IN: Mathematical Population Dynamics: Analysis of Heterogeneity Vol 2: Carcinogenesis and Cell & Tumor Growth
Arino O, Axelrod D, Kimmel M. (eds)
Wuerz Publications, Winnipeg, Manitoba, Canada
1995 p283-96

21

Modeling breast cancer recurrence[1]

Thu Hoàng[2]
Université René Descartes
15 rue des Sts Pères,
75006 Paris, France
Phone: 33(1) 4286–2116
FAX: 33(1) 4286–0402
e-mail: hoang@citi2.fr

Alexander Tsodikov
St Petersburg Technical University
Polytechnicheskaya ul.29
St Petersburg, Russia

Andrej Yakovlev
St Petersburg Technical University
Polytechnicheskaya ul.29
St Petersburg, Russia

Bernard Asselain
Institut Curie
26 rue d'Ulm
75005 Paris, France

Abstract

Studies of cancer treatment by surgery and/or radiotherapy often use time to local failure as an indicator of treatment efficiency. We model local recurrence of breast cancer as the result of two competing risks which we relate to two populations of tumor cells. By maximum likelihood estimation, we obtain the estimates of the expected numbers of clonogens for early and late recurrences and their progression time parameters. The estimates are coherent with clinical observations. The difference in the mechanisms of the two types of recurrence reflects the differential tumor aggressiveness.

Key words: Parametric model, breast cancer recurrence, competing risks, ML estimates.

[1]We are thankful to a referee for valuable. comments and suggestions leading to substantial improvements of the original manuscript.

[2]Research partially supported by ARC (Association pour la Recherche sur le Cancer).

21.1 Introduction

In recent years, conservative treatment of breast cancer by local surgery and/or radiotherapy has become a widely accepted alternative to mutilating mastectomy. The evaluation of such treatment techniques is often based on the risk of local recurrence [5,15]. Local recurrence refers to any tumor relapse within the treated breast. It represents either a true recurrence or a second tumor. In practice, its localization gives an indication of its origin. A recurrence at or in the immediate vicinity of the original tumor bed may be interpreted as a true recurrence whereas a recurrence elsewhere is considered frequently as a possible second cancer or a preexisting subclinical tumor at the time of the treatment.

Within 5 years of the primary treatment local failure occurs in 2% to 30% patients. From 5 years up to 20 years and perhaps beyond, the local recurrence rate is 5% to 25% [15]. Early recurrence seems to be related to metastasis [4,5,13]. Since it carries poorer prognosis than late recurrence it is considered as a marker of aggressive tumor biology.

In order to better understand the process of local tumor progression, we propose a modelling of the two types of recurrence dynamics observed in the clinical settings.

21.2 The model

At the end of the treatment, the cells that will propagate into a newly detectable tumor—we call them clonogens—are surviving neoplastic cells capable of giving rise to tumor regeneration. Consider the case when tumors are exposed to large single doses of radiation or chemotherapy. In this case it is natural to assume that the number N of clonogens prior to irradiation is very large but the probability p of their survival after the treatment is very low. If N is nonrandom, one may confidently consider the number ν of surviving clonogens as a Poisson random variable (r.v.) The probability η of tumor cure (no surviving clonogens) is given by

$$\eta = P(\nu = 0) = e^{-\theta} \tag{1}$$

where $\theta = Np$ is the mean number of clonogens surviving the treatment.

If N is random the situation is not simple except when N is also a Poisson variable—in which case formula (1) remains valid. Considering that cell proliferation might occur during the time intervals between successive fractions of radiation, in principle one can no longer expect the number of surviving clonogens to be Poisson. In a computer experiment, Tucker et al. [24] showed that deviations from Poisson statistics might result in a bias of about 10% for the estimate of the probability of cure in most standard treatment regimens. However in our view this (small) bias has been overestimated due to the chosen probability of cell division between consecutive fractions. This probability—set by the authors to be 0.4—is too high in view of:

- the typical mitotic cycle duration in tumors,
- the asynchronous entry of cells into the prereplicative period after irradiation,

- the radiation-induced block of DNA synthesis and mitosis which frequently exceeds the one-day interval chosen by Tucker et al. (see the discussion in Yakovlev and Zorin [25]).

Therefore, we rejoin most authors in maintaining the assumption that the number of remaining clonogens is a Poisson variable [18,19,20,21,22]. We proceed from this assumption to develop our model of tumor recurrence.

