Bulletin of the Seismological Society of America, Vol. 86, No. 3, pp. 691–700, June 1996

Statistical Estimation of Seismic Hazard Parameters: Maximum Possible Magnitude and Related Parameters

by V. F. Pisarenko, A. A. Lyubushin, V. B. Lysenko, and T. V. Golubeva

Abstract The problem of statistical estimation of earthquake hazard parameters is considered. The emphasis is on estimation of the maximum regional magnitude, M_{max}, and the maximum magnitude, $M_{max}(T)$, in a future time interval T and quantiles of its distribution. Two estimators are suggested: an unbiased estimator with the lowest possible variance and a Bayesian estimator. As an illustration, these methods are applied for the estimation of M_{max} and related parameters in California and Italy.

Introduction

Over the past few decades, much effort has been focused on obtaining realistic assessments of seismic hazard (see Lamarre *et al.,* 1992 and references cited therein for a detailed historic survey of the problem). An important component of the hazard assessment is the determination of maximum magnitude, and associated uncertainty.

The notion of apparent magnitude was introduced by Tinti and Mulargia (1985) (see also Kijko and Sellevoll, 1992). The apparent (i.e., observed) magnitude $\bar{M}$ is equal to "true" magnitude M, disturbed by a random error ε:

$$\bar{M} = M + \varepsilon.$$

The probability distribution of ε depends on the particular earthquake catalog and can be modeled by a Gaussian, uniform, or some other distribution function. The authors mentioned above assumed that the true magnitude distribution obeys the Gutenberg–Richter law, but this problem can be viewed otherwise. Namely, the Gutenberg–Richter law was empirically established using apparent magnitudes. The notion of true magnitude is rather vague and even slightly mysterious: we cannot flatly assert that there exists a true magnitude value M that is measured with a random error ε. The uncertainty of magnitude is formed not only by measurement error but it is also an inherent feature of earthquake source description. Thus, the apparent magnitude characterizes not only earthquake size (energy) but some nuisance factors (noise and others). Fortunately, both apparent and true magnitudes either satisfy or do not satisfy Gutenberg–Richter law simultaneously and with the same slope within the whole magnitude range, except at values approaching M_{max}. We study the problem of estimating M_{max} for two cases: $\bar{M} = M$ and $\bar{M} = M + \varepsilon$. We start with the assumption that all seismic parameters but M_{max} are known without error. However unrealistic this assumption seems to be, it can be accepted in some practical situations, as we shall show later. Under this assumption, we derive a statistically optimal M_{max} estimator, based on a catalog of arbitrary size n. This estimator is unbiased and has the lowest possible variance for any finite n. We point out that this result is valid for an arbitrary magnitude-frequency law. The explicit form of solution allows us to analyze the role of each seismic parameter in the M_{max}-estimation problem and to draw some practical conclusions. Then we generalize these results for estimation of an arbitrary function of M_{max}. As an example of such a function, we study the probability distribution function of $M_{max}(T)$—maximal magnitude that will occur in a future time interval $(t, t + T)$. Some of these results were obtained by Pisarenko (1991) and Pisarenko and Lysenko (1994).

Next we consider the most general case when uncertainties are present both in magnitude and in all other seismic parameters. In this case, there is no unbiased optimal estimator of M_{max}, and we suggest a Bayesian approach to estimate M_{max} and related parameters.

Statistical Estimation of M_{max} (Exact Magnitudes and Known Parameters)

We consider first the situation when magnitude is known exactly. Suppose $M_1, \ldots, M_n$ is a sample of main-shock magnitudes that can be considered as independent, identically distributed random values. If the earthquake catalog contains aftershocks, it is necessary to eliminate them by one of known procedures (see Gardner and Knopoff, 1974; Molchan and Dmitrieva, 1993). We assume further that the magnitude probability distribution function $F(x; \theta)$ (magnitude-frequency law) has the general form

$$F(x; \theta) = \begin{cases} G(x)/G(\theta), & M_0 \leqq x \leqq \theta, \\ 0, & x > \theta, \end{cases} \qquad (1)$$

where $\theta = M_{max}$ is the maximum possible magnitude (unknown parameter to be estimated) and M_0 is the threshold of completeness (known parameter). $G(x)$ is given by

$$G(x) = \int_{M0}^{x} g(u)\,du, \tag{2}$$

where $g(u)$ is some positive integrable function.

The probability density $f(x;\theta)$ has the form

$$f(x;\theta) = \begin{cases} f(x)/G(\theta), & M_0 \leqq x \leqq \theta, \\ 0, & x < M_0, \quad x > \theta. \end{cases}$$

For example, the function $g(u) = 10^{-bu}$ corresponds to the well-known Gutenberg–Richter law. We stress that our considerations are valid for any arbitrary magnitude-frequency law satisfying equations (1) and (2). Equations (1) and (2) imply that all parameters of seismic process (such as the earthquake activity rate and the "slope" of magnitude-frequency law) are known, with the exception of $M_{\max}$. We keep this assumption in the present section: later we shall consider uncertainty on the other parameters.

The likelihood function $L(\theta)$ for this situation is

$$L(\theta) = G^{-n}(\theta)\, g(M_1) \ldots g(M_n) H(\theta - \mu_n), \tag{3}$$

where $\mu_n = \max_{1 \leqq i \leqq n} M_i$. $H(x)$ is the Heavyside function:

$$H(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geqq 0. \end{cases}$$

Only one factor in equation (3) depends both on the unknown parameter θ and on the sample values $M_1, \ldots, M_n$. This factor is the function $H(\theta - \mu_n)$. Thus, μ_n is a sufficient statistic for parameter θ (see Rao, 1945). It follows from (3) that the θ value giving the maximum of L is $\theta = \mu_n$: i.e., μ_n is a maximum likelihood (m.l.) estimator of the maximum magnitude.

