where w_1 and w_2 are unknown. The hypothesis $H_0(w_1=w_2=w)$, where w is a specified number, can be tested by observing the variable $\zeta(n_1,n_2)$, defined by (4), which — if the assumption (c) and the hypothesis H_0 are satisfied — is asymptotically normal

$$N\left[\frac{n_2-n_1}{(n_1+n_2)^p}\,w^{1-p};\ \frac{\sqrt{1-w}}{(n_1+n_2)^{p-0,5}}\,w^{0,5-p}\right].$$

EXAMPLE 4. The variable $\xi(\lambda)$ is distributed according to the negative binomial law, given by the formula

$$P[\xi(\lambda)=j]=(-1)^j\binom{-\lambda}{j}w^j(1-w)^\lambda,$$

where $j=0,1,2,\ldots,\ \lambda>0,\ 0<w<1$. Here we have

$$m(\lambda)=\frac{w\lambda}{1-w};\qquad \sigma^2(\lambda)=\frac{w\lambda}{(1-w)^2}.$$

The variable $\zeta(\lambda_1,\lambda_2)$ defined by (4) — if the assumption (c) is satisfied — is asymptotically normal

$$N\left[\frac{\lambda_2-\lambda_1}{(\lambda_1+\lambda_2)^p}\,\frac{w^{1-p}}{(1-w)^{1-p}};\ \frac{[(\lambda_1+\lambda_2)w]^{0,5-p}}{(1-w)^{1-p}}\right].$$

Our theorem can thus be applied in particular to testing parametric hypothesis concerning Pascal variables since they have a negative binomial distribution with an integer value λ.

STATISTICAL ESTIMATION OF PARAMETERS IN MARKOV PROCESSES

BY

O. LANGE (WARSZAWA)

1. Methods of estimation. Consider a simple Markov process with the transition function

$$f(t_0,x_0;t_k,x_k;\Theta_1,\Theta_2,\ldots). \tag{1.1}$$

The transition function expresses the conditional probability (for discrete processes), or the conditional probability density (for continuous processes), that the random variable $\xi(t)$ will assume the value x_k at the moment t_k if its value is x_0 at the moment t_0. This function contains certain parameters $\Theta_1,\Theta_2,\ldots$ the values of which have to be determined from statistical observation.

In Markov processes this can be done by the method of maximum likelihood, which consists in choosing the estimators of the parameters $\Theta_1,\Theta_2,\ldots$ so as to maximize the probability or probability density of an observed set of realizations of the stochastic process. The method of maximum likelihood can be applied in several ways.

If the realizations of the stochastic process can be repeated many times (as, for instance, in the laboratory or in industrial production) we take n independent realizations of the process and perform on each realization a pair of observations at the moments, say, $t_0^{(r)}$ and $t_k^{(r)}$. The superscript r stands for the r-th realization $(r=1,2,\ldots,n)$. Denote by $x_i^{(r)}$ the result of the observation carried out on the r-th realization at the moment $t_i^{(r)}$, where $i=0,k$. Since the pairs of observations are independent, their likelihood function is

$$L_1=\prod_{r=1}^{n}f(t_0^{(r)},x_0^{(r)};t_k^{(r)},x_k^{(r)};\Theta_1,\Theta_2,\ldots). \tag{1.2}$$

The estimators of $\Theta_1,\Theta_2,\ldots$, which will be denoted by $\hat\Theta_1,\hat\Theta_2,\ldots$, are determined from the condition $L_1=\max$.

This way of using the method of maximum likelihood will be called *cross section estimation*, or *space estimation* (over the space of realizations of the process). The estimators thus derived will be called *cross section* or *space estimators*.

icm©

In cases, however, where the realization of the stochastic process cannot be repeated (as, for instance, in meteorological processes, in processes of growth of human populations, in socio-economic processes) we have to use the method of maximum likelihood in a different way, which will be called *historical estimation* or *time series estimation*. The estimators thus obtained will be called, accordingly, *historical* or *time estimators*.

Historical or time series estimation consists in perfoming a number of observations on a single realization of the stochastic process. Let the observations be made at the moments $t_0^{(r)}, t_1^{(r)}, \ldots, t_k^{(r)}$. The results of the observations then form the time series $x_0^{(r)}, x_1^{(r)}, \ldots, x_k^{(r)}$. The superscript r serves to identify the realization on which the observations are performed; since, in this case, only one realization is accessible to observation, it may also be omitted. In view of the process being Markovian, the likelihood function of the observed time series is

$$L_2 = \prod_{i=1}^{k} f(t_{i-1}^{(r)}, x_{i-1}^{(r)}; t_i^{(r)}, x_i^{(r)}; \Theta_1, \Theta_2, \ldots). \tag{1.3}$$

The estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots$ are determined from the condition $L_2 = \max$.