Each surviving clonogenic cell possesses in the long run the capacity of giving rise to an overt tumor. Let X_i be the random time for the i-th clonogen to produce a detectable tumor. By analogy with the terminology accepted in carcinogenesis studies we call X_i the progression time. Nonnegative r.v.'s X_i for $i = 1, 2, ...$ are assumed to be independent and identically distributed (i.i.d.) with common distribution function $F(x)$. The time to tumor recurrrence can be defined as the random minimum

$$U = \min_{0 \le i \le \nu} X_i \ ,$$

where $X_0 = +\infty$ with probability one.

If ν is a Poisson r.v. independent of the sequence $X_1, X_2, ...$, the survival function $\bar{G}(t) = 1 - G(t)$ for the r.v. U can be easily obtained:

$$\bar{G}(t) = P(U \ge t) = \sum_{k=0}^{\infty} \frac{\theta^k}{k!} e^{-\theta} \big(1 - F(t)\big)^k = e^{-\theta F(t)} \ . \tag{2}$$

The key advantage of this model is to show explicitly the contribution of the two characteristics of tumor growth: the mean number of clonogens θ and the rate of their progression described by the function $F(t)$. Their estimation, if feasible, furnishes additional information on the biology of tumor recurrence, thereby offering a more refined interpretation of observational data. The survival function $\bar{G}$ corresponds to a substochastic distribution and its limiting value $\bar{G}(+\infty) = e^{-\theta}$ represents the probability of tumor cure (compare with formula (1)). Most parametric models implicitly assume a zero limiting survival probability [3,10]. The importance of allowing for a survival fraction in failure-time models has been recognized by several authors [8,11,14,16,17,27].

The hazard function $\lambda(t)$ defined with respect to $\bar{G}(t)$ is

$$\lambda(t) = \theta f(t) \tag{3}$$

where f is the density of the distribution F. If the progression time distribution F is unimodal, then the hazard function $\lambda(t)$ has a maximum. Note that the assumption on the exponentiality of F should be rejected since that would correspond to the unrealistic case of a monotone decreasing hazard.

To describe a possible heterogeneity of clonogens with respect to the progression time distribution, introduce k different types of tumor cells with distributions $F_j(t)$. Then the progression time distribution F is represented by a finite mixture

$$F(t) = \sum_{j=1}^{k} q_j F_j(t), \quad 0 < q_j < 1, \quad \sum_{j=1}^{k} q_j = 1 \ . \tag{4}$$

This mixture of distributions yields the independent competing risks model for the function $\bar{G}$, i.e.,

$$\bar{G}(t) = \prod_{j=1}^{k} \exp\big(-\theta_0 q_j F_j(t)\big) \tag{5}$$

where θ_0 is the expected total number of viable clonogens of various types existing in the treated tumor. Within the framework of this model the hazard functions λ_j are additive and

$$\lambda(t) = \theta_0 \sum_{j=1}^{k} q_j f_j(t) \ .$$

In view of the last formula, it is not surprising that the bimodal shape of the hazard function arises when breast cancer recurrences originate from two distinct populations of progenitor cells, i.e., the fast and slowly evolving clonogens, as shown below.

21.3 Estimation of parameters

Within the scheme of right independent censoring [10] the likelihood for a random sample of size n is of the form

$$L = \prod_{j} g(t_j) \prod_{k} \bar{G}_k(s_k) \ , \tag{6}$$

where $g(t) = -\bar{G}'(t)$ is the corresponding probability density function, t_j for $j = 1, ..., m$ (m random) represent the observed failure times and s_k for $k = 1, ..., n-m$ are the censored observations.

If one selects a two-parameter family of distributions to approximate the function F in (2), then there will be only 3 parameters to be estimated from the time-to-recurrence observations, the estimation of which is feasible by the maximization of the likelihood function L . Because of its flexibility, we choose F to be a gamma distribution with the density

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}, \quad \alpha > 1, \quad t \geq 0 \ ,$$

where α and β are shape and scale parameters respectively. The mean $\tau = \alpha/\beta$ and the standard deviation $\sigma = \sqrt{\alpha}/\beta$ of the progression time can be computed. This progression time model reflects to some extent the multistage structure of the tumor development process.

Another reason for such a choice is that finite mixtures of gamma distributions are identifiable [23,26], and so are the competing risks models of type (5) when applied to the description of tumors arising from a heterogeneous population of clonogens.

To maximize the log-likelihood $\ell = \log L(\theta, \alpha, \beta)$ with respect to the unknown parameters θ, α, β we use the following 3-step procedure.

