Physics of the Earth and Planetary Interiors, 61 (1990) 84–98
Elsevier Science Publishers B.V., Amsterdam – Printed in The Netherlands

Strategies in strong earthquake prediction

G.M. Molchan

International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Academy of Sciences of the U.S.S.R., Warshavskoye sh. 79, k.2, Moscow 113556 (U.S.S.R.)

Molchan, G.M., 1990. Strategies in strong earthquake prediction. Phys. Earth Planet. Inter., 61: 84–98.

The prediction problem of a stochastic point process in terms of loss function γ is considered. Losses depend on prediction errors: $\hat{n}$, the fraction of failures-to-predict; and $\hat{\tau}$, the fraction of alarm time. The structure of γ-optimal prediction strategies for some functions γ is described. This allows analysis of the announced prediction of strong earthquakes in the southern part of the San Andreas fault. We consider $(\hat{n}, \hat{\tau})$ as a property of prediction strategy and set problems on the description of 'prediction capability' and comparison of prediction algorithms.

In addition, the distribution of empirical prediction errors for the renewal process model and the type of attenuation of inter-event time distribution, are considered.

1. Introduction

In earthquake prediction the following problems have been raised and studied: the search for precursors, their selection and algorithmic synthesis, theoretical justification of precursors, and statistical estimate of prediction significance.

In this paper one more problem is accentuated, prediction goals, and hence the choice of the most effective strategy. The problem is important as it deals with formulation of the prediction problem itself. If prediction goals are fuzzy (which is the case at present), then practically all the suggested prediction strategies are incomparable and, consequently, are equally valuable.

The problem is that there are at least two significant types of errors in earthquake prediction: the fraction of failures-to-predict $\hat{n}$ and the fraction of space–time alarm $\hat{\tau}$. These errors were initially used for qualitative confirmation of the significance of prediction as related to Bernoulli trials with success probability $\hat{\tau}$. However, if we take into account the stochastic nature of earthquakes, then long-term errors $(\hat{n}, \hat{\tau})$ will become specifications of prediction strategy. However, their vectorial nature presents a difficulty in comparison of strategies. It should be pointed out that the choice of prediction algorithm does not solve the problem. Various users choose different algorithmic parameters (in addition to the predicted magnitude and the region boundary), which cause a multiplicity of strategies.

The theoretical uncertainty in comparison of strategies disappears after the hypothetical loss function $\gamma = f(\hat{n}, \hat{\tau})$ is introduced. A similar case is observed in the statistical theory of testing hypotheses. We shall describe below the structure of γ-optimal prediction strategies for any one-dimensional point process and for certain loss functions which have their analogues in testing hypotheses. This approach serves as a basis for analysis (see Section 3.2) of predictions on the San Andreas fault made by the Working Group on California Earthquake Probabilities (WGCEP; 1988). Finally, the problem of 'prediction capability' description of prediction algorithms is discussed.

Two problems stand out. The first is concerned with asymptotics of mean values, dispersions and covariations for prediction results $[\nu_\pi(t), \tau_\pi(t)]$ within the length interval t, where $\nu_\pi(t)$ is the number of successes and $\tau_\pi(t)$ is total alarm time for strategy π. In Section 3.3, the problem is solved for processes with independent intervals

0031-9201/90/$03.50

(renewal processes); here the general result of Smith (1955) is refined. The results obtained can be useful in comparative estimation of the strength of a real prediction, when the renewal process is taken as a reference point.

The second problem refers to the asymptotics of the inter-event interval distribution $F(x)$. Kagan and Knopoff (1987b) argued in favour of non-trivial self-similarity $F(x)$ in the domain of main values x. In Section 4 it is shown that this property disappears with natural limitations and asymptotics $F'(x)$ has ordinary exponential character.

2. Prediction problem

2.1. *Strategies*

Let $\{t_i\}$ be a time sequence of strong events in region G. We shall solve the problem of predicting events t_i using only the catalog $\{t_n\}$ without additional data on weak seismicity. Formalizing the situation, we assume that (Condition A) $\{t_n\}$ is a point process on the whole axis $(-\infty, +\infty)$ and inter-event intervals $\tau_i = t_i - t_{i-1}$ form an ergodic stationary sequence with the bounded average $E\tau < \infty$ and $\Pr\{\tau = 0\} = 0$.

Let

$$\mathscr{F}(t) = \left\{\tau_i : t_i \leqslant t;\ \tau_- = \min_{t_i \leqslant t}(t - t_i)\right\} \tag{1}$$

be the history of the point process up to the moment t (strictly speaking, σ-algebra of events generated by the denoted random values). The value τ determines the time elapsed since the last event up to the moment t. For convenience, let us adopt $\mathscr{F}_i$ for $\mathscr{F}(t_i) = \{\tau_n : n \leqslant i\}$.

Real prediction of the next event t_{i+1} is carried out in the following way: the observer regularly at a certain time-step makes a decision (which depends on the data received) on whether to prolong the alarm or to cancel it (Gabrielov et al., 1990). In the prediction considered by $\mathscr{F}(t)$ the rule of the alarm can be determined a priori at a moment $t_1 + 0$. As a result, this leads to the choice of a time set alarm $\Delta_{i+1} \subset (0, \infty)$ which can depend on the past $\mathscr{F}_i$. The event t_{i+1} is predicted if

$$t_{i+1} \in t_i + \Delta_{i+1}$$

At a moment t_{i+1} the alarm is called off and the current prediction cycle is terminated. The measure

$$\tau(\Delta_{i+1}) = \operatorname{mes}[(0, \tau_{i+1}) \cap \Delta_{i+1}]$$

defines the alarm time in the described cycle, and the indicator-function

$$\chi_{\Delta_{i+1}}(\tau_{i+1}) = \begin{cases} 1 & \tau_{i+1} \in \Delta_{i+1} \text{ (success)} \\ 0 & \tau_{i+1} \in \Delta^c_{i+1} \text{ (failure)} \end{cases}$$

defines the outcome of prediction. Here Δ^c is a complement of Δ on $(0, \infty)$.

Let $\{\Delta_i\}$ be called a 'pure' prediction strategy. If elements $\{\Delta_i\}$ are random, i.e. Δ_{i+1} is given with a certain conditional distribution of probabilities $\pi(\cdot \mid \omega_-)$, where ω_- is the sample of the past $\mathscr{F}_i$, we shall call such a strategy $\pi = \{\pi_i\}$ a 'mixed' strategy. If weak seismicity does not give information on strong events then the above-described real strategy is equivalent to the 'mixed' strategy. 'Mixed' strategies are interesting a priori if different regimes with random transitions are possible in the seismic process. It is then necessary to guess the regime and choose a corresponding pure strategy.

It is natural to consider only stationary strategies $\pi = \{\pi_i\}$ for a stationary point process. The measure π_{i+1} then depends only on the sample of the past $\{\tau_i, \tau_{i-1}, \ldots\}$ and does not depend on the number of a prediction cycle. Formally, $\pi_{i+1} = T\pi_i$, where T is the shift operator on $\{\tau_i\} : T\{\tau_i\} = \{\tau_{i+1}\}$. It is clear that the description of stationary strategies is reduced to that of prediction of one moment of time, t_1. Thus we shall often identify the term 'strategy' with Δ or the measure $\pi(\cdot \mid \mathscr{F}_0)$.

We now introduce two empirical prediction errors of N events $\{t_1, \ldots, t_n\}$ for strategy $\pi = \{\pi_i\}$ the empirical fraction of successes

$$\hat{\hat{n}}_\pi = 1/N \sum_{0 \leqslant i \leqslant N} \chi_{\Delta^c_{i+1}}(\tau_{i+1}) \tag{2}$$

where $\{\Delta_i\}$ is the sample of $\{\pi_i\}$ and Δ^c is a complement of Δ on the semi-axis $(0, \infty)$; and the empirical fraction of the alarm time

$$\hat{\hat{\tau}}_\pi = 1/N \sum_{0 \leqslant i \leqslant N} \tau(\Delta_{i+1}) \Big/ \left(1/N \sum_{0 \leqslant i \leqslant N} \tau_i\right) \tag{3}$$

For stationary strategies, the elements of each of the three sums in eqns. (2) and (3) form stationary sequences with the finite average value

$$\chi_\Delta(\tau) \leqslant 1,\ \tau(\Delta) \leqslant \tau,\ E\tau < \infty$$

Thus the individual ergodic theorem (Billingsley, 1965) can be applied to these sums. As a result we have:

2.1.1. Statement 1

Let π be an arbitrary stationary strategy. Then under condition A with probability 1

$$\hat{\mathring{n}}_\pi \to \mathring{n}_\pi = E\chi_{\Delta^c}(\tau_1),\ N \to \infty \tag{4}$$

$$\hat{\mathring{\tau}}_\pi \to \mathring{\tau}_\pi = E\tau(\Delta_1)/E\tau_1,\ N \to \infty \tag{5}$$

where $\Delta = \Delta(\omega_-)$, E is the expectation, i.e., averaging in (ω_-, τ_1) or in $(\omega_-, \tau_1, \Delta)$ for a pure or mixed strategy respectively.

