·方法介绍·

带辅助信息下广义加速危险率模型的估计研究*

长春工业大学 数学与统计学院(130012)　李　洋　王淑影△　马　睿

【提　要】 **目的**　以右删失数据为背景,研究如何在广义加速危险率模型下结合外部辅助信息,提高参数估计精度与应用效果。**方法**　将外部生存概率转化为无偏估计方程,利用经验似然法构建约束似然函数,采用Sieve方法和Bernstein多项式逼近非参数部分,通过R语言Maxlik函数进行参数估计和优化。**结果**　模拟结果表明,所提方法显著提升了参数估计的有效性。在黑色素瘤数据分析中,该方法提高了回归参数的精度,并有效识别了肿瘤尺寸的显著影响。**结论**　本研究提高了参数估计的精度,增强了生存分析的可靠性,有助于识别预后生物标志物,为黑色素瘤等疾病的研究提供方法支持。

【关键词】 右删失数据　广义加速危险率模型　辅助生存信息　经验似然　Sieve方法
【中图分类号】 O212　　【文献标志码】 A　　**DOI**　10.11783/j.issn.1002-3674.2026.03.031

生存分析研究失效时间及其统计规律,重点关注事件(如死亡、失效等)和寿命(从记录开始到事件发生的时间)。在实际数据分析中,由于失效时间难以精确测量,右删失数据较为常见。Freireich等[1]对急性白血病患者的临床试验将患者分为接受6-巯基嘌呤治疗组和安慰剂组,旨在比较两种方案对缓解时间的影响。试验结束时,部分患者仍保持缓解,导致其缓解时间右删失,而其他患者的缓解时间被准确记录。

现有回归分析方法处理删失数据会导致估计效率低下和删失信息浪费,生存分析专注于处理删失数据的方法。最早的生存分析关注生存时间的分布特征,未充分考虑协变量的影响。随着研究深入,参数模型和半参数模型逐渐发展。Cox的比例风险模型中危险率成比例的假设常不成立[2]。Bennett的比例优势模型假设危险率比收敛到常数1[3-4],而Lin等[5]的加性风险模型通过基准危险率与协变量相加来描述危险率。临床试验中,这些模型的适用性常因治疗组与对照组的危险率函数交叉而受限。为此,Chen等[6]提出了加速风险模型,放宽比例假设,适用于相同危险率的情况,Chen等[7]的广义加速危险率模型进一步整合了多种经典模型,展现出更大的灵活性。

随着大数据时代的到来,整合不同来源或类型的数据作为辅助信息变得尤为重要。Huang等[8]利用亚组生存概率作为辅助信息,显著提高了比例风险模型的估计效率。随后,这一方法被应用于其他生存模型,如加性风险模型、线性变换模型及考虑治愈特性的非混合治愈模型等[9-12]。尽管已有大量研究探讨了辅助信息下右删失数据和常见生存模型的参数估计问题,关于广义加速危险率模型的研究相对较少。因此,本文利用经验似然方法结合辅助信息,对广义加速危险率模型的参数估计方法进行优化,提高其估计效率,并增强模型在实际应用中的适用性。

方法介绍

1. 数据、模型和符号

对于右删失数据,设观测时间 $Y=\min(T,C)$,$\delta=I(T\leqslant C)$,其中 T 表示失效时间,C 表示删失时间,δ 是示性变量,$\delta=1$ 表示事件发生,$\delta=0$ 表示数据删失。研究样本为 n 个独立个体,定义样本观测集为 $O=\{O_i: O_i=(Y_i,\delta_i,Z_i), i=1,\cdots,n\}$。

考虑失效时间 T 服从广义加速危险率模型[7]:

$$\lambda(t|Z)=\lambda_0(te^{\beta^TZ})e^{\gamma^TZ} \tag{1}$$

其中,$\lambda_0(\cdot)$ 代表基线风险函数,Z 表示协变量,给定 Z 时,T 和 C 条件独立。β 和 γ 是感兴趣的 P 维回归参数,代表协变量对危险率的影响效应。当 $\beta=0$,模型为比例风险模型;当 $\beta=\gamma$,为加速失效时间模型;当 $\gamma=0$,为加速风险模型。

2. 辅助信息引入

令 $\Omega_m(m=1,2,\cdots,M)$ 表示第 m 个亚组,它的 t^* 年的生存概率已知。那么第 m 个亚组在 t^* 时刻的辅助信息记作[2]:

$$P(T>t_k^*|Z\in\Omega_m)=\pi_{km},k=1,\cdots,K,m=1,\cdots,M$$

根据重期望公式推得:

$$E\left[E\{I(T>t_k^*)I(Z\in\Omega_m)-\pi_{km}I(Z\in\Omega_m)|Z\}\right]$$
$$=E\left[I(Z\in\Omega_m)E\{I(T>t_k^*)-\pi_{km}|Z\}\right]$$
$$=0$$

