ROYAL NORWEGIAN COUNCIL FOR SCIENTIFIC AND INDUSTRIAL RESEARCH

PROCEEDINGS FROM THE SEMINAR ON

SEISMOLOGY AND SEISMIC ARRAYS

OSLO, 22–25 NOVEMBER 1971

Editors: E S Husebye and H Bungum

Arranged in connection with the opening of The Norwegian Seismic Array (NORSAR) 1972

SEISMIC EVENT DETECTION AND DISCRIMINATION - SOME STATISTICAL CONSIDERATIONS

R.T. LACOSS

Lincoln Laboratory
Massachusetts Institute of Technology
Lexington, Massachusetts

INTRODUCTION

This paper consists of a discussion of several related statistical ideas in the areas of seismic detection and discrimination. The intention is to demonstrate that detection and discrimination problems can be formulated in a relatively precise statistical language and that this can be used to avoid making sloppy assertions based upon data and to make it possible to compare and evaluate different systems or uses of seismic data.

The items covered are:

- A maximum likelihoood method for estimating seismicity and the incremental detection probability of any system used for detection.
- The distinction between incremental and cumulative detection thresholds.
- Indirect estimation of surface and body wave detection performance based upon background noise statistics.
- Conversion of surface wave detection thresholds to body wave thresholds and vice versa.

ESTIMATING SEISMICITY AND DETECTION CAPABILITY SIMULTANEOUSLY

Suppose that a seismic station or a network of stations is operated under a fixed set of rules for the detection of earthquakes for a period of time of duration T. Suppose also that only events which prove to be earthquakes from a preassigned region are retained and their body-wave magnitudes determined. The data from such an experiment is an integer, K, the total number of earthquakes recorded, and a list of K magnitudes, $m_1,....,m_K$, usually presented in the form of a cumulative histogram, showing the log of the number of recorded events having magnitude at least m, versus m. Using the available data, it should be possible to determine a) the seismicity of the region in question, or b) the detection capability of the station or network used, or c) both. In fact, one must almost always determine both seismicity and detection performance, although the chief objective may be to find only one of these quantities, since the other is rarely well enough known in advance to be modeled accurately. In order to reduce this problem to the estimation of specific parameters from the data, so that standard statistical estimation techniques may be applied, it is necessary to make specific assumptions about the random character of seismicity and the detection of weak events in background noise by the seismic network. For mathematical convenience, the well-known and tractable laws of Poisson and Gauss have been used here, and the results are only as valid as the underlying assumptions, although the method is quite general.

Suppose that earthquakes occur as a Poisson process, with rate depending upon magnitude and other parameters. Since in the type of experiment of interest here one is concerned only with total numbers of events which occur over a fixed period of time, we need make no claim that the Poisson process accurately represents the occurrence of events as a stochastic process, but only that the number of events recorded having magnitudes (and other parameters) in given ranges are Poisson variables. Let $\overline{N}_m$ be the mean number of events which occur during the period T having magnitude at least m. In addition to the Poisson assumption further we assume a linear relationship,

$$\log \overline{N}_m = a - bm,$$

between $\log \overline{N}_m$ (natural logarithm) and magnitude. If base-ten logarithms are used, a and b become $a' = a/\log 10$ and $b' = b/\log 10$.

Finally, the probability $\Pi(m)$, that the seismic system will detect an event having magnitude m is assumed to be an error function:

$$\Pi(m) = (2\pi\sigma^2)^{-\frac{1}{2}}\int_{-\infty}^{m} \exp - \frac{(m'-\mu)^2}{2\sigma^2} dm' = \Pi_o(m-\mu)$$

In terms of these assumptions, the general problem is the estimation of the four parameters a, b, μ, and σ from the K+1 data values $K, m_1, \ldots, m_K$. The quantities a and b define the seismicity of the region μ and σ the detection capabilities of the system.

Given the assumptions one can obtain the probability distribution of $K, m_1, \ldots m_K$ given a, b, μ, and σ. Taking the log of this gives the log likelihood function

$$\Lambda(K, m_1, \ldots, m_K; a, b, \mu, \sigma) =$$

$$-\bar{N} + K(a - b\langle m\rangle + \log b) + \sum_{i=1}^{K} \log\Pi(m_i)$$

where $\langle m\rangle$ is the arithmetic average of the observed magnitudes and $\bar{N}$ is the expected number of recorded events, which can be calculated. The maximum likelihood estimates of parameters are the values which maximize this function.

