PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 15 (1995), pp. 115–125

ON A WEATHER MODIFICATION PROBLEM OF PROFESSOR NEYMAN*

BY

DAVID R. BRILLINGER (BERKELEY, CALIFORNIA)

Currently in the period of dynamic indeterminism in science, there is hardly a serious piece of research which, if treated realistically, does not involve operations on stochastic processes. The time has arrived for the theory of stochastic processes to become an item of usual equipment of every applied statistician.

J. Neyman [16]

Abstract. In a 1975 paper Professor J. Neyman posed a problem concerning the average travel time of the effects of cloud seeding. That problem is here discussed in the context of stochastic point processes. An analysis and solution is provided for the data that motivated the problem. There is a review of some of the available work concerning velocity estimation for travelling disturbances.

1. Introduction. Professor Neyman was concerned with problems of weather modification throughout much of his career (see Neyman [18]). One of his last papers was on the topic and appeared in this journal (Neyman [19]). In Neyman and Scott [20] and Neyman [17] an interesting problem is posed. This work discusses some aspects of that problem.

A weather modification experiment to reduce hail, Grossversuch III, was carried out at Ticino, Switzerland. Each day a decision was made whether conditions were suitable to define an experimental day. If suitable, randomly, seeding was or was not carried out the following day. Seeding, if any, lasted from 730 to 2130 hours. Rainfall measurements were made in Zürich, about 120 km away from Ticino. Fig. 1 provides graphs of average hourly rainfall totals smoothed by a running mean of three, for the experimental days when

* This research was supported by the National Science Foundation Grant DMS-9300002 and Office of Naval Research Grant N00014-94-1.

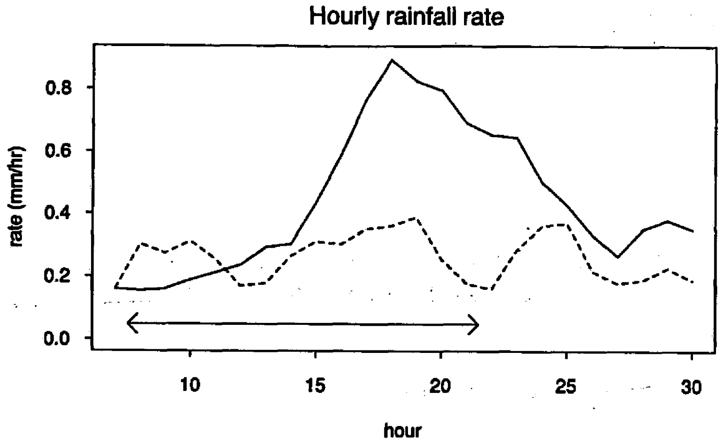

Fig. 1. Three-hour moving averages of hourly rainfall totals recorded at Zürich. The winds were from the south and there was a stability layer. The precipitation is in mm. The solid curve refers to seeded days, the dashed to unseeded. The seeding period is indicated by the arrow

a stability layer and southerly winds were present. The solid curve refers to days with seeding, and the dashed curve – to days without seeding. There were 53 experimental days with seeding and 38 without seeding. What Professor Neyman focussed on, in the Figure, was an apparent effect of seeding, starting at perhaps 1400 hours.

Neyman and Scott [20] write as follows:

> ... the curves ... represent averages of a number of independent realizations of certain stochastic processes. The 'seeded' curves are a sample from a population of one kind of processes and the 'not seeded' curve a sample from another. For an initial period of a number of hours ... the two kinds of processes coincide. Thereafter, at some unknown time T, the two processes may become different. Presumably, all the experimental days differ from each other, possibly depending on the direction and velocity of prevailing winds. Therefore, the time T must be considered as a random variable with some unknown distribution. The theoretical problem is to deduce the confidence interval for the expectation of T, ...

Dr. M. Schüepp (see Neyman [18]) speculates on the cause of the extra rainfall at Zürich as being Mediterranean moisture affected by seeding and brought north by southerly winds.

