

SOME STATISTICAL ASPECTS OF CATASTROPHIC RISKS

C. S. ANANTAPADMANABHAN

India

1. Risks of a catastrophic nature, such as war or earthquake, can differ, in a statistical sense, from other ordinarily insured perils only in respect of the distribution of the probability function. For example, although the probability of a loss occurring may be small, its magnitude once it occurs may be very large. One reason for the latter feature is that the sum insured may itself be heavy. But still more important is the fact that there would be strong positive correlation between similar contiguously situated risks. This is the main reason for the loss taking on a catastrophic character.

2. The fact that the probability of a loss is ordinarily small, leads to the result that the expected loss and hence the pure premium for the risk is small. Equally, and for the same reason, the "loss strain at risk" is large. If the latter could be considerably reduced—which can only be done by charging impracticably high premiums—the risk becomes more easily insurable.

3. If there are n risks of this nature, 1, 2, 3 . . . n and $v_i(x)$ is the conditional frequency function of the i th risk we may write

$$\int_{0+}^{\infty} v_i(x) dx = q_i$$

While we may imagine this as a frequency curve, the complementary probability p_i of a loss not occurring can be represented as a vertical line at the origin—the different methods of representation of p and q are actually convenient. With this representation, the expansion of the binomial product

$$\prod_{i=1}^n \{(1 - q_i) + \lambda q_i\} = \sum_{i=0}^n A_i \lambda^i$$

shows that the convolution of the n curves can be expressed as the sum of $n + 1$ functions

$$V_n(x) = \sum_{r=0}^n A_r(x). \tag{A}$$

where

$$(i) \ A_0(x) = \prod_{i=1}^n p_i$$

(ii) the term $A_1(x)$ comprises n separate curves of the form $v_i(x)$ each multiplied by its complementary probability, being of the form

$$A_1(x) = \sum_{i=1}^n p_1 p_2 p_3 \dots p_{i-1} p_{i+1} \dots p_n v_i(x)$$

(iii) the term $A_2(x)$ comprises $\binom{n}{2}$ terms each consisting of the convolution of the curves taken two at a time and multiplied by the complementary probability, being of the form

$$A_2(x) = \sum_{i,j} p_1 p_2 p_3 \dots p_{i-1} p_{i+1} \dots p_{j-1} p_{j+1} \dots p_n \int_{0+}^x v_i(\xi) v_j(x - \xi) d\xi$$

and so on.

4. We should particularly note that the frequency curves involved do not have a total frequency of unity, for according to our definition

$$\int_{0+}^{\infty} v_i(x) dx = q_i$$

Hence the convolution of two such curves has a total frequency equal to the corresponding products, so that

$$\int_{0+}^{\infty} \int_{0+}^z v_i(x) v_j(z - x) dx dz = q_i q_j$$

and the same rule applies for higher orders of convolution also. It also follows that each term in expression (A) above contains implicitly the product of $(n - r)$ q -ratios since it contains a convolution of $(n - r)$ frequency curves.

5. We know that the convolution of two independent normal distributions is itself normal; if x_1 and x_2 are distributed as $N(\mu_1\sigma_1^2)$ and $N(\mu_2\sigma_2^2)$ respectively, $x_1 + x_2$ is distributed as $N(\mu_1 + \mu_2\sigma_1^2 + \sigma_2^2)$. This property is obviously unaffected if the ordinate of each of the curves is reduced in fixed proportions, the proportion being the same for all ordinates on each curve. We may make use of this interesting property and assume as an approximation that the claim on risk i is distributed normally with mean S_i and standard deviation σ_i . While our present knowledge of the statistical fluctuations is too limited to presume any great inaccuracy in this approach, the normal curve converges sufficiently rapidly on both sides to enable us to expect confidently that the errors involved in ignoring negative values of the loss variable are negligible.

6. Let any two risks i and j have a *coefficient of contiguity* α_{ij} defined by the condition that if risk i becomes a loss, the probability of risk j becoming a loss through contiguity (and not by the direct effect of the peril) is α_{ij} . The probability of both i and j being lost in such circumstances is therefore $q_i p_j \alpha_{ij}$.

