science for a changing world

Probability Models for Estimation of Number and Costs of Landslides

by Robert A. Crovelli[1]

Open-File Report 00-249

2000

U.S. DEPARTMENT OF THE INTERIOR
U.S. GEOLOGICAL SURVEY

[1]U.S. Geological Survey, MS 939, Box 25046, Denver Federal Center, Denver, Colorado 80225

Abstract

The objective of this report is to describe the development of probability models for estimation of the number and costs of landslides during a specified time. Important philosophical ideas about natural processes and probability models are presented first. Then two probability models for the number of landslides that occur during a specified time are investigated: a continuous-time model (Poisson model) and a discrete-time model (binomial model). Estimation theory is developed for the estimation of the parameters of both of the models. The exceedance probability of one or more landslides during a specified time is formulated for both models. The estimation theory and probability formulation of the Poisson model are applied to the future occurrence of landslides in Seattle, Washington, using historical data from 1909 to 1997. Theoretical and numerical comparisons between the Poisson and binomial models are conducted that show the binomial model is an approximation to the Poisson model. An economic probability model is developed as an addition to the Poisson model for the estimation of the total damage from future landslides in terms of economic loss as costs in dollars. For illustrative purposes the economic probability model is applied to damaging landslides caused by El Nino rainstorms within the winter season 1997-98 in the San Francisco Bay region, California.

Philosophy of Probability Models

Natural Processes

Important philosophical ideas about natural processes:

- Determinism or the law of cause and effect is the doctrine that all events in the universe are deterministic: every event has a cause.

- At the scale of geologic and atmospheric hazards (e.g., landslides, earthquakes, floods, tsunamis, volcanoes, and storms), nature is deterministic: every hazardous event has a cause.

- A hazard process is a physical process involving the occurrence of point (hazardous) events in time.

- Beginning at some point in time, after a certain amount of time, the first hazardous event occurs. Then, after a certain amount of time, the second hazardous event occurs. And so forth. The time between hazardous events is certainly irregular.

- We cannot predict exactly when a hazard event will occur because of the limitations to our knowledge of nature.

- The limitations to our knowledge of nature are explained by the following: Heisenberg Uncertainty Principle and Godel's Theorem; chaos theory and fractal geometry; algorithmic and computational complexity; physical and financial constraints.

- Chaos is the *apparent* randomness from extremely complex behavior occurring in a deterministic process due to excessive sensitivity of an event to small changes in initial conditions.

Probability Models

Important philosophical ideas about probability models:

- Probability is a numerical measure of *our* uncertainty regarding nature.
- A probability model is a mathematical model that incorporates our uncertainty.
- Probability models are an approach to deal with the limitations to our knowledge of natural processes.
- Probability models are used for purposes of description and prediction of physical processes in nature.
- Randomness is an assumption of probability models, not natural processes. Hazards do not occur at random in nature, but they do occur at random in the models.
- It is not correct to say that a natural process follows a particular probability model. (This would be putting the cart before the horse.)
- We will always be uncertain of nature because of our limitations in understanding.

In summary, hazard processes are deterministic, but because of our limitations when studying hazards, we resort to probability models that incorporate our uncertainty.

Probability Models for Landslides

Consider the occurrence of landslides during a specified time in a particular area.

Denote

$N(t)$: Number of landslides that occur during time t in a particular area

We are interested in deriving a formula for calculating the probability of one or more landslides during a specified time t. That is,

$P\{N(t) \geq 1\}$

Two probability models for $N(t)$ will be investigated: first, a continuous-time model and second, a discrete-time model.

Poisson Model for Number of Landslides

The Poisson model is a continuous-time model consisting of the occurrence of random point-events (landslides) in ordinary time which is naturally continuous. The Poisson model is the most commonly used model for the occurrence of random point-events in time and has been used in modeling the occurrence of earthquakes.

Assumptions of the Poisson model:

- The numbers of events (landslides) which occur in disjoint time intervals are independent.
- The probability of an event occurring in a very short time interval is proportional to the length of the time interval. The probability of more than one event in such a short time interval is negligible.
- The probability distribution of the number of events remains the same for all time intervals of a fixed length.

It is important to acknowledge that these assumptions may not completely hold for the occurrence of landslides, especially the independence assumption. However, given a certain lack of understanding of the physical processes that control landslides, the Poisson model represents the best first-approximation model in attempting to model their occurrence. A first-approximation model is often applied in mathematical modeling when the assumptions are not completely satisfied by the physical process. Usually the first-approximation model is relatively easy to work with and is mathematically tractable. A more accurate model might be extremely complex and not mathematically tractable.

Poisson Distribution -- Probability of n landslides during time t:

$$P\{N(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \qquad n = 0, 1, 2, \ldots$$

where

λ: Rate of occurrence of landslides

Note that time t is specified, whereas rate λ is estimated.

