APPLICATIONES MATHEMATICAE
28,4 (2001), pp. 467–485

Ryszard Magiera (Wrocław)

Γ-MINIMAX SEQUENTIAL ESTIMATION FOR MARKOV-ADDITIVE PROCESSES

Abstract. The problem of estimating unknown parameters of Markov-additive processes from data observed up to a random stopping time is considered. To the problem of estimation, the intermediate approach between the Bayes and the minimax principle is applied in which it is assumed that a vague prior information on the distribution of the unknown parameters is available. The loss in estimating is assumed to consist of the error of estimation (defined by a weighted squared loss function) as well as a cost of observing the process up to a stopping time. Several classes of optimal sequential procedures are obtained explicitly in the case when the available information on the prior distribution is restricted to a set Γ which is determined by certain moment-type conditions imposed on the prior distributions.

1. Introduction. Let $(A(t), X(t))$, $t \geq 0$, (the time parameter t is continuous) be a Markov-additive process with state space $\mathbb{R} \times I$, where $I = \{1, \ldots, m\}$. It is assumed that the conditional distribution of $A(t)-A(s)$, given $X(u) = i$ for all $u \in [s, t]$, is given by the density

$$\exp[\upsilon_i x - f_i(\upsilon_i)(t-s)]$$

with respect to a σ-finite measure which may depend on the state i in general, and υ_i is a real parameter, $\upsilon_i \in \varUpsilon_i \subseteq \mathbb{R}$. This means that the sojourn time distributions belong to one-dimensional exponential families. Let $(\lambda_{ij})_{i,j=1}^{m}$ be the transition intensity matrix of the embedded m-state Markov chain $X(t)$.

The model of processes considered covers a class of Markov-additive processes which have important applications to queueing and data communi-

2000 *Mathematics Subject Classification*: Primary 62L12; Secondary 62F15, 62C20.

Key words and phrases: gamma-minimax estimation, Bayes estimation, minimax estimation, stopping time, sequential procedure, Markov-additive process.

cation models. They are used to model queueing-reliability systems, arrival processes in telecommunication networks, environmental data, neural impulses etc. A particularly important class of Markov-additive processes is the class of Markov-additive processes $(A(t), X(t))$ of arrivals, i.e., those with the additive component $A(t)$ taking values in the set of nonnegative integers. A typical example is that of arrivals in a queueing system.

Sequential estimation procedures of the form $\delta = (\tau, d(\tau))$ will be considered, where τ is a stopping time and $d(\tau)$ is an estimator based on the observation of the process up to τ. The parameter $\vartheta = (\lambda_{ij}, i, j = 1, \ldots, m; \upsilon_1, \ldots, \upsilon_m)$ of the Markov-additive process considered is unknown and the problem is to find optimal sequential procedures, i.e., optimal stopping times τ and the corresponding sequential estimators $d(\tau)$ for ϑ. It is supposed that if the observation is stopped at time τ and the estimate $d(\tau)$ is reported, then the loss incurred is

$$\mathcal{L}_\tau(\vartheta, d(\tau)) = \mathcal{L}(\vartheta, d(\tau)) + c(\tau),$$

where $\mathcal{L}(\vartheta, d(\tau))$ denotes the loss function (representing the error of estimation) and $c(\tau)$ is the cost function. The loss function is defined by a weighted squared error loss. The cost for a given procedure is determined by a function of one of the components of the Markov-additive process; for example, it is the cost depending on arrivals at a queueing system up to the moment of stopping.

Let π be a prior distribution on Θ. Then the Bayes risk of the sequential procedure $\delta = (\tau, d(\tau))$ is

$$\overline{\mathcal{R}}(\pi, \delta) = \int_{\Theta} E_\vartheta[\mathcal{L}_\tau(\vartheta, d(\tau))]\, d\pi(\vartheta).$$

If there is precise prior information on the distribution of the unknown parameter ϑ which can be described by a prior π, then usually the Bayes principle is used. If on the other hand no prior information is available, then the minimax principle can be applied. In this paper, to find optimal sequential estimation procedures, an intermediate approach between the Bayes and the minimax principle is chosen. The use of the Γ-minimax principle is appropriate if vague prior information is available which can be described by a subset Γ of the set Π of all priors. The problem is to find stopping times τ and the corresponding sequential estimators $d(\tau)$ subject to the minimax criterion: a sequential procedure $\delta^0 = (\tau^0, d^0)$ is said to be *Γ-minimax* if

$$\sup_{\pi\in\Gamma} \overline{\mathcal{R}}(\pi, \delta_0) = \inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} \overline{\mathcal{R}}(\pi, \delta),$$

i.e., if it minimizes the maximum of the total Bayes sequential risk when the set of prior distributions of the unknown parameter is restricted to a subset Γ of all priors. $\overline{\mathcal{D}}$ is the class of all sequential procedures δ having finite Bayes risk for each $\pi \in \Gamma$.

The set Γ is determined by certain moment-type conditions imposed on the prior distributions. The idea and tools are exhibited to obtain Γ-minimax sequential procedures for estimating important quantities of the unknown parameters of the Markov-additive process. As one of the tools for solving the problem, a minimax theorem, which is a considerable generalization of a theorem of Dvoretzky, Kiefer and Wolfowitz (1953), is provided for a general class of stochastic processes and a wide class of stopping times.

Several classes of Γ-minimax sequential procedures for estimating the unknown parameters of the Markov-additive process are presented. For example, a class of Γ-minimax sequential procedures is derived explicitly in the case when for a fixed state i the ratios of λ_{ij}, $j = 1, \ldots, m$, $j \neq i$, to $f_i'(v_i)$ are of interest. In particular, the results presented are applicable to the Markov-additive processes of arrivals most frequently appearing in the literature, i.e., to the Markov-modulated Poisson processes.

The results obtained constitute a generalization of the results given by Magiera (1999).

2. The model and sampling times. Set $I = \{1, \ldots, m\}$. In accordance with the definition given by Pacheco and Prabhu (1995), a process $(A(t), X(t))$, $t \geq 0$, with the continuous time parameter t, on the state space $\mathbb{R} \times I$ is said to be a *Markov-additive process* if

1) $(A(t), X(t))$, $t \geq 0$, is a Markov process;

2) the conditional distribution of $(A(s+t) - A(s), X(s+t))$, given $(A(s), X(s))$, depends only on $X(s)$.

The basic theory of Markov-additive processes is presented by Çinlar (1972), Ezhov and Skorokhod (1969), and Prabhu (1991), where a more general state space for $X(t)$ is considered. Recall (see Pacheco and Prabhu (1995)) the most important properties of these processes. The transition probability measure of a Markov-additive process is given by

$$(1)\qquad \begin{aligned} &P(A(s+t) \in B,\ X(s+t) = j \mid A(s) = y,\ X(s) = i) \\ &\qquad = P(A(s+t) - A(s) \in B - y,\ X(s+t) = j \mid X(s) = i) \end{aligned}$$

for $s, t > 0$, $i, j \in I$, $y \in \mathbb{R}$ and $B \in \mathcal{B}_{\mathbb{R}}$, where $\mathcal{B}_{\mathbb{R}}$ is the σ-algebra of Borel subsets of $\mathbb{R}$. Since $(A(t), X(t))$, $t \geq 0$, is Markov, it follows from (1) that the component $X(t)$ is Markov and that $A(t)$ has conditionally independent increments, given $X(t)$. For $0 \leq t_1 \leq \ldots \leq t_n$, $n > 2$, the increments

$$A(t_1) - A(0), A(t_2) - A(t_1), \ldots, A(t_n) - A(t_{n-1})$$

are conditionally independent given $X(0), X(t_1), \ldots, X(t_n)$. The process $A(t)$, which is, in general, not Markovian is called the *additive component* of the Markov-additive process. Markov renewal processes are discrete time versions of Markov-additive processes with the additive component taking values in $\mathbb{R}_+$.