7. With this definition let us first take the simple case of three risks 1, 2 and 3. In the ordinary case, when the risk due to contiguity is neglected, the distribution can be written in the form

$$\begin{aligned}
 I = & [p_1 p_2 p_3] \\
 & + [p_1 p_2 q_3 + p_1 q_2 p_3 + q_1 p_2 p_3] \\
 & + [p_1 q_2 q_3 + q_1 p_2 q_3 + q_1 q_2 p_3] \\
 & + [q_1 q_2 q_3]
 \end{aligned}
 \tag{B}$$

The terms in square brackets separate out the terms representing no loss, one loss, two losses and three losses respectively. When there is hazard due to contiguity the expansion becomes

$$\begin{aligned}
 I = & [p_1 p_2 p_3] \\
 & + [q_1(I - \alpha_{12})(I - \alpha_{13})p_2 p_3 + q_2(I - \alpha_{21})(I - \alpha_{23})p_1 p_3 \\
 & \quad + q_3(I - \alpha_{31})(I - \alpha_{32})p_1 p_2] \\
 & + [q_1 p_2 p_3 \alpha_{12}(I - \alpha_{13}) + q_1 p_2 p_3 \alpha_{13}(I - \alpha_{12}) \\
 & \quad + p_1 q_2 p_3 \alpha_{21}(I - \alpha_{23}) + p_1 q_2 p_3 \alpha_{23}(I - \alpha_{21}) \\
 & \quad + p_1 p_2 q_3 \alpha_{31}(I - \alpha_{32}) + p_1 p_2 q_3 \alpha_{32}(I - \alpha_{31}) \\
 & \quad + q_1 q_2 p_3 (I - \alpha_{13})(I - \alpha_{23}) + p_1 q_2 q_3 (I - \alpha_{21})(I - \alpha_{31}) \\
 & \quad + q_1 p_2 q_3 (I - \alpha_{12})(I - \alpha_{32})]
 \end{aligned}$$

$$\begin{aligned}
 &+ [q_1 p_2 p_3 \alpha_{12} \alpha_{13} + p_1 q_2 p_3 \alpha_{21} \alpha_{23} + p_1 p_2 q_3 \alpha_{31} \alpha_{32} \\
 &+ q_1 q_2 p_3 (\alpha_{13} + \alpha_{23} - \alpha_{13} \alpha_{23}) \\
 &+ q_1 p_2 q_3 (\alpha_{12} + \alpha_{32} - \alpha_{12} \alpha_{32}) \\
 &+ p_1 q_2 q_3 (\alpha_{21} + \alpha_{31} - \alpha_{21} \alpha_{31}) \\
 &+ q_1 q_2 q_3] \tag{C}
 \end{aligned}$$

8. In the frequency curve concept, if in addition to the q -terms we think of each p also as a frequency curve with a single value, we can say that each one of the terms in expression (C) represents a convolution of three frequency curves multiplied by a fraction involving α 's. But for these fractional multipliers, which only redistribute the terms, this expression is algebraically identical with the expansion (B) above. But a graphical study of the distribution is more instructive.

9. For illustrative calculations, three risks have been assumed, the distribution of each of which is assumed to follow the normal distribution. The parameters assumed are as follows:

Curve A	Mean = 100	$\sigma = 10$	$q = .15$
Curve B	Mean = 160	$\sigma = 15$	$q = .25$
Curve C	Mean = 180	$\sigma = 6$	$q = .20$

The three curves are shown graphically in Fig. 1.

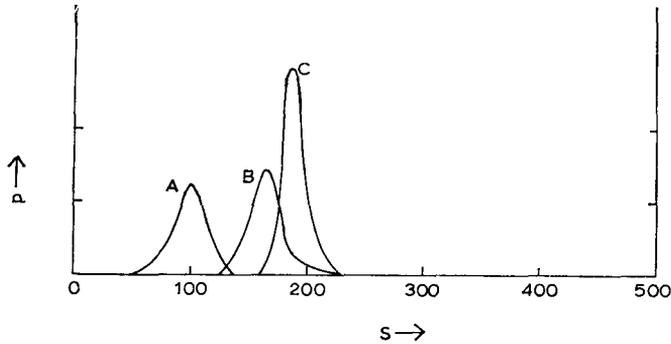
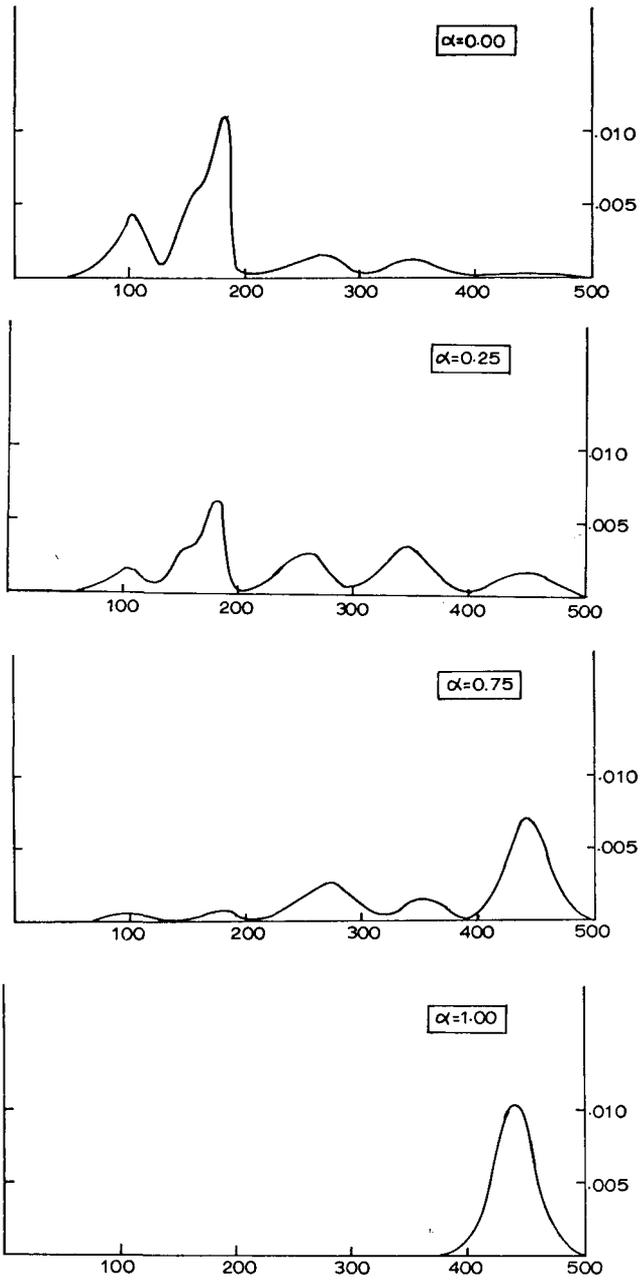