The prehistory of process $\mathscr{F}_0$ can be a significant part of the prediction and its empirical errors. It is very important that limit prediction errors do not depend on initial information $\mathscr{F}_0$. Thus $(\mathring{n}, \mathring{\tau})_\pi$ is a characteric of strategy π. This justifies the application of Bayesian methods of prediction which rely on a priori representation of the past.

2.2. Strategies compared

Let α_G be the loss for one failure-to-predict in a given region, and β_G be the loss for a unit of alarm time in the same region (subscript G specifies the region).

It is known that $\lambda_G = 1/E\tau_G$ determines the average number of events per time unit for a stationary point process. Thus $\alpha_G\lambda_G\mathring{n}_\pi$ defines long-term losses for failures-to-predict in region G related to the time unit. However, it also follows from statement 1 that

$$\left(N\hat{\mathring{n}}_\pi\right)\alpha_G / \sum_{0 \leqslant i \leqslant N} \tau_i \to \mathring{n}\alpha_G\lambda$$

The value of $\beta_G\mathring{\tau}_\pi$ determines the average losses per unit time connected with an alarm given in region G.

To arrange prediction strategies let us use some loss functions $\gamma = f(\mathring{n}, \mathring{\tau})$:

(a) long-term losses per time unit

$$\gamma_a = \alpha_G\lambda\mathring{n}_\pi + \beta_G\mathring{\tau}_\pi \tag{6}$$

(b) fraction of failures-to-predict $\mathring{n}_\pi$ or losses

$$\gamma_b = \alpha_G\lambda\mathring{n}_\pi$$

caused by them for strategies with the given level of the error

$$\mathring{\tau}_\pi \leqslant \epsilon$$

(c) fraction of the alarm time $\mathring{\tau}$ or losses

$$\gamma_c = \beta_G\mathring{\tau}_\pi$$

related to the alarm for strategies with the given level of the error $\mathring{n}_\pi$; $\mathring{n}_\pi \leqslant \epsilon$;

(d) $\gamma_d = \max(\mathring{n}_\pi, \mathring{\tau}_\pi)$

Let us say that strategy π_1 is better than π_2 relatively to γ if $\gamma_{\pi_1} < \gamma_{\pi_2}$. Strategy π is γ-optimal if it is the best one. Let the γ_d-optimal strategy be called a minimax strategy. In cases (b) and (c), strategies are comparable if they satisfy the same limits.

Let $F(x \mid \omega_-) = P\{\tau_1 < x \mid \mathscr{F}_0\}$ be the distribution of τ_1 under condition $\mathscr{F}_0$. To simplify the description of γ-optimal strategies we will introduce one condition of regularity on function $F(x \mid \omega_-)$. (Condition B): $\mathscr{F}(x \mid \omega_-)$ has density and levels of the hazard function

$$r(x \mid \omega_-) = F'(x \mid \omega_-)/[1 - F(x \mid \omega_-)] \tag{7}$$

have zero $\mathrm{d}F(x \mid \omega_-)$ − measure for almost all ω_-.

Condition B prohibits exponential sections for decreasing function $\bar{F}(x \mid \omega_-) = 1 - \mathrm{F}(x \mid \omega_-)$, i.e. at any interval I

$$\bar{F}(x) \neq a_I \exp(-b_I x),\ x \in I.$$

The value of $r(x \mid \omega_-)\,\mathrm{d}x$ corresponds to the probability of the event $\{t_0 + x < t_1 < t_0 + x + \mathrm{d}x\}$ on condition that the prehistory is known up to the moment $t = t_0 + x$, i.e. $t_1 - t_0 \geqslant x$.

2.2.1. Statement 2

Suppose that $\gamma = \gamma_i$, $i = a, b, c, d$ and the past $\mathscr{F}_0$ is known. Then, under condition B, γ-optimal strategies of prediction τ_1 are pure; the alarm set is defined by

$$\Delta(u_i) = \{x : r(x \mid \omega_-) > u_i\},\ i = a, b, c, d \tag{8}$$

where threshold $u_a = \beta_G/\alpha_G$; in the remaining cases, thresholds are defined by the following

equations:

$$\text{(b)}\quad E\int_{\Delta(u)}[1-F(x\mid\omega_-)]\,dx=\epsilon m \tag{9}$$

$$\text{(c)}\quad E\int_{\Delta(u)}dF(x\mid\omega_-)=1-\epsilon \tag{10}$$

The equation for threshold u_d is derived from the last two by exclusion of parameter ϵ.

There are equivalent γ_i-optimal strategies for $i=b,\ c$:

$\pi(u_i)=\{x: r(x\mid\omega_-)>u_i(\omega_-)\}$

where u_i are defined by the equations

$$\int_{\pi(u_b)}[1-F(x\mid\omega_-)]\,dx=\epsilon m(\omega_-)$$

and

$$\int_{\pi(u_c)}dF(x\mid\omega_-)=1-\epsilon$$

(Here $m(\omega_-)=E\{\tau_1\mid\mathscr{F}_0\}$ is the conditional average τ_1).

2.2.2. Proof

The proof of the statement is reduced to establishing analogy between the prediction problem and the problem of testing simple statistical hypotheses. It may therefore be of supplementary interest. Let $t_0=0$, $\Delta=\Delta(\omega_-)$ be the alarm set of pure strategy π, and ω_- be a realization of the past $\{t_i,\ i\leqslant 0\}$. According to (4)

$$\mathring{n}_\pi=E\chi_{\Delta^c}(\tau_1)=E\big(E\chi_{\Delta^c}(\tau_1)\mid\mathscr{F}_0\big)$$

$$=E\int_{\Delta^c(\omega_-)}dF(x\mid\omega_-)$$

or

$$\mathring{n}_\pi=\int_{D^c}dF(x\mid\omega_-)\mathscr{P}(d\omega_-)=\int_{D^c}d\Phi(x,\ \omega_-) \tag{11}$$

where $D=\{(x,\ \omega_-): x\in\Delta(\omega_-)\}$ and $\mathscr{P}(d\omega_-)$ is a measure corresponding to the point process in the interval $t\leqslant t_0$. Similarly,

$$\mathring{\tau}_\pi=\lambda E\int_0^{\tau_1}\chi_\Delta(u)\,du$$

$$=\lambda E\int_0^\infty\int_0^x\chi_\Delta(u)\,du\,dF(x\mid\omega_-)$$

$$=\lambda E\int_\Delta(1-F(x\mid\omega_-))\,dx$$

or

$$\mathring{\tau}_\pi=\lambda\int_D[1-F(x\mid\omega_-)]\mathscr{P}(d\omega_-)=\int_D d\Phi_1(x,\ \omega_-) \tag{12}$$

It is clear that $d\Phi$ and $d\Phi_1$ are probabilistic measures.

For example,

$$\int d\Phi_1=\lambda E\int_0^\infty[1-F_1(x\mid\omega_-]\,dx$$

$$=\lambda EE(\tau_1\mid\mathscr{F}_0)=\lambda E\tau_1=1$$

Relationships (11) and (12) allow us to consider the problem of minimization of γ_i, where $i=a$–d, from the viewpoint of testing statistical hypotheses. There is one observation $(x,\ \omega_-)$ which should be related to one of two distributions: Φ (hypothesis H_0) or Φ_1 (hypothesis H_1). Let D be the area of $(x,\ \omega_-)$ for accepting H_0 and let its complement D^c be the area for accepting hypothesis H_1. Then, according to eqns. (11) and (12), $\mathring{n}_\pi$ and $\mathring{\tau}_\pi$ will define errors of two kinds. In turn, the minimization problems γ_b or γ_c by means of pure strategies become equivalent to Neyman–Pearson problems when one of the errors is fixed and the other is optimized by the choice of the critical set D^c.