In Neyman [17] the problem is further described as:

> Now I come to the formulation of what I belive to be a novel theoretical statistical problem: *to produce a confidence interval (subject to some intelligible property of optimality) for the average time, say T, at which the effects of seeding over the target begin to be felt at a given locality for which the hourly precipitation amounts are available.*

The problem is addressed in this work by constructing an estimate of the distribution of travel times of individual effects moving from Ticino to Zürich.

The layout of the paper is the following: initially there is a review of some techniques that have been employed for velocity and delay estimation, next a description of the Grossversuch III Experiment, then some comments on the use of conceptual models, and finally an analysis and discussion. The principal analysis is by nonlinear least squares with the regression function given by expressions (6) and (7) below. The data employed in the analysis had to be read from Fig. 1 of [20].

2. Velocity estimation. The problem of velocity estimation, and the directly related one of delay estimation, has been much studied in science and technology. A broad variety of techniques have been proposed. The fact that the operation of motion is a translation is seized on in various of the methods. Carter [8] provides a collection of papers, mainly from an engineering viewpoint, addressing the topic.

In the simplest case one assumes constant velocity, inputs a pulse to the system of interest at a given time, and then measures the time at which the travelling pulse arrives. Naive forms of radar and sonar proceed in this fashion. In the case of an earthquake, the pulse may be viewed as the motion at the hypocenter of the earthquake. One estimates the origin time of the earthquake and then the arrival time at the observatory, and hence the travel time. Because of the presence of noise, difficulties often arise in measuring the actual arrival time (see Freedman [10]).

To handle the presence of noise, and the changing shape of the signal, researchers sometimes crosscorrelate the input and output and look for the lag at which the correlation is largest. The pulse may be a chirp, or a random sequence in some cases, e.g. in seismic exploration or in spread spectrum radar. Sometimes cross-spectral analysis is employed. An interesting application of that technique to the alignment of tree ring records is given in Foutz [9].

Turning to another field, a basic experiment of neuroscience involves applying a stimulus and recording a consequent response. A characteristic of the response is its latency, that is the time elapsing after stimulation application until a response is evident. Typically the stimulus has to be applied many times. Formal procedures of latency estimation are considered in Woody [27], Brillinger [4], Pham et al. [22].

Counts and point processes are other data types. The case of the times at which unlabelled vehicles of random velocities pass two measuring points may be studied via cross-spectral analysis (see Brillinger [3]). Lindley [13] develops estimates of the mean speed of particles from data consisting of total counts, at successive times, of the number of particles in a specified region. The idea is that if the particles are moving slowly, the counts will change slowly, whereas if the particles are moving quickly, the counts will change quickly. McDunnough [14] extends the problem to the case of counts available for several regions.

In an important class of circumstances, an array of sensors is employed and both velocity and direction are to be estimated. Briggs [2] was concerned with ionospheric movements and measurements made with an array of antennae. The spatial cross-correlation function was estimated at two times. The direction and velocity could then be estimated from the coordinates of the maximum cross-correlation. Arking et al. [1] estimated the cross-spectrum of two images in order to study speed and direction of cloud motion. Various statistical aspects of working with array data are developed in Cameron and Hannan [7], Thomson [24].

Brillinger [5] was concerned with estimating the joint distributions of several successive motions given locations of particles.

Difficulties that can arise in estimating velocity and delay include: the velocity may depend on time or frequency, a Doppler effect may be present due to relative motion, there may be signal-generated noise. Higher-order spectra can sometimes handle additive Gaussian noise effectively (see Nikias and Mendel [21]).

3. The data. The Grossversuch III Experiment ran from 1957 through 1963. It involved seeding thunder storm clouds in an attempt to reduce hail.