- Step 1: apply the random search algorithm [28] that requires the specification of a domain A containing the overall maximum but not a starting point for the optimization.

- Step 2: apply the Davidon-Fletcher-Powell algorithm [6] with the initial points provided by step 1. If the boundary of the set A is attained then go to step 3, otherwise step 2 gives the final solution.

- Step 3: apply the Zoutendijk algorithm [6] allowing for constraints which specify A.

In order to simplify the computations, we confine the search for the value of α that maximizes the log-likelihood ℓ to the set of positive integers, i.e., the problem $\max_{\theta,\alpha,\beta} \ell$ is replaced by

$$\max_{2\leq\alpha\leq\alpha_{\max}} \tilde{\ell} \quad \text{where} \quad \tilde{\ell} = \max_{\theta,\beta} \ell.$$

It suffices to take $\alpha_{\max} = 20$ to cover all reasonable values of the variation coefficient $1/\sqrt{\alpha}$, the smallest being equal to 0.2 .

The above-outlined numerical algorithm (see Appendix A for details) may readily be applied to the multicomponent model given by (5). Further insight into the properties of the estimates for finite samples is provided by statistical decision theory: It can be shown that the maximum likelihood estimator is close to a minimax estimator (see Appendix B).

21.4 Early and late recurrence in breast cancer

We apply model (5) to analyze data on breast cancer recurrence for 877 patients treated and followed at the Institut Curie from 1960 to 1988. Description of a sub-cohort is given by Fourquet *et al.* [5]. The data include the localization of the recurrences in terms of their occurence in the same quadrant as the primary tumor or elsewhere, the time to tumor recurrence and the censoring index value.

First consider the ipsilateral (treated) breast as a whole. We proceed from the independent censoring of the data caused by recurrences in the contralateral breast because there are grounds to believe that cancers in the two breasts develop independently of one another after the treatment [1,9]. Plots of the parametric estimate based on model (5) and the Kaplan-Meier disease-free curve are shown in Figure 21.1A. Figure 21.1B represents the parametric estimate and the Belayev kernel estimate [2] of the corresponding hazard function.

The model provides a good description of the data for $k = 2$, implying the existence of two competing populations of clonogens that give rise to the tumor recurrence. For k=1, the goodness of fit test developed recently by Hjort [7] rejects the null hypothesis at a significance level of less than .001. When we assume k=2, the significance level is aproximately .1, thereby indicating that the two competing risks model is consistent with the data.

The mean number of surviving clonogens and the progression time parameters for the two populations are presented in Table 21.1. The estimate of θ_2 is expected to be less precise than that of θ_1 because of the censoring effects, and this emerges in the corresponding asymptotic confidence intervals (Table 21.1). Note that both the life-table and the kernel estimates indicate the bimodal shape of the hazard function in this case.

It is generally accepted that a high rate of growth is indicative of the tumor aggressiveness, this being valid for early recurrences in all regions of the treated

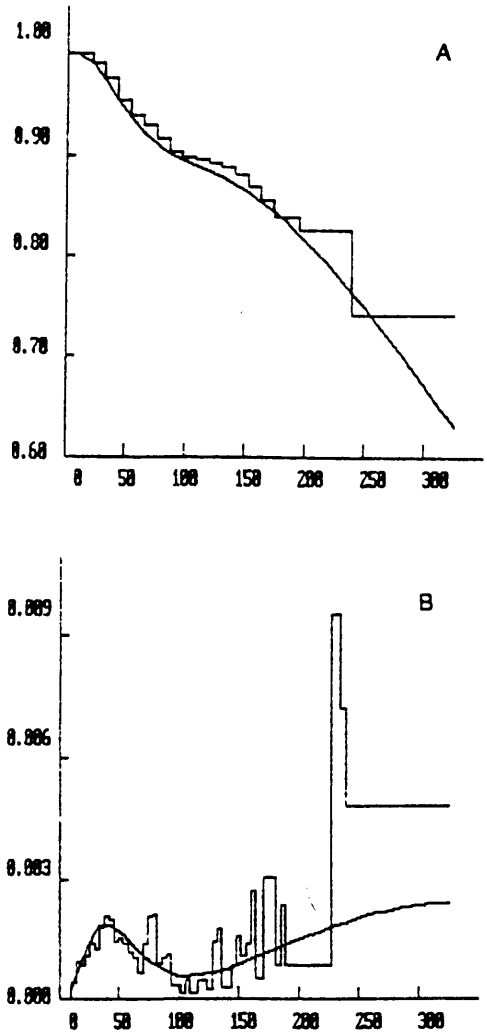

Figure 21.1: Disease-free and hazard curves for the ipsilateral breast cancer (A) Estimated disease-free curves. Solid line: parametric estimate, stepwise curve: Kaplan-Meier estimate. (B) Estimated hazard rates. Solid line: parametric estimate, stepwise curve: kernel estimate.

breast. The model makes it possible to relate early recurrences to those originating from the rapidly evolving subpopulation of clonogens.

