Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

P.J.C. Spreij

An on-line parameter estimation algorithm for counting process observations

Department of Operations Research and System Theory Report OS-R8406 May

AN ON-LINE PARAMETER ESTIMATION ALGORITHM FOR COUNTING PROCESS OBSERVATIONS

P.J.C. SPREIJ
Centre for Mathematics and Computer Science, Amsterdam

The parameter estimation problem for counting process observation is considered. It is assumed that the intensity of the counting process is adapted to the family of σ-algebras generated by the counting process itself. Furthermore we assume that the intensity depends linearly on some deterministic constant parameters. An on-line parameter estimation algorithm is then presented for which convergence is proved by using a stochastic approximation type lemma.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 60G55, 62M05, 93E12.
KEY WORDS & PHRASES: counting process, martingale, estimation, stochastic Lyapunov function.
NOTE: This report will be submitted for publication elsewhere.

Report OS-R8406
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. Introduction

Counting processes frequently occur as the observation processes in mathematical models for industrial processes or in biology, software engineering and nuclear medicine. Usually such a counting process can be considered as the output process of some stochastic system. The underlying state process then influences the counting process. The purpose is then to estimate this state given the observations. This problem is known as the filtering problem and has been investigated extensively [1].

The solution of this problem requires knowledge of the involved parameters, which means that one can compute the solution to the filtering problem in a practical case only if one knows the correct parameter values. Unfortunately, in many cases these are not known and are therefore to be estimated. This may happen before the processes start running on related additional information and/or on the basis of the observations. In the latter case some asymptotic results for off-line maximum likelihood estimation are available [3,4].

It is the purpose of the present paper to give a contribution to the on-line parameter estimation problem in a specific case. The approach undertaken has proven to be fruitful in for instance discrete time ARMAX processes [7] or continuous time Gaussian AR processes [6].

The paper is organized as follows. In section 2 we give some basic results for counting processes. In section 3 we give a heuristic derivation of our parameter estimation algorithm. Section 4 contains the convergence proof of the algorithm.

2. Preliminary results

2.1. We assume that we are given a complete probability space $(\Omega,\mathcal{F},P)$, a time set $T=[0,\infty)$ and a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions in the sense of [2]. All stochastic processes in the sequel are defined on $\Omega\times T$ and adapted to $\{\mathcal{F}_t\}_{t\geq 0}$. We study the case that we are given an observed process which is a counting process, that is a map $n:\Omega\times T\rightarrow\mathbb{N}_0$ which has only jumps of magnitude $+1$. Then it is known [1,2] that n is a submartingale and therefore admits the so called Doob-Meyer decomposition (w.r.t. $\{\mathcal{F}_t\}_{t\geq 0}$)

$$n_t = \Lambda_t + m_t \tag{2.1}$$

where $\Lambda:\Omega\times T\rightarrow\mathbb{R}$ is a predictable increasing process and m a local martingale. Now assume that Λ is an absolutely continuous process, say $\Lambda_t=\int_0^t\lambda_s\,ds$ then we can rewrite (2.1) as

$$dn_t=\lambda_t\,dt+dm_t \tag{2.2}$$

The process λ is called the intensity process.

A major problem for counting process observations is usually to identify the intensity process λ. This problem can be set up in two stages. In the first stage we have to solve a filtering problem. To be precise we have to determine $\hat\lambda_t=E(\lambda_t|\mathcal{F}_t^n)$, where $\mathcal{F}_t^n=\sigma\{n_s,s\leq t\}$. Then $\hat\lambda_t$ is the optimal (in the sense of mean squared error) estimate given the observations during $[0,t]\subset T$ and given the values of deterministic parameters. We can then replace (2.2) by the minimal decomposition of n (i.e. with respect to $\{\mathcal{F}_t^n\}$)

$$dn_t=\hat\lambda_t\,dt+d\overline{m}_t \tag{2.3}$$

where $\overline{m}$ is a local martingale adapted to $\{\mathcal{F}_t^n\}_{t\geq 0}$. In the second stage one looks for estimates of left unknown deterministic parameters. If one adopts the maximum likelihood criterion, (2.3) and the computation of $\hat\lambda_t$ appear to be crucial. The likelihood functional in this case is known [1, p. 174] to be

$$L_t=\exp[-\int_0^t(\hat\lambda_s-1)ds+\int_0^t\log\hat\lambda_{s-}dn_s]. \tag{2.4}$$

