The Probability Distribution of Extremes

I. Calderone
Facade Technology, 601 St. Kilda Road, Melbourne.

and

W. H. Melbourne
Department of Mechanical Engineering, Monash University.

Abstract

The probability distributions of extremes were investigated for various types of populations of random variables. The commonly used distributions for the extremes were compared with the actual distributions, in order to determine which distributions are the most appropriate to use for predicting the extremes of the original population. The investigation revealed that the appropriate probability distribution to use for the extremes, depends upon the nature of the original population. Consequently, a knowledge of the original population is necessary, in order to more accurately predict the extremes of the population. It is concluded that the lognormal distribution is the most appropriate to use for predicting the extremes of lognormal populations. Since interim results have shown that the probability distributions of wind speeds appear to be closely approximated by lognormal distributions, then it may be appropriate to use the lognormal distribution to predict extreme wind speeds.

1. INTRODUCTION

It has been previously found, [1], that the probability distribution of the positive pressure, on the windward wall of a building, is very closely represented by a lognormal distribution. It has also been found [2, 3] that a lognormal distribution gives a much closer representation of glass breakage data than a Weibull distribution. The Weibull distribution is the Type III extreme value distribution for minimum values. [4] Consequently, it has been commonly used to represent the distribution of the minimum breakage stresses.

The Type I extreme value distribution (also known as Gumbel's extreme value distribution), [4], is commonly used to represent the probability distribution of extreme wind gust velocities. The Type I distribution can also be used to represent the distribution of extreme minimum values. Gumbel (1954), [5], recognised that the Type I distribution is not an exact solution, but it gives a very close approximation to the exact distribution of normal extremes. Gumbel also recognised that the lognormal distribution is a close approximation to the Type I distribution.

For the purpose of determining which distributions are the most appropriate to use, for predicting the extremes of the original population, the exact distributions, for various types of populations of random variables, were compared with the Gumbel, Weibull and Lognormal distributions.

2. VARIOUS POPULATIONS OF RANDOM VARIABLES

The extremes were investigated, for a variety of probability distributions (see Fig. 1) as follows.

2.1 Uniform Distribution - This distribution has a constant probability density function across the range of values of the population and a uniformly increasing linear cumulative probability. The population is formed from random values having equal probability of occurrence.

2.2 Triangular Distribution - This distribution has a triangular probability density function across the range of values of the population and a symmetrical "S" shaped cumulative probability. The population is formed from the sum of two independent random variables.

2.3 Gaussian (Normal) Distribution - This distribution has a bell shaped probability density function across the range of values of the population and a symmetrical "S" shaped cumulative probability. The population is formed from the sum of many independent random variables.

2.4 Lognormal Distribution - This distribution has a skewed bell shaped probability density function across the range of values of the population and a skewed "S" shaped cumulative probability. The population is formed from the product of many independent random variables.

3. INVESTIGATION

For the purpose of this investigation, examples of the populations described above, were generated using the random number generator of Microsoft Quick BASIC. Samples of extreme values were taken from these populations, as follows.

Sample sizes of 2, 5 and 10 were used and the highest (and lowest) of each sample was taken, giving the extreme value distributions for:

(a) Highest (and lowest) of 2 values,

(b) Highest (and lowest) of 5 values,

(c) Highest (and lowest) of 10 values.

The exact probability distribution of the extremes was also determined, using the relationships given by Gumbel [5] as follows.

(a) The probability that the highest value falls short of x_n is

$$\Phi_n(x_n) = F^n(x_n) \qquad (1)$$

(b) The probability for the smallest value to be less than x_1 is:

$$\Phi_n(x_1) = 1-[1-F(x_1)]^n \qquad (2)$$

Close agreement was found between the actual distributions determined from equations (1) and (2) above, and those determined from the extreme values taken from the uniform distribution populations generated using the random number generator of Microsoft Quick BASIC (see Fig. 2). Consequently, equations (1) and (2) were used in this investigation, as this required less computing time.