在广义加速危险率模型下,条件生存函数为:

$$S(Y_i|Z_i)=\exp\left\{-\Lambda_0(Y_ie^{\beta^TZ_i})e^{\gamma^TZ_i-\beta^TZ_i}\right\} \tag{2}$$

进一步有:

$$E\left[I(Z\in\Omega_m)\left\{\exp\left\{-\Lambda_0(t_k^*e^{\beta^TZ_i})e^{\gamma^TZ_i-\beta^TZ_i}\right\}-\pi_{km}\right\}\right]=0$$

*基金项目:吉林省长白英才青年博士后项目(202442110)
△通信作者:王淑影,E-mail: wangshuying0601@163.com

定义估计函数为：

$$\Psi_{km}(Z,\beta,\gamma)=I(Z\in\Omega_m)\left\{\exp\left\{-\Lambda_0(t_k^* e^{\beta^T Z_i})e^{\gamma^T Z_i-\beta^T Z_i}\right\}-\pi_{km}\right\}$$

从而估计方程表示如下：

$$E\{\Psi_{km}(Z,\beta,\gamma)\}=0, k=1,\cdots,K, m=1,\cdots,M$$

3. 约束似然函数推导

本节构造带有辅助信息的约束似然函数。基于观测集O的全似然函数构造如下：

$$\begin{aligned}L(\theta,O)&=\prod_{i=1}^{n}f(Y_i|Z_i)^{\delta_i}S(Y_i|Z_i)^{1-\delta_i}dG(Z_i)\\&=\prod_{i=1}^{n}\left[\lambda(Y_i|Z_i)S(Y_i|Z_i)\right]^{\delta_i}S(Y_i|Z_i)^{1-\delta_i}dG(Z_i)\\&=\prod_{i=1}^{n}\lambda(Y_i|Z_i)^{\delta_i}S(Y_i|Z_i)dG(Z_i)\end{aligned}\tag{3}$$

其中，$f(\cdot)$表示失效时间的概率密度函数，$S(\cdot)$表示生存函数，$G(\cdot)$表示Z的分布函数。

使用经验似然方法[13-14]将无偏估计方程融入观测对数似然函数中，同时满足以下约束条件：

$$\max_{p_1,\cdots,p_n}\sum_{i=1}^{n}\log p_i,\ \text{s.t. } 0\leqslant p_i\leqslant 1,\ \sum_{i=1}^{n}p_i=1,\ \sum_{i=1}^{n}p_i\Psi(Z_i,\beta,\gamma)=0$$

其中，p_i表示样本Z_i的经验概率：

$$p_i=dG(Z_i), i=1,\cdots,n$$

$\Psi(Z_i,\beta,\gamma)$表示辅助信息的集合向量：

$$\Psi(Z_i,\beta,\gamma)=[\Psi_{11}(Z_i,\beta,\gamma),\cdots,\Psi_{KM}(Z_i,\beta,\gamma)]^T$$

根据拉格朗日乘子法，得到优化解：

$$p_i=\frac{1}{n}\frac{1}{\{1+\xi^T\Psi(Z_i,\beta,\gamma)\}}$$

其中，$\xi=(\xi_{11},\cdots,\xi_{KM})^T$是拉格朗日乘子向量，$\xi$满足方程组：

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\Psi(Z_i,\beta,\gamma)}{\{1+\xi^T\Psi(Z_i,\beta,\gamma)\}}=0$$

综上，通过引入辅助生存信息，得到约束对数似然函数如下：

$$\begin{aligned}l(\theta,O)&=\sum_{i=1}^{n}\left[\delta_i\lambda(Y_i|Z_i)+S(Y_i|Z_i)\right]+\sum_{i=1}^{n}\log dG(Z_i)\\&=\sum_{i=1}^{n}\left\{\delta_i\left[\gamma^T Z_i+\log\lambda_0(Y_i e^{\beta^T Z_i})\right]-\Lambda_0(Y_i e^{\beta^T Z_i})e^{\gamma^T Z_i-\beta^T Z_i}\right\}\\&-\sum_{i=1}^{n}\log\{1+\xi^T\Psi(Z_i,\beta,\gamma)\}\end{aligned}\tag{4}$$