Markov-additive processes of arrivals (i.e. ones for which $A(t)$ has the state space $\{0, 1, 2, \ldots\}$) were studied by Pacheco and Prabhu (1995). An important special case is the Markov-modulated Poisson process in which the rate λ of occurrence of Poisson events changes instantaneously with the change of state in $X(t)$. A survey of possible applications of Markov-additive processes of arrivals can be found in Pacheco and Prabhu (1995). For example, these processes have been used to model overflow from trunk groups, superpositioning of packeted voice streams, and input to Asynchronous Transfer Mode networks, which will be used in high-speed communication networks. Using Markov-additive processes, some theoretical queueing results were established by Neuts (1992).

Although there is a vast literature on Markov-additive processes and their use in modelling many stochastic phenomena, the statistical issues for these processes have not been the subject of much study. Nonparametric and Bayesian estimation for Markov renewal processes was considered by Gill (1980) and Phelan (1990 a,b). Some problems of parameter estimation for Markov-modulated Poisson processes were treated by Rydén (1994). Asymptotic normality of sequential and nonsequential maximum likelihood estimators for Markov renewal and Markov-additive processes was established by Stefanov (1995). A sequential estimation scheme based on "small" samples was considered by Fygenson (1991), where efficient (optimal in the sense of the Cramér–Rao–Wolfowitz lower bound) sequential procedures were investigated for Markov renewal processes.

In this paper we consider the problem of finding optimal, under minimax criterion, sequential procedures for estimating the parameters of Markov-additive processes in the case when the set of prior distributions of the parameters is restricted.

Let the conditional distribution of $A(t) - A(s)$, given $X(u) = i$ for all $u \in [s, t]$, be given by the density

$$\exp[\upsilon_i x - f_i(\upsilon_i)(t - s)] \tag{2}$$

with respect to a σ-finite measure which may depend on the state i in general, and υ_i is a real parameter, $\upsilon_i \in \Upsilon_i \subset \mathbb{R}$. Υ_i is assumed to be the interior of the natural parameter space of the exponential family given by (1). Assume that $f_i'(\upsilon_i) > 0$ for each $\upsilon_i \in \Upsilon_i$. Since formula (1) implies that the derivative $f_i'(\upsilon_i)$ is the mean value parameter of the increments for $A(t)$, this assumption covers the natural case of positive increments (arrivals) of the additive component $A(t)$. It is assumed that $X(0) = 1$ with probability 1.

The likelihood function corresponding to the observation of the process up to time t has the following form (see Stefanov (1995)):

$$(3)\quad L(t,\lambda,\upsilon) = \prod_{\substack{i,j=1\\ i\neq j}}^{m} \lambda_{ij}^{N_{ij}(t)} \exp[-S_i(t)\lambda_{ij}] \prod_{i=1}^{m} \exp[A_i(t)\upsilon_i - S_i(t) f_i(\upsilon_i)]$$
$$= \prod_{\substack{i,j=1\\ i\neq j}}^{m} \lambda_{ij}^{N_{ij}(t)} \prod_{i=1}^{m} \exp\{A_i(t)\upsilon_i - S_i(t)[\lambda_{ii} + f_i(\upsilon_i)]\},$$

where $\lambda = (\lambda_{ij})_{i,j=1}^m$ is the transition intensity matrix of the embedded m-state Markov chain $X(t)$, $\upsilon = (\upsilon_1, \ldots, \upsilon_m)$, $N_{ij}(t)$ is the number of transitions from state i to state j of the Markov process $X(s)$ in the time interval $[0,t]$, $S_i(t)$ is the sojourn time in state i of the process $X(s)$, and

$$A_i(t) = \sum_{n=1}^{\infty} \{[A(\eta_n^*(i)) - A(\eta_n(i))] \mathbf{1}_{[0,\infty)}[t - \eta_n^*(i)]$$
$$+ [A(t) - A(\eta_n(i))] \mathbf{1}_{(\eta_n(i),\eta_n^*(i))}(t)\},$$

where $\eta_n(i)$ is the nth consecutive time of first entrance of $X(s)$ to state i, and $\eta_n^*(i)$ is the exit time from state i after $\eta_n(i)$.

By the fundamental identity of sequential analysis (see Appendix), for any finite stopping time τ the sequential version of (3) has the following form:

$$(4)\quad L(\tau,\lambda,\upsilon)$$
$$= \prod_{\substack{i,j=1\\ i\neq j}}^{m} \lambda_{ij}^{N_{ij}(\tau)} \exp[-S_i(\tau)\lambda_{ij}] \prod_{i=1}^{m} \exp[A_i(\tau)\upsilon_i - S_i(\tau) f_i(\upsilon_i)]$$
$$= \prod_{\substack{i,j=1\\ i\neq j}}^{m} \lambda_{ij}^{N_{ij}(\tau)} \prod_{i=1}^{m} \exp\{A_i(\tau)\upsilon_i - S_i(\tau)[\lambda_{ii} + f_i(\upsilon_i)]\}$$
$$= \exp\Big\{ \sum_{\substack{i,j=1\\ i\neq j}}^{m} N_{ij}(\tau)\log\lambda_{ij} + \sum_{i=1}^{m} A_i(\tau)\upsilon_i - \sum_{i=1}^{m} S_i(\tau)[\lambda_{ii} + f_i(\upsilon_i)]\Big\}.$$

From the theoretical point of view (the possibility of exploiting the very useful tools associated with exponential families), and in view of possible applicability, the most relevant stopping times which can be taken into account in searching for optimal sequential estimation procedures are the following:

$$\tau_{J,s}^i = \inf\Big\{t : \sum_{j\in J} N_{ji}(t) = s\Big\}, \qquad s = 1, 2, \ldots,$$

$$\tau_{i,s}^i = \inf\{t : S_i(t) = s\}, \qquad s > 0,$$

$$\widetilde{\tau}^i_{i,s} = \inf\{t : A_i(t) = s\},$$
$$\tau^i_{iJ,s} = \inf\Big\{t : A_i(t) + \sum_{j\in J} N_{ji}(t) = s\Big\}, \qquad s = 1, 2, \ldots,$$

for each $i \in I$ and each $J \subseteq I$ such that $\sum_{j\in J} \lambda_{ji} > 0$. It is additionally assumed that $P_{\lambda,\upsilon}(\widetilde{\tau}^i_{i,s} < \infty) = 1$ and $P_{\lambda,\upsilon}(\tau^i_{iJ,s} < \infty) = 1$ for each (λ, υ). Considering, for example, the stopping time $\tau^i_{J,s}$, one should bear in mind that the following linear dependencies for the components of $N(\tau^i_{J,s})$ hold with probability 1:

$$\tag{5} \begin{gathered} \sum_{j\in J} N_{ji}(\tau^i_{J,s}) - s = 0, \\ \sum_{j=1,\, k\neq j}^{m} N_{jk}(\tau^i_{J,s}) - \sum_{j=1,\, k\neq j}^{m} N_{kj}(\tau^i_{J,s}) + \mathbf{1}_{(k)}(X(0)) - \mathbf{1}_{(k)}(X(\tau^i_{J,s})) = 0, \\ k = 1, \ldots, m-1, \end{gathered}$$

where $\mathbf{1}_{(k)}(X(0))$ equals 1 if $k = 1$, and 0 otherwise, whereas $\mathbf{1}_{(k)}(X(\tau^i_{J,s}))$ equals 1 if $k = i$, and 0 otherwise. The class of stopping times above was described by Stefanov (1995). By his Proposition 5.1, for each stopping time from this class the curved (in general) exponential family of (4) becomes a noncurved exponential family of order equal to the dimension of the parameter (λ, υ). The condition $P_{\lambda,\upsilon}(\widetilde{\tau}^i_{i,s} < \infty) = 1$ is satisfied, for example, if the process $A_i(t)$ is nonnegative with continuous trajectories and $A_i(t) \to \infty$ as $t \to \infty$, or if the process $A_i(t)$ is a Poisson process; in the latter case s must be a natural number (see Remark 5.1 in Stefanov (1995)). The condition $P_{\lambda,\upsilon}(\tau^i_{iJ,s} < \infty) = 1$ is satisfied, for example, in the case of Markov-modulated Poisson processes. The stopping times $\widetilde{\tau}^i_{i,s}$ and $\tau^i_{iJ,s}$ are new in comparison to those relevant in sequential estimation problems for finite-state Markov processes. The stopping times $\widetilde{\tau}^i_{i,s}$ and $\tau^i_{i,s}$ will be considered in the next section.