The exact probability distributions for the highest and lowest extreme values (for a sample size of 10) for each of the selected types of populations are shown in Fig. 3. The exact distributions were then compared with fitted Gumbel, Weibull and lognormal distributions as follows.

To determine the parameters for the fitted Weibull Distribution, the actual distributions were plotted on a Weibull probability chart (see Fig.4, showing the case for minimum extremes, for sample size of 10, from the Normal distribution).

To determine the parameters for the fitted Gumbel Distribution, the actual distributions were also plotted on a Gumbel probability chart (see Fig. 5, showing the case for the minimum extremes, for sample size of 10, from the Normal distribution).

To determine the parameters for the fitted lognormal distribution, the mean, median and mode of the actual distributions were found and used in the following equations, given by Aitchison and Brown, [6], for the lognormal distribution.

$$mean = e^{\mu+\frac{1}{2}\sigma^2} \qquad (3)$$

$$median = e^{\mu} \qquad (4)$$

$$mode = e^{\mu-\sigma^2} \qquad (5)$$

The parameters, σ and μ, determined using equations (3), (4) and (5) were then used in the 4 parameter lognormal distribution, given by:

$$f(x) = \frac{\exp\left(-0.5\left(\frac{\ln(k.x+c)-\mu}{\sigma}\right)^2\right)}{\sqrt{2\pi}\,\sigma\,(k.x+c)} \qquad (6)$$

The actual extreme value distributions, for the case of the highest (and lowest) of 10 values taken from various distributions, were compared with fitted distributions, as shown in Fig. 6 and Fig. 7 (a).

Also, to simulate extreme annual maxima, the case for the highest of 365 values was compared with fitted Gumbel and Lognormal distributions. (See in Fig. 7 (b)).

Examples of full scale wind speed measurements and maximum hourly wind gusts were also considered, as shown in Fig. 8.

4. DISCUSSION OF RESULTS

None of the commonly used extreme value distributions fitted the extremes for a uniform distribution.

For the case of a population formed from the sum of two independent random variables (a triangular distribution), it was found that a 3 parameter Weibull distribution fitted the actual distribution very closely (See Fig. 6 (a)). In this case, the 3 parameter Weibull distribution was much better than the 2 parameter Weibull, Gumbel or Lognormal distributions.

For the extreme minima, for the case of a Gaussian (normal) population, the Weibull and Gumbel plots showed that the cumulative probability distribution was distinctly curved for this case (see Fig. 4 and Fig. 5). Consequently, the probability distribution of the data could not be matched closely using the 2 parameter Weibull distribution or the Gumbel distribution.

However, it was again found that a 3 parameter Weibull distribution fitted the actual distribution very closely (See Fig. 6 (a)). In this case, the 3 parameter Weibull distribution was also found to be better than the 2 parameter Weibull, Gumbel or Lognormal distributions.

For the extreme minima, for the case of a Lognormal population, it was found that a 3 parameter Weibull distribution as well as the Gumbel and Lognormal distributions fitted the actual distribution reasonably well (See Fig. 6 (b)). In this case, it is not possible to readily determine which distribution is the best to use.

For the extreme maxima, for the case of a Lognormal population, it was found that a Lognormal distribution fitted the actual distribution noticeably better than a Gumbel distribution (See Fig. 7).

Since, it can be shown by the central limit theorem, that the distribution for the product of n positive random variables becomes approximately lognormal as n becomes large. This means that the lognormal distribution appears in nature when unrelated phenomena represented by positive random numbers combine multiplicatively (Pfeiffer and Schum, 1973).

Examples of distributions of full scale wind speeds and maximum hourly wind gusts are shown in Fig. 8. The lognormal distribution appears to fit the data reasonably well.

5. CONCLUSION

The investigation revealed that the appropriate probability distribution for the extremes, depends upon the nature of the original population. Consequently, a knowledge of the original population is necessary, in order to more accurately predict the extremes of the population. The lognormal distribution appears to be the most appropriate to use for predicting the extremes of lognormal populations. Since interim results have shown that the probability distributions of wind speeds appear to be closely approximated by lognormal distributions, then it may be appropriate to use the lognormal distribution to predict extreme wind speeds.