Now we use the result by Rao–Blackwell–Kolmogorov (Rao, 1945; Blackwell, 1947; Kolmogorov, 1950) that states: the conditional expectation (under fixed sufficient statistic μ_n) of any unbiased estimator $\hat{\theta}$ of parameter θ gives the unbiased estimator with the lowest possible variance. As an unbiased estimator $\hat{\theta}$, we can take one value, based on any single sample, e.g., M_1. It can be verified that the nonlinear function of M_1,

$$\hat{\theta}(M_1) = M_1 + G(M_1)/g(M_1), \tag{4}$$

is an unbiased estimate of θ; i.e., its expectation identically equals to θ:

$$E[\hat{\theta}(M_1)] \equiv \theta.$$

The distribution function $\Phi(x, y)$ of the pair (M_1, μ_n) is equal to

$$\Phi(x, y) = F^n(y;\theta)\, H(x - y) + F(x;\theta)\, F^{n-1}(y;\theta)\, H(y - x).$$

Now we can find the density function $\phi(x, y)$ of the pair (M_1, μ_n) and the density function $\phi(y)$ of μ_n:

$$\phi(x, y) = \frac{\partial^2 \Phi(x, y)}{\partial x \partial y}, \quad \phi(y) = \int_{-\infty}^{\infty} \phi(x, y)\,dx.$$

Thus, the conditional density function $\phi(x \mid y)$ of $M_1 = x$ under fixed $\mu_n = y$ is equal to

$$\phi(x \mid y) = \frac{\phi(x, y)}{\phi(y)} = \frac{H(y - x)}{G(y)} \left(1 - \frac{1}{n}\right) g(x) + \frac{\delta(x - y)}{n}. \tag{5}$$

Then the best unbiased estimator $\tilde{\theta}$ is given by averaging (4) with density (5):

$$\tilde{\theta}_n = \int_{M0}^{\theta} \hat{\theta}(x)\, \phi(x \mid \mu_n)\,dx = \mu_n + \frac{G(\mu_n)}{n\, g(\mu_n)} = \mu_n + \frac{1}{n\, f(\mu_n;\mu_n)}. \tag{6}$$

This result is well known in mathematical statistics (e.g., Kendall and Stuart, 1961, exercise 17.24, chapter 17).

Since the distribution function of μ_n is $F^n(x;\theta)$, it is easy to derive the variance of $\tilde{\theta}_n$:

$$\begin{aligned} \mathrm{var}(\tilde{\theta}_n) &= E(\tilde{\theta}_n^2) - [E(\tilde{\theta}_n)]^2 \\ &= \int_{M0}^{\theta} \left(x + \frac{1}{n\, f(x;x)}\right)^2 dF^n(x;\theta) - \theta^2 \\ &= \frac{1}{n^2} E\left(\frac{1}{f^2(\mu_n;\mu_n)}\right). \end{aligned} \tag{7}$$

The variance depends on unknown parameter θ, and therefore, equation (7) cannot be used directly. Fortunately, since μ_n is a sufficient statistic, any function of μ_n is an unbiased estimator of its expectation, having the lowest possible variance. Thus, we use as the best unbiased estimator of $\mathrm{var}(\tilde{\theta}_n)$ the following expression:

$$\widetilde{\mathrm{var}}(\tilde{\theta}_n) = \frac{1}{n^2 f^2(\mu_n;\mu_n)}. \tag{8}$$

Equations (6) through (8) can be generalized for estimation of an arbitrary function $\phi(\theta)$ of parameter θ. By similar reasoning, we get the best unbiased estimator $\tilde{\phi}(\theta)$:

$$\tilde{\phi}(\theta) = \phi(\mu_n) + \frac{\phi'(\mu_n)}{n\, f(\mu_n; \mu_n)}. \qquad (9)$$

The variance of $\tilde{\phi}$ and its best unbiased estimator are, respectively,

$$\mathrm{var}(\tilde{\phi}) = \frac{1}{n^2} E\left[\left(\frac{\phi'(\mu_n)}{f(\mu_n; \mu_n)}\right)^2\right], \quad \widetilde{\mathrm{var}}(\tilde{\phi}) = \frac{1}{n^2}\left(\frac{\phi'(\mu_n)}{f(\mu_n; \mu_n)}\right)^2. \qquad (10)$$

Now we shall generalize formulas (6) through (10) for the case when the earthquake sequence is a Poisson process with intensity parameter λ. Suppose we have in a time interval $[-\tau, 0]$ a random number of earthquakes ν_τ that obeys the Poisson distribution

$$\Pr(\nu_\tau = k) = \frac{\exp(-\lambda\tau)\,(\lambda\tau)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

We denote by μ_τ the maximum magnitude of the largest earthquake to occur in $[-\tau; 0]$ under condition $\nu_\tau \geqq 1$. Then for any function $\phi(\theta)$, the estimator

$$\tilde{\phi}_\tau(\theta) = \phi(\mu_\tau) + \frac{\phi'(\mu_\tau)}{\nu_\tau f(\mu_\tau; \mu_\tau)} \qquad (11)$$

is unbiased and has the lowest possible variance, given by

$$\mathrm{var}(\tilde{\phi}_\tau) = E\left[\frac{1}{\nu_\tau^2}\left(\frac{\phi'(\mu_\tau)}{f(\mu_\tau; \mu_\tau)}\right)^2\right]. \qquad (12)$$

Since $\nu_\tau \rightarrow \tau\lambda$ and $\mu_\tau \rightarrow \theta$ as $\tau \rightarrow \infty$, it can be shown that

$$\mathrm{var}(\tilde{\phi}_\tau) = \frac{1}{(\lambda\tau)^2}\left(\frac{\phi'(\tau)}{f(\mu_\tau; \mu_\tau)}\right)^2 + \frac{\mathrm{const}}{(\lambda\tau)^3}.$$

For finite τ, equation (12) is some function of the unknown parameter θ and cannot be calculated explicitly. But since μ_τ is a sufficient statistic for θ and since each function of a sufficient statistic is the best unbiased estimator of its own expectation, we get the best unbiased estimator for expression (12) or, rather, for its main term as $\tau \rightarrow \infty$:

$$\widetilde{\mathrm{var}}(\tilde{\phi}_\tau) = \frac{1}{\nu_\tau^2}\left(\frac{\phi'(\mu_\tau)}{f(\mu_\tau; \mu_\tau)}\right)^2. \qquad (13)$$

Now we would like to make some comments on the choice of the parameter M_0. Since increasing M_0 leads to decreasing λ_0 and at the same time to increasing $f(\theta; \theta)$, it can be shown from equation (13) that for $\lambda\tau \gg 1$, the variance practically does not depend on the choice of M_0. This conclusion is valid unless the magnitude-frequency law deviates significantly from a chosen parametric form, which is a truncated exponential.

We mentioned in the Introduction that estimation of θ [or $\phi(\theta)$] with known parameters b and λ can be considered realistic and adequate only when the existing uncertainties in b and λ do not change significantly the results of estimation. Now we consider this issue in more detail. We suppose that from an *a priori* study, uncertainty in b and λ can be characterized by their standard deviations σ_b and σ_λ.