其中，参数向量$\theta=\left((\beta^T,\gamma^T,\xi^T)^T,\Lambda_n\right)^T$。

4. Sieve极大似然估计

因为模型包含无穷维的非参数部分Λ_0，所以本节基于Sieve方法，采用Bernstein多项式对非参数部分进行拟合逼近。

首先，定义Sieve空间：

$$\Theta_n=\left\{\theta=(\eta,\Lambda_n)^T,\eta=(\beta^T,\gamma^T,\xi^T)^T\in B,\Lambda_n\in\Gamma_n\right\}$$

其中，参数空间：

$$B=\{\eta:\eta\in R^{p+p+K\times M},\|\eta\|\leqslant L, L\in(0,\infty)\}$$

函数空间：

$$\Gamma_n=\{\Lambda_n:\Lambda_n(t)=\sum_{k=0}^{N}\phi_k B_{k,N}(t),0\leqslant\phi_0\leqslant\cdots\leqslant\phi_N,\sum_{k=0}^{N}|\phi_k|\leqslant M_n, t\in[l,r]\}$$

其中，$t=Ye^{\beta^T Z}, l=\min(Ye^{\beta^T Z}), r=\max(Ye^{\beta^T Z})$，$\phi_0,\cdots,\phi_N$为Bernstein多项式系数，通过添加$0\leqslant\phi_0\leqslant\cdots\leqslant\phi_N$的约束，保证$\Lambda_0(t)$非负且单调不减的性质。$M_n=O(n^a), a>0$。$B_{k,N}(t)$是阶数为$N=O(n^v)$，$v\in(0,1)$的Bernstein基函数[15]：

$$B_{k,N}(t)=\binom{N}{k}\left(\frac{t-l}{r-l}\right)^k\left(1-\frac{t-l}{r-l}\right)^{N-k}$$

通过Sieve方法得到感兴趣参数估计量记为$\hat\eta=(\hat\beta^T,\hat\gamma^T)^T$。对$\hat\eta$的协方差估计，采用非参数bootstrap方法[16-17]，基于观测数据O，生成B个独立的bootstrap样本$O^{(b)}=\{O_i^{(b)};\ i=1,\ldots,n\}$，利用bootstrap样本得到$\hat\eta$的估计量记为$\hat\eta^{(b)}$，则$\hat\eta$协方差估计量可定义为：

$$\hat\Sigma=\frac{1}{B-1}\sum_{b=1}^{B}\left\{\hat\eta^{(b)}-\frac{1}{B}\sum_{b=1}^{B}\hat\eta^{(b)}\right\}^{\otimes 2}$$

其中，$a^{\otimes 2}=aa^T$。

模拟研究

本节通过数值模拟来评估所提方法在有限样本下的性能表现。失效时间T使用逆变换法从广义加速危险率模型中生成：

$$\lambda(t|Z)=\lambda_0(te^{\beta^T Z})e^{\gamma^T Z}$$

设置不同情形下的参数真值组合为：

$$\beta=(1.3,1.1),(-1.5,-1.2),(0.4,0.4)$$

$$\gamma=(1.4,1.2),(-1,-0.6),(1.2,1.2)$$

删失时间C服从(C_0,C_1)上的均匀分布，设置删失率约为10%和20%。基线风险函数设为两种形式，$\lambda_0(t)=\log(1+t)$和$\lambda_0(t)=1/(1+t)$。协变量Z_i为2维，第一种情况是$Z_1\sim B(1,0.5), Z_2\sim U(0,2)$。根据协变量值域划分亚组情形1：

$$\Omega_1=\{(Z_1,Z_2):Z_1=0,Z_2\leqslant 1\},\Omega_2=\{(Z_1,Z_2):Z_1=0,Z_2>1\}$$

$$\Omega_3=\{(Z_1,Z_2):Z_1=1,Z_2\leqslant 1\},\Omega_4=\{(Z_1,Z_2):Z_1=1,Z_2>1\}$$

第二种情况是$Z_1,Z_2\sim B(1,0.5)$。对应亚组情形2：

$$\Omega_1=\{(Z_1,Z_2):Z_1=0,Z_2=0\},\Omega_2=\{(Z_1,Z_2):Z_1=0,Z_2=1\}$$

$\Omega_3 = \{(Z_1, Z_2): Z_1 = 1, Z_2 = 0\}, \Omega_4 = \{(Z_1, Z_2): Z_1 = 1, Z_2 = 1\}$

针对两种情形分别选取生存时间点 $t_k^* = 0.5$ 和 $t_k^* = 1$。样本量设置为 $n = 200$ 和 $n = 400$，模拟次数为1000次，bootstrap抽样次数为50。

研究对比了加入辅助信息的约束经验似然（empirical likelihood，EL）估计与不考虑辅助信息的极大似然（maximum likelihood，ML）估计两种方法的估计效果。