Considering the estimation problem for some special functions of the unknown parameter (λ, υ) we exhibit the idea and tools for proving the Γ-minimaxity of sequential procedures determined by the stopping times from the class above.

3. Γ-minimax sequential procedures. Suppose that the unknown parameter vector is $\overline{\lambda}_i = (\lambda_i, \upsilon_i)$ (i fixed), where $\lambda_i = (\lambda_{i1}, \ldots, \lambda_{i,i-1}, \lambda_{i,i+1}, \ldots, \lambda_{im})$. Then the likelihood function is

$$\tag{6} L_i(\tau, \overline{\lambda}_i) = \prod_{j=1, j\neq i}^{m} \lambda_{ij}^{N_{ij}(\tau)} \exp\{A_i(\tau)\upsilon_i - S_i(\tau)[\lambda_{ii} + f_i(\upsilon_i)]\}.$$

3.1. It will be shown that if the ratios of λ_{ij} ($j = 1, \ldots, m$; $j \neq i$) to $f_i'(\upsilon_i)$ are of interest, then to estimate them one can use a sequential procedure defined by the following stopping time:

$$\tau_{a_i} = \inf\{t : A_i(t) = a_i\}, \quad a_i > 0, \tag{7}$$

where it is additionally assumed that $P_{\overline{\lambda}_i}(\tau_{a_i} < \infty) = 1$ for each $\overline{\lambda}_i \in \Lambda_i \times \Upsilon_i$, $\Lambda_i = (0, \infty)^{m-1}$. We will then show that the stopping time (7) determines a class of Γ-minimax sequential procedures under a weighted squared error loss and the cost depending on the value of the process $A_i(t)$ at the moment of stopping. In particular, for a Markov-additive arrivals process this cost will be a function of arrivals in a queueing system. For the stopping time τ_{a_i} the sequential likelihood function (6) takes the form

$$L_i(\tau_{a_i}, \overline{\lambda}_i) = \prod_{j=1, j\neq i}^{m} \lambda_{ij}^{N_{ij}(\tau_{a_i})} \exp\{a_i\upsilon_i - S_i(\tau_{a_i})[\lambda_{ii} + f_i(\upsilon_i)]\}. \tag{8}$$

The family (8) is an exponential family in which the dimension of the sufficient statistic equals the dimension of the unknown parameter. It then follows from the well known analytical properties of noncurved exponential families (see Barndorff-Nielsen (1978) or Brown (1986)) that the regularity conditions which allow one to differentiate twice under the integral sign with respect to the parameter $\overline{\lambda}_i$ in the identity $\int L_i(\tau_{a_i}, \overline{\lambda}_i)\, d\mu_{\tau_{a_i}} = 1$ are satisfied (μ_τ denotes a dominating measure in the fundamental identity of sequential analysis, i.e., $dP_{\lambda,\tau}/d\mu_\tau = L(\tau, \lambda)$—see Appendix). Thus, for the stopping time τ_{a_i} the following Wald identities hold:

$$E_{\overline{\lambda}_i} S_i(\tau_{a_i}) = \frac{a_i}{f_i'(\upsilon_i)}; \tag{9}$$

$$E_{\overline{\lambda}_i} N_{ij}(\tau_{a_i}) = \lambda_{ij} E_{\overline{\lambda}_i} S_i(\tau_{a_i}) = \frac{\lambda_{ij} a_i}{f_i'(\upsilon_i)}, \quad j \neq i; \tag{10}$$

$$E_{\overline{\lambda}_i}\Big[\sum_{j=1,\, j\neq i}^{m} N_{ij}(\tau_{a_i})\Big] = \lambda_{ii} E_{\overline{\lambda}_i} S_i(\tau_{a_i}) = \frac{\lambda_{ii} a_i}{f_i'(\upsilon_i)};$$

$$E_{\overline{\lambda}_i}\{[N_{ij}(\tau_{a_i}) - \lambda_{ij} S_i(\tau_{a_i})][N_{ik}(\tau_{a_i}) - \lambda_{ik} S_i(\tau_{a_i})]\} = 0, \quad k = 1, \ldots, m;\ j \neq k;$$

$$E_{\overline{\lambda}_i}\{[N_{ij}(\tau_{a_i}) - \lambda_{ij} S_i(\tau_{a_i})][a_i - S_i(\tau_{a_i}) f_i'(\upsilon_i)]\} = 0, \quad j \neq i;$$

$$E_{\overline{\lambda}_i}[N_{ij}(\tau_{a_i}) - \lambda_{ij} S_i(\tau_{a_i})]^2 = E_{\overline{\lambda}_i} N_{ij}(\tau_{a_i}) = \frac{\lambda_{ij} a_i}{f_i'(\upsilon_i)}, \quad j \neq i;$$

$$E_{\overline{\lambda}_i}\left[S_i(\tau_{a_i}) - \frac{a_i}{f_i'(\upsilon_i)}\right]^2 = \frac{f_i''(\upsilon_i)}{(f_i'(\upsilon_i))^2} E_{\overline{\lambda}_i} S_i(\tau_{a_i}) = \frac{a_i f_i''(\upsilon_i)}{(f_i'(\upsilon_i))^3}; \tag{11}$$

$$(12)\quad E_{\overline{\lambda}_i}\left\{\left[N_{ij}(\tau_{a_i})-\frac{\lambda_{ij}a_i}{f_i'(\upsilon_i)}\right]\left[S_i(\tau_{a_i})-\frac{a_i}{f_i'(\upsilon_i)}\right]\right\}$$
$$=\lambda_{ij}E_{\overline{\lambda}_i}\left[S_i(\tau_{a_i})-\frac{a_i}{f_i'(\upsilon_i)}\right]^2=\frac{\lambda_{ij}a_i f_i''(\upsilon_i)}{(f_i'(\upsilon_i))^3},\qquad j\neq i;$$

$$(13)\quad E_{\overline{\lambda}_i}\left[N_{ij}(\tau_{a_i})-\frac{\lambda_{ij}a_i}{f_i'(\upsilon_i)}\right]^2=\frac{a_i\lambda_{ij}}{f_i'(\upsilon_i)}\left\{1+\frac{\lambda_{ij}f_i''(\upsilon_i)}{(f_i'(\upsilon_i))^2}\right\},\qquad j\neq i.$$

Define

$$w(\overline{\lambda}_i)=\frac{1}{a_i}\Big[\sum_{j=1,j\neq i}^{m}\mathrm{Var}_{\overline{\lambda}_i}N_{ij}(\tau_{a_i})+\mathrm{Var}_{\overline{\lambda}_i}S_i(\tau_{a_i})\Big]$$
$$=\frac{f_i''(\upsilon_i)(\sum_{j=1,j\neq i}^m\lambda_{ij}^2+1)+(f_i'(\upsilon_i))^2\lambda_{ii}}{(f_i'(\upsilon_i))^3},$$
$$w_1(\overline{\lambda}_i)=\frac{1}{a_i^2}\Big[\sum_{j=1,j\neq i}^{m}(E_{\overline{\lambda}_i}N_{ij}(\tau_{a_i}))^2+(E_{\overline{\lambda}_i}S_i(\tau_{a_i}))^2\Big]$$
$$=\frac{\sum_{j=1,j\neq i}^m\lambda_{ij}^2+1}{(f_i'(\upsilon_i))^2},$$
$$\mathcal{S}(\overline{\lambda}_i)=\frac{w_1(\overline{\lambda}_i)}{w(\overline{\lambda}_i)}=\frac{f_i'(\upsilon_i)(\sum_{j=1,j\neq i}^m\lambda_{ij}^2+1)}{f_i''(\upsilon_i)(\sum_{j=1,j\neq i}^m\lambda_{ij}^2+1)+(f_i'(\upsilon_i))^2\lambda_{ii}}$$
$$=\left[\frac{f_i''(\upsilon_i)}{f_i'(\upsilon_i)}+\frac{f_i'(\upsilon_i)\lambda_{ii}}{\sum_{j=1,j\neq i}^m\lambda_{ij}^2+1}\right]^{-1}$$

and

$$\varrho_{0,i}=\max\{\varrho_i\geq 0:\sup_{\overline{\lambda}_i\in\Lambda_i\times\Upsilon_i}\varrho_i\mathcal{S}(\overline{\lambda}_i)\leq 1\}.$$

