APPENDIX

DISTRIBUTION THEORY OF SPILL INCIDENCE

1. The derivation of the predicted probability distribution

This appendix describes rigorously the derivation of the predicted probability distribution on spill occurrence given as equation 4, in subsection "Predicted Probability Distributions for a Fixed Class of Spills." The development is a Bayesian one; a good general description of these Bayesian inference techniques may be found in Box and Tiao, (1973, p. 1-73). The application of these methods to oilspill occurrence forecasting was proposed and described in Devanney and Stewart, (1974).

We will use the following terminology:

n = number of future spills,
t = future exposure,
λ = true rate of spill occurrence per unit exposure,
ν = number of spills observed in past,
τ = past exposure,
$f(n)$ = a marginal probability density on n, and
$f(n|y)$ = the conditional probability density of n given that the random variable $\gamma = y$.

Assume that spills occur at random with some intensity, $p(n)$:

$$P[n \text{ spills over exposure } t] = \frac{(\lambda t)^n e^{-\lambda t}}{n!} \quad \text{(A-1)}$$

Suppose that, in the absence of information about λ, we choose to represent our uncertainty about this parameter in the form of an "improper" prior density on λ:

$$f(\lambda|\text{ no data}) = 1/\lambda \quad \text{(A-2)}$$

This says, in effect, that with no spills ever having been observed, we place a good deal of faith on λ being equal to 0, although we allow *a priori* the possibility that it may be any positive number. This may seem artificial (as is often the case with Bayesian ignorance priors), but note that in any case all it takes is one observation of a spill to refute the notion that $\lambda = 0$. Our previous feelings in the absence of any data will be overwhelmed by minimal experimental evidence.

Suppose we then observe ν spills in τ exposure and wish to update our estimate of λ. The Bayesian approach is to represent our new estimate by a posterior density on λ derived from our ignorance prior density on λ combined with experimental evidence. This is accomplished through use of Bayes theorem:

$$f(\lambda|\nu,\tau) = \frac{f(\nu|\lambda,\tau) f(\lambda|\text{no data})}{f(\nu,\tau)}$$

$$= \frac{f(\nu|\lambda,\tau) f(\lambda|\text{no data})}{\int_0^\infty f(\nu|\lambda,\tau) f(\lambda|\text{no data}) d\lambda}$$

$$= \frac{\left[\frac{(\lambda\tau)^\nu e^{-\lambda\tau}}{\nu!}\right] \frac{1}{\lambda}}{\int_0^\infty f(\nu|\lambda,\tau) f(\lambda|\text{no data}) d\lambda}$$

$$= \frac{\left[\frac{(\lambda\tau)^\nu e^{-\lambda\tau}}{\nu!}\right] \frac{1}{\lambda}}{\int_0^\infty \frac{(\lambda\tau)^r e^{-\lambda\tau}}{\nu!} \frac{1}{\lambda} d\lambda}$$

$$= \frac{(\lambda\tau)^\nu e^{-\lambda\tau} \frac{1}{\lambda}}{\int_0^\infty x^{\nu-1} e^{-x} dx}$$

$$= \frac{(\lambda\tau)^{\nu-1} e^{-\lambda\tau}\tau}{(\nu-1)!}$$

This is the density on λ in Devanney and Stewart (1974, p. 28) "through which our past spill experience enters the analysis." It is, in Bayesian terms, the posterior density on λ.

If we were to gather more evidence, this posterior would now become the prior, and the same reasoning would apply:

$$f(\lambda|\nu_2,\tau_2) = \frac{f(\nu_2|\lambda,\tau_2) \times f(\lambda|\nu_1,\tau_1)}{\int_0^\infty f(\nu_2|\lambda,\tau_2) f(\lambda|\nu_1,\tau_1) d\lambda} \quad \text{(A-4)}$$

$$= \frac{e^{-\lambda(\tau_1+\tau_2)}(\lambda(\tau_1+\tau_2))^{\nu_1+\nu_2-1}(\tau_1+\tau_2)}{(\nu_1+\nu_2-1)!}$$

Note that this is exactly the same density on λ we would have obtained by adding the two exposures, τ_1, and τ_2, and the two numbers of spills, ν_1 and ν_2, and treating it all as one piece of data.