The loss function γ_a differs only in constant normalization from the loss function typical for the Bayesian approach to hypothesis testing. Indeed, let us assume that

$$p_0=\alpha_G\lambda/L,\ p_1=\beta_G/L$$

are a priori distributions of hypotheses H_0 and H_1, $L=\alpha_G\lambda+\beta_G$. Then the normalized loss function

$$\gamma_a/L=p_0\mathring{n}_\pi+p_1\mathring{\tau}_\pi$$

will define the average error in testing the hypothesis H_0 against H_1 with the critical set D^c.

Finally, the problem of minimization of γ_d is dual to the minimax approach in testing hypotheses: the problem is to find D with $\max(\mathring{n}_\pi,\ \mathring{\tau}_\pi)=\min$.

The established duality of the problems allows us to use the corresponding results of the theory of hypothesis testing (Borovkov, 1984; p. 281). In all

the above problems the optimal criteria are based on the likelihood ratio

$$\frac{d\Phi(x, \omega_-)}{d\Phi_1(x, \omega_-)} = \frac{1}{\lambda} \frac{F'(x \mid \omega_-)\mathscr{P}(d\omega_-)}{1 - F(x \mid \omega_-)\mathscr{P}(d\omega_-)}$$

$$= \lambda^{-1} r(x \mid \omega_-)$$

In this case γ_i-optimal solutions to the optimization problems $i = a$–d should be looked for in the class of the following criteria:

$$D = \{(x, \omega_-) : d\phi/d\Phi_1 > u\}$$

if the Φ measure of sets $\{(x, \omega_-)$: $d\phi/d\Phi_1 = u\}$ is zero for any u (condition B). For example, for γ_a

$$\Delta_{\text{opt}} = \{x: d\phi/d\Phi_1 > p_1/p_0\}$$

$$= \{x : r(x \mid \omega_-) > \beta_G/\alpha_G\}$$

By analogy, we obtain eqns. (8)–(10) for the threshold u. The thresholds u_i, $i = b$, c, d are found from equations $\mathring{\tau}_\pi = \epsilon$ for b, $\mathring{n}_\pi = \epsilon$ for c and $\mathring{\tau}_\pi = \mathring{n}_\pi$ for d.

The critical sets (8) are γ_i-optimal in testing hypotheses among all randomized decision functions $\{\varphi\}$ which can accept hypothesis H_0 with probability $p = p(x, \omega_-)$, $p \in [0.1]$. It is clear that class $\{\varphi\}$ contains mixed strategies of the prediction problem. Hence, γ_i-optimal strategies are pure.

Let us consider strategies of the kind $\Delta(u) = \{x : r(x \mid \omega_-) > u(\omega_-)\}$ with random thresholds u. If the threshold is chosen from condition

$$\int_{\Delta(u)} dF(x \mid \omega_-) = 1 - \epsilon$$

the strategy $\Delta(u)$ will obey the condition $\mathring{n}_\pi = \epsilon$ and will minimize the values

$$\epsilon_1(\omega_-) = \int_{\Delta(u)} [1 - F(x \mid \omega_-)]/m(\omega_-)\, dx$$

where

$$m(\omega_-) = \int_0^\infty [1 - F(x \mid \omega_-)]\, dx$$

However, the error

$$\mathring{\tau}_\Delta = \lambda E \epsilon_1(\omega_-) m(\omega_-)$$

and, therefore, strategy $\Delta(u)$ is γ_c-optimal. Similarly, it can be shown that equation $\epsilon_1(\omega_-) = \epsilon$ determines the threshold $u(\omega_-)$ for γ_b-optimal strategy.

2.2.3. Comments

(1) In fact, strategies such as (8) are γ-optimal for any continuous convex function $\gamma = f(\mathring{n}, \mathring{\tau})$, which increases on each ray $\mathring{n} = k\mathring{\tau}$, $k \geqslant 0$.

(2) Function $r(t \mid \omega_-)$ is the optimal short-term prediction (Vere-Jones, 1978; Kagan and Knopoff, 1977, 1987a): $r(x \mid \omega)\, dx$ is the best unbiased estimator for the number of events in the interval dt, $t = t_0 + x$, under the condition $\mathscr{F}(t)$. Statement 2 gives the optimality of $r(t \mid \omega_-)$ on the whole for prediction which does not use weak seismicity. Levels of r determine long-term predictions. These levels are regulated by one loss function, i.e. by prediction goals.

According to the Kolmogorov–Wiener concept of time series prediction, the long-term predictor of the point process is determined by the Smith renewal function $H(x \mid \omega_-)$

$$H(x \mid \omega_-) = E\{\nu(t_0, t_0 + x)/\mathscr{F}_0\},\ x > 0$$

At the moment t_0, the differential of this function $dH(x \mid \omega_-)$ predicts the number of events in the dx vicinity of each succeeding moment of time.

(3) It is easy to see that if r corresponds to the conditional distribution of τ_{i+1} under the finite past $\omega_-^n = (\tau_j,\ i \geqslant j > i - n)$, then sets

$$\Delta(u) = \{x : r(x \mid \omega_-^n) > u\}$$

preserve γ-optimality among strategies which use the finite past data ω_-^n. It is natural to call such strategies γ-optimal strategies in a weak sense. They are optimal for processes with independent intervals τ_i, τ_j, $|i - j| > n$.

3. The renewal process

3.1. Examples of strategies

Let us consider the simplest stochastically periodic point process which has independent and equally distributed time intervals τ_i, i.e. the renewal process with distribution function $f(x) = \Pr\{\tau < x\}$. It includes both an absolutely random Poisson sequence $F'(x) = \lambda \exp(-\lambda x)$, and a

predictable periodic sequence $F'(x) = \delta(x - x_0)$. These extremities present a range of possibilities in prediction of the renewal processes.

A random renewal process seems to be an appropriate model for rare events in some local catalog (a typical object for a practically interesting prediction), or to reflect the main elements of the 'seismic cycle' for a large tectonic region. At the same time, the well-known mathematical results on superimposition of the renewal processes show (see, for example, Daley and Vere-Jones (1988)) that the total catalog of rare local events can have all the properties of Poisson flow in time. In such a case, events from different places should be in weak dependence.

The spread index is a useful parameter of the process

$$J = \lim_{t \to \infty} \frac{\text{Var } \nu(t)}{E\nu(t)} = \frac{\text{Var } \tau}{(E\tau)^2} \tag{13}$$

where $\nu(t)$ is the number of events in the interval $(0, t)$ and Var is the dispersion. Deviations of J from unity reflect on average the tendency of a point process of 'clustering' ($J > 1$), and to periodicity ($J < 1$, 'scattering'), as related to the Poisson chaos.

The γ-optimal strategies for renewal processes are defined by sets

$$\Delta = \{x : r(x) = F'(x)/[1 - F(x)] > u\}$$

Δ is the interval for the unimodal hazard function $r(x)$. It is unbounded, $\Delta = (x, \infty)$, if $r(x)$ is an increasing function (the so-called property of 'positive ageing'; this term is used in the description of endurance failures in materials). As is known (Cox and Lewis, 1966), in this case $J < 1$, so when events tend to scattering, announcement of an alarm should be delayed. For decreasing function $r(x)$ ('negative ageing'), $\Delta_{\text{opt}} = (0, x)$, i.e. it is preferable to give alarm at the beginning of a prediction cycle if events in the process tend to clustering.

In earthquake prediction there is a tendency to reduce the duration of the alarm. However, there are some exceptions: prediction in Kamchatka, based on seismic gaps (Fedotov et al., 1977), and the method of long-range aftershocks (Prozorov and Rantsman, 1972). Thus, if density $F'(x)$ is unimodal, with a well-expressed peak, its short vicinity seems a natural choice for an alarm set. Generally, such a decision is not correct, as an alarm after the modal point $F'(x)$ will be quickly cancelled if $F'(x)$ decreases sufficiently rapidly. Otherwise, the announcement of unbounded alarm is not always to the detriment of a prediction strategy. Some examples of optimal alarm sets are given below.