Finally, situations may occur where it is possible to perform historical observations on a set of independent realizations of the same stochastic process. Let there be n such realizations and $k+1$ observations performed on each at the moments $t_0^{(r)}, t_1^{(r)}, \ldots, t_k^{(r)}$ respectively. Denoting, as before, by $x_i^{(r)}$ the result of the observation performed on the r-th realization at the moment $t_i^{(r)}$, we have the following observation matrix:

$$\begin{pmatrix} x_0^{(1)} & x_1^{(1)} & \ldots & x_k^{(1)} \\ x_0^{(2)} & x_1^{(2)} & \ldots & x_k^{(2)} \\ \ldots & \ldots & \ldots & \ldots \\ x_0^{(n)} & x_1^{(n)} & \ldots & x_k^{(n)} \end{pmatrix}. \tag{1.4}$$

The rows of the matrix are time series corresponding to the various realizations of the process, the columns are cross sections of observations performed on different realizations. It should be noted that the observations corresponding to a given column need not be simultaneous, for the $k+1$ observations performed on each realization may be effected on different realizations at different moments. Thus the $t_i^{(r)}$ corresponding to a given subscript i but to different superscripts r may be different.

The likelihood function of the above observation matrix is

$$L = \prod_{r=1}^{n} \prod_{i=1}^{k} f(t_{i-1}^{(r)}, x_{i-1}^{(r)}; t_i^{(r)}, x_i^{(r)}; \Theta_1, \Theta_2, \ldots). \tag{1.5}$$

The estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots$ are determined from the condition $L = \max$. This way of determining the estimators will be called *complete estimation* and the estimators thus obtained will be called *complete estimators*.

Complete estimators maximize the probability or probability density of the whole observation matrix (1.4), while historical and cross section estimators maximize only the probability or probability density of a particular row or column, respectively. Historical estimation and cross section estimation may thus be treated as special cases of complete estimation corresponding to $n=1$ and $k=1$, respectively. We shall, therefore, henceforth consider the general case of complete estimation.

In the present paper we shall consider the statistical estimation of parameters in the following elementary Markov processes: the simple Poisson process, the Gaussian process with stationary independent increments which is usually called the Brownian motion process, the linear "birth process" and the linear "death process". Finally we shall consider the case of estimating transition probabilities in simple Markov chains.

2. The simple Poisson process. For the simple Poisson process the transition function (1.1) takes the form

$$f(t_i - t_{i-1}, z_i; \lambda) = \frac{[\lambda(t_i - t_{i-1})]^{z_i}}{z_i!} \exp[-\lambda(t_i - t_{i-1})]. \tag{2.1}$$

Here

$$z_i = x_i - x_{i-1} \tag{2.2}$$

denotes the number of changes of state occuring during the period $t_i - t_{i-1}$; x_i and x_{i-1} denote the number of changes of state taking place during the periods from 0 to t_i and from 0 to t_{i-1}, respectively. Obviously x_i and x_{i-1} are integers. The number of changes of state expected during the period $t_i - t_{i-1}$ is

$$Ez_i = \lambda(t_i - t_{i-1}),$$

where λ is a constant. The simple Poisson process is a homogeneous process with independent increments.

The logarithm of the likelihood function is

$$\begin{aligned} \log L &= \sum_{r=1}^{n} \sum_{i=1}^{k} \log f(t_i^{(r)} - t_{i-1}^{(r)}, z_i^{(r)}; \lambda) \\ &= \sum_{r=1}^{n} \sum_{i=1}^{k} [z_i^{(r)} \log \lambda + z_i^{(r)} \log(t_i^{(r)} - t_{i-1}^{(r)}) - \log(z_i^{(r)}!) - \lambda(t_i^{(r)} - t_{i-1}^{(r)})]. \end{aligned} \tag{2.3}$$

icm©

Putting

$$\frac{\partial \log L}{\partial \lambda} = 0$$

and taking account of (2.2), we find the estimator of λ

$$(2.4) \qquad \hat{\lambda} = \frac{\sum_{r=1}^{n} (x_k^{(r)} - x_0^{(r)})}{\sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})}.$$

The estimator $\hat{\lambda}$ depends only on the results of the observations performed at the moments $t_0^{(r)}$ and $t_k^{(r)}$ and on the length of the periods $t_k^{(r)} - t_0^{(r)}$ elapsing between these observations. The results of observations carried out at intermediate moments $t_1^{(r)}, t_2^{(r)}, \ldots, t_{k-1}^{(r)}$ do not affect the value of the estimator. Thus it is sufficient to perform on each realization of the process only one pair of observations, all intermediate observations are redundant.