6. REFERENCES

1. Calderone, I., Cheung, J.C.K., Melbourne, W.H., J. Wind Eng. Ind. Aerodyn, Vol. 53 (1994) pp 247-259.

2. Calderone, I., Melbourne, W.H, Proc., 3rd Workshop on Wind Engineering, Brisbane, July 8-9, 1993, pp E5-9.

3. Calderone, I., Melbourne, W.H, Austceram 94 Conference, Sydney, 25-27 July 1994.

4. Hahn, G.J. & Shapiro, S.S., Statistical Models in Engineering, John Wiley and Sons, New York, p111.

5. Gumbel, E.J., Statistical Theory of Extreme Values and Some Practical Applications. National Bureau of Standards, USA, 1954.

6. Aitchison, J. and Brown, J.A.C, The Lognormal Distribution, Cambridge University Press, 1969. p9.

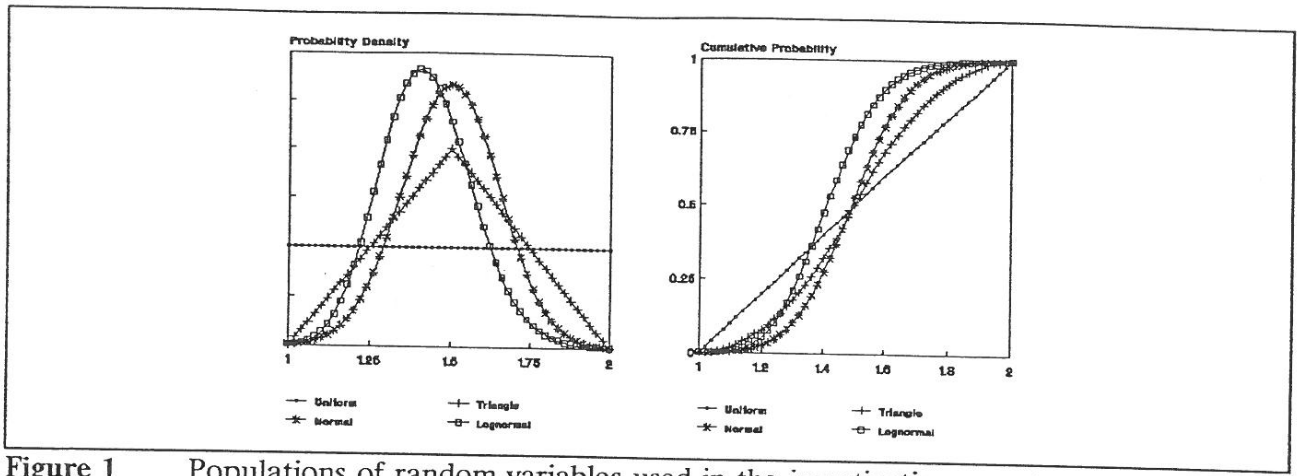

Figure 1 Populations of random variables used in the investigation.

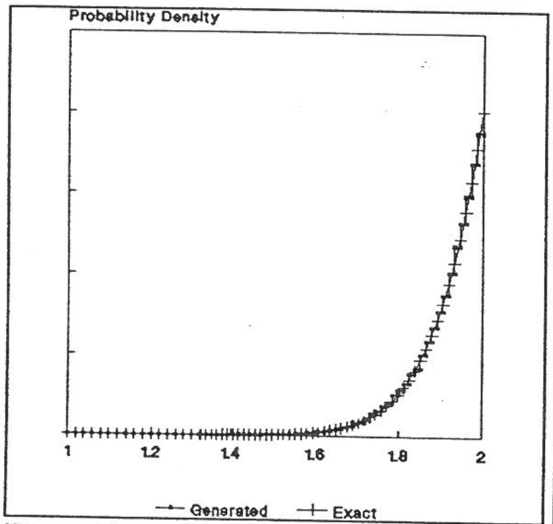

Figure 2 Extreme value distribution for highest of 10 from a uniform distribution.