In the Bayesian approach to the problem of finding optimal estimation procedures, it is important to give a certain characterization of the prior distributions on $\Lambda_i\times\Upsilon_i$, which should be conjugate to the family (6). We assume certain regularity conditions to hold for the exponential family of distributions of the additive component $A(t)$ of the process considered. Namely, suppose that for any $\varrho_i>\varrho_{0,i}$ and $\beta_i>0$ the following conditions are satisfied:

$$(i1)\quad \int_{\Upsilon_i} f_i''(\upsilon_i)(f_i'(\upsilon_i))^2\exp[\varrho_i\upsilon_i-\beta_i f_i(\upsilon_i)]\,d\upsilon_i<\infty,$$

$$(i2)\quad \int_{\Upsilon_i}(f_i'(\upsilon_i))^{-1}\exp[\varrho_i\upsilon_i-\beta_i f_i(\upsilon_i)]\,d\upsilon_i<\infty,$$

$$(i3)\quad \int_{\Upsilon_i}\frac{d}{d\upsilon_i}\{[f_i''(\upsilon_i)+f_i'(\upsilon_i)(\varrho_i-\beta_i f_i'(\upsilon_i))]\exp[\varrho_i\upsilon_i-\beta_i f_i(\upsilon_i)]\}\,d\upsilon_i=0.$$

The natural prior distribution density of the parameter $\overline{\lambda}_i$ is proportional to

$$(14)\qquad g(\overline{\lambda}_i; r_i, \alpha_i) := \prod_{j=1,\, j\neq i}^{m} \lambda_{ij}^{r_{ij}-1} \exp(-r_{ii}\lambda_{ij}) f_i'(\upsilon_i) \exp[\alpha_i \upsilon_i - r_{ii} f_i(\upsilon_i)],$$

$r_i = (r_{i1}, \ldots, r_{im})$, and it is proper for all $r_{ij} > 0$, $j = 1, \ldots, m$, and each $\alpha_i > 0$ (the integral of (14) with respect to $d\upsilon_i$ is proportional to α_i / r_{ii}; for results relating to conjugate priors for exponential families of processes see Magiera and Wilczyński (1991)).

In solving the problem of finding Γ-minimax sequential procedures, modified priors π^* of the parameter $\overline{\lambda}_i$ will be considered which are defined according to the density

$$(15)\qquad g^*(\overline{\lambda}_i; r_i, \alpha_i) = C(r_i, \alpha_i) w(\overline{\lambda}_i) g(\overline{\lambda}_i; r_i, \alpha_i).$$

It is easy to see that under condition $(i1)$ there exists a norming constant $C(r_i, \alpha_i)$ such that formula (15) represents a probability distribution for all $r_{ij} > 0$, $j = 1, \ldots, m$, and each $\alpha_i > \varrho_{0,i}$. Conditions $(i2)$ and $(i3)$ are needed to derive finite posterior expected loss under the weighted squared error defined below.

The family of priors π^* on $\Lambda_i \times \Upsilon_i =: \overline{\Lambda}_i$ defined by (15) and satisfying the conditions $(i1)$–$(i3)$ will be denoted by $\mathcal{E}^*(r_i, \alpha_i)$. Let E^* stand for the expectation with respect to the distribution π^*. Making use of the regularity conditions on the family $\mathcal{E}^*(r_i, \alpha_i)$ one can obtain the following identities:

$$(16)\qquad \alpha_i E^*\left[\frac{\lambda_{ij}}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right] = r_{ij} E^*\left[\frac{1}{w(\overline{\lambda}_i)}\right], \quad j \neq i,$$

$$(17)\qquad \alpha_i E^*\left[\frac{1}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right] = r_{ii} E^*\left[\frac{1}{w(\overline{\lambda}_i)}\right]$$

and

$$(18)\qquad E^*\left\{\frac{\sum_{j=1,\, j\neq i}^{m} [r_{ij} f_i'(\upsilon_i) - \alpha_i \lambda_{ij}]^2 + [r_{ii} f_i'(\upsilon_i) - \alpha_i]^2}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right\} = \alpha_i.$$

Let the loss function be defined by

$$(19)\qquad \mathcal{L}(\overline{\lambda}_i, d_i) = \frac{1}{w(\overline{\lambda}_i)}\left\{\sum_{\substack{j=1\\ j\neq i}}^{m} \left[d_{ij} - \frac{\lambda_{ij}}{f_i'(\upsilon_i)}\right]^2 + \left[d_{ii} - \frac{1}{f_i'(\upsilon_i)}\right]^2\right\},$$

where $d_i = (d_{i1}, \ldots, d_{im})$, and let the cost function $c(\cdot)$ depend only on the value of the process $A_i(t)$ at the moment of stopping.

In the following Theorem 1 a class of Γ-minimax sequential estimation procedures is established in the case when a special moment condition is imposed on the set of all prior distributions.

THEOREM 1. *Let Γ be a class of all distributions π on $\Lambda_i \times \Upsilon_i$ for which*

$$E[\mathcal{S}(\overline{\lambda}_i)] = M_i \tag{20}$$

with $M_i^{-1} > \varrho_{0,i}$. If there exists a_i^* such that

$$\frac{1}{M_i^{-1} + a_i^*} + c(a_i^*) = \min_{a_i} \left[\frac{1}{M_i^{-1} + a_i} + c(a_i)\right],$$

then the sequential procedure $\delta_{a_i^*} = (\tau_{a_i^*}, d_i^0(\tau_{a_i^*}))$ with $\tau_{a_i^*}$ defined by (7) and

$$d_i^0(\tau_{a_i^*}) = \frac{1}{M_i^{-1} + a_i^*} \times (N_{i1}(\tau_{a_i^*}), \ldots, N_{i,i-1}(\tau_{a_i^*}), S_i(\tau_{a_i^*}), N_{i,i+1}(\tau_{a_i^*}), \ldots, N_{im}(\tau_{a_i^*}))$$

is Γ-minimax.

Proof. Using identities (9)–(13) we find that the risk associated with the estimation error and corresponding to a sequential procedure $\delta_{a_i} = (\tau_{a_i}, d_i^0(\tau_{a_i}))$ with τ_{a_i} defined by (7) and

$$d_i^0(\tau_{a_i}) = \frac{1}{M_i^{-1} + a_i} \times (N_{i1}(\tau_{a_i}), \ldots, N_{i,i-1}(\tau_{a_i}), S_i(\tau_{a_i}), N_{i,i+1}(\tau_{a_i}), \ldots, N_{im}(\tau_{a_i}))$$

is

$$\begin{aligned}
&\mathcal{R}_0(\overline{\lambda}_i, \delta_{a_i}) \\
&\quad = \frac{1}{w(\overline{\lambda}_i)}\Bigg\{ \sum_{\substack{j=1\\ j\neq i}}^{m} E_{\overline{\lambda}_i}\left[\frac{N_{ij}(\tau_{a_i})}{\varrho_{0,i} + a_i} - \frac{\lambda_{ij}}{f_i'(\upsilon_i)}\right]^2 + E_{\overline{\lambda}_i}\left[\frac{S_i(\tau_{a_i})}{\varrho_{0,i} + a_i} - \frac{1}{f_i'(\upsilon_i)}\right]^2\Bigg\} \\
&\quad = \frac{1}{(M_i^{-1} + a_i)^2}\left\{a_i + \frac{\mathcal{S}(\overline{\lambda}_i)}{M_i^2}\right\}.
\end{aligned}$$

Thus for $\pi \in \Gamma$ the Bayes risk is

$$\overline{\mathcal{R}}_0(\pi, \delta_{a_i}) = E[\mathcal{R}_0(\overline{\lambda}_i, \delta_{a_i})] = \frac{1}{M_i^{-1} + a_i}, \tag{21}$$

and this Bayes risk is independent of $\pi \in \Gamma$.

