Title
On earthquake predictability measurement: Information score and error diagram

Journal
Pure and Applied Geophysics, 164(10)

ISSN
0033-4553

Author
Kagan, Yan Y.

Publication Date
2007-10-01

DOI
10.1007/s00024-007-0260-1

Peer reviewed

On earthquake predictability measurement: information score and error diagram

Yan Y. Kagan[1*]
[1] Department of Earth and Space Sciences
University of California, Los Angeles, California, USA

June 13, 2008

Abstract

We discuss two methods for measuring the effectiveness of earthquake prediction algorithms: the information score based on the likelihood ratio and error diagrams. For both of these methods, closed form expressions are obtained for the renewal process based on the gamma and lognormal distributions. The error diagram is more informative than the likelihood ratio and uniquely specifies the information score. We derive an expression connecting the information score and error diagrams. We then obtain the estimate of the region bounds in the error diagram for any value of the information score. We discuss how these preliminary results can be extended for more realistic models of earthquake occurrence.

KEY WORDS: Earthquake prediction, Statistical methods, Seismicity, Renewal processes, Information score, Error diagram

1 Introduction

In a recent article, Jordan (2006) argued that more objective, rigorous, quantitative methods for testing earthquake prediction schemes need to be developed. Particularly, he asked "What is the intrinsic predictability of the earthquake rupture process?" To contribute to this inquiry we discuss two methods currently used to measure the performance of earthquake prediction programs.

The first method is the likelihood ratio procedure which has long been used for statistical analysis of random processes. In particular, Kagan and Knopoff (1976; 1987), Kagan (1991), Ogata (1999), Kagan and Jackson (2000), Imoto (2004), Rhoades and Evison (2006), and Helmstetter ***et al.*** (2006) have applied this likelihood method for earthquake occurrence studies. Kagan and Knopoff (1977) first proposed calculating the information score for earthquake predictability based on the likelihood ratio.

The second method is related to the Relative Operating Characteristic (ROC) used in weather prediction efforts (Jolliffe and Stephenson, 2003), where the success rate of an event prediction is compared against the false alarm rate (***ibid.***, p. 69; see also Holliday ***et al.***, 2005). Since periodic (diurnal, annual) effects are strong in weather prediction, such a method has broad applications; we can compare the above characteristics of a forecast system for one-day or one-year alarm periods. But in earthquake prediction, there

is no natural time scale for forecasting, so the time interval is arbitrary. Therefore, if the alarm duration is increased, both criteria approach the trivial result: all events are predicted with no false alarms.

Molchan (1990) modified this method as an error diagram to predict random point processes. Molchan and Kagan (1992) and Molchan (1997; 2003) also review the error diagram method and its applications. Error diagrams are actively used to evaluate earthquake prediction algorighms. Recently, McGuire *et al.* (2005), Kossobokov (2006), and Baiesi (2006) and others used this method for this purpose. Kagan and Jackson (2006) commented on Kossobokov's analysis and discussed the application of error diagrams to predicting earthquakes.

In this paper we consider the simplest stochastic models of earthquake occurrence – the renewal processes (Davis *et al.*, 1989; Daley and Vere-Jones, 2003) based on the gamma and lognormal distributions. The reason for using these models is that the closed form information score expressions exist presently only for these processes (Daley and Vere-Jones, 2004; Bebbington, 2005). We simulate these processes to test both these formulae and our simulation procedure and thereafter to test the derived forward and inverse relations between the information scores and the error diagrams. Although the simple renewal models are widely used for representing large earthquake occurrences on specific faults, the data available presently are rather scarce. We discuss extending our results to more complex earthquake occurrence models.

Extensive earthquake catalog data are available that can be modeled using more sophisticated schemes. A more complete and rigorous investigation of such models is a task for future work.

2 Information score

Kagan and Knopoff (1977) suggested measuring the effectiveness of earthquake prediction algorithm by first evaluating the likelihood ratio to test how well does a model approximate earthquake occurrence. In particular, they estimated the information score, $\hat{I}$, per one event by

$$\hat{I} = \frac{\ell - \ell_0}{N} = \frac{1}{N} \sum_{i=1}^{N} \log_2 \frac{p_i}{\pi_i}, \tag{1}$$

where $\ell - \ell_0$ is the log-likelihood ratio, $\log_2$ was used to obtain the score in bits of information, N is the number of earthquakes in a catalog, p_i is the probability of earthquake occurrence according to a stochastic model, conditioned by the past:

$$p_i = \text{Prob}\{\text{an event in } (t,\, t+\Delta) \,|\, I(t)\}, \tag{2}$$

where $I(t)$ is the past history of the process up to the moment t, and π_i is a similar probability of event occurrence for a Poisson process.

One information bit would mean that uncertainty of earthquake occurrence is reduced on average by a factor of 2 by using a particular model.

Here the 'average' needs to be understood as a geometric mean (Vere-Jones, 1998). For long catalogs ($N \to \infty$)

$$\lim_{N\to\infty} \hat{I} \to I = E\left(\log_2 \frac{p_i}{\pi_i}\right), \tag{3}$$

where E is the mathematical expectation (Vere-Jones, 1998; Daley and Vere-Jones, 2003).

For a renewal (i.e., with independent intervals) process the information score can be calculated as (Daley and Vere-Jones, 2003, their equation 7.6.16)

$$I = m\,(1 - \log m + H), \tag{4}$$

where m is the intensity (rate) of a renewal process and the entropy function H is

$$H = \int_0^\infty f(x) \log f(x)\, dx, \tag{5}$$

where $f(x)$ is a probability density function (pdf) for interevent times.

The entropy function (5) has been calculated in closed form for two distributions, gamma and lognormal (Daley and Vere-Jones, 2004; Bebbington, 2005). Imoto (2004) obtained information score estimates for the lognormal, gamma, and several other distributions. For this purpose (5) was integrated numerically.

The gamma distribution has the pdf

$$f(x) = \frac{\lambda^\kappa x^{\kappa-1}}{\Gamma(\kappa)} \exp(-\kappa x), \tag{6}$$

where Γ is the gamma function, κ is a shape parameter, λ is a scale parameter, and $0 < \lambda < \infty$; $0 < \kappa < \infty$; $0 \leq x < \infty$ (Evans *et al.*, 2000). If $\kappa = 1$, then the process is the Poisson one, $\kappa < 1$ characterizes the occurrence of clustered events.