Having done all this, if we desire the density of the phenomenon (oilspill occurrence) given our current uncertainty about λ, we take the average of the Poisson densities weighted according to the posterior on λ:

$$f(n|t,\nu,\tau) = \int_0^\infty f(n|\lambda,t) f(\lambda|\nu,\tau) d\lambda$$

$$= \int_0^\infty \frac{(\lambda t)^n e^{-\lambda t}}{n!} \frac{[\lambda(\tau_1+\tau_2)]^{\nu_1+\nu_2-1}(\tau_1+\tau_2)}{(\nu_1+\nu_2-1)!}$$

$$= \frac{(n+\nu-1)!\, t^n \tau^\nu}{n!(\nu-1)!(t+\tau)^{n+\nu}} \quad \text{(A-5)}$$

This is the negative binomial distribution given as equation 4, in the subsection "Predicted Probability Distributions for a Fixed Class of Spills."

2. Moment-generating functions

Results in the remainder of this appendix depend on the use of generating functions. Some standard results from probability theory will be reviewed.

If X is a discrete random variable with $P(X=n)=P_n$, the generating function of X (Feller, 1957, p. 249) is

$$\Phi_X(s)=\sum_{n=0}^{\infty} p_n s^n \qquad \text{(A-6)}$$

Moment generating functions for some common distributions used in this analysis are as follows:
Bernoulli random variable with probability p of "success":

$$\Phi_X(s)=1-p+ps \qquad \text{(A-7)}$$

Poisson random variable with mean λt:

$$\Phi_Y(s)=\exp(\lambda t(s-1))$$

Negative binomial random variable with mean $\nu t/\tau$ and variance $\nu t/\tau(1+t/\tau)$

$$\Phi_N(s)=\left(\frac{\tau}{t+\tau-ts}\right)^{\nu} \qquad \text{(A-8)}$$

If X_k is a sequence of random variables with $P(X_k=n)\,P_{kn}$, and X is a random variable such that $P(X=n)=p_n$, in order that $p_{kn}=p$ for any fixed n, it is necessary and sufficient that

$$\Phi_{X_k}(s)=\Phi_X(s) \qquad \text{(A-9)}$$

for all s in [0, 1] (Feller, 1957, p. 262).

If $Z=X+Y$, and X and Y are independent, then

$$\Phi_Z(s)=\Phi_X(s)\Phi_Y(s) \qquad \text{(A-10)}$$

(Feller, 1957, p. 250). If X_i, $i=1, 2, 3, \ldots$, are independent and identically distributed,

$Z=\sum_{i=1}^{N} X_i$, and N is independent of the X_i, then

$$\Phi_Z(s)=\Phi_N(\Phi_X(s)) \qquad \text{(A-11)}$$

(Feller, 1957, p. 268).

3. Convergence of the negative binomial to the Poisson

Let N be the number of spills in an exposure t, and assume (following the first part of this appendix) that N is a Poisson random variable with generating function

$$\Phi_N(s)=\exp(\lambda t(s-1)), \qquad \text{(A-12)}$$

and that the predicted number of spills N' is a negative binomial random variable with generating function

$$\Phi_{N'}(s)=\left(\frac{\tau}{t+\tau-ts}\right)^{\nu}, \qquad \text{(A-13)}$$

where ν is the number of spills observed in the past in the course of exposure τ. If the Poisson model holds, then the Law of Large Numbers guarantees that as $\tau\to\infty$ then $\nu/\tau\to\lambda$. Suppose we had adopted the negative binomial model. Then

$$\Phi_{N'}(s)=\left(\frac{\tau}{t+\tau-ts}\right)^{\nu}$$

$$=\left(\frac{\nu/\lambda}{t+\nu/\lambda-ts}\right)^{\nu} \qquad \text{(A-14)}$$

and, as τ grows larger,

$$\Phi_{N'}(s)=\left(\frac{1}{\lambda t/\nu+1-\frac{\lambda t}{\nu}s}\right)^{\nu}$$

$$=\left(1+\frac{\lambda t}{\nu}(1-s)\right)^{-\nu} \qquad \text{(A-15)}$$

which approaches

$$\Phi_{N'}(s)=\exp\left(\lambda t(s-1)\right) \qquad \text{(A-16)}$$

as τ (and hence ν) grows larger.

