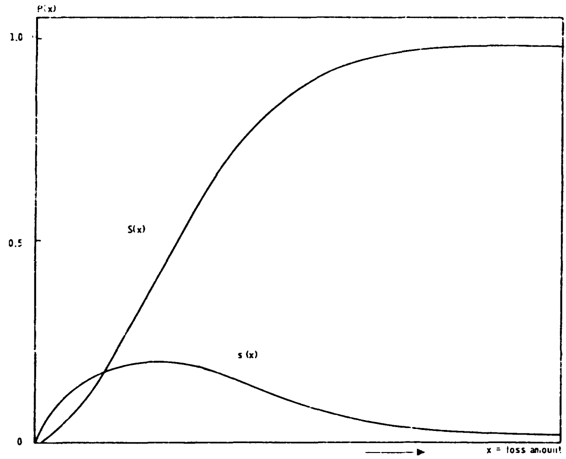
A more than adequate review of the types of distributions used in the past to model S(x) is contained in a recent paper by Benktander[5]. The following are the models that have appeared most often in the recent literature:

Exponential Distribution

Beard, Pentikainen and Pesonen[4] suggest that since loss distributions show the highest frequency for the small losses, the frequency declining

* The fire frequency phenomenon, $P_r(t)$ has been treated with success by several authors as a Poisson process with rate of occurrence σ = observed fire frequency for a fixed time interval T.[4]

FIGURE 2
DISTRIBUTION OF FIRE LOSS AMOUNT



with increasing loss size, the exponential function would provide at least a first approximation for S(x).
That is:

$$S(x) = 1-e^{-\nu x} \qquad (\text{for } x \geq 0) \tag{3}$$

where: mean $= 1/\nu$

variance $= 1/\nu^2$

However, the authors are prompt to point out that the use of an exponential function as a model for S(x) can only be a crude approximation to the truth and rarely is applicable.

Log-Normal Distribution

The log-normal distribution model is favored by various studies for a diverse variety of types of insurance. Among others, Beard[2] uses it for analyzing fire losses in Denmark; Benckert[6] for industrial and non-industrial fire losses and business interruption and accident insurance in Sweden; Ferrara[7] for industrial losses in Italy; and Ramachandran[8] for fire losses in England.

The probability density function (the cumulative distribution cannot be written in a close form expression) for the amount of loss would be (Figure 3):

$$s(x) = \frac{dS(x)}{dx} = \frac{1}{x\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(\log x - m)^2\} \tag{4}$$

where:

$$\text{mean} = \exp\{m + \frac{\sigma^2}{2}\}$$

$$\text{variance} = \{\exp\{2m + \sigma^2\}\} \cdot \{\exp\{\sigma^2\} - 1\}$$

Pareto Distribution

A significant number of researchers suggest the Pareto distribution as a plausible model for fire loss amount. Benckert and Sternberg[9] postulated the Pareto law for the distribution of fire losses in Swedish homes. Mandelbrot[10] derived the Pareto law from the assumption that the probability of the fire increasing its intensity at any instant of time is constant. More recently, Anderson[11] used the Pareto distribution to model fire losses in the Scandinavian countries and Benckert and Jung[12] used both the Pareto and the log-normal models for the distribution of fire insurance claims in Swedish dwellings.

The Pareto cumulative distribution would be (Figure 4).

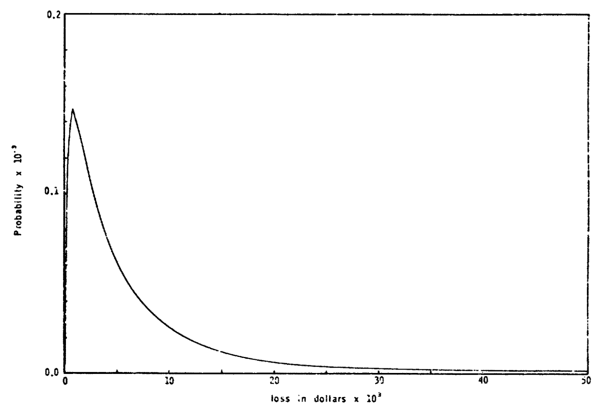
$$S(x) = 1 - \{\frac{x}{x_o}\}^{-\alpha} \qquad \text{for } x \geq x_o$$

where:

$$\text{mean} = \frac{x_o\alpha}{\alpha-1} \qquad \text{for} \quad \alpha > 1$$

$$\text{variance} = \frac{x_o^2 d}{(\sigma-1)^2(\sigma-2)} \qquad \text{for} \quad \sigma > 2$$

