Geophys. J. R. astr. Soc. (1970) **21**, 307–322.

Some Applications of the Maximum Likelihood Method in Seismology

V. F. Pisarenko

(Received 1970 February 9)

Summary

The standard procedure of maximum likelihood estimation is stated. This procedure is applied to derive maximum likelihood estimators in some seismological problems, namely amplitude and phase corrections, group and phase velocities of surface waves and derivatives of travel-time curves $dt/d\Delta$. The formulas for confidence regions for these functions are obtained.

1. Introduction

The maximum likelihood method is intensively used in mathematical statistics. Advantages of this method consist in its universality and in the fact that its estimators are the most efficient under rather general conditions. The maximum likelihood method has been applied successfully in some seismological problems [see **(1)–(5)**]. As a rule this method gave better results than other methods, and its only disadvantage consisted in larger computations.

In this paper we shall state the standard maximum likelihood procedure for parameter estimation and illustrate its application in several seismological problems. When applying this procedure we encountered a difficulty rather typical in seismology: observed data often contain, besides parameters we are interested in, some incidental parameters, whose number increases as the volume of data increases. The maximum likelihood estimators in such situation are not guaranteed to be consistent. In our case we have proved the consistency and asymptotic normality of the maximum likelihood estimators (m.l.e.) directly. The question whether the m.l.e. in that case are efficient or not is open.

2. Standard maximum likelihood estimation

Let us suppose that a sample $X_1, \ldots, X_n$ of random variables is given, whose probability distribution depends on parameters $\alpha_1, \ldots, \alpha_m$. The parameters can vary in some domain A. Let us suppose further that the density function of random variables under question exists and denote it by $f(z_1, \ldots, z_n/\alpha_1, \ldots, \alpha_m)$. Substituting for the arguments of this density the sample $X_1, \ldots, X_n$ we get:

$$L = f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m). \tag{1}$$

Function (1) (or sometimes its logarithm) is called the likelihood function for the parameters $\alpha_1, \ldots, \alpha_m$.

The m.l.e.'s of the parameters $\alpha_1, \ldots, \alpha_m$, according to this definition, are equal to those values $\hat{\alpha}_1, \ldots, \hat{\alpha}_m$ that give the maximum value to the function (1). We shall consider only the situations where the maximum of the function (1) over $\alpha_1, \ldots, \alpha_m$ under fixed $X_1, \ldots, X_n$ is achieved with probability 1 in a unique point. If the function (1) is differentiable, then its maximum in the domain A is achieved in the point where

$$\frac{\partial f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)}{\partial \alpha_k} = 0, \quad k = 1, \ldots, m. \tag{2}$$

Sometimes it is more convenient from the computational point of view to consider the system of equations that is equivalent to (2):

$$\frac{\partial \log f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)}{\partial \alpha_k} = 0, \qquad k = 1, \ldots, m. \tag{3}$$

The equations (3) are called the likelihood equations. Thus the problem of finding the m.l.e. for $\alpha_1, \ldots, \alpha_m$ is reduced either to solving the system (3) or to finding directly the maximum of the function (1). From the computational point of view this problem is often rather complicated, in particular when m is large. Some methods of solving such problems can be found in **(6)–(8)**. Under rather general regularity conditions on the density function (1) the m.l.e. $\hat{\alpha}_1, \ldots, \widehat{\alpha_m}$ are consistent, asymptotically-normal and asymptotically-efficient [see e.g. **(7, 9)**]. This means that, as $n \to \infty$, the estimators $\hat{\alpha}_1, \ldots, \widehat{\alpha_m}$

(a) converge in probability to the true values of parameters $\alpha_1, \ldots, \alpha_m$,

(b) are asymptotically-normal variables after normalization of order $n^{-\frac{1}{2}}$, and

(c) the variances of their limit normal distribution are less than the corresponding variances of any other regular estimators.

The covariance matrix B of the limit normal law for $\sqrt{n}(\hat{\alpha}_1 - \alpha_1), \ldots, \sqrt{n}(\hat{\alpha}_m - \alpha_m)$ can be found from the following condition. If we denote an element of the inverse matrix B^{-1} by β_{kj}, then

$$\beta_{kj} = E\left\{\frac{\partial \log f}{\partial \alpha_k} \cdot \frac{\partial \log f}{\partial \alpha_j}\right\}; \qquad k, j = 1, \ldots, m, \tag{4}$$

where the symbol E means mathematical expectation, i.e. the result of averaging with density $f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)$;

$$\beta_{kj} = \int_{-\infty}^{\infty}\!\!\int \frac{\partial \log f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)}{\partial \alpha_k} \cdot \frac{\partial \log f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)}{\partial \alpha_j}$$
$$\times f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)\, dX_1, \ldots, dX_n. \tag{5}$$

The matrix $\{\beta_{kj}\}$ is called *Fisher's information matrix.*

The limit density function $\phi(y_1, \ldots, y_m)$ of the random variables

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_1), \ldots, \sqrt{n}(\hat{\alpha}_m - \alpha_m)$$

can be expressed by means of β_{kj}:

$$\phi(y_1, \ldots, y_m) = \frac{1}{(2\pi)^{m/2} \det B^{\frac{1}{2}}} \cdot \exp\left\{-\frac{1}{2} \sum_{k,j=1}^{m} \beta_{kj} y_k y_j\right\}. \tag{6}$$

In order to obtain a confidence region for the true values $\alpha_1, \ldots, \alpha_m$ one can consider the statistic χ^2:

$$\chi^2 = n \sum_{k,\, j=1}^{m} \beta_{kj}(\hat{\alpha}_k - \alpha_k)(\hat{\alpha}_j - \alpha_j). \tag{7}$$

The distribution of this statistic tends, as $n \to \infty$, to a χ^2-distribution with m degrees of freedom. Consequently the inequality

$$\chi^2 \leqslant \chi_m^{\,2}(\gamma), \tag{8}$$

where $\chi_m^{\,2}(\gamma)$ is the quantile of level γ for χ^2-distribution with m degrees of freedom, specifies a confidence region of confidence level γ for $\alpha_1, \ldots, \alpha_m$. This region is an ellipsoid, because the matrix $\{\beta_{kj}\}$ is positive-definite. In order to be able to use the confidence region (8) in practice one needs β_{kj}, and β_{kj} depends according to (5) on unknown parameters $\alpha_1, \ldots, \alpha_m$. This seems to be a vicious circle. A way out consists of substituting for the unknown $\alpha_1, \ldots, \alpha_m$ in (5) the approximate values $\hat{\alpha}_1, \ldots, \hat{\alpha}_m$:

$$\hat{\beta}_{kj} = \int_{-\infty}^{\infty} \!\!\ldots\! \int \frac{\partial \log f(X_1, \ldots, X_n/\hat{\alpha}_1, \ldots, \hat{\alpha}_m)}{\partial \hat{\alpha}_k} \cdot \frac{\partial \log f(X_1, \ldots, X_n/\hat{\alpha}_1, \ldots, \hat{\alpha}_m)}{\partial \hat{\alpha}_j}$$
$$\times f(X_1, \ldots, X_n/\hat{\alpha}_1, \ldots, \hat{\alpha}_m)\, dX_1, \ldots, dX_n. \tag{9}$$

The statistic $\hat{\chi}^2$:

$$\hat{\chi}^2 = n \sum_{k,\, j=1}^{m} \hat{\beta}_{kj}(\hat{\alpha}_k - \alpha_k)(\hat{\alpha}_j - \alpha_j), \tag{10}$$

like χ^2, tends to a χ^2-distributed variable with m degrees of freedom, so that one can take a confidence region of confidence level γ in the form:

$$\hat{\chi}^2 \leqslant \chi_m^{\,2}(\gamma). \tag{11}$$

Let us consider now the case when n is not too large, so that the asymptotic theory stated above does not yet hold.