For the moment, we restrict our analysis to the exponential Gutenberg–Richter law, although similar reasoning can be applied to other, nonexponential frequency-magnitude distributions. So, we assume that

$$g(u) = 10^{-bu} = \exp(-\beta u), \quad \beta = b\ln(10), \qquad (14)$$
$$G(u) = [\exp(-\beta M_0) - \exp(-\beta u)]/\beta.$$

Then we derive from equation (11) the θ estimator:

$$\tilde{\theta}_\tau = \mu_\tau + \frac{\exp[\beta(\mu_\tau - M_0)] - 1}{\nu_\tau\beta}.$$

We shall assume that σ_β and σ_λ are small enough, so they perturb this estimator by a factor of $(1 + \delta)$, where δ is a sufficiently small number.

If this rule is applied to $\tilde{\theta}_\tau$, *the results are as follows:*

$$\delta_\beta \cong \frac{(\mu_\tau - M_0)\exp[\beta(\mu_\tau - M_0)]}{\mu_\tau\nu_\tau\beta}\sigma_\beta, \qquad (15)$$
$$\delta_\lambda \cong \frac{\exp[\beta(\mu_\tau - M_0)]}{\lambda\mu_\tau\nu_\tau\beta}\sigma_\lambda.$$

If we take as a practical example $\nu_\tau = 50$, $\mu_\tau - M_0 = 2$, $\beta = 2$, $\mu_\tau = 7.5$, $\delta_\beta < 0.05$, and $\delta_\lambda < 0.05$, then restrictions on σ_β and σ_λ are

$$\sigma_\beta < 0.34; \quad \sigma_b < 0.15; \quad \sigma_\lambda/\lambda < 0.69.$$

These restrictions are not too severe and often can be met in practice. The general case of uncertainty on β and λ is treated below.

Estimation of Parameters Connected with M_{max}

If we assume the earthquake sequence to be Poissonian and independent of magnitude, then it is easy to derive the probability distribution of $M_{max}(T)$ under the condition of $\nu_T \geqq 1$ (see Epstein and Lomnitz, 1966; Kijko and Sellevoll, 1989):

$$\Phi_T(x; \theta) = \Pr(M_{max}(T) < x) = \frac{\exp[\lambda T F(x; \theta)] - 1}{\exp(\lambda T) - 1}. \qquad (16)$$

Equation (16) for fixed x can be considered a function of θ only, and the estimation method described in the previous section can be applied. Similarly, the density function at fixed x, i.e., $\partial\Phi_T(x;\theta)/\partial x$, can be estimated, as well as that of any other interesting parameter connected with the $M_{\max}(T)$ distribution. We restrict ourselves to $\Phi_T(x;\theta)$ and its quantiles x_a, i.e., the roots of the following equation:

$$\Phi_T(x_a;\theta) = a, \quad 0 < a < 1. \tag{17}$$

It should be noted that all mentioned parameters are functions of θ, and their simple estimators could be obtained by inserting the θ estimator into corresponding function. This method corresponds to the maximum likelihood (m.l.) principle: any function of an m.l. estimator is an m.l. estimator of this function. But since equations (16) and (17) are nonlinear, substituting the θ estimator will lead to biased estimators with variances larger than given by equation (12), while equation (11) gives an unbiased estimator with the lowest variance. Asymptotically, as $\tau \to \infty$, these estimators are of course equivalent, but for finite τ, their difference can be appreciable. In our examples (see below), this difference sometimes went up to 0.75 of a magnitude unit.

Using equation (17), one gets the following equation for the a quantile x_a:

$$x_a = x_a(\theta) = h\left[\frac{G(\theta)}{\lambda T}\ln(1 + a[\exp(\lambda T) - 1])\right],$$

where $h[\cdot]$ is the inverse function with respect to $G(x)$. Thus, from equations (11) and (16), one gets the best unbiased estimators for $\Phi_T(x;\theta)$ and $x_a(\theta)$:

$$\tilde{\Phi}_T(x;\theta) = \Phi_T(x;\mu_\tau) + \frac{\frac{\partial}{\partial\theta}\Phi_T(x;\mu_\tau)}{\nu_\tau f(\mu_\tau;\mu_\tau)}$$

$$= \Phi_T(x;\mu_\tau) - \frac{\lambda T\, F(x;\mu_\tau)}{\nu_\tau}\left[\Phi_T(x;\mu_T) + \frac{1}{\exp(\lambda T) - 1}\right], \tag{18}$$

$$\tilde{x}_a(\theta) = \tilde{x}_a(\mu_\tau) + \frac{x_{a,\theta}(\mu_\tau)}{\nu_\tau f(\mu_\tau;\mu_\tau)},$$

where $x_{a,\theta}(\mu_\tau)$ is the following derivative:

$$x_{a,\theta}(\mu_\tau) = \left.\frac{\partial x_a(\theta)}{\partial\theta}\right|_{\theta=\mu_\tau}.$$

The variances corresponding to these estimators are given by equation (13):

$$\widetilde{\operatorname{var}}(\tilde{\Phi}_T) = \frac{1}{\nu_\tau^2}\left[\frac{\frac{\partial}{\partial\theta}\Phi_T(x;\mu_T)}{f(\mu_\tau;\mu_\tau)}\right]^2, \tag{19}$$

$$\widetilde{\operatorname{var}}(\tilde{x}_a) = \frac{1}{\nu_\tau^2}\left[\frac{x_{a,\tau}(\mu_\tau)}{f(\mu_\tau;\mu_\tau)}\right]^2. \tag{20}$$

For the sake of brevity, the quantile formula given by equation (18) is written out only for the Gutenberg–Richter law (see equation 14):

$$\tilde{x}_a(\theta) = M_0 - \frac{1}{\beta}\ln[1 - \kappa(1 - \exp[-\beta(\mu_\tau - M_0)])] + \frac{\kappa(1 - \exp[-\beta(\mu_\tau - M_0)])}{\beta\nu_\tau[1 - \kappa(1 - \exp[-\beta(\mu_\tau - M_0)])]},$$

where

$$\kappa = \frac{\ln(1 + a[\exp(\lambda T) - 1])}{\lambda T}.$$

The variance estimators, given by equations (19) and (20), are equal to

$$\widetilde{\operatorname{var}}(\tilde{\Phi}_T) = \frac{1}{\nu_\tau^2}\left[\lambda T\, F(x;\mu_\tau)\left(\Phi_T(x;\mu_\tau) + \frac{1}{\exp(\lambda T) - 1}\right)\right]^2,$$

$$\widetilde{\operatorname{var}}(\tilde{x}_a) = \frac{1}{\nu_\tau^2}\left[\frac{\kappa(1 - \exp[-\beta(\mu_\tau - M_0)])}{\beta(1 - \kappa[1 - \exp(-\beta[\mu_\tau - M_0])])}\right]^2.$$