Fig. 1

10. The frequency curves of the conjoint distributions on the assumption that all the α 's are equal have been worked out for values of $\alpha = 0.00, 0.25, 0.75$ and 1.00 . These curves are shown in



S →

Fig. 2

Fig. 2. It will be seen that as α increases in value, the frequency curve is pushed outwards more and more away from the origin, though the areas under the curves all remain equal, being obviously equal to

$$1 - p_1 p_2 p_3$$

In the extreme case of $\alpha = 1.00$ only the combined convolution of the three curves remains, expressing the obvious fact that loss of one risk implies the sure loss of all the three risks.

11. There is no accepted convention for measuring the hazardousness of a risk, but it will be readily conceded that if such a measure were agreed to, it would have the following features:

(i) The hazardousness must be zero at the origin, and it must increase rapidly with the Sum insured. Alternatively if catastrophic risks only are to be evaluated, upto a critical Sum insured S the hazardousness would be zero and beyond that point the hazardousness would rapidly increase with the Sum insured.

(ii) For any particular value of Sum insured, the hazardousness must increase with the probability of loss.

12. If $h(s)ds$ represents the hazardousness at point s the hazardousness implicit in a distribution can be measured by

$$H = \int_0^{\infty} h(s)ds$$

It will be obvious that with the above conventions, the hazardousness of the distributions shown in Fig. 2 rapidly increases with increase in the value of α .

13. The general formula in the simple case where there is no risk due to contiguity is, of course, easily written:

Let

$$f(x|x_i\sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2}\frac{(x-x_i)^2}{\sigma_i^2}}$$

$$\prod_{i=1}^n p_i = P_n$$

Then the frequency function $V_n(x)$ is

$$\begin{aligned}
 V_n(x) &= \sum_{r=0}^n A_r(x) \\
 &= P_n + \sum_i f(x|x_i) \frac{\sigma_i^2}{p_i} \frac{P_n}{p_i} + \\
 &\quad + \sum_{i,j} f(x|x_i + x_j) \frac{\sigma_i^2 + \sigma_j^2}{p_i p_j} \frac{P_n}{p_i p_j} + \\
 &\quad + \sum_{i,j,k} f(x|x_i + x_j + x_k) \frac{\sigma_i^2 + \sigma_j^2 + \sigma_k^2}{p_i p_j p_k} \frac{P_n}{p_i p_j p_k} + \dots
 \end{aligned}$$

The effect of introducing the idea of loss from contiguity is to re-distribute a part of each term $A_r(x)$ except $A_0(x)$ and $A_n(x)$ to all terms subsequent to it. The evaluation of the resultant expression is algebraically practicable though admittedly laborious except in the simplest cases as in paras 9 and 10 above.

14. In the discussion so far, we have assumed the variables of the distributions to be mutually independent even though there is the risk of loss due to contiguity. But in practice, we can expect that a heavy loss in one risk is more likely to produce a heavy loss in a contiguously situated risk. If x_1 and x_2 are normally distributed, and at the same time x_1 and x_2 are also positively correlated, a well known statistical result tells us that $x_1 + x_2$ is still distributed normally but with a higher variance, namely in the form

$$N(\mu_1 + \mu_2 \quad \sigma_1^2 + \sigma_2^2 + 2\sigma_{12})$$

where σ_{12} is the covariance between x_1 and x_2 . The conclusions which we have reached earlier are therefore unaffected, except that such correlations tend to flatten the distributions.