For the gamma renewal process, normalized to have the mean equal to 1, i.e., $\lambda = \kappa$, the information score is (Daley and Vere-Jones, 2004, their equation 14)

$$I(\kappa) = [\log \kappa + (\kappa - 1)\,\psi(\kappa) - \kappa - \log \Gamma(\kappa)\,] / \log(2)\,, \tag{7}$$

where ψ is the digamma function (Abramowitz and Stegun, 1972, Eq. 6.3.1). If $\kappa = 1$, $I(\kappa) = 0$. This function is shown in Fig. 1. The small κ-values correspond to a clustered process, the large values correspond to a quasi-periodic one.

The lognormal distribution has the pdf

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\,[\,\log(x) - \mu\,]^2\right\}\,, \tag{8}$$

where σ is a shape parameter, μ is a scale parameter and $0 < \mu < \infty$; $0 < \sigma < \infty$; $0 \leq x < \infty$ (Evans *et al.*, 2000). For a renewal process normalized to have the mean interevent time equal to 1.0

$$\hat{\lambda} = \exp(-\mu - \sigma^2/2) = 1.0\,, \tag{9}$$

or $\mu = -\sigma^2/2$, the information score for the lognormal distribution is (Bebbington, 2005, p. 2303)

$$I(\sigma) = \left[\frac{1+\sigma^2}{2\,\log(2)} - \log_2\left(\sigma\sqrt{2\pi}\right)\right] . \qquad (10)$$

This function also is shown in Fig. 1. Contrary to the gamma distribution plot, the small σ-values correspond to a quasi-periodic process, the large values to a clustered one. The sequence with the parameter value $\sigma = 1$ is the closest to the Poisson process, its information score is at minimum, but is still non-zero, $I \simeq 0.117$ bits.

Fig. 2 displays a simulation of the lognormal renewal process for $\sigma = 1.86$. According to (10), the average information score for such a process would be 1 bit. An initial alarm with the duration 0.1 is declared after each event. If a new event occurs during a declared alarm, the alarm time is extended accordingly (*cf.* Stark, 1997), thus individual alarms may significantly exceed the length of the initial one. These alarms are shown at the bottom of the plot.

3 Error diagrams

The error diagram for evaluating how well a prediction program performs was first suggested by Molchan (1990). For any prediction algorithm, the diagram plots the fraction of alarm time, τ, versus the fraction of failures to

predict, ν. For the optimal prediction algorithm the lower hull of points is concave (Molchan, 1997; 2003).

Fig. 3 shows a sketch of the error diagram. The diagonal of the square corresponds to the random guess prediction strategy: an alarm is declared independently of the process past history (Molchan, 1991; Molchan and Kagan, 1992). The other curves are for hypothetical prediction algorithms. Two dashed line curves are error diagram curves consisting of two segments, the lower curve is for a certain forecast strategy and the upper curve is an 'antipodal' prediction, obtained by a reversal of alarm declaration (*ibid.*).

The information score or Shannon's information for line curves, consisting of straight line segments, can be calculated, extending arguments by Molchan (1991) or Molchan and Kagan (1992), as

$$I = \sum_{j=1}^{n+1} (\nu_j - \nu_{j+1}) \times \log_2 \left[-\frac{\nu_j - \nu_{j+1}}{\tau_j - \tau_{j+1}} \right] . \qquad (11)$$

where n is the number of line segments and τ_j and ν_j are the coordinates of line segment ends. Eq. 11 shows that the information score in Fig. 3 is the same for the forecast (dashed curve) and for its antipodal prediction ($I = 1$ bit). The score is 0, 0.84, and 0.63 bits for solid line, dash-dotted and dotted line curves, respectively.

The error diagram curve for a clustered renewal process can be calculated as follows. An alarm is declared after each event for a fixed initial time (Δt). We normalize the initial alarm window w by dividing it by the mean

interevent time, $\bar{T}$, i.e., making w dimensionless

$$w = \frac{\Delta t}{\bar{T}} . \tag{12}$$

As we discussed above, for both distributions the mean interval $\bar{T}$ is adjusted to be 1.0. If a new event occurs during a declared alarm, the alarm time is extended by w. Such an alarm declaration strategy produces an optimal prediction curve for a clustered process, in which the hazard rate is monotonically decreasing (Molchan, 1991; 2003).

The fraction of failures to predict is

$$\nu(w) = 1 - \int_0^w f(x)\, dx . \tag{13}$$

The fraction of total alarm time is

$$\tau(w) = w\, \nu(w) + \int_0^w x\, f(x)\, dx . \tag{14}$$

The first right-hand term in (14) is the average alarm duration, if no event occurs in the w interval. The second term is the average alarm length if an event or events occur during a declared alarm.

For the lognormal and gamma renewal processes the variables $\nu(w)$ and $\tau(w)$ can be found in a closed form. For the gamma distribution

$$\nu(w) = 1 - \gamma(\kappa, \kappa w)/\Gamma(\kappa) , \tag{15}$$

and

$$\tau(w) = w\, \nu(w) + \gamma(1+\kappa, \kappa w)/\Gamma(1+\kappa) , \tag{16}$$

where γ is the incomplete gamma function (Abramowitz and Stegun, 1972, Eq. 6.5.2).

For the lognormal distribution

$$\nu(w) \;=\; \frac{1}{2}\left\{1 - \mathrm{erf}\left[\frac{\sigma^2 + 2\log(w)}{2\sigma\sqrt{2}}\right]\right\}\,, \tag{17}$$

and

$$\tau(w) \;=\; w\,\nu(w) \;+\; \frac{1}{2}\left\{1 + \mathrm{erfc}\left[\frac{\sigma^2 - 2\log(w)}{2\sigma\sqrt{2}}\right]\right\}\,. \tag{18}$$

Here erf and erfc are the error function and its complementary function, respectively.