2.2 The model.

From here on we assume that $\hat\lambda$ has a special structure

$$\hat\lambda_t=p^T\phi_t \tag{2.5}$$

where $p \in \mathbb{R}^m$ is the vector of unknown parameters and $\phi:\Omega \times T \to \mathbb{R}$ is a process adapted to $\{\mathcal{F}_t^n\}_{t \geq 0}$. The minimal decomposition (2.3) now becomes

$$dn_t = p^T \phi_t \, dt + d\overline{m}_t. \tag{2.6}$$

Plugging (2.5) into (2.4) and writing $L_t(p)$ instead of L_t in order to express the dependence of the likelihood functional on p, we get

$$L_t(p) = \exp[-p^T \int_0^t \phi_s \, ds + t + \int_0^t \log(p^T \phi_{s-}) dn_s]. \tag{2.7}$$

3. Derivation of the algorithm

In this section we state a parameter estimation algorithm for the model (2.5), (2.6). The proof that the parameter estimates given by this algorithm indeed converge to the true parameter value will be given in section 4. The algorithm is constucted in such a way that the estimates $\hat{p}_t$ of p approximately maximize the likelihood functional (2.7), or equivalently, minimize $J_t(\cdot)$ given by

$$J_t(p) = p^T \int_0^t \phi_s \, ds - \int_0^t \log(\phi_{s-}^T p) dn_s. \tag{3.1}$$

After posing the algorithm we present a heuristic derivation.

3.1 Algorithm.

Consider the model (2.5), (2.6). A recursive maximum likelihood parameter estimation algorithm is given by

$$d\hat{p}_t = R_t \phi_{t-} (dn_t - \phi_t^T \hat{p}_t \, dt), \; \hat{p}_0 \tag{3.2}$$

$$dR_t = -R_t \phi_t \phi_t^T R_t \, dt, \; R_0 \tag{3.3}$$

The interpretation is that for each t $\hat{p}_t$ approximately minimizes $J_t(\cdot)$ as stated above and that R_t is up to a multiplicative scalar factor an approximation of the second derivative of $J_t(\cdot)$. Thus (3.2), (3.3) can be considered as a quasi Newton scheme for minimizing the family of functions $\{J_t(\cdot)\}_{t \geq 0}$. Observe that it follows from (3.3) that R_t stays a positive definite matrix when the initial value R_0 is chosen to be symmetric and positive definite.

3.2. To understand the algorithm (3.2), (3.3) it is useful to consider first a non-stochastic situation. To be precise let $J:\mathbb{R}_+ \times \mathbb{R}^m \to \mathbb{R}$, $J \in C^2(\mathbb{R}_+ \times \mathbb{R}^m, \mathbb{R})$ such that $J(t,\cdot):\mathbb{R}^m \to \mathbb{R}$ has a unique minimum, attained for say $x(t)$. Under some regularity conditions it then follows from the implicit function theorem that the function $t \to x(t)$ satisfies the differential equation

$$dx(t) = -\left[\frac{\partial^2}{\partial x^2} J(t, x(t))\right]^{-1} \frac{\partial^2}{\partial x \, \partial t} J(t, x(t)) dt. \tag{3.4}$$

Let us now return to our estimation problem, that is finding the value $\hat{p}_t$ that for each t minimizes (3.1). For an evolution equation for $\hat{p}_t$ one then tries to find an equation like (3.4). However the functional J of (3.1) does not satisfy the desired smoothness conditions and therefore one has to look for something related to (3.4). Our choice is

$$d\hat{p}_t = -\left[J_{t-}''(\hat{p}_{t-})\right]^{-1} \partial_t J_t'(\hat{p}_t) \tag{3.5}$$

where prime denotes partial differentiation with respect to p and ∂_t means the partial forward differential operator with respect to t. In order to fully specify the algorithm we also need recursive expressions for $J_t''(\hat{p}_t)$ and $J_t'(\hat{p}_t)$. Later on we will establish almost sure convergence of the family $\{\hat{p}_t\}_{t \geq 0}$ to the true parameter value p_0.

From (3.1) we get by formal differentiation

$$J_t'(p)=\int_0^t \phi_s\,ds - \int_0^t \frac{\phi_{s-}}{p^T\phi_{s-}}dn_s, \quad hence \tag{3.6}$$

$$\partial_t J_t'(p)=\phi_t\,dt - \frac{\phi_{t-}}{p^T\phi_{t-}}dn_t. \tag{3.7}$$

Define $k_t=\dfrac{\phi_t}{\hat{p}_t^T\phi_t}$ and $Q_t=[J_t''(\hat{p}_t)]^{-1}$. Using these expressions and (3.7) we can rewrite (3.5) as

$$d\hat{p}_t=Q_{t-}k_{t-}(dn_t-\phi_t^T\hat{p}_t\,dt) \tag{3.8}$$

The next problem is the finding of a recursion for Q_t. It turns out that an exact equation for Q_t cannot be obtained for $p\in\mathbb{R}^m$ with $m\geqslant 2$ and that certain approximations are not satisfactory in that these cause problems in analyzing the convergence properties of the algorithm.