3.1.1. Example 1

F is a uniform distribution on $[0, 2m]$. Then

$$r(x) = (2m - x)_+^{-1}, \quad J = 1/3$$

The minimax interval, i.e. the γ_d-optimal strategy is the set $\Delta_{\text{opt}} = (km, \infty)$, where km is the point of 'golden section' of $[0, 2m]$: $k = 3 - \sqrt{5} \simeq 0.76$ (an unexpected optimal property of the 'golden section'!). The errors are $\mathring{n} = \mathring{\tau} \simeq 38\%$.

3.1.2. Example 2

F is a γ-distribution with average m and index J:

$$F'(x) = cx^{1/J} \exp(-x/Jm), \quad x \geqslant 0$$

TABLE 1

Parameters of minimax strategy for γ-distribution

	Strategy							
	$\Delta = \{x : x < km\}$		$\Delta = \{x : x > km\}$					
Spread index $J = \sigma^2/m^2$ [a]	2	1	0.4	0.3	0.2	0.1	0.05	0.02
k	0.71	0.69	0.72	0.73	0.74	0.77	0.81	0.87
Prediction errors $\mathring{n} = \mathring{\tau}$ (%)	40	50	39	36	31	25	21	19

[a] m, average; σ^2, dispersion.

This distribution is unimodal, and is almost Gaussian with $J \ll 1$. The hazard function $r(t)$ is monotonic, thus the γ-optimal alarm sets are $\Delta = (0, x)$ for $J > 1$ and $\Delta = (x, \infty)$ for $J < 1$. $\varkappa = 1$ corresponds to the unpredictable Poisson process; here condition B is incorrect: $r(x) = m^{-1}$, $x \geqslant 0$. Thus the optimal open sets Δ could take any form and contain any point.

Table 1 demonstrates prediction possibilities by means of γ_d-optimal (minimax) strategy for γ-distribution. The minimax strategy has the following form:

$$\Delta = \{x: \operatorname{sgn}(1 - J) \cdot [x - k(J)m] > 0\}$$

Dimensionless parameter $k(J)$ turned out to be unexpectedly stable: over a wide range of index $J \in [0.05,2]$, the value of $K(J) = 3/4 \pm 0.06$.

An example can demonstrate stability not only of the strategy but also of the prediction results. We approximate a γ-distribution by a uniform distribution with the same average and value of J. It is possible for $J = 1/3$. We use the γ_d-optimal strategy $\Delta = (0.76m, \infty)$ in this approximation (see example 1) for prediction in the true model F. The pseudo-optimal strategy leads to errors $\mathring{n}_* = 40\%$ and $\mathring{\tau}_* = 34.6\%$. The average error is $(\mathring{n}_* + \mathring{\tau}_*)/2 = 37\%$, against true minimax errors $\mathring{n} = \mathring{\tau} = 36\%$. The agreement is very good for such a crude approximation of true distribution.

3.2. *Prediction on the San Andreas fault*

Nishenko and Buland (1987) analysed 53 intervals τ_i in 14 regions with high seismicity. Out of three types of distributions $F_G(x)$, Normal, Weibull and Lognormal, preference was given to the third type. The following model is accepted for inter-event intervals for tectonic regions $\{G\}$

$$\tau = m_G \exp\left(v_G \xi - v_G^2/2\right) \tag{14}$$

where ξ is a standard normal random value $N(0, 1)$. $m_G = E\tau_G$ and v_G^2 is the variation of $\ln \tau$.

The spread index for this model is

$$J = \exp\left(v_G^2\right) - 1$$

The WGCEP (1988) applied model (14) for prediction of the strongest earthquakes on the San Andreas fault. Table 2 contains data on eight of the fault segments. Prediction is determined by conditional probabilities

$$R_\delta = P\{\nu(\tau_-, \tau_- + \delta) \geqslant 1 \mid \tau \geqslant \tau_-\}$$
$$= [F(\tau_- + \delta) - F(\tau_-)]/[1 - F(\tau_-)]$$

of occurrence of an event on segments G for forthcoming (after 1988) $\delta = 30$ yr; τ_- is the time since the last event in G. According to the WGCEP calculations the most dangerous of the denoted eight segments is now the Parkfield (P) region, where $R_\delta > 0.9$ against $R_\delta < 0.4$ for the other segments (see Table 2).

Using this method of comparing the fault segments almost always Parkfield is diagnosed as the most dangerous. Actually, for any segment except P, $R > 0.9$ for $0 \leqslant \tau \leqslant 800$ yr. Moreover, the R value in Parkfield can reach the level of 0.5 only in the impossible case of quiescence for 4000 yr.

TABLE 2

Conditional probability R_δ of major earthquakes along segments of the San Andreas fault, 1988–2018

Fault segment	1	2	3	4	5	6	7	k	k [a]	10
San Francisco Peninsula	90	7	1906	196	0.38	0.176	2032	0.75	8.8	0.30
Santa Cruz Mountains	35	6.5	1906	136	0.44	0.185	2007	0.74	8.4	0.30
Parkfield	30	6	1966	21	0.9	0.056	1983	0.81	37	0.19
Cholame	55	7	1857	159	0.3	0.281	1972	0.72	4.8	0.34
Carrizo	145	8	1857	296	0.1	0.137	2082	0.76	13	0.28
Mojave	100	7.5	1857	162	0.3	0.168	1979	0.75	9	0.29
San Bernardino Mountains	100	7.5	1812	198	0.2	0.36	1951	0.7	3.4	0.36
Coachella Valley	100	7.5	1680	256	0.4	0.09	1880	0.78	19	0.25

1, length (km); 2, expected magnitude; 3, date of most recent event; 4, expected recurrence time (yr); 5, R_δ; 6, spread index, J; 7, start of alarm; 8, 9, k and k [a] are time alarm thresholds; 10, Errors $\mathring{n} = \mathring{\tau}$.

[a] Except for the last four columns, the data are from WGCEP, (1988).

Otherwise, the R-test causes a steady alarm in P, but no alarms in other segments.

Let us consider the prediction in model (14) from this viewpoint. The hazard function $r(x)$ of the lognormal distribution is unimodal, with $r(0) = r(\infty) = 0$. Thus optimal alarm sets form finite intervals $\Delta = \{x: km_G < x < k_* m_G\}$. For minimax strategies, normalized thresholds Δ, k and k^* are presented in Table 2. Upper thresholds k^* are determined by the choice of the model $F(x)$ and are practically unlimited. They change within the range 700–5000 yr. That is no longer important because if upper thresholds are substituted for ∞, then prediction errors differ by 0.001 for observed values of $J \leqslant 0.4$. k threshold values and prediction errors $\epsilon = \mathring{n} = \mathring{\tau}$ proved to be stable in the sense of the model choice. Functions $k(J)$ and $\epsilon(J)$ stay almost unchanged when a lognormal distribution was replaced by Weibull or γ distribution, with $\Delta = \{x: x > k(F)\}$ (compare Tables 1 and 2).

The above circumstances justify the choice of a minimax strategy according to which five out of eight fault segments should be in a state of alarm by 1988, in particular, the region P—since 1983—and the Coacella Valley (CV)—since 1880 (!)—although according to R_δ, alarm is impossible in CV. However, as related to the minimax strategy, the concentration of U.S. scientific efforts in region P is justified. According to the minimax strategy, a maximum of successes $\delta m_G^{-1}(1 - \mathring{n})$ for the considered period δ is expected in region P.

The analysis becomes more complicated if we take into account absolute losses for the whole region G, i.e., losses of the type

$$\gamma_a = \alpha_G m_G^{-1} \mathring{n} + \beta_G \mathring{\tau}$$

Concretization of α_G, β_G is not unambiguous. We can suppose that

$$\alpha_G = \alpha_M(I);\ \alpha_G = \tilde{\alpha}_M m_G(II);\ \beta_G = \beta_M |G|$$

where subscript M shows the dependence of parameters on the magnitude of predicted events.

Losses from one failure-to-predict of magnitude M are proportional to area $Q(M)$ of the destruction zone if the destruction objects are homogeneous and uniformly distributed in space. If we account for recurrence frequency $p(M)$ then

$$\alpha_M \propto \int_{\Delta M} Q(M) p(M)\, \mathrm{d}M$$

Usually $Q(M) \propto 10^{b_1 M}$, $p(M) \propto 10^{-bM}$ and $b_1 \simeq b$. Thus $\alpha_M \propto |\Delta M|$.