Contrary to similar expressions (15–16) for the gamma distribution, equations (17–18) do not specify an optimal prediction strategy. In the optimal strategy the alarm is declared when the hazard function or the pdf exceeds a certain level (Molchan and Kagan, 1992; Molchan, 1997; 2003). For the lognormal distribution the hazard function increases initially but eventually decreases again. However, for the highly clustered version of the process $\sigma = 1.86$, the initial increase is concentrated at very small time intervals, hence the error curve is close to be optimal. For example, if $\sigma = 1.86$, only for $\tau < 0.0003$ the error curve is above the random guess strategy, i.e., the forecast curve is not optimal. However, these considerations are irrelevant for the problem we are considering here: the relation between an error curve and the information score. As shown in Fig. 3 and Eq. 3 even the antipodal curve yields the same value of I as the real forecast curve. Our numerical

experiments show that (17–18) for any value of σ specify a curve which, when processed by (3) or by Eq. 19 below, yields the information score as in (10).

Fig. 4 shows an example of the error diagram for the lognormal renewal process. The theoretical curves (17, 18) are compared to the simulation results for 33 choices of the alarm duration w. The alarm duration starts with 0.001 and then increases logarithmically with the factor value 1.52 until it reaches the total length of a series (1000 units). We deliberately use relatively short simulated sequences to show random fluctuations.

4 Relation between the error diagram and information score

The information score can be calculated for an error diagram curve as an extension of (11) to continuous concave curves

$$I = \int_0^1 \log_2 \left(-\frac{\partial \nu}{\partial \tau} \right) d\nu \,. \tag{19}$$

It is helpful to have an estimate of the region boundaries for curves corresponding to a specific value of the information score. Such an estimate can be obtained if the prediction scheme in the error diagram consists of two linear segments (*cf.* Fig. 3) with the slopes

$$D_1 = -\frac{\partial \nu_1}{\partial \tau_1} = -\frac{\Delta \nu_1}{\Delta \tau_1}, \tag{20}$$

for the first segment and D_2 for the second segment;

$$D_1 \geq 1\,, \text{ and } D_2 \leq 1\,, \tag{21}$$

given a curve concavity. $D_1 = D_2 = D_0 = 1$ is the random guess strategy (Molchan, 1990).

For the assumed information score I the envelope curve is defined by the equation

$$D_1 \left[\frac{\nu}{\nu - 1 - D_1} \right]^{\nu} = -2^I . \tag{22}$$

By solving this equation for any value of D_1, one obtains the ν-value for the contact point of two linear segments, $\tau = (\nu - 1)/D_1$.

As an example, Fig. 5 displays several two line segments that correspond to the prediction schemes with an information score that equals 1 bit. The envelope curve is also shown. This curve delineates the lower boundary of all possible error curves with the information score $I = 1$ bit. Numerical experiments show that any concave (not straight) line connecting the point (τ, ν) with (0, 1) or (1, 0) would have a larger information score (I). A similar result can be obtained using the Hölder-Jensen inequality. The upper boundary is, in principle, the random guess line, D_0: if $D_1 \to \infty$, the resulting curve can be as close to this line as needed.

We also show in this diagram the results of simulating a mixture of two Poisson processes with the rates differing by a factor 44.4. This factor was adjusted to obtain the information score 1 bit for a renewal process in which intervals have been selected randomly from each sequence. The simulation results are similar to theoretical curves having two straight line segments.

Fig. 6 displays the curves for the renewal processes with gamma and lognormal distributions. The information score again is taken to be 1 bit. Curves for both sequences, clustered and quasi-periodic, are shown. All four curves are within the region specified by (22) and the random guess line.

When simulating or computing curves for quasi-periodic sequences, alarm declaration is reversed, i.e., it is declared after the elapsed w time period following an event. Alternatively, an alarm strategy is the same as in clustered sequences producing an antipodal prediction (Molchan and Kagan, 1992; Molchan, 1997). Then a curve is rotated 180° around the center of symmetry $[\tau = 1/2;\ \nu = 1/2]$.

The curves behavior in Fig. 6 is difficult to see in the neighborhood of the point $\tau = 0$, $\nu = 1$, so we display the curves in a semi-log format (Fig. 7). The clustered gamma distribution ($\kappa = 0.329$) is seen as approaching $\nu = 1$ line more slowly than the other curves. This means that for even very small alarm time intervals, some prediction capability is still available for this model. The lower envelope of (22) also has non-zero values for the fraction of predicted events $(1 - \nu)$ for small τ, i.e., although the curve asymptotically approaches the purely random forecasts, this convergence is very slow.

5 Discussion

In this work we use simulations as well as previously and newly derived analytic expressions for renewal stochastic processes to test four relations:

(a) between the pdf's of the gamma and lognormal distributions and the information score;

(b) between the pdf's of these distributions and error curves;

(c) between the error curve and the information score;

(d) the inverse relation between the information score and the error curve.

Clearly, from the theoretical and the simulation results described above, the error diagram represents a much more complete picture of the stochastic point process than does the likelihood analysis. Using the diagram curve one can calculate the information score for a renewal process sequence. The score also imposes some limits on the diagram region where curves are located, but Figs. 5 and 6 show that these limitations are rather broad. By specifying a more restricted class of point processes to approximate an earthquake occurrence, the relation between these two methods can likely be made more precise.

As we mentioned in the Introduction section, the available information is limited to few or even only one historical large earthquake on a single tectonic structure. Because of this the renewal models, analyzed in this work, may have only a minimal application in the statistical analysis of seismicity patterns. However, even with these qualifications, a significant number of publications apply the renewal or similar models (Davis ***et al.***, 1989; Molchan, 1990, 1991; Imoto, 2004; Jackson and Kagan, 2006) for earthquake occurrence description. Therefore, it is important to have a quantitative compar-

ison of various prediction techniques applied to such models. It is even more important to extend these results to more complex models of earthquake occurrence: for example, taking into account insufficient earthquake data, various modeling assumptions, multidimensionality of the earthquake process, and so on. Below we will offer a few comments on how the techniques discussed above can be used to forecast earthquake rupture.

