Annals of Operations Research 9(1987)545–560

COMPUTING RISK PROFILES FOR COMPOSITE LOW-PROBABILITY HIGH-CONSEQUENCE EVENTS

D.I. HEIMANN

Transportation Systems Center, U.S. Department of Transportation, Cambridge, Massachusetts 02142, USA

and

T.S. GLICKMAN

Whiting School of Engineering, The Johns Hopkins University, Baltimore, Maryland 21218, USA

Abstract

Low-probability high-consequence events play an important role is assessing the risk of catastrophic loss. Their risk profiles, however, can be difficult to obtain. This paper obtains the risk profiles of low-probability high-consequence events where the final consequence results from a number of intermediate events. Called composite events, these events occur, for example, in accidents releasing hazardous material. The structure of composite events is described and risk profile equations developed, both on a per-event and per-annum basis. The many extremely low valued terms, especially in the tail of the risk profile, make calculation nontrivial. Accordingly, algorithms are developed to compute these equations. In addition, formulae for means and variances are obtained, and an illustrative example is provided.

Keywords and phrases

Risk analysis, computational methods, probability, algorithms.

1. Introduction

An important consideration one needs to take into account in risk analysis is the probability of incurring a catastrophic loss. This is particularly true when addressing risks from events which occur infrequently but have severe consequences when they do occur, such as airplane crashes, nuclear accidents, toxic chemical accidents, earthquakes, or hurricanes. Such low-probability, high-consequence events have a far higher perceived impact, both on the affected groups and the general public, than events which occur more frequently, but with less severity per event, such as non-epidemic disease or automobile accidents.

Because of the low probability of such events, their historical data is very sparse. Due to this very limited data, the sample sizes of statistical analyses would be very small, leading to high variances and poor (in the sense of imprecise) risk estimates. Without further knowledge, the likelihood and severity of low-probability, high-consequence events are thus difficult to estimate from historical data alone.

If the hazardous event has a special structure, however, that structure may be exploited to find reasonable risk estimates which could not otherwise be obtained. An event which has such a structure is what we shall refer to as a composite event, that is, one where the final consequence is the result of a number of intermediate events, for each of which enough data or analytic insight exists to obtain a conditional probability distribution of its effects. The risk profile (a function showing the likelihood of any given level of consequence) of a composite event can then be obtained through convolution of the intermediate probability expressions. One can also similarly investigate the sensitivity of this risk to changes in the underlying parameters, such as changes in the type of hazard or the underlying economic activity creating the hazard exposure.

Carrying out these computations requires great care. The part of the risk profile which corresponds to high consequence levels consists of extremely low-valued probabilities, which are in turn the sums of many terms with even lower values. Such computations are susceptible to round-off error and computer underflow. Moreover, they may show a high degree of sensitivity to the sequence of computation and the efficiency of the algorithm. Failure to take these considerations into account can lead to inordinately large computation times or incorrect results (e.g. risk profiles whose probabilities sum to less than one).

This paper describes a technique to obtain the risk profile of a composite low-probability high-consequence event by exploiting its special structure. The composite event consists of the following sequence: (1) an initial event occurs, (2) intermediate actions take place, (3) a resulting event occurs, and (4) the resulting event causes damage. For example, the storage of poisonous gas at a pesticide manufacturing plant, the production of munitions at an arsenal, the storage of LNG in a port facility, and the consumption of plutonium in a nuclear power plant all represent composite events. In a hazardous materials release incident, an initiating accident (initial event) creates a puncture or overpressure in a container (intermediate event), causing some or all of its contents to be spilled (intermediate event), so that the surrounding persons, property, and environment are exposed to the effects of toxicity, fire, or explosion (resulting event). More generally, the event sequence is descriptive of other situations as well, such as the risk posed to electrical equipment by the release of carbon fibers from composite materials.

The consequences of a composite event such as a hazardous materials release incident are commonly (but not necessarily) measured in such terms as lives lost, injuries sustained, value of property destroyed, or level of groundwater contaminated. A risk profile usually takes the form of an upward cumulation of the risk density, which is the probability distribution of the magnitude of a selected consequence. Kaplan and Garrick [1] have provided a useful description of the concepts involved in computing and interpreting risk profiles. Fiksel and Rosenfield [2] present an approach for modeling risk profiles of a few specific types of hazardous materials release incidents. A related approach was used in a detailed risk analysis of derailments of trains carrying bulk chemicals, as described by Glickman and Rosenfield [3].