In view of the reproductive property of the Poisson distribution the sampling distribution of

$$\sum_{r=1}^{n} (x_k^{(r)} - x_0^{(r)}) = \hat{\lambda} \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})$$

is given by the probability function

$$(2.5) \qquad \frac{\left[\lambda \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})\right]^{\hat{\lambda} \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})}}{\left[\hat{\lambda} \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})\right]!} \exp\left[-\lambda \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})\right].$$

Consequently, the expectation of $\hat{\lambda}$ is

$$(2.6) \qquad E\hat{\lambda} = \lambda$$

and the variance of $\hat{\lambda}$ is

$$(2.7) \qquad V\hat{\lambda} = \frac{\lambda}{\sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})}.$$

The estimator $\hat{\lambda}$ is thus unbiased and consistent. Its sampling variance depends only on the number of realizations and the length of the periods $t_k^{(r)} - t_0^{(r)}$, and is not affected by observations at intermediate moments. Introducing the average period between the pairs of observations performed on each realization

$$(2.8) \qquad t_k - t_0 = \frac{1}{n} \sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)}),$$

we have

$$(2.9) \qquad V\hat{\lambda} = \frac{\lambda}{n(t_k - t_0)}.$$

The efficiency of the estimator, therefore, can be increased either by augmenting the number of realizations considered or by lengthening the average period between the two observations carried out on each realization. Additional observations at intermediate moments are useless.

3. The Brownian motion process. In the Brownian motion process the transition function (1.1) is

$$(3.1) \qquad f(t_i - t_{i-1}, z_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi(t_i - t_{i-1})}} \exp\left\{-\frac{[z_i - \mu(t_i - t_{i-1})]^2}{2\sigma^2(t_i - t_{i-1})}\right\},$$

where z_i is the change of state taking place during the period $t_i - t_{i-1}$. Denoting the state at the moments t_i and t_{i-1} by x_i and x_{i-1}, respectively, we have

$$(3.2) \qquad z_i = x_i - x_{i-1}.$$

Here x_i and x_{i-1} may be real numbers.

The expected change of state during the period $t_i - t_{i-1}$ is

$$Ez_i = \mu(t_i - t_{i-1})$$

and the variance of the change of state during that period is

$$E[z_i - \mu(t_i - t_{i-1})]^2 = \sigma^2(t_i - t_{i-1}),$$

where μ and σ^2 are constants. The Brownian motion process is thus a Gaussian process with stationary independent increments.

The logarithm of the likelihood function is

$$(3.3) \qquad \log L = \sum_{r=1}^{n} \sum_{i=1}^{k} \log f(t_i^{(r)} - t_{i-1}^{(r)}, z_i^{(r)}; \mu, \sigma^2)$$

$$= \sum_{r=1}^{n} \sum_{i=1}^{k} \left\{-\log\sigma - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log(t_i^{(r)} - t_{i-1}^{(r)}) - \frac{[z_i^{(r)} - \mu(t_i^{(r)} - t_{i-1}^{(r)})]^2}{2\sigma^2(t_i^{(r)} - t_{i-1}^{(r)})}\right\}.$$

Putting

$$\frac{\partial \log L}{\partial \mu} = 0$$

and taking account of (3.2), we obtain the estimator of μ

$$\hat{\mu} = \frac{\sum_{r=1}^{n} (x_k^{(r)} - x_0^{(r)})}{\sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})}. \tag{3.4}$$

The estimator of σ^2 is obtained by putting

$$\frac{\partial \log L}{\partial \sigma^2} = 0$$

and substituting $\hat{\mu}$ for μ in this equation. We obtain

$$\hat{\sigma}^2 = \frac{1}{nk} \sum_{r=1}^{n} \sum_{i=1}^{k} \frac{[(x_i^{(r)} - x_{i-1}^{(r)}) - \hat{\mu}(t_i^{(r)} - t_{i-1}^{(r)})]^2}{t_i^{(r)} - t_{i-1}^{(r)}}. \tag{3.5}$$