It is questionable whether the data on quadrants may be considered in the same way. If there is a mutual dependence between the tumor cell clones evolving in the regions of the treated breast, the censoring mechanism can no longer be assumed independent. Should this be the case then both the nonparametric and parametric estimates would not be appropriate for estimating the corresponding marginal survival functions nor would the Hjort test be tenable for testing the goodness of fit.

To elucidate this point, we analyze the data on quadrants as if there were independent censoring of the time-to-tumor observations. Figures 21.2 and 21.3 exhibit the disease-free and hazard curves for the same and the other quadrants respectively. Again, we proceed from the two competing risks model to provide a better description of the data.

Parameters for different populations of clonogens responsible for tumor recurrence in all the quadrants of the ipsilateral breast are given in Table 21.2. The rapidly developing population manifests itself in the same and in the other quadrants, thereby indicating that there is no perfect correspondence between the temporal characteristics, i.e., the proliferative potential of clonogens, and their location

Table 21.1: Asymptotic likelihood inference on the ipsilateral breast cancer recurrence.

Parameter	Maximum Likelihood Estimate	Asymptotic .95 Confidence Interval
θ_1	0.11	0.08 , 0.14
α_1	4.00	3.52 , 4.48
β_1	0.076	0.064 , 0.088
θ_2	1.07	0.00 , 3.19
α_2	5.00	1.33 , 8.67
β_2	0.012	0.00 , 0.27

Table 21.2: Maximum likelihood estimates of the parameters for model of tumor recurrence. Legends. # clonogens: expected number of clonogens, time: mean progression time in months, std deviation: standard deviation of progression time

	# clonogens		time		std deviation	
localization	θ_1	θ_2	τ_1	τ_2	σ_1	σ_2
ipsilateral breast	0.11	1.07	53	431	26	193
same quadrant	0.07	5.17	59	1048	34	468
other quadrant	0.05	0.34	50	315	29	157

in the treated breast. Recall that the analysis of the pooled data for the ipsilateral breast also reveals such fast growing population of tumor cells.

At the same time, the progression time parameters of slowly evolving clonogens differ widely and the values of θ_2 for the two regions (see Tables 21.1 and 21.2) are too far from summing up to the value for the entire breast. This clearly demonstrates that our independence assumptions are likely to break down for this type of data.

When analyzing the contralateral breast cancer in a similar manner, one meets with the problem of discriminating between a true recurrence and a new cancer of the same histological type and localization. This problem will be addressed in a future paper.

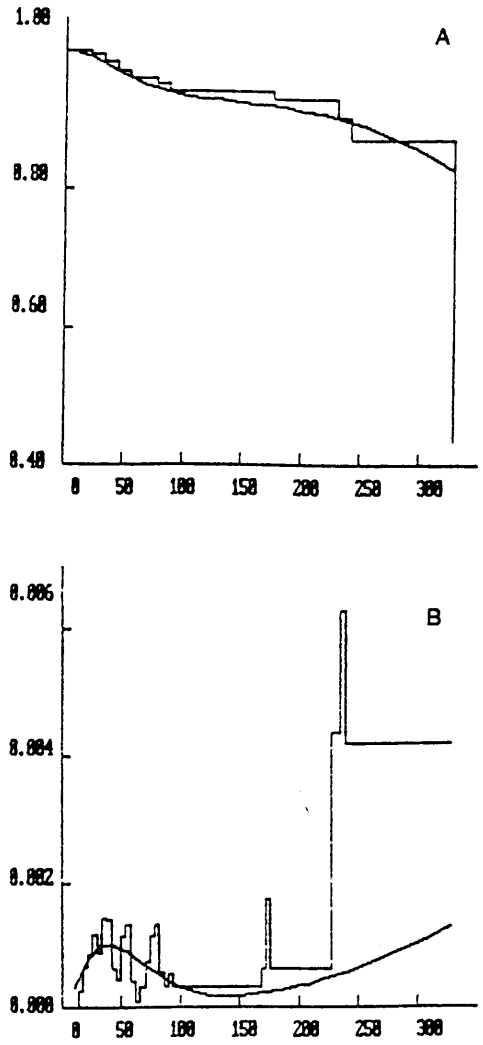

Figure 21.2: Disease-free and hazard curves for the same quadrant of the ipsilateral breast (A) Estimated disease-free curves. Solid line: parametric estimate, stepwise curve: Kaplan-Meier estimate. (B) Estimated hazard curves. Solid line: parametric estimate, stepwise curve: kernel estimate.

