Nonparametric and Dual Likelihood in Survival Analysis

by

Per Aslak Mykland[1]

TECHNICAL REPORT NO. 395

Department of Statistics
The University of Chicago
Chicago, Illinois 60637

September 1994

[1] Research supported in part by National Science Foundation grant DMS 93-05601.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292 and DMS 87-03942 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

Some key words and phrases: Ancillarity, Bartlett correction, Discrete time calculus.

Nonparametric and Dual Likelihood in Survival Analysis

by

Per Aslak Mykland[1]
The University of Chicago

Summary

Recent results have shown that one can often construct dual likelihoods to analyze nonparametric likelihood ratio statistics. We will investigate a number of important statistics in survival analysis from this perspective. Findings include results on Bartlett correctibility and ancillarity. In part of the discussion, we use a discrete time calculus for survival analysis, which we believe can be useful also in other contexts.

1 Research supported in part by National Science Foundation grant DMS 93-05601.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292 and DMS 87-03942 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

Some key words and phrases: Ancillarity, Bartlett correction, Discrete time calculus.

1. Introduction

One of the more interesting findings about Dual Likelihood is that there are some instances where the dual likelihood ratio statistic coincides with the likelihood ratio statistic in a nonparametric likelihood. This is the case for empirical likelihood with estimating equations and for point process type likelihoods when estimating the cumulative hazard (see Mykland [1993, Section @@@]).

The importance of such a coincidence, when it occurs, is that it implies that the nonparametric LR statistic automatically retains a number of key properties of the parametric LR statistic, notably Bartlett correctibility, and similarly for the signed square root of the LR statistic.

The purpose of this paper is to further explore how this connection works in the case of the point process likelihoods that can be found in survival analysis. We begin by showing that the dual LR statistic in Aalen's regression (Aalen (1980, 1989)) corresponds to a nonparametric LR statistic (Section 2). The results in Section @@@ of Mykland (1993) is a special case of this. Another special case is the comparison of two survival distributions at a point, as well as the comparison of two quantiles (Section 3).

A question of some importance is what happens when there are additional parameters. As is clear from sections 2-3, the infinite dimensional nuisance parameter representeds by the cumulative hazard does not cause any problem, but other nuisance parameters might. In Section 4, we look at the one sample problem with Cox covariates. It will there turn out that the nonparametric likelihood does not have a dual, and one cannot expect highter order properties to hold. The first order χ^2 approximation remains valid, however.

What about other properties of likelihood? An important result in parametric inference, for example, is that the LR statistic is locally sufficient up to and including $O(n^{-1})$ (McCullagh (1984)). This remains, obviously, broadly true for dual likelihood, if one considers the alternatives implied by that likelihood (Mykland [1993, Section @@@]). In general, however, it is not clear that this would be the case when considering the alternatives in the (original) nonparametric likelihood. For the one sample survival analysis problem we discuss this question in Section 7, and show that, at least in this instance, the LR statistic is locally sufficient to the relevant order.

The conversion of nonparametric likelihood problems into dual parametric ones is a useful tool for analyzing nonparametric likelihood ratio (NPLR) statistics, providing results, for example, on correction factors and ancillarity.

This is particularly so in survival analysis, and it is the purpose of this paper to discuss some important test statistics in this area from the point of view provided by this technique. We shall mostly consider the comparison of two survival curves (Sections 2-3), but also the one-sample problem with nuisance parameters (Section 4). Cox regression will also be seen to be a type of dual likelihood (Section 5).

In connection with the latter, we discuss in Section 5 a discrete time calculus for survival analysis, which we think can be quite generally useful. An example is given of a variance calculation using this method, and then we apply the method to Cox regression.

Finally, we use dual likelihood to investigate ancillarity in Section 6 and in the Appendix.

The use of dual likelihood in survival analysis has earlier been studied by Mykland (1993) in the context of estimating the survival curve of a single population (with no nuisance parameter), and also in connection with Aalen's regression model (Aalen (1980, 1989)). Also, it is known that logrank statistics can be turned into LR statistics in a Cox model, see Example VII.2.4 (p. 487–488) in Andersen, Borgan, Gill and Keiding (1993). This relates to the construction in Section 5.