Estimation for Nonuniform Magnitude Distribution

We now generalize our estimation problem for the case when the magnitude distribution (magnitude-frequency law) depends on time. In other words, instead of assuming equation (1), we assume that the ith term of the sample $M_1, \ldots, M_n$ has a distribution function given by

$$F_i(x;\theta) = \begin{cases} G_i(x)/G_i(\theta), & M_{0i} \leqq x \leqq \theta, \\ 0, & x > \theta, \quad i = 1, \ldots, n. \end{cases}$$

By the above reasoning, it can be shown that for an arbitrary function $\phi(\theta)$, there exists an unbiased estimator $\tilde{\phi}$ with the lowest possible variance:

$$\tilde{\phi} = \phi(\mu_n) + \phi'(\mu_n)\left[\sum_{j=1}^{n} f_j(\mu_n;\mu_n)\right]^{-1}. \tag{21}$$

Its variance is equal to

$$\mathrm{var}(\tilde{\phi}) = E\left[\phi'(\mu_n)\left(\sum_{j=1}^{n} f_j[\mu_n; \mu_n]\right)^{-1}\right]^2, \quad (22)$$

and the best unbiased estimator of this variance is

$$\widetilde{\mathrm{var}}(\tilde{\phi}) = \left[\phi'(\mu_n)\left(\sum_{j=1}^{n} f_j[\mu_n; \mu_n]\right)^{-1}\right]^2. \quad (23)$$

Equations (21) and (23) can be used to combine several subcatalogs with different thresholds of completeness. For example, suppose we have a catalog consisting of two subcatalogs, covering two different time spans:

first subcatalog: k; b_1; M_{01}; $f_1(x; \theta)$; $M_1, \ldots, M_k$

second subcatalog: l; b_2; M_{02}; $f_2(x; \theta)$; $M_{k+1}, \ldots, M_{k+l}$.

Using (21), we can get three θ estimators, corresponding respectively to the first subcatalog, second subcatalog, and the full catalog:

$$\tilde{\theta}_1 = \mu_1 + [k f_1(\mu_1; \mu_1)]^{-1}; \quad (24)$$
$$\mu_1 = \max(M_1, \ldots, M_k);$$
$$\tilde{\theta}_2 = \mu_2 + [l f_2(\mu_2; \mu_2)]^{-1}; \quad (25)$$
$$\mu_2 = \max(M_{k+1}, \ldots, M_{k+l});$$
$$\tilde{\theta}_3 = \mu_3 + [k f_1(\mu_3; \mu_3) + l f_2(\mu_3; \mu_3)]^{-1}; \quad (26)$$
$$\mu_3 = \max(\mu_1, \mu_2).$$

Using equation (23), one can estimate the contribution of each subcatalog to the combined estimator (26) and to its variance. It turns out that the combined estimator has less variance as compared with variance of any linear combination of the estimators θ_1 and θ_2. This follows from the inverse quadratic dependence of estimator variance on the size of the catalog.

As an example, we consider two subcatalogs of Southern Italy covering the areas $36°30' \leqq \phi \leqq 39°50'$ and $14°30' \leqq \lambda \leqq 17°20'$, taken from Kijko and Sellevoll (1989) and Tinti and Mulargia (1984). The Gutenberg–Richter law (equation 14) is assumed. The catalog characteristics are given in Table 1. The slope of the magnitude-frequency law was estimated by Kijko and Sellevoll (1989) as $\tilde{\beta} = 1.93$. From equations (21) and (23), we get

$$\tilde{\theta}_1 = 7.28 \pm 0.68; \quad \tilde{\theta}_2 = 7.03 \pm 0.43; \quad \tilde{\theta}_3 = 6.86 \pm 0.26.$$

Performing optimal weighting of $\tilde{\theta}_1$ and $\tilde{\theta}_2$, with weights proportional to the inverse variances, we get the following estimate:

Table 1
The Characteristics of the Italian Catalog by Tinti and Mulargia (1984)

	Subcatalog 1	Subcatalog 2
Time period	1717–1818	1819–1979
Number of earthquakes	7	38
Maximum observed magnitude	6.6	6.6
Threshold of completeness	5.4	4.8

$$\tilde{\theta} = 7.10 \pm 0.36.$$

Comparing this standard deviation with 0.26, we see, as pointed out above, that the effect of joint use of independent subcatalogs is not additive.

We conclude that in similar situations old subcatalogs can sometimes be used in θ estimation despite their uncertainties.

Estimation with Uncertainties in Magnitude and b, λ Parameters

Now we consider the most general case when uncertainties are present both in magnitude and in other parameters. In this case, there exists no sufficient statistics but a trivial one; i.e., the whole sample $M_1, \ldots, M_n$ and the approach described above is not applicable. To estimate parameters in this situation, we use the Bayesian approach (see Ibragimov and Hasminskii, 1979; Dong *et al.*, 1984; Lamarre *et al.*, 1992; Savage, 1994). We assume that any observed magnitude $\bar{M}$ is the sum of a true magnitude M and some random error ε with known probability density:

$$\bar{M} = M + \varepsilon. \quad (27)$$

The observed magnitude $\bar{M}$ is called apparent magnitude (Tinti and Mulargia, 1984).

We restrict ourselves to the Gutenberg–Richter law, to facilitate the exposition. The results can be generalized for an arbitrary magnitude-frequency law, since the Bayesian approach is essentially numerical and does not need any specific analytic properties of distributions.

We assume that the true distribution $F_t(x; \beta; \theta)$ depends on time and at the moment t has the form

$$F_t(x; \beta; \theta) = \begin{cases} 0, & x < M(t), \\ \dfrac{1 - \exp[-\beta x + \beta M(\mathrm{t})]}{1 - \exp[-\beta\theta + \beta M(t)]}, & M(t) \leqq x \leqq \theta, \\ 1, & x > \theta. \end{cases}$$

The dependence of the lower threshold $M(t)$ on t is assumed known. It describes the evolution of this threshold in the given catalog. Distribution $F_t(x; \beta; \theta)$ depends on t only through $M(t)$.

The Poissonian rate $\lambda(t)$ depends on t, too, because $M(t)$ varies in time. In order to describe this dependence, we denote by λ_0 the rate corresponding to the magnitudes above some fixed threshold M_0. Then we get

$$\lambda(t) = \lambda_0 \frac{1 - \exp[-\beta\theta + \beta M(t)]}{1 - \exp(-\beta\theta + \beta M_0)}. \quad (28)$$

Equation (28) describes as well the periods of absence of registration that correspond to $M(t) = \theta$, $\lambda(t) = 0$.

