

# 1

## *Statistics of Earthquakes*

*David R. Brillinger*

---

### 1.1 Introduction

Earthquake statistics are facts recorded in the aftermath of earthquakes. They are the concern of seismologists, geologists, engineers, government officials, and statisticians. Statisticians are involved because of the large amount and many forms of data that become available following an event. The associated substantive field, seismology, has been called the “science based on data called seismograms” (Aki and Richards [1]). A seismogram is a recorded time series of the displacement, velocity, or acceleration experienced by a particle of the Earth. Figure 1 presents an example of this basic datum. It is a record made at Berkeley of the January 17, 1995 earthquake that caused over 5000 casualties and much damage in Kobe, Japan. One notes a variety of wiggles and fluctuations of varying amplitudes. The seismologist attaches physical significance to such features and reads off specific quantitative values, e.g., arrival times of various waves. In the figure the first 9 min or so is noise; then the signal begins to arrive.

Statistical methods have played an important role in seismology for many years. Perhaps this is due to the pathbreaking efforts of Harold Jeffreys [4]. Of Jeffreys’ work, Hudson [18] has written: “The suc-

cess of the Jeffreys–Bullen travel time tables was due in large part to Jeffreys’ consistent use of sound statistical methods.” In particular, Jeffreys’ methods were robust and resistant.

Statistics enters for a variety of reasons. The basic quantity of concern is often a probability or *risk*. The data sets are massive and of many types. There is substantial inherent variability and measurement error. Models need to be fitted and revised. Inverse problems, with the basic parameters defined indirectly, need to be solved. Experiments need to be designed. Sometimes the researcher must fall back on simulations. It can be noted that new statistical techniques appear to find immediate application in seismology particularly and in geophysics generally. As well, problems arising in the field of seismology have led to the development of new statistical techniques.

The field of seismology is largely observational. It underwent the “digital revolution” in the fifties and continually poses problems exceeding the capabilities of the day’s computers. Researchers have turned up a variety of empirical laws [24], which prove useful for extrapolation to situations with few data. Physical theories find important application [1]. The subject matter developed leads to hazard estimation

## 2 Statistics of Earthquakes

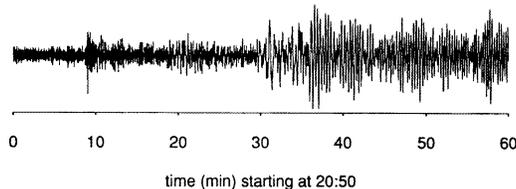


Figure 1: Honshu event of January 17, 1995.

[50], improved seismic design [32], earthquake prediction [28], determination of insurance premiums [9], and general knowledge of the structure of the Earth [53].

### 1.2 Problems

Specific problems addressed by earthquake researchers include the detection, location, and quantification of seismic events; risk assessment; prediction of earthquakes; the distinguishing of earthquakes from nuclear explosions; and learning about the Earth's interior for example, (the determination of wave velocities as a function of depth). Statistical methodology is employed on all of these problems. The researchers are interested in structural questions such as: How should seismometers be laid out in a network [45]? How can one predict earthquakes [28,39]? Are layer boundaries flat or bumpy [37]? Is activity on different faults associated? Further, in problems of risk assessment, there is a need for attenuation laws providing the falloff of strength of earthquake effect with distance from the seismic source. Algorithms are needed for automatically detecting the onset of a strong earthquake and thence shutting down a nuclear reactor at high risk. General references providing basic seismological background include Aki and Richards [1], Bullen and Bolt [10], and Bolt [5].

### 1.3 Types of Data

Measurements may be made close to the seismic source or far away. The recording of data may be continual, as at an observatory, or brief, as when strong-motion instruments are triggered by substantial motion. The data processed may be the seismogram  $Y(t), t = 0, \dots, T - 1$ , as in the figure, or they may be quantities derived from the seismogram. Quantitative values read off a seismogram include the first arrival time, direction of first motion, first-motion amplitude, signal duration, maximum overall amplitude, and oscillation periods. Characteristic features may be noted to infer the individual arrival times of superposed waves of different types; see Simon [41]. Alternatively there may be arrays of seismometers with the instruments arranged in such a fashion that an earthquake signal may be seen traveling and changing shape. The data recorded at location  $(x_j, y_j)$  can then be denoted  $Y_j(t), t = 0, \dots, T - 1$ , with  $j$  labeling locations.