2. Comparing two survival distributions at a point

Suppose the cumulative hazards of two populations are given by $\Lambda_1(t)$ and $\Lambda_2(t)$, that numbers at risk are, respectively, $Y_1(t)$ and $Y_2(t)$, and let failures be denoted by $S_1, S_2, \ldots$ and $T_1, T_2, \ldots$. It is easy to see (Andersen, Borgan, Gill and Keiding (1993, Section II.7)) that the nonparametric log likelihood is given by

$$\ell_N(\Lambda_1, \Lambda_2) = \sum_{S_i \le t} \log \Lambda_1\{S_i\} + \sum_{T_i \le t} \log \Lambda_2\{T_i\} - \int_0^t Y_1(s) d\Lambda_1(s) - \int_0^t Y_2(s) d\Lambda_2(s) + C \tag{2.1}$$

where C is random but a function of the data only. Note that one arrives at (2.1) by first assuming that the Λ's are continuous, and then maximizing slightly to obtain that they are, in fact, discrete.

(This is a standard procedure; in the one-sample problem the MLE becomes the Nelson-Aalen estimator.)

Suppose one wishes to test the null hypothesis of equality at a point, *i.e.*, $\Lambda_1(t) = \Lambda_2(t)$, for fixed t. The NPLR statistic is then given by

$$\tfrac{1}{2} W_{NP} = \max \ell_N(\Lambda_1, \Lambda_2) - \max_{\Lambda_1(t)=\Lambda_2(t)} \ell_N(\Lambda_1, \Lambda_2). \tag{2.2}$$

To actually implement this, one would use the Lagrangian

$$\ell_L(\Lambda_1, \Lambda_2, \mu) = \ell_N(\Lambda_1, \Lambda_2) - \mu(\Lambda_1(t) - \Lambda_2(t)). \tag{2.3}$$

In the spirit of dual likelihood, however, we maximize instead (2.3) subject to fixed μ. This gives rise (by simple differentiation) to maximizers

$$\hat{\Lambda}_{1,\mu}\{S_i\} = (Y_1(S_i) + \mu)^{-1} \tag{2.4}$$

and

$$\hat{\Lambda}_{2,\mu}\{T_i\} = (Y_2(T_i) - \mu)^{-1}. \tag{2.5}$$

$\mu = 0$ corresponds to unrestricted maximization.

Substituting (2.4)–(2.5) into (2.3) gives

$$\ell_L(\hat{\Lambda}_{1,0}\hat{\Lambda}_{2,0}, 0) - \ell_L(\hat{\Lambda}_{1,\mu}, \hat{\Lambda}_{2,\mu}, \mu) = \ell_{\mathrm{DUAL}}(\mu), \tag{2.6}$$

where,

$$\ell_{\mathrm{DUAL}}(\mu) = \sum_{S_i \le t} \log\left(1 + \frac{\mu}{Y_1(S_i)}\right) + \sum_{T_1 \le t} \log\left(1 - \frac{\mu}{Y_2(T_i)}\right). \tag{2.7}$$

The crucial facts are now the following. The right-hand side of (2.6) is a log likelihood, in fact, the dual log likelihood corresponding to the score martingale

$$\sum_{S_i \le t} \frac{1}{Y_1(S_i)} - \sum_{T_i \le t} \frac{1}{Y_2(T_i)} - (\Lambda_1(t) - \Lambda_2(t)) \tag{2.8}$$

(the term involving the cumulative hazards vanish under H_0 at time t).