21.5 Appendix A. Estimation procedure

21.5.1 Random Search

The algorithm is based on Zigliavskij's theory of random search [28]. Let $\Xi = \{\xi_1, \xi_2, \ldots, \xi_n\}$ be a sample from the parametric space Ω and $\{\eta_1, \eta_2, \ldots, \eta_n\}$ the corresponding sample of function values where, for $i = 1, 2, ...,$ $\eta_i = l(\xi_i)$ are i.i.d. with distribution function $U(t)$. Statistical inference on the value of $\tilde{\ell} = \max_{\xi_i} \ell(\xi_i)$ is based on the $k+1$ maximal order statistics $\eta_{(n)}, \eta_{(n-1)}, \ldots, \eta_{(n-k)}$.

Denote $V(v) = 1 - U(\tilde{\ell} - 1/v)$. Assume $k = k(n)$, $\lim_{n\to\infty} k^2/n = 0$ and $\lim_{v\to\infty} V(tv)/V(v) = t^{-\alpha}$, for $v > 0$ and $0 < \alpha < \infty$. The latter assumption is typical in the theory of extreme values; It usually holds for not very exotic functions. The statistic $\lfloor \tilde{\ell} - \eta_{(n)} \rfloor / \lfloor \eta_{(n)} - \eta_{(n-k)} \rfloor$ converges in distribution to a random variable

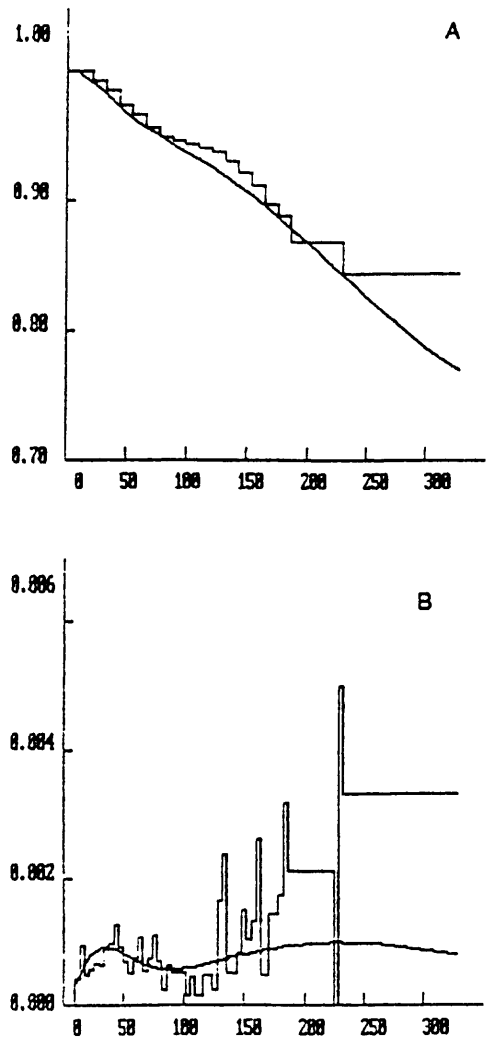

Figure 21.3: Disease-free and hazard curves for other quadrants of the ipsilateral breast. (A) Estimated disease-free curves: Solid line: parametric estimate, stepwise curve: Kaplan-Meier estimate. (B) Estimated hazard curves. Solid line: parametric estimate, stepwise curve: kernel estimate.