The time series recorded at one location are typically trivariate (two horizontal and one vertical components),  $\{Y_1(t), Y_2(t), Y_3(t), t = 0, \dots, T - 1\}$ . Derived values such as an event's origin time, location, and magnitude— $\{\tau_k, (x_k, y_k, z_k), M_k, k = 1, 2, \dots, K\}$ , with  $k$  indexing events—may be collected into so-called catalogs for geographic regions of interest. These catalogs can date back centuries and are a fundamental tool of seismological research. The data may be binary, as in signs

of first wave motion (+ or −, corresponding to up or down).

Sometimes seismological data are not based directly on seismograms. For example, they may be subjective assessments of damage following an event. The modified Mercalli (MM) intensities are used for such a description. They are ordinal-valued. The description of MM intensity VI starts: “Felt by all; many frightened and run outdoors. Some heavy furniture moved . . .,” while that of intensity VII starts: “Everybody runs outdoors. Damage negligible in buildings of good design and construction; slight to moderate in well-built ordinary structures; considerable in poorly built . . .”

### 1.4 Models

Models range from the naive—e.g. the exponential distribution for magnitudes (the so-called Gutenberg–Richter relation)—to the massive and sophisticated. Those commonly employed include simple binomial, Gaussian, Poisson, complex spatial-temporal, and branching. Seismic engineers, for example, proceed by developing stochastic models for the response of a building to seismic input, while seismologists may model the Earth’s interior as random to handle the omnipresent irregularities [19].

An exceedingly broad range of stochastic models have been employed by earthquake researchers. These provide effective summarization of the data and allow addressing questions of interest. For example the sequence of times  $\{\tau_k\}$  of earthquake occurrence in a given region may be viewed as corresponding to part of a realization of a stochastic point process. It becomes a marked point process or jump process  $\{(\tau_k, M_k)\}$  when there is a value (mark) associated with each event time. This could be the event’s magnitude or seismic moment. A basic point-process parameter,

the rate, tells how many earthquakes may be expected in a unit time interval. There are other parameters to describe temporal dependence. A random field or spatial process,  $Y(x, y)$ , can be envisaged as describing realized values of (say) maximum displacements occurring at locations  $(x, y)$  on the Earth’s surface during the course of an earthquake. A fluctuating displacement value in time and space,  $Y(x, y, t)$ , may be viewed as a spatial-temporal process. A branching process may correspond to crack or geological-fault formation or underlie the times and locations of events [47,48].

There are many uses made of the models of elementary statistics such as multiple regression and nonlinear regression. The generalized linear model is beginning to be employed for data that are counts or proportions or are necessarily positive.

### 1.5 Statistical Methods

Reviews of statistical techniques applied to earthquake data are provided by Jeffreys [21], who describes the methods employed through the mid sixties, Vere-Jones and Smith [49], who provide a review of many contemporary instances up through 1980, and Vere-Jones [48], who presents more recent work.

Much of modern seismological research is based on the spectral analysis of seismograms; see Reference 2. Other specific statistical methods that have been employed are maximum likelihood [16], errors in variables [14], robust regression [6], nonlinear regression [6], probit analysis [49], Fourier inference [20], discrimination [43], array analysis [40,42], point processes [12,17,25,30,46], moment functions [22], inverse problems [36], bootstrap [27], and sensitivity analysis [38].

The smoothness-priors approach to non-stationary data [26] leads to plausible plots of time-varying frequency content

## 4 Statistics of Earthquakes

of seismic signals. Reference 31 presents maximum-likelihood state-based methods for handling the data of reflection seismology; Reference 13 indicates how the EM method may be employed to deconvolve pulses hidden in seismic traces. The non-Gaussianity of seismograms is taken advantage of in higher-order moment analysis [15]. Researchers [33,34,35,23] have carried out a variety of likelihood-based analyses of earthquake times as a point process\*. An important conceptual development is the systems approach of breaking down a circumstance into components as in problems of seismic risk analysis [11]. References 7, 8 present a variety of statistical analyses of earthquake data.

### 1.6 The Literature

The principal journals of the field include the *Bulletin of the Seismological Society of America*, *Journal of Geophysical Research*, *Geophysical Journal of the Royal Astronomical Society*, *Geophysical Research Letters*, and *Mathematical Geology*.

The field of seismology has always been remarkable for the speed with which the data are shared. Nowadays catalogs and waveforms may be obtained directly from many sites through the Internet. One list of computer addresses is given in Reference 29.