In the discussion to follow, the computational aspects of these approaches are generalized to cover all types of composite damaging release incidents, applying also to other composite low-probability, high-consequence events in which a random initiating event results in a random, quantifiable effect on a vulnerable entity randomly distributed over an exposed area. For the sake of clarity, and without loss of generality, our discussion will address the fatality consequence of a hazardous materials release. The risk profile models, expressed both on a per incident and per annum basis, are developed in sect. 2. The techniques to actually compute these risk profiles, as well as formulae to directly compute the mean per accident and per annum risks and their variances, are shown in sect. 3. Section 4 shows an illustrative example of a risk profile computation.

2. The composite-event risk profile models

Let the per incident risk density, i.e. the probability of x fatalities due to a hazardous materials release incident, be $f(x)$. Let also the per annum risk density, i.e. the probability of x fatalities in total due to all the hazardous materials release incidents of interest during a year, be $u(x)$. The per incident risk profile $F(x)$ is thus the sum from x to infinity on $f(\cdot)$, while the per annum risk profile $U(x)$ is the sum from x to infinity on $u(\cdot)$.

The persons in the exposed area are of two distinct types: (1) the general population (third-party bystanders), and (2) the professional population (employees and emergency responders). The function $f(x)$ is thus the convolution of $g(x)$, the probability of x fatalities in the general population due to a hazardous materials release incident, with the probability distribution of the number of fatalities among the professional population. The function $u(x)$ is determined from the convolution of the function $f^{(n)}(x)$, the probability of x fatalities due to n hazardous materials release incidents (i.e. the n-fold convolution of $f(x)$ with itself), with the probability distribution of n such incidents in a year.

Let r_{ijk} be the joint probability that an incident involves the release of hazardous material type i, and occurs under operating condition type j, when the general population exposure is in stratum k. Let $g_{ijk}(x)$ be the conditional probability of x fatalities given such an incident. Then we have

$$g(x) = \sum_{i,j,k} r_{ijk}\, g_{ijk}(x)\,. \tag{1}$$

The function $g_{ijk}(x)$ depends in turn on $p_j(q)$, the conditional probability that not more than quantity q of the hazardous material is spilled when the incident occurs under operating condition j, and $h_{ik}(x|q)$, the conditional probability that there are x fatalities in the general population, given hazardous material type i, population density stratum k, and quantity-spilled q. Based upon these values, $g_{ijk}(x)$ is given by

$$g_{ijk}(x) = \int_0^\infty h_{ik}(x|q)\, \mathrm{d}p_j(q)\,. \tag{2}$$

Now assume that the distribution $h_{ik}(x|q)$ is given by $h(x\,;T_{ik}(q))$, a Poisson distribution with mean $T_{ik}(q)$, where $T_{ik}(q)$ is the conditional mean number of fatalities among the general population for a release quantity q of hazardous material type i in population-density stratum k. Equation (2) therefore becomes

$$g_{ijk}(x) = \int_0^\infty h(x\,;T_{ik}(q))\, \mathrm{d}p_j(q). \tag{3}$$

The mean number of fatalities $T_{ik}(q)$ is given by

$$T_{ik}(q) = v_i\, \rho_k\, A_i(q)\,,$$

where v_i is the mortality rate of the general population in the lethal area to a release of hazardous material type i, ρ_k is the density of the general population in stratum k, and $A_i(q)$ is the average size of the lethal area for a release of quantity q of hazardous material type i.

3. Computing the composite-event risk profile

3.1. LINEARIZATION

Both the quantity-released probability distribution $p_j(q)$ and the lethal-area function $A_i(q)$ are usually known only for a set of discrete points $\{q_z\}$, $z = 0, \ldots, M$, i.e. $p_j(q)$ is defined at the points $\{(q_z, p_{jz})\}$, where $p_{jz} = p_j(q_z)$, and $A_i(q)$ is defined at the points $\{(q_z, A_{iz})\}$, where $A_{iz} = A_i(q_z)$ (note that $q_0 = p_{j0} = A_{i0} = 0$). We therefore approximate the continuous functions $p_j(q)$ and $T_{ik}(q) = v_i\, \rho_k\, A_i(q)$ with piecewise-linear functions $p_j^\star(q)$ and $T_{ik}^\star(q)$. At the points $\{q_z\}$, $z = 0, \ldots, M$, we have

$$p_j^\star(q_z) = p_j(q_z) \equiv p_{jz}$$

$$T_{ik}^\star(q_z) = T_{ik}(q_z) = T_{ikz} \equiv v_i\, \rho_k\, A_{iz}\,, \qquad z = 0, \ldots, M,$$

while for all $q \neq q_z$, the functions $p_j^\star(q)$ and $T_{ik}^\star(q)$ are linear.

