

A FAST ALGORITHM FOR THE NONPARAMETRIC MAXIMUM LIKELIHOOD ESTIMATE IN THE COX-GENE MODEL

I-Shou Chang¹, Chi-Chung Wen², Yuh-Jenn Wu³ and Che-Chi Yang⁴

¹*National Health Research Institutes,*

²*Tamkang University,* ³ *Chung Yuan Christian University and*

⁴ *Lunghwa University of Science and Technology*

Abstract: The Cox model with the gene effect for age at onset was introduced and studied by Li, Thompson and Wijsman (1998) and Li and Thompson (1997). This paper concerns the numerical performance of the nonparametric maximum likelihood estimate of the environmental effects and the genetic effect in this model. Based on the self-consistency equations derived from the score functions, we propose a fast iterative algorithm for the computations of the nonparametric maximum likelihood estimate and its asymptotic variance. Simulation studies conducted using these algorithms indicate that the profile likelihood-based normal approximations for the estimates are valid with reasonable sample sizes, and the bootstrap methods work well also for smaller sample sizes, and are computationally feasible.

Key words and phrases: Age at onset, bootstrap, frailty, gene effect, profile likelihood, self-consistency equations.

1. Introduction

Analysis of familial diseases with variable age at onset is common in human genetics. Several papers have considered regression models where age at onset depends on observed individual environment effects and unobserved gene effect. See, for example, Vaupel, Manton and Stallard (1979), Meyer and Eaves (1988), Elston and George (1989), Abel and Bonney (1990), Mack, Langholz and Thomas (1990), Gauderman and Thomas (1994), Yashin and Iachine (1995), Li and Thompson (1997), etc. In particular, Li et al. (1998) and Siegmund and McKnight (1998) proposed maximum likelihood estimation of the gene effect for age at onset with semiparametric models for right censored data. These models study genetic effect at an individual level and extend classical semiparametric hazard models by the introduction of a binary frailty allowing for shared frailty within families. In order to implement this likelihood approach, Li et al. (1998) and Siegmund and McKnight (1998) studied Monte Carlo methods based on the

EM algorithm, then did some simulation studies. The model in Li et al. (1998) is called the Cox-gene model and is the main concern of the present paper.

Chang, Hsiung, Wang and Wen (2005) gave a theoretical justification of the maximum likelihood estimates in Li et al. (1998) and some of those in Siegmund and McKnight (1998). More precisely, it indicated conditions under which the parameters are identifiable and the estimate is consistent, asymptotically normal and efficient. Furthermore, it suggested that the characterization of the score functions developed for the asymptotic theory provides an alternative approach to the calculation of the nonparametric maximum likelihood estimates, henceforth NPMLE, and that asymptotic variances can be estimated consistently using the theory of observed profile information. In this paper, we present numerical methods suggested by this theory and use them to illustrate the above likelihood methods through simulation studies.

The basis of our algorithms for the computations of the NPMLE and the asymptotic variance is a class of self-consistency equations derived from the score functions. That these are fast algorithms, as compared with the EM-algorithm implemented in Li et al. (1998) and Siegmund and McKnight (1998), is seen clearly in the simulation studies. The simulation studies in Li et al. (1998) involved one replicate only, and those in Siegmund and McKnight (1998) did not calculate variances.

To make the discussion more focused, we only consider families consisting of siblings. For this type of family, Chang et al. (2005) showed that the theory is valid when there are three or more siblings in each family and, in case there are observable environmental effects, it is valid even if there is only one member in each family. Our simulation studies indicate clearly that the normal approximation of the NPMLE holds with reasonable sample sizes.

To indicate that the likelihood methods do not require a fixed family size, we show by simulations that asymptotic normality seems also to hold for families with varying sizes. We note that Chang et al. (2005) assumes every family has the same number of individuals.

To take advantage of the speed of our algorithms, we also conducted bootstrap studies, and found that the bootstrap methods may provide better confidence interval when sample size is smaller and the normal approximation is not adequate.

This paper is organized as follows. Section 2 recapitulates the main results in Chang et al. (2005) and presents the self-consistency equations that lead to the algorithms presented in Section 3. Section 4 presents the simulation studies, that examine the normal approximations for the NPMLE, the performance of the bootstrap methods for the Cox-gene model, and other properties of the

algorithms. Section 5 has some discussion. We give results concerning the convergence of the algorithms in the Appendix.

2. The Model and the Score

The notations and assumptions in this paper are consistent with those in Chang et al. (2005). In order to have a concise presentation, some of the regularity conditions are not stated here.

Let T_{ik} , C_{ik} , and Z_{ik} denote, respectively, the age at onset, the censoring time, and the covariate of the i th individual in the k th family. Here $i = 1, \dots, m_k$, $k = 1, \dots, K$. Let g_{ik} be the genotype of the i th individual in the k th family at a certain locus. Assume that there are two alleles at this locus and denote them by A and a . Thus the genotype takes one of the three values, aa , aA or AA . Let $S_{ik} = I_{[g_{ik}=aA \text{ or } AA]}$ denote the susceptibility type. Denote by q_0 the population allele frequency of A .

Let $T_k = (T_{1k}, \dots, T_{m_k k})$, $C_k = (C_{1k}, \dots, C_{m_k k})$, $Z_k = (Z_{1k}, \dots, Z_{m_k k})$ and $S_k = (S_{1k}, \dots, S_{m_k k})$. Assume that, given Z_k and S_k , $T_{1k}, \dots, T_{m_k k}$, $C_{1k}, \dots, C_{m_k k}$ are conditionally independent, and for an individual having covariate z and susceptibility type s , the hazard of disease onset at age t is

$$\lambda_0(t) \exp(\beta_0^T z + \mu_0 s). \tag{1}$$

Here, $\lambda_0(\cdot)$ is a non-negative deterministic baseline function, $\beta_0 \in \mathfrak{R}^D$, $Z_{ik} \in \mathfrak{R}^D$, and $\mu_0 > 0$. Assume further that (C_k, Z_k) and S_k are independent, and that the distribution of (C_k, Z_k) does not involve $\lambda_0(\cdot)$, β_0 , μ_0 , and q_0 . Assume Z_{ik} is non-degenerate for every i and k . Let $X_{ik} = T_{ik} \wedge C_{ik}$, the minimum of T_{ik} and C_{ik} , and $\delta_{ik} = I_{[T_{ik} \leq C_{ik}]}$.