Similarly to the estimator $\hat{\lambda}$ in the Poisson process, the estimator $\hat{\mu}$ is independent of the results of observations carried out at moments intermediate between $t_0^{(r)}$ and $t_k^{(r)}$. By virtue of the reproductive property of the normal distribution it is normally distributed with expectation

$$E\hat{\mu} = \mu \tag{3.6}$$

and variance

$$V\hat{\mu} = \frac{\sigma^2}{\sum_{r=1}^{n} (t_k^{(r)} - t_0^{(r)})}. \tag{3.7}$$

The estimator $\hat{\mu}$ is thus unbiased and consistent. Its sampling variance depends only on the number of realizations and on the length of the periods $t_k^{(r)} - t_0^{(r)}$, and is not affected by additional observations at intermediate moments. By writing it in the form

$$V\hat{\mu} = \frac{\sigma^2}{n(t_k - t_0)}, \tag{3.8}$$

where $t_k - t_0$ is, as in (2.8), the average period between the pairs of observations performed on each realization, we find that the sampling variance of $\hat{\mu}$ is inversely proportional to the number of realizations taken into account and to the average period mentioned.

Unlike $\hat{\mu}$, the estimator $\hat{\sigma}^2$ depends on the results of the observations at all moments in the intervals $t_k^{(r)} - t_0^{(r)}$, as well as on the choice of these moments. It is distributed according to the χ^2 law with $nk-1$ degrees of freedom. In view of the known properties of the χ^2 distribution, the expectation and the sampling variance of $\hat{\sigma}^2$ are, respectively,

$$E\hat{\sigma}^2 = \frac{nk-1}{nk} \sigma^2, \tag{3.9}$$

$$V\hat{\sigma}^2 = \frac{2\sigma^4}{nk-1}. \tag{3.10}$$

Thus the estimator $\hat{\sigma}^2$ it not unbiased. An unbiased estimator, however, can be obtained by taking

$$\frac{nk}{nk-1} \hat{\sigma}^2. \tag{3.11}$$

Because of (3.10) $\hat{\sigma}^2$, as well as the expression (3.11), are consistent estimators. Their efficiency increases (roughly) in proportion both to the number of realizations considered and to the number of observations performed on each realization.

4. Testing hypotheses. Since the sampling distribution of the estimator $\hat{\lambda}$ in the simple Poisson process and of the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ in the Brownian motion process are known, hypotheses concerning values of the corresponding parameters can be tested by means of the Neyman-Pearson procedure.

In the simple Poisson process and in the Brownian motion process the number of changes of state or the magnitude of the change of the state, respectively, occurring in not overlapping time intervals are independent. Consequently the χ^2 criterion can be applied to test the hypothesis that one or several observed time series are realizations of a simple Poisson process or of a Brownian motion process. Furthermore the hypothesis that a set of observed time series are realizations of the same Poisson or Brownian motion process, *i. e.* of a process with the same parameter values, can be tested by the usual procedure of analysis of variance.

5. Linear "birth" and "death" processes. The transition function of the linear "birth process" is

$$f(t_{i-1}, x_{i-1};\ t_i, x_i; \alpha) = \binom{x_i - 1}{x_i - x_{i-1}} e^{-\alpha x_{i-1}(t_i - t_{i-1})} [1 - e^{-\alpha(t_i - t_{i-1})}]^{x_i - x_{i-1}}. \tag{5.1}$$

Here x_i and x_{i-1}, which must be integers, denote the number of individuals in the "populations" at the moments t_i and t_{i-1}, respectively. The probability of any one individual "giving birth" to a new individual during the infinitesimal period dt is αdt, where α is a constant. It can be shown that the number of individuals expected at the moment t_i is

$$Ex_i = x_{i-1} e^{\alpha(t_i - t_{i-1})}. \tag{5.2}$$

The logarithm of the likelihood function is

$$\begin{aligned}\log L = \sum_{r=1}^{n} \sum_{i=1}^{k} \log f(t_{i-1}^{(r)}, x_{i-1}^{(r)}; t_i^{(r)}, x_i^{(r)}; \alpha) = \sum_{r=1}^{n} \sum_{i=1}^{k} \Bigg\{ \log \binom{x_i^{(r)} - 1}{x_i^{(r)} - x_{i-1}^{(r)}} \\ - \alpha x_{i-1}^{(r)} (t_i^{(r)} - t_{i-1}^{(r)}) + (x_i^{(r)} - x_{i-1}^{(r)}) \log \left[1 - e^{-\alpha(t_i^{(r)} - t_{i-1}^{(r)})} \right] \Bigg\}.\end{aligned} \tag{5.3}$$