Definition of recurrence intervals $\{T_i, i = 1, 2, \ldots, n\}$:

T_1: Time until the first landslide

T_i: Time between the $(i-1)$st and the ith landslide for $i > 1$

Note that n landslides will have n recurrence intervals.

Theorem – Recurrence intervals $\{T_i, i = 1, 2, \ldots, n\}$ are independent identically distributed exponential random variables having mean recurrence interval (μ) equal to the reciprocal of the rate of occurrence, i.e., $\mu = 1/\lambda$.

For landslides, the mean recurrence interval (μ) is the average time interval between landslides.

Note

$\lambda = 1/\mu$

Variance of T_i is

$V[T_i] = 1/\lambda^2 = \mu^2$

Probability of a recurrence interval being greater than time t

$$P\{T_i > t\} = P\{N(t) = 0\} = e^{-\lambda t} = e^{-t/\mu}$$

Probability of one or more landslides during time t (exceedance probability)

$$P\{N(t) \geq 1\} = 1 - P\{N(t) = 0\} = 1 - e^{-\lambda t} = 1 - e^{-t/\mu}$$

Note

If t is fixed and $\mu \to \infty$, then $P\{N(t) \geq 1\} \to 0$.

If μ is fixed and $t \to \infty$, then $P\{N(t) \geq 1\} \to 1$.

Mean or expected value of $N(t)$ is

$E[N(t)] = \lambda t = t/\mu$

Note that the smaller the μ, the larger the $E[N(t)]$.

Variance of $N(t)$ is

$V[N(t)] = \lambda t = t/\mu$

Estimation of the Parameters λ and μ in the Poisson Model

A Poisson model having an unknown rate λ is to be observed for a fixed time t^*.

We want to determine a statistic that is a good estimator of the parameter λ.

It can be shown (Ross, 1972) that the maximum likelihood estimator of λ, denoted by R, is given by

$$R = \frac{N(t^*)}{t^*}$$

Mean or expected value of R is

$$E[R] = E[N(t^*)/t^*] = \lambda t^*/t^* = \lambda$$

(R is an unbiased estimator of λ)

Variance of R is

$$V[R] = V[N(t^*)/t^*] = \lambda t^*/(t^*)^2 = \lambda/t^*$$

Theorem – The statistic R is the unique minimum variance unbiased estimator of λ.

Since $\mu = 1/\lambda$, a statistic that is a good estimator of the parameter μ is

$$M = 1/R = \frac{t^*}{N(t^*)}$$

Consider another statistic as an estimator of the parameter μ

$$M' = \frac{\sum_{i=1}^{N(t^*)} T_i}{N(t^*)}$$

where

T_i: The ith observed recurrence interval ($i = 1, 2, \ldots, N(t^*)$)

Because

$$\sum_{i=1}^{N(t^*)} T_i \le t^*$$

The estimator M' will tend to be biased low and underestimate μ.

Example

Suppose a record of the occurrence of landslides for the past 100 years (t^*) showed that 5 landslides (n) occurred.

An estimate of λ would be

$r = n/t^* = 5/100 = 1/20 = 0.05$

Hence, we expect landslides at a rate of 0.05 per year.

An estimate of μ would be

$m = 1/r = t^*/n = 100/5 = 20$

Therefore, we expect the mean recurrence interval to be 20 years.

From either of these estimates of the parameters λ and μ, we could calculate the probability of one or more landslides during a future time t

$$P\{N(t) \geq 1\} = 1 - e^{-\lambda t} = 1 - e^{-t/\mu}$$

Using $m = 20$ and specifying $t = 50$, we get

$$P\{N(50) \geq 1\} = 1 - e^{-50/20} = 0.918$$

There is a 91.8% chance of one or more landslides occur during the next 50 years.

Application

Seattle, Washington, has kept records of landslide occurrence from 1909 to present. Records from 1909 to 1997 ($t^* = 88.4$ years) were analyzed to determine landslide density using a moving count circle approach (Coe and others, 2000). This analysis showed that n landslides occurred within each count circle, where n ranged from 0 to 30. The Poisson model was applied to these data, and the results are given in table 1.

Table 1. Poisson model for percent chance of one or more landslides in Seattle, Washington, during a specified time