The likelihood function can be maximized in two steps. Setting partials of Λ with respect to a and b equal to zero we obtain

$$\hat{a} = \log K + \hat{b}\mu + \log\hat{b} + \log\int_{-\infty}^{\infty} \eta(o,\sigma) e^{-bm} dm$$

and

$$1/\hat{b} = \tfrac{1}{2}(\langle m\rangle - \mu)\left[1 + \sqrt{1 + \frac{4\sigma^2}{(\langle m\rangle - \mu)^2}}\right]$$

where $\eta(o,\sigma^2)$ is a Gaussian p.d.f with variance σ^2 and zero mean.

Substituting back into Λ gives

$$\Lambda(K,m_1,\ldots,m_K;\hat{a},\hat{b},\mu,\sigma) =$$
$$K\left[\log K-2+\tfrac{1}{2}(\hat{b}\sigma)^2+\log\hat{b}\right]+\sum_{i=1}^{K}\log\Pi_o(m_i-\mu,\sigma).$$

This is a function of μ and σ and can be easily maximized. We have chosen simply to contour the function on the μ,σ plane and find the peak. Of course once μ and σ are found a and b are also known. It would have been possible to use a computer to do a four parameter maximization, but we believe the two step process gave a little more insight into the situation.

Given that an experiment produced a total of K recorded events, then the probability that N_m, the number having magnitude at least m, is equal to k is given by the binomial distribution with success probability given by $p = \bar{N}_m/\bar{N}$. The value of $\bar{N}_m$ is

$$\bar{N}_m = e^{a-bm}\operatorname{erf}\left(\frac{m-\mu}{\sigma}\right)$$
$$+e^{a-b\mu+\frac{b^2\sigma^2}{2}}\left[1-\operatorname{erf}\left(\frac{m-\mu+b\sigma^2}{\sigma}\right)\right]$$

and the expected total number of detected events is

$$\bar{N} = \bar{N}_{-\infty} = e^{a-b\mu+\frac{b^2\sigma^2}{2}}$$

These equations, and the binomial distribution, can be used to compare experimental cumulative distributions with $\bar{N}_m$ and to indicate confidence intervals on the data.

As a demonstration of the maximum likelihood method we have used the first 2000 events of 1968 reported by the National Ocean Survey (NOS) to estimate μ, σ, a' and b'. The primed (base 10)