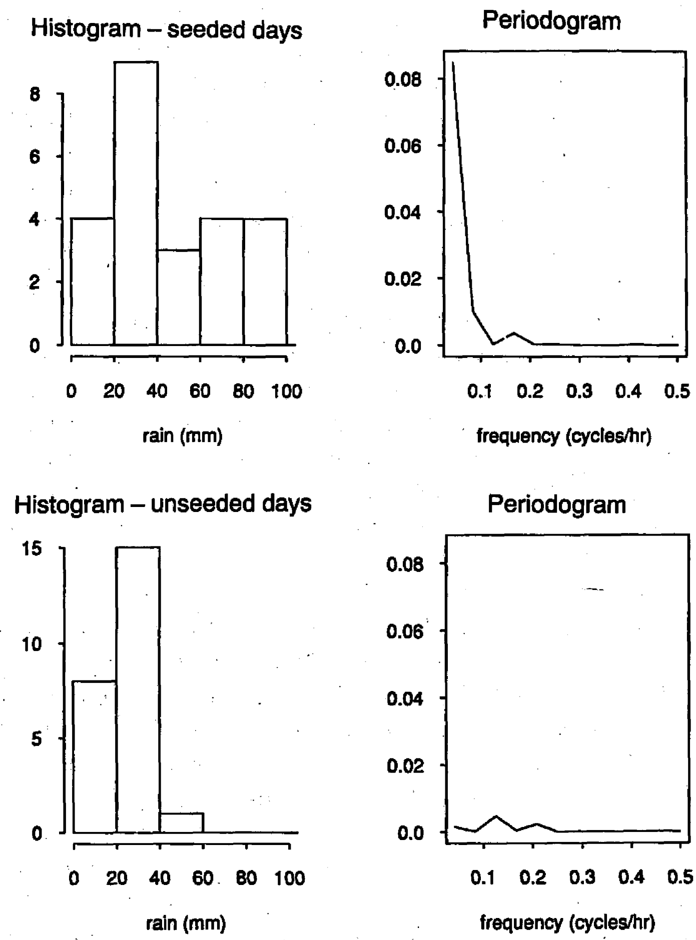

Fig. 2. Histograms and periodograms of the seeded and unseeded data graphed in Fig. 1

Seeding was by silver iodide dispersed from 20 hilltop-based generators. The idea was that silver iodide was a nucleating agent of ice crystals, which then may coalesce to form droplets. The burners were 5 minutes on, followed by 10 or 15 minutes off throughout a 14-hour period. There were 292 experimental days over the seven experimental years.

The measured data employed by Professor Neyman were Zürich hourly rainfall totals, $Y_l(t)$ with $t = 7, \ldots, 30$ indexing time in hours and with $l = 1, \ldots$ indexing experimental days with seeding. There were similar data for unseeded experimental days. These last, control data are basic for assessing the presence of a seeding effect.

Initial data analyses suggested considering separately the days with a stability layer (as revealed by a radiosonde at Milan) and southerly winds. There were 58 such experimental days that were seeded and 38 that were not. Fig. 1 graphs the three-hour running mean in both the seeded and the unseeded cases. The running mean and averaging operations were important because plots for a single day were highly variable. The data of the Figure are what were employed in the analysis of Section 5 below. Fig. 2 presents histograms and periodograms of the data. There are clear differences between the seeded and unseeded cases, as were apparent in Fig. 1. The periodograms provide an indication of the extent of serial correlation in the data. This is necessary to determine uncertainties of estimates. There is no suggestion of strong dependence in the unseeded case.

4. Conceptual models. Professor Neyman often emphasized the importance of constructing conceptual stochastic models to develop statistical analyses and to address scientific questions. Starting with LeCam [12], conceptual models, based on point processes and their smoothings, have proved useful in modelling rainfall (see, e.g., Waymire et al. [26] and Phelan [23]). A naïve model, pertinent to the problem at hand, is the following.

Consider particles at Ticino that move off with a possibility of leading to a rain particle at Zürich. The particles could be cells that become rain cells. Suppose that the particles are born at Ticino at the times σ_j of a point process M of rate $p_M(t)$. Suppose that the velocities with which the particles move to Zürich have a distribution and are independent of the process M. If the j-th particle has velocity v_j and the distance to be travelled is Δ, then its travel time is $u_j = \Delta/v_j$. Let N denote the point process of arrival times $\tau_j = \sigma_j + u_j$ of the particles at Zürich. (N is a Neyman–Scott process with clusters of size 1.) One can write, symbolically,

$$(1) \qquad \frac{d}{dt}M(t) = \sum_j \delta(t-\sigma_j), \quad \frac{d}{dt}N(t) = \sum_j \delta(t-\sigma_j-u_j),$$

where $\delta(\cdot)$ is the Dirac delta. If the travel times have density $f(\cdot)$, following from (1), the rate of the process N will be

$$p_N(t) = \int p_M(t-u)f(u)du. \tag{2}$$

Also, if M is a Poisson process, then so will be N (see Vere-Jones [25]).