Now, referring to Theorem 5 we show that the constant Bayes risk of (21) is the limit of a sequence of posterior risks associated with a family of prior distributions and the corresponding a_i-Bayes estimators.

Let π_ε^*, $\varepsilon > 0$, be the family of prior distributions $\pi^*(\overline{\lambda}_i; r_i, \alpha_i) \in \mathcal{E}^*(r_i, \alpha_i)$ with

$$r_{ij} = \varepsilon M_{0,ij}, \quad j = 1, \ldots, m; \qquad \alpha_i = \frac{1}{M_i}\Big(1 + \varepsilon \sum_{j=1}^{m} M_{0,ij}^2\Big),$$

where

$$M_{0,ij} = \begin{cases} E^*\left[\dfrac{\lambda_{ij}}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right], & j \neq i, \\ E^*\left[\dfrac{1}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right], & j = i, \end{cases}$$

$j = 1, \ldots, m$. Using (16)–(18) we obtain

$$E^*[\mathcal{S}(\overline{\lambda}_i)] = E^*\left[\frac{\sum_{j=1,\, j\neq i}^m \lambda_{ij}^2 + 1}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right] = M_i,$$

so that $\pi_\varepsilon^* \in \Gamma$.

Set $Z_i(t) = (N_{i1}(t), \ldots, N_{i,i-1}(t), S_i(t), N_{i,i+1}(t), \ldots, N_{im}(t))$. Let τ be any finite stopping time with respect to $\mathcal{F}_t = \sigma\{(Z_i(s), A_i(s)), s \leq t\}, t \geq 0$. It is easy to see that the family $\mathcal{E}^*(r_i, \alpha_i)$ is the conjugate one. Consequently, the posterior probability distribution $\pi^*(\overline{\lambda}_i \mid Z_i(\tau) = z_i,\, A_i(\tau) = a_i)$ of the parameter $\overline{\lambda}_i$, given $Z_i(\tau) = z_i = (n_{i1}, \ldots, n_{i,i-1}, s_i, n_{i,i+1},\ \ldots,\ n_{im})$ and $A_i(\tau) = a_i$, is determined by $\pi^*(\overline{\lambda}_i; \widetilde{r}_i, \widetilde{\alpha}_i)$, where $\widetilde{r}_i = r_i + z_i$ and $\widetilde{\alpha}_i = \alpha_i + a_i$. Let

$$\widetilde{\mathcal{R}}(\pi^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i),\ d_i(z_i, a_i)) = \int_{\overline{\Lambda}_i} \mathcal{L}(\overline{\lambda}_i, d_i)\, d\pi^*(\overline{\lambda}_i \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i)$$

be the posterior risk associated with the prior π^* and an estimator $d_i(z_i, a_i)$. Taking into account (16)–(18), a standard calculation shows that this risk attains its minimum if

$$d_{ij} = d_{ij}^* = \begin{cases} \dfrac{r_{ij} + n_{ij}}{\alpha_i + a_i}, & j \neq i, \\ \dfrac{r_{ii} + s_i}{\alpha_i + a_i}, & j = i. \end{cases}$$

Thus the a_i-Bayes estimator $d_{\varepsilon,i}^* = (d_{\varepsilon,i1}^*, \ldots, d_{\varepsilon,im}^*)$ for π_ε^* is defined by

$$d_{\varepsilon,ij}^* = \begin{cases} \dfrac{\varepsilon M_{0,ij} + N_{ij}}{M_i^{-1}(1 + \varepsilon \sum_{j=1}^m M_{0,ij}^2) + a_i}, & j \neq i, \\ \dfrac{\varepsilon M_{0,ii} + S_i}{M_i^{-1}(1 + \varepsilon \sum_{j=1}^m M_{0,ij}^2) + a_i}, & j = i. \end{cases}$$

The posterior risk associated with π_ε^* and d_ε^* is

$$\widetilde{\mathcal{R}}(\pi_\varepsilon^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i),\ d_{\varepsilon,i}^*) = \frac{1}{M_i^{-1}(1 + \varepsilon \sum_{j=1}^{n_i} M_{0,ij}^2) + a_i}.$$

Thus we obtain (see formula (21))

$$\overline{\mathcal{R}}_0(\pi, \delta_{a_i}) = \lim_{\varepsilon \to 0} \widetilde{\mathcal{R}}(\pi_\varepsilon^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i),\ d_{\varepsilon,i}^*).$$

Now Theorem 5 with $W(z_i, a_i) = a_i$ for each (z_i, a_i), $z_i \in \{1, 2, \ldots\}^{i-1} \times (0, \infty) \times \{1, 2, \ldots\}^{m-i}$, $a_i > 0$, and $K(a_i) = 1/(M_i^{-1} + a_i)$ yields the desired result. ■

In the following Theorem 2 the Γ-minimax sequential estimation procedure is presented in the case when the prior information on the parameter (λ_i, υ_i) is determined by three special moment conditions for this parameter.

THEOREM 2. *Let Γ be the class of all distributions π of the parameter (λ_i, υ_i) for which*

$$E\left[\frac{\lambda_{ij}}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right] = M_{0,ij}, \quad j \neq i; \qquad E\left[\frac{1}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right] = M_{0,ii},$$
$$E\left[\frac{\lambda_{ij}^2}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right] = M_{1,ij}, \quad j \neq i; \qquad E\left[\frac{1}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right] = M_{1,ii},$$
$$E\left[\frac{1}{w(\overline{\lambda}_i)}\right] = M_{2,i},$$

with $\overline{M}_i := \sum_{j=1}^m (M_{1,ij} M_{2,i} - M_{0,ij}^2) > 0$ and $M_{2,i}\overline{M}_i^{-1} > \varrho_{0,i}$. If there exists $a_i^* > 0$ such that

$$\frac{1}{M_{2,i}\overline{M}_i^{-1} + a_i^*} + c(a_i^*) = \min_{a_i}\left[\frac{1}{M_{2,i}\overline{M}_i^{-1} + a_i} + c(a_i)\right],$$

then the sequential procedure $\delta_{a_i^*} = (\tau_{a_i^*}, d_i^0(\tau_{a_i^*}))$ with $\tau_{a_i}^*$ defined by (7) *and*

$$d_{ij}^0(\tau_{a_i^*}) = \begin{cases} \dfrac{M_{0,ij} + \overline{M}_i N_{ij}(\tau_{a_i^*})}{M_{2,i} + \overline{M}_i a_i^*}, & j \neq i, \\[2ex] \dfrac{M_{0,ii} + \overline{M}_i S_i(\tau_{a_i^*})}{M_{2,i} + \overline{M}_i a_i^*}, & j = i, \end{cases}$$

is Γ-minimax.

Proof. A fairly laborious calculation shows that for the sequential procedure $\delta_{a_i} = (\tau_{a_i}, d_i^0(\tau_{a_i}))$ the risk is

$$\mathcal{R}_0(\overline{\lambda}_i, \delta_{a_i}) = \frac{1}{(M_{2,i}\overline{M}_i^{-1} + a_i)^2} \times \left\{a_i + \frac{\sum_{j=1,\, j\neq i}^m [M_{0,ij} f_i'(\upsilon_i) - M_{2,i}\lambda_{i,j}]^2 - [M_{0,ii} f_i'(\upsilon_i) - M_{2,i}]^2}{\overline{M}_i^2 w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right\},$$

and consequently, for $\pi \in \Gamma$, the Bayes risk is

$$\overline{\mathcal{R}}_0(\pi, \delta_{a_i}) = \frac{1}{M_{2,i}\overline{M}_i^{-1} + a_i}, \tag{22}$$

which is independent of $\pi \in \Gamma$.