Assume that the (T_k, C_k, Z_k, S_k) are independent, $k = 1, \dots, K$. The statistical problem is to estimate $(\Lambda_0, \beta_0, \mu_0, q_0)$ based on $\{X_k, \delta_k, Z_k | k = 1, \dots, K\}$, where $X_k = (X_{1k}, \dots, X_{m_k k})$, $\delta_k = (\delta_{1k}, \dots, \delta_{m_k k})$, and $\Lambda_0(t) = \int_0^t \lambda_0(u) du$.

The NPMLE $(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})$ studied in Li et al. (1998), Siegmund and McKnight (1998) and Chang et al. (2005) is the maximizer of the likelihood

$$L_K(\Lambda, \beta, \mu, q) \equiv \prod_{k=1}^K \sum_{s \in S} p(s, q) \left(\prod_{i=1}^{m_k} [\Delta \Lambda(X_{ik}) e^{\beta^T Z_{ik} + \mu S_{ik}}]^{\delta_{ik}} \exp[-\Lambda(X_{ik}) e^{\beta^T Z_{ik} + \mu S_{ik}}] \right). \tag{2}$$

Here $\Delta \Lambda(t) = \Lambda(t) - \Lambda(t-)$, Λ is a non-decreasing function with $\Lambda(0) = 0$, $p(s, q) \equiv p(s_1, \dots, s_{m_k}, q)$ is the probability that the susceptibility vector takes the value (s_1, \dots, s_{m_k}) when the dominant allele A has frequency q , and S is the set of all possible values of $S_k = (S_{1k}, \dots, S_{m_k k})$. We note that $p(s, q)$ depends

on the family structure and is calculated here under the assumptions of random mating and Mendelian segregation. To simplify the notation, we suppress the subscript k in $p(s, q)$. For the purpose of computation, the parameter space we consider is $\Theta_c = \{ (\Lambda, \beta, \mu, q) \mid \Lambda \in \mathcal{L}_c, \beta \in \mathcal{B}, \mu \in \mathcal{U}, q \in \mathcal{Q} \}$. Here, $\mathcal{L}_c = \{ \Lambda : [0, \tau] \rightarrow [0, \infty) \mid \Lambda(0) = 0, \Lambda(\tau) \leq c, \Lambda \text{ is non-decreasing and right continuous} \}$ for some $\tau > 0$ and $c > 0$; \mathcal{B} and \mathcal{U} are compact subsets of \mathfrak{R}^D and $(0, \infty)$, respectively; and \mathcal{Q} is a closed subinterval of $(0, 1)$. The true parameters β_0, μ_0 , and q_0 are assumed to be interior points of \mathcal{B}, \mathcal{U} , and \mathcal{Q} , respectively, and $\Lambda_0(\tau)$ is assumed to be less than c .

It follows from Theorem 4.1 and Theorem 5.1 in Chang et al. (2005) that $\sqrt{K} \left((\hat{\beta}_K - \beta_0)^T, \hat{\mu}_K - \mu_0, \hat{q}_K - q_0 \right)^T$ is asymptotically normal with mean 0 and variance Σ^{-1} , and that Σ can be estimated consistently as follows. Let

$$M_K(\beta, \mu, q) = \frac{1}{K} \sup_{\Lambda \in \mathcal{L}_c} \log L_K(\Lambda, \beta, \mu, q). \tag{3}$$

Then $\nu^T \Sigma \nu$ is approximately

$$-2 \frac{M_K((\hat{\beta}_K^T, \hat{\mu}_K, \hat{q}_K)^T) + \gamma_K \nu_K - M_K((\hat{\beta}_K^T, \hat{\mu}_K, \hat{q}_K)^T)}{\gamma_K^2}, \tag{4}$$

for every sequence ν_K , in \mathfrak{R}^{D+2} , converging in probability to ν , and for every sequence γ_K satisfying $(\sqrt{K} \gamma_K)^{-1} = O_p(1)$ and $\gamma_K = o_p(1)$.

We now introduce notations so that we can state several equations that are useful in proposing the algorithms. These equations are derived from the score functions.

Let

$$f_k(\Lambda, \beta, \mu, q, s) = p(s, q) \left(\prod_{i=1}^{m_k} e^{\mu s_i \delta_{ik}} \exp[-\Lambda(X_{ik}) e^{\beta^T Z_{ik} + \mu s_i}] \right),$$

$$W_K(\Lambda, \beta, \mu, q; u) = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{m_k} \frac{\sum_s f_k(\Lambda, \beta, \mu, q, s) e^{\beta^T Z_{ik} + \mu s_i}}{\sum_s f_k(\Lambda, \beta, \mu, q, s)} I_{(0, X_{ik}]}(u),$$

$$G_K(u) = \frac{1}{K} \sum_{k=1}^K \sum_{i=1}^{m_k} I_{[T_{ik}, \infty)}(u \wedge C_{ik}),$$

$$b_k(\Lambda, \beta, \mu, q, s) = f_k(\Lambda, \beta, \mu, q, s) / p(s, q).$$

Knowing that $\frac{\partial}{\partial q} p(s, q)$ is a polynomial in q , we delete the monomials in $\frac{\partial}{\partial q} p(s, q)$ having negative coefficients and denote the resulting polynomial by $p'_+(s, q)$. Let $p'_-(s, q) = p'_+(s, q) - \frac{\partial}{\partial q} p(s, q)$. Let $e_1^T = (1, 0, 0, \dots, 0)$, $e_2^T = (0, 1, 0, \dots, 0)$, \dots , $e_D^T = (0, 0, 0, \dots, 1)$. The NPMLE $(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})$ satisfies the following.

Proposition 1.