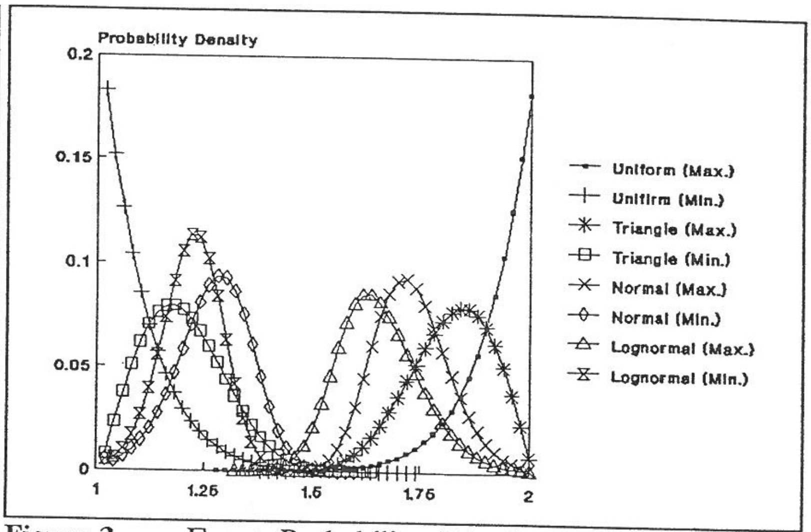

Figure 3 Exact Probability Distributions of Extremes (highest and lowest) of various populations.

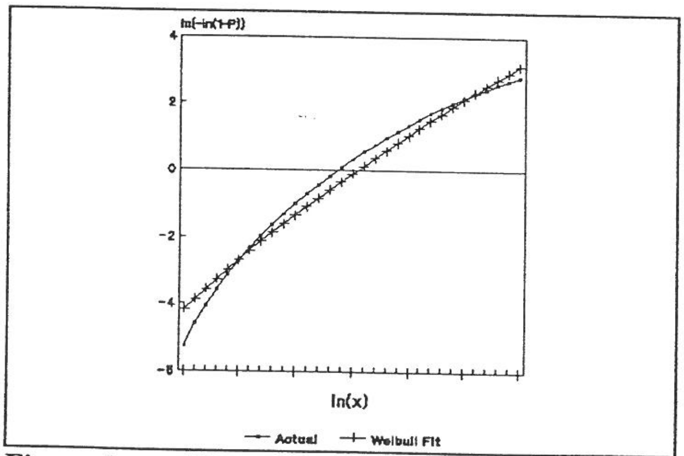

Figure 4 Weibull plot for fitting extreme minimums of a normal distribution.

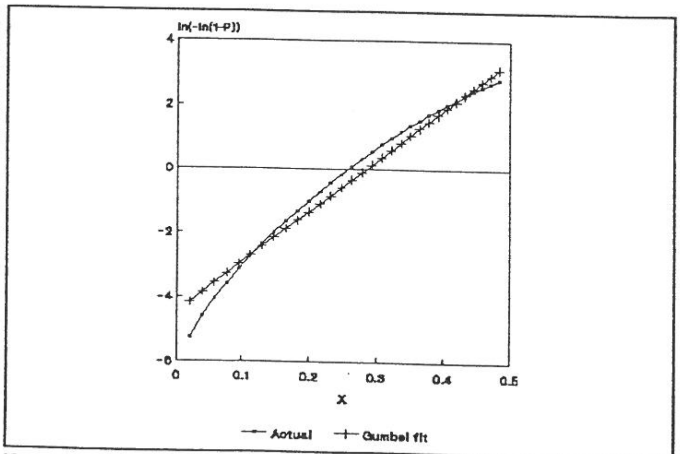

Figure 5 Gumbel plot for fitting extreme minimums of a normal distribution.