We suppose that the parameters $\alpha_1, \ldots, \alpha_m$ are random variables themselves and vary from experiment to experiment according to some *a priori* density function $\psi(\alpha_1, \ldots, \alpha_m)$. Such a situation can occur, e.g. when $\alpha_1, \ldots, \alpha_m$ are signal parameters on a seismogram and vary in a random manner from earthquake to earthquake. In this situation one can find a confidence region for $\alpha_1, \ldots, \alpha_m$ from the *a posteriori* distribution of these parameters, i.e. from their conditional distribution under the condition that the observed sample is $X_1, \ldots, X_n$. According to Bayes formula [see e.g. (**9**)] we have for this *a posteriori* density:

$$g(\alpha_1, \ldots, \alpha_m/X_1, \ldots, X_n)$$
$$= \frac{f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m).\psi(\alpha_1, \ldots, \alpha_m)}{\int_{-\infty}^{\infty} \!\ldots\! \int f(X_1, \ldots, X_n/\bar{\alpha}_1, \ldots, \bar{\alpha}_m).\psi(\bar{\alpha}_1, \ldots, \bar{\alpha}_m)\, d\bar{\alpha}_1, \ldots, d\bar{\alpha}_m}. \tag{12}$$

Having the function g it is possible to choose in the multidimensional space of the parameters $\alpha_1, \ldots, \alpha_m$ a region $\mathfrak{A}$, such that the integral of the function g over $\mathfrak{A}$ would have the prescribed confidence level γ. The choice of such a region $\mathfrak{A}$ is not unique, but one usually chooses it near maximum values of g, i.e. one finds a constant C so that

$$\int \cdots \int_{g>c} g(\alpha_1, \ldots, \alpha_m/X_1, \ldots, X_n)\, d\alpha_1, \ldots, d\alpha_m = \gamma. \tag{13}$$

Now the region $\mathfrak{A}$ defined by the inequality

$$g(\alpha_1, \ldots, \alpha_m/X_1, \ldots, X_n) > C \tag{14}$$

is a confidence region of confidence level γ for $\alpha_1, \ldots, \alpha_m$. The main difficulty in using formulas (12)–(14) (besides computational difficulties) is caused usually by ignorance (or incomplete knowledge) of the *a priori* distribution $\psi(\alpha_1, \ldots, \alpha_m)$. The function ψ can sometimes be estimated from previous experiments. Sometimes it is desirable to manage without, as follows. The formula (12) shows that the *a posteriori* density g is the likelihood function f weighted with weight ψ and normalized. According to standard maximum likelihood theory the likelihood function tends to a δ-function as $n \to \infty$, the δ-function being concentrated in the vicinity of the true values of $\alpha_1, \ldots, \alpha_m$. On the other hand it is often known in advance that the function ψ varies smoothly as compared to the δ-type function f. Because of the normalization (12) it does not matter what the function ψ is and it can be put equal to a constant. Then we have

$$g(\alpha_1, \ldots, \alpha_m/X_1, \ldots, X_n) = \frac{f(X_1, \ldots, X_n/\alpha_1, \ldots, \alpha_m)}{\int_{-\infty}^{\infty} \cdots \int f(X_1, \ldots, X_n/\bar{\alpha}_1, \ldots, \bar{\alpha}_m)\, d\bar{\alpha}_1, \ldots, d\bar{\alpha}_m}. \tag{15}$$

It should be noted however, that the parameters $\alpha_1, \ldots, \alpha_m$ cannot always be considered as random variables, because sometimes they do not vary from experiment to experiment or vary systematically.

There is a general method of deriving confidence regions when the likelihood function is known [see (7)]. It should be noticed however that as a rule this method involves very cumbersome computations, particularly when n and m are not too small. For every fixed vector $(\alpha_1, \ldots, \alpha_m)$ we choose a region $C(\alpha_1, \ldots, \alpha_m)$ in X-space so that the condition

$$(X_1, \ldots, X_n) \in C(\alpha_1, \ldots, \alpha_m) \tag{15'}$$

would occur with probability γ, i.e.

$$\int \cdots \int_{(z_1, \ldots, z_n) \in C(\alpha_1, \ldots, \alpha_m)} f(z_1, \ldots, z_n/\alpha_1, \ldots, \alpha_m)\, dz_1, \ldots, dz_n = \gamma$$

for every $(\alpha_1, \ldots, \alpha_m)$.

Now we insert in (15′) the observed sample $X_1, \ldots, X_n$ and try all vectors $(\alpha_1, \ldots, \alpha_m)$. The set of those vectors $(\alpha_1, \ldots, \alpha_m)$ for which (15′) holds will constitute the confidence region of confidence level γ.

We mention briefly methods of deriving confidence regions, not connected with maximum likelihood estimators. These methods are not universal. Their general idea is to find some statistic $T(X_1, \ldots, X_n; \alpha_1, \ldots, \alpha_m)$ whose distribution does not depend on $\alpha_1, \ldots, \alpha_m$ and can be evaluated. The examples of such statistics are

(1) Student's statistic for an unknown mean value of a normal population;

(2) χ^2-statistic for an unknown variance of a normal population;

(3) Multidimensional version of Student's statistic–Hotelling statistic [see (10)]; and

(4) Bartlett–Wilks–Linnik statistic for parameters of multidimensional linear regression [see **(10, 11)**].

Knowing such a statistic T, it is easy to construct a confidence region of γ-level for $\alpha_1, \ldots, \alpha_m$ by inverting the inequalities

$$T_1 < T(X_1, \ldots, X_n;\ \alpha_1, \ldots, \alpha_m) < T_2, \tag{16}$$

where T_1 and T_2 are chosen so that the inequalities (16) would hold with probability γ;

$$\mathscr{P}\{T_1 < T(X_1, \ldots, X_n;\ \alpha_1, \ldots, \alpha_m) < T_2\} = \gamma. \tag{17}$$