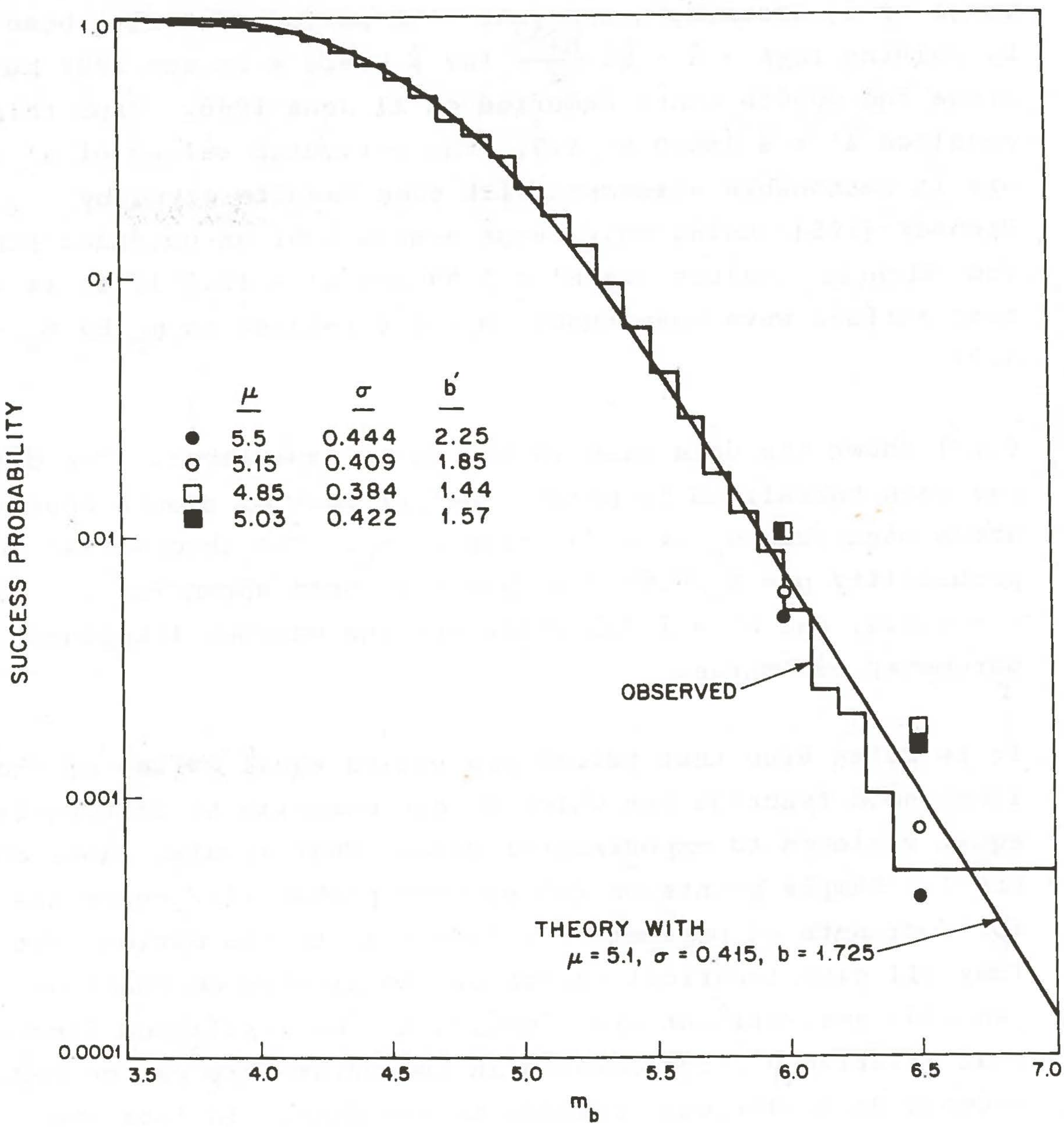

Fig 1. Observed (2000 U.S.C. and G.S. events) and theoretical cumulative histograms.

the solid line on Fig 2. The experimental data are also shown in the figure. By using tables of the cumulative binomial distribution for 90 trials and various success probabilities integers were determined which most nearly satisfied the equations

$$\sum_{k=0}^{k_5} \text{Prob } \{N_m=k/N=K\} = 5\%$$

seismicity parameters were estimated to be consistent with standard seismological usage. The likelihood function in this example is maximized by $\hat{\mu} = 5.1$, $\hat{\sigma} = 0.415$, and $\hat{b}' = 1.725$. The corresponding value of $\hat{a}$, assuming a one-year time period, has also been computed by solving $\log K = \hat{a} - \hat{b}\hat{\mu} + \frac{\hat{b}^2\hat{\sigma}^2}{2}$ for $\hat{a}$ where K is not 2000 but 4500 since the 2000th event occurred on 11 June 1968. From this we obtained $\hat{a}' = \hat{a}\log 10 = 11.9$. The estimated values of a' and b' are in reasonable agreement with some results given by Richter (1954) using only large events over an extended time period. The Richter values are b' = 1.59 and a' = 12.2 if it is assumed that surface wave magnitudes, M_s, are related to m_b by $M_s = 1.59m_b - 3.97$.

Fig 1 shows the data used in the above experiment. The data has been normalized to obtain the per cent of events observed above magnitude m_b as a function of m_b. The theoretical success probability $p = \bar{N}_m/\bar{N}, \bar{N}=2000$, has also been shown for $\mu = 5.1$, $\sigma = 0.415$, and b' = 1.725 which are the maximum likelihood parameter estimates.

It is often true that parameters giving equal values of the likelihood function but which do not maximize it do roughly equal violence to experimental data. That is also shown on Fig 1. Sample points on the success probability curve are shown for four sets of parameters in addition to the optimum set. They all give identical values of the likelihood function. One can also see, without even looking at the likelihood function, that relatively large changes in the parameters can be made without doing serious violence to the data. In fact the maximum of the likelihood function was a ridge structure which allowed large changes in parameters to give very small changes in likelihood. In a sense this lack of sensitivity can be interpreted as a lack of precision in estimating the parameters.