Also, since

$$\frac{\partial}{\partial\mu}\ell_{\mathrm{DUAL}}(\mu) = \sum_{S_i \le t} \frac{1}{Y_1(S_i)+\mu} - \sum_{T_i \le t} \frac{1}{Y_2(T_i)-\mu}$$
$$= \hat{\Lambda}_{1,\mu}(t) - \hat{\Lambda}_{2,\mu}(t) \tag{2.9}$$

and since $\ell_{\mathrm{DUAL}}(\mu)$ is clearly concave, it follows that the maximizer $\hat{\mu}$ of ℓ_{DUAL} is such that the constraint in the null hypothesis is satisfied. Hence

$$\tfrac{1}{2}W_{NP} = \max_{\mu} \ell_{\mathrm{DUAL}}(\mu), \tag{2.10}$$

and so the NPLR inherits the properties of a one-dimensional parameter LR statistic. In particular, under weak regularity conditions, W_{NP} is asymptotically χ^2_1, and it can be Bartlett corrected to be $\chi^2_1 + O(n^{-3/2})$. The same goes for the signed square root of W_{NP}. See McCullagh (1987, Ch. 7) for details.

Since,

$$\frac{\partial^p}{\partial\mu^p}\,\ell_{\mathrm{DUAL}}(0) = (p-1)!\left((-1)^{p-1}\sum_{S_i \le t}\frac{1}{Y_1(S_i)^p} - \sum_{T_i \le t}\frac{1}{Y_2(T_i)^p}\right),$$

the Bartlett correction factor can easily be worked out with the help of formula (7.13) (p. 212) in McCullagh (1987) and the methods used in Mykland and Ye (1992).

3. Comparison of two quantiles

An often relevant question is to compare the medians or other quantiles of the two survival distributions. One approach to this would be to use an estimated (rather than fixed) t in the previous section, *i.e.*, a $\hat{t}$ for which $\Lambda(\hat{t}) \approx -\log(q)$ when looking for the q'th quantile. One way to achieve this is to set

$$\hat{t} = \inf_{s} \hat{\Lambda}(s) \ge -\log(q), \tag{3.1}$$

where $\hat{\Lambda}$ is the joint Nelson-Aalen estimator for the two populations. The arguments in the previous section (apart from the actual computation of the Barlett correction) go through so long as $\hat{t}$ is a stopping time (see p. 4 of Jacod and Shiryaev (1987)), since the martingale property is preserved under stopping. This is true for (3.1), and it will also be true for a number of other constructions.

Note that, under the null hypothesis, $\hat{t}$ is consistent by the consistency of $\hat{\Lambda}$ and the continuity of Λ_1 and Λ_2.

4. A case of nuisance parameters

Directly obtaining a likelihood from maximizing a Lagrangian does by no means always work. The exercise nonetheless provides a useful starting point for further analysis. Nuisance parameters in empirical likelihood have been investigated from this angle by Qin & Lawless (1984). We do a somewhat similar analysis here for the case of estimating a baseline hazard in Cox regression, the regression parameters being seen as the nuisance parameters.

It should be emphasized that both Qin & Lawless (1994) and the current section only give rise to a χ^2 asymptotic distribution. Higher order likelihood properties are unlikely to hold, as suggested by Lazar and Mykland (1994)'s work on empirical likelihood.

Suppose that we have n patients, and that patient no. i has hazard rate on the proportional form

$$\lambda_i(\beta, t) = \lambda_0(t) r_i(\beta, t), \tag{4.1}$$

where, for example (Cox (1972, 1975), see also Andersen, Borgan, Gill and Keiding (1993)),

$$r_i(\beta, t) = Y_i(t) \exp\left(\sum_{\alpha} \beta^{\alpha} Z_{\alpha,i}(t) \right),$$

$Y_i(t)$ being the indicator of patient $\#i$ being under observation, and $Z_{\alpha,i}$ being predictable covariates. The β's are the regression parameters.

We shall in the following assume that interest is focused on inference for the baseline cumulative hazard at time t,

$$\Lambda_0(t) = \int_0^t \lambda_0(s) ds. \tag{4.2}$$

In analogy to (2.1), the nonparametric log likelihood is given by

$$\begin{aligned} \ell_N(\Lambda_0, \beta) = & \sum_{T_i \le t} \log(r_i(\beta, T_i)\Lambda_0\{T_i\}) \\ & - \int_0^t \left(\sum r_i(\beta, s) d\Lambda_0(s) \right) + C \end{aligned} \tag{4.3}$$