Poisson Model for Number of Landslides in Seattle					R.A. Crovelli		
Percent Chance of One or More Landslides During a Specified Time							
Past (years):	88.4						
Number of Landslides	Mean Recurrence Interval (years)	Time (years) 1	5	10	25	50	100
1	88.4	1.124847	5.499124	10.69585	24.63336	43.19869	67.73612
2	44.2	2.237042	10.69585	20.24768	43.19869	67.73612	89.59042
3	29.46666667	3.336726	15.60679	28.77786	57.19076	81.67369	96.64146
4	22.1	4.424041	20.24768	36.39567	67.73612	89.59042	98.91641
5	17.68	5.499124	24.63336	43.19869	75.68379	94.08722	99.65039
6	14.73333333	6.562115	28.77786	49.27407	81.67369	96.64146	99.8872
7	12.62857143	7.613149	32.69446	54.69964	86.18808	98.09231	99.96361
8	11.05	8.65236	36.39567	59.5449	89.59042	98.91641	99.98826
9	9.822222222	9.679882	39.89335	63.87191	92.15465	99.3845	99.99621
10	8.84	10.69585	43.19869	67.73612	94.08722	99.65039	99.99878
11	8.036363636	11.70038	46.32227	71.18701	95.54374	99.80142	99.99961
12	7.366666667	12.69362	49.27407	74.2688	96.64146	99.8872	99.99987
13	6.8	13.67568	52.06356	77.02097	97.46878	99.93593	99.99996
14	6.314285714	14.6467	54.69964	79.47877	98.09231	99.96361	99.99999
15	5.893333333	15.60679	57.19076	81.67369	98.56224	99.97933	100
16	5.525	16.55609	59.5449	83.63385	98.91641	99.98826	100
17	5.2	17.4947	61.76957	85.38434	99.18333	99.99333	100
18	4.911111111	18.42276	63.87191	86.94761	99.3845	99.99621	100
19	4.652631579	19.34038	65.85864	88.34368	99.53612	99.99785	100
20	4.42	20.24768	67.73612	89.59042	99.65039	99.99878	100
21	4.20952381	21.14477	69.51035	90.70381	99.73651	99.99931	100
22	4.018181818	22.03177	71.18701	91.69812	99.80142	99.99961	100
23	3.843478261	22.9088	72.77147	92.58607	99.85033	99.99978	100
24	3.683333333	23.77595	74.2688	93.37906	99.8872	99.99987	100
25	3.536	24.63336	75.68379	94.08722	99.91499	99.99993	100
26	3.4	25.48112	77.02097	94.71964	99.93593	99.99996	100
27	3.274074074	26.31934	78.28462	95.28442	99.95171	99.99998	100
28	3.157142857	27.14814	79.47877	95.78879	99.96361	99.99999	100
29	3.048275862	27.96761	80.60726	96.23922	99.97257	99.99999	100
30	2.946666667	28.77786	81.67369	96.64146	99.97933	100	100

Binomial Model for Number of Landslides

Costa and Baker (1981) give a probability model that they used in flood hazard analyses for modeling the occurrence of floods. The Costa-Baker model was also used by Keaton and others (1988) and Lips and Wieczorek (1990) in modeling the occurrence of debris flows. The Costa-Baker model was given without any derivation as follows, written in the notation of this paper:

$$P\{N(t) \geq 1\} = 1 - (1 - 1/\mu)^t$$

with the mean recurrence interval $\mu = 1/p$ and t is number of years

where

p: Probability of a flood in any one year

The Costa-Baker model is a crude model in that it divides time into fixed discrete increments (one-year increments). It is designed for large values of t and μ.

The Costa-Baker model is actually an example of the binomial model.

The binomial model is a discrete-time model consisting of the occurrence of random point-events (landslides) in discrete time; that is, time is partitioned into a series of discrete increments of the same length and within each increment a single point-event (landslide) may or may not occur.

Assumptions of the binomial model:

- There are t independent "trials" (relatively small time-increments of fixed length).
- Each trial results in a "success" (landslide) or a "failure" (no landslide).
- The probability of success, p, remains the same from trial to trial.

Binomial Distribution -- Probability of n landslides during discrete time t:

$$P\{N(t) = n\} = C_n^t p^n (1-p)^{t-n} \qquad n = 0, 1, 2, \ldots$$

where

$$C_n^t = \frac{t!}{n!(t-n)!}$$

Note that time t is specified, whereas probability p is estimated.

Definition of recurrence intervals $\{T_i, i = 1, 2, \ldots, n\}$:

T_1: Number of time increments until the first landslide

T_i: Number of time increments between the $(i-1)$st and until the ith landslide for $i > 1$

Note that n landslides will have n recurrence intervals.

Theorem – Recurrence intervals $\{T_i, i = 1, 2, \ldots, n\}$ are independent identically distributed geometric random variables having mean recurrence interval (μ) equal to the reciprocal of the probability of success, i.e., $\mu = 1/p$.

For landslides, the mean recurrence interval (μ) is the average time interval between landslides.

Since

$$p = 1/\mu$$

the larger the mean recurrence interval (μ), the smaller the probability of a landslide in any one year (p). Also, if $\mu < 1$, then $p > 1$ which is not allowed. Therefore, the binomial model has the restriction $\mu \geq 1$.