Under this approximation $g_{ijk}(x)$, as expressed in eq. (3), is replaced by

$$g_{ijk}^\star(x) = \int_0^\infty h(x; T_{ik}^\star(q))\, \mathrm{d}p_j^\star(q). \tag{4}$$

Using the definition of $p_j^\star(q)$ and defining the variable $T = T_{ik}^\star(q)$, we obtain from (4)

$$g_{ijk}^\star(x) = \sum_{z=0}^{M} (\Delta p_{jz} / \Delta T_{ikz}) \int_{T_{ikz}}^{T_{ik,z+1}} h(x; T)\, \mathrm{d}T, \tag{5}$$

where $\Delta p_{jz} = p_{j,z+1} - p_{jz}$ and $\Delta T_{ikz} = T_{ik,z+1} - T_{ikz}$.

Remark

For completeness, we define $p_{j,M+1}$ as 1 (thus, $\Delta p_{jM} = 1 - p_{jM}$) and ΔT_{ikM} as 2 $(\Delta p_{jM})\,(\Delta T_{ik,M-1} / \Delta p_{j,M-1})$, i.e. the slope of the mortality cumulative distribution function for $q = q_M^+$ is assumed to be one-half of the slope for $q_{M-1} \leqslant q \leqslant q_M$ (that is, $[\Delta p_{jM} / \Delta T_{ikM}] = \frac{1}{2}[\Delta p_{j,M-1} / \Delta T_{ik,M-1}]$).

3.2. EVALUATING THE COMPOUND POISSON INTEGRAL

To evaluate the integral in (5), we integrate by parts. Let $u = T^x/x!$ and $\mathrm{d}v = \mathrm{e}^{-T}\mathrm{d}T$. Then $\mathrm{d}u = T^{(x-1)}/(x-1)!$ and $v = -\mathrm{e}^{-T}$, so that

$$\int h(x; T)\, \mathrm{d}T = \int \mathrm{e}^{-T}\, T^x/x!\, \mathrm{d}T = -\,\mathrm{e}^{-T}\, T^x/x! + \int \mathrm{e}^{-T}\, T^{(x-1)}/(x-1)!\, \mathrm{d}T. \tag{6}$$

Let

$$Q(x, T) = \int e^{-T} T^x/x!\, dT.$$

Then eq. (6) becomes

$$Q(x, T) = \begin{cases} -e^{-T} T^x/x! + Q(x-1, T), & x \geqslant 1 \\ -e^{-T}, & x = 0 \end{cases} \tag{7}$$

and eq. (5) becomes

$$g^{\star}_{ijk}(x) = \sum_{z=0}^{M} (\Delta p_{jz}/\Delta T_{ikz})\, \Delta Q(x, T_{ikz}), \tag{8}$$

where $\Delta Q(x, T_{ikz}) = Q(x, T_{ik,z+1}) - Q(x, T_{ikz})$. From (1), the unconditional risk density $g(x)$ for the general population is thus represented as

$$g(x) \approx \sum_{i,j,k} r_{ijk}\, g^{\star}_{ijk}(x), \tag{9}$$

where $g^{\star}_{ijk}(x)$ is given in (8). Since the operating condition j does not enter into the calculation for $Q(x, T)$, the sum over j can be evaluated separately, yielding

$$g(x) \approx \sum_{i,k,z} (\Delta p_{ikz}/\Delta T_{ikz})\, \Delta Q(x, T_{ikz}), \tag{10}$$

where

$$\Delta p_{ikz} = \sum_{j} r_{ijk}\, \Delta p_{jz}\,.$$

Because $Q(x, T)$ is defined recursively in x, eq. (10) is evaluated in this manner as well. For computational efficiency, the calculation of $Q(x, T)$ is represented as

$$Q(x, T) = \begin{cases} \text{COEF}(x, T) + Q(x-1, T), & x \geqslant 1 \\ -e^{-T}, & x = 0 \end{cases} \tag{11}$$

$$\text{COEF}(x, T) = \begin{cases} (T/x)(\text{COEF}(x-1, T)), & x \geqslant 1 \\ -e^{-T}, & x = 0. \end{cases} \tag{12}$$

However, the following underflow problem occurs in computing $Q(x, T)$: The values T_{ikz} in (10) represent the mean fatalities due to a release in quantity level z and population density stratum k. If the quantity (and hence the lethal area) and the population density are large, the fatalities will be significant as well. For catastrophic releases, T_{ikz} will be of the order of 100 or greater. In these cases, the values $Q(0, T_{ikz}) = \mathrm{COEF}\ (0, T_{ikz}) = -\exp(-T_{ikz})$ will soon be too small in magnitude for the arithmetic units of the computer to process, causing an underflow condition which in turn causes both $Q(0, T_{ikz})$ and COEF $(0, T_{ikz})$ to be regarded as zero. This, of course, will result in eqs. (11) and (12) yielding $Q(x, T_{ikz}) = 0$ for all x, a most undesirable result.