We observe a sample of apparent magnitudes $x_1, \ldots, x_\nu$ in a time interval $(-\tau, 0)$, where earthquake occurrences at times $t_1, \ldots, t_\nu$ are generated by a Poisson process with time-varying rate $\lambda(t)$. Below, we denote this sample by $\mathbf{x}$. According to the Bayesian approach, we assume parameters $(\lambda_0, \beta, \theta)$ to be random values with known *a priori* density $p(\lambda_0, \beta, \theta)$. This assumption represents a well-known weak point of the Bayesian approach, but a consoling fact is that the final estimators depend weakly on this *a priori* distribution, and for large samples, this dependence practically vanishes. Thus, we have a set of four random values $(\lambda_0, \beta, \theta, \mathbf{x})$, one of them, $\mathbf{x}$, being vectorial. Three values $(\lambda_0, \beta, \theta)$ are nonobservable, and $\mathbf{x}$ is observed. We denote by $f(\lambda_0, \beta, \ldots)$ the probability density of random values indicated by its arguments $\lambda_0, \beta, \ldots$, whereas the conditional density, say, of λ_0 under given $\mathbf{x}$, is denoted by $f(\lambda_0 \mid \mathbf{x})$, where the argument of condition is separated by the vertical bar. Our aim is to find an *a posteriori* (conditional) density $f(\lambda_0, \beta, \theta \mid \mathbf{x})$ under given $\mathbf{x}$. Knowing this function, we can estimate any function $\phi(\lambda_0, \beta, \theta)$ of nonobservable parameters:

$$\tilde{\phi} = \iiint \phi(\lambda_0, \beta, \theta) f(\lambda_0, \beta, \theta \mid \mathbf{x})\, d\lambda_0\, d\beta\, d\theta. \quad (29)$$

Equation (29) is evaluated numerically. The consistency of Bayesian estimators and their optimal properties as sample size tends to infinity are well known (Ibragimov and Hasminskii, 1979).

From Bayes formula, we have

$$f(\mathbf{x} \mid \lambda_0, \beta, \theta) = C \exp\left(-\int_{-\tau}^{0} \bar{\lambda}(t)\, dt \right) \prod_{k=1}^{\nu} \bar{\lambda}(t_k) f_k(x_k \mid \beta, \theta), \quad (30)$$

where C is some normalizing constant; $\bar{\lambda}(t)$ is the Poissonian rate, corrected for magnitude uncertainty; and $f_k(x_k \mid \beta, \theta)$ is the *a posteriori* density of the apparent magnitude x_k, occurring at time t_k.

The corrected rate $\bar{\lambda}(t)$ corresponds to apparent magnitudes, whereas $\lambda(t)$ corresponds to true magnitudes. In case of a homogeneous catalog, $\bar{\lambda}(t)$ does not depend on t, and the term $\exp(-\bar{\lambda}\tau)\bar{\lambda}^\nu$ in equation (30) is proportional to the Poisson probability $\exp(-\bar{\lambda}\tau)\bar{\lambda}^\nu/\nu!$. For the sake of simplicity, we assume further on that the random error distribution is uniform:

$$n_t(x) = \begin{cases} \dfrac{1}{2\Delta(t)}, & |x| \leqq \Delta(t), \\ 0, & |x| > \Delta(t), \end{cases} \quad (31)$$

where $\Delta(t)$ is the known function characterizing the evolution of the magnitude uncertainty in the given catalog.

It can be shown (see Molchan and Podgayetskaya, 1973) that

$$\bar{\lambda}(t) = \lambda(t)\, [\exp(\beta\Delta[t]) - \exp(-\beta\Delta[t])]/2\beta\Delta(t). \quad (32)$$

In accordance with equation (27), the density of apparent magnitude at time t_k under condition $M \geqq M(t_k)$ is the convolution of true magnitude density and error density (31) taken for $x \geqq M(t_k)$:

$$f_k(x \mid \beta, \theta) = \begin{cases} C_k(\beta, \theta)\, h(\beta\Delta_k) \exp(-\beta x), & M(t_k) \leqq x \leqq \theta - \Delta_k, \\ \dfrac{C_k(\beta, \theta)}{2\beta\Delta_k} [\exp(\beta\Delta_k - \beta x) - \exp(-\beta\theta)], & \theta - \Delta_k \leqq x \leqq \theta + \Delta_k, \\ 0, \quad x < M(t_k), & x > \theta + \Delta_k, \end{cases} \quad (33)$$

where $\Delta_k = \Delta(t_k)$ and $h(x) = [\exp(x) - \exp(-x)]/2x$; $C_k(\beta, \theta)$ is a normalizing constant given by

$$C_k(\beta, \theta) = \frac{2\Delta_k \beta^2 \exp(\beta\theta)}{\exp[\beta(\theta + \Delta_k - M[t_k])] - \exp[\beta(\theta - \Delta_k - M[t_k])] - 2\beta\Delta_k}.$$

Thus, we can express $f(\mathbf{x} \mid \lambda_0, \beta, \theta)$ using equations (30) through (33). According to the Bayes formula, we have

$$f(\lambda_0, \beta, \theta, \mathbf{x}) = f(\mathbf{x} \mid \lambda_0, \beta, \theta) p(\lambda_0, \beta, \theta), \quad (34)$$

$$f(\lambda_0, \beta, \theta \mid \mathbf{x}) = f(\lambda_0, \beta, \theta, \mathbf{x})/f(\mathbf{x}), \quad (35)$$

where $f(\mathbf{x})$ is the *a priori* density of $\mathbf{x}$:

$$f(\mathbf{x}) = \iiint f(\lambda_0, \beta, \theta, \mathbf{x})\; d\lambda_0\; d\beta\; d\theta. \quad (36)$$

Collecting equations (32) through (36), we finally get

$$f(\lambda_0, \beta, \theta \mid \mathbf{x}) = \exp\left(-\int_{-\tau}^{0} \bar{\lambda}(t)\,dt\right) \prod_{k=1}^{\nu} \bar{\lambda}(t_k) f_k(x_k \mid \beta, \theta) p(\lambda_0, \beta, \theta) \Big/ \iiint \exp\left(-\int_{-\tau}^{0} \bar{\lambda}(u)\,du\right) \prod_{j=1}^{\nu} \bar{\lambda}(t_j) f_j(x_j \mid z, w) p(y, z, w)\; dy\; dz\; dw. \quad (37)$$