The estimator $\hat{\alpha}$ is obtained from the equation

$$\frac{\partial \log L}{\partial \alpha} = 0. \tag{5.4}$$

In order to obtain a workable solution of (5.4) we shall assume that all observations are carried out at equal intervals of length τ. We have then

$$t_i^{(r)} - t_{i-1}^{(r)} = \tau$$

for all r's and i's, and we obtain

$$e^{\hat{\alpha}\tau} = \frac{\sum_{r=1}^{n} \sum_{i=1}^{k} x_i^{(r)}}{\sum_{r=1}^{n} \sum_{i=1}^{k} x_{i-1}^{(r)}}, \tag{5.5}$$

whence

$$\hat{\alpha} = \frac{1}{\tau} \left(\log \sum_{r=1}^{n} \sum_{i=1}^{k} x_i^{(r)} - \log \sum_{r=1}^{n} \sum_{i=1}^{k} x_{i-1}^{(r)} \right). \tag{5.6}$$

This result was obtained by David G. Kendall[1]).

For the linear "death process" the transition function is

$$f(t_{i-1}, x_{i-1}; t_i, x_i; \beta) = \binom{x_{i-1}}{x_i} e^{-\beta x_{i-1}(t_i - t_{i-1})} [e^{\beta}(t_i - t_{i-1}) - 1]^{x_{i-1} - x_i} \tag{5.7}$$

---

[1]) D. G. Kendall, *Stochastic Processes and Population Growth*, Journal of the Royal Statistical Society, Series B, 9 (1949), p. 250.

where x_i and x_{i-1} have the same meaning as before. Note, however, that in the "death process" $x_i \leqslant x_{i-1}$, whereas in the "birth process" $x_i \geqslant x_{i-1}$. The constant β is defined by βdt being the probability that any one individual "dies" during the infinitesimal period dt. The number of individuals expected at the moment t_i is

$$Ex_i = x_{i-1} e^{-\beta(t_i - t_{i-1})}. \tag{5.8}$$

By a procedure similar as in the "birth process" we find that the estimator $\hat{\beta}$ satisfies the relation

$$e^{\hat{\beta}\tau} = \frac{\sum_{r=1}^{n} \sum_{i=1}^{k} x_{i-1}^{(r)}}{\sum_{r=1}^{n} \sum_{i=1}^{k} x_i^{(r)}}, \tag{5.9}$$

whence

$$\hat{\beta} = \frac{1}{\tau} \left(\log \sum_{r=1}^{n} \sum_{i=1}^{k} x_{i-1}^{(r)} - \log \sum_{r=1}^{n} \sum_{i=1}^{k} x_i^{(r)} \right). \tag{5.10}$$

The exact sampling distributions of $\hat{\alpha}$ sand $\hat{\beta}$ are as yet unknown. Neither do we know whether these estimators are unbiased (**P 129**). However, the known asymptotic properties of maximum likelihood estimators allow us to find their asymptotic sampling variance. The observed realizations of the process being independent, we have, for $n \to \infty$, asymptotically

$$V\hat{\alpha} = -\frac{1}{E \dfrac{\partial^2 \log L}{\partial \alpha^2}}. \tag{5.11}$$

From (5.3) we find, treating $t_i^{(r)} - t_{i-1}^{(r)} = \tau = \text{const}$,

$$\frac{\partial^2 \log L}{\partial \alpha^2} = \frac{\tau^2 e^{\alpha\tau}}{(e^{\alpha\tau} - 1)^2} \sum_{r=1}^{n} (x_k^{(r)} - x_0^{(r)}). \tag{5.12}$$

By virtue of (5.2) we have

$$Ex_k^{(r)} = x_0^{(r)} e^{\alpha T},$$

where $T = k\tau$ is the total observation period. Consequently,

$$E \frac{\partial^2 \log L}{\partial \alpha^2} = -\frac{\tau^2 e^{\alpha\tau}(e^{\alpha T} - 1)}{(e^{\alpha T} - 1)^2} \sum_{r=1}^{n} x_0^{(r)}$$

icm©

and, according to (5.11),

$$V\hat{\alpha} = \frac{1}{N_0(e^{\alpha T}-1)} \cdot \frac{(e^{\alpha\tau}-1)^2}{\tau^2 e^{\alpha\tau}};$$

where $N_0 = \sum_{r=1}^{n} x_0^{(r)}$ is the total initial "population" of all the realizations considered.

This can be brought into the form

$$\text{(5.13)} \qquad V\hat{\alpha} = \frac{\alpha^2}{N_0(e^{\alpha T}-1)} \left(\frac{\sinh \frac{1}{2}\alpha \frac{T}{k}}{\frac{1}{2}\alpha \frac{T}{k}} \right)^2,$$

a formula which was obtained by D. G. Kendall[2]).