The stability of estimates is also related to the fluctuation in experimental success probabilities which might be obtained for a typical experiment. Such fluctuations have been investigated using a short run of N.O.S. data (90 events from two PDE cards). The short run of data was used to emphasize the fluctuations. The likelihood function showed a very broad minimum, but the parameter values b' = 1.65, $\mu = 4.9$, and $\sigma = 0.39$ were near the minimizing set. The curve of the success probability for these parameter values is plotted as

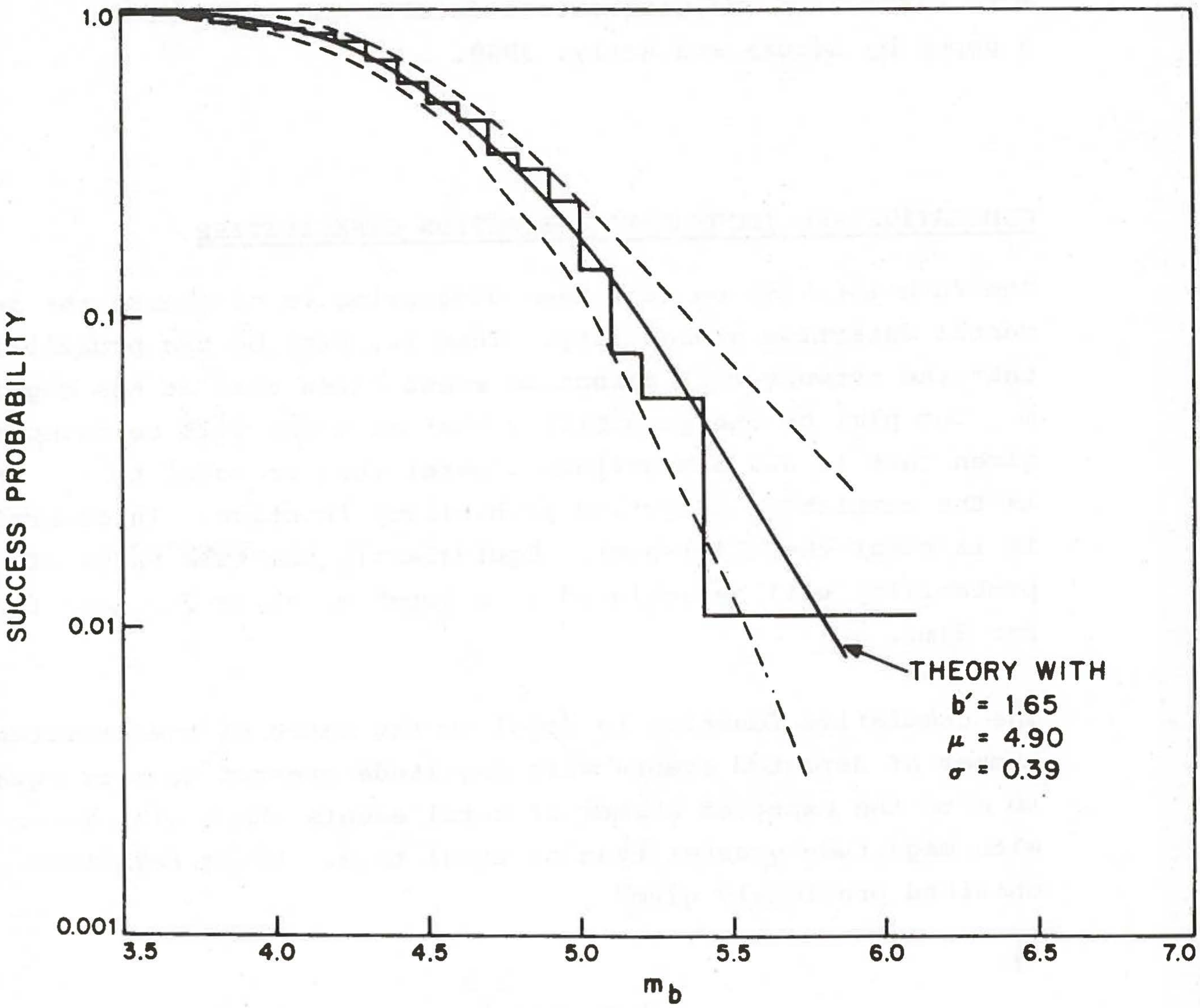

Fig 2. Observed (90 U.S.C. and G.S. events) and theoretical histograms and 90% confidence intervals.

and

$$\sum_{k=k_{95}}^{K} \mathrm{Prob}\{N_m=k/N=K\} = 5\%$$

In other words, k_5 and k_{95} determine the 5% tails. Normalizing k_5 and k_{95} as a fraction of K, we obtain the dashed curves in Fig 2. It appears that the model gives a reasonable picture both of the average seismicity and the fluctuations to be expected about the mean.

More discussion and statistical details can be found in a paper by Lacoss and Kelly, 1969.