FIGURE 3
LOGNORMAL PROBABILITY DENSITY FUNCTION



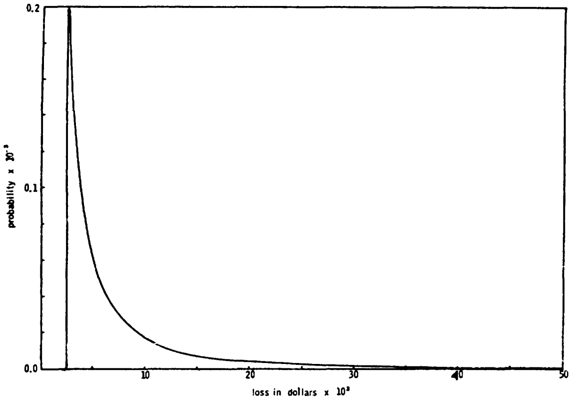
Other Distributions

Other distributions that have been less commonly suggested for modeling loss amounts include: the log-Pearson Type I, also tested by Beard,[2] an infinitive variance left-bounded stable distribution postulated by Holcomb[13]; and a gamma distribution of the logarithm of loss used by Hewitt[14] to fit auto property damage losses.

The Failure Rate Function and the Distribution of Fire Loss Amounts

Researchers on the area of Reliability Theory, the subfield of Applied Probability Theory and Statistics concerned with the development of

FIGURE 4
PARETO PROBABILITY DENSITY FUNCTION



mathematical models and methods directed towards the solution of problems in predicting, estimating or optimizing the life distribution of systems,[15] have in the past addressed the problem of distinguishing among various nonsymmetrical probability distributions given sparse observations at the right hand tail. When attempting to rely on actual observations of time to failure of equipment for inference from the data the type of probability distribution that best models the phenomenon, one faces the problem that the general type of distributions that are most appropriate (nonsymmetric, skewed to the left) are different from each other at the right hand tail, where observations are sparse.

Reliability theory appeals to a concept that permits one to distinguish between the different distribution functions on the basis of a physical consideration: the concept of the "failure rate", also known in the literature of reliability as the "hazard rate", in actuarial studies as "force of mortality" and in extreme-value theory as "intensity function". Good references to the subject are found in books by Barlow and Proschan[16, 17] and Mann, Schafer and Singpurwalla[15].

If $S_T(t)$ is the cumulative distribution function of the time to-failure (i.e. time to extinguishment) of random variable T (i.e. fire), and $s_T(t)$ is its probability density function, then the failure rate, $r(t)$, is defined as:

$$r(t) = \frac{s_T(t)}{1-S_T(t)} \qquad (6)$$

$r(t)dt$ is the conditional probability that the system fails (i.e. fire stops) in the time interval $(t, t + dt)$, given that it was functioning (i.e. burning) up until time t. A distribution $S(t)$ is called IFR (i.e., with increasing failure rate) if $r(t)$ is increasing in t, and DFR (i.e. with decreasing failure rate) if $r(t)$ is decreasing in t.[16]

If one assumes that the duration of a fire can be considered a negative exponential process for a particular type of buildings of a size much larger than the observed average loss, the cumulative distribution of the fire duration would be given by:

$$S_T(t) = P\{T \leq t\} = 1-e^{-\nu t} \tag{7}$$

$$\text{Where:} \quad \nu = \frac{1}{E(T)}$$

Let x_t be the loss associated with a fire of duration t. It is assumed that for a homogenous group of risks the fire loss increases exponentially with the duration of the fire.*
Therefore

$$x_t \propto e^{kt} \tag{8}$$

where the coefficient k would vary for different homogenous groups. Expression (8) can be rewritten as

$$x_t = Ce^{kt} \tag{9}$$

where C is the constant of proportionality and from boundary conditions,

when $t = 0$.