- 1) The derived equations apply to very long processes ($N \to \infty$), they need to be extended for a small number of predicted earthquakes when random fluctuations could modify results. For example, in Fig. 4, the direct application of (11) may give a zero or a negative argument for a logarithm. If we form a concave hull of simulated points, the obtained simulation curve would be lower than the theoretical estimate. Molchan (1990, Eqs. 18 and 19) derived asymptotic expressions for statistical estimates of τ and ν for sufficiently long renewal processes. This result can be used to evaluate uncertainties in error curve estimates as in Fig. 4.

- 2) Our analysis assumes that earthquake process fits the models exactly. In reality we work with models which are at best imperfect approximations to real stochastic relations between earthquakes. Daley and Vere-Jones (2004, p. 301) suggest that the information score is at a maximum if the true model is used for the process description. Harte and Vere-Jones (2005, p. 1240) also write that "the [ν-τ] curve for a wrong model always lies above the curve for the true model." If the data do not match the model, the obtained score $\hat{I}$

(3) should be below the maximum, thus the error curve would be above the corresponding lower envelope (see Figs. 5, 6). However, the actual shape of the curve would depend on the employed model pdf form, and hence may differ significantly from the true model curve. In Figs. 5 and 6 various curves corresponding to $I = 1$ bit are quite different and it is possible that a curve corresponding to a lesser information score may in some τ-interval be below a curve with a higher I. This problem needs to be investigated more closely.

- 3) Earthquake prediction algorithms are tested against the Poisson hypothesis even though earthquakes do not follow this distribution. Testing against weak null hypothesis biases the results (Stark, 1997). More realistic reference distributions need to be introduced in the measurements of the earthquake prediction efficiency.

- 4) Additionally, a more appropriate stochastic model for earthquake occurrence is not a renewal but a branching process (Hawkes and Oakes, 1974; Kagan and Knopoff, 1976; Ogata, 1999) which captures the important feature of seismicity, its clustering. Renewal processes can yield clustering features, but in contrast to branching models their clustering is symmetric in time. Earthquake occurrence is highly time asymmetric, there are few, if any, foreshocks, and many aftershocks. These events often exhibit secondary clustering.

- 5) Moreover, earthquakes occur not only in time. Their spatial coordinates, earthquake size, and focal mechanisms need to be taken into account

in actual prediction efforts (Kagan and Jackson, 2000). Introducing new variables complicates the calculation of the information score and the error diagram. Molchan and Kagan (1992) have done some preliminary work in determining error diagrams for multidimensional processes. Kagan and Jackson (2000) and Helmstetter *et al.* (2006) have shown how to evaluate the effectiveness of spatial smoothing for seismic hazard maps.

The results of statistical analysis of the central California catalog (Kagan and Knopoff, 1987; Kagan, 1991; Molchan and Kagan, 1992) suggest that (19) is valid for multidimensional processes as well. In Fig 8 we display the error curve obtained for the catalog, its processing by (11) yields $\hat{I} =$ 1.86 bits. As in Fig. 7 the model exhibits a significant predictive power even for very small space-time alarm volumes. The information score obtained by the likelihood procedure is 1.58 bits (Kagan and Knopoff, 1987; Kagan, 1991). There are many potential sources for the score discrepancy: random fluctuations due to the limited size of the catalog, possible biases in handling boundary effects, etc. As we mentioned above (item 1), random fluctuations may increase the estimate of $\hat{I}$, obtained from the error diagram.

- 6) Another challenge in dealing with earthquake prediction is the fractal nature of most distributions controlling earthquakes (Kagan, 2006). Since these distributions approach infinity for small time and distance intervals, the value of the information score is not well defined (see Helmstetter *et al.*, 2006). Similarly, the error diagram curve would start to approach the point of

the ideal prediction ($\tau = 0$, $\nu = 0$) for earthquake catalogs of high location accuracy and extending to small time intervals after a strong earthquake. Clearly both predictability measurement methods need to be significantly modified to apply to fractally distributed seismicity forecasting.

- 7) None of the above-mentioned models would be likely to allow an analytic computation of the information score or the error diagram curves. Therefore, the solution to most of the problems listed in this section can only be obtained by simulation. Most likely simulations need to be started for simpler models and then model complexity could be increased. If the simulations could not be tested against closed form expressions, the results would be less reliable. Nevertheless, we hope that such simulations would show that Eqs. 19 and 22 which describe the relationship between two measures of earthquake predictability turn out to have a general applicability beyond the simple models analyzed in this paper.

Finally, I would like to mention that Eq. 19 was derived, using heuristic arguments, exemplified in Fig. 3, in December 1991 – January 1992. Since that time I have privately sent these preliminary results to many researchers interested in the problem. Recently, Harte and Vere-Jones (2005) published a similar formula (see the first right-hand term in their Eq. 18) for a model of the discrete-time point process. Actually, for this term $d\tau$ can be cancelled and since the second term disappears for a continuous process, their equation is becoming almost identical to our (19). However, they did not explore the

connection between the error diagram properties and the information score or any constraints the information score would impose on the diagram.

Acknowledgments

I appreciate partial support from the National Science Foundation through grants EAR 04-09890, and DMS-0306526, as well as from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. I am very grateful to P. Stark of UC Berkeley who sent me MATLAB programs employed in his (Stark, 1997) paper. With appropriate modifications these programs were used in some calculation reported above. I also thank D. D. Jackson, I. V. Zaliapin, F. Schoenberg, and J. C. Zhuang of UCLA, D. Vere-Jones of Wellington University and P. Stark for very useful discussions. Reviews by Rodolfo Console and by two anonymous reviewers have been very helpful in revising the manuscript. I thank Kathleen Jackson for significant improvements in the text. Publication 1058, SCEC.

References

[1] Abramowitz, M. and I. A. Stegun (1972), *Handbook of Mathematical Functions*, Dover, NY, pp 1046.

[2] Baiesi, M. (2006), Scaling and precursor motifs in earthquake networks, *Physica A*, **360**(2), 534-542.

[3] Bebbington, M. S. (2005), Information gains for stress release models, *Pure Appl. Geophys.*, **162**(12), 2299-2319.

[4] Daley, D. J., and Vere-Jones, D. (2003), *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, 2-nd ed., Vol. 1, pp. 469.