Thus, if the Poisson model is correct, the analyst will be led to the Poisson model as enough data accumulates even while formally adopting the negative binomial model. Spill incidence could be modeled quite simply and directly using the Poisson distribution with λ set equal to ν/τ. This convergence to the true model is an example of "Bayesian consistency." The advantage of the negative binomial model, as derived through the Bayesian methodology of this appendix, is that it incorporates the uncertainty about λ for a finite exposure τ, since ν/t will never equal λ exactly. The uncertainty is reflected in a broader distribution on spill incidence due to the larger variance of the negative binomial distribution—a wider range of spill incident totals has non-neglible probability. The variance of N' is $\frac{\nu t}{\tau}\left(1+\frac{t}{\tau}\right)$, the variance of N is $\frac{\nu t}{\tau}$, and the difference is $\nu(\frac{t}{\tau})^2$. Thus the increase in uncertainty (as measured by the difference in variances) is proportional to the squared ratio of estimated future exposure, t, to observed past exposure, τ.

This is only one measure of the closeness of the two models, of course. Of more interest is a direct comparison of the summary features presented in the Oilspill Risk Analysis Model of the U.S. Geological Survey, particularly in calculating the probability that no spills will occur. The expectations of N and N' are the same under the two models, $\nu t/\tau$. Let

$$P_p = P[0 \text{ spills}|\text{Poisson model}] = e^{-\nu t/\tau} \quad \text{(A-17)}$$

$$P_{nb} = P[0 \text{ spills}|\text{Negative binomial model}) =$$

$$\left(\frac{1}{1+t/\tau}\right)^{\nu}. \quad \text{(A-18)}$$

Consequently, dividing the two equations and applying a Taylor's expansion yields

$$\ln\frac{P_{nb}}{P_p} = -\left(-\frac{\nu t}{\tau}\right) + \left(-\frac{\nu t}{\tau} + {}^1/_2\,\nu\Theta^2\right)$$

$$= +{}^1/_2\nu\Theta^2, \quad \text{(A-19)}$$

where

$$0 \leqslant \Theta \leqslant (t/\tau)^2 \quad \text{(A-20)}$$

or

$$1 \leqslant \frac{P_{nb}}{P_p} \leqslant e^{1/2\nu(t/\tau)^2}. \quad \text{(A-21)}$$

Thus the difference in the probability of no spills occurring under two models (and hence the difference in the probabilities of one or more spills) is again directly related to the size of $(t/\tau)^2$.

4. Distribution of the total number of spills from multiple sources

Let N_1 and N_2 be negative binomial distributed random variables with generating functions

$$\Phi_{N_1}(s) = \left(\frac{\tau_1}{t_1+\tau_1-t_1 s}\right)^{\nu_1}, \quad \text{(A-22)}$$

$$\Phi_{N_2}(s) = \left(\frac{\tau_2}{t_2+\tau_2-t_2 s}\right)^{\nu_2}. \quad \text{(A-23)}$$

Then, if $N = N_1 + N_2$,

$$\Phi_N(s) = \Phi_{N_1}(s)\Phi_{N_2}(s), \quad \text{(A-24)}$$

following equation A-10. In general, this will not be a simple distribution. However, if $t_1 = t_2 = t$ and $\tau_1 = \tau_2 = \tau$, then

$$\Phi_N(s) = \left(\frac{\tau}{t+\tau-ts}\right)^{\nu_1+\nu_2}, \quad \text{(A-25)}$$

so N is distributed as a negative binomial random variable with mean

$$\lambda = (\nu_1+\nu_2)\frac{t}{\tau} \quad \text{(A-26)}$$

and variance

$$\sigma^2 = (\nu_1+\nu_2)\frac{t}{\tau}\left(1+\frac{t}{\tau}\right). \quad \text{(A-27)}$$

5. Distribution of the number of hits

Let N, the total number of spills, be distributed as above, that is, negative binomial with parameters ν, t, τ. For each spill that occurs, associate a random variable X which takes the value 1 if a specified event occurs (such as the spill hitting land and 0 otherwise. Let X be a Bernoulli random variable.

$$P(X=1) = P \quad \text{(A-28)}$$

Let T be the total number of events that occur when spills originate from a single source,

$$T = \sum_{i=1}^{N} X_i \quad \text{(A-29)}$$

From section 2 of this appendix, the generating function of T is

$$\Phi_T(s) = \Phi_N(\Phi_X(s)) \quad \text{(A-30)}$$

From equations A-6 and A-8

$$\Phi_T(s) = \left(\frac{\tau}{t+\tau-t(1-p+ps)}\right)^{\nu}$$

$$= \left(\frac{\tau}{pt+\tau-pts}\right)^{\nu} \quad \text{(A-31)}$$

Thus, the distribution of T, the number of events, is in turn negative binomial, but with parameters ν, pt, and τ.

USGS ☆U.S. GOVERNMENT PRINTING OFFICE: 1982—576-034/116