where T_i are the death times that are observed, and C is random but a function of the data only. This can be decomposed

$$\ell_N(\Lambda_0, \beta) = \ell_{\text{COX}}(\beta) + \ell_{\text{REST}}(\Lambda_0, \beta) + C, \tag{4.4}$$

where $\ell_{\text{COX}}(\beta)$ is the usual log proportional hazards likelihood (Andersen, Borgan, Gill and Keiding (1993), formula (7.2.7), p. 483),

$$\ell_{\text{COX}}(\beta) = \sum_{T_i \le t} \ln r_i(\beta, T_i) - \sum_{T_i \le t} \ln \overline{r}(\beta, T_i)$$

$(\overline{r}(\beta, t) = \sum_{i=1}^{n} r_i(\beta, t))$, and

$$\ell_{\text{REST}}(\Lambda_0, \beta) = \sum_{T_i \le t} \ln \Lambda_0\{T_i\} + \sum_{T_i \le t} \ln \overline{r}(\beta, T_i) - \sum_{T_i \le t} \overline{r}(\beta, T_i)\Lambda_0\{T_i\}.$$

Now suppose one wants to test the null hypothesis $\Lambda_0(t) = \theta$. The natural nonparametric likelihood ratio statistic is then

$$W_{NP} = 2(\max \ell_N(\Lambda_0, \beta) - \max_{\Lambda_0(t)=\theta} \ell_N(\Lambda_0, \beta)). \tag{4.5}$$

To actually implement this, one would use the Lagrangian

$$\ell_L(\Lambda_0, \beta_0, \mu) = \ell_N(\Lambda_0, \beta) - \mu(\Lambda_0(t) - \theta). \tag{4.6}$$

As in Section 2, we maximize (4.6) subject to fixed μ (rather than fixed θ). We also wish to retain the β's, and so define $\hat{\Lambda}_{\beta,\mu}$ as the *arg max* of (4.6) subject to fixed β and μ (the result, obviously, does not depend on θ). Note that $\hat{\Lambda}_{\beta,0}$ is the unconstrained maximizer.

By simple differentiation, the maximization gives $\Lambda_{\beta,\mu}\{T_i\} = (\overline{r}(\beta, T_i) + \mu)^{-1}$, whence

$$\begin{aligned}\ell_L(\hat{\Lambda}_{\beta,\mu}, \beta, \mu) &= \ell_{\text{COX}}(\beta) - \sum_{T_i \le t} \ln\left(1 + \frac{\mu}{\overline{r}(\beta, T_i)}\right) - \sum_{T_i \le t} \frac{\overline{r}(\beta, T_i)}{\overline{r}(\beta, T_i) + \mu} - \mu \sum_{T_i \le t} \frac{1}{\overline{r}(\beta, T_i) + \mu} + \theta\mu \\ &= \ell_{\text{COX}}(\beta) - \ell_{\text{DUAL}}(\beta, \mu) - \#(T_i \le t),\end{aligned} \tag{4.7}$$

where

$$\ell_{\text{DUAL}}(\beta,\mu) = -\theta\mu + \sum_{T_i \le t} \ln\left(1 + \frac{\mu}{\overline{r}(\beta, T_i)}\right) \tag{4.8}$$

Note that (4.8) is the dual log likelihood for testing $\Lambda_0(t) = \theta$ if the β's are known (cf. Mykland (1993)).

Similarly to Section 2, since

$$\begin{aligned}\frac{\partial}{\partial\mu}\ell_{\text{DUAL}}(\beta,\mu) &= -\theta + \sum_{T_i \le t} \frac{1}{\mu + \overline{r}(\beta, T_i)} \\ &= -\theta + \hat{\Lambda}_{\beta,\mu} \end{aligned} \tag{4.9}$$

and since ℓ_{DUAL} is strictly concave in μ for given β,

$$\max_{\beta,\mu} \ell_L(\hat{\Lambda}_{\beta,\mu}, \beta, \mu) = \max_{\Lambda_0(t)=\theta,\beta} \ell_N(\Lambda_0, \beta). \tag{4.10}$$