(i)

$$\hat{\Lambda}(t) = \int_0^t \frac{1}{W_K(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}; u)} dG_K(u); \tag{5}$$

(ii)

$$\hat{\beta}^T e_j = \log \left(\frac{\sum_{k=1}^K \sum_{i=1}^{m_k} \delta_{ik} e_j^T Z_{ik}}{\sum_{k=1}^K \sum_{i=1}^{m_k} \frac{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s) (e_j^T Z_{ik} \hat{\Lambda}(X_{ik}) \exp(\hat{\beta}^T (Z_{ik} - e_j) + \hat{\mu} s_i))}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}} \right) \tag{6}$$

for $j = 1, \dots, D$;

(iii)

$$\hat{\mu} = \log \left(\frac{\sum_{k=1}^K \sum_{i=1}^{m_k} \frac{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s) (\delta_{ik} s_i)}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}}{\sum_{k=1}^K \sum_{i=1}^{m_k} \frac{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s) (s_i \hat{\Lambda}(X_{ik}) \exp(\hat{\beta} Z_{ik} + \hat{\mu}(s_i - 1)))}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}} \right); \tag{7}$$

(iv)

$$\hat{q} = \frac{\sum_{k=1}^K \frac{\sum_s \hat{q} p'_+(s, \hat{q}) b_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}}{\sum_{k=1}^K \frac{\sum_s p'_-(s, \hat{q}) b_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}}. \tag{8}$$

Remarks. We note that part (i) in Proposition 1 is just the first part of Lemma 2.1 in Chang et al. (2005), and (ii), (iii), and (iv) are derived, respectively, from the score functions (2.4), (2.5), and (2.6) there. Although the proof of Proposition 1 is straightforward and hence omitted, we would like to point out that it is the basis of the following algorithms, the main ideas of which are explained in the Appendix. In fact, part (ii), (iii), and (iv) of Proposition 1 are motivated by Proposition 3 in the Appendix. Specifically, suppose the score function is represented as the derivative η'_1 of a function η_1 , with $\eta'_1(\alpha_0) = 0$ and $\alpha_0 > 0$. If $\eta'_1 = \eta_2 - \eta_3$ for two positive functions η_2 and η_3 , α_1 is close to α_0 , and $\alpha_{J+1} = \alpha_J[\eta_2(\alpha_J)/\eta_3(\alpha_J)]$, then α_J converges to α_0 under certain regularity conditions. With suitably chosen η_2 and η_3 , we get (ii), (iii), and (iv). Consider (iv), for example. We set

$$\eta_2 = \sum_{k=1}^K \frac{\sum_s \hat{q} p'_+(s, \hat{q}) b_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}, \quad \eta_3 = \sum_{k=1}^K \frac{\sum_s p'_-(s, \hat{q}) b_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}{\sum_s f_k(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}, s)}.$$

Then $\eta_2 - \eta_3$ is the score function and step (5) in Algorithm 3.1 is an implementation of $\alpha_{J+1} = \alpha_J[\eta_2(\alpha_J)/\eta_3(\alpha_J)]$. We note that although the Newton-Ralphson

method is also an iterative procedure and applicable here, our (unreported) simulation studies indicate that it is very slow.

3. Algorithm

Let $\mathbf{L}(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})(t)$, $\mathbf{B}_j(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})$, $\mathbf{M}(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})$, and $\mathbf{Q}(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q})$ denote, respectively, the right hand sides of (5), (6), (7) and (8). Let $\mathbf{B}(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}) = (\mathbf{B}_1(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}), \dots, \mathbf{B}_D(\hat{\Lambda}, \hat{\beta}, \hat{\mu}, \hat{q}))^T$.

3.1. Algorithm for Estimating $\hat{\Lambda}$, $\hat{\beta}$, $\hat{\mu}$, \hat{q} , and $\hat{\Sigma}$

- (1) Choose starting values Λ_1 , β_1 , μ_1 , and q_1 .
- (2) Set $J = 1$.
- (3) $\beta_{J+1} = \mathbf{B}(\Lambda_J, \beta_J, \mu_J, q_J)$.
- (4) $\mu_{J+1} = \mathbf{M}(\Lambda_J, \beta_{J+1}, \mu_J, q_J)$.
- (5) $q_{J+1} = \mathbf{Q}(\Lambda_J, \beta_{J+1}, \mu_{J+1}, q_J)$.
- (6) $\Lambda_{J+1}(t) = \mathbf{L}(\Lambda_J, \beta_{J+1}, \mu_{J+1}, q_{J+1})(t)$.
- (7) $J = J + 1$.
- (8) Repeat (3) for a suitable number N of iterations once there is evidence of convergence.
- (9) The estimates of Λ , β , μ , q are given by $\hat{\Lambda} = \Lambda_N$, $\hat{\beta} = \beta_N$, $\hat{\mu} = \mu_N$, $\hat{q} = q_N$, respectively.
- (10) Set $\gamma_K = 1/\sqrt{K}$.
- (11) The (i, j) -entry of $\hat{\Sigma}$ is

$$- \left[M_K((\hat{\beta}^T, \hat{\mu}, \hat{q})^T + \gamma_K e_i + \gamma_K e_j) - M_K((\hat{\beta}^T, \hat{\mu}, \hat{q})^T + \gamma_K e_i) \right. \\ \left. - M_K((\hat{\beta}^T, \hat{\mu}, \hat{q})^T + \gamma_K e_j) + M_K((\hat{\beta}^T, \hat{\mu}, \hat{q})^T) \right] / \gamma_K^2.$$

3.2. Algorithm for Computing $M_K(\beta, \mu, q)$

- (1) Choose a starting value Λ_1 , set $L_1 = \log L_K(\Lambda_1, \beta, \mu, q)$.
- (2) Set $J = 1$.
- (3) $\Lambda_{J+1}(t) = \mathbf{L}(\Lambda_J, \beta, \mu, q)(t)$, set $L_{J+1} = \log L_K(\Lambda_{J+1}, \beta, \mu, q)$.
- (4) $J = J + 1$.
- (5) Repeat (3) for a suitable number N of iterations once there is evidence of convergence.
- (6) $M_K(\beta, \mu, q) = L_N/K$.

4. Simulation Studies

The purpose of this section is to assess the numerical performance of the NPMLE computed using the algorithms in Section 3. There are four parts in this section. The first part examines the normal approximation claimed in

Chang et al. (2005), the second part studies the bootstrap methods, the third part gives some idea of the dependence of the estimates on the starting values used in the algorithms, and the fourth part remarks on the computations needed for large pedigrees.

All the data are generated according to (1) with $\int_0^t \lambda_0(s) ds = \log(100/(100 - t))$, for $t \in [0, 100)$, and several different values of β_0, μ_0 , and q_0 ; the censoring variable is uniform(0,100); and $P(Z_{ik} = 0) = P(Z_{ik} = 1) = 1/2$. Except for the studies in Table 2, the number of iterations in using Algorithm 3.1 is set at 100, and that for Algorithm 3.2 is 10. In Table 2, we consider bootstrap coverage and the number of iterations in using Algorithm 3.1 is set at 50.

4.1. Normal Approximation

Our simulation studies indicate that the normal approximations for $\hat{\beta}, \hat{\mu}$, and \hat{q} are generally quite satisfactory. They also suggest that a relatively larger sample size is needed when q_0 is near 0 or μ_0 is small, and a much smaller sample size is needed in case there are more siblings in each family. Among other things, Table 1A seems to suggest that if the total number of individuals is fixed, studies having larger family size tend to provide better estimates in terms of mean-squared error and confidence interval coverage. In fact, the purpose of Table 1A is to demonstrate this phenomenon.