Denote by π_0^* the prior distribution $\pi^*(\overline{\lambda}_i; r_i, \alpha_i) \in \mathcal{E}^*(r_i, \alpha_i)$ with

$$r_{ij} = \frac{M_{0,ij}}{\overline{M}_i}, \quad j = 1, \ldots, m, \qquad \alpha_i = \frac{M_{2,i}}{\overline{M}_i},$$

where

$$M_{0,ij} = \begin{cases} E^*\left[\dfrac{\lambda_{ij}}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right], & j \neq i, \\ E^*\left[\dfrac{1}{w(\overline{\lambda}_i) f_i'(\upsilon_i)}\right], & j = i, \end{cases}$$

for $j = 1, \ldots, m$.

Using (16)–(18) we obtain

$$E^*\left[\frac{1}{w(\overline{\lambda}_i)}\right] = M_{2,i}, \qquad E^*\left[\frac{\lambda_{ij}^2}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right] = M_{1,ij}, \quad j \neq i,$$

$$E^*\left[\frac{1}{w(\overline{\lambda}_i)(f_i'(\upsilon_i))^2}\right] = M_{1,ii},$$

so that $\pi_0^* \in \Gamma$.

The posterior risk $\widetilde{\mathcal{R}}(\pi_0^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i), d_i(z_i, a_i))$ associated with the prior π_0^* and an estimator $d_i(z_i, a_i)$ attains its minimum if

$$d_{ij} = d_{0,ij}^* = \left\{ \begin{array}{ll} \dfrac{M_{0,ij} + \overline{M}_i n_{ij}}{M_{2,i} + \overline{M}_i a_i}, & j \neq i, \\ \dfrac{M_{0,ii} + \overline{M}_i s_i}{M_{2,i} + \overline{M}_i a_i}, & j = i, \end{array} \right\} =: d_{ij}^0(z_i, a_i).$$

Taking into account (18) leads to the following form of the posterior risk associated with π_0^* and $d_{0,i}^*(z_i, a_i)$:

$$\widetilde{\mathcal{R}}(\pi_0^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i),\ d_{0,i}^*(z_i, a_i)) = \frac{1}{M_{2,i}\overline{M}_i^{-1} + a_i}.$$

Since

$$\overline{\mathcal{R}}_0(\pi, \delta_{a_i}) = \widetilde{\mathcal{R}}(\pi_0^*(\cdot \mid Z_i(\tau) = z_i,\ A_i(\tau) = a_i),\ d_{0,i}^*(z_i, a_i))$$

(see formula (22)), the theorem follows from Theorem 5 by putting $W(z_i, a_i) = a_i$ for each (z_i, a_i), $z_i \in \{1, 2, \ldots\}^{i-1} \times (0, \infty) \times \{1, 2, \ldots\}^{m-i}$, $a_i > 0$ and $K(a_i) = 1/(M_{2,i}\overline{M}_i^{-1} + a_i)$ for each $a_i > 0$. ∎

Note that the sequential procedure $\delta_{a_i^*} = (\tau_{a_i^*}, d_i^0(\tau_{a_i^*}))$ is the only minimax procedure under the loss function given by (19) and the assumed cost function.

Example. In particular, conditions $(i1)$–$(i3)$ are satisfied in the case when the conditional density of $A(t) - A(s)$ given $X(u) = i$, $u \in [s, t]$, is the Poisson density with intensity μ_i, i.e., $f_i(\upsilon_i) = \exp(\upsilon_i)$, where $\upsilon_i = \log \mu_i$.

Since in this case $f_i''(\upsilon_i) = f_i'(\upsilon_i)$ for each $\upsilon_i \in \Upsilon_i = (-\infty, \infty)$, it is easily seen that for the Markov modulated Poisson process, $\varrho_{0,i} = 1$. The condition (20) takes the form

$$E\left\{\left[1 + \frac{\mu_i \lambda_{ii}}{\sum_{j=1, j\neq i}^m \lambda_{ij}^2 + 1}\right]^{-1}\right\} = M_i,$$

where $M_i < 1$. ■

3.2. It will be shown below that to estimate the intensities λ_{ij} ($j = 1, \ldots, m$; $j \neq i$) and the mean value parameter $f_i'(\upsilon_i)$ one can apply a sequential procedure defined by

$$\tau_{s_i} = \inf\{t : S_i(t) = s_i\}, \qquad s_i > 0. \tag{23}$$

For this stopping time the sequential likelihood function (6) takes the form

$$L_i(\tau_{s_i}, \overline{\lambda}_i) = \prod_{j=1, j\neq i}^{m} \lambda_{ij}^{N_{ij}(\tau_{s_i})} \exp\{A_i(\tau_{s_i})\upsilon_i - s_i[\lambda_{ii} + f_i(\upsilon_i)]\}.$$

Suppose that for any $\varrho_i > 0$ and $\beta_i > 0$ the following conditions are satisfied:

$$\int_{\Upsilon_i} f_i''(\upsilon_i) \exp[\varrho_i \upsilon_i - \beta_i f_i(\upsilon_i)]\, d\upsilon_i < \infty, \tag{$s1$}$$

$$\int_{\Upsilon_i} (f_i'(\upsilon_i))^2 \exp[\varrho_i \upsilon_i - \beta_i f_i(\upsilon_i)]\, d\upsilon_i < \infty, \tag{$s2$}$$

$$\int_{\Upsilon_i} \frac{d}{d\upsilon_i}\{[\varrho_i - \beta_i f_i'(\upsilon_i)] \exp[\varrho_i \upsilon_i - \beta_i f_i(\upsilon_i)]\}\, d\upsilon_i = 0. \tag{$s3$}$$

The natural prior distribution of the parameter $\overline{\lambda}_i$ is proportional to

$$\prod_{j=1,\, j\neq i}^{m} \lambda_{ij}^{r_{ij}-1} \exp(-\alpha_i \lambda_{ij}) \exp[r_{ii}\upsilon_i - \alpha_i f_i(\upsilon_i)], \tag{24}$$

and it is proper for all $r_{ij} > 0$, $j = 1, \ldots, m$, and each $\alpha_i > 0$ (the integral of (24) with respect to $d\upsilon_i$ is finite for each $r_{ii} > 0$ and $\alpha_i > 0$). The modified prior, i.e., (24) multiplied by the sum $\sum_{j=1,\, j\neq i}^m \lambda_{ij} + f_i''(\upsilon_i) = \lambda_{ii} + f_i''(\upsilon_i)$, is proper under condition ($s1$) for all $r_{ij} > 0$, $j = 1, \ldots, m$, and each $\alpha_i > 0$.

Let the loss function be defined by

$$\mathcal{L}(\overline{\lambda}_i, d_i) = [\lambda_{ii} + f_i''(\upsilon_i)]^{-1}\Big[\sum_{j=1,\, j\neq i}^{m} (d_{ij} - \lambda_{ij})^2 + (d_{ii} - f_i'(\upsilon_i))^2\Big]$$

and let the cost function $c(\cdot)$ depend only on the value of the process $S_i(t)$ at the moment of stopping.

Taking advantage of Theorem 5, in an analogous way to Theorem 1 one obtains the following result.