$$x_0 = C \tag{10}$$

where x_0 is the minimum discernable loss associated with fire of minimum duration defined as $t = 0$.
Therefore

$$x_t = x_0\, e^{kt} \tag{11}$$

One can now drop the subscript from x_t since there is a one to one correspondence between the loss amount and the time of fire,
thus

$$\frac{x}{x_0} = e^{kt} \qquad \text{for} \quad x \geq x_0 \tag{12}$$

Taking ln in (12) and replacing in (7) one finds that the distribution of fire loss amount would follow the Pareto law:

$$S(x) = 1-e^{-\left(\frac{\nu}{k}\right)\ln\left(\frac{x}{x_0}\right)} = 1 - \left(\frac{x}{x_0}\right)^{-\nu^*} \qquad \text{for} \quad x \geq x_0 \tag{13}$$

$$\text{where} \quad \nu^* = \frac{\nu}{k}$$

But (7) implies a constant probability of survival for the fire (constant failure rate)

---

* The assumption of exponential fire growth has been empirically tested by Baldwin[18] and others. It is more accurate for small fires than for large ones, where the conditions of the fire boundary tend to become more significant (availability of combustibles, oxygen, fire fighting).

$$r(t) = \frac{\nu e^{-\nu t}}{e^{-\nu t}} = \nu \tag{14}$$

This observation could be true only (if ever) during the early part of the development of a fire, since once the fire is in full development, supplies of fuel and oxygen will be gradually exhausted and fire-fighting probably would have started.[19] Therefore a model that would reflect a gradual decrease in the probability of survival of the fire (increasing failure rate) would appear to be better fitted to reflect the behavior of the fire process.

Given a functional form of r(t), the probability distribution of fire duration implied by the failure rate could be easily determined:

$$\text{since: } \frac{d}{dt} S_T(t) = s_T(t) \tag{15}$$

equation (6) can be rewritten as:

$$r(t) = \frac{\frac{d}{dt} S_T(t)}{1-S_T(t)} = -\frac{d}{dt} \{ \ln (1 - S_T(t)) \} \tag{16}$$

integrating $\int_0^t r(\tau)\, d\tau = \ln \{ 1 - S_T(t) \}$ (17)

from which one may derive

$$S_T(t) = 1-e^{-\int_0^t r(\tau)\, d\tau} \tag{18}$$

where r(t) again is the failure rate, conditional rate of extinguishing the fire at time t, given that the fire is still in progress. Expression (7) is a special case of (18) for $r(t) = \nu$ constant.

The author has stated that a distribution with an increasing failure rate seems more intuitively appealing than one with a constant or decreasing failure rate; hence, assume r(t) to be a monotonically increasing function of t. The simplest function to test would be

$$r(t) = a + bt \qquad \text{for } a > 0,\ b > 0 \tag{19}$$

although r(t) could be generalized to be a n-degree polynomial in t or an exponential function*. If one substitutes expression (19) into expression (18) one obtains:

$$S_T(t) = 1 - e^{-(at + \frac{b}{2} t^2)} \tag{20}$$

and if one incorporates the assumption that fire loss grows exponentially with time, (12), a new cumulative distribution function for the fire loss amount is obtained:

$$S(x) = 1- \left(\frac{x}{x_o}\right)^{-(\alpha + \beta \ln x)} \qquad \text{for } x \geq x_o \tag{21}$$

---

* Ramachandran[20] has recently suggested an exponentially increasing failure rate $r(t) = e^{\alpha+\beta t}$. This, when coupled with the assumption of exponential growth with time of fire loss amount, leads to a Weibull distribution for x, the fire loss amount.