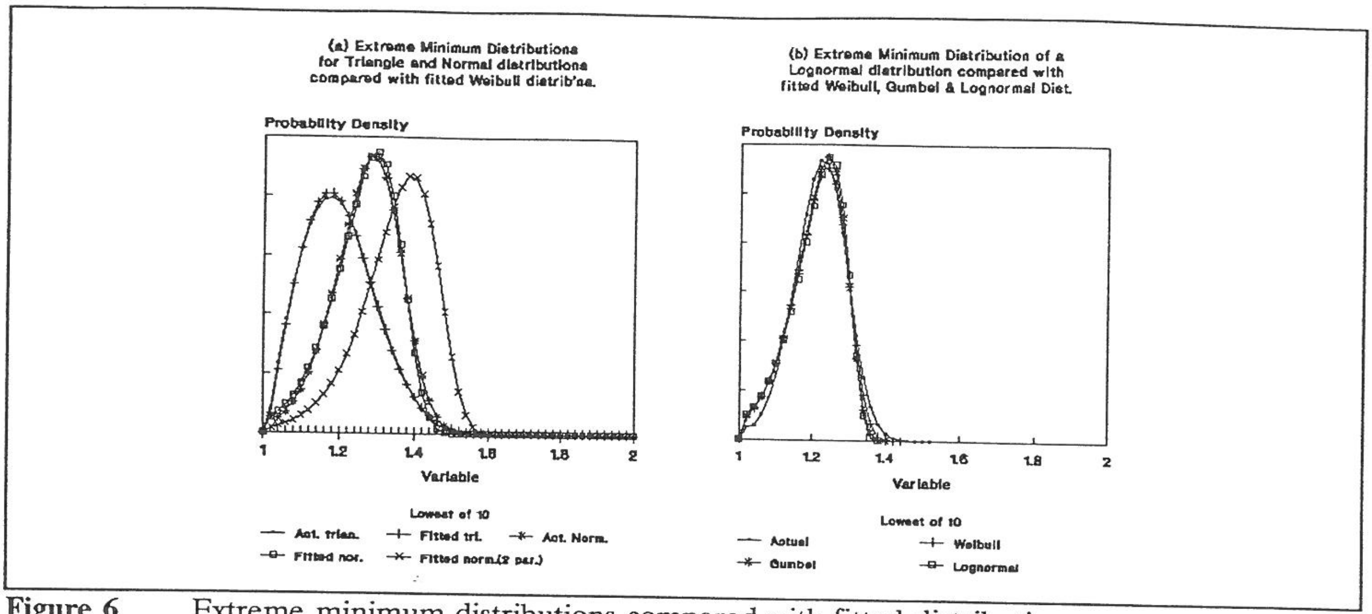

Figure 6 Extreme minimum distributions compared with fitted distributions.