In [12], [26], and [23], the rainfall intensity is a smoothing of a point process, say of M or N. Suppose that a time τ_j of N is associated with a mark R_j giving a corresponding amount of rain that falls. One can write

$$\frac{d}{dt}X(t) = \sum_j R_j\delta(t-\tau_j) \tag{3}$$

with $X(t) = \int_0^t dX(v)$ representing the cumulative amount of rain collected from time 0 to time t. If the R_j are independent of N and have mean μ_R, then, following (2), the intensity of rainfall at time t is

$$p_X(t) = \mu_R \int p_M(t-u)f(u)du. \tag{4}$$

The expected amount of rain from time 0 to t is given by

$$\mathrm{E}\{X(t)\} = \int_0^t p_X(v)dv.$$

These formulae will be employed to develop a regression function for analysis.

The running mean of order 3 of the hourly totals may be written in the form

$$\tfrac{1}{3}(X(t+1)-X(t-2))$$

and its expected value equals

$$\frac{1}{3}\int_{t-2}^{t+1} p_X(v)dv = \frac{1}{3}\mu_R \int_{t-2}^{t+1}\int p_M(v-u)f(u)dudv. \tag{5}$$

One wishes to estimate $f(\cdot)$ of (5) given information on the processes M and N.

5. Results. To proceed, the seeding rate $p_M(t)$ will be taken to be constant on the time interval from 730 to 2130 hours and to be 0 otherwise. It will be assumed that the velocities of travel of the particles are independent gamma variables with shape parameter s. Write the travel time U as θ/W with W gamma, having scale 1 and shape s. Let $F_\theta(\cdot)$ denote the distribution function of U. It will be assumed that the hourly rate of rainfall, unrelated to seeding, is α.

With $p_M(t) = C$ for $A < t < B$, where $A = 7.5$ and $B = 21.5$, one has the regression function

$$\mathrm{E}\{Y(t)\} = \alpha + \frac{C}{3}\mu_R\left[\int_{t-2-A}^{t+1-A} F_\theta(u)du - \int_{t-2-B}^{t+1-B} F_\theta(u)du\right] \tag{6}$$

in the case of seeding and $E\{Y(t)\} = \alpha$ in the case of no seeding. With the assumed velocity distribution, (6) may be evaluated in terms of $G_s(\cdot)$, the distribution function of the gamma, making use of the expression

$$\int_0^x F_1(v)dv = x\left[1-G_s\left(\frac{1}{x}\right)\right]-\frac{\Gamma(s-1)}{\Gamma(s)}\left[1-G_{s-1}\left(\frac{1}{x}\right)\right]. \tag{7}$$

Estimates of the unknowns $\mu = \theta/(s-1)$ (the average travel time), s, α, $\beta = C\mu_R$ are determined by ordinary least squares, weighting the seeded terms by 53 and the unseeded by 38, the respective numbers of cases averaged. Fig. 3