Also, obviously, $\max \ell_L(\hat{\Lambda}_{\beta,0}, \beta, 0) = \max \ell_N(\Lambda_0, \beta)$, hence if W_{NP} is the nonparametric log likelihood ratio (LR) statistic for testing $\Lambda_0(t) = \theta$, it satisfies

$$\tfrac{1}{2}W_{NP} = \max_{\beta}\Big(\ell_{\text{COX}}(\beta) - \ell_{\text{DUAL}}(\beta, 0)\Big) - \max_{\beta}\Big(\ell_{\text{COX}}(\beta) - \max_{\mu} \ell_{\text{DUAL}}(\beta,\mu)\Big). \tag{4.11}$$

(Keep in mind that $\ell_{\text{DUAL}}(\beta, 0) = 0$.)

The quantity $\ell(\beta,\mu) = \ell_{\text{COX}}(\beta) - \ell_{\text{DUAL}}(\beta,\mu)$ is not a log likelihood; it is easy to verify that (at $\mu = 0$)

$$E\dot{\ell}_{\beta_\alpha}\dot{\ell}_{\mu} = 0 \neq -E\ddot{\ell}_{\beta_\alpha,\mu} \tag{4.12}$$

$$E\dot{\ell}_{\mu}\dot{\ell}_{\mu} = E\ddot{\ell}_{\mu\mu} \tag{4.13}$$

and

$$E\dot{\ell}_{\beta_\alpha}\dot{\ell}_{\beta_\gamma} = -E\ddot{\ell}_{\beta_\alpha,\beta_\gamma} \tag{4.14}$$

where only the last identity is consistent with likelihoodness (this is the Cox regression part). Note that

$$E\ddot{\ell}_{\beta_\alpha,\mu} = E\int_0^t \frac{\overline{r}_{\beta_\alpha}(s)}{\overline{r}^2(s)}\, dN(s) \tag{4.15}$$

which in the vast majority of instances is nonzero.

To see what happens asymptotically, we let Z_1 be the limit in law of $\dot{\ell}_\mu/\sqrt{n}$, Z_α the limit in law of $\dot{\ell}_{\beta_\alpha}/\sqrt{n}$, $\kappa_{r,s} =$ limit of $\mathrm{Cov}(\dot{\ell}_r, \dot{\ell}_s)/n$ and $\kappa_{rs} =$ limit of $E\ddot{\ell}_{rs}/n$. We assume that the relevant limits exist, and that the Z_r are jointly Gaussian. It is then easy to see that

$$\begin{aligned} W_{NP} &\xrightarrow{\mathcal{L}} Z_\alpha \kappa^{\alpha,\beta} Z_\beta + Z_r \kappa^{rs} Z_s \\ &= Z^r \kappa_{r\alpha} \kappa^{\alpha,\beta} \kappa_{\beta s} Z^s + Z^r \kappa_{rs} Z^s \\ &= (Z^1)^2 \kappa_{1r} \kappa^{r,s} \kappa_{s1} \end{aligned} \tag{4.16}$$

where $\kappa^{\alpha,\beta}$, $\kappa^{r,s}$ and κ^{rs} are the inverses of $\kappa_{\alpha,\beta}$, $\kappa_{r,s}$ and κ_{rs}, respectively, and where $Z^r = \kappa^{rs} Z_s$. On the other hand,

$$\mathrm{Var}(Z^1) = \kappa^{1r} \kappa_{r,s} \kappa^{s1}, \tag{4.17}$$

so that W_{NP} converges in law to $c\chi^2_1$ where $c = \kappa_{1r}\kappa^{r,s}\kappa_{s1}/\kappa^{1r}\kappa_{r,s}\kappa^{s1}$. Since

$$\begin{aligned} \kappa_{\alpha r} \kappa^{r,s} \kappa_{s1} &= -\kappa_{\alpha,\beta} \kappa^{\beta,\gamma} \kappa_{\gamma 1} + \kappa_{\alpha 1} \kappa^{1,1} \kappa_{11} \\ &= -\kappa_{\alpha 1} + \kappa_{\alpha 1} = 0, \end{aligned} \tag{4.18}$$