To address this, we note from (11) and (12) that COEF (x, T) is increasing in x for all $x < T$. In addition, for large T (i.e. greater than 50), the Poisson distribution with mean T is negligibly different from the normal distribution with mean and variance T, so that the average value $\overline{\mathrm{COEF}}$ of COEF (x, T) over the interval $0 \leqslant x \leqslant T$ satisfies the following relation

$$\sum_{x=0}^{T-1} \mathrm{COEF}\,(x, T) = T * \overline{\mathrm{COEF}} \approx \mathcal{N}(T; T, T) = 1/2,$$

where $\mathcal{N}(x; \mu, \sigma^2)$ is the cumulative probability distribution function of the normal distribution, with mean μ and variance σ^2. Thus, $\overline{\mathrm{COEF}} \cong 1/2T$.

Since COEF (x, T) is increasing in x for $0 \leqslant x < T$ and the average value $\overline{\mathrm{COEF}}$ is large enough to avoid underflow, there exists a value $x_0 < T$ such that the values COEF (x, T), $x_0 \leqslant x < T$, will also be large enough to avoid underflow. x_0 is thus defined as

$$x_0 = \min_x \left\{ \mathrm{e}^{-T}\, T^x/x! \geqslant \epsilon \right\},$$

where ϵ is the smallest quantity which the computer regards as nonzero.

If $x_0 = 0$ (i.e. if $\mathrm{e}^{-T} \geqslant \epsilon$ or, equivalently, if $T \leqslant \ln(1/\epsilon)$), then the computation of (11) and (12) is straightforward. If $x_0 > 0$, however, two problems arise:

(a) The computation of COEF (x_0, T) is not straightforward. We already know (since $x_0 > 0$) that e^{-T} is small enough to cause underflow, so that a naive approach would yield COEF $(x_0, T) = 0$. Furthermore, if x_0 is large, then both $x_0!$ and T^{x_0} would be large enough to cause overflow.

(b) The value of x_0 is not known a priori. The naive approach to determine it would be to calculate the quantity COEF $(x, T) = -\mathrm{e}^{-T}\, T^x/x!$ for $x = 0$, then for $x = 1$, then for $x = 2$, etc., until a value for x is reached for which COEF (x, T) does not underflow. If x_0 is large, however (which can easily happen if T itself is large), this becomes computationally very inefficient and time-consuming. We therefore need a reasonable upper bound for x_0 (an upper bound, so that it is known a priori that for x equal to the upper bound, COEF (x, T) will not underflow since $x > x_0$; and a reasonable upper bound, so that the quantities COEF (x, T) for x between x_0 and the upper bound, which are not taken into account in eq. (12), are in fact negligible).

To address the first problem, we take logarithms of the various terms in COEF (x_0, T) to obtain

$$\text{COEF}\,(x_0, T) = -\exp(x_0 \ln T - T - \sum_{k=1}^{x_0} \ln k). \tag{13}$$

None of the terms in the argument of the "exp" function in (13) exhibit overflow or underflow problems, and since $\epsilon \leqslant -\text{COEF}\,(x_0, T) < 1$, neither does the overall expression.

To address the second problem, we note that the quantity $-\text{COEF}\,(x, T) = \mathrm{e}^{-T}\, T^x/x!$ is the probability density function for a Poisson distribution with mean T. When T is large, as observed above, this distribution can be accurately represented by a normal distribution with mean and variance T. Let $x_0^\star$ be equal to a value seven standard deviations below the mean, i.e.

$$x_0^\star = T - 7\sqrt{T}. \tag{14}$$

Then

- Since T is large, and specifically since $T \geqslant 50$, $x_0^\star$ is positive.
- Since $x_0^\star$ is seven standard deviations below the mean, we have that

  $$\mathrm{e}^{-T}\, T^{(x_0^\star)}/x_0^\star! \approx \mathcal{N}'(-7)/\sqrt{T} \approx 10^{-12} > \epsilon,$$

  where $\mathcal{N}'(z)$ is the value of the probability density function corresponding to the normal distribution $\mathcal{N}(z) = \mathcal{N}(z; 0, 1)$. Thus, no underflow occurs for the Poisson distribution at $x_0^\star$ (note that $\epsilon = 10^{-38}$ on the computer we used, and on virtually all computers, ϵ is many orders of magnitude below 10^{-12}), and therefore $x_0^\star$ is an upper bound for x_0.
- Since

  $$\sum_{x=0}^{x_0-1} \mathrm{e}^{-T}\, T^x/x! \approx \mathcal{N}(-7) \approx 10^{-12},$$

  the accuracy lost by not taking into account the values of COEF arising from $x = 0, 1, \ldots, x_0^\star - 1$ is negligible.