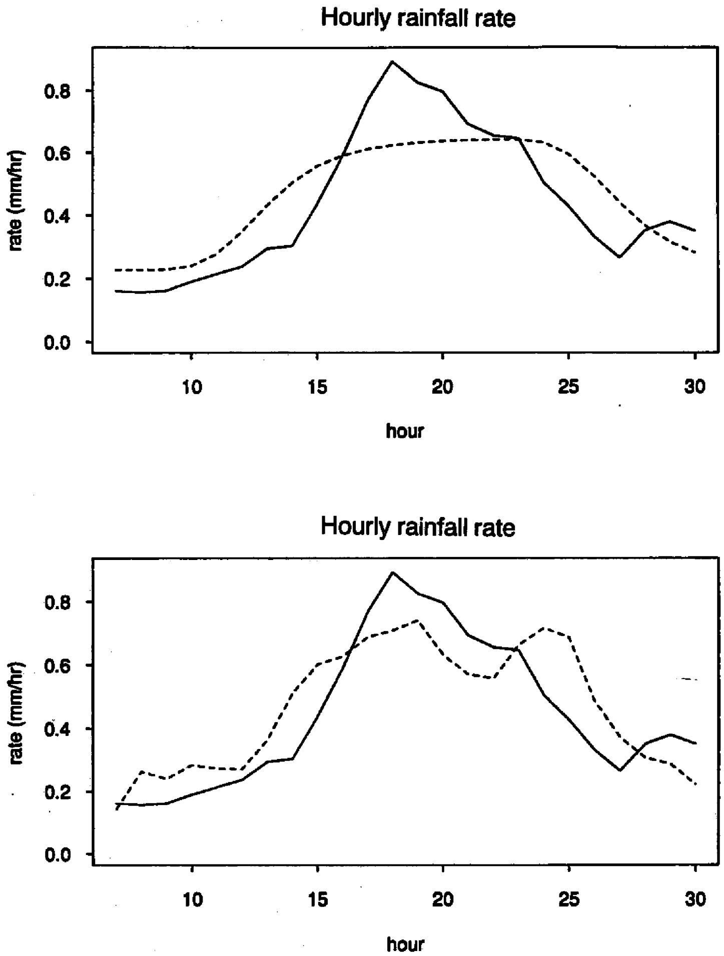

Fig. 3. The top panel provides the original data (solid line), and the fitted curve (dashed line) corresponding to velocities from a gamma distribution. The bottom panel is the fitted curve plus the adjusted unseeded fluctuations

presents the data (solid curve) and the fitted (dashed) curve. The estimates obtained are:

$$\hat{\mu} = 5.50\,\mathrm{hr}, \quad \hat{s} = 7.24, \quad \hat{\alpha} = 0.23, \quad \hat{\beta} = 5.85.$$

The estimate $\hat{\mu}$ is perhaps something like the value Professor Neyman was seeking. (The approximate standard error of $\hat{\mu}$, ignoring serial dependence, is 0.76 hr. The results of Hannan [11] can be used to obtain an approximate standard error in the case of stationary errors.) By taking the distance travelled to be 120 km, the estimated average velocity is 25.33 km/hr. The fit does not appear to be closely reflecting the sharpness of the peak at 1800 hours. This may reflect the inappropriateness of assuming $p_M(\cdot)$ to be constant throughout the seeding period or perhaps be a consequence of the inherent variability. To examine this last, the fluctuations of the unseeded days have been added to the fitted curve – see the lower panel of Fig. 3. The fit now appears plausible. Fig. 4 is the estimated inverse gamma density of the travel times, from which, e.g., means and intervals may be determined.

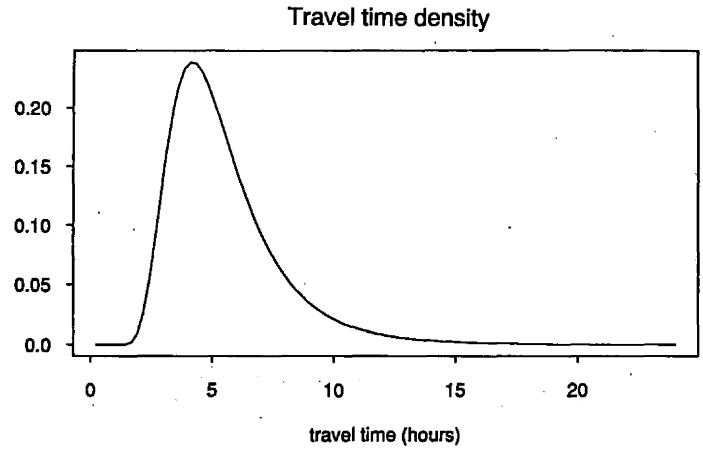

Fig. 4. The estimated travel time density resulting from assuming a gamma distribution for velocities