By an analogous procedure we get asymptotically

$$\text{(5.14)} \qquad V\hat{\beta} = \frac{\beta^2}{N_0(1-e^{-\beta T})} \left(\frac{\sinh \frac{1}{2}\beta \frac{T}{k}}{\frac{1}{2}\beta \frac{T}{k}} \right)^2.$$

The sampling variance of the estimators $\hat{\alpha}$ and $\hat{\beta}$ is thus inversely proportional to the total initial "population" of all the realizations considered. It also decreases rapidly with the length of the observation period T. Furthermore, it depends on the number of observations performed on each realization (which is $k+1$). From (5.13) and (5.14) it is seen immediately that the variance decreases with the number k, reaching asymptotically a maximum value for $k\to\infty$ when the squared factor becomes unity. Thus the estimators $\hat{\alpha}$ and $\hat{\beta}$ become most efficient under conditions of continuous observation of each realization of the process.

6. The influence of intermediate observations. The estimators $\hat{\alpha}$ and $\hat{\beta}$ in the linear "birth" and "death" processes are, so to speak, on the opposite pole of the estimators $\hat{\lambda}$ and $\hat{\mu}$ in the simple Poisson process and the Brownian motion process, respectively. The latter two, as we know, depend only on the first and on the last observation performed on each realization. So does their efficiency. Neither their value nor their efficiency is affected by intermediate observations.

[1]) D. G. Kendall, loc. cit. [2]), p. 250.

In order that an estimator $\hat{\Theta}$ be independent of the intermediate observations the derivative of the logarithm of the likelihood function must be separable into two factors

$$\text{(6.1)} \qquad \frac{\partial \log L}{\partial \Theta} = g \cdot h(x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(n)}; x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)}; t_0^{(1)}, t_0^{(2)}, \ldots, t_k^{(n)}; t_k^{(1)}, t_k^{(2)}, \ldots, t_k^{(n)}; \Theta)$$

in such a way that the factor g may depend on all or some of the observations $x_0^{(r)}, x_1^{(r)}, \ldots, x_k^{(r)}$ and corresponding moments $t_0^{(r)}, t_1^{(r)}, \ldots, t_k^{(r)}$, as well as on other parameters, but does not depend on the parameter Θ (or if it depends the equation $g=0$ has no admissible solution for Θ; for instance the solution is complex while Θ is postulated to be real), whereas the factor h, which depends on the parameter Θ, depends only on the first and the last observation performed on each realization, *i. e.* on $x_0^{(r)}, x_k^{(r)}$ and $t_0^{(r)}, t_k^{(r)}$ $(r=1,2,\ldots,n)$. This is obviously sufficient as well as necessary.

We assume that the equation

$$\text{(6.2)} \qquad h(x_0^{(1)}, x_0^{(2)}, \ldots, x_0^{(n)}; x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(n)}; t_0^{(1)}, t_0^{(2)}, \ldots, t_0^{(n)}; t_k^{(1)}, t_k^{(2)}, \ldots, t_k^{(n)}; \Theta) = 0$$

is uniquely solvable with regard to Θ. The solution yields the estimator $\hat{\Theta}$. This estimator does not depend on the intermediate observations.

The condition (6.1) implies

$$\text{(6.3)} \qquad \log L = gH(\Theta) + C,$$

where $H(\Theta)$ is the primitive function of h with regard to Θ, and g and C do not depend on Θ.

The condition (6.3) is satisfied with regard to the parameter λ in the simple Poisson process. It is satisfied in the Brownian motion process with regard to the parameter μ but not with regard to the parameter σ^2. It is not satisfied with regard to the parameters α and β, respectively, in the linear "birth process" and the linear "death process".

The conditions for independence from intermediate observations of the sampling variance of the estimator $\hat{\Theta}$ require separate investigation. In view of the relation

$$\text{(6.4)} \qquad V\hat{\Theta} = -\frac{1}{E\dfrac{\partial^2 \log L}{\partial \Theta^2}},$$

which holds asymptotically for $nk\to\infty$, we find that the sampling va-

icm©

riance of $\hat{\Theta}$ does not depend on the intermediate observations, at least asymptotically for $n \to \infty$, when

$$(6.5) \qquad E \frac{\partial^2 \log L}{\partial \Theta^2}$$

is a function of $x_0^{(r)}, x_k^{(r)}$ and $t_0^{(r)}, t_k^{(r)}$ $(r=1,2,\ldots,n)$ only. This condition is necessary as well as sufficient.

