Likelihood Inference for Mixtures: Geometrical and Other Constructions of Monotone Step-Length Algorithms

Dankmar Bohning

Biometrika, Volume 76, Issue 2 (Jun., 1989), 375-383.

Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28198906%2976%3A2%3C375%3ALIFMGA%3E2.0.CO%3B2-B

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Biometrika is published by Biometrika Trust. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Biometrika
©1989 Biometrika Trust

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR

http://www.jstor.org/
Thu Mar 14 05:15:24 2002

Biometrika (1989), **76**, 2, *pp.* 375–83
Printed in Great Britain

Likelihood inference for mixtures: Geometrical and other constructions of monotone step-length algorithms

By DANKMAR BÖHNING
Department of Epidemiology, Free University Berlin, 1000 Berlin 45, Federal Republic of Germany

SUMMARY

This paper considers algorithms for the maximum likelihood estimator of the mixing distribution of a family of parametric densities. Specific interest is devoted to the monotonicity of the step-length of the solution. In § 2 we introduce the idea of estimating the area above the second derivative curve, and in Theorem 1 we relate area-overestimation to the monotonicity of the associated algorithm. In § 3 we apply Theorem 1 to the monotonicity analysis of well-known algorithms as well as to the construction of a new class of monotone step-length choices which is particularly simple in the mixture setting. Numerical comparisons and refinements in keeping the monotonic step-length from becoming too conservative are given. Theorem 3 characterizes conditions under which the Newton–Raphson step is monotonic and, if these conditions fail to hold, states that the regula falsi step is monotonic.

Some key words: Area estimation; Curvature of mixture likelihood; Monotonicity; Vertex direction method; Vertex exchange method.

1. INTRODUCTION

We consider a parametric family of densities $f(x, \theta)$ with respect to a sample space $X \subseteq \mathbb{R}_p$ and a parameter space $\Theta \subseteq \mathbb{R}_r$. Assuming population heterogeneity we observe a random variable with the mixture marginal density

$$f(x, P) = \int_\Theta f(x, \theta) P(d\theta).$$

Here P is an unknown probability measure corresponding to the distribution of θ in the population. This is the usual setting of nonparametric mixture modelling (Lindsay, 1983). A general introduction is given by Titterington, Smith & Makov (1985). Mallet (1986) extends the frame to include random nonlinear regression.

The problem is to estimate P via maximum likelihood. Given observations $x_1, \ldots, x_N$ find the nonparametric maximum likelihood estimator $\hat{P}$ which maximizes the likelihood

$$L(P) = \Pi_i \int_\Theta f(x_i, \theta) P(d\theta) = \Pi_i L_i(P), \tag{1}$$

with $L_i(P) = \int f(x_i, \theta) P(d\theta)$, where the integral is over Θ, denoting the ith contribution to the likelihood. The products in (1) are from 1 to N. Observe that $l(P) = \log L(P)$ is a concave functional on the set $\mathscr{P}$ of all probability measures on X. This formulation points out the similarity to the D-optimal design problem, where D stands for determinant (Fedorov, 1972; Silvey, 1980); many well-known theorems can be easily rederived for the mixture setting. Let P_θ be the one-point measure which puts all its mass 1 at θ and define

$$Q(s) = \log L\{(1-s)P + sP_\theta\}, \quad s \in [0, 1].$$

Then the first derivative with respect to s is

$$Q'(0) = d(\theta, P) - N,$$

with

$$d(\theta, P) = \sum_{i=1}^{N} \frac{f(x_i, \theta)}{L_i(P)}.$$

The Kiefer–Wolfowitz equivalence theorem for mixtures (Lindsay, 1983) states that $\hat{P}$ is a nonparametric maximum likelihood estimator if and only if $\max_\theta d(\theta, \hat{P}) = N$.