3. Derivation of m.l.e.'s in some seismological problems

3.1 *Relative amplitude and phase corrections*

Let us denote the complex Fourier transforms of records at two seismic stations by

$$\left.\begin{aligned} x_k(\omega) &= s_k(\omega)+n_k(\omega) \\ y_k(\omega) &= H(\omega).s_k(\omega)+m_k(\omega) \end{aligned}\right\} k = 1, \ldots, N, \tag{18}$$

where $s_k(\omega)$ is the spectrum of the signal of the k-th earthquake at the first station, $H(\omega)\cdot s_k(\omega)$ is the spectrum of the signal of the k-th earthquake at the second station; $n_k(\omega)$, $m_k(\omega)$ are Fourier transforms of the noise at the first and second stations respectively. The total number of earthquakes under consideration is N. The noises are assumed to be Gaussian stationary independent processes with correlation functions $B(\tau)$ and $R(\tau)$ for the first and second station respectively. Mean values of the noises are zero. The function $H(\omega)$ is the frequency response of a linear filter connecting undisturbed signals at the first and second stations. For example, in the case of the determination of the interstation phase velocity of surface waves the argument of the complex function $H(\omega)$ is equal to the phase difference at frequency ω. Let us fix for a while a frequency ω and derive the m.l.e. for $|H(\omega)| = r$ and $\arg H(\omega) = \theta$. The likelihood function is

$$\log L = -N \log (2\pi)^2 \sigma^4 z - \frac{1}{2\sigma^2} \sum_{k=1}^{N} |x_k - s_k|^2 - \frac{1}{2\sigma^2 z} \sum_{k=1}^{N} |y_k - Hs_k|^2, \tag{19}$$

where z stands for the ratio of the noise variances at the second and first stations for frequency ω. If the time interval of the records is $(-T, T)$ then $z(\omega)$ can be written in the form

$$z(\omega) = \int_{-T}^{T}\!\!\int R(t-s)\, e^{i\omega(t-s)} dt\, ds \Big/ \int_{-T}^{T}\!\!\int B(t-s)\, e^{i\omega(t-s)} dt\, ds. \tag{20}$$

The unknown complex signals s_k are incidental parameters whose number increases as $N \to \infty$; we are interested only in the parameters r, θ.

Representing complex values on a plane we can give the following geometrical reformulation of our problem (see Fig. 1). There are unknown vectors s_k; they are subjected to a rotation through an angle θ and to a dilatation (compression) by a factor r. However both original vectors s_k and transformed vectors $H.s_k$ are known only approximately. We have to provide estimators for the angle θ and the dilatation coefficient r, using N pairs of approximate vectors.

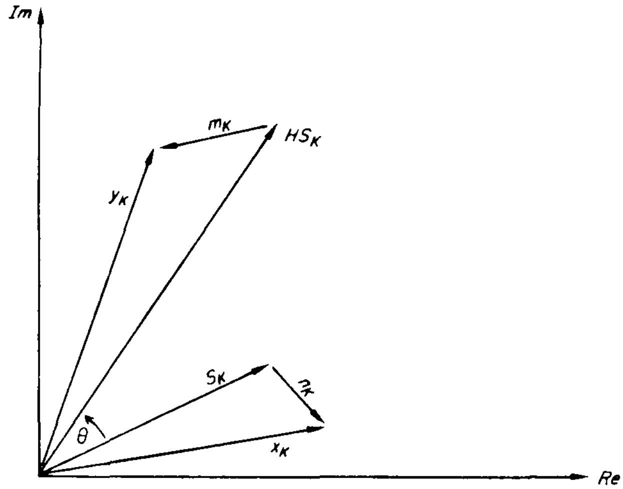

FIG. 1. Geometrical representation of s_k and Hs_k, where $H = re^{i\theta}$.

It is easy to find max L over all $r, \theta, s_1, \ldots, s_N$. Let us introduce complex vectors, with their inner product and norm, according to the formulas:

$$\left.\begin{aligned} x &= (x_1, \ldots, x_N); & (x, y) &= \sum_{k=1}^{N} x_k y_k^* \\ y &= (y_1, \ldots, y_N); & \|x\|^2 &= (x, x) = \sum_{k=1}^{N} |x_k|^2 \end{aligned}\right\} \tag{21}$$

Then the m.l.e.'s of r and θ have the form

$$\hat{r} = d + \surd(d^2 + z); \tag{22}$$

$$\hat{\theta} = \tan^{-1}(a/b); \tag{23}$$

where

$$d = \frac{\|y\|^2 - z\|x\|^2}{2|(x, y)|}; \qquad a = \operatorname{Im}(y, x); \qquad b = \operatorname{Re}(y, x).$$

The estimators $\hat{r}$, $\hat{\theta}$ are consistent, asymptotically-normal and uncorrelated. This can be proved, though not from known theorems of m.l.e. properties, which are not applicable because of the incidental parameters. These results must be proved individually, using the asymptotic normality of (y, x), $\|x\|^2$, $\|y\|^2$, which behave like sample moments of independent and identically distributed random variables.

Asymptotic confidence intervals for r, θ have the following form

$$\mathscr{P}\{|r - \hat{r}| + t_\gamma . \delta\} \cong 2\gamma - 1 \tag{24}$$

$$\mathscr{P}\left\{|\theta - \hat{\theta}| < \frac{t_\gamma}{\hat{r}} . \delta\right\} \cong 2\gamma - 1 \tag{25}$$

where t_γ is the quantile of level γ of the normal distribution. The statistic δ is defined by the following expressions:

$$\delta^2 = N^{-1} P^{-2} [2z + (z + \hat{r}^2).P] \tag{26}$$

$$P = \frac{2(z + \hat{r}^2)}{\hat{r}[(\hat{r}^2 . \|x\|^2 + \|y\|^2 . (a^2 + b^2)^{\frac{1}{2}} - 2\hat{r}]}. \tag{27}$$

The details of the derivation of (22)–(27), and their examination by means of some artificial examples with pseudo-random numbers can be found in **(12)**, **(13)**.

Let us consider now the questions of the confidence region for the values of the function $H(\omega)$ at some prescribed points $\omega_1, \ldots, \omega_M$ and of the smoothing of $\hat{H}(\omega)$. We shall need the following formulas for noise covariances $n(\omega_j)$, $n(\omega_e)$; $m(\omega_j)$, $m(\omega_e)$:

$$B_{je} = En(\omega_j)\, n^*(\omega_e) = \int_{-T}^{T}\!\!\int B(t-s) \exp(i\omega_j t - i\omega_e t)\, dt\, ds$$

$$= T^2 \int_{-\infty}^{\infty} \beta(\omega)\, W_T(\omega-\omega_j)\, W_T(\omega-\omega_e)\, d\omega$$

$$R_{je} = Em(\omega_j)\, m^*(\omega_e) = \int_{-T}^{T}\!\!\int R(t-s) \exp(i\omega_j t - i\omega_e t)\, dt\, ds$$

$$= T^2 \int_{-\infty}^{\infty} \rho(\omega)\, W_T(\omega-\omega_j)\, W_T(\omega-\omega_e)\, d\omega,$$

where

$$W_T(\omega) = \frac{\sin \omega T}{\omega T}$$

and $\beta(\omega)$, $\rho(\omega)$ are Fourier transforms of $B(t)$, $R(t)$. Supposing we have m.l.e.'s $\hat{H}_1 = \hat{H}(\omega_1), \ldots, \hat{H}_M = \hat{H}(\omega_M)$ we want to smooth them using some complex weights W_{kj}:

$$\tilde{H}(\omega_k) = \tilde{H}_k = \sum_{j=1}^{M} W_{kj} \hat{H}_j, \qquad k = 1, \ldots, M. \tag{28}$$