Each study in this subsection consists of 1,000 random samples of different sample sizes. For each sample, we first use the algorithms in Section 3 to compute $\hat{\beta}, \hat{\mu}, \hat{q}$, and $\hat{\Sigma}$, and then use the asymptotic normality to get a 0.95 confidence interval. The number of these 1,000 samples for which the true parameter falls in its 0.95 confidence interval is recorded; these numbers are then used to get the 95% confidence interval coverage, which is expressed in percentiles in the seventh, eighth and ninth columns of Table 1. We use CI_β, CI_μ , and CI_q to indicate the columns for β, μ , and q , respectively. The percentages of the samples for which the true parameter β_0, μ_0 , and q_0 falls below (above) its 0.95 confidence interval are also contained in these columns and denoted by $L_\beta (U_\beta), L_\mu (U_\mu)$, and $L_q (U_q)$ respectively. Each row in Table 1 represents the results for one simulation study scenario. We use Scen- a to denote the study presented in the a th row in Table 1. Here $a = 1, \dots, 15$.

Each family in our studies consists of siblings. The second and third columns are, respectively, the sample size K and the number m of siblings in each family. The next three columns contain, respectively, the true parameter values β_0, μ_0 and q_0 used in the data generation. The tenth, eleventh, and twelfth columns give the sample mean, sample standard deviation (SD), averaged standard deviation computed by profile likelihood (SD^{Prof}) of the 1,000 estimates, and sample mean-squared error (MSE) for β_0, μ_0 , and q_0 , respectively.

We note that the family sizes in the last two rows in Table 1B are random, $2 + \text{Bi}(4, 0.3)$ where $\text{Bi}(4, 0.3)$ is the binomial random variable with parameters 4 and 0.3, implying that the average family size is 3.2.

The starting values used in the algorithms for each study in this section are $\Lambda_1(t) = t/80$, $\beta_1 = 0.9$, $\mu_1 = 3.55$, and $q_1 = 0.16$.

In some of the studies, μ_0 is equal to 4.6052. We note that $\exp(4.6052) = 100$, which is the genetic relative risk used in Siegmund and McKnight (1998), and studied in Mack et al. (1990).

Figure 1 presents the Q-Q plots for the experiment in Scen-14. Figure 1 strongly suggests an asymptotic normality theory for family with varying sizes, which was not discussed in Chang et al. (2005).

Table 1. 0.95 confidence interval coverage, expressed in percentile, sample mean, sample standard deviation (SD), averaged standard deviation computed by profile likelihood (SD^{Prof}), and sample mean-squared error (MSE), for 15 simulation study scenarios.

Scen	K	m	β_0	μ_0	q_0	$L_\beta(\%)$ $CI_\beta(\%)$ $U_\beta(\%)$	$L_\mu(\%)$ $CI_\mu(\%)$ $U_\mu(\%)$	$L_q(\%)$ $CI_q(\%)$ $U_q(\%)$	mean(β) (SD) (SD^{Prof}) MSE	mean($\hat{\mu}$) (SD) (SD^{Prof}) MSE	mean(\hat{q}) (SD) (SD^{Prof}) MSE
1	150	2	1	3	0.1	12.3 85.5 2.2	11.3 87.3 1.4	11.7 87.0 1.3	1.0131 (0.1885) (0.1895) 0.0357	3.0081 (0.4876) (0.5432) 0.2378	0.0996 (0.0302) (0.0374) 0.0009
2	100	3	1	3	0.1	6.1 92 1.9	5.8 92.3 1.9	4.8 93.6 1.6	1.0164 (0.1843) (0.1846) 0.0342	3.0157 (0.4487) (0.4785) 0.2016	0.0998 (0.0295) (0.0361) 0.0009
3	200	2	1	3	0.1	7.5 90.5 2.0	5.4 92.2 2.4	6.6 91.6 1.8	1.0154 (0.1658) (0.1632) 0.0277	2.9955 (0.4082) (0.4601) 0.1666	0.1006 (0.0272) (0.0310) 0.0007
4	133	3	1	3	0.1	3.2 94.6 2.2	3.2 95.4 1.4	2.4 96.5 1.1	1.0011 (0.1573) (0.1593) 0.0247	3.0250 (0.3788) (0.3935) 0.1441	0.0992 (0.0247) (0.0294) 0.0006
5	300	2	1	3	0.1	4.8 93.4 1.8	3.0 94.0 3.0	3.6 94.6 1.8	1.0117 (0.1324) (0.1324) 0.0177	3.0043 (0.3505) (0.3624) 0.1229	0.1015 (0.0222) (0.0245) 0.0005
6	200	3	1	3	0.1	2.4 94.2 3.4	2.2 95.4 2.4	2.0 96.1 1.9	1.0023 (0.1323) (0.1285) 0.0175	3.0037 (0.3055) (0.3042) 0.0933	0.0999 (0.0212) (0.0229) 0.0004
7	450	2	1	3	0.1	3.6 93.6 2.8	2.6 95.4 2.0	1.4 95.8 1.8	1.0031 (0.1088) (0.1071) 0.0118	2.9996 (0.2755) (0.2839) 0.0759	0.1005 (0.0170) (0.0189) 0.0003
8	300	3	1	3	0.1	2.6 95.4 2.0	2.0 95.3 2.7	2.1 95.1 2.8	1.0023 (0.1025) (0.1046) 0.0105	3.0133 (0.2456) (0.2426) 0.0605	0.1005 (0.0169) (0.0179) 0.0003

Table 1B.