Variance of T_i is

$$V[T_i] = (1-p)/p^2 = \mu(\mu - 1)$$

Probability of a recurrence interval being greater than time t

$$P\{T_i > t\} = P\{N(t) = 0\} = (1-p)^t = (1-1/\mu)^t$$

Probability of one or more landslides during time t (exceedance probability)

$$P\{N(t) \geq 1\} = 1 - P\{N(t) = 0\} = 1-(1-p)^t = 1-(1-1/\mu)^t$$

where the last expression is the Costa-Baker model.

Note that $1-p$ is raised to the t power under the assumption of independence.

Mean or expected value of $N(t)$ is

$$E[N(t)] = tp = t/\mu$$

Variance of $N(t)$ is

$$V[N(t)] = tp(1-p) = (t/\mu)(1 - 1/\mu)$$

Estimation of the Parameters p and μ in the Binomial Model

A binomial model having an unknown probability p is to be observed for a fixed time t^*.

We want to determine a statistic that is a good estimator of the parameter p.

It can be shown that the maximum likelihood estimator of p, denoted by F, is given by the relative frequency of occurrence

$$F = \frac{N(t^*)}{t^*}$$

Recall

t^*: Number of one-year increments

$N(t^*)$: Number of one-year increments in which a landslide occurred

Mean or expected value of F is

$E[F] = E[N(t^*)/t^*] = t^*p/t^* = p$

(F is an unbiased estimator of p)

Theorem – The statistic F is the unique minimum variance unbiased estimator of p.

Since $\mu = 1/p$, a statistic that is a good estimator of the parameter μ is

$$M^* = 1/F = \frac{t^*}{N(t^*)}$$

Example

Suppose a record of the occurrence of landslides for the past 100 years (t^*) showed that 5 landslides (n) occurred. Assume that the 5 landslides occur in 5 individual one-year increments.

An estimate of p would be

$f = n/t^* = 5/100 = 1/20 = 0.05$

Hence, we expect the probability of a landslide in any one-year increment to be 0.05.

An estimate of μ would be

$m^* = 1/f = t^*/n = 100/5 = 20$

Therefore, we expect the mean recurrence interval to be 20 years.

From either of these estimates of the parameters p and μ, we could calculate the probability of one or more landslides during a future time t

$$P\{N(t) \geq 1\} = 1 - (1 - p)^t = 1 - (1 - 1/\mu)^t$$

Using $m = 20$ and specifying $t = 50$, we get

$$P\{N(50) \geq 1\} = 1 - (1 - 1/20)^{50} = 0.923$$

There is a 92.3% chance of one or more landslides during the next 50 years.

The Binomial Model is an Approximation to the Poisson Model

Poisson model

$$P\{N(t) \geq 1\} = 1 - e^{-t/\mu} = 1 - (e^{-1/\mu})^t$$

Binomial model

$$P\{N(t) \geq 1\} = 1 - (1 - 1/\mu)^t$$

Now compare $e^{-1/\mu}$ and $1 - 1/\mu$

Recall the exponential series

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \cdots \qquad (x \text{ real})$$

Then

$$e^x \cong 1 + x \qquad (-1 < x < 1)$$

Let

$x = -1/\mu$

Thus

$1 - 1/\mu$ is equal to the first two terms of the exponential series for $e^{-1/\mu}$.

Hence

$e^{-1/\mu} \cong 1 - 1/\mu$

Therefore

$1-(e^{-1/\mu})^t \cong 1-(1-1/\mu)^t$

For a numerical comparison between the Poisson model and the binomial model see tables 2 (Poisson model) and 3 (binomial model) which give the results from the application of each model to a generic data set. The binomial model significantly over estimates the exceedance probabilities for relatively short mean recurrence intervals (a few years) and short periods of time. For example, when the mean recurrence interval is two years and the specified time is one year, the exceedance probability is equal to 50% using the binomial model, whereas it is equal to 39.3% using the Poisson model. The difference between the two models becomes negligible for longer mean recurrence intervals and longer time periods. This obviously is significant to hazards because events with short mean recurrence intervals and short time periods (less than 25 years) play a major role in determining the degree of hazard.

Theorem – The binomial distribution is an approximation to the Poisson distribution.
Given

Poisson distribution with parameters
t: Specified time (number of years)
λ: Rate of occurrence of events (landslides)

Binomial distribution with parameters
ν: Number of "trials" (number of time increments)
p: Probability of occurrence of a "success" (landslide) in any trial

When ν tends to infinity, and p tends to zero, but means $\lambda t = \nu p$ remain constant,

then the binomial distribution approaches the Poisson distribution.

For a proof of this theorem see Walpole and Myers (1989).

Table 2. Poisson model for percent chance of one or more landslides during a specified time based on a generic data set.