On the other hand the case of $p\in\mathbb{R}^1$ is easy to handle and it will be illustrative for the multivariable case. In this case (3.6) reads as

$$J_t'(p)=\int_0^t \phi_s\,ds - \frac{dn_t}{p}, \quad hence \tag{3.9}$$

$$J_t''(p)=\frac{n_t}{p^2}. \tag{3.10}$$

Therefore Q_t becomes $\dfrac{\hat{p}_t^2}{n_t}$ and with $k_t=\dfrac{1}{\hat{p}_t}$ (3.8) reads as

$$d\hat{p}_t=\frac{\hat{p}_{t-}}{n_{t-}}(dn_t-\phi_t\hat{p}_t\,dt). \tag{3.11}$$

Observe that $\hat{p}_t=\dfrac{n_t}{\Phi_t}$, where $\Phi_t=\int_0^t\phi_s\,ds$ satisfies (3.11) and this value for $\hat{p}_t$ is also found by directly minimizing (3.1). One can prove that $\hat{p}_t$ given by (3.11) converges to the true parameter value, using the method of section 4.

Applying the stochastic calculus rule to $Q_t=\dfrac{\hat{p}_t^2}{n_t}$ one can verify that Q_t satisfies

$$dQ_t=-2Q_t^2k_t\phi_t\,dt+Q_{t-}^2k_{t-}^2\,dn_t. \tag{3.12}$$

Returning to the multivariate case $p\in\mathbb{R}^m$, $m\geqslant 2$ one would like to extend (3.12) in order to obtain an evolution equation for Q_t. This suggests

$$dQ_t=-2Q_tk_t\phi_t^TQ_t\,dt+Q_{t-}k_{t-}k_{t-}^TQ_{t-}\,dn_t. \tag{3.13}$$

One hopes that (3.8) together with (3.13) constitutes the desired algorithm. Although (3.8), (3.13) yield some appealing properties compared to the scalar case $p\in\mathbb{R}^1$ such as $\hat{p}_t=Q_t\Phi_t$, $\hat{p}_t^TQ_t^{-1}\hat{p}_t=n_t$ and $\Phi_t^T\hat{p}_t=n_t$ we were not able to prove the desired convergence properties. The major bottleneck was the verification of the technical condition (see (4.5))

$$\int_0^\infty Q_{t-}k_{t-}k_{t-}^TQ_{t-}\,dn_t<\infty, \tag{3.14}$$

which is however a trivial excercise if $p\in\mathbb{R}^1$. The main cause of this technical problem was the term $\phi_t^T\hat{p}_t$ in the denominator of k_t. Therefore we tried to incorporate this term in Q_t such that $Q_tk_t=R_t\phi_t$, for a matrix valued process R_t and the idea was then to find an equation for R_t.

Inspection of the case $p\in\mathbb{R}^1$, neglection of the derivatives of ϕ and the formula $Q_tk_t=R_t\phi_t$ then leads from (3.13) to

$$dR_t = -R_t \phi_t \phi_t^T R_t \, dt. \tag{3.3}$$

4. Convergence proof

In this section we present a convergence proof for the algorithm (3.2), (3.3) which establishes almost sure convergence of the parameter estimates to the true parameter value. The proof is completely in the spirit of the proofs in [6,7]. We begin with stating an important technical lemma, which is a simple version of a more general result in [6], that in turn can be considered as the continuous time counterpart of a result in discrete time stochastic approximation [5].

4.1 Lemma. *Let x, a, b be nonnegative stochastic processes and m a local martingale such that $x = a - b + m$ and assume that*
(i) a and b are increasing processes with $a_0 = b_0 = 0$,
(ii) $\exists c \in \mathbb{R}_+$ such that $\Delta a \leqslant c$ a.s.
(iii) $\lim_{t\to\infty} a_t < \infty$ a.s.
Then
(a) $\lim_{t\to\infty} x_t$ exists and is finite a.s.
(b) $\lim_{t\to\infty} b_t$ is finite a.s.