it follows that the matrix $\kappa_{ar}\kappa^{r,s}\kappa_{sb}$ is block diagonal, and hence $c = 1$. Hence

$$W_{NP} \xrightarrow{\mathcal{L}} \chi^2_1. \tag{4.19}$$

5. The Partial History of the Process

There are two types of filtrations occurring in survival analysis. There is the standard, "counting process", full filtration $(\mathcal{F}_t)$ — the history of the process. There is also a discrete time filtration, however, given by $G_0, G_{\frac{1}{2}}, G_1, G_{\frac{3}{2}}, \ldots = \mathcal{F}_0, \mathcal{F}_{T_1-}, \mathcal{F}_{T_1}, \mathcal{F}_{T_2-}$, which we could call a discrete time filtration or the *partial history of the process.* A fair amount of work has been done with this filtration, mostly implicitly, notably by Cox (1972, 1975), Næs (1982) and Wong (1986). The approach is often thought of as "heuristic", but it is perfectly rigorous, as seen in Theorem II.7.1 (p. 96) in Andersen, Borgan, Gill and Keiding (1993), the results going back to Jacod (1975).

The partial history approach has several advantages: it permits the treatment of ties, it makes more processes predictable, and one can use discrete time limit theorems to show asymptotic results.

The approach is useful when comparing populations or estimating covariates. For example, in the setting of Section 2, if $N_{i,t}$ is the number of observed failures up to time t, then the compensator in the "partial" G-filtration is given by

$$d\langle N_i \rangle_t = \frac{Y_i(t)}{Y_1(t) + Y_2(t)}\, d(N_1 + N_2). \tag{5.1}$$

Our purpose in introducing this filtration is twofold. On the one hand, it gives rise to a useful calculus which is different from the standard point process calculus. In addition, it ties in with our previous discussion of dual likelihoods, as it will represent Cox regression as a kind of dual likelihood.

We begin by illustrating the use of this calculus on a variance calculation, and we then go on to the likelihood applications.

The variance of logrank statistics. Consider the test statistic $Z_k(t)$ given in formula (5.22) in Andersen, Borgan, Gill and Keiding [(1993), p. 346]. Obviously,

$$\langle Z_k \rangle_t = 0.$$

Furthermore (still using $\langle , \rangle$ to refer to the partial filtration)

$$\begin{aligned}
\Delta\langle Z_k, Z_j \rangle_T &= K(T)^2 \operatorname{Cov}(Z_k, Z_j \mid \mathcal{F}_{T-}) \\
&= K(T)^2 \operatorname{Cov}(N_k, N_j \mid \mathcal{F}_{T-}) \\
&= K(T)^2 \left(-\frac{Y_k(T) Y_j(T)}{Y.(T)^2} + \delta_{kj}\, \frac{Y_k(T)}{Y.(T)} \right) \Delta N.(T),
\end{aligned}$$

i.e.,

$$\langle Z_k, Z_j \rangle_t = \int_0^t K(t)^2 \left(-\frac{Y_k(s) Y_j(s)}{Y.(s)^2} + \delta_{kj}\, \frac{Y_k(s)}{Y.(s)} \right) dN.(s).$$

This is the same formula as (5.25) on p. 347 of Andersen, Borgan, Gill and Keiding, but derived very differently. There, it is an estimated expected quantity, here it is an observed quantity. In fact, if one creates a dual Cox-type likelihood (see below), it is the observed information under the null hypothesis.

Finally, it is interesting to ponder that the above estimate of covariance seems to deal perfectly well with ties.

One should compare what happens in the full information (counting process) theory, cf. Example V.22 (p. 352–354) in Andersen, Borgan, Gill and Keiding (1983). The estimate of variance (formula (5.2.10)) is quite different from the formula given here. It should be noted that this is due to the fact that the model in their example is different from the conditional hypergeometric model used in our derivations.