Scen	K	m	β_0	μ_0	q_0	$L_\beta(\%)$	$L_\mu(\%)$	$L_q(\%)$	mean($\hat{\beta}$)	mean($\hat{\mu}$)	mean(\hat{q})
						$CI_\beta(\%)$	$CI_\mu(\%)$	$CI_q(\%)$	(SD)	(SD)	(SD)
						$U_\beta(\%)$	$U_\mu(\%)$	$U_q(\%)$	(SD ^{Prof})	(SD ^{Prof})	(SD ^{Prof})
									MSE	MSE	MSE
9	1400	1	1	3	0.1	3.1	16.3	6	1.0076	3.0184	0.1018
						94	74.7	92.4	(0.0897)	(0.3025)	(0.0187)
						2.9	9	1.6	(0.0886)	(0.2713)	(0.0182)
									0.0081	0.0918	0.0004
10	2400	1	1	3	0.1	2.8	12.5	4.5	1.0031	3.0118	0.1008
						94.4	78.5	93.2	(0.0678)	(0.2287)	(0.0133)
						2.8	9	2.3	(0.0676)	(0.2319)	(0.0141)
									0.0046	0.0524	0.0002
11	300	3	1	4.6052	0.05	2.1	4	0.7	1.0077	4.6319	0.0495
						96.1	93.2	98.5	(0.0902)	(0.3830)	(0.0087)
						1.8	2.8	0.8	(0.0950)	(0.3649)	(0.0107)
									0.0082	0.1474	0.0001
12	500	3	1	4.6052	0.01	2.3	3.4	1.1	1.0082	4.5278	0.0105
						94.9	90.3	97.5	(0.0732)	(0.6233)	(0.0042)
						2.8	6.3	1.4	(0.0714)	(0.4115)	(0.0065)
									0.0054	0.3945	0.0000
13	700	3	1	4.6052	0.01	2.4	3.2	1.1	1.0009	4.5955	0.0103
						95.6	92.5	98.7	(0.0571)	(0.4565)	(0.0037)
						2.0	4.3	0.2	(0.0602)	(0.3340)	(0.0050)
									0.0033	0.2085	0.0000
14	300	3.2	1	3	0.1	2.5	2.4	1.6	0.9991	3.0033	0.0994
						95.1	95.9	96	(0.1032)	(0.2210)	(0.0159)
						2.4	1.7	2.4	(0.1004)	(0.2257)	(0.0171)
									0.0107	0.0489	0.0003
15	400	3.2	1	4.6052	0.01	2.6	3.3	2.3	1.0024	4.5421	0.0107
						95.2	90.9	96.5	(0.0770)	(0.6650)	(0.0051)
						2.2	5.8	1.2	(0.0770)	(0.4186)	(0.0071)
									0.0059	0.4462	0.0000

4.2. Bootstrap

These studies confirm the expected benefits of the bootstrap (Efron and Tibshirani (1993)): when the normal approximation is satisfactory, bootstrap methods seem to perform as well; when the normal approximation fails, bootstrap methods may still offer reasonable solutions.

The bootstrap results are reported in Table 2. In Table 2, the first column is the sample size; the second, fourth, and sixth columns report, respectively, the same statistics reported in the seventh, eighth, and ninth columns in Table 1; the third, fifth, and seventh columns are, respectively, the bootstrap counterparts of the second, fourth, and sixth columns; the eighth, ninth, and tenth columns, respectively, report the same statistics in the tenth, eleventh, and twelfth columns in Table 1, except for the replacement of MSE by SD^{Boot} , which is the averaged standard deviation calculated by the bootstrap method. We note that the bootstrap sample size is 1,000 and the coverage is based on 200 replicates.

It is seen from Table 2 that normal approximation and bootstrap are comparable and satisfactory for β , normal approximation is less satisfactory than the bootstrap for μ and q ; in particular, SD^{Boot} is much closer to SD for μ than is SD^{Prof} .