### References

1. Aki, K. and Richards, P. G. (1980). *Quantitative Seismology*. Freeman, San Francisco.
2. Bath, M. (1974). *Spectral Analysis in Geophysics*. Elsevier, Amsterdam.
3. Bolt, B. A. (1960). The revision of earthquake epicentres, focal depths and origin-times using a high speed computer. *Geophys. J. R. Astron. Soc.*, **3**, 433–440.
4. Bolt, B. A. (1989). Sir Harold Jeffreys (1891–1989). *Bull. Seismol. Soc. Amer.*, **79**, 2006–2011.
5. Bolt, B. A. (1993). *Earthquakes*. Freeman, New York.
6. Bolt, B. A. and Abrahamson, N. A. (1982). New attenuation relations for peak and expected accelerations of strong ground motion. *Bull. Seismol. Soc. Amer.*, **72**, 2307–2321.
7. Brillinger, D. R. (1988). Some statistical methods for random process data from seismology and neurophysiology. *Ann. Statist.*, **16**, 1–54.
8. Brillinger, D. R. (1989). Some examples of the statistical analysis of seismological data. In *Observatory Seismology*, J. J. Litehiser, ed. University of California Press, Berkeley, pp. 266–278.
9. Brillinger, D. R. (1993). Earthquake risk and insurance. *Environmetrics*, **4**, 1–21.
10. Bullen, K. E. and Bolt, B. A. (1985). *An Introduction to the Theory of Seismology*, 4th ed. Cambridge University Press, Cambridge.
11. Cornell, C. A. (1968). Engineering seismic risk analysis. *Bull. Seismol. Soc. Amer.*, **58**, 1583–1606.
12. Cornell, C. A. and Winterstein, S. R. (1988). Temporal and magnitude dependence in earthquake recurrence models. *Bull. Seismol. Soc. Amer.*, **78**, 1522–1537.
13. Der, Z. A., Lees, A. C., McLaughlin, K. L., and Shumway, R. H. (1992). Deconvolution of short period and teleseismic and regional time series. In *Environmental and Earth Sciences*, A. T. Walden and P. Guttorp, eds. Halstead, New York, pp. 156–188.
14. Ganse, R. A., Amemiya, Y., and Fuller, W. A. (1983). Prediction when both variables are subject to error, with application to earthquake magnitudes. *J. Amer. Statist. Ass.*, **78**, 761–765.
15. Giannakis, G. B. and Mendel, J. M. (1986). Tomographic wavelet estimation via higher-order statistics. *Proc. Fifty-Sixth Internat. Conf. Soc. Explor. Geophys.*, Houston, Tex., pp. 512–514.