THEOREM 3. *Let $\varGamma$ be the class of all distributions π on $\varLambda_i \times \varUpsilon_i$ for which*

$$E\left[\frac{\sum_{j=1,\,j\neq i}^{m}\lambda_{ij}^2+(f_i(\upsilon_i))^2}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_i$$

with $M_i^{-1}>0$. If there exists s_i^* such that

$$\frac{1}{M_i^{-1}+s_i^*}+c(s_i^*)=\min_{s_i}\left[\frac{1}{M_i^{-1}+s_i}+c(s_i)\right],$$

then the sequential procedure $\delta_{s_i^*}=(\tau_{s_i^*},d_i^0(\tau_{s_i^*}))$ with $\tau_{s_i^*}$ defined by (23) *and*

$$\begin{aligned} d_i^0(\tau_{s_i^*}) = {} & \frac{1}{M_i^{-1}+s_i^*}\\ & \times (N_{i1}(\tau_{s_i^*}),\ldots,N_{i,i-1}(\tau_{s_i^*}),A_i(\tau_{s_i^*}),N_{i,i+1}(\tau_{s_i^*}),\ldots,N_{im}(\tau_{s_i^*}))\end{aligned}$$

is $\varGamma$-minimax.

Arguments analogous to those of Theorem 2 yield the following result.

THEOREM 4. *Let $\varGamma$ be the class of all distributions π of the parameter ϑ for which*

$$\begin{aligned} &E\left[\frac{\lambda_{ij}}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_{0,ij},\quad j\neq i, && E\left[\frac{f_i'(\upsilon_i)}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_{0,ii},\\ &E\left[\frac{\lambda_{ij}^2}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_{1,ij},\quad j\neq i, && E\left[\frac{(f_i'(\upsilon_i))^2}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_{1,ii},\\ &E\left[\frac{1}{\lambda_{ii}+f_i''(\upsilon_i)}\right]=M_{2,i}, \end{aligned}$$

with $\overline{M}_i:=\sum_{j=1}^{m}(M_{1,ij}M_{2,i}-M_{0,ij}^2)>0$ and $M_{2,i}\overline{M}_i^{-1}>0$. If there exists $s_i^*>0$ such that

$$\frac{1}{M_{2,i}\overline{M}_i^{-1}+s_i^*}+c(s_i^*)=\min_{s_i}\left[\frac{1}{M_{2,i}\overline{M}_i^{-1}+s_i}+c(s_i)\right],$$

then the sequential procedure $\delta_{s_i^*}=(\tau_{s_i^*},d^0(\tau_{s_i^*}))$ with $\tau_{s_i}^*$ defined by (23) *and*

$$d_{ij}^0(\tau_{s_i^*})=\begin{cases}\dfrac{M_{0,ij}+\overline{M}_iN_{ij}(\tau_{s_i^*})}{M_{2,i}+\overline{M}_is_i^*}, & j\neq i,\\[2ex] \dfrac{M_{0,ii}+\overline{M}_iA_i(\tau_{s_i^*})}{M_{2,i}+\overline{M}_is_i^*}, & j=i,\end{cases}$$

is $\varGamma$-minimax.

For example, conditions ($s1$)–($s3$) are satisfied for the Poisson density with intensity μ_i, i.e., if $f_i(\upsilon_i)=\exp(\upsilon_i)$, $\upsilon_i=\log\mu_i$.

REMARK 1. In the limit case of Theorem 1 when $1/M_i\to\varrho_{0,i}$ (resp. Theorem 3 when $1/M_i\to 0$) or in the limit case of Theorem 2 when

$M_{2,i}/\overline{M}_i \to \varrho_{0,i}$ (resp. Theorem 4 when $M_{2,i}/\overline{M}_i \to 0$), the expressions for the Γ-minimax sequential estimation procedures imply the forms for the minimax procedures obtained by Magiera (1999).

The idea and tools are presented to solve the problem considered in the case when only one row of the matrix (λ, υ) is unknown. We then use the likelihood function defined by (6) which has a much simpler form than that of (4). For estimable functions involving elements of more than one row of the matrix (λ, υ) the problem can be solved using an appropriate stopping time, but this requires more intricate calculations.

Considering, for example, a three-state Markov-additive process one can use the stopping time

$$\tau^1_{J,s} = \inf\{t : N_{21}(t) + N_{31}(t) = s\},$$

$s = 1, 2, \ldots$; $J = \{2, 3\}$, $\sum_{j\in J} \lambda_{j1} > 0$. One should then take into account the following relations (see formulas (5)):

$$\begin{aligned} &N_{21}(\tau^1_{J,s}) + N_{31}(\tau^1_{J,s}) = s,\\ &N_{21}(\tau^1_{J,s}) + N_{31}(\tau^1_{J,s}) = N_{12}(\tau^1_{J,s}) + N_{13}(\tau^1_{J,s}),\\ &N_{12}(\tau^1_{J,s}) + N_{32}(\tau^1_{J,s}) = N_{21}(\tau^1_{J,s}) + N_{23}(\tau^1_{J,s}), \end{aligned}$$

which reduce the exponential family (4) to a noncurved exponential model for estimating the parameters associated with all the rows of the matrix (λ, υ).

4. Appendix. Let $X(t)$, $t \in \mathcal{T}$, be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, P_\lambda)$ with values in $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$, where $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = \{0, 1, 2, \ldots\}$ and λ is a parameter with values in an open set $\Lambda \subseteq \mathbb{R}^n$. Let $P_{\lambda,t}$ denote the restriction of P_λ to $\mathcal{F}_t = \sigma\{X(s) : s \le t\}$. Suppose that the family $P_{\lambda,t}$, $\lambda \in \Lambda$, is dominated by a measure μ_t which is the restriction of a probability measure μ to $\mathcal{F}_t$, and that the density functions (likelihood functions) have the form

$$\frac{dP_{\lambda,t}}{d\mu_t} = L(Y(t), t, \lambda),$$

where $Y(t)$, $t \in \mathcal{T}$, is a process with values in $(\mathbb{R}^k, \mathcal{B}_{\mathbb{R}^k})$ and adapted to the filtration $\mathcal{F}_t$, $t \in \mathcal{T}$, and $L(\cdot, \cdot, \lambda)$ is a continuous function. In the case of continuous time it is supposed that the process $Y(t), t \in \mathcal{T}$, has Skorokhod paths, i.e., paths which are right-continuous and have left-hand limits.

Let τ be a stopping time relative to $\mathcal{F}_t$, $t \in \mathcal{T}$, such that $P_\lambda(\tau < \infty) = 1$ for each $\lambda \in \Lambda$. The random variable $(Y(\tau), \tau)$ is a sufficient statistic for λ relative to $\mathcal{F}_\tau$. Let $P_{\lambda,\tau}$ and μ_τ denote the restrictions of P_λ and μ to $\mathcal{F}_\tau$.

By the fundamental identity of sequential analysis, $P_{\lambda,\tau} \ll \mu_\tau$ and

$$\frac{dP_{\lambda,\tau}}{d\mu_\tau} = L(Y(\tau), \tau, \lambda).$$

Suppose that, in addition to the loss associated with the estimation error, the statistician incurs a cost of observation of the process. Let $\mathcal{L}(\lambda, d(Y(t), t), Y(t), t)$ be the loss function determining the loss incurred by the statistician if λ is the true value of the parameter and d is the estimator chosen by him having observed $Y(t)$ at the moment of stopping t. Denote by $c(Y(t), t)$ the cost function which represents the cost of observing the process up to time t.

Sequential procedures of the form $\delta = (\tau, d)$ for estimating the value of a function of λ will be considered, where τ is a finite stopping time with respect to $\mathcal{F}_t$, $t \in \mathcal{T}$, and $d = d(Y(\tau), \tau)$ is an $\mathcal{F}_\tau$-measurable random variable. The risk of the sequential procedure δ corresponding to the estimation error is

$$\mathcal{R}_0(\lambda, \delta) = E_\lambda \mathcal{L}(\lambda, d(Y(\tau), \tau), Y(\tau), \tau).$$

The total risk function of the sequential procedure $\delta = (\tau, d)$ is defined by

$$\mathcal{R}(\lambda, \delta) = \mathcal{R}_0(\lambda, \delta) + E_\lambda c(Y(\tau), \tau).$$

By the fundamental identity of sequential analysis the expectations (risks) are well defined for randomly stopped processes.