CUMULATIVE AND INCREMENTAL DETECTION CAPABILITIES

The function $\Pi(m)$ we have been discussing is of course the incremental detection probability. That is, $\Pi(m)$ be the probability that the network will detect an event given that it has magnitude m. Let p(m) be the probability that an event will be detected given that is has a magnitude greater than or equal to m. This is the cumulative detection probability function. In general it is clear that $\Pi(m)<p(m)$. Equivalently the same value of probability will be achieved at a lower m value for p(m) than for $\Pi(m)$.

The cumulative function is equal to the ratio of the expected number of detected events with magnitude greater than or equal to m to the expected number of total events which will occur with magnitude greater than or equal to m. Using relations obtained previously gives

$$p(m) = \Pi(m)+e^{b(m-\mu)+\frac{b^2\sigma^2}{2}}\left[1-\Pi(m+b\sigma^2)\right].$$

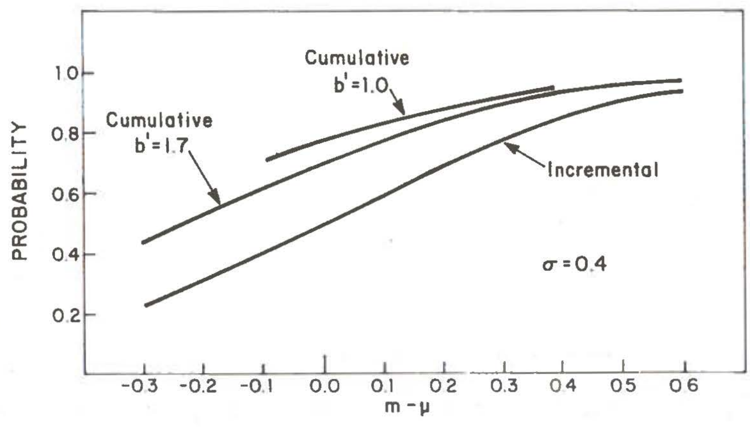

Fig 3. Cumulative and incremental detection probabilities.

Fig 3 shows the cumulative and incremental probabilities as a function of $m-\mu$ for some typical values of σ and b'. We note that $p(m)=\Pi(m+\Delta)$ when Δ is typically 0.2 to 0.3. Thus cumulative threshold for a system, at any success probability is 0.2 to 0.3 magnitude units less than the incremental threshold. This does depend upon the shape of

$\Pi(m)$, as determined by σ, and the seismicity parameter b'. In general one can see that p(m) must depend on the shape of the seismicity curve and in that way can be misleading. That is, the value of p(m) can be quite high for an m with low $\Pi(m)$ and dealing with p(m) would tend to ignore the fact that near m the system is not operating at a satisfactory success level.

DETECTION PERFORMANCE INFERRED FROM NOISE OBSERVATIONS

The discussion to this point has assumed a network or system operating and gathering event statistics which can be used to evaluate the system. The false alarm problem is essentially ignored. One can also make statements about capability, including something about false alarms, by considering background noise statistics. I will do this for a simple station which declares a detection each time the envelope of a filtered seismogram passes some threshold.

Suppose W is the bandwidth of the seismogram in Hz and that an independent observation of its envelope can be taken every 1/W seconds. Let p(c) denote the probability that a specific sample will exceed the value c with no real signal present. That is p(c) is a false alarm probability. Using this notation the expected number of false alarms per day ($\overline{\text{F.A.P.D.}}$) is 86400 Wp(c) if independent observations are made every 1/W seconds.

If the seismic noise is Gaussian zero mean and the bandwidth W is not too large, the distribution of the envelope is Rayleigh. Specifically if σ^2 is the noise power then

$$p(c) = e^{-c^2/2\sigma^2}$$

so

$$\overline{\text{F.A.P.D.}} = 86400 W e^{-c^2/2\sigma^2}$$

This can be used to relate false alarms, bandwidth and total noise power to a proposed detection threshold. In some cases it may be desired to obtain $\overline{\text{F.A.P.D.}}$ as a function of power spectral density. In that case, as long as noise power does not fluctuate too much, we can take $\sigma^2 = W\sigma_o^2$ where σ_o^2 is the power spectral density.

Some numbers may be of interest. Suppose W = 0.3, typical of some SP data and we want $\overline{\text{F.A.P.D.}} < (100), (10)$ then $p < (0.0038), (0.00038)$. This requires $c > 3.3\sigma$ and $c > 4.0\sigma$.