As can easily be verified, the condition (6.5) is satisfied for the estimator $\hat{\lambda}$ in the simple Poisson process and for the estimator $\hat{\mu}$ in the Brownian motion process. In these cases (6.4) is satisfied not asymptotically but exactly. In the Brownian motion process the sampling variance of $\hat{\sigma}^2$ depends on the number of intermediate observations, but is independent of their timing. This can also be seen directly from (3.10).

7. Relation to least squares estimation. The maximum likelihood estimators obtained in this paper are identical with the corresponding least squares estimators. Using the same notation as before, we find the following.

The estimator $\hat{\lambda}$ for the simple Poisson process can be obtained by minimizing the expression

$$(7.1) \qquad \sum_{r=1}^{n} \sum_{i=1}^{k} \frac{(z_i^{(r)} - E z_i^{(r)})^2}{t_i^{(r)} - t_{i-1}^{(r)}},$$

where $E z_i^{(r)} = \lambda (t_i^{(r)} - t_{i-1}^{(r)})$.

The estimator $\hat{\mu}$ for the Brownian motion process can be obtained by minimizing the expression

$$(7.2) \qquad \sum_{r=1}^{n} \sum_{i=1}^{k} \frac{(z_i^{(r)} - E z_i^{(r)})^2}{t_i^{(r)} - t_{i-1}^{(r)}},$$

where $E z_i^{(r)} = \mu (t_i^{(r)} - t_{i-1}^{(r)})$.

The estimator $\hat{\sigma}^2$ is then the mean square of the residuals of the least squares estimation, *i. e.*

$$(7.3) \qquad \sigma^2 = \frac{1}{nk} \sum_{r=1}^{n} \sum_{i=1}^{k} \frac{[z_i^{(r)} - \hat{\mu}(t_i^{(r)} - t_{i-1}^{(r)})]^2}{t_i^{(r)} - t_{i-1}^{(r)}}.$$

The estimators $\hat{\alpha}$ and $\hat{\beta}$ for the linear "birth process" and linear "death process" can be obtained by minimizing the expression

$$(7.4) \qquad \sum_{r=1}^{n} \sum_{i=1}^{k} \frac{(x_i^{(r)} - E x_i^{(r)})^2}{x_{i-1}^{(r)}},$$

where $E x_i^{(r)} = x_{i-1}^{(r)} e^{\alpha\tau}$ in the "birth process" and $E x_i^{(r)} = x_{i-1}^{(r)} e^{-\beta\tau}$ in the "death process".

Notice should be taken of the fact that in the expression (7.1) and (7.2) the squares of the deviation from the expected value are weighted by the reciprocals of the length of time $t_i^{(r)} - t_{i-1}^{(r)}$ elapsing between successive observations. In the expression (7.4), instead, the squares of the deviations from the expected value are weighted by the reciprocal of the size of the "population" at the beginning of each successive observation period. In other words, in (7.1) and (7.2) the squares of the deviations are taken as "per unit of time", whereas in (7.4) they are taken as "per unit of population".

8. Estimation of transition probabilities in simple Markov chains. Finally, we consider the problem of estimating transition probabilities in simple Markov chains with a finite number of states. We assume the transition probabilities to be stationary.

Denote the states by the numbers $1,2,\ldots,s$ and denote by p_{ij} the probability of transition from state i to state j. The probabilities of transition form the transition matrix

$$(8.1) \qquad \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1s} \\ p_{21} & p_{22} & \cdots & p_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ p_{s1} & p_{s2} & \cdots & p_{ss} \end{pmatrix}.$$

In this matrix

$$0 \leqslant p_{ij} \leqslant 1 \quad \text{for} \quad i,j = 1,2,\ldots,s$$

and

$$(8.2) \qquad \sum_{j=1}^{s} p_{ij} = 1 \quad \text{for} \quad i = 1,2,\ldots,s.$$

Let N changes of state be observed and denote by m_{ij} the observed frequency of changes from state i to state j. Denote further

$$(8.3) \qquad n_i = \sum_{j=1}^{s} m_{ij},$$

i. e. the total frequency of changes starting from state i. Obviously

$$\sum_{i=1}^{s} n_i = N.$$

We have the observation matrix

$$(8.4) \qquad \begin{pmatrix} m_{11} & m_{12} & \cdots & m_{1s} \\ m_{21} & m_{22} & \cdots & m_{2s} \\ \cdot & \cdot & \cdot & \cdot \\ m_{s1} & m_{s2} & \cdots & m_{ss} \end{pmatrix}.$$