The mean-square deviation of the smoothed function $\tilde{H}_k$ from the true value $H_k = H(\omega_k)$ is

$$E|\tilde{H}_k - H_k|^2 = \left|\sum_{j=1}^{M} W_{kj} H_j - H_k\right|^2 + \sum_{j,e=1}^{M} W_{kj} W_{ke}^* b_{je}, \tag{29}$$

where $\{b_{je}\}$ stands for the asymptotic covariance of the m.l.e.'s $\hat{H}_j$ and $\hat{H}_e$. The first term on the right-hand side of (29) is the systematic bias of the smoothed estimator $\tilde{H}_k$; the second one is the variance of the random deviation of $\tilde{H}_k$. What are the weights W_{kj} minimizing (29)? After some algebraic computations we get:

$$W_{kj} = H_k \sum_{e=1}^{M} \beta_{je}^* H_e^*, \tag{30}$$

where $\{\beta_{je}\}$ is the inverse of the matrix

$$\{b_{je} + H_j H_e^*\}, \qquad j, e = 1, \ldots, M. \tag{31}$$

It may be noticed by the way that similarly we could minimize an arbitrary linear combination of the two terms on the right-hand side of (29). In the same way one can find the weights for smoothing amplitudes $\hat{r}_1 = |\hat{H}_1|, \ldots, \hat{r}_M = |\hat{H}_M|$ and/or phases $\hat{\theta}_1 = \arg \hat{H}_1, \ldots, \hat{\theta}_M = \arg \hat{H}_M$ if one is interested separately in amplitudes and/or phases. In that case it is more natural to take real weights. Let us consider for example the smoothing of phases with real weights v_{kj}:

$$\tilde{\theta}(\omega_k) = \tilde{\theta}_k = \sum_{j=1}^{M} v_{kj} \hat{\theta}_j, \qquad k = 1, \ldots, M. \tag{32}$$

The optimum weights minimizing the mean-square deviation are defined by

$$v_{kj} = \theta_k \sum_{e=1}^{M} \gamma_{je}\theta_e, \tag{33}$$

where $\{\gamma_{je}\}$ is the inverse of the matrix

$$\{C_{je}+\theta_j\theta_e\} \tag{34}$$

and $\{C_{je}\}$ is the asymptotic covariance-matrix of the m.l.e.'s $\hat{\theta}_j$ and $\hat{\theta}_e$. Now we give approximate formulas for the asymptotic covariances of the m.l.e.'s

$$\hat{H}_j; \quad \hat{r}_j = |\hat{H}_j|; \quad \hat{\theta}_j = \arg \hat{H}_j.$$

In these formulas indices j, e refer (as before) to the frequencies ω_j, ω_e respectively.

$$\operatorname{cov}(\hat{r}_j, \hat{r}_e) \cong \frac{r_j r_e}{NC_{jj}C_{ee}(r_j^2+z_j)(r_e^2+z_e)}\Big[R_{je}^2+2R_{je}\,Re\,(H_jH_e^*C_{je})+z_jz_eB_{je}^2$$

$$+2z_jz_eB_{je}\,Re\,C_{je}\Big] + \frac{\cos(\theta_j-\theta_e)(r_j^2-z_j)(r_e^2-z_e)}{2NC_{jj}C_{ee}(r_j^2+z_j)(r_e^2+z_e)}$$

$$\times [Re\,(R_{je}+H_jH_e^*B_{je})\,C_{je}+B_{je}R_{je}], \tag{35}$$

$$\operatorname{cov}(\hat{\theta}_j, \hat{\theta}_e) \cong \frac{\cos(\theta_j-\theta_e)}{2Nr_jr_eC_{jj}C_{ee}}[Re\,(R_{je}+H_jH_e^*B_{je})\,C_{je}+B_{je}R_{je}], \tag{36}$$

$$\operatorname{cov}(\hat{r}_j, \hat{\theta}_j) \cong \frac{\sin(\theta_e-\theta_j).(r_j^2-z_j)}{2Nr_eC_{jj}C_{ee}(r_j^2+z_j)}[Re\,(R_{je}+H_jH_e^*B_{je})\,C_{je}+B_{je}R_{je}], \tag{37}$$

$$\operatorname{cov}(\hat{H}_j, \hat{H}_e) \cong \frac{H_jH_e^*}{r_jr_e}\operatorname{cov}(\hat{r}_j, \hat{r}_e) - \frac{H_jH_e^*}{r_j}\operatorname{cov}(\hat{r}_j, \hat{\theta}_e)$$

$$- \frac{H_j^*H_e}{r_e}\operatorname{cov}(\hat{\theta}_j, \hat{r}_e)+H_j^*H_e\operatorname{cov}(\hat{\theta}_j, \hat{\theta}_e), \tag{38}$$

where

$$C_{je} = \frac{1}{N}\sum_{k=1}^{N} s_k(\omega_j)\,s_k^*(\omega_e). \tag{39}$$

Knowing the covariance matrices $\{\operatorname{cov}(\hat{H}_j, \hat{H}_e)\}$, $\{\operatorname{cov}(\hat{r}_j, \hat{r}_e)\}$, $\{\operatorname{cov}(\hat{\theta}_j, \hat{\theta}_e)\}$ it is easy to construct a confidence region for a set $H_1, \ldots, H_M$ (or for $r_1, \ldots, r_M$ or for $\theta_1, \ldots, \theta_M$). The sum

$$\sum_{j,e=1}^{M} \mathscr{X}_{je}(H_j-\hat{H}_j)(H_e-\hat{H}_e)^* \tag{40}$$

has a χ^2-distribution with $2M$ degrees of freedom. Here $\mathscr{X}_{je}$ stands for the inverse matrix of $\{\operatorname{cov}(\hat{H}_j, \hat{H}_e)\}$. The confidence region with confidence level γ for $H_1, \ldots, H_M$ is

$$\sum_{j,e=1}^{M} \mathscr{X}_{je}(H_j-\hat{H}_j).(H_e-\hat{H}_e)^* \leqslant \chi_{2M}^2(\gamma). \tag{41}$$

For a set of phases $\theta_1, \ldots, \theta_M$ we get a similar confidence region

$$\sum_{j,e=1}^{M} \lambda_{je}(\theta_j-\hat{\theta}_j)(\theta_e-\hat{\theta}_e) \leqslant \chi_M^2(\gamma), \tag{42}$$

where $\{\lambda_{je}\}$ is the inverse matrix of $\{\text{cov}(\hat{\theta}_j, \hat{\theta}_e)\}$. The sum in (42) has M degrees of freedom, because all its terms are real, while the sum (41) has $2M$ degrees of freedom, because all its terms are complex. For $r_1, \ldots, r_M$ we have a confidence region

$$\sum_{j,e=1}^{M} \mu_{je}(r_j - \hat{r}_j)(r_e - \hat{r}_e) \leqslant \chi_M{}^2(\gamma), \tag{43}$$

where $\{\mu_{je}\}$ is the inverse matrix of $\{\text{cov}(\hat{r}_j, \hat{r}_e)\}$. The formulas (35)–(39) contain the unknown quantities $H_1, \ldots, H_M, C_{je}$. One can substitute for them in (41)–(43), (30), (33) their consistent estimators $\hat{H}_1, \ldots, \hat{H}_M$ and

$$\hat{C}_{je} = \frac{1}{2N} \sum_{k=1}^{N} \left(\frac{x_k{}^{(j)} y_k{}^{*(e)}}{\hat{H}_e^*} + \frac{x_k{}^{*(e)} y_k{}^{(j)}}{\hat{H}_j} \right), \tag{44}$$

where $x_k{}^{(j)}$, $y_k{}^{(j)}$ are the values of the spectra of the k-th earthquake at the j-th frequency ω_j for the first and second stations respectively.