The above is fine for a surveillance situation and is typical of short period data handling. A slight modification will handle the typical LP situation where it is known that an event occurred and we wish to measure its surface waves (usually Rayleigh waves) but do not wish to assign a surface wave value which in fact is just background noise level.

Given an SP detection and location, suppose we can predict that the Rayleigh wave will arrive in some interval of T seconds. There are L = TW independent observations of the envelope in that interval and if any one exceeds our threshold we will declare a measurement. With no signal present the false alarm probability is the probability that at least one of the L observations will exceed the threshold being used. This false alarm probability is

$$P(c) = 1 - F^L(c)$$

where c is the threshold and F is the cumulative distribution of the noise envelope. Using the Rayleigh envelope distribution which was used previously we obtain

$$P(c) = 1 - \left(1 - e^{-\frac{c^2}{2\sigma^2}}\right)^L .$$

Since this must be small to avoid identifying an explosion as an earthquake, we can actually approximate it very well by

$$P(c) \simeq Le^{-c^2/2\sigma^2}.$$

Fig 4 shows the false alarm probability for values of L from 1 (window = 40 secs) to 9 (window = 300 secs). The normalized threshold is c/σ. The advantage of a short window can be clearly seen. For example if c/σ is set to 3, then the false alarm probability is 1.0% for a 40 second window and 10.0% for a 360 second window. In practice, unless more than one site is used for confirmation, a false alarm probability considerably less than 1.0% would probably be required. The use of multiple sites can also be considered in a statistical framework but is beyond the scope of this paper.

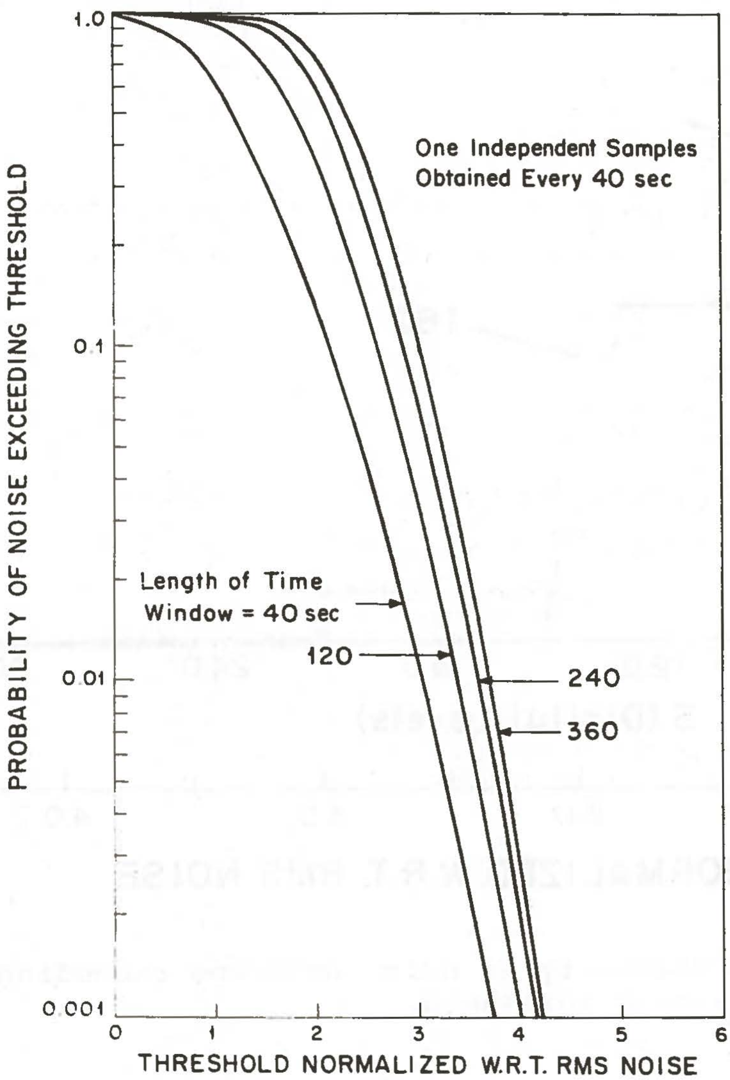

Fig 4. Theoretical probability of noise envelope exceeding a given level versus threshold.

Fig 5 shows experimental data obtained by using chirp filtered LASA beams of noise. The agreement with the theory curves is quite good. The use of a chirp filter is not important for the statistics but can be significant in practice since by compressing the signal, as well as increasing the signal-to-noise ratio, it can allow smaller time windows. A similar discussion of LP false alarms will be found in a paper by Capon et al, 1969.