with distribution function $F_k(u) = 1 - \left(1 - \left(u/(1+u)\right)^{\alpha}\right)^k$, for $u \geq 0$. The asymptotic one–sided $(1-\gamma)$–confidence interval for $\tilde{\ell}$ is

$$\left[\eta_{(n)} \ , \ \eta_{(n)} + r(k,\gamma)\big(\eta_{(n)} - \eta_{(n-k)}\big)\right] \tag{A1}$$

where $r(k,\gamma) = \lfloor (1-\gamma)^{-1/\alpha} - 1 \rfloor^{-1}$. Its asymptotic length is equal to $(\tilde{\ell} - \theta_n)\phi(k,\gamma)$ where $\theta_n = \inf\{v : 1 - U(v) \leq 1/n\}$ and $\phi(k,\gamma) = r(k,\gamma)\,\lfloor \big(\Gamma(k+1+1/\alpha)/\Gamma(k+1/\alpha)\big) \rfloor$ with $\lim_{k\to\infty} \phi(k,\gamma) = (-\ln\gamma)^{1/\alpha}$. Note that for $k = 10$, the limit of ϕ is practically attained.

The unknown parameter α may be replaced by its consistent and asymptotically unbiased estimate $\hat{\alpha} = \ln(k/m)/\ln\lfloor \big(\eta_{(n)} - \eta_{(n-k)}\big)/\big(\eta_{(n)} - \eta_{(n-m)}\big) \rfloor$. For $\alpha > 1$ and $k \to \infty$ such that $k^2/n \to 0$ and $m/k \to \tau$, where $0 < \tau < 1$, $\lim_{k\to\infty} E(\hat{\alpha} - \alpha)^2 = \alpha^2(1-\tau)/(\tau \ln^2 \tau)$. Note that the latter function is minimal at $\tau \approx .2$, i.e., $m \approx k/5$.

Now, suppose we have a sample Ξ in some subset of the parametric space Ω. Let M_0 be a record of ℓ in the whole parameter space. To test the hypothesis that the subset contains the global maximum, we use (A1) and reject this hypothesis if

$$M_0 > \eta_{(n)} + r(k,\gamma)\big(\eta_{(n)} - \eta_{(n-k)}\big) \qquad (A2)$$

where the right hand side of inequality (A2) is computed for the subset. Again, for $n \approx 100$, the optimal value is almost attained with $k = 10$. If we have a nested sequence of subsets $\Omega \supset \Omega_1 \supset \Omega_2 \supset \ldots$ then the probability that the global maximum is not lost is at least $1 - \gamma$.

21.5.2 Zoutendijk algorithm

In the search for

$$\tilde{\ell} = \max_{\{\xi \in \Omega\}} \ell(\xi) \qquad (A3)$$

let ξ_k be the vector obtained at the k-th iteration. Consider the solution s_k to the maximization problem

$$\max_{\{\tilde{\xi}_k : \xi_k + \tilde{\xi}_k \in \Omega\}} \big[\ell(\xi_k) + \operatorname{grad}\big(\ell(\xi_k)\big)\,\tilde{\xi}_k\big] \qquad (A4)$$

The solution to the problem

$$\max_{\{\Delta : \xi_k + \Delta s_k \in \Omega\}} \ell(\xi_k) + \Delta s_k$$

gives the vector $\xi_{k+1} = \xi_k + \Delta s_k$ for the next iteration. This method is the best among optimization methods with constraints using the gradient vector or its approximation [6]. Although problem (A4) may prove to be as serious as problem (A3), it may be solved by the simplex method when Ω is a polygon as in this case.

Remark. Usually the global minimum of the log-likelihood function is located in the inner point of the parametric space where it is approximated much better by a quadratic function than by a linear one. The Fletcher-Powell algorithm approximates the function linearly at first but as the number of iterations grows the approximation becomes quadratic (it converges to Newton algorithm). With this algorithm calculations are much faster than with the gradient type algorithms and it does not require the second derivatives as well. For these reasons, we go through step 2 before step 3.

21.6 Appendix B. An optimal estimator for $G(t)$.

Consider the following statistical problem. Let U be a positive r.v. whose distribution function belongs to a parametric family $\{P(t/\sigma);\ \sigma > 0\}$ with σ denoting an unknown scale parameter. Let $t_1, ..., t_n$ be the sample of size n from this distribution. The problem is to estimate the survival function $Q(t/\sigma) = 1 - P(t/\sigma)$ from the sample $t_1, ..., t_n$. The conventional approach to this problem implies the construction of an appropriate estimator $\hat{\sigma}_n = \hat{\sigma}_n(t_1, ..., t_n)$ for the unknown parameter σ with subsequent use of the function $Q(t/\hat{\sigma}_n)$ as an estimator for $Q(t/\sigma)$. However, it seems more natural to search for a statistic $\hat{Q}(t) = \hat{Q}(t; t_1, ..., t_n)$ which estimates the survival function directly but not necessarily through $Q(t/\hat{\sigma}_n)$.