EL方法下的约束对数似然函数为：

$$l_{EL}(\theta,O) = \sum_{i=1}^{n}\left[\delta_i\lambda(Y_i|Z_i) + S(Y_i|Z_i)\right] + \sum_{i=1}^{n}\log dG(Z_i)$$

$$= \sum_{i=1}^{n}\left\{\delta_i\left[\gamma^T Z_i + \log\lambda_0(Y_i e^{\beta^T Z_i})\right] - \Lambda_0(Y_i e^{\beta^T Z_i}) e^{\gamma^T Z_i - \beta^T Z_i} - \sum_{i=1}^{n}\log\{1 + \xi^T\Psi(Z_i,\beta,\gamma)\}\right\}$$

EL方法利用协变量 Z 的经验分布，通过经验似然方法结合外部辅助信息与似然函数。通过引入辅助信息来增加样本信息量，在保持模型结构不变的前提下，通过约束优化提高参数估计的效率。

ML方法下的对数似然函数为：

$$l_{ML}(\theta,O) = \sum_{i=1}^{n}\left[\delta_i\lambda(Y_i|Z_i) + S(Y_i|Z_i)\right]$$

$$= \sum_{i=1}^{n}\left\{\delta_i\left[\gamma^T Z_i + \log\lambda_0(Y_i e^{\beta^T Z_i})\right] - \Lambda_0(Y_i e^{\beta^T Z_i}) e^{\gamma^T Z_i - \beta^T Z_i}\right\}$$

ML方法依赖样本数据构建对数似然函数，直接通过优化该函数来估计模型参数。该方法的估计结果在样本量充分大时较为稳健，但在样本量有限情况下，估计精度可能会有所下降。

本部分数值模拟的主要流程包括：首先生成符合模型假设的右删失数据集；接着，基于生成的数据构建两种方法（EL和ML）下的对数似然函数；之后利用R软件中maxLik函数进行参数估计和优化；最后，计算评价指标以评估两种方法的估计效果。

评估指标包括经验偏差（bias）、估计值的样本标准差、估计标准误的平均值和95%的经验覆盖概率（coverage probability，CP）。在亚组情形1下，基于 $\lambda_0(t) = \log(1 + t)$，变换不同删失率，结果见表1。在亚组情形2下，变换不同形式的基线风险函数，结果见表2和表3。

ML和EL估计绝大多数bias小于0.05，SE与SD接近，CP绝大多数接近95%。当样本量从200增加到400，各项评价指标呈现出降低趋势，且在低删失率下，各项指标结果相对更好。此外，EL方法下离散协变量对应参数的SE相比于ML方法显著降低，说明对于广义加速危险率模型，通过对离散协变量分组得到的辅助信息在提高参数估计精度方面更为有效。

此外，考虑Bernstein多项式阶数取 $N = 3, 4, 5, 6$，bootstrap抽样次数取 $B = 50, 100, 200$ 的情况。根据表4结果，在不同 N 和 B 选择下，两种估计方法依然表现出较好的稳健性，且EL方法始终优于ML方法。

图1~4展示了 $n = 400$ 时，ML与EL方法对于 Λ_0 的估计。图中的黑色实线表示真实累积基线风险曲线，灰色实线为每次估计的累积基线风险曲线，黑色虚线则表示估计值取平均后的累积基线风险曲线。结果表明，与ML方法相比，EL方法的变异性更低，说明EL方法显著提高了参数估计的准确性和稳定性。

表1 $\lambda_0(t) = \log(1 + t)$，亚组情形1下EL和ML估计结果

参数	n	方法	删失率10%				删失率20%			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=1.3	200	ML	−0.035	0.210	0.212	0.962	−0.016	0.197	0.222	0.969
		EL	0.023	0.187	0.187	0.939	0.041	0.197	0.179	0.935
	400	ML	−0.037	0.144	0.165	0.967	−0.021	0.166	0.173	0.959
		EL	0.017	0.179	0.168	0.949	0.009	0.162	0.157	0.940
β_2=1.1	200	ML	−0.025	0.352	0.348	0.958	0.002	0.333	0.355	0.961
		EL	0.033	0.221	0.220	0.937	0.028	0.218	0.205	0.935
	400	ML	−0.045	0.287	0.293	0.962	−0.010	0.274	0.290	0.962
		EL	0.021	0.200	0.193	0.939	0.027	0.192	0.175	0.930
γ_1=−1.4	200	ML	0.052	0.169	0.171	0.963	0.033	0.157	0.179	0.972
		EL	−0.044	0.215	0.215	0.948	−0.076	0.232	0.208	0.926
	400	ML	0.060	0.126	0.132	0.930	0.042	0.127	0.135	0.960
		EL	0	0.177	0.176	0.937	−0.033	0.174	0.172	0.941
γ_2=1.2	200	ML	0.030	0.281	0.281	0.966	0.013	0.258	0.287	0.967
		EL	−0.033	0.214	0.246	0.958	−0.069	0.269	0.244	0.936
	400	ML	0.053	0.221	0.232	0.964	0.028	0.214	0.229	0.967
		EL	−0.014	0.210	0.201	0.937	−0.033	0.226	0.206	0.949