The problem just stated can be generalized to the case when there is an array of $(k+1)$ stations and we want to determine the relative amplitude and phase corrections $H_1(\omega), \ldots, H_k(\omega)$ of the k stations with respect to one central station. For simplicity we shall consider such a generalization only for one frequency ω. Let $x_i(\omega)$ stand for the Fourier-spectrum of the k-th earthquake at the central station and $y_i{}^{(1)}(\omega), \ldots, y_i{}^{(k)}(\omega)$ for the Fourier spectra of the i-th earthquake at the rest of the k stations. The likelihood function for a fixed frequency ω is

$$\log L = C - \frac{1}{2\sigma^2} \|x - s\|^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{k} \frac{\|y^{(j)} - H_j s\|^2}{z_j}, \tag{45}$$

where C is a constant, and the other notations are as above. We have to maximize L over all complex vectors $s = (s_1, \ldots, s_N)$ and complex numbers $H_1, \ldots, H_k$. It can be shown that this problem is equivalent to the problem of the determination of the normalized eigen-vector of the Hermitian matrix

$$g_{je} = \sum_{i=1}^{k} y_j{}^{(i)}(y_e{}^{(i)})^*(z_j z_e)^{-\frac{1}{2}} + x_j x_e{}^*, \qquad j, e = 1, \ldots, N \tag{46}$$

corresponding to the largest eigen-value. Denoting this vector by u_e we get the m.l.e. for $H_k(\omega)$:

$$\hat{H}_j(\omega) = \sum_{e=1}^{N} y_e{}^{(j)} u_e{}^* z_e{}^{-1}, \qquad j = 1, \ldots, k. \tag{47}$$

It should be noted that an estimator of $H_j(\omega)$ can be got by means of formulas (22), (23) and using only the spectra of the central and j-th stations $x_i(\omega)$, $y_i{}^{(j)}(\omega)$, $i = 1, \ldots, N$. The latter estimator would be worse (its mean square deviation is larger) than the estimator (47).

3.2 *Group and phase velocities of surface waves*

Let us consider now an application of the m.l.e. method to the determination of group and phase velocities of surface waves. We suppose that from a preliminary analysis [e.g. by means of the very efficient procedures developed by M. Landisman *et al.*, (17)–(18)] it is known that a seismogram $x(t)$, $(-T \leqslant t \leqslant T)$ contains M modes of surface waves of some type. Let $\xi_k(\omega)$ denote the spatial frequency of the k-th mode. It is well known that the group and phase velocities of the k-th mode are defined by the equations:

$$1/v_k(\omega) = \frac{d\xi_k(\omega)}{d\omega}; \qquad C_k(\omega) = \frac{\omega}{\xi_k(\omega)}. \tag{48}$$

We now introduce for every function $\xi_k(\omega)$ its own parametrisation, i.e. we assume that $\xi_k(\omega)$ can be represented with sufficient accuracy in the form

$$\xi_k(\omega) = \sum_{j=1}^{m_k} \alpha_{kj}\phi_j(\omega), \tag{49}$$

where $\phi_j(\omega)$ are known functions (for example $\phi_j(\omega) = \omega^j$), and α_{kj} are unknown coefficients. Having $\xi_k(\omega)$ in the form (49) it is possible to calculate by means of known methods a synthetic seismogram $y_k(t;\ \alpha_{kj})$ for the k-th mode. The mode $y_k(t;\ \alpha_{kj}, \beta_{kj})$ can depend also on some incidental parameters β_{kj} corresponding to the source and the medium. Now we assume that the observed seismogram $x(t)$ has the form

$$x(t) = \sum_{k=1}^{M} y_k(t;\ \alpha_{kj}, \beta_{kj}) + n(t), \qquad -T \leqslant t \leqslant T, \tag{50}$$

where $n(t)$ is Gaussian noise with correlation function $B(t, s)$ and zero mean. Let $x(t)$ be digitalized in points $t_1, \ldots, t_N$. Then the likelihood function for the unknown parameters $\{\alpha_{kj}, \beta_{kj}\}$ is

$$\log L = C - \tfrac{1}{2} \sum_{e, j=1}^{N} \gamma_{ej} \left[x(t_e) - \sum_{k=1}^{M} y_k(t_e;\ \alpha_{kl}\ \beta_{kl})\right] [x(t_j) - \sum_{k=1}^{M} y_k\,(t_j;\ \alpha_{kl}, \beta_{kl})], \tag{51}$$

where C is a constant and $\{\gamma_{ej}\}$ is the inverse matrix of $\{B(t_e, t_j)\}$.

Now one can apply the standard method of m.l.e. stated above and find confidence regions for $\{\alpha_{kj}\}$, and consequently for $v_k(\omega)$, $C_k(\omega)$. However, finding the maximum of L in (57) is very difficult from the computational point of view. Therefore it is reasonable to attempt to apply the m.l.e. method not directly to the seismogram, but to functions obtained from the seismogram after some processing. Such an attempt is the method of the contour diagram in the period-velocity plane for the determination of group velocity, due to M. Landisman *et al.* [see **(14)–(17)**]. In this method a function of two variables $y(t, \omega)$ is obtained from a seismogram $x(t)$. Here ω is the frequency of interest and t is the 'arrival time of period $2\pi/\omega$'; the value of t is proportional to the inverse of the group velocity $v(\omega)$ since $v(\omega) = R/t(\omega)$, where R is the distance from the source to the station. We assume now that our seismogram $x(t)$ contains only one mode and we try to derive a confidence region for $v(\omega_1), \ldots, v(\omega_M)$. We shall consider a so-called frequency version of the determination of contour diagrams. In this version the complex function $y(\tau, \omega)$ is calculated:

$$y(\tau, \omega) = \frac{1}{\pi} \int_0^\infty X(\lambda).H\left(\frac{\omega-\lambda}{\omega}\right) \exp(-i\lambda\tau)\, d\lambda, \tag{52}$$

where $X(\lambda)$ is the Fourier transform of $x(t)$ over the time interval in question $(-T \leqslant t \leqslant T)$; $H(\omega)$ is the 'frequency window'. M. Landisman *et al.* (cit. above) took the window in the form:

$$H(\omega) = \exp(-\alpha\omega^2). \tag{53}$$

The values of $|y(\tau, \omega)|$ are inscribed in the plane (τ, ω) for some prescribed grid of arguments (τ_j, ω_k); then contour lines of constant levels of $|y(\tau, \omega)|$ are drawn and the crest line is marked. We assume that the spectrum of the observed seismogram has the form:

$$X(\omega) = s(\omega) + n(\omega), \tag{54}$$

where $s(\omega)$ is the Fourier transform of the mode of surface wave under consideration and $n(\omega)$ is the Fourier transform of a realisation of Gaussian noise with spectral power density $\phi(\omega)$. Both transforms are taken over time interval $(-T, T)$. We have

$$|y(\tau,\omega)|^2 = \frac{1}{\pi^2}\left\{\int_0^\infty [\operatorname{Re} s(\lambda)\cos\lambda\tau + \operatorname{Im} s(\lambda)\sin\omega\tau].H\left(\frac{\omega-\lambda}{\omega}\right)d\lambda\right.$$
$$\left. + \int_0^\infty [\operatorname{Re} n(\lambda)\cos\lambda\tau + \operatorname{Im} n(\lambda)\sin\lambda\tau].H\left(\frac{\omega-\lambda}{\omega}\right)d\lambda\right\}^2$$
$$+ \frac{1}{\pi^2}\left\{\int_0^\infty [-\operatorname{Re} s(\lambda)\sin\lambda\tau + \operatorname{Im} s(\lambda)\cos\lambda\tau].H\left(\frac{\omega-\lambda}{\omega}\right)d\lambda\right.$$
$$\left. + \int_0^\infty [-\operatorname{Re} n(\lambda)\sin\lambda\tau + \operatorname{Im} n(\lambda)\cos\lambda\tau].H\left(\frac{\omega-\lambda}{\omega}\right)d\lambda\right\}^2. \tag{55}$$