If we assume that the *a priori* density $p(\lambda_0, \beta, \theta)$ is a constant within a parallelepiped

$$\lambda_1 \leqq \lambda_0 \leqq \lambda_2; \quad \beta_1 \leqq \beta \leqq \beta_2; \quad \theta_1 \leqq \theta \leqq \theta_2,$$

then equation (37) is simplified:

$$f(\lambda_0, \beta, \theta \mid \mathbf{x}) = \exp\left(-\int_{-\tau}^{0} \bar{\lambda}(t)\,dt\right) \prod_{k=1}^{\nu} \bar{\lambda}(t_k) f_k(x_k \mid \beta, \theta) \Big/ \int_{\lambda_1}^{\lambda_2} dy \int_{\beta_1}^{\beta_2} dz \int_{\theta_1}^{\theta_2} dw \exp\left(-\int_{-\tau}^{0} \bar{\lambda}(u)\,du\right) \prod_{j=1}^{\nu} \bar{\lambda}(t_j) f_j(x_j \mid z, w). \quad (38)$$

Equations (37) through (38) give the desired *a posteriori* density of nonobservable parameters $(\lambda_0, \beta, \theta)$ given the observed vector $\mathbf{x}$.

Examples

We have used the catalogs for California and Southern Italy to illustrate our method. California was divided into four parts, shown in Figure 1, and Southern Italy was bounded by latitude 41.5. The characteristics of the subcatalogs are given in Table 2. Since uncertainties in λ, b parameters give rise to perturbation characteristics δ_β, δ_λ that are less than 0.06, we can apply unbiased estimators given by equation (11). The results of estimation are shown in Table 2. We see that in all cases, the standard deviations of Bayesian estimators are reasonable, varying from 0.33 up to 0.46. For unbiased estimators, we get one case (*Ca*3) with very large standard deviation (0.96) while all other standard deviations range from 0.4 to 0.54. It is evident that the unbiased estimator for *Ca*3 is rather uncertain, because of the large standard deviation, although its value 8.06 does not seem suspect. On the whole, we have good agreement between the unbiased and Bayesian estimators. So, we may conclude that the unbiased estimator can be used if δ_β, δ_λ are not too large (less, than, say, 0.05). The Bayesian procedure is more stable but, of course, more time-consuming.

Having analyzed quantile estimators, we can deduce that for $\alpha = 0.9$ and $T \leqq 50$, they are as a rule less than the estimators of $M_{\max}$. Only for $T \geqq 100$ do these quantiles become comparable with $M_{\max}$. The scatter of quantile estimators is rather low; e.g., for $T = 30$, their standard deviations range from 0.26 to 0.29 (see Table 2).

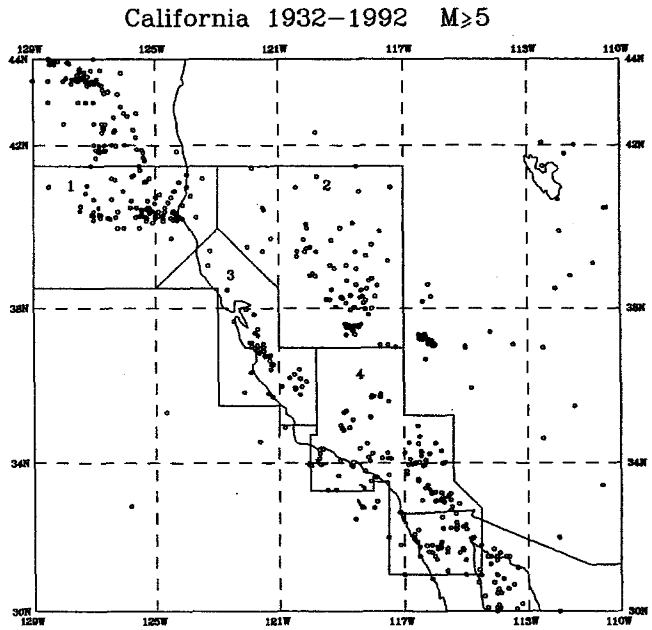

Figure 1. Subdivision of California into four regions for regional study of seismic hazard parameters. Shallow $M \geqq 5.0$ mainshocks are shown, 1932–1992.

The family of probability densities of $M_{\max}(T)$ and "tail" probabilities $P\{M_{\max}(T) > M\}$ for true and apparent maximal magnitude $M_{\max}(T)$ are shown in Figures 2 and 3. The apparent and true magnitude distributions differ little if T is not very large, with the true magnitude being strictly higher. The average of maximal deviations

$$\max_x [F(x) - \bar{F}(x)]$$

is $0.08 \pm 0.04 (T = 10)$ and $0.13 \pm 0.07 (T = 50)$. Also, we compared the quantiles of apparent and true magnitude distributions. The results are shown in Table 3. We see that only for a high significance level α ($\alpha = 0.98$) and large T ($T = 50$) does this difference became appreciable (0.25).

We would like to point out that by including random error on magnitude, one can get an estimate of the true magnitude maximum less than the maximum of the sample $M_1, \ldots, M_n$, while usually seismologists believe that the true maximum magnitude exceeds $\max(M_1, \ldots, M_n)$.

Discussion

We have suggested two methods for estimation of $\theta = M_{\max}$: unbiased and Bayesian estimators. The latter are, in general, more stable, but the former are much simpler. In good situations, their efficiency is almost the same. The Bayesian approach needs an *a priori* distribution for unknown parameters. It is an undesirable but unfortunately inherent

Table 2
The Characteristics of Subcatalogs of Shallow ($h \leqq 100$ km) Mainshocks of California (1932–1992) and Southern Italy (1900–1994) and Results of Estimation

Region	Number of earthquakes $M \geqq 5.0$	M_{max} observed	b	λ	Unbiased θ estimate	Bayes θ estimate	Bayes $x_{0.9}$ estimate, $T = 30$
*Ca*1	94	7.2	0.88	1.33	7.65 ± 0.45	7.85 ± 0.40	7.58 ± 0.29
*Ca*2	85	7.2	0.90	1.25	7.74 ± 0.54	7.91 ± 0.43	7.56 ± 0.26
*Ca*3	52	7.1	0.98	0.68	8.06 ± 0.96	7.86 ± 0.46	7.06 ± 0.27
*Ca*4	54*	7.7	0.75	0.82	8.09 ± 0.40	8.02 ± 0.33	7.98 ± 0.28
Southern Italy	44	7.1	0.76	0.48	7.59 ± 0.49	7.78 ± 0.42	7.40 ± 0.29

*Number of earthquakes in *Ca*4 is taken for $M \leqq 5.6$.