To characterize the losses caused by using $\hat{Q}(t)$ as an estimator for $Q(t/\sigma)$, introduce the loss function

$$L(\hat{Q}; \sigma) = \frac{1}{\sigma} \int_0^\infty \left(\hat{Q}(t) - Q(t/\sigma)\right)^2 dt \ .$$

The expectation

$$R_\sigma \hat{Q} = E_\sigma L(\hat{Q}; \sigma)$$

is called the risk of the estimator $\hat{Q}$. It is assumed that function $Q(t) = Q(t; t_1, ..., t_n)$ is measurable with respect to $t_1, ..., t_n$ for almost all t , and also measurable with respect to t for almost all $t_1, ..., t_n$.

A statistic $\hat{Q}(t)$ is called a proper estimator for the survival function $Q(t/\sigma)$ if for every $a > 0$

$$Q(at; at_1, ..., at_n) = Q(t; t_1, ..., t_n) \ .$$

It is easy to show that the risk R_σ of a proper estimator does not depend on the parameter σ . We are interested in finding an estimator that minimizes the risk R_σ over the class of proper estimators. Assume that the c.d.f. $P(t)$ is absolutely continuous and has a finite first moment, and let $p(t)$ stand for the density of P.

Theorem (see [12]). *Under the above assumptions,*

$$Q^*(t) = \frac{\int_0^\infty Q(tu) u^n \prod_{j=1}^n p(t_j u) du}{\int_0^\infty u^n \prod_{j=1}^n p(t_j u) du} \qquad (B1)$$

is optimal in the class of proper estimators of the survival function $Q(t/\sigma)$, $\sigma > 0$. It is minimax in the class of parametric estimators.

Let t_j for $j = 1, ..., m$ and s_k for $k = 1, ..., n - m$ represent the observed failure times and the censored observations respectively. A statistic $\hat{Q}_{ci}(t) = \hat{Q}_{ci}(t; t_{j_1}, ..., t_{j_m}; s_{k_1}, ..., s_{k_{n-m}})$ is called a proper estimator for the survival function $Q(t/\sigma)$, $\sigma > 0$, under independent right censorship, if for every $a > 0$

$$\hat{Q}_{ci}(at; at_{j_1}, ..., at_{j_m}; as_{k_1}, ..., as_{k_{n-m}}) = \hat{Q}_{ci}(t; t_{j_1}, ..., t_{j_m}; s_{k_1}, ..., s_{k_{n-m}}) \ .$$

The corresponding optimal estimator Q^*_{ci} can be obtained by using (6) instead of

$$Q^*_{ci}(t) = \frac{\int_0^\infty Q(tu)u^m \prod_j p(t_j u) \prod_k Q(s_k u)du}{\int_0^\infty u^m \prod_j p(t_j u) \prod_k Q(s_k u)du} . \qquad (B2)$$

Consider the survival function for the duration of tumor latency given by (2). Assume that $\theta = 1/\sigma$ is an unknown parameter and F a prescribed distribution function. Despite the fact that σ is not a scale parameter of the function $\bar{G}(t)$ one can introduce a new variable $\xi = F(t)$ and reduce the model to the one considered above. In the case of a censored sample, it follows from (B2) that the optimal estimator of $\bar{G}$ is

$$\bar{G}^*_{ci}(t) = \left(1 + \frac{F(t)}{\sum_j F(t_j) + J\sum_k F(s_k)}\right)^{-(m+1)} . \qquad (B3)$$

If the observed survival is high, one may expect that the estimator given i (B3) will be numerically close to the maximum likelihood estimator

$$\tilde{\bar{G}}(t) = \exp\left(-\frac{mF(t)}{\sum_j F(t_j) + \sum_k F(s_k)}\right) .$$

The above result gives one more reason to use the maximum likelihood method for estimating the parameters in models (2) and (5).

21.7 References

[1] Asselain, B., Fourquet, A., Hoang, T., Myasnikova, C. and Yakovlev, A.Yu. (1993). Testing the independence of competing risks: an application to the analysis of breast cancer recurrence. *Biometrical Journal*, 36, pp. 465–473.

[2] Belayev, Yu. K. (1987). *Statistical Methods of Incomplete Data Analysis in Reliability.* Znanie, Moscow (in Russian).

[3] Cox, D.R. and Oakes, D. (1983). *Analysis of Survival Data.* Chapman and Hall, London.