4.2. Here is our main result:

Theorem *Consider the algorithm (3.2), (3.3). Let p_0 be the true parameter value. Let $\tilde{p}_t = \hat{p}_t - p_0$ and let $\psi_t = \phi_t^T \phi_t$, $\Psi_t = \int_0^t \psi_s \, ds + tr(R_0^{-1})$.*
Assume:
(i) $\lim_{t\to\infty} \Psi_t = \infty$ a.s.
(ii) $\int_0^\infty \Psi_t^{-2} \psi_t \phi_t \, dt < \infty$ a.s.
(iii) $\lim_{t\to\infty} \Psi_t^{-1} \int_0^t \psi_s \psi_s^T as = C$, where $C \in \mathbb{R}^{m \times m}$ is positive definite a.s. Then
(a) $\lim_{t\to\infty} \hat{p}_t = p_0$ a.s.
(b) $\lim_{t\to\infty} \Psi_t^{-1} \int_0^t (\phi_s^T \tilde{p}_s)^2 ds = 0$ a.s.

Proof. From (3.2), (3.3) it follows that

$$d\tilde{p}_t = R_t \phi_{t-}(dn_t - \phi_t^T \hat{p}_t \, dt) = R_t \phi_{t-}(d\overline{m}_t - \phi_t^T \tilde{p} dt) \tag{4.1}$$

$$dR_t^{-1} = \phi_t \phi_t^T dt \tag{4.2}$$

Define the Lyapunov like process $u_t = \tilde{p}_t^T R_t^{-1} \tilde{p}_t + \int_0^t (\tilde{p}_s^T \phi_s)^2 ds$. Applying the stochastic calculus rule to u_t, we obtain

$$du_t = 2\tilde{p}_{t-}^T \phi_{t-} d\overline{m}_t + \phi_{t-}^T R_t \phi_{t-} dn_t. \tag{4.3}$$

Observe that $\Psi_t = tr(R_t^{-1})$. Define $w_t = u_t \Psi_t^{-1}$, then

$$dw_t = -\Psi_t^{-1} w_t \psi_t \, dt + \phi_t^T R_t \phi_t \Psi_t^{-1} p_0^T \phi_t \, dt + dm_{1t}, \tag{4.4}$$

where m_1 is a local martingale. We want now to apply lemma 4.1 to equation (4.4). Becuase u, w, Ψ are positive, we then see that the only thing we have to check is assumption 4.1.iii,

$$\int_0^\infty \phi_t^T R_t \phi_t \Psi_t^{-1} p_0^T \phi_t \, dt < \infty \tag{4.5}$$

To that end, let $\rho_t = trR_t$. Let γ_{it} be one of the eigenvalues of R_t^{-1}, then $\lim_{t\to\infty}\Psi_t^{-1}\gamma_{it} = c_i > 0$ by assumption (iii) of the theorem. Hence $\gamma_{it} = c_i\Psi_t(1+o(1))$, $(t\to\infty)$. Now γ_{it}^{-1} is an eigenvalue of R_t, $\gamma_{it}^{-1} = c_i^{-1}\Psi_t^{-1}(1+o(1)),(t\to\infty)$. Hence $\rho_t = \Psi_t^{-1}(\Sigma c_i^{-1}+o(1))(t\to\infty)$, or $\rho_t = o(\Psi_t^{-1}),(t\to\infty)$. Recall that for a positive definite matrix A, $x^TAx \leqslant x^Tx.tr(A)$ and $x^TA^2x \leqslant x^Tx(tr(A))^2$. Then

$$\int_0^\infty \phi_t^T R_t \phi_t \Psi_t^{-1} p_0^T \phi_t dt = \int_0^\infty \phi_t^T R_t R_t^{-1} R_t \phi_t \Psi_t^{-1} p_0^T \phi_t dt \leqslant$$

$$\leqslant \int_0^\infty \phi_t^T R_t^2 \phi_t p_0^T \phi_t dt \leqslant \int_0^\infty \phi_t^T \phi_t \rho_t^2 p_0^T \phi_t =$$

$$p_0^T \int_0^\infty \psi_t \rho_t^2 \phi_t dt = p_0^T \int_0^\infty \phi_t \psi_t O(\Psi_t^{-2}) dt < \infty, \textit{ by assumption (ii)}.$$