Evidently, model $\alpha_G - \mathrm{II}$ can occur in insurance. Let the insurance rates be independent of location on the fault. Then insurance payments are proportional to time. However, the period between failures-to-predict is proportional to m_G. Thus we can assume that compensation for an object's destruction is also proportional to m_G. Consequently, $\alpha_G \propto m_G$. However, these arguments do not account for the efficiency of investments.

For methodical purposes we assume $\alpha_M = \alpha = \tilde{\alpha}_M$, $\beta_M = \beta$; $k = \beta/\alpha$. The minimum of loss function γ_a is determined by the value

$$\gamma_* = \alpha_G m_G^{-1} \left\{ 1 - \int_0^\infty [r_0(u) - k m_G |G|]_+ \times [1 - F_0(u)] \right\} \mathrm{d}u$$

where $x_+ = (x + |x|)/2$, F_0 is the distribution of τ with $E\tau = 1$, i.e.

$$F_0(x) = \Phi[\ln(x/v) + (v/2)]$$

$\Phi(x)$ is a standard Gaussian distribution, and $r_0(x)$ is a hazard function for $F_0(x)$. Value γ_* is an increasing function of parameter k:

$$\frac{\mathrm{d}}{\mathrm{d}k}\gamma_* = \alpha |G| \mathring{\tau} \geqslant 0 \tag{15}$$

Moreover,

$$\gamma_* = \alpha_G^* \begin{cases} k|G|[1 + o(1)] & k = \beta/\alpha \ll 1 \\ 1/m_G & k = \beta/\alpha \geqslant r_0^* / (m_G |G|) \end{cases} \tag{16}$$

where

$$r_0^* = \max r_0(x)$$

Hence we can obtain a qualitative conclusion of optimal losses, if we use real data on the segments' dimensions and periods m_G (see Table 2). We

summarize our conclusions as follows. Under condition $\alpha_G = \alpha$, in the Parkfield region expected losses are minimal if $\beta/\alpha < k_1$ and maximal if $\beta/\alpha > k_2 (k_1 < k_2)$. Under other conditions $\alpha_G = \alpha m_G$, losses in Parkfield are not greater than in other regions with any α/β.

For both models, when $\alpha_G = \alpha$ the simplest 'optimistic' strategy (no alarm) is γ_a-optimal if $\beta/\alpha \geqslant \max r_0^*/|G|m_G$. However, the simplest pessimistic strategy (endless alarm) cannot be γ_a-optimal in any segment if $\beta > 0$.

3.3. *Distribution of prediction results*

Prediction, based on the renewal process model, can serve as a standard for comparing prediction strategies. Thus it is necessary to know the distribution of prediction results (the number of successes and length of the alarm time) for the pure strategy Δ, which is applied to the renewal process in the interval $(t_0, t_0 + t) = J_t$. This problem is decided for $t \gg 1$ under the condition that t_0 is an event of the point process. However, the final point, $t_0 + t$, should not necessarily coincide with process events. This means that the alarm is interrupted by the final point as well. Thus the number of events on the interval ν_t is becoming random, which complicates the solution to the above problem.

The following asymptotics for $H(t) = E\nu_t$, $t \gg 1$ are used below (Smith, 1955):

$$H(t) = (t + \mu_1)/m_1 + o(1), \ \mu_1 = m_2/2m_1 = E\tau_- \tag{17}$$

where $m_k = E\tau^k$. It is supposed here and in further analysis that $m_2 < \infty$ and that the measure dF is not concentrated on the lattice $\{h_0 + kh_-, k = 0, \pm 1, \ldots.\}$.

Statement 3

Let $\nu_\Delta(t)$ be the number of successes and $\tau_\Delta(t)$ be the alarm time in the renewal process prediction for the interval J_t with the help of the strategy defined by alarm set Δ. Then

(a) vector $[\nu_\Delta(t), \tau_\Delta(t)] = W_t$ is asymptotically normal with $t \to \infty$

(b) its average has the asymptotic:

$$\begin{aligned} E\nu_\Delta(t) &= H(t)E\chi_\Delta(\tau) - E\tau\chi_\Delta(\tau)/m_1 + o(1) \\ E\tau_\Delta(t) &= H(t)E\Delta(\tau) - E\tau_-\chi_\Delta(\tau_-) + o(1) \end{aligned} \tag{18}$$

where τ_- (see (1)) is a random variable with distribution density $f_1(x) = [1 - F(x)]/m_1$ and moments $\mu_k = m_{k+1}[(k+1)m_1]^{-1}$; $\Delta(x) = \int_0^x \chi_\Delta(u)\, du$.

If $m_3 < \infty$ then the covariation matrix of vector W_t has the form

$$\begin{aligned} &\operatorname{cov}[\nu_\Delta(t), \tau_\Delta(t)] \\ &\quad = H(t) \operatorname{cov}\Big[\chi_\Delta(\tau) - \frac{\tau}{m_1}E\chi_\Delta(\tau), \\ &\qquad \Delta(\tau) - \frac{\tau}{m_1}E\Delta(\tau)\Big] + C + o(1) \end{aligned} \tag{19}$$

The elements of the constant matrix C are as follows.

$$\begin{aligned} c_{00} &= (m_{0\Delta})^2\left[\tfrac{3}{4}m_2^2/m_1^4 - \tfrac{2}{3}m_3/m_1^3\right] \\ &\quad + 2\frac{m_{0\Delta}}{m_1^2}(m_{2\Delta} - m_2 m_{1\Delta}/m_1) \\ &\quad + m_{1\Delta}/m_1(m_{1\Delta}/m_1 - 1) \\ c_{11} &= 2\mu_{0\Delta}(\mu_{2\Delta} - \mu_2\mu_{0\Delta}) + (\mu_{1\Delta} - \mu_1\mu_{0\Delta})^2 \\ &\quad + 2(\mu_{1\Delta} - \mu_1\mu_{0\Delta})E[\Delta(\tau_-) - \mu_{0\Delta}\tau_-] \\ &\quad - 2E\tau_-\Delta(\tau_-)[\chi_\Delta(\tau_-) - \mu_{0\Delta}] \\ c_{01} &= 1/m_1(m_{0\Delta}\mu_1 - m_{1\Delta}) \\ &\quad \times [2\mu_{0\Delta}\mu_1 - \mu_{1\Delta} - E\Delta(\tau_-)] \\ &\quad + m_{0\Delta}/m_1(\mu_{2\Delta} - \mu_2\mu_{0\Delta}) \\ &\quad - m_{0\Delta}/m_1\mu_1(\mu_{1\Delta} - \mu_1\mu_{0\Delta}) \\ &\quad + \mu_{0\Delta}/m_1[m_{2\Delta} - E\Delta(\tau)\tau\chi_\Delta(\tau)] \\ &\quad + m_{0\Delta}/m_1[E\Delta(\tau_-)\tau_- - \mu_{0\Delta}\mu_2] \end{aligned}$$

where $m_{k\Delta} = E\tau^k\chi_\Delta(\tau)$ and $\mu_{k\Delta} = E\tau_-^k\chi_\Delta(\tau_-)$.

The proof of the statement can be found in Appendix A. The main terms of the asymptotics (18) and (19) and the fact that vector $W(t) = [\nu_\Delta(t), \tau_\Delta(t)]$ is asymptotically normal proceeds from the common results of Smith (1955) for accumulation processes. Vector $W(t)$ belongs to such processes. The asymptotic results of Smith are insufficient for real catalogs, thus they have

been refined up to the order of o(1) when $t \to \infty$. Refinement of the asymptotic for average values (18) seems to be of greatest practical interest. This is proved by calculations for the renewal process with a γ-distribution parameterized by average m and index J (see example 2). A strategy of the kind $\Delta = \{x: x > 3/4\ m\}$ and observation interval $t = 10m$, allowing on average for 10–11 events, were considered. When $J \in [0.05, 0.5]$, corrections c_{ii} to the main terms of the asymptotics (18) made up $-(9-14)\%$, and in (19) they were $\pm 1.5\%$ for ν_Δ and $-(4-7)\%$ for τ_Δ.