This theorem suggests the following algorithmic approach to find $\hat{P}$. Consider convex combinations $(1-s)P + sP_\theta$ with θ maximizing $d(., P)$ for current value P. This method is called the vertex direction method since we are always moving in the direction of a vertex P_θ of the probability simplex $\mathcal{P}$. An alternative to the vertex direction method is the vertex exchange method, defined for $s \in [0, 1]$ by the iteration $P + sP(\theta^*)(P_\theta - P_{\theta^*})$, where θ maximizes $d(., P)$, and θ^* minimizes $d(., P)$ in the support of P. Clearly, if $s = 1$, $P + P(\theta^*)(P_\theta - P_{\theta^*})$ exchanges the 'bad' support point θ^* with the new point θ. There are other alternatives to the vertex direction method such as the projected gradient (Wu, 1978) or the projected Newton method (Atwood, 1976, 1980). These procedures are compared in detail empirically with examples from optimal design and mixture likelihood by Böhning (1985) with results in favour of the projected Newton and the vertex exchange method; in terms of convergence rate, the Newton was slightly superior to the vertex exchange method, whereas the vertex exchange method is superior in numerical complexity. Silvey, Titterington & Torsney (1978) suggest an algorithm for optimal designs on a finite design space, of which the analogue in the mixture setting with finite, fixed support is an EM algorithm (Dempster, Laird & Rubin, 1977). In § 2 we offer a class of monotonic step-length choices for any feasible direction h; here feasible means $P + sh \in \mathcal{P}$ for all $s \in [0, 1]$. We observe an important geometric property of the mixture likelihood function. Define $Q(s) = \log L(P + sh)$, $s \in [0, 1]$. We have

$$Q''(s) = -\sum_{i=1}^{N} \left\{ \frac{\int f(x_i, \theta) h(d\theta)}{\int f(x_i, \theta) P + sh(d\theta)} \right\}^2 \leq 0,$$

$$Q^{(iv)}(s) = -6 \sum_{i=1}^{N} \left\{ \frac{\int f(x_i, \theta) h(d\theta)}{\int f(x_i, \theta) P + sh(d\theta)} \right\}^4 \leq 0,$$

showing that Q is concave and has a concave second derivative.

2. AREA ESTIMATION AND MONOTONICITY

In this section we offer another interpretation of the Newton–Raphson algorithm as an area estimation algorithm. Consider

$$A(s) = \int_0^s Q''(t)\, dt = Q'(s) - Q'(0),$$

the area above the curve Q'' from 0 to s. Specifically, for the line maximizer $\hat{s}$ we have

$$A(\hat{s}) = Q'(\hat{s}) - Q'(0) = -Q'(0). \quad (2)$$

Equation (2) allows an interesting perspective: although we do not know $\hat{s}$, we do know $A(\hat{s})$, the area above Q'' from 0 to $\hat{s}$: it is $-Q'(0)$.

Algorithms differ in the way they estimate $A(s)$. Let $a(s)$ denote the estimate of $A(s)$. Then many well-known algorithms can be reproduced by equating $a(s)$ with the 'true' $A(\hat{s}) = -Q'(0)$ and solving it for s. The estimating equation is

$$a(s) = A(\hat{s}) = -Q'(0). \tag{3}$$

As an example consider the Newton–Raphson algorithm. It uses the rectangular estimate

$$a_{\mathrm{NR}}(s) = sQ''(0)$$

namely the box with height $Q''(0)$ and length s. Equation (3) then gives the usual Newton–Raphson iterate. We are able to give a simple condition for the step-length to be monotonic. It connects the idea of area overestimation with the monotonicity of an algorithm.

THEOREM 1. *Let*

$$a(s) \leqslant A(s) \quad \textit{for all } s \in [0, 1]. \tag{4}$$

Then s defined by the estimating equation (3) *is monotonic.*

Proof. We have $a(s) \leqslant A(s)$ for all s. In particular this is true for s satisfying (3). Denote this by s^*. Then $A(s^*) \geqslant a(s^*) = A(\hat{s})$. Since A is strictly decreasing we have $0 \leqslant s^* \leqslant \hat{s}$, implying $Q(0) \leqslant Q(s^*)$. □

Note that $A(s) < 0$. Geometrically, the condition (4) means that the area above the curve is always smaller than the estimated area. In the case of a_{NR} this condition is met if $\inf_t Q''(t) = Q''(0)$.

COROLLARY. *Let area satisfy the continuity condition $a(s) \to 0$ if $s \to 0$. Also, let $\{P_n\}$ be any sequence of the type $P_{n+1} = P_n + s_n h_n$, where h_n is either the vertex direction or the vertex exchange method, and s_n is the step-length choice according to* (3). *If s_n meets the overestimating condition* (4) *then $l(P_n) \to l(\hat{P})$ monotonically.*

Proof. Monotonicity is clear. Suppose $\lim l(P_n) < l(\hat{P})$ as $n \to \infty$. Then also $\sup_\theta d(\theta, P_n)$ is bounded away from N. Consider the second-order Taylor expansion

$$l(P_n + sh_n) - l(P_n) = sQ_n'(0) + \tfrac{1}{2}s^2 Q_n''(s^*),$$

with $s^* \in [0, s]$ and $Q_n(s) = l(P_n + sh_n)$. Since $Q_n'(0) = -a(s_n)$ is bounded away from its lower bound 0, s_n is also bounded below by some positive constant. This finally forces $l(P_n) \to \infty$, which is impossible since l is bounded above. □