Hence, $x_0^* = T - 7\sqrt{T}$ is in fact a reasonable upper bound for x_0, and thus a reasonable lower limit from which to begin the computation of COEF (x, T) and thus $Q(x, T)$.

It should also be noted that COEF (x, T) monotonically approaches zero as $x \to \infty$. Therefore, the computational efficiency will significantly improve with negligible loss of accuracy by stopping the recursive calculation of COEF (x, T) and $Q(x, T)$ when COEF (x, T), for $x > T$, decreases below a specified cutoff value. The algorithm turns out to be very sensitive to the value of this cutoff parameter, however, which must therefore be a very small value. In a sample run of the model which we carried out on a computer with an underflow limit of $\epsilon = 10^{-38}$, a cutoff value of 10^{-30} yielded accurate results, but a cutoff value of 10^{-15}, while still very small in comparison to a total probability of unity, nevertheless caused significant inaccuracy in the resulting probability density function (the density function summed to a value significantly less than 1).

3.3. COMPUTATION OF THE DISCRETE CONVOLUTIONS

The risk density for the general population is combined with that for the professional population to obtain the overall risk density for a hazardous material release incident. The probability of the number of professional fatalities is assumed to follow a Poisson distribution denoted by $h(y; \bar{Y})$, where $\bar{Y}$, the mean of the distribution, is the average number of professional fatalities in a hazardous material release incident, as estimated from historical data.

The per-incident risk density is given by

$$f(x) = g \star h(x) = \sum_{y=0}^{x} g(x-y)\,h(y;\bar{Y}), \tag{15}$$

where $g \star h$ denotes the discrete convolution of g and h. The per-annum risk density for the region is

$$u(x) = \sum_{n=0}^{\infty} f^{(n)}(x)\,h(n;\bar{N}), \tag{16}$$

assuming that the number of incidents per year is Poisson distributed with mean $\bar{N}$, i.e. the probability of n releases occurring in a year is $h(n;\bar{N})$. By using Fourier transforms, particularly the Fast Fourier Transform (FFT) [4], $u(x)$ is computed through the transform equation

$$u^{\#}(s) = \sum_{n=0}^{N} (f^{\#}(s))^n\, h(n;\bar{N}), \tag{17}$$

where $f^{\#}(s)$ and $u^{\#}(s)$ are the transforms of $f(x)$ and $u(x)$, respectively, and where N is chosen so that the probability of more than N releases per year is negligibly small. By using FFT and eq. (17), the computation time is proportional to $NX \ln X$ instead of NX^2, where X is the largest number of fatalities to be considered. This represents a very large computational saving when X is large.

Note that the calculation in (17) is best carried out by means of Horner's method, i.e.

$$H_n = \begin{cases} h(N;\bar{N}), & n = N, \\ H_{n+1} f^{\#}(s) + h(n;\bar{N}), & n = N-1, \ldots, 0, \end{cases}$$

from which we obtain $u^{\#}(s) = H_0$. This approach requires only N multiplications, rather than the $2N - 1$ multiplications required by the direct calculation of (17). This cuts roughly in half the total number of multiplications required to compute $u^{\#}(s)$.

3.4. MEANS AND VARIANCES

In this section, we develop formulae to obtain accurate means and variances of the per-incident and per-annum fatalities without having to first compute the risk densities themselves. This is a definite advantage, especially if, as is frequently the case, one needs only the means and variances in order to answer the risk assessment question under consideration. Even if the actual risk densities are in fact required, independently-obtained means and variances provide accuracy checks and help determine how many points of the risk densities need to be computed.