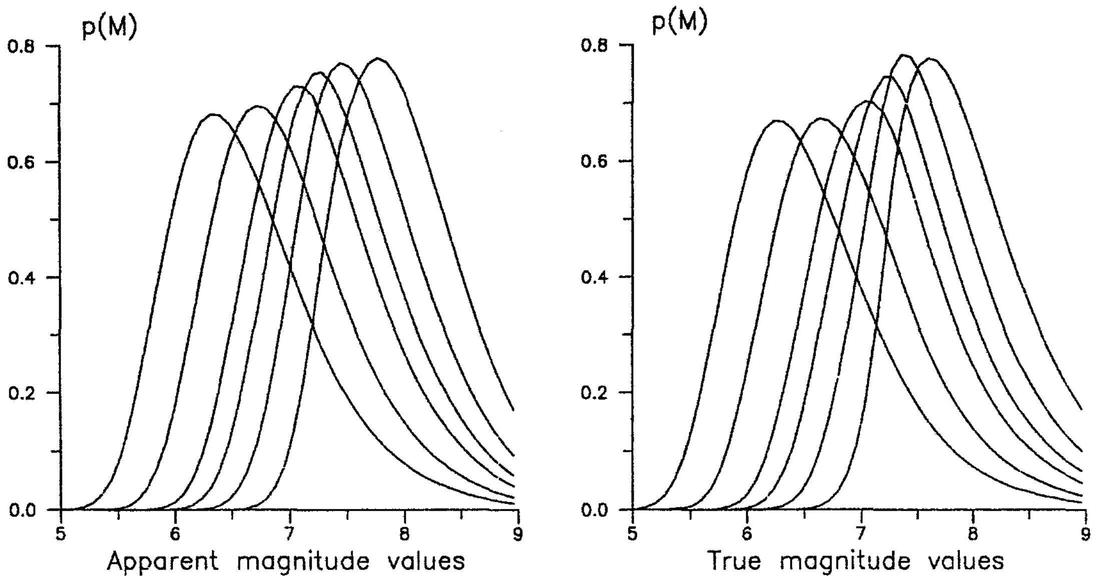

Figure 2. Statistical characteristics of seismic hazard parameters of the *Ca*4 region in California. Probability densities of $M_{max}(T)$. $T = 5, 10, 20, 30, 50$, and 100 yr (from left to right). Magnitude uncertainty Δ is 0.5.

Table 3
Differences between Quantiles of Apparent and True $M_{max}(T)$ Distributions, Averaged on Subcatalogs of Table 2

Significance Level a	(x_a[apparent $M_{max}(T)$] − x_a[true $M_{max}(T)$]) $T = 10$	$T = 50$
0.50	0.08 ± 0.01	0.10 ± 0.02
0.90	0.10 ± 0.02	0.15 ± 0.03
0.98	0.16 ± 0.03	0.25 ± 0.05

property of the Bayesian approach. However, *a posteriori* estimators depend weakly on this *a priori* distribution. The example of θ estimation in the *Ca*3 region shows that uncertainty of the unbiased estimator can sometimes be rather high. This case is explained, in our opinion, by a pronounced deviation of the empirical magnitude-frequency law from the straight-line Gutenberg–Richter dependence: the empirical curve is higher at large magnitudes.

The bias and standard deviation of estimators were evaluated by a parametric version of the standard bootstrap method (Efron and Tibshirani, 1986). We investigated numerically the influence of magnitude noise on the estimation of θ and related parameters. We can conclude that for moderate error levels ($\tilde{\Delta} \leqq 0.5$), the difference between the estimated distributions of apparent and true magnitudes is negligible. Say, for $T = 50$ and for all considered subcatalogs of California and Italy, 0.9 quantiles differ by more than 0.15 of a magnitude unit. For lesser quantile levels, the differences were still lower.

If noise level becomes larger ($\tilde{\Delta} \geqq 1.0$), then the difference between apparent and true estimators is more appreciable, in particular for large numbers of observed earthquakes.

The seismicity can be described also by an alternative characteristic, namely, by the probability of occurrence of an $M \geqq 7$ earthquake (Savage, 1994). But in our opinion, this characteristic is less adequate and full. On the one hand, the probability of an $M \geqq 7$ can be obtained from the $M_{max}(T)$-distribution function. On the other hand, it does not include any weaker earthquake with $M < 7$, although an M

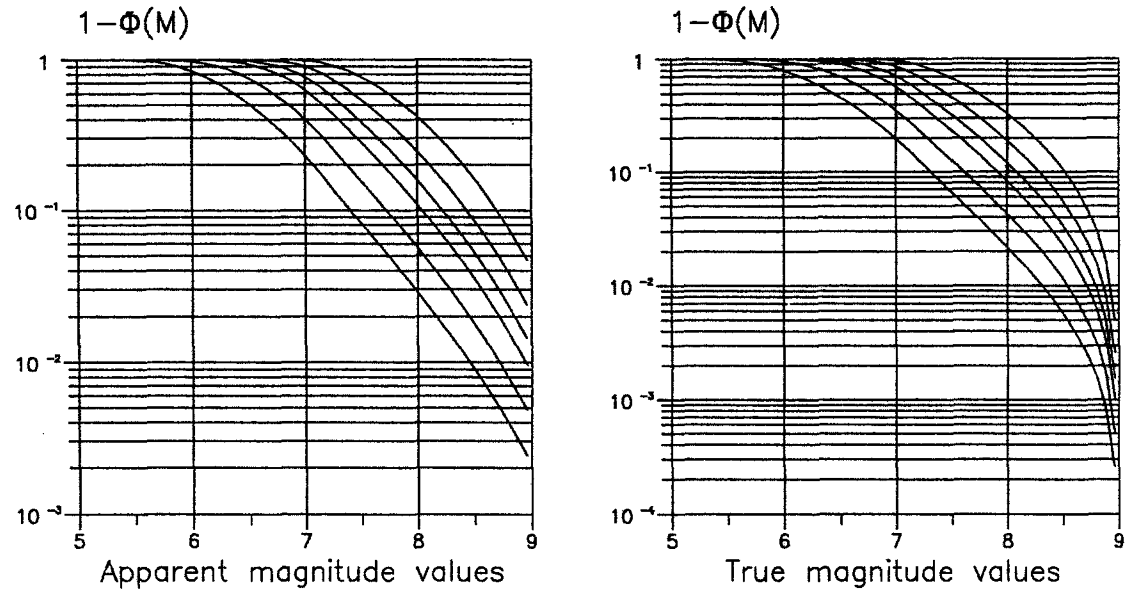

Figure 3. Statistical characteristics of seismic hazard parameters of the *Ca*4 region in California. "Tail" probabilities $1 - \Phi(M) = \Pr(M_{max}(T) > M)$. $T = 5, 10, 20, 30, 50$, and 100 yr (from left to right). Magnitude uncertainty Δ is 0.5.