3. SPECIFIC AREA ESTIMATORS

As another example let us consider the secant method step-length

$$s_{\mathrm{sec}} = -Q'(0)/\{Q'(1) - Q'(0)\}.$$

The secant method assumes that $A(s)$ can be estimated as the sth point of $A(1) = Q'(1) - Q'(0)$:

$$a_{\mathrm{sec}}(s) = sA(1).$$

Equation (3) leads again to s_{sec}. For the secant step-length to be monotonic, condition (4) takes the form

$$s\{Q'(1)-Q'(0)\} \leq Q'(s)-Q'(0)$$

or

$$(1-s)Q'(0)+sQ'(1) \leq Q'(s).$$

So, condition (4) applied to a_{sec} is equivalent to Q' being concave, a condition which is easy to check. However, it does not hold in general for our case, but see § 5.

Clearly, the estimating technique used by the Newton–Raphson algorithm is not necessarily monotonic as can be demonstrated in a simple picture. Next, we offer two modifications of a_{NR} leading to monotonic behaviour in general. Condition (4) shows the way. In our case, $Q''(s)$ is concave and thus, it attains its minimum either at 0 or 1. So, we construct the box by choosing its height by $M = \min\{Q''(0), Q''(1)\}$: $a_{\text{box}}(s) = sM$. Equation (3) leads to $s_{\text{box}} = -Q'(0)/M$. Clearly, the technique is more general in nature. However, in other problems we might not gain a lot, since we are replacing one minimum problem by another.

Above we have used a very rough approximation of a concave function, namely a horizontal line going through the minimum of Q'' on the line segment $[0, 1]$. Why not approximate Q'' by the line which connects $Q''(0)$ and $Q''(1)$? The area of the resulting trapezoid with height s is given by

$$a_{\text{trp}}(s) = sQ''(0) + \tfrac{1}{2}s^2\{Q''(1)-Q''(0)\}. \tag{5}$$

Equation (3) leads to

$$sQ''(0) + \tfrac{1}{2}s^2\{Q''(1)-Q''(0)\} + Q'(0) = 0$$

or

$$s^2 + 2\tilde{A}s + \tilde{B} = 0, \tag{6}$$

with

$$\tilde{A} = 2Q''(0)/\{Q''(1)-Q''(0)\}, \quad \tilde{B} = 2Q'(0)/\{Q''(1)-Q''(0)\}$$

assuming $Q''(1) \neq Q''(0)$. There are two cases. If $\tilde{A} > 0$ then $Q''(1) < Q''(0)$ and $\tilde{B} < 0$. Thus $\tilde{A}^2 - \tilde{B} > \tilde{A}^2$ and the upper zero of (6) has to be taken:

$$s_{\text{trp}} = -\tilde{A} + (\tilde{A}^2 - \tilde{B})^{\frac{1}{2}} > 0. \tag{7a}$$

If $\tilde{A} < 0$ then $Q''(1) > Q''(0)$ and $\tilde{B} > 0$. Thus $\tilde{A}^2 - \tilde{B} < \tilde{A}^2$ and the lower solution

$$s_{\text{trp}} = -\tilde{A} - (\tilde{A}^2 - \tilde{B})^{\frac{1}{2}} > 0 \tag{7b}$$

of (6) has to be used. If $Q''(0) = Q''(1)$ then $s_{\text{trp}} = s_{\text{NR}}$.

Table 1 summarizes various step-length choices. The last step-length formula is suggested and proved to be monotonic by Mallet (1986) in connexion with the vertex direction

Table 1. *Various area estimators and corresponding step-length choices*

Method	$a(s)$ – estimator	Step-length	Cond. for monoton.
NR	$sQ''(0)$	$s_{\text{NR}} = -Q'(0)/Q''(0)$	$Q''(0) = \inf, Q''(t)$
sec	$sA(1)$	$s_{\text{sec}} = -Q'(0)/A(1)$	Q' concave
box	$s \min\{Q''(0), Q''(1)\}$	$s_{\text{box}} = -Q'(0)/\min\{Q''(0), Q''(1)\}$	none
trp	(5)	(7)	none
Mallet	$sN\{1-Q'(0)-N\}$	$s_M = \dfrac{Q'(0)}{N\{Q'(0)+N-1\}}$	h, vertex direction

method for mixtures. The formula is inspired by the close link to the D-optimal design problem. However, whereas in the D-optimal design problem the corresponding formula provides the line maximizer, we will see in § 4 that s_M can be far from optimality.