Mean fatalities per incident and per annum

According to eq. (15), $f(x) = g \star h(x)$, so that $\bar{f} = \bar{g} + \bar{h}$. Since $h(x)$ is a Poisson distribution, $\bar{h} = \bar{Y}$. From eq. (9),

$$\bar{g} \approx \sum_{i,j,k} r_{ijk} \bar{g}^{\star}_{ijk}, \tag{18}$$

where $\bar{g}^{\star}_{ijk}$ is the mean of $g^{\star}_{ijk}(x)$. It therefore remains to find $\bar{g}^{\star}_{ijk}$. From eq. (5),

$$\bar{g}^{\star}_{ijk} = \sum_{z=0}^{M} (\Delta p_{jz}/\Delta T_{ikz}) \int_{T_{ikz}}^{T_{ik,z+1}} \sum_{x=0}^{\infty} x h(x;T)\, \mathrm{d}T. \tag{19}$$

But since $h(x;T)$ is a Poisson distribution, the integrand is simply T. Hence, after some simplification,

$$\bar{g}^{\star}_{ijk} = \sum_{z=0}^{M} 1/2(\Delta p_{jz})(T_{ik,z+1} + T_{ikz}). \tag{20}$$

Therefore

$$\bar{g} \approx \sum_{i,j,k} r_{ijk} \sum_{z=0}^{M} 1/2(\Delta p_{jz})(T_{ik,z+1} + T_{ikz}). \tag{21}$$

The mean number of total fatalities per accident $\bar{f}$ is therefore given by

$$\bar{f} \approx \bar{Y} + \sum_{i,j,k} r_{ijk} \sum_{z=0}^{M} 1/2(\Delta p_{jz})(T_{ik,z+1} + T_{ikz}). \tag{22}$$

From eq. (16)

$$\bar{u} = \sum_{n=0}^{\infty} \left(\sum_{x=0}^{\infty} x f^{(n)}(x) \right) h(n;\bar{N}) = \sum_{n=0}^{\infty} (n\bar{f}) h(n;\bar{N}) = \bar{f}\bar{N}. \tag{23}$$

The mean number of fatalities per annum $\bar{u}$ is therefore given by

$$\bar{u} \approx \bar{N}(\bar{Y} + \sum_{i,j,k} r_{ijk} \sum_{z=0}^{M} 1/2(\Delta p_{jz})(T_{ik,z+1} + T_{ikz})). \tag{24}$$

Variance of fatalities per incident and per annum

Since $f(x) = g \star h(x)$ is the probability distribution of the sum of two random variables, the variance of $f(x)$ is the sum of the respective variances of $g(x)$ and $h(x)$, i.e.

$$\sigma_f^2 = \sigma_g^2 + \sigma_h^2 . \tag{25}$$

Since $h(y;\bar{Y})$ is Poisson, $\sigma_h^2 = \bar{Y}$. As for σ_g^2, we have

$$\sigma_g^2 = \overline{g^2} - \bar{g}^2 , \tag{26}$$

where $\bar{g}$ is given by eq. (21). $\overline{g^2}$, the second moment of $g(x)$, is given by

$$\overline{g^2} \approx \sum_{i,j,k} r_{ijk} \overline{g^{2}_{ijk}}^{\star}, \tag{27}$$

where

$$\overline{g^{2}_{ijk}}^{\star} = \sum_{z=0}^{M} (\Delta p_{jz}/\Delta T_{ikz}) \int_{T_{ikz}}^{T_{ik,z+1}} \sum_{x=0}^{\infty} x^2 h(x;T)\,\mathrm{d}T. \tag{28}$$

The inner sum in (28) is equal to $T + T^2$. The variance of the number of fatalities per accident is therefore given as

$$\sigma_f^2 \approx \bar{g} + \left[\sum_{i,j,k} r_{ijk} \sum_{z=0}^{M} (1/3)(\Delta p_{jz})(T^2_{ik,z+1} + T_{ik,z+1}T_{ikz} + T^2_{ikz})\right] - \bar{g}^2 + \bar{Y}, \tag{29}$$

where $\bar{g}$ is given by eq. (21).

The number of fatalities per year U is a random variable which depends on the random variable N, the number of hazardous material release incidents in a year. According to the compound-variance formula (see Parzen [5]), the variance of U is given by

$$\begin{aligned}\sigma_u^2 &= E(\mathrm{Var}\,(U|N)) + \mathrm{Var}\,(E(U|N)) \\ &= \sigma_f^2 E(N) + \bar{f}^2\,\mathrm{Var}\,(N).\end{aligned}$$

Since N is a Poisson random variable, $E(N) = \mathrm{Var}\,(N) = \bar{N}$. Therefore, the variance of the number of fatalities per annum is given by

$$\sigma_u^2 = \bar{N}(\sigma_f^2 + \bar{f}^2), \tag{30}$$

where σ_f^2 is given by eq. (29) and $\bar{f}$ by eq. (22).