Let us denote the first, second, third and fourth integrals on the right-hand side of (55) by $f_1(\tau,\omega)$, $n_1(\tau,\omega)$, $f_2(\tau,\omega)$, $n_2(\tau,\omega)$ respectively. Further, let $\tau_0(\omega)$ stand for that value of τ where the function $f_1^2(\tau,\omega)+f_2^2(\tau,\omega)$ achieves its maximum, and let $\tau(\omega)$ stand for that value of τ where the function $|y(\tau,\omega)|^2$ achieves its maximum over τ. Now we state an essential assumption about the smallness of the noise: we assume that derivatives $\partial n_1/\partial\tau$, $\partial n_2/\partial\tau$ are so small that in the vicinity of $\tau_0(\omega)$

$$\tau_0(\omega)-|\tau_0(\omega)-\tau(\omega)| < \tau < \tau_0(\omega)+|\tau_0(\omega)-\tau(\omega)|$$

the function $f_1^2(\tau,\omega)+f_2^2(\tau,\omega)$, with probability close to 1, can be approximated with sufficient accuracy by a parabola of the second degree in $(\tau_0-\tau)$. In such a case it is easy to show that

$$\tau(\omega) \cong \tau_0(\omega) - \frac{f_1(\tau_0,\omega).n_1'(\tau,\omega)+f_2(\tau_0\,\omega).n_2'(\tau,\omega)}{\frac{\partial^2}{\partial\tau^2}[f_1^2(\tau,\omega)+f_2^2(\tau,\omega)]|_{\tau=\tau_0}}. \tag{56}$$

Thus the presence of noise results in shifting the argument of the maximum of $|y(\tau,\omega)|^2$ over τ, as compared to the argument of the maximum of $f_1^2+f_2^2$ (when the noise is absent), by some random value, which is equal to the second term on the right-hand side of (56). Moreover $\tau_0(\omega)$ itself has a bias relative to the function $R/v(\omega_k)$. This bias $b(\omega) = \tau_0(\omega)-R/v(\omega)$ can be evaluated by means of substituting for $s(\omega)$ in $f_1(\tau,\omega)$ and $f_2(\tau,\omega)$ some proper theoretical spectrum of the mode in question and comparing $\tau_0(\omega)$ and $R/v(\omega)$.

Let us denote the random term on the right-hand side of (56) by $\varepsilon(\omega)$, and the covariance cov $[\varepsilon(\omega_k), \varepsilon(\omega_j)]$ by d_{kj}. Now we are able to write the likelihood function for $v(\omega_1), \ldots, v(\omega_M)$,

$$\log L = C - \tfrac{1}{2}\sum_{k,j=1}^{M}\gamma_{kj}\left[\tau(\omega_k)-b(\omega_k)-\frac{R}{v(\omega_k)}\right]\left[\tau(\omega_j)-b(\omega_j)-\frac{R}{v(\omega_j)}\right], \tag{57}$$

where C is a constant, $\{\gamma_{kj}\}$ is the inverse matrix of $\{d_{kj}\}$. There is no difficulty of principle in finding covariances d_{kj}. It ought to be taken into account only that the noise spectra $\operatorname{Re} n(\omega)$ and $\operatorname{Im} n(\omega)$ constituting $n_1'(\tau,\omega)$, $n_2'(\tau,\omega)$ can be expressed in the following form [see **(19)**]:

$$\operatorname{Re} n(\omega) = \int_0^\infty dg(\lambda)\left[\frac{\sin(\lambda-\omega)T}{\lambda-\omega}+\frac{\sin(\lambda+\omega)T}{\lambda+\omega}\right], \tag{58}$$

$$\operatorname{Im} n(\omega) = \int_0^\infty dh(\lambda) \left[\frac{\sin(\lambda-\omega)T}{\lambda-\omega} - \frac{\sin(\lambda+\omega)T}{\lambda+\omega} \right], \tag{59}$$

where $dg(\lambda)$, $dh(\lambda)$ are uncorrelated stochastic spectral measures with orthogonal increments, connected with the spectral power density $\phi(\lambda)$ by the following relations:

$$E|dg(\lambda)|^2 = E|dh(\lambda)|^2 = \phi(\lambda)\, d\lambda. \tag{60}$$

Knowing the likelihood function (57) one can construct a confidence region for $v(\omega_1), \ldots, v(\omega_M)$ because the quadratic form on the right-hand side of (57) has a χ^2 distribution with M degrees of freedom. We point out once more that expression (57) has been derived under the assumption of smallness of the noise. But if the noise is not small then it seems to be unreasonable to use the method of the contour diagram at all.

Similar considerations can be used for the determination of the interstation phase velocity based on two records $x_1(t)$ and $x_2(t)$ at two stations. The phase difference $\Delta\phi(\omega)$ in this case is:

$$\Delta\phi(\omega) = \arg y_2(\tau, \omega) - \arg y_1(\tau, \omega) \tag{61}$$

and an estimator of the phase velocity for frequency ω is

$$\hat{C}(\omega) = \frac{\omega \Delta R}{2\pi \Delta\phi(\omega)}, \tag{62}$$

where ΔR is the distance between the stations. As for the case of the group velocity under the condition of smallness of the noise, the disturbed functions $\arg y_1(\tau, \omega)$, $\arg y_2(\tau, \omega)$ can be linearized and a confidence region for values of the phase velocity $C(\omega_1), \ldots, C(\omega_M)$ can be constructed. At the end of this paragraph we shall consider the construction of confidence regions for group velocity, obtained by the method of peaks and troughs [see (**18**)].

Let the arrival times of visible periods $T_1, \ldots, T_M$ on a seismogram $x(t)$ be $t_1, \ldots, t_M$. We assume that the noise $n(t)$ is small enough, so that it shifts the peak of period T_k insignificantly to a value Δt_k and that in the vicinity of t_k the signal peak can be approximated with sufficient accuracy by a parabola of the second degree. Denoting the true time of arrival of the peak of T_k by $\bar{t}_k$ we can write:

$$x(t) = s(t) + n(t), \qquad x'(t_k) = s'(t_k) + n'(t_k) = 0, \tag{63}$$

$$s'(t_k) \cong s(\bar{t}_k) + s'(\bar{t}_k) . (t_k - \bar{t}_k) + s''(\bar{t}_k) . \frac{(t_k - \bar{t}_k)^2}{2}. \tag{64}$$

Substituting (64) into (63) and taking into account that $s'(\bar{t}_k) = 0$ we get

$$t_k = \bar{t}_k - \frac{n'(t_k)}{s''(\bar{t}_k)}. \tag{65}$$

Denoting the covariance of the derivative of the noise $n'(t_k)$ at the points t_k, t_j by d_{kj}, we obtain the likelihood function for $v(T_1), \ldots, v(T_M)$,

$$\log L = C - \tfrac{1}{2} \sum_{k, j=1}^{M} \gamma_{kj} \left[t_k - \frac{R}{v(T_k)} \right] . \left[t_j - \frac{R}{v(T_j)} \right] . s''(\bar{t}_k) s''(\bar{t}_j), \tag{66}$$

where C is a constant and $\{\gamma_{kj}\}$ is the inverse matrix of $\{d_{kj}\}$. A confidence region for $v(T_1), \ldots, v(T_M)$ can be easily derived from (66), because the quadratic form on the right-hand side of (66) has a χ^2-distribution with M degrees of freedom. The curvatures $s''(\bar{t}_k)$ in (66) ought to be estimated by some proper procedure.