4. Empirical comparisons

Suppose θ can attain only 3 values, $\theta \in \{1, 2, 3\}$, and $x \in \{1, 2, 3, 4\}$,

$$[f(x, 1)] = (0{\cdot}60 \;\; 0{\cdot}30 \;\; 0{\cdot}05 \;\; 0{\cdot}05)^{\mathrm{T}},$$

$$[f(x, 2)] = (0{\cdot}05 \;\; 0{\cdot}15 \;\; 0{\cdot}30 \;\; 0{\cdot}50)^{\mathrm{T}},$$

$$[f(x, 3)] = (0{\cdot}01 \;\; 0{\cdot}08 \;\; 0{\cdot}21 \;\; 0{\cdot}70)^{\mathrm{T}}.$$

Further suppose that the number of times we have observed $x = 1, 2, 3, 4$ are respectively 15, 10, 20, 55 so that $N = 100$. Then $\hat{P}$ is $(0{\cdot}2102, 0{\cdot}0424, 0{\cdot}7473)^{\mathrm{T}}$ (Böhning, 1985).

Choose as search direction the vertex exchange method; for computational details see the Appendix. The three step-length choices under consideration are s_{box}, s_{sec} and s_{trp}. The results are given in Table 2. Recall that $\max_\theta d(\theta, P_n) = N$ means that P_n is a maximum likelihood estimate. We are concentrating on convergence rate. Because of the nice computational forms there are practically no differences in terms of computational complexity. Clearly, trp and sec are superior to box.

Table 2. $\max_\theta d(\theta, P_n)$ *for the vertex exchange method with three different step-length choices*

n	s_{box}	s_{sec}	s_{trp}	n	s_{box}	s_{sec}	s_{trp}
1	120·11	120·11	120·11	8	115·18	102·93	101·67
2	119·30	113·71	113·79	10	114·02	102·12	100·61
4	117·81	107·34	106·40	15	111·53	100·51	101·05
6	116·43	104·43	103·03	20	109·89	100·27	100·06

Let us close this section with some remarks concerning the vertex direction method. If we use the vertex direction method with step-length choice according to Mallet, we have at $n = 20\,000$ $P_n = (0{\cdot}2027, 0{\cdot}1097, 0{\cdot}6856)^{\mathrm{T}}$, which is still far away from $\hat{P}$. At that step, $d(\theta_n, P_n) = 100{\cdot}29$. The vertex direction method with step-length choice according to s_{trp} does better. We have 4 digits of accuracy after about 1000 iterations. Nevertheless, the overall performance of the vertex direction method is not satisfying. A referee raised the point that using the vertex direction method a monotonic step-length will always be feasible, e.g. less than 1, since $\hat{s}$ is less than 1, whereas this would not always be obvious for general h. In fact, when the vertex exchange method is used, $\hat{s}$ is sometimes larger than 1. Since Q is concave the optimal step-length is $s = 1$, defining an exchange step in this case. In addition, $\hat{s} \geq 1$ is easily detected by $Q'(1) \geq 0$. If, however, $\hat{s} < 1$ then any step-length satisfying (4) has to be feasible.

5. Halley's correction and the weighted curvature average

The modifications of the Newton–Raphson step-length, as discussed in § 3, often slow down the convergence rate. This is particularly the case for the estimator a_{box} as can be seen in the first row of Table 2. The question is: how can we overestimate $A(s)$ and be simultaneously as close as possible to $A(s)$ to achieve quick convergence? Richardson

(1988) points out that rectangular area estimators can be better than trapezoidic estimators. We now consider the construction of an optimal rectangular area estimator.