The logarithm of the likelihood function of this matrix is, according to the multinomial law,

$$\log L = \log(N!) - \sum_{i=1}^{s}\sum_{j=1}^{s}(m_{ij}!) - \sum_{i=1}^{s}\sum_{j=1}^{s} m_{ij}\log p_{ij}. \tag{8.5}$$

The estimators $\hat{p}_{ij}$ are found by putting

$$\frac{\partial \log L}{\partial p_{ij}} = 0 \tag{8.6}$$

subject to the side relations (8.2). Introducing the Lagrange multipliers $l_1, l_2, \ldots, l_s$, we arrive at the equations

$$m_{ij} = l_j p_{ij} \qquad (i,j=1,2,\ldots,s). \tag{8.7}$$

Summing over j and taking into account (8.3) as well as (8.2), we find

$$l_i = n_i \qquad (i=1,2,\ldots,s)$$

and, consequently,

$$\hat{p}_{ij} = \frac{m_{ij}}{n_i} \qquad (i,j=1,2,\ldots,s). \tag{8.8}$$

The estimator of the probability of transition from state i to state j is is the relative frequency of changes issuing in state j among all changes starting from state i. This result was first obtained by V. I. Romanovski[3])

The expectation and the sampling variance of $\hat{p}_{ij}$ are, respectively,

$$E\hat{p}_{ij} = p_{ij} \quad \text{and} \quad V\hat{p}_{ij} = \frac{1}{n_i} p_{ij}(1-p_{ij}). \tag{8.9}$$

The estimator $\hat{p}_{ij}$ is thus unbiased and consistent.

The observation matrix (8.4) being given, we can by virtue of (8.8) estimate the transition matrix (8.1). By means of the χ^2 criterion the hypothesis can be tested that an observation matrix (8.4) is the result of the realization of a simple Markov chain with some theoretical transition matrix (8.1). In this case

$$\chi^2 = \sum_{i=1}^{s}\sum_{j=1}^{s}\frac{(m_{ij}-n_i p_{ij})^2}{n_i p_{ij}}, \tag{8.10}$$

the number of degrees of freedom being s^2-1.

[3]) В. И. Романовский, *Дискретные цепи Маркова*, Москва-Ленинград 1949, p. 393.

SUR LE COLORIAGE DES GRAPHS

PAR

J. MYCIELSKI (WROCŁAW)

Un *graph fini* est un ensemble fini (situé dans l'espace euclidien à trois dimensions) de points dont certains sont joints par des arcs simples qui n'ont pas de points communs (sauf — peut-être — de points finaux)[1]). Un graph composé de trois points joints deux-à-deux s'appelle un *triangle*.

Colorier un graph au moyen de n couleurs — veut dire — peindre chacun de ses points par une de ces couleurs, les points joints par un arc simple étant peints de couleurs différentes.

Il est évident qu'un graph qui contient m points joints deux-à-deux ne peut pas être colorié par moins que m couleurs. Je me propose de démontrer dans cette note qu'un théorème inverse serait faux. En effet je vais prouver le théorème suivant:

Théorème. *Pour chaque nombre naturel n il existe un graph fini ne contenant aucun triangle, qui ne peut pas être colorié au moyen de n couleurs.*

Démonstration. Pour $n=1$, deux points joints par un arc simple sont un exemple d'un tel graph.

Désignons par $a_1, \ldots, a_m$ les points d'un graph A ne contenant pas de triangle, qui ne peut pas être colorié par n couleurs. Nous allons construire un graph A^* qui ne contient pas de triangle et qui ne peut pas être colorié par $n+1$ couleurs.

Désignons par $a_{k_{i1}}, \ldots, a_{k_{ir}}$ la suite de tous les points du graph A qui sont joints au point a_i.

Choisissons dans l'espace un ensemble de $m+1$ points: $a_0', a_1', \ldots, a_m'$ qui est disjoint avec le graph A.

Joignons par des arcs simples chacun des points $a_{k_{i1}}, \ldots, a_{k_{ir}}$ au point a_i' (pour $i=1,\ldots,m$). Joignons encore chacun des points $a_1', \ldots, a_m'$ au point a_0'.

[1]) Il est évident qu'un graph fini constitue un modèle d'une relation symétrique et antiréflexive définie dans un ensemble fini. Ainsi cet article pourrait être écrit en termes d'algèbre de relations, comme p. ex le travail de K. Zarankiewicz, *Sur les relations symétriques dans l'ensemble fini*, Colloquium Mathematicum 1 (1947), p. 10-14.