4. Illustrative example

To illustrate the risk profile computation scheme we have discussed, we shall use data from Glickman and Rosenfield [3] on the release of hazardous materials in train derailments. The data provide for fifteen strata of density of exposed population (so that $k = 1, 2, \ldots, 15$), sixteen classes of release quantity (so that $z = 0, 1, \ldots, 15$), five different operating conditions corresponding to five levels of track quality (so that

Table 1

Levels of population density

Level (k)	Density (ρ_k) (per sq. mile)
1	250
2	750
3	1250
4	1750
5	2250
6	2750
7	3250
8	3750
9	4250
10	4750
11	5250
12	5750
13	8000
14	12 500
15	25 000

Table 2

Levels of release quantity

Level (z)	Quantity (q_z) (gallons)
0	0
1	1000
2	5000
3	9000
4	15 000
5	19 000
6	25 000
7	35 000
8	45 000
9	55 000
10	75 000
11	95 000
12	125 000
13	175 000
14	225 000
15	275 000

Table 3

Probability of an incident of type i

i	$\Sigma_j \Sigma_k r_{ijk}$
1	0.1043
2	0.3032
3	0.4022
4	0.0005
5	0.0811
6	0.0003
7	0.0045
8	0.1039

Table 4

Representative values for lethal areas (sq. km.) (based on release of flammable compressed gases)

Spill size (gal)	Scenario (probability) Flame jet (0.30)	Pool fire (0.28)	Fireball (0.20)	Rocketing (0.20)	Vapor cloud (0.02)
1 – 100	1.9×10^{-4}	–	–	6.0×10^{-4}	4.4×10^{-4}
100 – 1000	1.9×10^{-4}	2.0×10^{-4}	1.2×10^{-4}	6.0×10^{-4}	6.1×10^{-4}
1000 – 5000	1.9×10^{-4}	1.3×10^{-3}	5.0×10^{-4}	6.0×10^{-4}	1.1×10^{-3}
5000 – 10 000	1.9×10^{-4}	3.0×10^{-3}	9.4×10^{-4}	6.0×10^{-4}	1.2×10^{-3}
10 000 – 25 000	1.9×10^{-4}	6.6×10^{-3}	1.7×10^{-3}	6.0×10^{-4}	1.7×10^{-3}
25 000 – 50 000	1.9×10^{-4}	1.3×10^{-2}	2.8×10^{-3}	6.0×10^{-4}	2.1×10^{-3}
50 000 – up	1.9×10^{-4}	2.6×10^{-2}	4.4×10^{-3}	6.0×10^{-4}	2.3×10^{-3}

Table 5

Risk probabilities and means for illustrative example

Result	Probability of occurring: per incident $[f(x)]$	Probability of occurring: per year $[u(x)]$
No fatalities	0.945	0.015
At least 1 fatality	0.550	0.985
At least 10 fatalities	7.6×10^{-4}	0.190
At least 50 fatalities	2.6×10^{-5}	2.6×10^{-3}
At least 100 fatalities	8.1×10^{-6}	7.0×10^{-4}
At least 500 fatalities	2.4×10^{-7}	5.3×10^{-6}