Poisson Model for Number of Landslides				R.A. Crovelli		
Percent Chance of One or More Landslides During a Specified Time						
Mean Recurrence	Time (years)					
Interval (years)	1	5	10	25	50	100
1	63.21206	99.32621	99.99546	100	100	100
2	39.34693	91.7915	99.32621	99.99963	100	100
5	18.12692	63.21206	86.46647	99.32621	99.99546	100
10	9.516258	39.34693	63.21206	91.7915	99.32621	99.99546
20	4.877058	22.11992	39.34693	71.34952	91.7915	99.32621
50	1.980133	9.516258	18.12692	39.34693	63.21206	86.46647
100	0.995017	4.877058	9.516258	22.11992	39.34693	63.21206
200	0.498752	2.469009	4.877058	11.75031	22.11992	39.34693
500	0.1998	0.995017	1.980133	4.877058	9.516258	18.12692
1000	0.09995	0.498752	0.995017	2.469009	4.877058	9.516258
2000	0.049988	0.249688	0.498752	1.24222	2.469009	4.877058
5000	0.019998	0.09995	0.1998	0.498752	0.995017	1.980133
10000	0.01	0.049988	0.09995	0.249688	0.498752	0.995017

Table 3. Binomial model for percent chance of one or more landslides during a specified time based on a generic data set.

Binomial Model for Number of Landslides				R.A. Crovelli		
Percent Chance of One or More Landslides During a Specified Time						
Mean Recurrence	Time (years)					
Interval (years)	1	5	10	25	50	100
1	100	100	100	100	100	100
2	50	96.875	99.90234	100	100	100
5	20	67.232	89.26258	99.62221	99.99857	100
10	10	40.951	65.13216	92.82102	99.48462	99.99734
20	5	22.62191	40.12631	72.26104	92.3055	99.40795
50	2	9.60792	18.29272	39.65353	63.58303	86.73804
100	1	4.900995	9.561792	22.21786	39.49939	63.39677
200	0.5	2.475125	4.888987	11.77798	22.16874	39.42296
500	0.2	0.996008	1.982096	4.88182	9.525318	18.14332
1000	0.1	0.499001	0.995512	2.470229	4.879437	9.520785
2000	0.05	0.24975	0.498876	1.242529	2.469619	4.878247
5000	0.02	0.09996	0.19982	0.498802	0.995116	1.980329
10000	0.01	0.04999	0.099955	0.2497	0.498777	0.995066

Probability Model for Costs of Landslides

Damage due to landslides will be taken in the form of economic loss as costs in dollars. However, the theory below would also apply to other types of damage as in the case of human loss in deaths.

Given

$$Y(t) = \sum_{i=1}^{N(t)} X_i$$

where

$N(t)$: Number of landslides that occur during time t in a particular area

$N(t)$ has a Poisson distribution with rate λ.

X_i: The amount of damage (cost) from the ith landslide

The X_i ($i = 1, 2, \ldots$) are independent and identically distributed random variables which are also independent of $N(t)$.

$Y(t)$: The total amount of damage (costs) from all of the landslides during time t

Then

Mean or expected value of $Y(t)$ is

$$\mu_Y = E[Y(t)] = \lambda t E[X]$$

Variance of $Y(t)$ is

$$\sigma_Y^2 = V[Y(t)] = \lambda t\{ V[X] + (E[X])^2\}$$

From an observed number of landslides n, the sample mean cost M_X is an estimator of $E[X]$ where

$$M_X = \frac{\sum_{i=1}^{n} X_i}{n}$$

The sample variance S_X^2 is an estimator of $V[X]$ where

$$S_X^2 = \frac{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}{n(n-1)}$$

In the case that only observed estimates of the minimum value of X, Min(X), and maximum value of X, Max(X), are available, then an estimator of the standard deviation of X would be

$$S_X = \frac{Max(X) - Min(X)}{6}$$

The divisor of 6 is based on plus and minus three standard deviations from the mean for a range of 6 standard deviations.

The Pareto probability distribution is possibly a good approximate distribution for the random variable X.

Crovelli (1992) showed that the lognormal probability distribution is a good approximate distribution for the type of random variable $Y(t)$. Hence, the fractiles of $Y(t)$ can be approximated by using the lognormal distribution. As derived in Crovelli (1992), the characterizing parameters of the lognormal distribution, namely mu (μ^*) and sigma (σ^*), can be calculated from the mean μ_Y and standard deviation σ_Y of a lognormal random variable Y as follows

$$\mu^* = \ln\left(\frac{\mu_Y^2}{\sqrt{\mu_Y^2 + \sigma_Y^2}}\right)$$

$$\sigma^* = \sqrt{\ln(\sigma_Y^2 / \mu_Y^2 + 1)}$$

Knowing the lognormal characterizing parameters, the lognormal fractiles can be calculated from the formula

$$F100\alpha = e^{\mu^* + z_\alpha \sigma^*} \quad 0 \leq \alpha \leq 1$$

Where Z is a standard normal random variable and $P\{Z > z_\alpha\} = \alpha$.