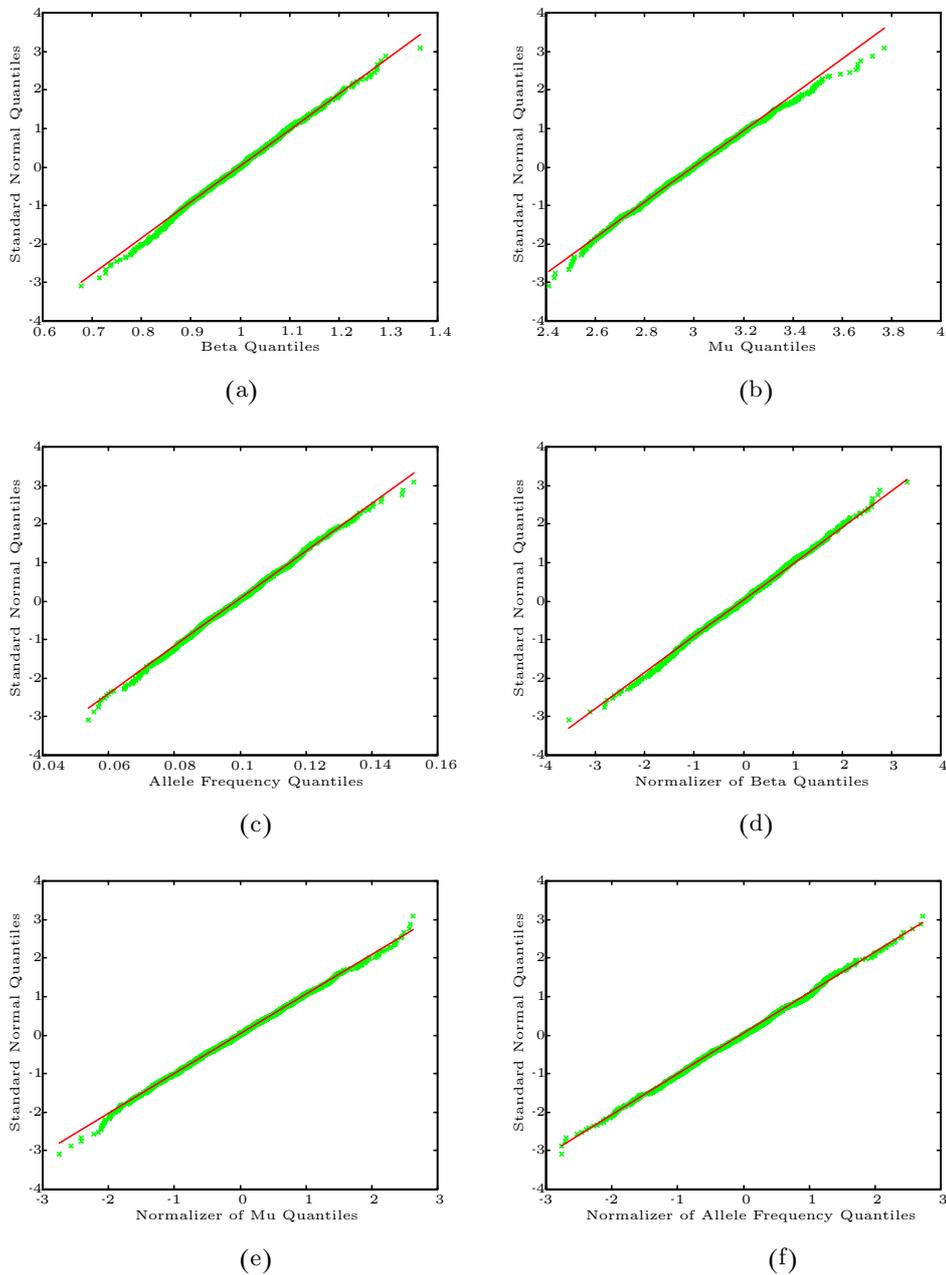


Figure 1. Q-Q plots for the study in Scen-14. (a) Q-Q plot of $\hat{\beta}$ vs. standard normal. (b) Q-Q plot of $\hat{\mu}$ vs. standard normal. (c) Q-Q plot of \hat{q} vs. standard normal. (d) Q-Q plot of normalizer of $\hat{\beta}$ vs. standard normal. (e) Q-Q plot of normalizer of $\hat{\mu}$ vs. standard normal. (f) Q-Q plot of normalizer of \hat{q} vs. standard normal.

Table 2. Comparison of normal approximation and the bootstrap methods for $m = 3, \beta_0 = 1, \mu_0 = 4.6052, q_0 = 0.01$.

K	$L_\beta(\%)$	$L_\beta^{Boot}(\%)$	$L_\mu(\%)$	$L_\mu^{Boot}(\%)$	$L_q(\%)$	$L_q^{Boot}(\%)$	mean($\hat{\beta}$)	mean($\hat{\mu}$)	mean(\hat{q})
	$CI_\beta(\%)$	$CI_\beta^{Boot}(\%)$	$CI_\mu(\%)$	$CI_\mu^{Boot}(\%)$	$CI_q(\%)$	$CI_q^{Boot}(\%)$	(SD)	(SD)	(SD)
	$U_\beta(\%)$	$U_\beta^{Boot}(\%)$	$U_\mu(\%)$	$U_\mu^{Boot}(\%)$	$U_q(\%)$	$U_q^{Boot}(\%)$	(SD ^{Prof})	(SD ^{Prof})	(SD ^{Prof})
							(SD ^{Boot})	(SD ^{Boot})	(SD ^{Boot})
100	2.8	2.9	10.7	0.0	15.0	2.5	1.0217	3.6796	0.0182
	95.3	95.8	51.1	78.4	72.0	97.3	(0.1624)	(1.8351)	(0.0189)
	1.9	1.3	38.2	21.6	13.0	0.2	(0.1639)	(0.6250)	(0.0232)
							(0.1718)	(1.5103)	(0.0195)
150	3.0	3.6	7.7	0.0	9.7	2.3	1.0156	4.0204	0.0149
	95.5	95.6	64.6	82.6	83.2	97.7	(0.1312)	(1.4715)	(0.0115)
	1.5	0.8	27.7	17.4	7.1	0.0	(0.1327)	(0.5932)	(0.0173)
							(0.1377)	(1.3693)	(0.0147)
200	3.5	3.1	5.9	0.1	8.2	1.3	1.0082	4.1378	0.0134
	92.2	93.9	70.3	91.0	86.9	98.7	(0.1201)	(1.2763)	(0.0087)
	4.3	3.0	23.8	8.9	4.9	0.0	(0.1122)	(0.5044)	(0.0141)
							(0.1181)	(1.2669)	(0.0121)

4.3. Starting Values

Since the NPMLE and the algorithms we use are local in nature, the estimates may depend on the starting values β_1, μ_1 , and q_1 . Based on our simulation studies, it seems that the dependence on the starting values is not a serious problem for β and μ , and even for q we need only set the starting value less than 0.5. Using data of Scen-8, we report the results regarding the dependence on q_1 . With the starting values for Λ_1, β_1 , and μ_1 being fixed, we calculate $\hat{\beta}, \hat{\mu}, \hat{q}$ for $q_1 = 0.01, 0.02, \dots, 0.99$. Plots for $(q_1, \hat{\beta}), (q_1, \hat{\mu}), (q_1, \hat{q})$, and a plot for q_1 vs. log-likelihood are presented in Figure 2.

4.4. Remarks on Pedigree Size

Examining (2) closely, we know that the numerical performance of our method depends on pedigree structure only through the number of individuals in the pedigree and the susceptibility probability $p(s, q)$. Because $p(s, q)$ is a polynomial in q , the computation for sibship data is no easier than that for other pedigree data with the same number of individuals in one pedigree. With this understanding, we consider only sibship data in the following.

We report in Table 3 the average computing time needed for analyzing one set of simulated sibship data under the parameter values $\beta_0 = 1, \mu_0 = 3, q_0 = 0.1$

and $\Lambda_0(t) = \log(100/(100 - t))$, and using the algorithms in Section 3. More precisely, for each replicate, we calculate the NPMLE and their asymptotical variances; we report the average of the computing times based on ten replicates. It seems clear from Table 3 that the computing time needed depends mainly on the total number of individuals in the study, not on the size of the pedigree. This, together with Table 1A, indicates that it is more desirable to conduct studies having relatively larger pedigree when the total number of individuals in the study is fixed. We do not recommend a study of only one pedigree, because this offers little possibility to study the variance of the estimates.