表2　亚组情形2下，$\lambda_0(t)=\log(1+t)$的EL和ML估计结果

参数	n	方法	删失率10%				删失率20%			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=−1.5	200	ML	0.004	0.298	0.298	0.959	−0.012	0.364	0.344	0.951
		EL	0.014	0.175	0.197	0.940	0.019	0.190	0.207	0.939
	400	ML	0.001	0.276	0.260	0.935	0.030	0.329	0.311	0.930
		EL	0.039	0.144	0.155	0.943	0.059	0.149	0.168	0.932
β_2=−1.2	200	ML	0.032	0.307	0.289	0.945	0.014	0.336	0.325	0.945
		EL	0.043	0.167	0.193	0.944	−0.025	0.177	0.197	0.939
	400	ML	0.041	0.278	0.251	0.929	−0.016	0.307	0.294	0.946
		EL	0.052	0.140	0.148	0.926	0.001	0.132	0.157	0.950
γ_1=−1	200	ML	0.025	0.218	0.229	0.964	0.039	0.260	0.261	0.947
		EL	0.025	0.196	0.225	0.951	0.017	0.205	0.229	0.951
	400	ML	0.039	0.200	0.191	0.934	0.022	0.234	0.232	0.950
		EL	0.058	0.172	0.173	0.912	0.043	0.164	0.176	0.939
γ_2=−0.6	200	ML	−0.003	0.218	0.223	0.966	0.023	0.240	0.253	0.958
		EL	−0.002	0.183	0.210	0.940	0.032	0.214	0.214	0.947
	400	ML	0.006	0.194	0.187	0.945	0.052	0.218	0.218	0.943
		EL	0.019	0.155	0.166	0.944	0.060	0.163	0.174	0.936

表3　亚组情形2下，$\lambda_0(t)=1/(1+t)$的EL和ML估计结果

参数	n	方法	删失率10%				删失率20%			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=0.4	200	ML	−0.043	0.584	0.584	0.962	−0.003	0.589	0.565	0.951
		EL	0.001	0.136	0.105	0.903	0.039	0.139	0.136	0.942
	400	ML	−0.014	0.394	0.432	0.979	−0.015	0.416	0.459	0.972
		EL	−0.011	0.101	0.089	0.927	0.013	0.136	0.122	0.922
β_2=0.4	200	ML	−0.051	0.570	0.577	0.959	−0.006	0.588	0.579	0.959
		EL	0.010	0.114	0.103	0.913	0.045	0.133	0.136	0.929
	400	ML	−0.009	0.400	0.433	0.971	−0.015	0.437	0.451	0.967
		EL	−0.003	0.102	0.090	0.934	0.023	0.154	0.125	0.925
γ_1=1.2	200	ML	0.027	0.193	0.194	0.950	0.021	0.200	0.196	0.951
		EL	0.022	0.143	0.127	0.939	−0.019	0.155	0.166	0.954
	400	ML	0.026	0.154	0.143	0.931	0.011	0.145	0.148	0.954
		EL	0.022	0.102	0.099	0.938	−0.004	0.143	0.139	0.950
γ_2=1.2	200	ML	0.028	0.193	0.194	0.954	0.022	0.192	0.197	0.960
		EL	0.004	0.146	0.131	0.949	−0.032	0.154	0.165	0.948
	400	ML	0.027	0.150	0.144	0.947	0.007	0.154	0.148	0.949
		EL	−0.001	0.111	0.100	0.950	−0.026	0.142	0.138	0.949

实例分析

本节对SEER数据库中的黑色素瘤数据展开分析。黑色素瘤是一种源自皮肤黑色素细胞的恶性肿瘤，全球发病率逐年上升。早期发现和治疗的患者生存率较高，一旦扩散至淋巴结或器官，生存概率将大幅降低。本研究旨在揭示影响黑色素瘤患者生存时间的关键临床特征，为早期干预和个性化治疗提供依据[18-19]。定义失效事件为因黑色素瘤致死，失效时间为疾病确诊至死亡的时间。研究包含2008—2018年49804例黑色素瘤患者的生存数据，随机抽取300例作为小样本数据。对失效时间构建广义加速危险率模型$\lambda(t|Z)=\lambda_0(te^{\beta_1 Z_1+\beta_2 Z_2})e^{\gamma_1 Z_1+\gamma_2 Z_2}$，其中$Z_1$表示肿瘤尺寸，作为离散协变量，取值为1时表示肿瘤尺寸大于4mm。Z_2表示年龄，作为连续型协变量并进行标准化处理。

