The Annals of Statistics
1982, Vol. 10, No. 3, 1006–1008

CONVERGENCE OF SIMAR'S ALGORITHM FOR FINDING THE MAXIMUM LIKELIHOOD ESTIMATE OF A COMPOUND POISSON PROCESS

BY DANKMAR BÖHNING

Free University Berlin

Simar (1976) suggested an iteration procedure for finding the maximum likelihood estimate of a compound Poisson process, but he could not show convergence. Here the more general case of maximizing a concave functional on the set of all probability measures is treated. As a generalization of Simar's procedure, an algorithm is given for solving this problem, including assumptions to ensure convergence to an optimum. Finally, it is shown that Simar's functional fulfills these assumptions.

1. Introduction and formulation of the problem. Simar (1976) considers the maximum likelihood (ML) estimation of a compound Poisson process. This problem leads to the maximization of the concave functional ϕ_p defined by

$$\phi_p(\delta) = \sum_{i=0}^{N} \alpha_i \log\left\{\int_T e^{-t} t^i \delta(dt)\right\}, \tag{1.1}$$

where $\alpha_0, \cdots, \alpha_N$ are known, positive real numbers summing to 1, T is equal to $[0, \infty)$, and $\delta \in \Delta$, the set of all probability measures (p.m.'s) on T. More generally, we consider the maximization of a concave functional $\phi: \Delta \to \mathbb{R} \cup \{-\infty\}$, where T is now an arbitrary topological space, and Δ is the set of all p.m. on its associated Borel sigma-algebra.

Let $\Delta^+ = \{\delta \in \Delta \mid \phi(\delta) > -\infty\}$, and, for $K \in \mathbb{R}$, $\Delta^K = \{\delta \in \Delta \mid \phi(\delta) \geq K\}$. Also let δ_t be the unit measure at $t \in T$. Note that for $\phi = \phi_p$, $\Delta^+ = \Delta \backslash \{\delta_0\}$. Let $S(\delta)$ denote the support of $\delta \in \Delta$.

ASSUMPTIONS. (i) There exists $\hat{\delta} \in \Delta^+$ such that $\phi(\hat{\delta}) \geq \phi(\delta)$ for all $\delta \in \Delta$. (ii) For all $\delta \in \Delta^+$ there exists $\beta^+(\delta, t) \in (0, 1]$ such that

$$(1 - \beta)\delta + \beta\delta_t \in \Delta^+, \quad \beta \in [0, \beta^+].$$

REMARK. Assumption (ii) allows the definition of the directional derivative

$$\Phi_\phi(\delta, \delta_t) = \lim_{\beta \to 0+} \beta^{-1}\{\phi((1 - \beta)\delta + \beta\delta_t) - \phi(\delta)\} \tag{1.2}$$

for any $\delta \in \Delta^+$ and any vertex direction δ_t.

ASSUMPTION. (iii) For all $\gamma > 0$, $K \in \mathbb{R}$ there exists $\beta_0 = \beta_0(\gamma, K) \in (0, 1)$ such that

$$\Phi_\phi(\delta, \delta_t) \geq \gamma \quad \text{implies} \quad \phi((1 - \beta)\delta + \beta\delta_t) - \phi(\delta) \geq \beta\frac{\gamma}{2}$$

for all $\beta \in [0, \beta_0(\gamma, K)]$, all $\delta \in \Delta^K$, and all $t \in T$. From here on we sometimes write just Φ instead of Φ_ϕ. Note that

$$\sup_{t \in T} \phi(\delta, \delta_t) \geq \phi(\hat{\delta}) - \phi(\delta), \quad \delta \in \Delta^+. \tag{1.3}$$

2. A general algorithm and its convergence. The following algorithm and the associated convergence theorem is taken from Böhning (1981). For

$$\phi(\delta) = \log \det\left[\int_T tt'\delta(dt)\right\}, \; T \subset \mathbb{R}^k,$$

this algorithm is discussed by Silvey and Titterington (1973).

Received October 1981; revised January 1982.

AMS 1970 *subject classifications.* Primary 62G05; secondary 62M99.

Key words and phrases. Compound Poisson process, iterative procedures, maximum likelihood method

ALGORITHM 1. ($\phi = \phi_p$: Simar's algorithm)

1. Choose $\delta_1 \in \Delta^+$, set $n = 1$.
2. Find $t_n \in T$ such that $\Phi_\phi(\delta_n, \delta_{t_n}) = \sup_{t \in T} \Phi_\phi(\delta_n, \delta_t)$.
   If $\Phi_\phi(\delta_n, \delta_{t_n}) = 0$ stop.
3. Choose $\delta_{n+1} \in \Delta(n) = \{\delta \in \Delta^+ \mid S(\delta) \subset S(\delta_n) \cup \{t_n\}\}$ such that $\phi(\delta_{n+1}) = \sup_{\delta \in \Delta(n)} \phi(\delta)$
4. Set $n = n + 1$, and go to 2.