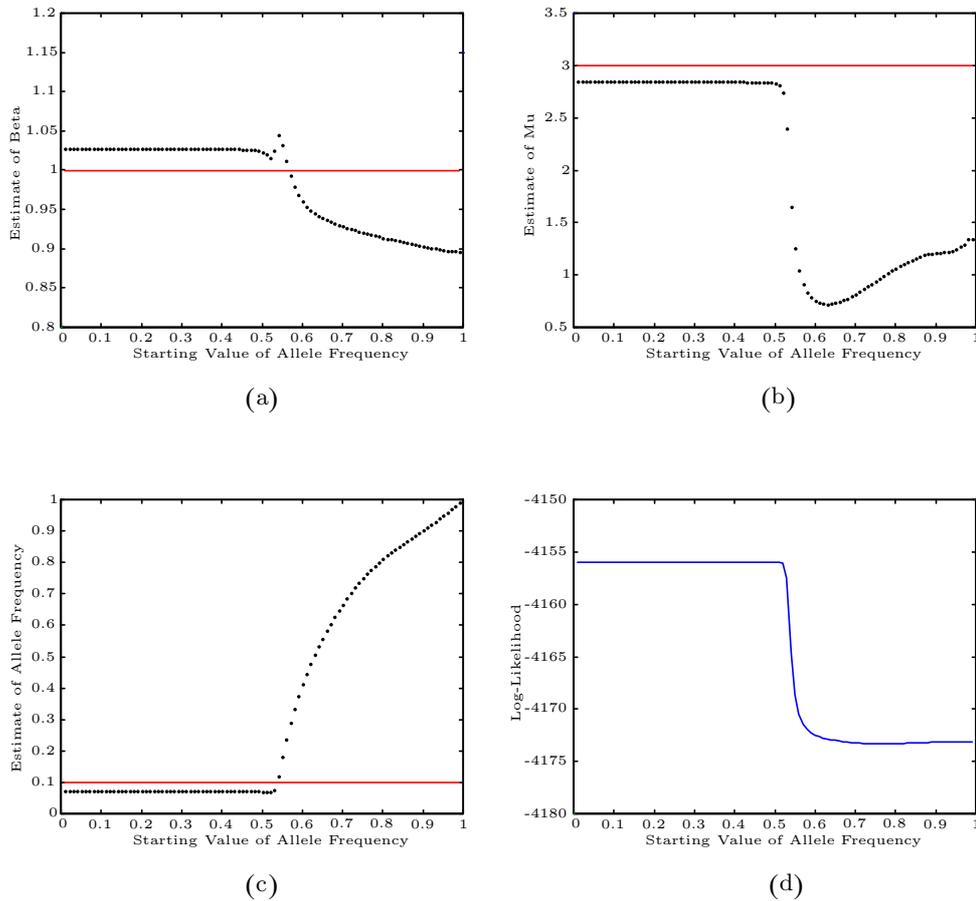


Figure 2. Starting value study in Scen-8. Dotted lines are the estimates; solid lines are true parameters. (a) Plot of q_1 vs. $\hat{\beta}$. (b) Plot of q_1 vs. $\hat{\mu}$. (c) Plot of q_1 vs. \hat{q} . (d) Plot of q_1 vs. log-likelihood.

Table 3. Computing time needed for sibship data when $\beta_0 = 1, \mu_0 = 3, q_0 = 0.1$, and $\Lambda_0(t) = \log(100/(100 - t))$. Here K is the number of families and m is the number of individuals in one family.

K	m	CPU time in seconds
300	4	53.57
400	3	53.56
300	5	73.74
500	3	74.75
300	6	98.30
600	3	99.23

5. Discussion

We have presented a fast algorithm for computing the NPMLE through the Cox-gene model, and used it to study the likelihood theory and the bootstrap methods for the Cox-gene model through simulation studies. Our simulation studies indicate that the normal approximations of the NPMLE work well with reasonable sample sizes. In case of smaller sample sizes for which normal approximation does not work well, we find that bootstrap methods provide a useful alternative.

The algorithms we use in this paper are based on the self-consistency equations derived from the score functions. We studied several other algorithms based on the score functions and found that the one in Section 3 is the best in terms of both accuracy and speed. In particular, the algorithms that replace *ii*), *iii*), and *iv*) in Proposition 1 by Newton-Ralphson methods do not perform comparably. In fact, among the computation procedures we studied, the one in the Section 3 is the only one that makes variance estimation and bootstrap methods feasible. The software, prepared with *Matlab*, is available from the author upon request. We have also prepared software for nuclear families consisting of parents and children.

Although this method seems satisfactory for the Cox-gene model of Li et al. (1998), we understand this work represents only a initial study toward the understanding of the mechanisms of genetic diseases. Serious efforts are needed to take into account multiple genes and environmental factors.

Acknowledgements

We are grateful to Professor Naisyin Wang and two referees for comments that led to improvement of the paper. This work is partially supported by NSC Grant 93-3112-B-400-007-Y.

Appendix

In this Appendix, we provide some heuristic ideas concerning the convergence of Algorithm 3.1 and 3.2. We do not present a rigorous and complete statement, as this is laborious. We assume $m_k = m$ for every k to ease the presentation.

Proposition 2. For given $(\Lambda_1, \beta, \mu, q)$ in its domain, let

$$\Lambda_{J+1}(t) = \int_0^t \frac{1}{W_K(\Lambda_J, \beta, \mu, q; u)} dG_K(u)$$

for $J = 1, 2, \dots$, and $t \in [0, \tau]$. Then there exists Ω_1 with $P(\Omega_1) = 1$ such that for every ω in Ω_1 , there exists a constant $K(\omega)$ such that for every $K > K(\omega)$, $\Lambda_J(\cdot)$ has a convergent subsequence, and the limit $\tilde{\Lambda}$ of any of its convergent subsequences satisfies

$$\tilde{\Lambda}(t) = \int_0^t \frac{1}{W_K(\tilde{\Lambda}, \beta, \mu, q; u)} dG_K(u). \quad (\text{A.1})$$

Proof. It follows from the definition of W_K that there exists a constant $c_2 > 0$ such that

$$W_K(\Lambda, \beta, \mu, q; u) \geq \frac{c_2}{K} \sum_{k=1}^K \sum_{i=1}^m 1_{(0, X_{ik}]}(u)$$

for every K and for every $(\Lambda, \beta, \mu, q, u)$ in its domain. Hence, using the Law of Large Numbers, we get $W(\Lambda, \beta, \mu, q; u) \geq c_2 \sum_{i=1}^m P(X_{i1} \geq \tau)$, where W is the limit of W_K . This indicates that

$$\int_0^\tau \frac{1}{W(\Lambda, \beta, \mu, q; u)} dG(u) < \infty, \quad (\text{A.2})$$

for every (Λ, β, μ, q) in its domain. Using (A.2) and Lemma 3.3 in Chang et al. (2005), we know that there exists a constant c_3 such that

$$\lim_{K \rightarrow \infty} \int_0^\tau \frac{1}{W_K(\Lambda, \beta, \mu, q; u)} dG_K(u) < c_3, \quad (\text{A.3})$$

for every (Λ, β, μ, q) in its domain. It follows from (A.3) that there exists Ω_1 with $P(\Omega_1) = 1$ such that for every ω in Ω_1 , there is a constant $K(\omega)$ such that for every $K > K(\omega)$,

$$\int_0^\tau \frac{1}{W_K(\Lambda, \beta, \mu, q; u)} dG_K(u) < c_3.$$

Thus we know from Helly's Lemma that $\Lambda_J(\cdot)$ has a convergent subsequence, and hence, (A.1) holds. This completes the proof.

Note that the convergence of Algorithm 3.2 follows from Proposition 2. Although Proposition 2 does not by any means imply the convergence of Algorithm 3.1, it together with the following Proposition 3 does provide relevant information. We note that there are extensions to Proposition 3, and they may also be useful in other places.