We consider the cubic approximation of $Q(s)-Q(0)$,

$$sQ'(0)+\tfrac{1}{2}Q''(0)s^2+\tfrac{1}{6}Q'''(0)s^3,$$

which is maximized at s fulfilling

$$Q'(0)+Q''(0)s+\tfrac{1}{2}Q'''(0)s^2=0. \tag{8}$$

If we replace the third derivative by the secant approximate $Q''(1)-Q''(0)$, we are back at equation (6). We note in passing that the trapezoidic estimation technique can be viewed as an approximate third-order method. Instead of looking at the two solutions of (8), we rewrite (8) as

$$s=-Q'(0)/[Q''(0)+\tfrac{1}{2}s\{Q''(1)-Q''(0)\}]. \tag{9}$$

We achieve the Halley correction (Gander, 1985) by replacing s on the right-hand side of (9) with the Newton–Raphson correction $-Q'(0)/Q''(0)$. More generally, one may consider

$$s_t=-Q'(0)/[Q''(0)+t\{Q''(1)-Q''(0)\}],$$

showing that the Halley correction is a specific weighted average in the curvatures at $s=0$ and $s=1$; namely $t=-Q'(0)/\{2Q''(0)\}$. Other averages are possible such as $t=\tfrac{1}{2}$. The question arises as to whether there is an optimal t in the sense that

$$Q''(0)+t\{Q''(1)-Q''(0)\}=A(1). \tag{10}$$

Statistically speaking, the question is: when do two averages coincide, one with respect to a uniform distribution, the other with respect to a distribution which puts all its mass at the border? Geometrically, this question can be stated as: when does a box of length 1 and height between $Q''(0)$ and $Q''(1)$ exist such that its area is equal to the area above Q'' from 0 to 1.

THEOREM 2. *If*

$$M=\max\{Q''(0),\,Q''(1)\}\geq A(1), \tag{11}$$

then there exists a feasible solution t to (10) *given by*

$$t=\{Q'(1)-Q'(0)-Q''(0)\}/\{Q''(1)-Q''(0)\}.$$

Proof. First suppose that $M=Q''(0)$. From (11) follows

$$Q''(0)\geq Q'(1)-Q'(0)$$

or $Q'(1)-Q'(0)-Q''(0)\leq 0$. In addition $M=Q''(0)>Q''(1)$ implying that $t\geq 0$. Of course the minimum $M^*=Q''(1)\leq Q'(1)-Q'(0)$ or

$$Q''(1)-Q''(0)\leq Q'(1)-Q'(0)-Q''(0),$$

or $t\leq 1$.

A second case is that $M=Q''(1)$. From (11) it follows that $Q''(1)\geq Q'(1)-Q'(0)$ or

$$Q''(1)-Q''(0)\geq Q'(1)-Q'(0)-Q''(0),$$

that is $t \leq 1$. Also $M^* = Q''(0) \leq Q'(1) - Q'(0)$ implying

$$Q'(1) - Q'(0) - Q''(0) \geq 0$$

or, $t \geq 0$. □

The condition (11) is somehow contrary to our construction principles used so far. For the two boxes with length 1 and height $Q''(0)$ and $Q''(1)$, respectively, the condition states that the smaller one of these two boxes has to have an area not larger than the area above Q'' from 0 to 1. A typical situation, in which (11) fails to hold, occurs if $\max_t Q''(t)$ is attained in the interior of $[0, 1]$ and is much larger than $\max\{Q''(0), Q''(1)\}$. In this case, even the celebrated trapezoidic approximation would give quite a conservative step-length.

Note that the optimal curvature average brings us back to an old method, the secant step-length:

$$Q''(0) + t\{Q''(1) - Q''(0)\} = Q'(1) - Q'(0),$$

with t as in Theorem 2.

We conclude the paper with a central theorem. It not only clarifies under which circumstances the Newton–Raphson step is monotonic, it also provides as an alternative the secant method with guaranteed monotonicity if these circumstances fail to hold.

THEOREM 3. (i) *If* $Q''(0) \leq Q''(1)$ *then* s_{NR} *is monotonic.*
Let now $Q''(0) > Q''(1)$.
(ii) *If* $Q''(0) < A(1)$ *then* s_{NR} *is monotonic.*
(iii) *If* $Q''(0) \geq A(1)$ *then the optimal curvature average exists and is given by*

$$Q''(0) + t\{Q''(1) - Q''(0)\} = Q'(1) - Q'(0).$$

In addition, $s_t = s_{sec}$ *is monotonic.*

Proof. The theorem is proved if we have verified (3) of Theorem 1 for each of the three cases. For (i) and (ii) this is obvious. To prove (iii) we have to verify condition (3), that is

$$s[Q''(0) + t\{Q''(1) - Q''(0)\}] = s\{Q'(1) - Q'(0)\} \leq A(s)$$

for all s. Suppose there exists s^* such that $s^*\{Q'(1) - Q'(0)\} > A(s^*)$. It follows that this inequality has to be true even for all $s \in [s^*, 1]$ leading to a contradiction for $s = 1$. □