= 6.9 earthquake is almost as dangerous as an $M = 7$ one. Besides, we have incorporated in the $M_{max}(T)$ estimation all associated uncertainties, unlike the approach in the cited article.

It is interesting to compare our estimates of M_{max} based on a 60-yr catalog with magnitudes of the largest historical earthquakes in California: Fort Tejon, 09.01.1857, $M = 8.2$ (region *Ca*4); Owens Valley, 26.03.1872, $M = 8.3$ (region *Ca*4); and San Francisco, 18.04.1906, $M = 8.3$ (region *Ca*3) (Iacopi, 1964). All three historical earthquake magnitudes are close enough to our estimates, and the corresponding differences are less than the estimated standard deviations. This fact is not very surprising for the *Ca*4 region, since maximum observed magnitude in our catalog for this region is 7.7. But it is noteworthy for the *Ca*3 region, where maximum observed magnitude in our catalog is 7.1.

In order to get reasonable estimators of λ, b parameters, it is enough to use an instrumental catalog for the period of 50 yr or more. Historical catalogs where intensities are converted into magnitudes by some empirical relation are not necessary for λ, b estimation. The situation with θ estimation is quite different. We find that additional catalogs (including historical ones) are useful for M_{max} estimation despite their magnitude uncertainty.

We have described our two methods for estimating maximum magnitudes. It goes without saying that these methods are easily extended to the distribution of seismic moment, m_0, connected with magnitude M by functional relation:

$$M = a \log(m_0) + b,$$

where a and b are some constants.

We would like to stress once more that the suggested methods are applicable not only to the Gutenberg–Richter magnitude-frequency law but to any arbitrary law. For example, the Gaussian distribution for M or gamma distribution for m_0 (see Kagan, 1991) could be used as well.

Acknowledgments

The authors are grateful to I. Savin and T. Rukavishnikova for their assistance in preparing the manuscript.

The research described in this publication was made possible in part by Grant Number MJF000 from the International Science Foundation and by Grant Number 94-232 from INTAS.

References

Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation, *Ann. Math. Statist.* **18,** 105–110.

Dong, W. M., H. C. Shah, and A. B. Bao (1984). Utilization of geophysical information in Bayesian seismic hazard model, *Soil Dyn. Earthquake Eng.* **3,** 103–111.

Efron, B. and R. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy, *Statist. Sci.* **1,** no. 1, 54–75.

Epstein, B. and C. Lomnitz (1966). A model for the occurrence of large earthquakes, *Nature* **211,** 954–956.

Gardner, J. K. and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? *Bull. Seism. Soc. Am.* **64,** 1363–1367.

Iacopi, R. (1964). *Earthquake Country,* Lane Book Company, Menlo Park, California, 192 pp.

Ibragimov, I. A. and R. Z. Hasminskii (1979). *Asymptotical Theory of Estimation,* Nauka, Moscow, 527 pp. (in Russian). English translation: Ibragimov, I. and R. Hasminskii (1981). *Statistical Estimation, Asymptotic Theory,* Springer-Verlag, New York.

Kagan, Y. Y. (1991). Seismic moment distribution, *Geophys. J. Int.* **106,** 123–134.

Kendall, M. and A. Stuart (1961). The advanced theory of statistics, in *Inference and Relationship,* Vol. 2, Griffin, London, 899 pp.

Kijko, A. and M. A. Sellevoll (1989). Estimation of earthquake hazard parameters from incomplete data files. Part I. Utilization of extreme and complete catalogs with different threshold magnitudes, *Bull. Seism. Soc. Am.* **79,** 645–654.

Kijko, A. and M. A. Sellevoll (1992). Estimation of earthquake hazard parameters from incomplete data files. Part II. Incorporation of magnitude heterogeneity, *Bull. Seism. Soc. Am.* **82,** 120–134.

Kolmogorov, A. (1950). Unbiased estimators, *Izvestiya AN SSSR,* Ser. Mat., 303–326, (in Russian). English translation in *Selected Works of A. N. Kolmogorov, Vol. II, Probability Theory and Mathematical Statistics,* A. N. Shiryayev (Editor), Kluwer Academic Publishers (Dordbrecht/Boston/London), 1992, 597 pp.

Lamarre, M., B. Townshed, and H. C. Shah (1992). Application of the bootstrap method to quantify uncertainty in seismic hazard estimates, *Bull. Seism. Soc. Am.* **82,** 104–119.

Molchan, G. M. and V. M. Podgayetskaya (1973). Parameters of global seismicity, in *Computational Seismology,* Iss. 6, Nauka, Moscow, 44–46 (in Russian).

Molchan, G. M. and O. D. Dmitrieva (1993). An aim approach to aftershock identification problem, in Seismicity and seismic zoning of Northern Europe-Asia, Iss. 1, *Inst. Phys. Earth Rus. Ac. Sci.,* 62–69 (in Russian).

Pisarenko, V. F. (1991). Statistical evaluation of maximum possible earthquake, *Phys. Solid Earth* **27,** 757–763.

Pisarenko, V. F. and V. B. Lysenko (1994). Probability distribution of the maximal earthquake, that may occur in a given time interval, submitted to *J. Earthquake Prediction Res.*

Rao, C. R. (1945). Information and accuracy attainable in estimation of statistical parameters, *Bull. Calcutta Math. Soc.* **7,** 81–91.

Savage, J. C. (1994). Probability of one or more $M \geqq 7$ earthquakes in southern California in 30 years, *Geophys. Res. Lett.* **21,** no. 4, 313–315.

Tinti, S. and F. Mulargia (1984). Probability of occurrence of large earthquakes in California and eastern Sicily on the basis of extreme value theory, *Ann. Geophys.* **2,** 529–536.

Joint Institute of the Physics of the Earth
Russian Academy of Sciences
Bolshaya Gruzinskaya 10
123242 Moscow, Russia
(A.A.L.)

International Institute of Earthquake Prediction Theory
and Mathematical Geophysics
Russian Academy of Sciences
Warshavskoe shosse 79, korp. 2
113556 Moscow, Russia
(V.F.P., V.B.L., T.V.G.)

Manuscript received 9 May 1995.