If the times $t_1, \ldots, t_M$ are smoothed by means of weights W_{kj}:

$$\hat{t}_k = \sum_{j=1}^{M} W_{kj} t_j \tag{67}$$

then the covariances of the smoothed values $\hat{t}_1, \ldots, \hat{t}_m$ are equal to

$$\operatorname{cov}(\hat{t}_k \hat{t}_{kj}) = \sum_{e,m=1}^{M} W_{kj} W_{jm} d_{em} s''(\bar{t}_e).s''(\bar{t}_m). \tag{68}$$

The systematic bias $\hat{t}_k$ as compared with true value $\bar{t}_k$ is

$$E\hat{t}_k - \bar{t}_k = \sum_{j=1}^{M} W_{kj} \bar{t}_j - \bar{t}_k. \tag{69}$$

The weights minimizing the mean-square deviation $E(\hat{t}_k - \bar{t}_k)^2$ are specified by

$$\sum_{j=1}^{M} W_{kj} [d_{jm} s''(\bar{t}_j) s''(\bar{t}_m) + \bar{t}_j . \bar{t}_m] = \bar{t}_k . \bar{t}_m; \qquad m = 1, \ldots, M. \tag{70}$$

3.3 *Derivative of travel-time curve $dt/d\Delta$*

We are given arrival times of N earthquakes at two stations, $t_1^{(1)}, \ldots, t_N^{(1)}$; $t_1^{(2)}, \ldots, t_N^{(2)}$, and corresponding epicentral distances $\Delta_1^{(1)}, \ldots, \Delta_N^{(1)}$; $\Delta_1^{(2)}, \ldots, \Delta_N^{(2)}$. These values have been measured with errors. We denote the corresponding exact values by

$$\tau_1^{(1)}, \ldots, \tau_N^{(1)}; \quad \tau_1^{(2)}, \ldots, \tau_N^{(2)}; \quad \delta_1^{(1)}, \ldots, \delta_N^{(1)}; \quad \delta_1^{(2)}, \ldots, \delta_N^{(2)}.$$

We suppose further that the following relations hold true:

$$t_k^{(2)} - t_k^{(1)} = \tau_k^{(2)} - \tau_k^{(1)} + \xi_k; \tag{71}$$

$$\frac{\Delta_k^{(2)} + \Delta_k^{(1)}}{2} = \frac{\delta_k^{(2)} + \delta_k^{(1)}}{2} + \eta_k; \qquad k = 1, \ldots, N \tag{72}$$

$$\Delta_k^{(2)} - \Delta_k^{(1)} = \delta_k^{(2)} - \delta_k^{(1)} + \zeta_k; \tag{73}$$

where ξ_k, η_k, ζ_k are independent Gaussian variables with zero mean and variances f^2, g^2, h^2 respectively. It is to be noted that, although the left sides of (72), (73) are functions of the same random variables $\Delta_k^{(2)}$, $\Delta_k^{(1)}$, the assumption of independence of η_k and ζ_k is nevertheless quite natural. It is due to the fact that the sum and the difference of random variables with equal variances are always uncorrelated. We suppose next that in the range of epicentral distances of interest the true derivative of the travel-time curve, $dt/d\Delta$, is represented with sufficient accuracy by a parabola of the m-th degree:

$$\frac{dt}{d\Delta} = \sum_{e=0}^{m} a_e \Delta^e. \tag{74}$$

Lastly we suppose that our stations are so close to each other that the following approximation is true:

$$\frac{\tau_k^{(2)} - \tau_k^{(1)}}{\delta_k^{(2)} - \delta_k^{(2)}} = \sum_{e=0}^{m} a_e \left(\frac{\delta_k^{(2)} + \delta_k^{(1)}}{2}\right)^e. \tag{75}$$

Now we put down the likelihood function:

$$\log L = C - \frac{1}{2f^2}\sum_{k=1}^{N}\left[t_k^{(2)} - t_k^{(1)} - (\delta_k^{(2)} - \delta_k^{(1)}).\sum_{e=0}^{m} a_e \left(\frac{\delta_k^{(2)}+\delta_k^{(1)}}{2}\right)^e\right]^2$$

$$-\frac{1}{2g^2}\sum_{k=1}^{N}\left[\frac{\Delta_k^{(2)}+\Delta_k^{(1)}}{2} - \frac{\delta_k^{(2)}+\delta_k^{(1)}}{2}\right]^2 - \frac{1}{2h^2}\sum_{k=1}^{N}[\Delta_k^{(2)} - \Delta_k^{(1)} - \delta_k^{(2)} + \delta_k^{(1)}]^2. \quad (76)$$

One can apply the maximum likelihood method to (76) to estimate the unknown parameters $a_0, \ldots, a_m$ and the incidental parameters $\delta_k^{(2)} - \delta_k^{(1)}$, $\frac{1}{2}(\delta_k^{(2)} + \delta_k^{(1)})$, $k = 1, \ldots, N$. It should be noted that in the case $m > 1$ the m.l.e. can be biased.

An alternative approach to this problem could consist in using some *a priori* density for the incidental parameters $\delta_k^{(2)} - \delta_k^{(1)}$, $\frac{1}{2}(\delta_k^{(2)} + \delta_k^{(1)})$, say $\phi(x)$, $\psi(y)$. Then the density of

$$t_k^{(2)} - t_k^{(1)};\quad \Delta_k^{(2)} - \Delta_k^{(1)};\quad \delta_k^{(2)} - \delta_k^{(1)};\quad \tfrac{1}{2}(\delta_k^{(2)} + \delta_k^{(1)});\quad \tfrac{1}{2}(\Delta_k^{(2)} + \Delta_k^{(1)})$$

is

$$C\exp\left\{-\frac{1}{2f^2}\left[t_k^{(2)} - t_k^{(1)} - (\delta_k^{(2)} - \delta_k^{(1)}).\sum_{e=0}^{m} a_e\left(\frac{\delta_k^{(2)}+\delta_k^{(1)}}{2}\right)^e\right]^2\right.$$

$$-\frac{1}{2g^2}\left[\frac{\Delta_k^{(2)}+\Delta_k^{(1)}}{2} - \frac{\delta_k^{(2)}+\delta_k^{(1)}}{2}\right]^2$$

$$\left.-\frac{1}{2h^2}[\Delta_k^{(2)} - \Delta_k^{(1)} - \delta_k^{(2)} + \delta_k^{(1)}]^2\right\}.\phi(\delta_k^{(2)} - \delta_k^{(1)}).\psi\left(\frac{\delta_k^{(2)}+\delta_k^{(1)}}{2}\right). \quad (77)$$