[5] Daley, D. J., and Vere-Jones, D. (2004), Scoring probability forecasts for point processes: The entropy score and information gain, *J. Applied Probability*, **41A**, 297-312.

[6] Davis, P. M., D. D. Jackson, and Y. Y. Kagan (1989), The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seismol. Soc. Amer.*, **79**(5), 1439-1456.

[7] Evans, M., N. Hastings, and B. Peacock (2000), *Statistical Distributions*, 3rd ed., New York, J. Wiley, 221 pp.

[8] Harte, D., and Vere-Jones, D. (2005), The entropy score and its uses in earthquake forecasting, *Pure Appl. Geophys.*, **162**(6-7), 1229-1253.

[9] Hawkes, A. G. and Oakes, D. (1974), A cluster process representation of a self-exciting process, *J. Appl. Prob.*, **11**, 493-503.

[10] Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006), Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. Seismol. Soc. Amer.*, **96**(1), 90-106.

[11] Holliday, J. R., K. Z. Nanjo, K. F. Tiampo, J. B. Rundle, D. L. Turcotte (2005), Earthquake forecasting and its verification, *Nonlinear Processes Geophys.*, **12**(6), 965-977.

[12] Imoto, M. (2004), Probability gains expected for renewal process models, *Earth Planets Space*, **56**, 563-571.

[13] Jackson, D. D., and Y. Y. Kagan (2006), The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes: lessons for the future, *Bull. Seismol. Soc. Amer.*, **96**(4B), S397-S409.

[14] Jolliffe, I. T. and D. B. Stephenson, Eds. (2003), *Forecast Verification: a Practitioner's Guide in Atmospheric Science*, J. Wiley, Chichester, England, 240 pp.

[15] Jordan, T. H. (2006), Earthquake predictability, brick by brick, *Seismol. Res. Lett.*, **77**(1), 3-6.

[16] Kagan, Y. Y. (1991), Likelihood analysis of earthquake catalogues, *Geophys. J. Int.*, **106**, 135-148.

[17] Kagan, Y. Y. (2006), Why does theoretical physics fail to explain and predict earthquake occurrence?, in: *Lecture Notes in Physics*, **705**, pp. 303-359, P. Bhattacharyya and B. K. Chakrabarti (eds.), Springer Verlag, Berlin–Heidelberg.

[18] Kagan, Y. Y., and D. D. Jackson (2000), Probabilistic forecasting of earthquakes, *Geophys. J. Int.*, **143**(2), 438-453.

[19] Kagan, Y. Y., and D. D. Jackson (2006), Comment on 'Testing earthquake prediction methods: "The West Pacific short-term forecast of earthquakes with magnitude MwHRV $\geq$ 5.8"' by V. G. Kossobokov, *Tectonophysics*, **413**(1-2), 33-38.

[20] Kagan, Y., and L. Knopoff (1976), Statistical search for non-random features of the seismicity of strong earthquakes, *Phys. Earth Planet. Inter.*, **12**(4), 291-318.

[21] Kagan, Y., and L. Knopoff (1977), Earthquake risk prediction as a stochastic process, *Phys. Earth Planet. Inter.*, **14**(2), 97-108.

[22] Kagan, Y. Y., and L. Knopoff (1987), Statistical short-term earthquake prediction, *Science*, **236**, 1563-1567.

[23] Kossobokov, V. G. (2006), Testing earthquake prediction methods: "The West Pacific short-term forecast of earthquakes with magnitude MwHRV $\geq$ 5.8", *Tectonophysics*, **413**(1-2), 25-31.

[24] McGuire, J. J., Boettcher, M. S., and Jordan, T. H. (2005), Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults, *Nature*, **434**(7032), 457-461; Correction – *Nature*, **435**(7041), 528.

[25] Molchan, G. M. (1990), Strategies in strong earthquake prediction, *Phys. Earth Planet. Inter.*, **61**(1-2), 84-98.

[26] Molchan, G. M. (1991), Structure of optimal strategies in earthquake prediction, *Tectonophysics*, **193**(4), 267-276.

[27] Molchan, G. M. (1997), Earthquake prediction as a decision-making problem, *Pure Appl. Geoph.*, **149**(1), 233-247.

[28] Molchan, G. M. (2003), Earthquake prediction strategies: A theoretical analysis, In: Keilis-Borok, V. I., and A. A. Soloviev, (Eds) *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, Springer, Heidelberg, 208-237.

[29] Molchan, G. M., and Y. Y. Kagan (1992), Earthquake prediction and its optimization, *J. Geophys. Res.*, **97**(B4), 4823-4838.

[30] Ogata, Y. (1999), Seismicity analysis through point-process modeling: A review, *Pure Appl. Geophys.*, **155**(2-4), 471-507.

[31] Rhoades, D. A., and F. F. Evison (2006), The EEPAS forecasting model and the probability of moderate-to-large earthquakes in central Japan, *Tectonophysics* **417**(1-2), 119-130.

[32] Stark, P. B. (1997), Earthquake prediction: the null hypothesis, *Geophys. J. Int.*, **131**(3), 495-499.

[33] Vere-Jones, D. (1998), Probabilities and information gain for earthquake forecasting, *Computational Seismology*, **30**, Geos, Moscow, 248-263.