首先，对黑色素瘤的总体数据进行探索性分析，以获取生存数据的基本特征。图5和图6显示，肿瘤尺寸对总体数据的生存状况有显著影响，但在小样本下，生存曲线置信区间交叉，与总体曲线走势不一致。

接下来，对总体数据采用ML方法分析，并对样本量为300的数据分别使用ML和EL方法分析，估计回归模型参数并对比效果。

表5结果显示，ML方法在小样本下未能检测出肿瘤尺寸的影响，与总体数据结果不符，反映出ML方法在小样本中的局限性。相反，EL方法在年龄和肿瘤尺寸上都是显著的，与总体数据分析一致，并提高了估计的有效性。结果与Han等[10]关于肿瘤尺寸显著性的分析一致，表明EL方法充分利用辅助信息，提供了更精确的参数估计并能更好地识别预后生物标志物。

表 4 不同阶数 N 和抽样次数 B 下 EL 和 ML 估计结果

参数	n	方法	B=50,N=3				B=50,N=4			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=0.4	200	ML	−0.042	0.547	0.604	0.976	−0.029	0.578	0.597	0.966
		EL	0.011	0.167	0.158	0.942	0.009	0.135	0.106	0.910
β_2=0.4	200	ML	−0.022	0.532	0.588	0.968	−0.025	0.541	0.571	0.970
		EL	0.020	0.182	0.153	0.923	0.010	0.135	0.106	0.921
γ_1=1.2	200	ML	0.051	0.198	0.193	0.936	0.036	0.182	0.194	0.956
		EL	0.051	0.169	0.175	0.944	0.025	0.132	0.128	0.939
γ_2=1.2	200	ML	0.059	0.195	0.194	0.945	0.029	0.197	0.193	0.940
		EL	0.031	0.199	0.182	0.947	0.010	0.153	0.132	0.946
参数	n	方法	B=50,N=5				B=50,N=6			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=0.4	200	ML	−0.021	0.537	0.552	0.962	0.049	0.478	0.476	0.944
		EL	0.016	0.105	0.083	0.912	0.032	0.092	0.080	0.921
β_2=0.4	200	ML	0.001	0.552	0.527	0.956	0.029	0.516	0.492	0.950
		EL	0.025	0.102	0.083	0.913	0.040	0.091	0.078	0.920
γ_1=1.2	200	ML	0.023	0.197	0.191	0.957	0.027	0.183	0.184	0.956
		EL	0.012	0.118	0.112	0.946	−0.002	0.127	0.112	0.933
γ_2=1.2	200	ML	0.028	0.201	0.191	0.944	0.027	0.185	0.187	0.951
		EL	−0.008	0.125	0.113	0.942	−0.029	0.108	0.111	0.940
参数	n	方法	B=100,N=4				B=200,N=4			
			bias	SD	SE	CP	bias	SD	SE	CP
β_1=0.4	200	ML	−0.030	0.559	0.595	0.961	−0.043	0.565	0.597	0.971
		EL	0.006	0.103	0.108	0.937	0.007	0.124	0.109	0.928
β_2=0.4	200	ML	−0.021	0.551	0.583	0.971	−0.029	0.590	0.587	0.965
		EL	0.008	0.113	0.107	0.930	0.016	0.117	0.107	0.927
γ_1=1.2	200	ML	0.030	0.195	0.194	0.945	0.025	0.197	0.195	0.959
		EL	0.021	0.125	0.132	0.959	0.029	0.122	0.133	0.957
γ_2=1.2	200	ML	0.033	0.193	0.193	0.953	0.028	0.191	0.195	0.964
		EL	0.003	0.123	0.134	0.961	0.005	0.142	0.135	0.950

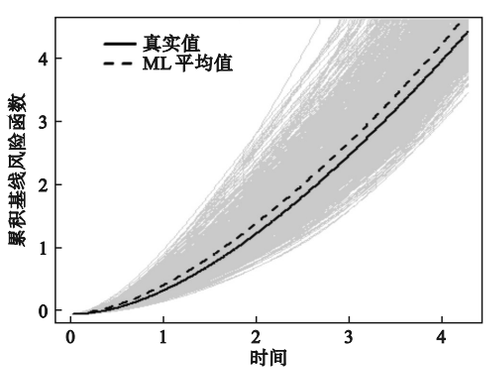


图 1 n=400,基于 $\lambda_0(t)=\log(1+t)$ 的 Λ_0 估计图(ML)