For example, two fractiles of interest in this report are

$$F95 = e^{\mu^* - 1.645\sigma^*} \quad and \quad F5 = e^{\mu^* + 1.645\sigma^*}$$

There is a 95% chance of exceeding $F95$, and a 5% chance of exceeding $F5$. Together, the low value of $F95$ and the high value of $F5$ form a range of values that is a 90%

prediction interval for $Y(t)$, the total costs from landslides during a specified time, (at a 90% confidence level).

The reverse problem would be to find the probability of exceeding a specified amount in economic loss due to landslides in a particular area during a specified time. That is, given y_α, find α such that

$$P\{Y(t) > y_\alpha\} = \alpha$$

Normalizing

$$z_\alpha = \frac{\ln y_\alpha - \mu^*}{\sigma^*}$$

Now, from z_α, find α such that $P\{Z > z_\alpha\} = \alpha$.

The aggregation of the total amounts of damage (costs) from landslides in k areas:

$$W = \sum_{i=1}^{k} Y_i$$

where Y_i: The total amount of damage (costs) from landslides in the ith area.

Mean or expected value of W

$$E[W] = \sum_{i=1}^{k} E[Y_i]$$

Variance of W under the assumption of independence of the Y_i

$$V[W] = \sum_{i=1}^{k} V[Y_i]$$

Variance of W under the assumption of perfect positive correlation of the Y_i

$$V[W] = \left(\sum_{i=1}^{k} \sqrt{V[Y_i]} \right)^2$$

Also, under the assumption of perfect positive correlation, the fractiles are additive. That is,

$$F_W 100\alpha = \sum_{i=1}^{k} F_i 100\alpha$$

Rough rather than rigorous mathematical definitions of independence and perfect positive correlation are the following:

- Two random variables are *independent* if they are not related in that knowing the value of one variable does not help in predicting the value of the other variable.

- Two random variables are *perfect positively correlated* if they are positively related in that a large value of one variable is associated with a large value of the other variable. Also, a small value of one variable is associated with a small value of the other variable.

The normal probability distribution is a good approximate distribution for this type of random variable *W* because of the well-known Central Limit Theorem of probability theory.

Example

Suppose a record of the occurrence of landslides for the past year (t^*) showed that 40 landslides (n) occurred.

An estimate of λ would be

$r = n/t^* = 40/1 = 40$

We expect landslides at a rate of 40 per year.

An estimate of μ would be

$m = 1/r = t^*/n = 1/40 = 0.025$

We expect the mean recurrence interval to be 0.025 years.

Suppose that from the 40 observed landslides, the sample mean cost $m_X = 0.5$ million dollars and the sample standard deviation $s_X = 0.1$ million dollars.

An estimate of the mean or expected value of $Y(t)$ during the specified time $t = 5$ years, that is, $E[Y(5)]$, is

$m_Y = rt(m_X) = (40)(5)(0.5) = 100$

Hence, a mean estimate of the total costs from landslides during the next 5 years is 100 million dollars.

An estimate of the variance of $Y(5)$, $V[Y(5)]$, is

$$s_Y^2 = rt(m_X^2 + s_X^2) = (40)(5)[(0.5)^2 + (0.1)^2] = 52$$

An estimate of the standard deviation of $Y(5)$ is

$$s_Y = 7.21$$

Estimates of the characterizing parameters of the lognormal distribution, namely mu (μ^*) and sigma (σ^*), are

$$m^* = \ln\left(\frac{m_Y^2}{\sqrt{m_Y^2 + s_Y^2}}\right) = \ln\left(\frac{(100)^2}{\sqrt{(100)^2 + (7.21)^2}}\right) = 4.60$$

$$s^* = \sqrt{\ln(s_Y^2 / m_Y^2 + 1)} = \sqrt{\ln[(7.21)^2 / (100)^2 + 1]} = 0.072$$

Estimates of the 95[th] fractile and the 5[th] fractile of $Y(5)$, namely $F95$ and $F5$, are

$$f95 = e^{m^* - 1.645 s^*} = e^{4.60 - 1.645(0.072)} = 88.37$$

$$f5 = e^{m^* + 1.645 s^*} = e^{4.60 + 1.645(0.072)} = 111.99$$

Therefore, a low estimate of the total costs from landslides during the next 5 years is 88 million dollars. There is a 95% chance of exceeding 88 million dollars. A high estimate of the total costs from landslides during the next 5 years is 112 million dollars. There is a 5% chance of exceeding 112 million dollars.

Application

The direct costs assessed to landslides for each county in the 10-county San Francisco Bay region, California, listed in Godt (1999) will be used to illustrate the probabilistic methodology developed above (see table 4). The damaging landslides were caused by El Nino rainstorms within the winter season 1997-98. An economic landslide hazard assessment of each county is performed as in the case of the previous example. Then the ten counties are aggregated for comparison under the two assumptions: independence and perfect positive correlation.