Expression (21), as Dumouchel and Olshen[21] noted, closely approximates a log-normal model. In this discussion, it is called the quasi-log-normal distribution.

Some Statistical Evidence

Head[22] fitted the cumulative probability of loss obtained from over 130,000 fire losses which occurred over a two-year span, against linear and quadratic functions of x and then again against linear and quadratic functions of lnx, including:

$$\ln y = a + b\,(\ln x) \tag{22}$$

and

$$\ln y = a + b(\ln x) + c(\ln^2 x) \tag{23}$$

He states that the best fit was yielded by the equation:

$$\ln y = 2.5972 - .9909 \ln x - .1226 \ln^2 x \tag{24}$$

where y was the fraction of losses larger than x:

$$y = 1 - S(x) \tag{25}$$

Note that equation (24) is equivalent to the quasi-log-normal distribution derived in the preceding section:

$$S(x) = 1 - \left(\frac{x}{x_o}\right)^{-(\alpha+\beta \ln x)} \qquad \text{For } x \geq x_o \tag{26}$$

where:

$$\alpha = 2.5972/\ln(x_o) = 1.2464$$

$$\beta = .1226$$

$$x_o = \exp\{(\alpha - .9909)/\beta\} = 8.0346$$

and expression (22) is equivalent to the Pareto distribution

$$S(x) = 1 - \left(\frac{x}{x_o}\right)^{-\nu} \tag{27}$$

where

$$\nu = -b$$

$$x_o = \exp\{-a/b\}$$

Ramachandran[19] found that, for fire losses in the United Kingdom, the quasi-log-normal distribution

$$\phi(x) = 1 - S(x) = x^{-.0778 - .1052 \log X} \tag{28}$$

fitted his data on textile mill fires better than the Pareto distribution (for $x_o = 1$)

$$\phi(x) = 1 - S(x) = x^{-.4291} \tag{29}$$

The author[23] fitted both the quasi-log-normal and the Pareto models to estimated cumulative probability distributions of industrial fire losses generated through a threshold linear logit model for different uniform types of risks. As it was the case with Head and Ramachandran, the estimation method used was a least squares regression on the linearized forms of the models, expressions (22) and (23), respectively.* For sprinklered fire resistive buildings serviced by professional fire departments and subjectively rated "Better than Fair" by an insurer's inspector, he obtained the quasi-log-normal distribution

$$S(x) = 1 - \{\frac{x}{28.22}\}^{0.645 - 0.098 \ln x} \tag{30}$$

and the Pareto distribution

$$S(x) = 1 - \{\frac{x}{1212.9}\}^{-0.75} \tag{31}$$

Figure (5) illustrates both cumulative distributions. They appear similar except that the quasi-log-normal presents a less dangerous tail (it approaches 1 faster than the Pareto). If one plots the estimated data on log-probability paper and compares it with the best-fit-log-normal distribution, as is done in Figure (6), it can be noticed that the log-normal distribution tends to underestimate the tail of the observed data. On the other hand, if one plots the estimated data on logarithmic paper and compares it with the best-fit- Pareto distribution, as is done in Figure (7), it can be noticed that the Pareto distribution tends to overestimate the tail of the observed data. From observation of the two plots it is not clear which distribution provides an overall best fit.

It is not surprising that one cannot say, based purely on statistics, which distribution provides the best fit, for that problem after all was the motivation for this paper. However, given that both distributions, provide a statistically good fit, one can discern among them by how well they satisfy the physical argument regarding fire growth. That the Pareto distribution appears to overestimate the tail of the actual distribution strengthens the beliefs as to the failure rate of the fire phenomenon. The Pareto distribution implies a constant probability of survival of the fire. Hypothesized in the previous section was that fire has an increasing failure rate, that is, its probability of survival decreases with time. If this hypothesis is correct, the Pareto distribution should overestimate the tail of the actual distribution of fire loss amounts. That is exactly what is observed in Figure 7. Thus, basing the arguments both on the statistical fit and on the hypothesis regarding the failure rate of the fire phenomenon, it would be appropriate to select the log-normal distribution of fire loss amount when analysing fire loss data.