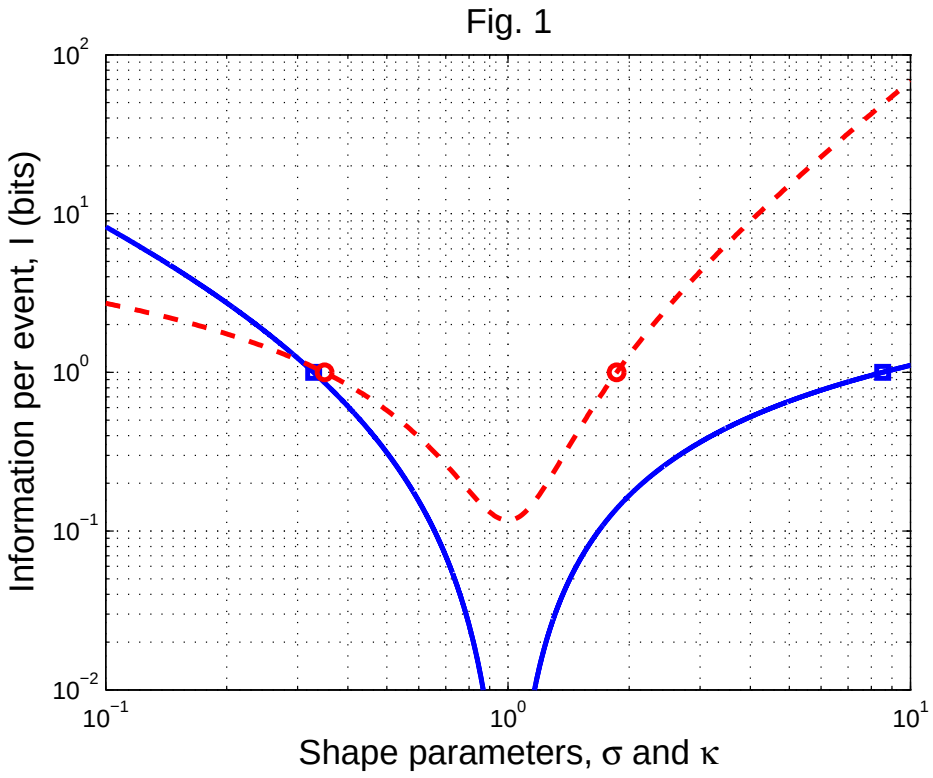

Figure 1: Dependence of the information score on the shape parameters κ and σ for the gamma (solid line) and lognormal (dashed line) renewal process. Two squares and circles show the curves position for $I = 1$ bit.

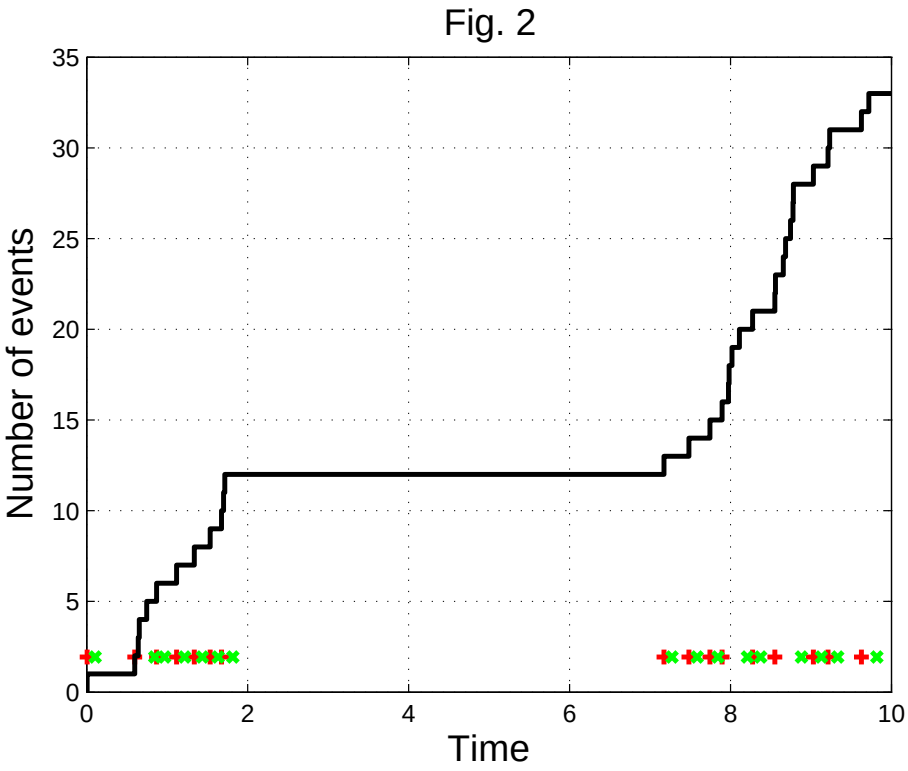

Figure 2: Parts of a realization of lognormal distributed renewal process. Solid line – cumulative number of events. Crosses – beginning of alarms, 'x's – end of alarms. In this example, 33 events are simulated with the shape parameter $\sigma = 1.86$ (see Eq. 8). After each event an alarm with duration 0.1 is issued. 17 events fall into 16 alarms, i.e., they are successfully predicted. The duration of alarms is 23.4 % of the total time.

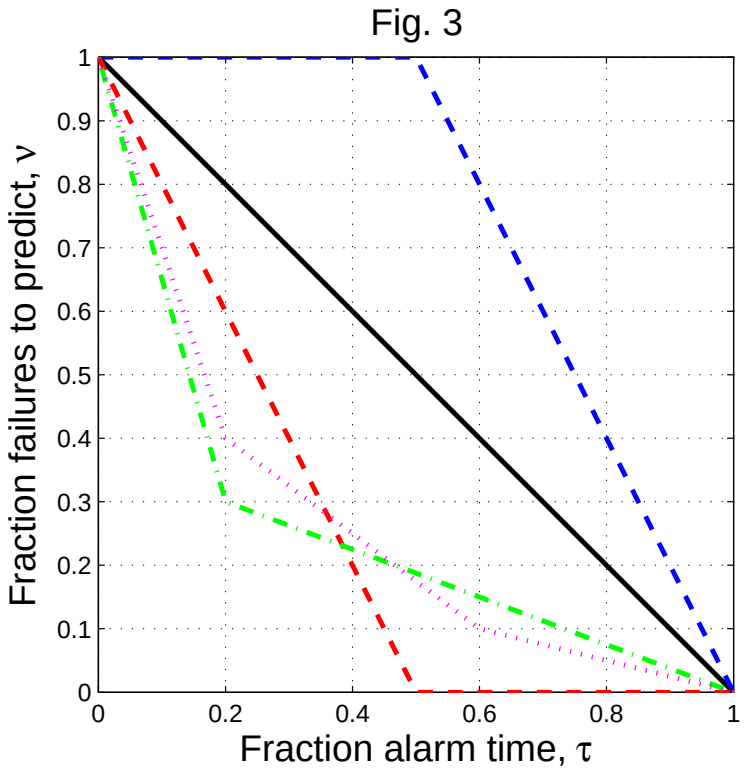

Figure 3: Error diagram example.