Hence the conditional density of $t_k^{(2)} - t_k^{(1)}$ under fixed $\Delta_k^{(2)} - \Delta_k^{(1)}$, $\frac{1}{2}(\Delta_k^{(2)} + \Delta_k^{(1)})$ is

$$W(t_k^{(2)} - t_k^{(1)}/\Delta_k^{(2)} - \Delta_k^{(1)};\ \tfrac{1}{2}(\Delta_k^{(2)} + \Delta_k^{(1)}))$$

$$= \frac{\displaystyle\iint_{-\infty}^{\infty}\exp\left\{-\frac{1}{2f^2}\left[t_k^{(2)} - t_k^{(1)} - x\sum_{e=0}^{m} a_e y^e\right]^2 - \frac{1}{2g^2}\left[\frac{\Delta_k^{(2)}+\Delta_k^{(1)}}{2} - y\right]^2 - \frac{1}{2h^2}[\Delta_k^{(2)} - \Delta_k^{(1)} - x]^2\right\}.\phi(x).\psi(y)\,dx\,dy}{\displaystyle\iiiint_{-\infty}^{\infty}\exp\left\{-\frac{1}{2f^2}\left[t_k^{(2)} - t_k^{(1)} - x\sum_{e=0}^{m} a_e y^e\right]^2 - \frac{1}{2g^2}[u-y]^2 - \frac{1}{2h^2}[v-x]^2\right\}.\phi(x).\psi(y)\,dx\,dy\,du\,dv} \quad (78)$$

Taking the product of

$$W\left(t_k^{(2)} - t_k^{(1)}/\Delta_k^{(2)} - \Delta_k^{(1)};\ \frac{\Delta_k^{(2)}+\Delta_k^{(1)}}{2}\right)$$

over all k we get the conditional density of all $t_k^{(2)} - t_k^{(1)}$, $k = 1, \ldots, N$ under fixed

$$\Delta_k^{(2)} - \Delta_k^{(1)},\quad \frac{\Delta_k^{(2)} - \Delta_k^{(1)}}{2},\quad k = 1, \ldots, N:$$

$$\prod_{k=1}^{N} W\left(t_k^{(2)} - t_k^{(1)}/\Delta_k^{(2)} - \Delta_k^{(1)};\ \frac{\Delta_k^{(2)}+\Delta_k^{(1)}}{2}\right). \quad (79)$$

The values of $a_0, \ldots, a_m$ giving a maximum to (79) are the conditional m.l.e.'s. A confidence region for $a_0, \ldots, a_m$ can be derived with the help of the density (79) as it was done above in the second paragraph.

Problems of such kinds are very difficult from the computational point of view. They are related to the so-called confluence analysis [see **(20)**, **(21)**], which differs from regression analysis in the fact that there are errors both in the values of observed function and in the values of its arguments.

4. Conclusion

In this paper the standard method of m.l.e. has been stated, including the construction of confidence regions based upon m.l.e. This method has been applied to derive formulas for m.l.e.'s in some seismological problems: determination of relative amplitude and phase corrections; determination of group and phase velocities of surface waves; estimation of the derivative of the travel-time curve. The main theoretical difficulty when applying the method of maximum likelihood consists in finding an adequate probabilistic setting of the problem and in finding a likelihood function for the parameters of the problem. This difficulty is not always easy to overcome. For example, so far no reasonable likelihood function has been found for arrival times corresponding to a multiloop travel-time curve.

When using the method of m.l.e. in practice one encounters also the real difficulty of computation (determination of the maximum of a function of many variables). Sometimes this difficulty is aggravated by the presence of many incidental parameters. Nevertheless the application of the method of maximum likelihood either directly to observed seismograms or to some 'semi-manufactured' functions derived from seismograms leads in practice to good results.

Institute of Physics of the Earth,
Moscow.

References

(1) Capon, J., Greenfield, R. & Kolker, R., 1967. Multidimensional maximum likelihood processing of a large aperture seismic array, *Proc. Instn elect. electron. Eng.*, **55**, 192–211.

(2) Vilkovich, E., Molchan, G. & Keilis-Borok, V. Statistical questions of estimation of surface effects connected with seismicity (in Russian), Vychislitelnaya Seismologia, Issue No. 5 (in press).

(3) Goltzman, F., Zolotukhina, L., Latyshev, K., Khalfin, L., Khalfina, N. & Chuguyeva, V., 1965. Statistical problems in interpretation of seismic data (in Russian), *Trudy Matem. Inst. AN SSSR*, Vol. LXXIX, 160–181.

(4) Jelankina, T., Keilis-Borok, V., Pisarenko, V. & Pyatetzky-Shapiro, I. Earthquake mechanism solution on a computer (in Russian), Vychislitelnaya Seismologia, Issue No. 5 (in press).

(5) Flinn, E., 1965. Confidence regions and error determinations for seismic event location, *Rev. Geophys.*, **3**, 157–185.

(6) Rastrigin, L., 1968. *Statistical methods of Search* (in Russian), Moscow.

(7) Rao, C., 1965. *Linear statistical inference and its applications*, New York–London–Sydney.

(8) Van Der Waerden, B., 1967. *Mathematische Statistik*, Berlin–Göttingen–Heidelberg.

(9) Cramer, H., 1946. *Mathematical methods of statistics*, Princeton University Press, Princeton.

(10) Anderson T., 1957. *An introduction to multi-variate statistical analysis*, New York.

(11) Linnik, Yu, 1962. *The least square method and the foundations of the theory of data processing* (in Russian), Moscow.

(12) Pisarenko, V. Statistical estimation of amplitude and phase corrections, *Geophys. J. R. astr. Soc.*, **20**, 89.

(13) Jelankina, T. & Pisarenko, V. Amplitude and phase corrections based on data of various accuracy (in Russian), Vychislitelnaya Seismologia (in press).

(14) Dziewonski, A., Landisman, M., Bloch, S., Satô, Y. & Asano, S., 1968. Progress report on recent improvement in the analysis of surface wave observations, *J. Phys. Earth*, **19**, Spec. Issue, 1–26.

(15) Landisman, M., Dziewonski, A. & Satô, Y. Recent improvements in the analysis of surface wave observations, *Geophys. J. R. astr. Soc.* (in press).

(16) Dziewonski, A., Bloch, S. & Landisman, M., 1969. A technique for the analysis of transient seismic signals, *Bull. seism. Soc. Am.*, **59**, 427–444.

(17) Bloch, S. & Hales, A., 1969. New techniques for the determination of surface wave phase velocities, *Bull. seism. Soc. Am.*, **58**, 1021–1034.

(18) Ewing, M., Jardetzky, F. & Press, F., 1957. *Elastic waves in layered media*, McGraw-Hill Book Co.

(19) Cramer, H. & Leadbetter, M., 1967. *Stationary and related stochastic processes*, New York–London.

(20) Frish, R., 1934. *Statistical confluence analysis by means of complete regression system*, Oslo.

(21) Klepikov, N. & Sokolov, S., 1964. *The analysis and design of experiments by means of maximum likelihood method* (in Russian), Moscow.