Then from lemma 4.1 we conclude that w and $\int_0^\infty w_s \Psi_s^{-1} \psi_s ds$ almost surely converge. We claim that $\lim_{t\to\infty} w_t = 0$ a.s. If not, there exists a subset of Ω with positive probability and an $\epsilon > 0$, such that $\lim_{t\to\infty} w_t \geqslant 2\epsilon$ on this subset. But then we also have on the same subset

$$\int_0^\infty \Psi_t^{-1} w_t \psi_t dt \geqslant \epsilon \int_0^\infty \Psi_t^{-1} \psi_t dt = \Big[log(\Psi_t)\Big]_0^\infty = \infty, \textit{ by assumption (i)}.$$

But this contradicts the second assertion of lemma 4.1. Since w is the sum of two positive quantities we have both

$$\lim_{t\to\infty} \Psi_t^{-1} \int_0^t (\tilde{p}_s^T \phi_s)^2 ds = 0 \ a.s. \text{ and} \tag{4.6}$$

$$\lim_{t\to\infty} \tilde{p}_t^T \frac{R_t^{-1}}{\Psi_t} \tilde{p}_t = 0 \ a.s. \tag{4.7}$$

Because of assumption (iii) we know that $\liminf_{t\to\infty} \Psi_t^{-1} R_t^{-1} = C > 0$, hence $\lim_{t\to\infty} \tilde{p}_t = 0$ a.s.

4.3 Example. If $\phi: T \to \mathbb{R}^2$, $\phi(t) = [1, sint+1]$, then the conditions of the theorem are satisfied. The matrix C in assumption (iii) becomes $\frac{1}{5}\begin{bmatrix} 2 & 2 \\ 2 & 3 \end{bmatrix}$.

5. Remarks

5.1 Clearly, condition 4.2 (iii) is sufficient to identify all the components of p_0. But it seems that one cannot do without. The strict positive definiteness of C is lost in either of the following situations that are worked out for $p_0 \in \mathbb{R}^2$. Let $\phi = [\phi_1, \phi_2]$ and let $\lim_{t\to\infty} \frac{\phi_{1t}}{\phi_{2t}} = 0$. Let $p_0 = [p_{01}, p_{02}]^T$. Then one cannot expect to identify p_{01}. For suppose $dn_{it} = p_{oi}\phi_{it} dt + dm_{it}, i = 1,2$, and let $n_t = n_{1t} + n_{2t}$. Then eventually all the observations of n_t are almost entirely those of n_{2t}, which doesn't yield much information about p_{o1}. Indeed C now becomes $\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Similarly if $\lim_{t\to\infty} \frac{\phi_{1t}}{\phi_{2t}} = c \in (0,\infty)$, one can only expect to identify $cp_{01} + p_{02}$.

5.2 Condition 4.2 (iii) appears as a technical condition, necessary for the proof of theorem 4.2. It seems however to be related to

$$\lim_{t\to\infty} \frac{1}{p_0^T \Phi_t} \int_0^t \frac{\phi_s \phi_s^T}{p_0^T \phi_s} ds > 0 \ a.s., \tag{5.1}$$

where $\Phi_t = \int_0^t \phi_s ds$. Here (5.1) has an appealing interpretation. To see this, define a normalized version of (3.1) by

$$H_t(p) = \frac{1}{p_0^T \Phi_t} J_t(p). \tag{5.2}$$

Then minimization of $H_t(\cdot)$ is equivalent with minimization of $J_t(\cdot)$. One can easily check that for large t $H_t''(p)|_{p=p_0}$ can be approximated by (5.1). Hence (5.1) says that for $t \to \infty$ p_0 is indeed a minimum point of $H_t(\cdot)$.

References

1 P.Brémaud, *Point Processes and Queues,* Springer, New York, 1981.

2 Dellacherie, C. & P.A. Meyer, *Probabilités et Potentiel,* Chap V-VIII, Hermann, Paris, 1980.

3 Yu. Lin'kov, *Estimates of parameters of counting processes,* Problems of information Transmission 18 (1982), pp. 63-76.

4 Ogata, Y., *The asymptotic behaviour of maximum likelihood, estimators for stationary point processes,* Ann. Inst. Statist. Math. 30 (1978), Part A pp. 243-261.

5 Robbins, H. & D. Siegmund, *A convergence theorem for nonnegative almost supermartingales and some applications,* in: *Optimizing methods in Statistics,* (J.S. Rustagi Ed.), pp. 233-256, Academic Press, New York, 1971.

6 Schuppen, J.H. van *Convergence Results for continuous-time adaptive Stochastic Filtering Algorithms,* J. Math. Anal. Appl. 96 (1983), pp. 209-225.

7 Solo, V., *The convergence of AML,* IEEE Trans. Automatic. Control 24 (1979), 958-962.