Proposition 3. Let $\eta_1 : (a, b) \rightarrow \Re$ be a function possessing bounded continuous second derivative. Let $\alpha_0 > 0$ be an isolated local maximum of η_1 and $\eta_1''(\alpha_0) < 0$. Let η_2 and η_3 be two positive and continuously differentiable functions satisfying $\eta_1' = \eta_2 - \eta_3$. Then there exist constants $\varepsilon > 0$ and $n_0 \geq 0$, such that if $|\alpha_1 - \alpha_0| < \varepsilon$ and $\alpha_{J+1} = \alpha_J[(\eta_2(\alpha_J) + n_0)/(\eta_3(\alpha_J) + n_0)]$ for $J = 1, 2, \dots$, then α_J converges to α_0 .

Remarks. The idea behind Proposition 3 is simple and goes as follows. If $0 < \alpha_J < \alpha_0$, then we would like to have $\alpha_{J+1} > \alpha_J$. If α_J is in a suitable neighborhood of α_0 and $0 < \alpha_J < \alpha_0$, then $\eta_2(\alpha_J) - \eta_3(\alpha_J) = \eta_1'(\alpha_J) > 0$, and hence $\alpha_{J+1} > \alpha_J$. Similar comments can be made for the case $\alpha_J > \alpha_0$.

Proof. Let $g_n(x) = [(\eta_2(x) + n)/(\eta_3(x) + n)]$. Using $g_n'(\alpha_0) = (\eta_2'(\alpha_0) - \eta_3'(\alpha_0))/(\eta_3(\alpha_0) + n) < 0$, we let $n_0 > 0$ satisfy $g_{n_0}'(\alpha_0) \geq (2/3) \cdot [(-c-1)/(\alpha_0+1)]$ for some $0 < c < 1$. Let $0 < \varepsilon < 1$ satisfy $\alpha_0 > \varepsilon$, $[\alpha_0 - \varepsilon, \alpha_0 + \varepsilon] \subset (a, b)$, and $|g_{n_0}'(x) - g_{n_0}'(\alpha_0)| < (1/2)|g_{n_0}'(\alpha_0)|$ for every x in $[\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]$. Then

$$\frac{1}{2}g_{n_0}'(\alpha_0) > g_{n_0}'(x) > \frac{3}{2}g_{n_0}'(\alpha_0) \geq \frac{-c-1}{\alpha_0 + \varepsilon} \tag{A.4}$$

for every x in $[\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]$.

Using the equation $[(\alpha_{J+1})/(\alpha_J)] - 1 = g_{n_0}(\alpha_J) - g_{n_0}(\alpha_0)$ and the Mean-Value Theorem, we have

$$\alpha_{J+1} - \alpha_0 = (\alpha_J - \alpha_0)(1 + g_{n_0}'(\tilde{\alpha}_J)\alpha_J) \tag{A.5}$$

for some $\tilde{\alpha}_J$ lying between α_J and α_0 .

Using (A.4), we have $g_{n_0}'(\tilde{\alpha}_1)\alpha_1 \geq [(-c-1)/(\alpha_0 + \varepsilon)]\alpha_1 > -c-1$, and hence, $1 + g_{n_0}'(\tilde{\alpha}_1)\alpha_1 \geq -c > -1$. Using (A.4) again, we have $1 + g_{n_0}'(\tilde{\alpha}_1)\alpha_1 \leq 1 + c_1\alpha_1 \leq 1 + c_1(\alpha_0 - \varepsilon) < 1$, where $0 > c_1 = \max_{x \in [\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]} g_{n_0}'(x)$. These, combined with (A.5), imply that $|\alpha_2 - \alpha_0| < c_2|\alpha_1 - \alpha_0|$ for some $0 < c_2 < 1$. Doing these recursively, completes the proof.

References

Abel, L. and Bonney, G. E. (1990). A time-dependent logistic hazard function for modelling variable age of onset in analysis of familial diseases. *Genetic Epidemiology* **7**, 391-407.

- Chang, I. S., Hsiung, C. A., Wang, M. C. and Wen, C. C. (2005). An asymptotic theory for the nonparametric maximum likelihood estimator in the Cox-gene model. *Bernoulli* **11**, 863-892.
- Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- Elston, R. C. and George, V. T. (1989). Age of onset, age at examination, and other covariates in the analysis of family data. *Genetic Epidemiology* **6**, 217-220.
- Gauderman, W. J. and Thomas, D. (1994). Censored survival models for genetic epidemiology: A Gibbs sampling approach. *Genetic Epidemiology* **11**, 171-188.
- Li, H. and Thompson, E. A. (1997). Semiparametric estimation of major gene and family-specific random effects for age of onset. *Biometrics* **53**, 282-293.
- Li, H., Thompson, E. A. and Wijsman, E. M. (1998). Semiparametric estimation of major gene effects for age of onset. *Genetic Epidemiology* **15**, 279-298.
- Mack, W., Langholz B. and Thomas, D. C. (1990). Survival models for familial aggregation of cancer. *Environmental Health Perspectives* **87**, 27-35.
- Meyer, J. M. and Eaves, L. J. (1988). Estimating genetic parameters of survival distributions: A multifactorial model. *Genetic Epidemiology* **5**, 265-275.
- Siegmund, K. and McKnight, B. (1998). Modeling hazard functions in families. *Genetic Epidemiology* **15**, 147-171.
- Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. *Demography* **16**, 439-454.
- Yashin, A. I. and Iachine, I. A. (1995). Genetic analysis of durations: Correlated frailty model applied to survival of Danish twins. *Genetic Epidemiology* **12**, 529-538.

Institute of Cancer Research and Div. of Biostatistics and Bioinformatics, National Health Research Institutes, 35, Keyan Road, Zhunan Town, Miaoli County 350, Taiwan.

E-mail: ischang@nhri.org.tw

Department of Mathematics, Tamkang University, 151 Ying-chuan Road Tamsui, Taipei County 251, Taiwan.

E-mail: ccwen@mail.tku.edu.tw

Department of Applied Mathematics, Chung Yuan Christian University, 200, Chung Pei Rd., Chung Li 320, Taiwan.

E-mail: yuhjenn@cycu.edu.tw

Department of Information Management, Lunghwa University of Science and Technology, 300, Sec.1, Wanshou Rd., Guishan, Taoyuan 333, Taiwan.

E-mail: yangzq@ihu.edu.tw

(Received January 2005 accepted March 2006)