An alternate analysis can be based on the method of moments as follows. From (4) one sees that

$$\int p_X(t)dt = \mu_R \int p_M(t)dt,$$
$$\int t p_X(t)dt = \mu_R\left[\int t p_M(t)dt + \int p_M(t)dt \int u f(u)du\right],$$

suggesting how to estimate the mean $\int u f(u)du$ of the travel time distribution. With α the mean of the control data, the next estimate is based on

$$\int t p_X(t)dt \cong \sum_t \left(t-\frac{1}{2}\right)|Y(t)-\hat{\alpha}|_+ \left(\sum_t |Y(t)-\hat{\alpha}|_+\right)^{-1} = 19.87,$$

where $|Y(t)|_+$ is 0 if $Y(t)$ is less than its standard error (estimated from the control data) and is $Y(t)$ otherwise. Since $\int tp_M(t)dt \cong (B+A)/2 = 14.5$, the estimated average travel time by this method is $19.87-14.50 = 5.37$ hours. In a similar fashion the variance may be estimated by $12.07-14^2/12 = -4.26$. This negative estimate could be due to a high level of uncertainty or evidence that the model is incorrect.

A third analysis to consider involves deconvolving an empirical version of (4) to obtain an estimate of $f(\cdot)$. Fig. 5 was prepared to consider this possibility.

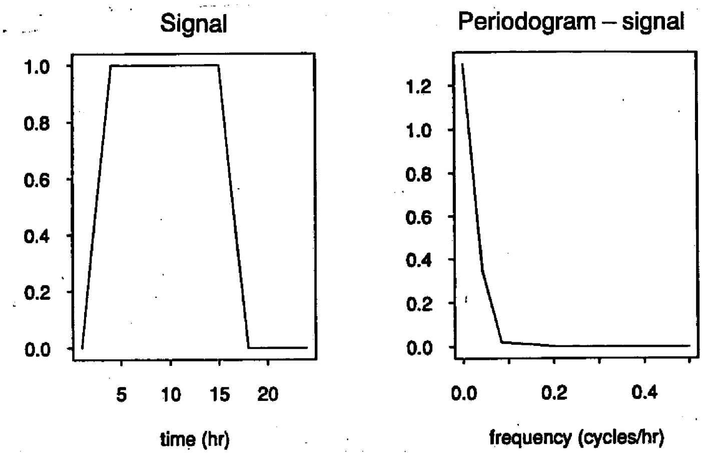

Fig. 5. The left-hand display is the signal, including the effects of aggregation and the running mean. The right-hand display is the corresponding periodogram

The left-hand display provides the signal modified for the effects of aggregation and for the running mean. Most of the mass of its periodogram (given in the right-hand display) is close to 0 indicating that there is little hope in proceeding this way. It may be compared with the right-hand display of Fig. 2. The sampling interval is too large and the running mean has suppressed too much of the information.

6. Discussion. Two solutions have been put forward for Professor Neyman's problem. The first solution involved keeping careful track of the effects of aggregation and smoothing in developing a regression function. The machinery of stochastic processes generally and point processes particularly handles this quite effectively. It is assumed that the effects of seeding are additive and that the travel times have a random distribution. The second solution is via the method of moments. There are difficulties in working with the data set that motivated the problem. Information was lost in forming the hourly totals and the running mean. This speaks against the use of more sophisticated deconvolution techniques or the use of nonparametric change-point techniques (e.g. Müller [15]) that might have proved useful had the original data been available.

The importance of controls and randomization is stressed throughout Professor Neyman's work and in Brillinger et al. [6]. Here it allows effective estimation of α in (6) and the assessment of uncertainty, as in Fig. 3.

Acknowledgement. I have been fortunate to have spent parts of my career with some of the grand people of Statistics, Jerzy Neyman among them. He was the gentleman of statisticians.

REFERENCES

[1] A. Arking, R. C. Lo and A. Rosenfeld, *A Fourier approach to cloud motion estimation*, J. Appl. Met. 17 (1978), pp. 735–744.

[2] B. H. Briggs, *On the analysis of moving patterns in geophysics — I. Correlation analysis*, J. Atmos. Terr. Physics 30 (1777–1794), pp. 1777–1794.

[3] D. R. Brillinger, *Cross-spectral analysis of processes with stationary increments including the $G/G/\infty$ queue*, Ann. Probab. 2 (1974), pp. 815–827.