Let π be a prior distribution of the parameter on the space $(\Lambda, \mathcal{B}_\Lambda)$ and let Γ be a set of prior distributions. Suppose that $\mathcal{R}(\lambda, \delta)$ is a $\mathcal{B}_\Lambda$-measurable function of λ. Then the Bayes risk is defined by

$$\overline{\mathcal{R}}(\pi, \delta) = \int_\Lambda \mathcal{R}(\lambda, \delta)\, \pi(d\lambda) = \overline{\mathcal{R}}_0(\pi, \delta) + \int_\Lambda E_\lambda c(Y(\tau), \tau)\, d\pi(\lambda),$$

provided the integral exists, where

$$\overline{\mathcal{R}}_0(\pi, \delta) = \int_\Lambda \mathcal{R}_0(\lambda, \delta)\, d\pi(\lambda).$$

Below, only such sequential procedures $\delta = (\tau, d)$ will be considered for which $\overline{\mathcal{R}}(\pi, \delta) < \infty$ for each $\pi \in \Gamma$. The class of all sequential procedures satisfying this condition will be denoted by $\overline{\mathcal{D}}$. The problem is to find optimal stopping rules τ and the corresponding estimators $d(\tau) = d(Y(\tau), \tau)$ subject to the minimax criterion: a sequential procedure $\delta_0 = (\tau_0, d_0)$ is said to be *Γ-minimax* if

$$\sup_{\pi \in \Gamma} \overline{\mathcal{R}}(\pi, \delta_0) = \inf_{\delta \in \overline{\mathcal{D}}} \sup_{\pi \in \Gamma} \overline{\mathcal{R}}(\pi, \delta).$$

This problem can be interpreted as the problem of choosing a minimax strategy in the game of the statistician against the nature where the payoff function is the Bayes risk associated with the given loss function $\mathcal{L}$ and the cost function c. The strategy of the nature is the prior distribution π of the parameter λ. It is assumed that the set of all strategies of the nature is reduced to a set Γ.

Let $\pi(\cdot \mid Y(\tau) = y,\ \tau = t)$ denote the posterior distribution of the parameter λ given $Y(\tau) = y$, $\tau = t$. The *posterior risk* corresponding to π and an estimator d is determined by

$$\widetilde{\mathcal{R}}(\pi(\cdot \mid Y(\tau) = y,\ \tau = t),\ d) = \int_{\Lambda} \mathcal{L}(\lambda, d(y,t), y, t)\, d\pi(\lambda \mid Y(\tau) = y,\ \tau = t).$$

An estimator $d^* = d^*(y,t)$ is called (y,t)-*Bayes* for π if

$$\widetilde{\mathcal{R}}(\pi(\cdot \mid Y(\tau) = y,\ \tau = t),\ d^*) = \inf_{d} \widetilde{\mathcal{R}}(\pi(\cdot \mid Y(\tau) = y,\ \tau = t),\ d)$$

for all $(y,t) \in U = \mathbb{R}^k \times \mathcal{T}$.

Denote by $W(y,t)$ a measurable mapping from $(U, \mathcal{B}_U)$ into $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

A minimax theorem will be presented which is a slight extension of Theorem 1 of Wilczyński (1985) and a considerable generalization of the familiar tool given by Dvoretzky, Kiefer and Wolfowitz (1953) in statistical decision theory.

THEOREM 5. *Assume that the cost function is of the form* $c(y,t) = c(W(y,t))$. *Suppose that there exists a sequence of priors* $\pi_n \in \Gamma$, $n = 1, 2, \ldots$, *of the parameter* λ *for which there are corresponding* (y,t)-*Bayes estimators* d_n^* *such that*

$$\liminf_{n\to\infty} \widetilde{\mathcal{R}}(\pi_n(\cdot \mid Y(\tau) = y,\ \tau = t),\ d_n^*) = K(W(y,t))$$

for each $(y,t) \in U$, *where* K *is a real-valued measurable function defined on* $\mathbb{R}$. *Moreover, assume that* $K(\cdot) + c(\cdot)$ *attains its minimum over* $\mathcal{Z}$ *at a point* z^*, *where* $\mathcal{Z}$ *is the set of values of the process* $W(Y(t), t)$, $t \in \mathcal{T}$. *If*

$$\tau_{z^*} = \inf\{t \in \mathcal{T} : W(Y(t), t) = z^*\}$$

is a finite stopping time for each $\lambda \in \Lambda$, *and if there exists an estimator* $d_{z^*}(\tau_{z^*}) = d_{z^*}(Y(\tau_{z^*}), \tau_{z^*})$ *such that*

$$\sup_{\pi \in \Gamma} \overline{\mathcal{R}}(\pi, \delta_{z^*}) \le K(z^*) + c(z^*),$$

where $\delta_{z^*} = (\tau_{z^*}, d_{z^*}(\tau_{z^*}))$, *then the sequential procedure* δ_{z^*} *is* Γ-*minimax under the loss function* $\mathcal{L}(\lambda, d(Y(t), t), Y(t), t)$ *in the class of all sequential procedures* $\delta = (\tau, d(\tau)) \in \overline{\mathcal{D}}$.

The proof of the theorem is the same as that of Theorem 1 of Wilczyński (1985), the only difference being that the more general form $\mathcal{L}(\lambda, d(Y(t), t), Y(t), t)$ of the loss function is taken into account and instead of $\mathcal{R}(\lambda, \delta)$ one should consider $\overline{\mathcal{R}}(\pi, \delta)$.

References

O. Barndorff-Nielsen (1978), *Information and Exponential Families*, Wiley, New York.

L. Brown (1986), *Fundamentals of Statistical Exponential Families*, IMS, Hayward, CA.

E. Çinlar (1972), *Markov additive processes. I*, Z. Wahrsch. Verw. Gebiete 24, 85–93.

A. Dvoretzky, J. Kiefer and J. Wolfowitz (1953), *Sequential decision problems for processes with continuous time parameter. Problems of estimation*, Ann. Math. Statist. 24, 403–415.

I. Ezhov and A. Skorokhod (1969), *Markov processes with homogeneous second component. I*, Teor. Veroyatn. Primenen. 14, 3–14.

M. Fygenson (1991), *Optimal sequential estimation for semi-Markov and Markov renewal processes*, Comm. Statist. Theory Methods 20, 1427–1444.

R. Gill (1980), *Nonparametric estimation based on censored observation of a Markov renewal process*, Z. Wahrsch. Verw. Gebiete 53, 97–116.

R. Magiera (1999), *Minimax sequential procedures for Markov-additive processes*, Stochastic Models 15, 871–888.

R. Magiera and M. Wilczyński (1991), *Conjugate priors for exponential-type processes*, Statist. Probab. Lett. 12, 379–384.

M. Neuts (1992), *Models based on the Markovian arrival process*, IEICE Trans. Comm. E75-B, 1255–1265.

A. Pacheco and N. Prabhu (1995), *Markov-additive processes of arrivals*, in: J. H. Dshalalow (ed.), Advances in Queueing: Theory, Methods, and Open Problems, Chap. 6, CRC Press, Boca Raton, 167–194.

M. Phelan (1990a), *Bayes estimation from a Markov renewal process*, Ann. Statist. 18, 603–616.

M. Phelan (1990b), *Estimating the transition probabilities from censored Markov renewal processes*, Statist. Probab. Lett. 10, 43–47.

N. Prabhu (1991), *Markov renewal and Markov-additive processes*, Technical Report 984, College of Engineering, Cornell Univ.

T. Rydén (1994), *Parameter estimation for Markov modulated Poisson processes*, Comm. Statist. Stochastic Models 10, 795–829.

V. Stefanov (1995), *Explicit limit results for minimal sufficient statistics and maximum likelihood estimators in some Markov processes: exponential families approach*, Ann. Statist. 23, 1073–1101.

M. Wilczyński (1985), *Minimax sequential estimation for the multinomial and gamma processes*, Zastos. Mat. 18, 577–595.

Institute of Mathematics
Wrocław University of Technology
Wybrzeże Wyspiańskiego 27
50-370 Wrocław, Poland
E-mail: magiera@im.pwr.wroc.pl

Received on 11.7.2001 (1587)