3.4. *The example of an artificial catalog*

Gabrielov (1984) described the block model of lithosphere dynamics, and an earthquake catalog was generated (Gabrielov et al., 1990) in the form juxtaposed with the Californian one. The catalog contains 92 events with $M \geqslant 6.4$ for a period of 720 yr.

Parameters of distribution function $F(x)$ in (18) and (19) have been estimated by means of an empirical unsmoothed estimate of function F for this catalog. The minimax strategy is used for prediction, with an alarm set $\Delta = (6,17)$ yr $= (0.76m, 2.2m)$. It was applied to three intervals with a length of 240 yr. The normalized values of successes

$$\nu_{\text{norm}} = [\nu_\Delta(t) - E\nu_\Delta(t)]/\sqrt{[\text{Var}\ \nu_\Delta(t)]}$$

were -0.94, 0.55 and 0.92.

Analogous values of the alarm time in the same intervals are $\tau_{\text{norm}} = -0.51$, 1.34 and 0.16. Prediction results $[\nu_\Delta(t), \tau_\Delta(t)]$ are normalized by the parameters of the renewal process, and the results obtained do not contradict them. This is an argument in favour of the weak dependence of recurrence intervals τ_i in Gabrielov's model.

The minimax strategy described above and the 'best' prediction strategy obtained by means of algorithm CN (Keilis-Borok and Rotwain, 1990) has the same average prediction error: $(\mathring{n} + \mathring{\tau})/2 = 40\%$ for the artificial catalog. It is interesting that these strategies use different information: the first uses information on strong events, and the second uses information on weak seismicity. Apparently these weak prediction results again demonstrate chaotic behavior in Gabrielov's model.

3.5. *Stability of strategies*

We considered various models of distribution $F(x)$: uniform, γ-distribution, lognormal and empirical for the artificial catalog. In cases of practical interest the minimax strategy showed unexpected stability. It is close to the strategy of the form

$$\Delta = \{x: x \geqslant 0.76m\},\ J < 1$$

where m is the average and J is the spread index. The threshold $0.76m$ corresponds to the 'golden section' of support for uniform distribution with the average m. Thus the problems of strategy stability and prediction results (see example 2) should be the objects of special study.

4. Tail behavior for recurrence time distribution

There are practically no data on the tail of distribution of τ for strong events. Some attempts have been made to overcome difficulties by means of models. Kagan and Knopoff (1987b) considered an earthquake as an event of 'dumping' of the accumulated stress σ_t in a seismic region when σ_t reaches the critical state σ. The τ-distribution is simulated by a random walk

$$\sigma(t + \Delta) = \sigma(t) + \Delta\xi(t),\ \sigma(0) = x,\ \sigma(t) < \sigma \tag{20}$$

where ξ_t is the Wiener process with average ct and dispersion $\mathrm{d}^2 t$. The non-zero average stress rate $c > \sigma$ can be related to plate tectonics; the diffusion part of $\Delta\sigma$ can be associated with random events of 'loading' or 'dumping' which are realized by creep and by weak seismicity inside/outside the region. When the critical state σ is achieved at a random moment τ, the stress drops to the initial level $\sigma(0)$ and the system starts to function anew.

As is already known (Skorohod, 1964), the described renewal model has the recurrence time distribution

$$F'(t \mid c) = \left[|\sigma - x|/d\sqrt{(2\pi)}\right] t^{-3/2} \times \exp\left\{-\tfrac{1}{2}\left[|\sigma - x|/\sqrt{(t)} - c\sqrt{(t)}\right]^2/d^2\right\},\quad t > 0 \tag{21}$$

The function distribution $F(t\,|0)$ has a self-similar character of attenuation with $t \gg 1$, i.e. $F'(t) \propto t^{-3/2}$. Among distributions of τ with fixed average value $E\tau = (\sigma - x)/c$ the distribution $F(t\,|\,c)$ is the nearest to $F(t\,|0)$ in the entropy distance: $\mathrm{dist}(F, F_0) = \int(\log \mathrm{d}F/\mathrm{d}F_0)\,\mathrm{d}F$. Self-similarity F with exponent $-3/2$ or proximity to this self-similarity are the main arguments in favour of model (20). It turns out that this property disappears when natural limits are set; the accumulated stresses in the region cannot drop below a certain zero threshold. The modified model is described by the relation

$$\sigma(t+\Delta) = \max[0,\ \sigma(t) + \Delta\xi(t)],\quad \sigma(0) = x,\ \sigma(t) < \sigma \tag{22}$$

Models of type (22) occur in the theory of stochastic storage processes. However, the author could not find the τ-distribution for the described case in the literature. The τ-distribution can be derived as follows. Let

$$V(t\,|\,x) = P\{\sigma(s) < \sigma \text{ for all } s \in (0,\ t);\ \sigma(0) = x\} = P(\tau > t)$$

Standard methods of the Markov processes (Skorohod, 1964) show that $V(t, x)$ satisfies the parabolic equation

$$\left(\frac{\partial}{\partial t} - c\frac{\partial}{\partial x} + 1/2d^2\frac{\partial^2}{\partial x^2}\right)V = 0,\ t > 0,\ x \in (0, \sigma)$$

with boundary conditions

$$\frac{\partial}{\partial x}V(t, 0) = 0;\ V(0, x) = 1,\quad x \in (0, \sigma);\ V(t, \sigma) = 0$$

By means of the Laplace transform

$$Z_\lambda(x) = \int_0^\infty \mathrm{e}^{-\lambda t}V(t, x)\,\mathrm{d}t = (1 - E\,\mathrm{e}^{-\lambda\tau})/\lambda$$

the problem is reduced to an ordinary differential equation

$$\lambda Z_\lambda(x) - 1 = \left(cD + \frac{d^2}{2}D^2\right)Z_\lambda(x),\ D = \frac{d}{dx}$$

$$Z_\lambda(\sigma) = 0 = Z_\lambda(0)$$

Hence

$$\Phi_x(\lambda) = E\,\mathrm{e}^{-\lambda\tau} = \frac{\lambda_+ \exp^{(-\lambda_- x)} - \lambda_- \exp^{(-\lambda_+ x)}}{\lambda_+ \exp^{(-\lambda_- \sigma)} - \lambda_- \exp^{(-\lambda_+ \sigma)}} \tag{23}$$

where

$$d\cdot\lambda_\pm = c/d \pm \sqrt{\left[2\lambda + (c/d)^2\right]}$$

Equation (23) defines all τ-moments. It can be inversed according to Jordan's lemma, i.e. the τ-distribution density can be expressed by poles of $\Phi_x(\lambda)$ in the semi-plan $\mathrm{Re}\,\lambda < 0$. As a result we have the following.

Statement 4

(1) In model (22), with $c \geqslant 0$, the distribution density τ has an exponential asymptotic, $F'(t) = 0$ [$\exp^{(-\rho_1 t)}$] To be more precise

$$F'(t) = \sum_{n \geqslant 1} a_n \exp(-\rho_n t) \tag{24}$$

where

$$\rho_n \equiv \rho_n(\sigma) = \begin{cases} 1/2(c/d)^2\left[(x_n/k)^2 + 1\right], & c > 0 \\ 1/2(d/\sigma)^2\pi^2(n - 1/2)^2, & c = 0 \end{cases} \tag{25}$$

Here $k \equiv k(\sigma) = c\sigma/d^2$ and $x_n \equiv x_n(k)$ are positive roots of the equation

$$k \sin x/x = -\cos x,\ k \geqslant 0$$

i.e.

$$x_n = \pi(n - 1/2) + k\cdot O(1/n)$$

Coefficients

$$a_n = \operatorname*{Res}_{\lambda = -\rho_n} \Phi_x(\lambda)$$

In particular, if $c = 0$ then

$$a_n = (-1)^{n-1}(d/\sigma)^2\pi(n - 1/2) \times \cos[\pi(n - 1/2)x/\sigma]$$

Series (24) converges fast because $a_n = \mathrm{O}(n)$ and $\rho_n = \mathrm{O}(n^2)$, $n \to \infty$.