[4] Fisher, B., Anderson, S., Fisher E. *et al.* (1991). Significance of ipsilateral breast tumour recurrence after lumpectomy. *The Lancet* **338**, 327–331

[5] Fourquet, A., Campana, F., Zafrani, B. *et al.* (1989). Prognostic factors of breast recurrence in the conservative management of early breast cancer: A 25-year follow-up. *Int. J. Radiat. Oncol. Biol. Phys.* **17**, 719–725.

[6] Himmelblau, D.M. (1972). *Applied Nonlinear Programming.* McGraw-Hill Book Company, Austin, Texas.

[7] Hjort, N. (1990). Goodness of fit tests in models for life history based on cumulative hazard rates. *Ann. Statist.* **18**, 1221–1258.

[8] Hoel, D.G. (1972). A representation of mortality data by competing risks. *Biometrics* **28**, 475–488 .

[9] Holmberg, L., Adami, H.O., Elbom, A., Bergstræm, A., and Lindgren, A. (1988). Prognosis in bilateral breast cancer. Effects of time interval between first and second primary tumours. *Journal of Cancer* **58**, 191–194.

[10] Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons, New York.

[11] Klebanov, L.B., Rachev, S.T. and Yakovlev, A.Yu. (1992). A stochastic model of radiation carcinogenesis: latent time distributions and their properties. *Mathematical Biosciences,* **113**, 51–75.

[12] Klebanov, L.B., Rachev, S.T., and Yakovlev, A.Yu. (1992). On the parametric estimation of survival functions. Submitted to *Statistics and Decisions.*

[13] Kurtz, J.M., Amalric, R., Brandone, H. *et al.* (1990). The prognostic significance of late local recurrence after breast conserving therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **18**, 87–93.

[14] Laurence, A.F. and Morgan, B.J.T. (1989). Observations on a stochastic model for quantal assay data. *Biometrics* **45**, 733–744.

[15] Lonning, P.E. (1991). Treatment of early breast cancer with conservation of the breast - a review. *Acta Oncologica* **30** (7), 779–792.

[16] Miller, R.G. (1981). *Survival Analysis.* John Wiley, New York.

[17] Morgan, B.J.T., North, P.M., and Pack, S.E. (1986). Collaboration between university and industry. *The Statistical Consultant in Action.* Hand, D.J. and Everitt, B.S., Eds. Cambridge University Press, Cambridge, 134–151.

[18] Munro, T.R. and Gilbert, C.W. (1961). The relation between lethal doses and radiosensitivity of tumor cells. *Br. J. Radiol* **34**, 246–251.

[19] Porter, E.H. (1980). The statistics of dose/cure relationships for irradiated tumours Part I. *Br. J. Radiol.* **53**, 210–227.

[20] Porter, E.H. (1980). The statistics of dose/cure relationships for irradiated tumours Part II. *Br. J. Radiol.* **53**, 336–345.

[21] Suit, H.D., Sedlacek, R., Fagundes, L., Goitein, M., and Rothman, K.J. (1978) Time distributions of recurrences of immunogenic and nonimmunogenic tumors following local irradiation. *Radiat. Res.* **73**, 251–266.

[22] Suit, H.D., Shalek, R.J., and Wette, R. (1965). Radiation response of C3H mammary carcinoma. *Cellular Radiation Biology.* Williams & Wilkins, Baltimore, 514–530.

[23] Teicher, H. (1961). Identifiability of finite mixtures. *Ann. Math. Statist.* **32**, 244–248.

[24] Tucker, S.L., Thames, H.D., and Taylor, J.M.G. (1990). How well is the probability of tumor cure after fractionated irradiation described by Poisson Statistics? *Radiat. Res.* **124**, 273–282.

[25] Yakovlev A.Yu. and Zorin A.V. (1988). *Computer Simulation in Cell Radiobiology.* Lecture Notes in Biomathematics **74**, Springer Verlag, Berlin.

[26] Yakowitz, S.J. and Spragins, J.D. (1968). On the identifiability of finite mixtures. *Ann. Math. Statist.* **39**, 209–214.

[27] Yamaguchi, K. (1992). Accelerated failure-time regression models with a regression model of surviving fraction: an application to the analysis of permanent employment in Japan. *J. Am. Stat. Assoc.* **87** (418), 284–292.

[28] Zhigljavsky, A. (1992). *Theory of Global Random Search.* Kluwer, Dordrecht.