16. Guttorp, P. and Hopkins, D. (1989). On estimating varying  $b$  values. *Bull. Seismol. Soc. Amer.*, **76**, 889–895.
17. Hawkes, A. G. and Adamopoulos, L. (1973). Cluster models for earthquakes—regional comparisons. *Bull. Int. Statist. Inst.*, **45**, 454–461.
18. Hudson, J. A. (1981). Mathematics in seismology. *J. Inst. Math. Appl.*, **17**, 34–39.
19. Hudson, J. A. (1982). Uses of stochastic models in seismology. *Geophys. J. R. Astron. Soc.*, **69**, 649–657.
20. Ihaka, R. (1993). Statistical aspects of earthquake source parameter estimation in the presence of signal generated noise. *Commun. Statist. A*, **22**, 1425–1440.
21. Jeffreys, H. (1967). Statistical methods in seismology. In *International Dictionary of Geophysics*, K. Runcorn, ed. Pergamon, London, pp. 1398–1401.
22. Kagan, Y. Y. (1981). Spatial distribution of earthquakes: the four point moment function. *Geophys. J. R. Astron. Soc.*, **67**, 719–733.
23. Kagan, Y. Y. (1991). Likelihood analysis of earthquake catalogs. *Geophys. J. R. Astron. Soc.*, **106**, 135–148.
24. Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relations in seismology. *Bull. Seismol. Soc. Amer.*, **65**, 1073–1095.
25. Kiremidjian, A. S. and Suzuki, S. (1987). A stochastic model for site ground motions from temporally dependent earthquakes. *Bull. Seismol. Soc. Amer.*, **77**, 1110–1126.
26. Kitagawa, G. and Gersch, W. (1996). *Smoothness Priors Analysis of Time Series*. Springer, New York.
27. Lamarre, M., Townsend, B. and Shah, H. C. (1992). Application of the bootstrap method to quantify uncertainty in seismic hazard estimates. *Bull. Seismol. Soc. Amer.*, **82**, 104–119.
28. Lomnitz, C. (1994). *Fundamentals of Earthquake Prediction*. Wiley, New York.
29. Malone, S. (1993). Seismology and the information super-highway. *Seismol. Res. Lett.*, **6**, 28–30.
30. McGuire, R. K., Toro, G. R., Veneziano, D., Cornell, C. A., Hu, Y. X., Jin, Y., Shi, Z., and Gao, M. (1992). *Non-stationarity of historical seismicity in China. Proc. Tenth World Conf. on Earthquake Engineering*, Balkema, Rotterdam, pp. 287–292.
31. Mendel, J. M. (1983). *Seismic Deconvolution: An Estimation-Based Approach*. Academic, New York.
32. Naeim, F., ed. (1989). *The Seismic Design Handbook*. Van Nostrand, New York.
33. Ogata, Y. (1983). Likelihood analysis of point processes and its application to seismological data. *Bull. Int. Statist. Inst.*, **50**, 943–961.
34. Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Ass.*, **83**, 9–27.
35. Ogata, Y. and Katsura, K. (1986). Point-process models with linearly parametrized intensity for application to earthquake data. *J. Appl. Probab.*, **23A**, 291–310.
36. O’Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, **1**, 501–527.
37. Pulliam, R. J. and Stark, P. B. (1993). Bumps on the core–mantle boundary—are they facts or artifacts? *J. Geophys. Res. Solid Earth*, **98**, 1943–1955.
38. Rabinowitz, N. and Steinberg, D. M. (1991). Seismic hazard sensitivity analysis: a multiparameter approach. *Bull. Seismol. Soc. Amer.*, **81**, 796–817.
39. Rhoades, D. A. (1986). Predicting earthquakes. In *The Fascination of Statistics*, R. J. Brook et al., eds. Dekker, New York, pp. 307–319.
40. Shumway, R. H. (1983). Replicated time-series regression: an approach to signal estimation and detection. In *Time Series in the Frequency Domain*, D. R.

## 6 *Statistics of Earthquakes*

- Brillinger and P. R. Krishnaiah, eds. North-Holland, Amsterdam, pp. 383–408.
41. Simon, R. B. (1981). *Earthquake Interpretations*. Williams Kaufmann, Los Altos.
42. Thomson, P. J. (1992). Signal estimation using stochastic velocity models and irregular arrays. *Ann. Inst. Statist. Math.*, **44**, 13–25.
43. Tjøstheim, D. (1981). Multidimensional discrimination techniques—theory and application. In *Identification of Seismic Sources—Earthquake or Underground Explosion*, E. S. Husebye and S. Mykkeltveit, eds. Reidel, Dordrecht, pp. 663–694.
44. Udias, A. (1989). Development of fault-plane studies for mechanism of earthquakes. In *Observatory Seismology*, J. J. Litehiser, ed. University of California Press, Berkeley, pp. 243–256.
45. Uhrhammer, R. A. (1982). The optimal estimation of earthquake parameters, *Phys. Earth Planetary Interiors*, **30**, 105–118.
46. Veneziano, D. and Van Dyck, J. (1987). Statistical analysis of earthquake catalogs for seismic hazard. In *Stochastic Approaches to Earthquake Engineering*, Y. K. Lin and R. Minai, eds. Springer, New York, pp. 385–427.
47. Vere-Jones, D. (1970). Stochastic models of earthquake occurrence. *J. R. Statist. Soc. B*, **32**, 1–62.
48. Vere-Jones, D. (1992). Statistical methods for the description and display of earthquake catalogs. In *Statistics in Environmental and Earth Sciences*, A. T. Walden and P. Guttorp, eds. Halstead, New York, pp. 220–246.
49. Vere-Jones, D. and Smith, E. G. C. (1981). Statistics in seismology. *Commun. Statist. Theory Methods A*, **10**, 1559–1585.
50. Wesnousky, S. G. (1986). Earthquakes, quaternary faults, and seismic hazard in California. *J. Geophys. Res.*, **91**, 12587–12631.

David R. Brillinger  
Department of Statistics, University of  
California, Berkeley CA USA  
brill@stat.berkeley.edu