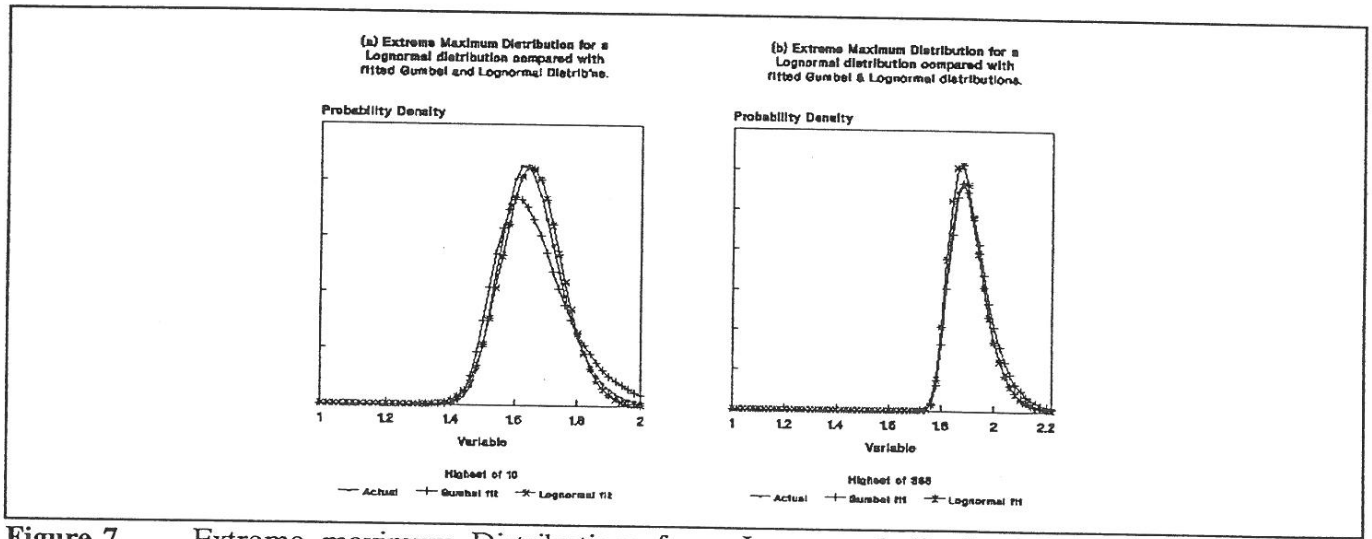

Figure 7 Extreme maximum Distributions for a Lognormal distribution compared with fitted Gumbel and Lognormal distributions.

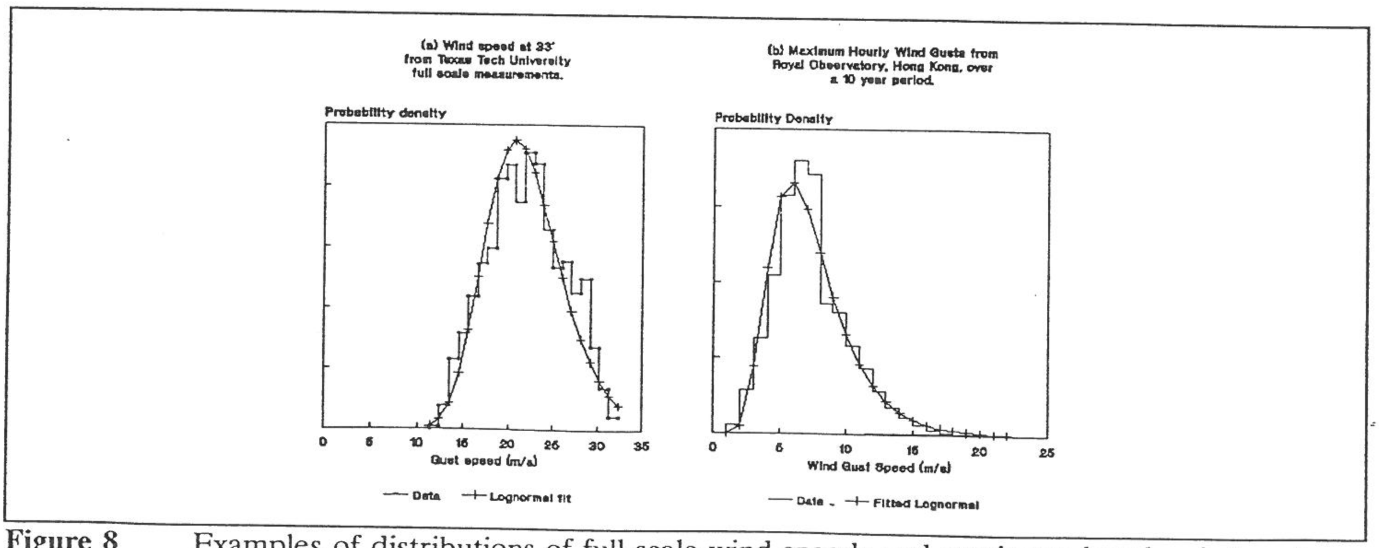

Figure 8 Examples of distributions of full scale wind speeds and maximum hourly wind gusts.