Our construction principles so far are using the knowledge of Q' and Q'' at $s = 0$ and $s = 1$. Yet another way would be adaptive. Suppose we use in a first step s_{NR}. If $Q(s_{NR}) \geq Q(0)$ then set $s_{new} = s_{NR}$. If, however, $Q(s_{NR}) < Q(0)$ we use one of our monotonic step-length choices discussed above, with $Q''(1)$ replaced by $Q''(s_{NR})$. For example, use the box-estimator. Then, we set

$$s_{new} = s_{box} = -Q'(0)/\min\{Q''(0), Q''(s_{NR})\}$$

leading again to a monotonic step-length. The computational expense is a little greater here.

ACKNOWLEDGEMENTS

Some ideas in this paper were initiated while the author was visiting associate professor at the Department of Statistics, Pennsylvania State University. My thanks go especially to Dr Bruce Lindsay for pointing out the concave curvature property of the mixture likelihood. Some general applications of curvature techniques are given by Böhning & Lindsay (1988). In addition, I thank the referee for many useful comments. This research was supported by the German Research Foundation.

APPENDIX

Computational forms

Let

$$\begin{aligned} Q(s) &= \textstyle\sum_i \log \int f(x_i, .)\, d(P+sh) \\ &= \textstyle\sum_i \log \left\{ \int f(x_i, .)\, dP + s \int f(x_i, .)\, dh \right\} \\ &= \textstyle\sum_i \log (L_i + sH_i), \end{aligned}$$

with $H_i = \int f(x_i, .)\, dh$. For the vertex direction method, $h = P_\theta - P$, we have

$$H_i = f(x_i, \theta) - \int f(x_i, .)\, dP$$

and for the vertex exchange method, $h = P(\theta^*)(P_\theta - P_{\theta^*})$,

$$H_i = P(\theta^*)\{f(x_i, \theta) - f(x_i, \theta^*)\}.$$

Ignoring indices we have in general

$$Q'(s) = \textstyle\sum H/(L+sH), \quad Q''(s) = -\sum \{H/(L+sH)\}^2,$$

and specifically

$$Q'(0) = \textstyle\sum H/L, \quad Q''(0) = -\sum (H/L)^2,$$

$$Q'(1) = \textstyle\sum H/(L+H), \quad Q''(1) = -\sum \{H/(L+H)\}^2.$$

Write $\tilde{a} = H/L$ and arrive at the computationally simple relations

$$Q'(0) = \textstyle\sum \tilde{a}, \quad Q'(1) = \sum \tilde{a}/(\tilde{a}+1),$$

$$Q''(0) = -\textstyle\sum \tilde{a}^2, \quad Q''(1) = -\sum \{\tilde{a}/(\tilde{a}+1)\}^2.$$

REFERENCES

ATWOOD, C. L. (1976). Convergent design sequences for sufficiently regular optimality criteria. *Ann. Statist.* **4**, 1124-38.
ATWOOD, C. L. (1980). Convergent design sequences for sufficiently regular optimality criteria, II. *Ann. Statist.* **8**, 894-913.
BÖHNING, D. (1985). Numerical estimation of a probability measure. *J. Statist. Plan. Inf.* **11**, 57-69.
BÖHNING, D. & LINDSAY, B. (1988). Monotonicity of quadratic approximation algorithms. *Ann. Inst. Statist. Math.* **40**, 641-63.
DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *J. R. Statist. Soc.* B **39**, 1-38.
FEDOROV, V. V. (1972). *Theory of Optimal Experiments.* New York: Academic Press.
GANDER, W. (1985). On Halley's iteration method. *Am. Math. Monthly* **92**, 131-4.
LINDSAY, B. G. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86-94.
MALLET, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika* **73**, 645-56.
RICHARDSON, G. D. (1988). Reconsidering area approximations. *Am. Math. Monthly* **95**, 754-5.
SILVEY, S. D. (1980). *Optimal Design.* London: Chapman and Hall.

SILVEY, S. D., TITTERINGTON, D. M. & TORSNEY, B. (1978). An algorithm for optimal designs on a finite design space. *Comm. Statist.* A **14**, 1379–89.

TITTERINGTON, D. M., SMITH, A. F. M. & MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions.* New York: Wiley.

WU, C. F. (1978). Some algorithmic aspects of the theory of optimal designs. *Ann. Statist.* **6**, 1286–301.

[*Received March* 1987. *Revised November* 1988]