* It should be pointed out that estimation of the parameters of Pareto and log-normal distributions obtained by least squares methods have been shown to be consistent[24] but tend to give biased estimates when fitted to cumulative data since the dependent values 1-S(x) are not independent random observations. Thus, when precise results are required, the results from least squares regression methods should be used as preliminary estimates to be improved, for example, through iterative techniques of Maximum Likelihood Estimation.

FIGURE 5

Cumulative Analytical Distribution of Fire Loss Amount for Sprinklered Fire Resistive Machine Shops Served by a Paid Fire Department and Rated Better than Fair

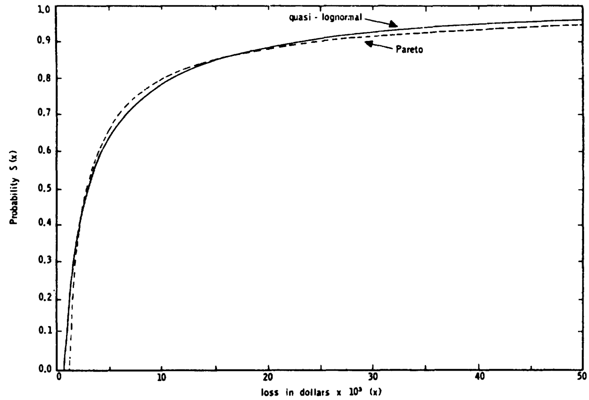


FIGURE 6

Estimated Cumulative Distribution of Fire Losses for Sprinklered Fire Resistive Machine Shops Served by a Professional Fire Department and Rated Better than Fair

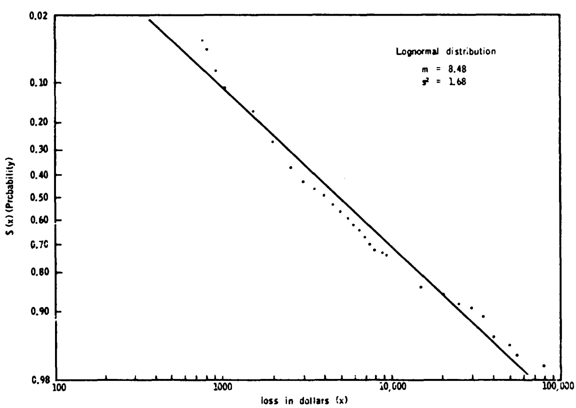
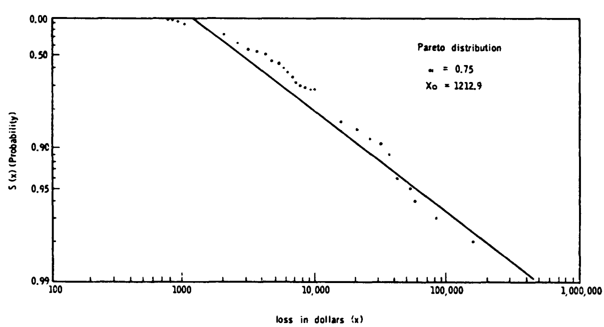


FIGURE 7

Estimated Cumulative Distribution of Fire Losses for Sprinklered, Fire Resistive Machine Shops Served by a Professional Fire Department and Rated Better than Fair



It should be clear that further research into the topic is required. The assumption of exponential growth with time of the fire loss amount should be explored further and alternative assumptions tested. Also, other IFR (increasing failure rate) distributions, such as the Weibull and the Fischer-Tippet extreme value distributions could be tested as plausible models for fire duration. Consideration also should be given to the study of the practical implications of the choice of probability distribution in fire insurance risk analysis and industrial risk management. It would be of interest for example, to study the sensitivity of the risk measures used in a particular analysis to changes in the assumption as to the failure rate of the phenomenon.