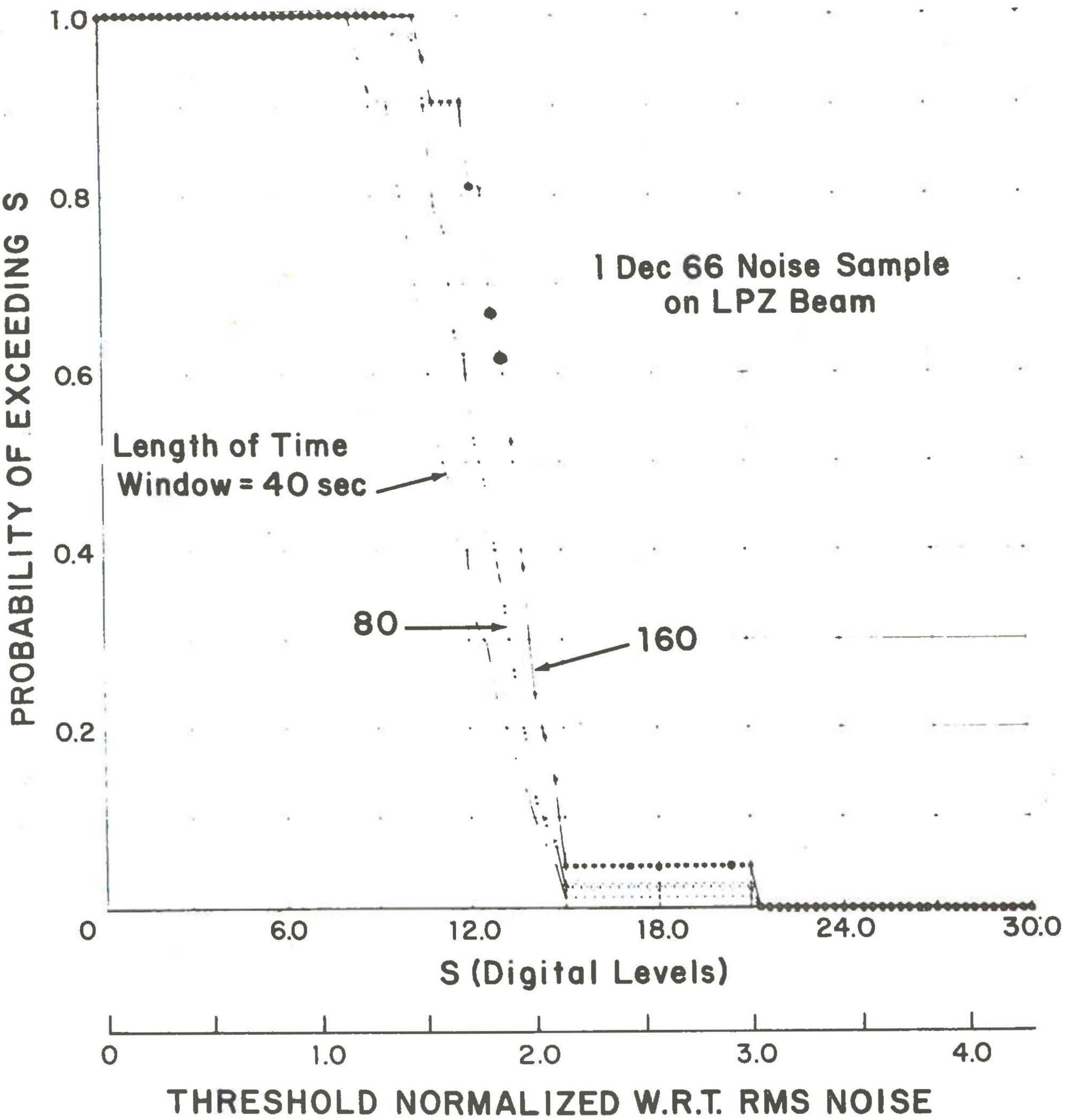

Fig 5. Experimental probability of noise envelope exceeding a given level versus threshold.

The preceding assumed σ^2 to be constant and is appropriate for a consideration of false alarms. However, suppose that a real system measures σ^2 and it changes slowly in time. The detection threshold can be set to some multiple of σ to maintain a constant false alarm probability. Since the threshold now moves a particular event may or may not be detected depending on the value of σ at the time of arrival. The detection probability can be determined only if the distribution of σ, now considered a random variable, is known.