[4] — *Some aspects of the analysis of evoked response experiments*, in: *Statistics and Related Topics*, D. A. Dawson, J. N. K. Rao and A. K. Md. Saleh (Eds.), North-Holland, 1981, pp. 155–168.

[5] — *Distributions of particle displacements via higher-order moment functions*, IEE PROC–F 140 (1993), pp. 390–394.

[6] — L. V. Jones and J. W. Tukey, *The Role of Statistics in Weather Resources Management, II. The Management of Weather Resources*, Washington Government Printing Office, 1978.

[7] M. A. Cameron and E. J. Hannan, *Measuring the properties of plane waves*, Math. Geol. 10 (1978), pp. 1–22.

[8] G. C. Carter (Editor), *Coherence and Time Delay Estimation*, IEEE Press, New York 1993.

[9] R. V. Foutz, *Estimation of a common group delay between two multiple time series*, J. Amer. Statist. Assoc. 75 (1980), pp. 779–788.

[10] H. W. Freedman, *The "little variable factor" — a statistical discussion of the reading of seismograms*, Bull. Seismol. Soc. Amer. 56 (1966), pp. 593–604.

[11] E. J. Hannan, *Non-linear time series regression*, J. Appl. Probab. 8 (1971), pp. 767–780.

[12] L. M. LeCam, *A stochastic description of precipitation*, Proc. Fourth Berk. Symp. Math. Statist. Probab. 3, J. Neyman (Ed.), University of California, Berkeley 1961, pp. 165–186.

[13] D. V. Lindley, *The estimation of velocity distribution from counts*, Proc. Internat. Congress. Math. Amsterdam 3, 1954, pp. 427–444.

[14] P. McDunnough, *Estimating the law of randomly moving particles by counting*, J. Appl. Probab. 16 (1979), pp. 25–35.

[15] H.-G. Müller, *Change-points in nonparametric regression analysis*, Ann. Statist. 20 (1992), pp. 737–761.

[16] J. Neyman, *Indeterminism in science and new demands on statisticians*, Amer. Statist. Assoc. 55 (1960), pp. 625–639.

[17] — *Problems of design and evaluation of rain making experiments*, in: *A Survey of Statistical Design and Linear Models*, J. N. Srivastava (Ed.), North-Holland, 1975, pp. 443–458.

[18] — *Experimentation with weather control*, in: *Applications of Statistics*, P. R. Krishnaiah (Ed.), North-Holland, 1977, pp. 1–25.

[19] — *Strong far-away effects of local cloud seeding*, Probab. Math. Statist. 2 (1982), pp. 89–107.

[20] — and E. L. Scott, *Rain stimulation experiments: design and evaluation*, Proc. WMO/IAMAP Sci. Conf. Weather Modification, WMO, Geneva 1974, pp. 449–457.

[21] C. L. Nikias and J. M. Mendel, *Signal processing with higher-order spectra*, IEEE Signal Proc. Mag (1993), pp. 10–37.

[22] D. T. Pham, J. Möcks, J. Köhler and T. Gasser, *Variable latencies of noisy signals: estimation and testing in brain potential data*, Biometrika 74 (1987), pp. 525–533.

[23] M. J. Phelan, *Point processes and inference for rainfall fields*, in: *Selected Proc. of the Sheffield Symposium on Applied Probability*, I. V. Basawa and R. L. Taylor (Eds.), Inst. Math. Statist., Hayward 1991, pp. 127–148.

[24] P. J. Thomson, *Signal estimation using an array of recorders*, Stochastic Process. Appl. 13 (1982), pp. 201–214.

[25] D. Vere-Jones, *Some applications of probability generating functionals to the study of input-output streams*, J. Roy. Statist. Soc. B 30 (1968), pp. 321–333.

[26] E. Waymire, V. K. Gupta and I. Rodriguez-Iturbe, *Spectral theory of rainfall intensity at the meso-β scale*, Water Resour. Res. 20 (1984), pp. 1453–1465.

[27] C. D. Woody, *Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals*, Med. Biol. Eng. 5 (1967), pp. 539–553.

Statistics Department
University of California
Berkeley, CA 94720, U.S.A.

Received on 18.1.1994