Table 4. Poisson model for total damage from landslides in San Francisco Bay region, California, during a specified time based on damaging landslides caused by El Nino rainstorms within the winter season 1997-98. (Spreadsheet has two panels.)

Probability Model for Total Damage from Landslides in San Francisco Bay Region								R.A. Crovelli			Panel 1 of 2	
	Past	No. of	Rate of	Mean	Reported	Cost of Landslide		Cost of Landslide		Specified	No. of Landslides	
	Time	Slides	Occur.	Re. In.	Costs	Minimum	Maximum	Mean	S. D.	Time	Mean	S. D.
County	(years)		(no./yr)	(years)	(millions of $)			(millions of $)		(years)		
Alameda	1	87	87	0.0115	20.02	0.001	3.5	0.23011	0.58317	5	435	20.8567
Contra Costa	1	119	119	0.0084	27	0.0018	3.5	0.22689	0.58303	5	595	24.3926
Marin	1	26	26	0.0385	2.54	0.004	0.75	0.09769	0.12433	5	130	11.4018
Napa	1	16	16	0.0625	1.12	0.015	0.1	0.07	0.01417	5	80	8.94427
San Mateo	1	41	41	0.0244	55	0.001	6	1.34146	0.99983	5	205	14.3178
Santa Clara	1	10	10	0.1	7.6	0.002	4.1	0.76	0.683	5	50	7.07107
San Francisco	1	5	5	0.2	4.1	0.05	2	0.82	0.325	5	25	5
Solano	1	6	6	0.1667	5	0.004	5	0.83333	0.83267	5	30	5.47723
Sonoma	1	7	7	0.1429	21	0.025	19.8	3	3.29583	5	35	5.91608
Santa Cruz	1	165	165	0.0061	14.68	0.00008	1.856	0.08897	0.30932	5	825	28.7228
Aggre. (p.p.c.)		482			158.06							
Aggre. (indep)		482			158.06							

Probability Model for Total Damage from Landslides in San Francisco Bay Region									Panel 2 of 2
	Estimates of Total Damage as Economic Loss During a Specified Time								
	Total Costs		Lognormal		Confidence	Total Costs		Specified	Chance
	Mean	S. D.	Mu	Sigma	Level	Low	High	Costs (SC)	Exceed SC
County	(millions of $)				(%)	(millions of $)		(10^6 $)	(%)
Alameda	100.1	13.07558	4.59771	0.130073	90	80.13872	122.9357	110	21.473582
Contra Costa	135	15.26065	4.898926	0.112683	90	111.4503	161.4626	150	16.075563
Marin	12.7	1.802868	2.531626	0.141251	90	9.967099	15.86258	17	1.6375497
Napa	5.6	0.638792	1.716303	0.113702	90	4.614854	6.708162	7	2.172339
San Mateo	275	23.95483	5.612991	0.086944	90	237.4558	316.0819	300	14.81854
Santa Clara	38	7.225265	3.619829	0.188453	90	27.38106	50.89711	55	1.9879655
San Francisco	20.5	4.410286	2.997803	0.212707	90	14.12478	28.43653	27	8.0584986
Solano	25	6.452391	3.186632	0.253946	90	15.94153	36.75724	35	7.3257959
Sonoma	105	26.36642	4.623387	0.24728	90	67.80577	152.9523	160	3.3847833
Santa Cruz	73.4	9.244752	4.288054	0.125455	90	59.24606	89.5153	85	10.892073
	790.3	108.4318			90	628.126	981.6094	875	21.736121
	790.3	43.30465			90	719.0702	861.5298	875	2.5237802

It is very important to realize that the winter season 1997-98 was an anomalous year and not representative of conditions in a typical year because the occurrence and costs of the landslides were considerably higher than normal. In actual practice a longer period of record of landslides covering multiple years and storms should be used to determine estimates of future landslide occurrence and costs. The main purpose of this application is to illustrate what could be done with the economic probability model that has been developed by using available data required by the model; the future estimates themselves are not meaningful. On the other hand, since *scenario planning* is becoming wide spread by various planners, it could be argued that this application might be used to represent a "worst-case scenario," and the future estimates themselves would be meaningful.

Summary

- The Poisson model is the most commonly used model for the occurrence of random point-events in time.
- The Costa-Baker model is an example of the general binomial model.
- The binomial model has the restriction $\mu \geq 1$.
- The binomial model is an approximation to the Poisson model.
- Estimation theory is developed for the estimation of the parameters of both of the models.
- The exceedance probability of one or more landslides during a specified time is formulated for both models.
- An economic probability model is developed for the estimation of the total damage from future landslides in terms of economic loss as costs in dollars.
- A summary of probability models for estimation of number and costs of landslides is given in table 5.