REFERENCES

1. Benktander, G. and Segerdahl, C. O., "On The Analytical Representation of Claim Distributions With Special Reference to Excess of Loss Reinsurance," *Transactions of the 16th International Congress of Actuaries*, Brussels, June 1960.
2. Beard, R. E., "Analytical Expressions of the Risks Involved in General Insurance," *Transactions of the 15th International Congress of Actuaries*, Volume 2, 1958.
3. Dickerson, O. D., Katti, S. K. and Hofflander, A. E., "Loss Distributions in Non-Life Insurance," *Journal of Insurance*, Volume 28, No. 3, 1961.
4. Beard, R. E., Pentikainen, T. and Pesonen, E., *Risk Theory*, Methuen and Co., Ltd., London, 1969.
5. Benktander, G., "The Distribution of Claims According to Size in Non-Life Insurance," unpublished paper, Zurich, February 1974.
6. Benckhert, L. G., "The Log-normal Model for the Distribution of One Claim," *ASTIN Bulletin*, Volume II, Part I, Amsterdam, 1962.
7. Ferrara, G., "Distributions des Sinistres Incendie Selon Leur Cout," *ASTIN Bulletin*, Volume VI, Part I, Amsterdam, September 1971.
8. Ramachandran, G., "Factors Affecting Loss-Multiple Regression Model With Extreme Values," Fire Research Station, Department of the Environment and Fire Offices' Committee, Fire Research Note 991, England, November 1973.
9. Benckert, L. G. and Sternberg, I., "An Attempt to Find An Expression for the Distribution of Fire Damage Acount," *Transactions of the 15th International Congress of Actuaries*. Vol. 2, 1958.
10. Mandelbrot, B., "Random Walks, Fire Damage and Other Paretian Risk Phenomena," *Operations Research*, Vol. XII, 1964.
11. Andersson, Hans, "An Analysis of the Development of the Fire losses in the Northern Countries After the Second World War," *ASTIN Bulletin*, Vol. VI, Part I, Amsterdam, 1971.
12. Benckert, L. G. and Jung, J., "Statistical Models of Claim Distributions in Fire Insurance," *ASTIN Bulletin*, Vol. VIII, Part I, Amsterdam, September 1974.
13. Holcomb, E. W., "Estimating Parameters of Stable Distributions with Application to Non-Life Insurance," Ph.D. Thesis, University of Tennessee, September 1973.

14. Hewitt, C. C., "Credibility for Severity," *Proceedings of the Casualty Actuarial Society,* Vol. LVII, 1970.
15. Mann, N. R., Schafer, R. E., and Singpurwalla, N. D., *Methods for Statistical Analysis of Reliability and Life Data,* John Wiley and Sons, New York, 1974.
16. Barlow, R. E. and Proschan, F., *Mathematical Theory of Reliability,* John Wiley and Sons, New York, 1965.
17. Barlow, R. E. and Proschan, F., *Statistical Theory of Reliability and Life Testing: Probability Models,* Holt, Rinehart and Winston, New York, 1975.
18. Baldwin, R. and North, M. A., "The Number of Sprinklers Opening and Fire Growth", *Fire Technology,* Boston, Massachusetts, November 1973.
19. Ramachandran, G., "The Poisson Process and Fire Loss Distribution", *XXXVIIth Session of the International Statistical Institute,* London, September 1969.
20. Ramachandran, G., "Extreme Value Theory and Large Fire Losses" *ASTIN Bulletin,* Vol. VII, Part 3, Amsterdam, March 1974.
21. DuMouchel, W. H. and Olshen, R. A., "On the Distribution of Claim Costs", Berkeley Actuarial Research Conference on Credibility, September 1974.
22. Head, G. L., *Insurance to Value,* Published for the S. S. Huebner Foundation for Insurance Education, University of Pennsylvania, by Richard D. Irwin, Inc., Homewood, Illinois, 1971.
23. Shpilberg, D., *Statistical Decomposition Analysis of Industrial Fire Loss,* Report R75-36, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Mass., October, 1975.
24. Johnson, N. L. and Kotz, S., *Continuous Univariate Distributions—2,* Houghton Mifflin Company, Boston, 1970.