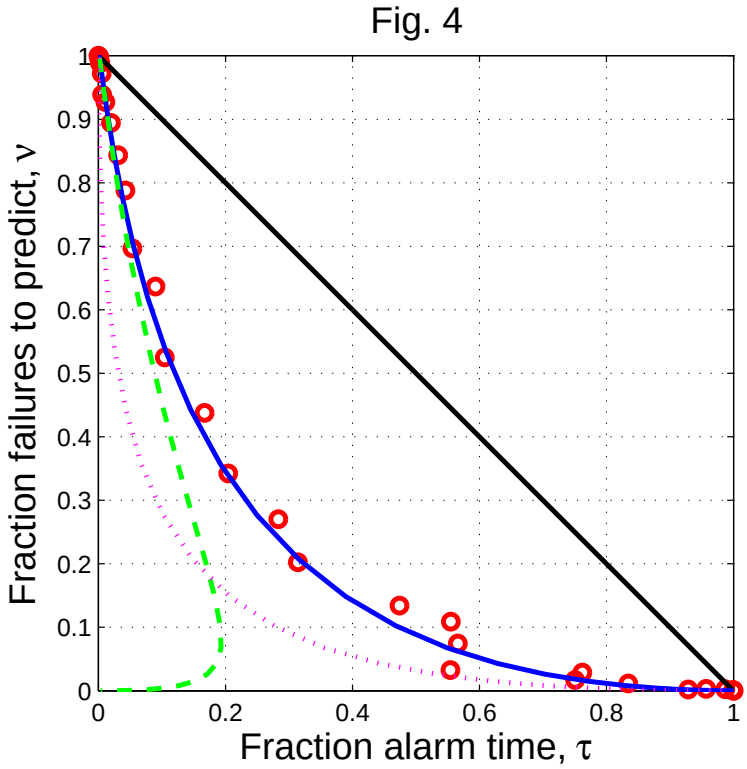

Figure 4: Error diagrams for the lognormal renewal process with $\sigma = 1.86$. The straight solid line is the strategy curve corresponding to a random guess. The left solid curve is calculated using (17–18), circles are the result of simulations. The dashed and dotted curves are the first and second right-hand terms in (18), respectively.

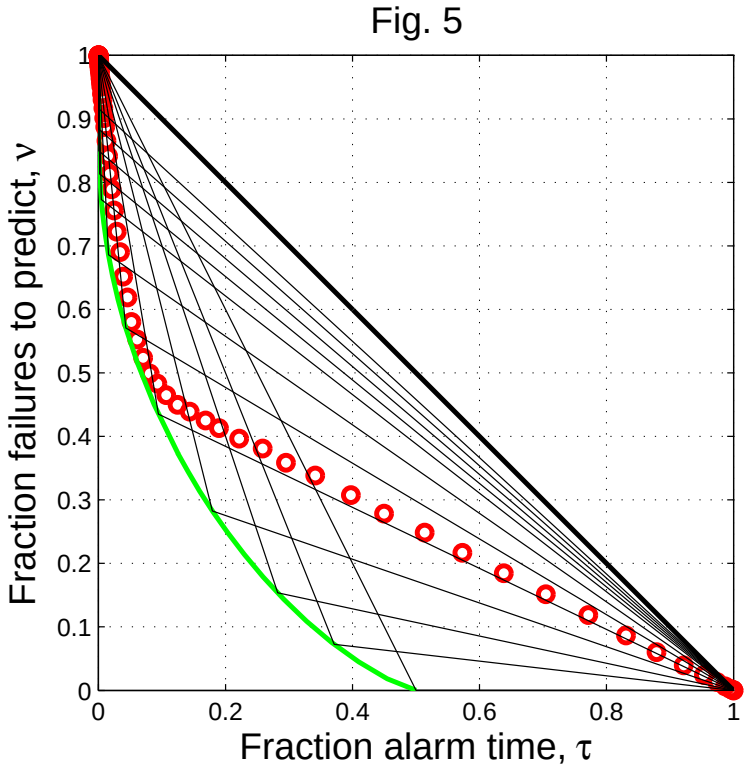

Figure 5: Error diagrams for renewal processes. The thick straight solid line corresponds to a random guess. Thin solid lines are for the curves with the information score 1 bit. The D_1-values (22) for the first segment starting from the right (or the second segment starting from the bottom) are 2, 2.5, 3, 4, 6, 10, 20, 50, 100, 250, 1000, and 10000. The left thick solid line is an envelope curve for these two-segment curves. Circles stand for simulating the Poisson renewal process with two states (see text).