Table 5. Summary of probability models for estimation of number and costs of landslides

Poisson Model for Number of Landslides (continuous-time model)

Random variables:	$N(t)$: Number of landslides	T: Recurrence interval
Probability distributions:	Poisson	Exponential
Parameters:	λ: Rate of occurrence	μ: Mean recurrence interval
Means or expected values:	$E[N(t)] = \lambda t$	$E[T] = \mu = 1/\lambda$
Standard deviations:	$S[N(t)] = (\lambda t)^{1/2}$	$S[T] = \mu$
Exceedance probabilities:	$P\{N(t) \geq 1\} = 1 - e^{-\lambda t}$	$P\{T > t\} = e^{-t/\mu}$
Estimators of parameters:	$R = N(t^*)/t^*$	$M = 1/R = t^*/N(t^*)$

Binomial Model for Number of Landslides (discrete-time model)

Random variables:	$N(t)$: Number of landslides	T: Recurrence interval
Probability distributions:	Binomial	Geometric
Parameters:	p: Probability of landslide	μ: Mean recurrence interval
Means or expected values:	$E[N(t)] = pt$	$E[T] = \mu = 1/p$
Standard deviations:	$S[N(t)] = [p(1-p)t]^{1/2}$	$S[T] = \mu(\mu - 1)$
Exceedance probabilities:	$P\{N(t) \geq 1\} = 1 - (1-p)^t$	$P\{T > t\} = (1 - 1/\mu)^t$
Estimators of parameters:	$F = N(t^*)/t^*$	$M = 1/F = t^*/N(t^*)$

Probability Model for Costs of Landslides

Random variables:	X: Cost of landslide	$Y(t)$: Total costs

$$Y(t) = \sum_{i=1}^{N(t)} X_i$$

Probability distributions:	Pareto	Lognormal
Means or expected values:	$E[X]$	$E[Y(t)] = E[N]E[X]$
Standard deviations:	$S[X]$	$S[Y(t)] = \{E[N](S[X])^2 + (E[X])^2(S[N])^2\}^{1/2}$

Random variable: W: Aggregation of total costs

$$W = \sum_{i=1}^{k} Y_i$$

Probability distribution: Normal

Mean or expected value:

$$E[W] = \sum_{i=1}^{k} E[Y_i]$$

Standard deviation:

$$S[W] = \left(\sum_{i=1}^{k} (S[Y_i])^2 \right)^{1/2} \textit{ under independence,}$$

$$S[W] = \sum_{i=1}^{k} S[Y_i] \quad \textit{under perfect positive correlation}$$

Conclusions

- The Poisson model is preferred over the binomial model (Costa-Baker model) because the Poisson model is a first-approximation model, and the binomial model is here an approximation of an approximation.

- The Poisson model has many useful properties and results that are mathematically tractable.

- The theory herein is applicable to many other hazard processes besides landslides, for example, earthquakes, floods, tsunamis, volcanoes, and storms.

- Not only are the probability models applicable for many types of natural hazards, but also for many types of damage in addition to economic loss, for example, human loss in deaths.

Acknowledgments

The author wishes to gratefully acknowledge the helpful reviews of Jeffrey A. Coe and Rex L. Baum, both in the Landslide Group of the Geologic Hazards Team at the U.S. Geological Survey.

References

Coe, J.A., Michael, J.A., Crovelli, R.A., and Savage, W.Z., 2000, Preliminary map showing landslide densities, mean recurrence intervals, and exceedance probabilities as determined from historic records, Seattle, Washington: U.S. Geological Survey Open-File Report 00-303, 25 p.

Costa, J.E. and Baker, V.R., 1981, Surficial geology – building with the earth: New York, Wiley & Sons, 498 p.

Crovelli, R.A., 1992, Probabilistic methodology for estimation of undiscovered petroleum resources in play analysis of the United States: Nonrenewable Resources, v. 1, no. 2, p. 153-162.

Godt, J.W., 1999, Maps showing locations of damaging landslides caused by El Nino rainstorms, winter season 1997-98, San Francisco Bay region, California: U.S. Geological Survey pamphlet to accompany miscellaneous field studies maps MF-2325-A-J, 13 p.

Keaton, J.R., Anderson, L.R., and Mathewson, C.C., 1988, Assessing debris flow hazards on alluvial fans in Davis County, Utah, *in* Fragaszy, R.J., ed., Twenty-fourth Annual Symposium on Engineering Geology and Soils Engineering: Pullman, Washington, Publications and Printing, Washington State University, p. 89-108.

Lips, E.W. and Wieczorek, G.F., 1990, Recurrence of debris flows on an alluvial fan in central Utah, *in* French, R.H., ed., Hydraulics/Hydrology of Arid Lands (H^2AL); Proceedings of the International Symposium: American Society of Civil Engineers, p. 555-560.

Ross, S.M., 1972, Introduction to probability models: New York, Academic Press, Inc., 273 p.

Walpole, R.E. and Myers, R.H., 1989, Probability and statistics for engineers and scientists: New York, Macmillan Publishing Company, 4th ed., 765 p.