	Per incident	Per year
Mean rate of fatalities	0.089	6.75
Variance	0.67	51.16

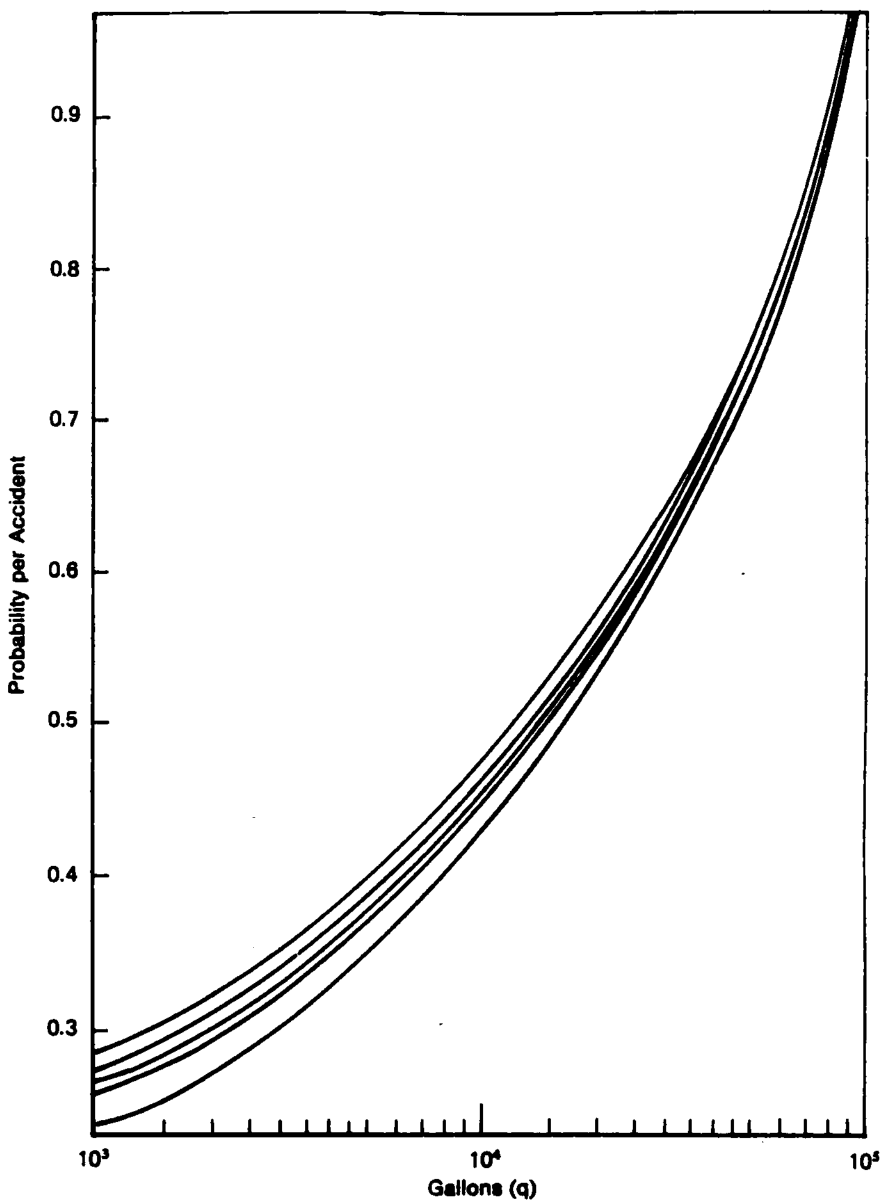

Fig. 1. Cumulative release-size probability (curves range from track quality 1 (top) to track quality 5 (bottom).

$j = 1, 2, \ldots, 5$), eight types of hazardous materials (so that $i = 1, 2, \ldots, 8$). Table 1 shows the values of the population densities ρ_k, while table 2 shows the values of the release quantities q_z. Table 3 provides a summary, for each value of i, of the joint probabilities r_{ijk} subtotaled over all j and k. The values of the release-quantity probabilities p_{jz} are plotted in fig. 1, while the values of the lethal area sizes $A_{iz} = A_i(q_z)$ are computed from

$$A_i(q_z) = \sum_s \prod_i(s)\, A_i(q_z|s),$$

where $\Pi_i(s)$ is the conditional probability of release scenario s for each value of i, and $A_i(q_z|s)$ is the average size of the lethal area for a release quantity q, given s, for each value of i. Representative values for scenario release probabilities and lethal areas appear in table 4. Other parameters of interest are the mortality rate υ_i, which is 0.002 for all i, the mean number of professional fatalities per derailment incident $\bar{Y} = 0.056$, and the mean number of derailment incidents per annum $\bar{N} = 79.75$.

From the p_{jz} and r_{ijk} values we compute p_{ikz}, and from the υ_i, ρ_k, and A_{iz} values we compute ΔT_{ikz} and $\Delta Q(x, T_{ikz})$. These results are then used to compute the general population fatality probability distribution $g(x)$ according to (10). From $g(x)$ and $\bar{Y}$ we then compute the overall per-incident fatality probability density $f(x)$ according to (15), and from $f(x)$ and $\bar{N}$ we in turn compute the per-annum fatality probability distribution $u(x)$ according to (16). Table 5 displays selected values of $f(x)$ and $u(x)$ (which also appear in [3]), as well as the means and variances $\bar{f}$, $\bar{u}$, σ_f^2, and σ_u^2.

Acknowledgement

We wish to acknowledge Dr. Peter H. Mengert, also at the Transportation Systems Center, as a co-developer of the concepts described in this paper.

References

[1] S. Kaplan and B. Garrick, On the quantitative definition of risk, Risk Analysis 1, 1(1981)11.
[2] J. Fiksel and D. Rosenfield, Probabilitistic models for risk assessment, Risk Analysis 2, 1(1982)1.
[3] T. Glickman and D. Rosenfield, Derailments and release of hazardous materials, Management Science 30, 4(1984)503.
[4] Special issue on Fast Fourier Transforms, IEEE Audio Transactions (1967).
[5] E. Parzen, *Stochastic Processes* (Holden-Day, San Francisco, 1962).