(2) In model (22) there is a useful representation for the recurrence time

$$\tau = \sum \epsilon_n\tau_n \tag{26}$$

where $\{\epsilon_n, \tau_n\}$ are jointly independent random variables,

$$P\{\tau_n > t\} = \exp(-\rho_n t)$$

and $\epsilon_n = 0$ or 1 with probability p_n or $1 - p_n$, where

$$p_n = \rho_n(\sigma)/\rho_n(x)$$

($\rho_n(\circ)$; see (25)). If $x = 0$ then $\epsilon_n \equiv 1$. (Representation (26) corresponds to infinite factorization of the characteristic function τ, based on zeroes and poles $\Phi_x(\lambda)$). In models (20) and (22), where $x = 0$, the distribution of τ is asymptotically normal with $k = c\sigma/d^2 \to \infty$, i.e.

$$P\{\sqrt{(k)}(\tau c/\sigma - 1) < 1\}$$

$$\to (2\pi)^{-1/2} \int_{-\infty}^{t} \exp^{(-u^2/2)}\, du$$

(3) In model (22), where $x = 0$, the average recurrence time is

$$E\tau = \sigma/c\left[1 - (1 - e^{-2k})/2k\right]$$

and the spread index

$$J = 1 - 2\frac{2(k-1)^2 + 1 - (2k+3)\, e^{-2k}}{(2k - 1 + e^{-2k})^2}$$

$$\simeq \frac{4k-5}{(2k+1)^2},\ k \to \infty$$

decreases as the function k from $J(0) = 2/3$ to 0 with $k \to \infty$, i.e. $J \leqslant 2/3$. (In model (20), $x = 0$, $E\tau = \sigma/c$ and $J = k^{-1} \geqslant 0$).

(4) In model (22), where $x = 0$, the hazard function

$$r(t) \simeq \rho_1 - \left|\frac{\rho_1 a_2}{\rho_2 a_1}\right|(\rho_2 - \rho_1) \exp^{[-(\rho_2 - \rho_1)t]},\ t \gg 1$$

i.e. the γ-optimal alarm sets $\Delta = \{t : r(t) > u\}$, and $u \ll \rho_1$ is unbounded. This is not true for model (20).

5. Conclusions

Analysis of predictions on the San Andreas fault reveals the complexity and ambiguity of the problem of the comparative hazard of segments of the fault. At present, critical analysis of WGCEP (1988) predictions in this region is aimed at the accuracy and fit of the applied model of strong events (Davis et al., 1989). However, it is equally important to specify the notion of 'hazard' itself and hence the prediction goals.

This paper describes optimal long-term prediction of a stationary point sequence in terms of the conditional hazard function $r(x/\omega\_)$ and the loss function. This approach is probably new in the theory of point process prediction. The space–time analog of r is known as the likelihood map of anticipated events or the short-term prediction map.

Vere-Jones (1978) identifies the problem of estimating $r(x/\omega\_)$ with the prediction problem itself. Such a solution of the prediction problem is always unstable, as the hazard function is rather sensitive to the fine structure of a seismic regime. The situation becomes more stable when the prediction goals are specified. This is confirmed by examples of the stability of strategies (section 3.5) and of the stability of prediction errors (section 3.1). It is natural to draw an analogy with time series prediction. It is common knowledge that coefficients of the linear predictor (it is the analog of $r(x \mid \omega\_)$) are very unstable when based on inexact correlations of a time series. However, that does not diminish the efficiency of prediction which is aimed at minimization of signal error on average, i.e. of its variation ϵ^2 (see the strict results given by Molchan (1982)). In the present analysis, losses $f(\mathring{n}, \mathring{\tau})$ from prediction errors are associated with ϵ^2.

The efficiency of prediction strategy is often measured by the value $k = (1 - \mathring{n})/\mathring{\tau}$ (Gusev, 1976). If k is related to a separate precursor, 1°g k approximately determines Shannon's quantity of information on the mainshock contained in the observed pattern. It is tempting to follow Aki (1981) and to consider Σ 1°g k as information obtained by several 'independent' precursors realized. In natural models such a value essentially overestimates the information. In fact, there are no non-trivial statistically independent precursors; at best, their false alarms are independent. Commonly, the value k does not reflect the gist of the problem. For example, for the 'optimistic strategy' $\mathring{n} = 1$, $\mathring{\tau} = 0$ and $k = 0/0$. This means that k can take any value in the vicinity of the trivial strategy. Thus the value of k should be applied to patterns when the degree of its informativeness is dis-

cussed. As far as prediction itself is concerned, i.e. the rules of using patterns to achieve certain goals, then its efficiency is defined by the degree of usefulness. This notion is larger than the preceding one is expressed here through a loss function. Utsu (1977) suggested the measurement of prediction quality by long-term prevented damage (in per cent).

The above statement is important for adequate formulation of the problem on the strength of prediction algorithms. Algorithm A usually has parameters p, and their concretization can give various strategies $\{S_A(p),\ p \in \mathscr{P}\}$ with a corresponding set of errors $\{\mathring{n}(p),\ \mathring{\tau}(p),\ p \in \mathscr{P}\} = D$. Usually, only one strategy of the set is taken into account, and is chosen subjectively. However, prediction goals are various and difficult to specify a priori. Thus not one but a number of possible strategies should be considered. Only those strategies which generate a lower boundary Γ_A of set D are important. To be more precise, errors $(\mathring{n},\ \mathring{\tau}) \in \Gamma_A$, if there is no strategy S_A for which $\mathring{n} \leqslant \mathring{n}_s$, and $\mathring{\tau} \leqslant \mathring{\tau}_s$, and at least one of these inequalities is strict. The curve Γ_A defines the predictive ability of algorithm A for the pair $(G,\ M)$. It allows comparison of algorithms irrespective of the prediction goals.

In practice, computational difficulties of Γ_A estimation are accompanied by statistical problems caused by insufficient samples of strong events. In such a case, requirements for prediction algorithms should be necessarily unified. The main requirement is to indicate a set of strategies of the algorithm which correspond to different thresholds of error $\mathring{\tau}$. For example, $\epsilon_i \leqslant \mathring{\tau} \leqslant \epsilon_{i+1}$, where $\epsilon = 0$, 5, 10, 20, 40 and 80%. The fraction of space–time alarm is chosen because of greater stability in its estimation. The graph of errors $(\mathring{n},\ \mathring{\tau})_\epsilon = \hat{\Gamma}_A$ with the statistics volume of strong events and the number of significant fitting parameters could serve as an 'empirical prediction capability' of the algorithm for region G and magnitude M. For insufficient data volume and an excess of fitting parameters, the behavior of graph $\hat{\Gamma}_A$ can be useful from the viewpoint of controlling the reliability of errors $(\mathring{n},\ \mathring{\tau})$. However, this requires further investigation.

It is noteworthy that density variations of predicted events' epicentres should be accounted for when the fraction of space–time alarm $\mathring{\tau}$ is calculated. Hypothetically, the density can also be reflected in weaker events.

Acknowledgements

This paper resulted from discussions on earthquake prediction at seminars given by Prof. V.I. Keilis-Borok in 1988. The author is grateful to all of the participants for fruitful stimulation.

Appendix A

Proof of Statement 3

Let $\{t_i\}$, $t_0 = 0$ be a renewal process with distribution function $F(x)$, $F(0) = 0$; Δ is an alarm set and $(0,\ t)$ is the observation interval.

We have the following possibilities for event t_1:

(1) With probability $\bar{F}(t) = 1 - F(t)$, $t_1 > t$. Then the number of successes $\nu_\Delta(t) = 0$ and the alarm time

$$\tau_\Delta(t) = \Delta(t) = \int_0^t \chi_\Delta(u)\, du$$

(2) with probabilities $dF(x)$, $t_1 \in (x,\ x + dx)$, $x < t$. Then

$$\nu_\Delta(t) = \chi_\Delta(x) + \nu_\Delta(t - x)$$

$$\tau_\Delta(t) = \Delta(x) + \tau_\Delta(t - x)$$

This leads to the equation for the characteristic function of the random vector $[\nu_\Delta(t),\ \tau_\Delta(t)]$:

$$\phi(t \mid \theta) = E \exp[-\theta_0 \nu_\Delta(t) - \theta_1 \tau_\Delta(t)]$$

That is

$$\phi(t \mid \theta) = \bar{F}(t) \exp[-\theta_1 \Delta(t)] + \int_0^t \exp[-\theta_0 \chi_\Delta(x) - \theta_1 \Delta(x)]\, \phi(t - x \mid \theta)\, dF(x) \tag{A1}$$

As

$$\left(-\frac{\partial}{\partial \theta_0}\right)^k \left(-\frac{\partial}{\partial \theta_1}\right)^l \phi(t \mid \theta) \mid_{\theta = 0}$$

$$= E\nu_\Delta^k \tau_\Delta^l(t) = m_{k,l}(t) \tag{A2}$$

eqn. (A1) generates equations for all moments of the vector $[\nu_\Delta(t), \tau_\Delta(t)]$. We are interested in the moments (A2) with $k + l \leqslant 2$.