Since seismic magnitude is a logarithmic function of signal amplitude, it may be convenient to consider the log of σ and the log of the envelope. Call this latter a. If the decision rule has been to accept an observation if the envelope was greater than $k\sigma$, then in terms of logs it is to do so if a $\geq \log \sigma + \log k$. The probability of detecting an event, given it has log amplitude a, is the probability that a - log k is greater than log σ. Assuming log σ to have a Gaussian distribution will give exactly the error function incremental detection probability introduced in the first Section.

Ideally the distribution of log σ, which will give detection probabilities, should be measured over a year's time since noise levels tend to have a yearly periodic component as well as daily and weekly fluctuations. In fact one could model the situation with an incremental detection probability which varies with the season rather than an average for the entire year. These fine points are of interest since the yearly average capability may be considerably different from the time local capability in certain seasons.

EXPRESSION OF M_s CAPABILITY IN TERMS OF m_b AND VICE VERSA

The last topic to be discussed involves expressing M_s detection capability in terms of m_b and vice versa. Only the first of these will be treated since the other is exactly the same in concept. Also only earthquakes will be discussed because the principles are the same for explosions.

The basic fact which cannot be ignored is that for a population of earthquakes the M_s and m_b values are not uniquely related. Thus one might imagine that

$$M_s = a\, m_b + b + r$$

where a,b are constants and r is a zero mean random variable which takes on different values for different events. Suppose that $\Pi(M)$ is the incremental detection probability for surface waves. It might be tempting to consider $\Pi(am_b+b)$ as a function of m_b, assuming a and b are known, as the probability of detecting a surface wave given that an event has body wave magnitude m_b. That is not correct, but it is not very difficult to obtain the correct expression.

Let $g(r)$ be the probability density function of r and assume $g(r) = g(-r)$. Then $g(M_s-am_b-b)\Pi(M_s)dM_s$ is the probability that the surface wave will be detected and the magnitude will be in the interval (M_s, M_s+dM_s) given m_b and the constants a,b. The integral of this is the desired probability and it can be recognized as a convolution using $g(r)=g(-r)$. The function $\Pi(M_s)$ is the convolution of $\dot{\Pi}(M_s)$ with a unit step function, u, so the function of interest looks like the impulse response of a linear system with three component filters with impulse responses g, $\dot{\Pi}$, and u. If Π is monotonic then $\dot{\Pi}$ is a probability density function and the function we desire can be thought of as the cumulative distribution function of the sum of two independent random variables with density function g and $\dot{\Pi}$.

Suppose that Π is the error function,

$$\Pi(m) = \frac{1}{\sigma_1\sqrt{2\pi}} \int_{-\infty}^{M-\mu} e^{-\frac{x^2}{2\sigma_1{}^2}} dx$$

and assume that r is Gaussian with variance $\sigma_2{}^2$. In this case the distribution of our imaginary random variable which is the sum of two independent Gaussian variables is also Gaussian with mean and variances given by the sums of those of the imaginary component random variables. Thus given m_b the probability that a surface wave will be measured is

$$P(m_b) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{am_b+b-\mu} e^{-\frac{z^2}{2\sigma^2}} dz$$

where

$$\sigma^2 = \sigma_1{}^2 + \sigma_2{}^2 .$$

Note that if m_b is such that $P(m_b) = 0.5$ then $P(m_b) = \Pi(am_b+b)$. However if $P(M_b) > 0.5$ then $P(m_b) < \Pi(am_b+b)$ because $\sigma > \sigma_1$. Similarly if $P(m_b) < 0.5$ then $P(m_b) > \Pi(am_b+b)$.

The point to remember is that if one is given a value of M_s and told that 90% of all events with M_s that size will be detected that cannot be directly expressed in terms of m_b by using $M_s = am_b+b$ and ignoring the scatter inherent in the M_s,m_b relationship. Also the distinction between cumulative and incremental detection levels must not be forgotten.

ACKNOWLEDGEMENTS

This work was sponsored by the Advanced Research Project Agency of the Department of Defense.

REFERENCES

1. J. Capon, R.J. Greenfield and R.T. Lacoss: Long-period signal processing results for the large aperture seismic array, Geophysics, Vol 34, No 3, June 1969.

2. C.F. Richter: Elementary Seismology, W.H. Freeman and Company, San Francisco, 1958.

3. R.T. Lacoss and E.J. Kelly: Estimation of Seismicity and Network Detection Capability, M.I.T. Lincoln Laboratory Technical Note 1969-41, 16 Sep 1969.