REMARK. For the sake of brevity we assume that Steps 2 and 3 are always well defined. For example, Step 2 is well defined if (a) T is compact and $t \to \Phi(\delta, \delta_t)$ is continuous, or (b) T is locally compact and $t \to \Phi(\delta, \delta_t)$ disappears at infinity. Recall that continuous $f: T \to \mathbb{R}$ is said to be *disappearing at infinity* if and only if for all $\varepsilon > 0$ there exists a compact set $T_\varepsilon \subset T$ s.t. $|f(t)| < \varepsilon$ for all $t \in T \backslash T_\varepsilon$. Since

$$\Phi_{\phi_p}(\delta, \delta_t) = \sum \{\alpha_i t^i e^{-t} / \int t^i e^{-t} \delta(dt)\} \dot{-} 1,$$

case (b) is true for Φ_{ϕ_p}.

CONVERGENCE THEOREM. If ϕ meets Assumptions (i), (ii), and (iii) then for any sequence (δ_n) created by Algorithm 1, we have that either (δ_n) is finite with its last element being maximal or,

$$\lim_{n \to \infty} \phi(\delta_n) = \phi(\hat{\delta}).$$

PROOF. $(\phi(\delta_n))$ converges since it is monotone increasing. Let ϕ^+ be its limit. Assume $\phi^+ < \phi(\hat{\delta})$. Then

$$\Phi(\delta_n, \delta_{t_n}) = \sup_{t \in T} \Phi(\delta_n, \delta_t) \geq \phi(\hat{\delta}) - \phi(\delta_n) \geq \phi(\hat{\delta}) - \phi^+ \geq \gamma > 0$$

for some suitable γ, using (1.3). Thus by assumption (iii)

$$\phi(\delta_{n+1}) - \phi(\delta_n) \geq \phi((1 - \beta_0)\delta_n + \beta_0 \delta_{t_n}) - \phi(\delta_n) \geq \frac{\gamma}{2} \beta_0 > 0$$

for all $n \in \mathbb{N}$, which contradicts the Cauchy property of $(\phi(\delta_n))$. □

COROLLARY. Let (δ_n) be any sequence constructed by Simar's algorithm. Then either (δ_n) is finite with its last element being the MLE, or

$$\lim_{n \to \infty} \phi_p(\delta_n) = \phi_p(\hat{\delta}).$$

PROOF. Since Simar has already shown the existence of a MLE, and since $\Delta^+ = \Delta \backslash \{\delta_0\}$, only Assumption (iii) has to be verified. For this purpose consider the set

$$\mathbf{M} = \{x \in \mathbb{R}^{N+1}: \int e^{-t} t^i \mu(dt) = x_{i+1},\ i = 0, \cdots, N, \quad \text{for some measure } \mu \text{ s.t. } \mu(T) \leq 1\}$$

which Simar has shown to be compact and convex. Additionally consider the two mappings

$$\mathbf{m}: \delta \in \Delta \to \mathbf{m}(\delta) = \left(\int e^{-t} \delta(dt), \cdots, \int e^{-t} t^N \delta(dt)\right)^T \in \mathbf{M}$$

and

$$\Psi: (x_0, \cdots, x_N)^T \in \mathbb{R}^{N+1} \to \Psi(x) = \textstyle\sum_{i=0}^N \alpha_i \log(x_i) \in \mathbb{R} \cup \{-\infty\}.$$

Note that Ψ is continuously differentiable on $M \backslash \{0\}$. From the mean value theorem we get

$$\begin{aligned}\phi_p((1 - \beta)\delta + \beta \delta_t) - \phi_p(\delta) &= \Psi((1 - \beta)\mathbf{m}(\delta) + \beta \mathbf{m}(\delta_t)) - \Psi(\mathbf{m}(\delta)) \\ &= \beta \nabla \Psi((1 - r\beta)\mathbf{m}(\delta) + r\beta \mathbf{m}(\delta_t))^T \mathbf{m}(\delta_t) - 1\end{aligned}$$

for some r, $0 \le r \le 1$. Thus

$$\phi_p((1-\beta)\delta + \beta\delta_t) - \phi_p(\delta) - \beta\Phi_{\phi_p}(\delta, \delta_t)$$
$$= \beta\{\nabla\Psi((1-r\beta)\mathbf{m}(\delta) + r\beta\mathbf{m}(\delta_t)) - \nabla\Psi(\mathbf{m}(\delta))\}^T\mathbf{m}(\delta_t).$$