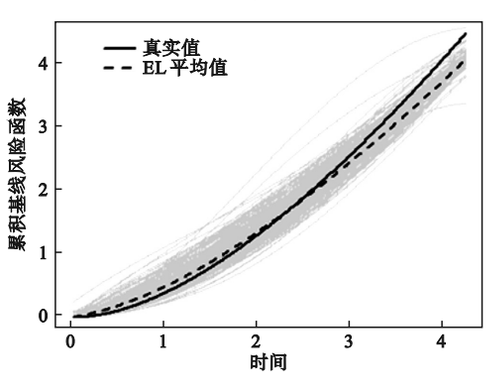


图 2 n=400,基于 $\lambda_0(t)=\log(1+t)$ 的 Λ_0 估计图(EL)

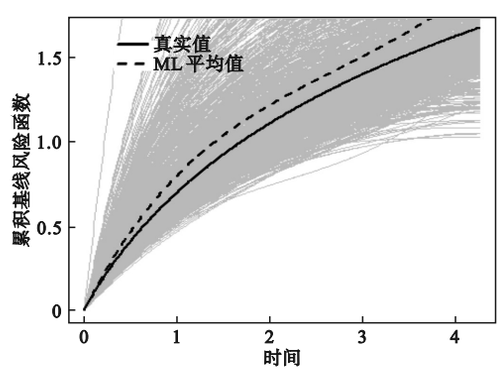


图 3 n=400,基于 $\lambda_0(t)=1/(1+t)$ 的 Λ_0 估计图(ML)

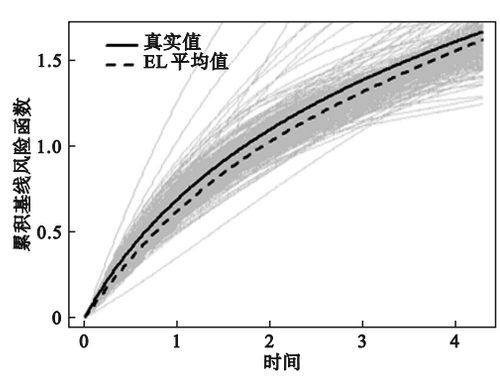


图 4 n=400,基于 $\lambda_0(t)=1/(1+t)$ 的 Λ_0 估计图(EL)

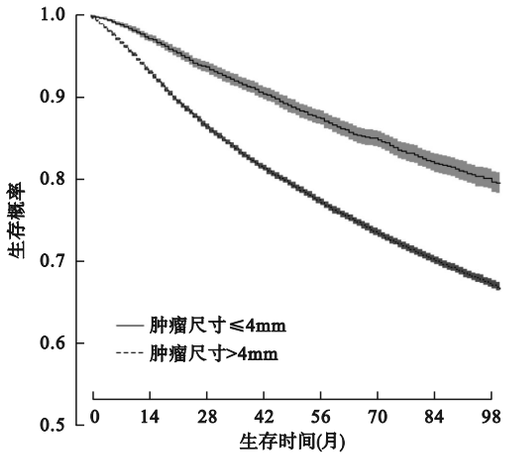


图5　总体数据下不同肿瘤尺寸黑色素瘤病例的生存曲线

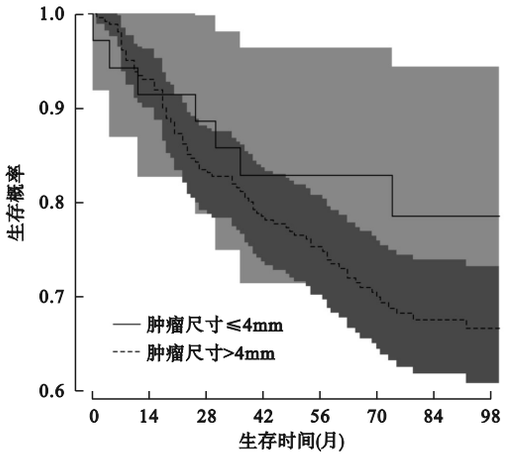


图6　小样本数据下不同肿瘤尺寸黑色素瘤病例的生存曲线

表5　黑色素瘤数据回归参数估计结果

项目	β_1			β_2			γ_1			γ_2		
	Est	SE	*P*	Est	SE	*P*	Est	SE	*P*	Est	SE	*P*
Aggregate	1. 508	0. 816	0. 032	2. 689	0. 807	<0. 001	0. 458	0. 146	<0. 001	4. 443	0. 386	<0. 001
ML	1. 929	1. 374	0. 080	3. 495	1. 354	0. 005	0. 453	0. 475	0. 170	4. 491	0. 946	<0. 001
EL	1. 959	0. 210	<0. 001	1. 945	0. 252	<0. 001	1. 986	0. 370	<0. 001	2. 111	0. 532	<0. 001