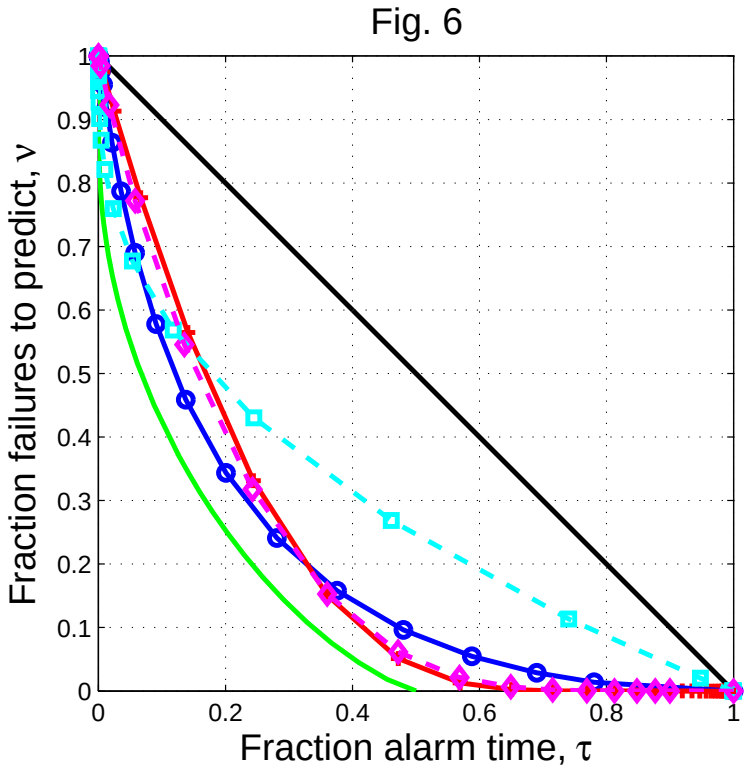

Figure 6: Error diagrams for renewal processes with the information score 1 bit. The straight solid line is the diagram curve corresponding to a random guess. The left solid line is an envelope curve for two-segment curves. Dashed curves with squares and with diamond signs are for the gamma distribution with $\kappa = 0.329$, and $\kappa = 8.53$, respectively. Solid curves with circles and with plus signs are for lognormal distribution with $\sigma = 1.86$, and $\sigma = 0.35$, respectively.

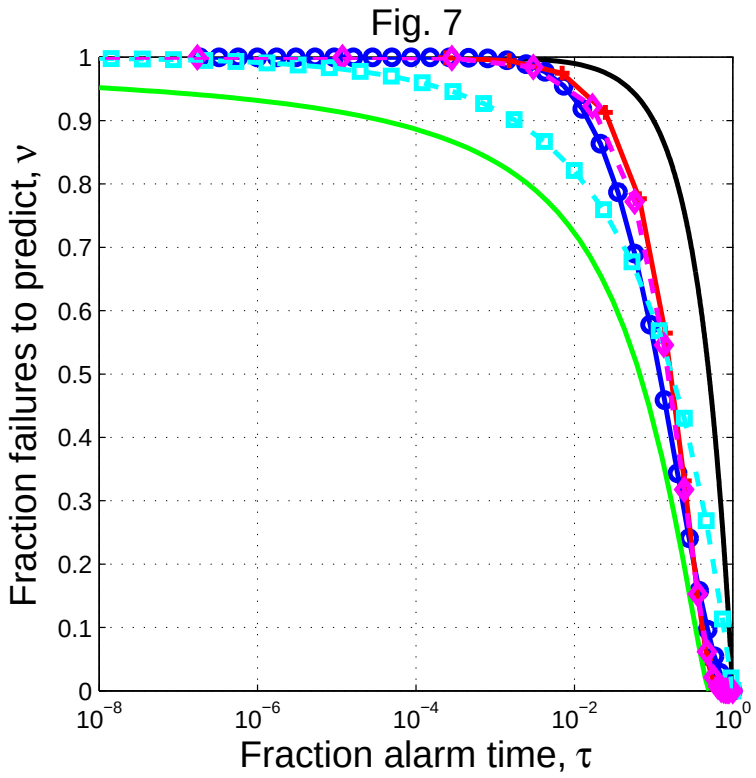

Figure 7: Same as Fig. 6 in semi-log format.

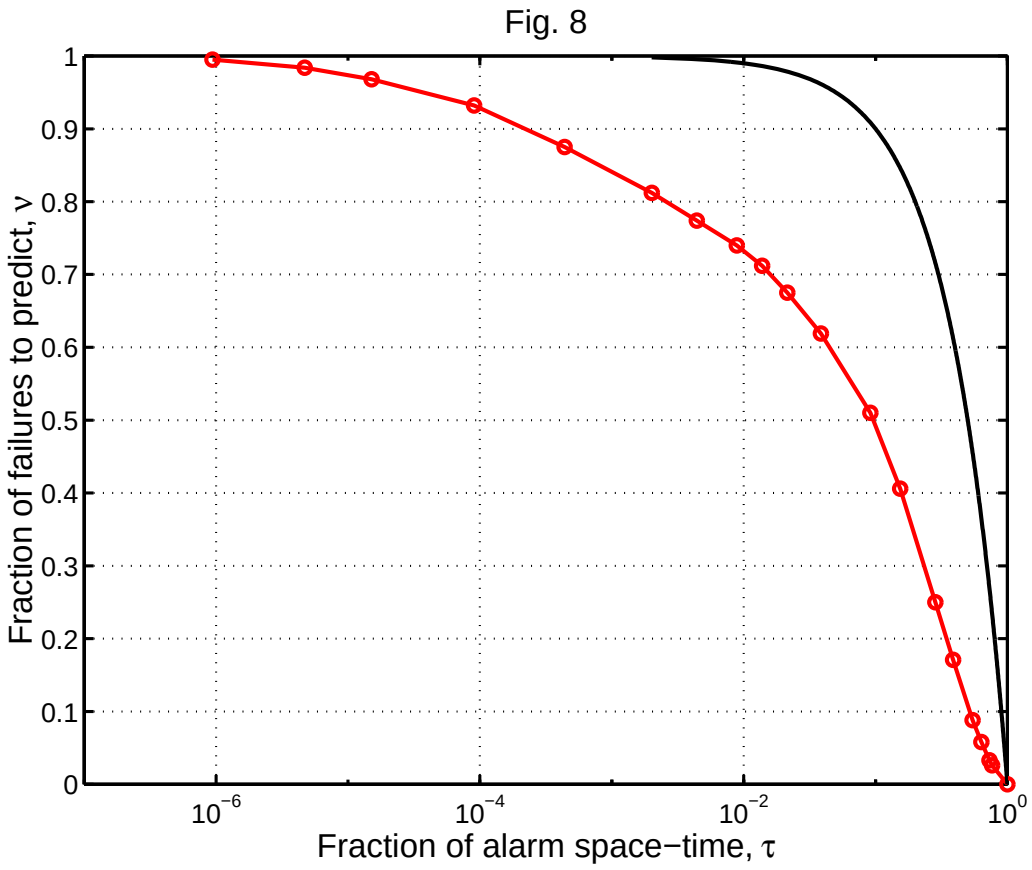

Figure 8: Error diagram (τ, ν): solid line – the strategy of random guess, solid line with asterisks – the error curve for the short-term prediction algorithm by Kagan and Knopoff (1987), as applied to seismicity of central California in the years 1971-1977.