Now, the problem is that if $\mathbf{m}(\delta) \in \mathbf{M}^K = \{x \in \mathbf{M} \mid \Psi(x) \ge K\}$, then $(1-r\beta)\mathbf{m}(\delta) + r\beta\mathbf{m}(\delta_t)$ need not be in $\mathbf{M}^K$. For this reason, we introduce the set $\mathbf{M}^{K,\varepsilon} = \{x \in \mathbf{M} \mid \Psi(x) \ge K - \varepsilon\}$ for $\varepsilon > 0$. This again is a compact set, on which $\nabla\Psi$ is uniformly continuous. We can thus find $\beta_0 = \beta_0(K, \gamma)$ such that for all $0 \le \beta \le \beta_0$

$$(1-r\beta)\mathbf{m}(\delta) + r\beta\mathbf{m}(\delta_t) \in \mathbf{M}^{K,\varepsilon}$$

and

$$\|\nabla\Psi((1-r\beta)\mathbf{m}(\delta) + r\beta\mathbf{m}(\delta_t)) - \nabla\Psi(\mathbf{m}(\delta))\| \le \frac{\gamma}{2S}$$

where $S = \sup\{\|\mu\| : \mu \in \mathbf{M}\}$. Altogether we get

$$|\phi_p((1-\beta)\delta + \beta\delta_t) - \phi_p(\delta) - \beta\Phi_{\phi_p}(\delta, \delta_t)| \le \frac{\beta\gamma}{2S}S = \beta\frac{\gamma}{2} \tag{2.1}$$

for all $\beta \in [0, \beta_0]$ and all $\delta \in \Delta^K$, $t \in T$. Assume that

$$\phi_p((1-\beta)\delta + \beta\delta_t) - \phi_p(\delta) \le \beta\frac{\gamma}{2}.$$

Then $-\Phi(\delta, \delta_t) \le -\gamma$ implies

$$\phi_p((1-\beta)\delta + \beta\delta_t) - \phi_p(\delta) - \beta\Phi_{\phi_p}(\delta, \delta_t) < \beta\frac{\gamma}{2} - \beta\gamma = -\frac{1}{2}\beta\gamma,$$

which contradicts (2.1). Thus $\Phi(\delta, \delta_t) \ge \gamma > 0$ implies $\phi_p((1-\beta)\delta + \beta\delta_t) - \phi_p(\delta) \ge \beta\frac{\gamma}{2}$ for all $\beta \in [0, \beta_0]$ and all $\delta \in \Delta^K$, $t \in T$, if we set $\beta_0(\gamma, K) = \beta_0$. □

REMARK. The argument used in the proof of the corollary is more general in nature. It applies for any convex linear mapping $\mathbf{m}: \Delta \to \mathbb{R}^N$, e.g., $\mathbf{m}\{(1-\beta)\delta + \beta\bar{\delta}\} = (1-\beta)\mathbf{m}(\delta) + \beta\mathbf{m}(\bar{\delta})$, $\beta \in [0, 1]$, $\delta, \bar{\delta} \in \Delta$ and any mapping $\psi: \mathbb{R}^N \to \mathbb{R} \cup \{-\infty\}$, which has the property that it is continuously differentiable on $\mathbf{M}^K$, and, $\phi = \psi \circ \mathbf{m}$. In optimal experimental design a multitude of optimality criteria fulfill the above assumption (Wu, 1978). For this reason the proof of the corollary is not simplified by using a second degree Taylor expansion.

Acknowledgment. I would like to thank Ben Torsney, University of Glasgow, for drawing my attention to Simar's article.

REFERENCES

BÖHNING, D. (1981). *Numerische Methoden in der optimalen Versuchsplanung.* Unpublished dissertation at the Department of Mathematics, Free University Berlin.

SILVEY, S. D. and TITTERINGTON, D. M. (1973). A geometric approach to optimal design theory. *Biometrika* **60** 21–32.

SIMAR, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist.* **4** 1200–1209.

WU, C. F. (1978). Some iterative procedures for generating nonsingular optimal designs. *Commun. Statist. A* **7** 1399–1412.

FREE UNIVERSITY BERLIN
KELCHSTRASSE
1000 BERLIN [WEST] 41
WEST GERMANY