讨　论

本文研究了右删失数据下融合辅助信息的广义加速危险率模型的参数估计问题。首先将外部生存概率转化为无偏估计方程，利用经验似然方法结合观测似然函数和无偏估计方程，进而构造出约束似然函数。基于Sieve方法，采用了Bernstein多项式逼近非参数部分。利用Maxlik函数进行了参数估计和优化。模拟结果显示，EL方法显著提高了参数估计的有效性。对黑色素瘤数据的分析表明，EL方法不仅提供了更精确的回归参数估计，还有效评估了小样本中肿瘤尺寸对生存状况的影响。

展望未来，本文方法可扩展至区间删失数据、相依删失数据等。当前研究重点是如何在右删失数据中引入辅助信息提升估计精度，并进一步探讨模型估计的理论性质。未来需研究应对数据异质性的方法，确保融合辅助信息时协变量效应估计更准确。

参　考　文　献

[1] Freireich EJ, Gehan E, Frei E, et al. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia: A model for evaluation of other potentially useful therapy [J]. Blood, 1963, 21(6): 699-716.

[2] Cox DR. Regression Models and Life-tables [J]. Journal of the Royal Statistical Society, Series C, 1972,34(2): 187-202.

[3] Bennett S. Analysis of Survival Data by the Proportional Odds Model [J]. Statistics in Medicine, 1983, 2(2): 273-277.

[4] 宇传华,余松林,徐勇勇. 比例优势模型实现ROC分析的方法及其应用前景分析[J]. 中国卫生统计,2005(5):293-296.

[5] Lin DY, Ying Z. Semiparametric Analysis of the Additive Risk Model [J]. Biometrika, 1994, 81(1): 61-71.

[6] Chen YQ, Wang MC. Analysis of accelerated hazards models [J]. Journal of the American Statistical Association, 2000, 95(450): 608-618.

[7] Chen YQ, Jewell NP. On a general class of semiparametric hazards regression models [J]. Biometrika, 2001, 88(3): 687-702.

[8] Huang CY, Qin J, Tsai HT. Efficient estimation of the Cox model with auxiliary subgroup survival information [J]. Journal of the American Statistical Association, 2016, 111(514): 787-799.

[9] Shang W, Wang X. The generalized moment estimation of the additive-multiplicative hazard model with auxiliary survival information [J]. Computational Statistics and Data Analysis, 2017, 112 (1): 154-169.

[10] Han B, Van KI, Wang X. Semiparametric estimation of the nonmixture cure model with auxiliary survival information [J]. Biometrics, 2022, 78(2): 448-459.

[11] Sheng Y, Sun Y, Huang CY, et al. Synthesizing external aggregated information in the penalized Cox regression under population heterogeneity [J]. Statistics in Medicine, 2021, 40(23): 4915-4930.

[12] Fang F, Long T, Shao J, et al. An Integrated GMM Shrinkage Approach with Consistent Moment Selection from Multiple External Sources [J]. Journal of Computational and Graphical Statistics, 2025, 34(4): 1670-1679.

[13] Owen AB. Empirical likelihood ratio confidence intervals for a single functional [J]. Biometrika, 1988, 75(2): 237-249.

[14] Qin J, Lawless J. Empirical likelihood and general estimating equations[J]. The Annals of Statistics, 1994, 22(1): 300-325.

[15] 张中文,许晴晴,王玖,等. 基于Bernstein多项式的半变系数组合诊断方法研究[J]. 中国卫生统计,2023,40(3):377-381.

[16] Efron B. Bootstrap methods: Another look at the jackknife [J]. The Annals of Statistics, 1979, 7(1): 1-26.

[17] 毛昂,曾子倩,魏敏,等. Bootstrap法与H-L法中位数差值的区间检验在非劣效试验中的模拟比较研究[J]. 中国卫生统计,2020,37(6):848-850.

[18] 沈其君,金宝相,闵捷,等. Cox模型及其在恶性黑色素瘤预后因素分析中的应用[J]. 中国卫生统计,1988(2):30-32.

[19] 陈明,郑肃政,黄煜,等. 基于SEER数据库的恶性雀斑样痣黑色素瘤术后预后列线图的构建与验证[J]. 医药前沿,2026,16(9):18-24.

（责任编辑：林美艺）