The asymptotics of these values with $t \to \infty$ is determined by the asymptotics of the Laplace transform $\phi(t \mid \theta)$ in the zero vicinity. We have

$$\hat{\phi}(s \mid \theta) = \int_0^\infty e^{-st} \phi(t \mid \theta)\, dt$$

Taking into account that (A1) is the convolution equation, we will obtain

$$\hat{\phi}(s \mid \theta) = \int_0^\infty e^{-st} \bar{F}(t) \exp[-\theta_1 \Delta(t)]\, dt \times \left\{1 - \int_0^\infty \exp[-st - \theta_0 \chi_\Delta(x) - \theta_1 \Delta(x)]\, dF(x)\right\}^{-1} \tag{A3}$$

We use for the asymptotics the average values $m_0(t) = E\nu_\Delta(t)$, and $m_1(t) = E\tau_\Delta(t)$. From (A1) and (A2), we have

$$m_i(t) = \int_0^t m_i(t-x)\, dF(x) + \phi_i(t), \quad i = 0, 1 \tag{A4}$$

where

$$\phi_0(t) = \int_0^t \chi_\Delta(x)\, dF(x) = E\chi_\Delta(\tau) - \int_t^\infty \chi_\Delta(x)\, dF_0(x), \quad F_0 \equiv F$$

$$\phi_1(x) = \bar{F}(t)\Delta(t) + \int_0^t \Delta(x)\, dF(x) = \int_0^t \bar{F}(x)\chi_\Delta(x)\, dx = E\Delta(\tau) - m\int_t^\infty \chi_\Delta(x)\, dF_1(x)$$

and

$$m = E\tau, \quad dF_1 = \bar{F}(x)m^{-1}\, dx$$

Functions $\phi(t)$ increase monotonically and have bounded limits $\phi_i(\infty)$.

Taking into account that the renewal function (19) $H(t) = E\nu[0, t)$ is a solution of the equation

$$H(t) = \int_0^t H(t-x)\, dF(x) + 1$$

$m_i(t) - \phi_i(\infty)H(t)$ is the solution of eqn. (A4), with the right-hand side $\tilde{\phi}_i = \phi_i - \phi_i(\infty)$. It has the following form:

$$m_i(t) - \phi_i(\infty)H(t) = \int_0^t \tilde{\phi}_i(t-x)\, dH(x) \to m^{-1} \times \int_0^\infty \tilde{\phi}_i(x)\, dx, \quad t \to \infty \tag{A5}$$

The last asymptotic proceeds from the 'key renewal theorem' (Smith, 1955). This theorem can be applied, as $\tilde{\phi}_i$ are monotonic and integrated (with $m_2 < \infty$):

$$\int_0^\infty \tilde{\phi}_i(x)\, dx = \int_0^\infty m^i \int_t^\infty \chi_\Delta(x)\, dF_i(x)\, dt = m^i \int_0^\infty x\chi_\Delta(x)\, dF_i(x), \quad i = 0, 1$$

Relationship (20) is proved.

To calculate the covariation matrix $[\nu_\Delta(t), \tau_\Delta(t)]$ we will use a formal method which relies on (A2) and (A3). The orders of asymptotics $m_{k,l}(t)$ and $m_{k,0}(t)\ m_{0,l}(t)$ with $t \to \infty$ are equal, i.e.

$$m_{k,l}(t) = a_{k,l}t^2 + b_{k,l}t + c_{k,l} + o(1), \quad k + 1 = 2 \tag{A6}$$

Thus their Laplace transform in the vicinity $s = 0$ has the form

$$\hat{m}_{k,l}(s) = 2a_{k,l}s^{-2} + b_{k,l}s^{-1} + c_{k,l} + o(1)$$

The coefficients a, b, c in (A6) are thus determined by the first terms of the Loran series for the function

$$\hat{m}_{k,l}(s) = \frac{\partial^2}{\partial\theta_0^k\, \partial\theta_1^l} - \hat{\phi}(s \mid 0), \quad k + 1 = 2$$

in the vicinity $s = 0$. It is then possible to find the asymptotics of the covariation matrix, because

$$\operatorname{cov}[\nu_\Delta(t), \tau_\Delta(t)] = \left|\left| m_{k,l}(t) - m_0^k(t)m_1^l(t) \right|\right|, \quad k + 1 = 2$$

The routine calculations are omitted here. Strict derivation of (19) requires finer techniques.

References

Aki, K., 1981. A probabilistic synthesis of precursory phenomena. Maurice Ewing Ser., 4: 566–574.

Billingsley, P., 1965. Ergodic Theory and Information. John Wiley, New York.

Borovkov, A.A., 1984. Mathematical Statistics: Parameters Estimation, Testing Hypotheses. Nauka, Moscow, 472 pp. (in Russian).

Cox, D.R. and Lewis, P.A.W., 1966. The Statistical Analysis of Series of Events. Methuen, London; John Wiley, New York.

Daley, D.J. and Vere-Jones, D., 1988. An Introduction to the Theory of Point Processes. Springer Series in Statistics. Springer-Verlag, New York, 702 pp.

Davis, P., Jackson, D. and Kagan, Y., 1989. The longer it has been since the last earthquake, the longer the expected time till the next? (in press).

Fedotov, S.A., Sobolev, G.A., Boldyrev, S.A. et al., 1977. Long- and short-term earthquake prediction in Kamchatka. Tectonophysics, 37: 305–321.

Gabrielov, A.M., Levshina, T.A. and Rotwain, I.M., 1990. Block model of the lithosphere dynamics and earthquake prediction. Phys. Earth Planet. Inter., 61: 18–28.

Gusev, A.A., 1976. The earthquakes—indications and prediction. Seismicity and Depth Structure of Siberia and Far East Nauka, Novosibirsk, pp. 241–247 (in Russian).

Kagan, Y. and Knopoff, L., 1987a. Statistical short-term earthquake prediction. Science, 236: 1563–15567.

Kagan, Y. and Knopoff, L., 1987b. Random stress and earthquake statistics: time dependence. Geophys. J.R. Astron. Soc., 88: 723–731.

Keilis-Borok, V.I. and Kossobokov, V.G., 1990. Premonitory activation of earthquake flow: algorithm M8. Phys. Earth Planet. Inter., 61: 74–84.

Keilis-Borok, V.I. and Rotwain, I.M., 1990. Diagnosis of time of increased probability of strong earthquakes in different regions of the world: algorithm CN. Phys. Earth Planet. Inter., 61: 57–73.

Molchan, G., 1982. Stability in the linear prediction problem of stationary processes. Mat. Izv. VUZ, 1(236): 42–52 (in Russian).

Nishenko, S. and Buland, R., 1987. A generic recurrence interval distribution for earthquake forecasting. Bull. Seismol. Soc. Am. 77 (4): 1382–1399.

Prozorov, A.G. and Rantsman, Y.Y., 1972. Earthquake statistics and morphostructures of Eastern Soviet Central Asia. Dokl. Akad. Nauk SSSR, 207: 12–15.

Skorohod, A.B., 1964. Random Processes with Independent Increments. Nauka, Moscow, pp. 278 (in Russian).

Smith, W.L., 1955. Regenerative stochastic processes. Proc. R. Soc. Lond. A, 232(1188): 6–31.

Utsu, T., 1977. Probabilities in earthquake prediction (in Japanese). Zisin J. Seismol. Soc. Jpn, No. 30: 179–185.

Vere-Jones, D., 1978. Earthquake prediction—a statistician's view. J. Phys. Earth, 26: 129–146.

Working Group on California Earthquake Predictions (WGCEP), 1988. Probabilities of large earthquakes occurring in California on the San Andreas fault. U.S. Geol. Surv. Open-File Rep. 88-398, 62 pp.