Cox Regression as a Dual Likelihood. With the above framework in mind, consider doing inference with the function

$$d = \sum_{T_i \leq t} \log(r_i(\beta, T_i)),$$

which in standard Cox regression would be

$$d = \sum_{T_i \leq t} \beta^\alpha Z_{\alpha,i}(T_i).$$

Suppose one wants to turn this into a likelihood. One way, cf. Mykland (1993, 1994), would be to construct

$$\Delta \log \text{ likelihood}(T) = d_T - \kappa(d_T \mid \mathcal{F}_{T-})$$

where κ is the conditional cumulant generating function. It is easy to see that

$$E(\exp(\Delta \log \text{ likelihood}(T)) \mid \mathcal{F}_{T-}) = 1.$$

The construction above is clearly the same as Cox regression, since $\overline{r}(\beta, T)$ is $\mathcal{F}_{T-}$-measurable, and since the Cox partial likelihood integrates to 1. Hence, in this case, Cox regression coincides with a dual-type likelihood.

An important consequence of this is that the well-known possibility of turning logrank statistics into LR statistics via the Cox likelihood (see, e.g., Example VII.2.4 (p. 487–488) in Andersen, Borgan, Gill and Keiding) is also consistent with the dual likelihood paradigm.

6. Ancillaries and dual likelihood

It is well-known in the parametric context that the Bartlett correction factor is stable with respect to second (and higher) order ancillaries, whence the LR statistic is independent of such ancillaries up to and including $O(n^{-1})$. For a discussion of this, see Section 8.4 (p. 236–238) of McCullagh (1987).

This result does not directly carry over to our current context, as the independence relates to ancillaries in the dual model, whereas we would really be interested in ancillaries in the primal model. It turns out, however, that the result still holds, and that the connection is quite straightforward.

In the presence of no nuisance parameters, an asymptotically normal quantity A is second order ancillary in the dual model if it is second order ancillary in the primal model. This holds both in the one-sample case (see Mykland (1993)) and in the two-sample case (Sections 2–3).

Hence the Bartlett factor will be stable with respect to second order ancillaries not only in the dual model, but also in the original nonparametric model. In addition, it follows that W_{NP} is independent of second order ancillaries (in both the primal and dual models) up to and including order $O(n^{-1})$.

The passage from nonparametric to dual ancillarity is derived in the Appendix in the one-sample case. The two-sample case is analogous.

The case treated in Section 4 can also to some extent be treated the same way, though here we only do a first order analysis. First order ancillarity in the nonparametric model means asymptotic uncorrelatedness with the Cox regression score martingales and with the martingale $M_t = N_t - \int_0^t \overline{r}(\beta, s) d\Lambda_0(s)$ (for the same reasons as those discussed in the Appendix).

In particular, it means uncorrelatedness with $\hat{\mu}$, *i.e.*, with Z^1 from (4.16). Hence, to first order, W_{NP} from (4.5) is independent of all first order ancillaries.

Acknowledgements

The author is grateful to Peter McCullagh for useful discussions, and Mitzi Nakatsuka for typing the manuscript. Thanks also to Andersen, Borgan, Gill and Keiding, whose book has been invaluable to me. This research was supported in part by National Science Foundation grant DMS 93-05601.

Appendix

We here carry out the promise from Section 7 to show that if A^r is second order ancillary in the primal (nonparametric) model, then it also is second order ancillary in the dual model.

Following Cox (1980) and McCullagh (1984, 1987, Section 8.3), an asymptotically normal vector A^r of order $O_p(n^{1/2})$ is second order ancillary in a likelihood model if, for all r, s

$$\mathrm{Cov}(A^r, \dot{\ell}) = o(n^{1/2}) \tag{A.1}$$

$$\mathrm{Cov}(A^r, \ddot{\ell}) + \mathrm{cum}(A^r, \dot{\ell}, \dot{\ell}) = o(n) \tag{A.2}$$

$$\mathrm{cum}(A^r, A^s, \dot{\ell}) = o(n). \tag{A.3}$$

Note that in the original definition, the orders were $O(1)$, $O(n^{1/2})$ and $O(n^{1/2})$, respectively, but (A.1)–(A.3) captures the same idea and is easier to work with for this purpose.

As mentioned in Section 7, we concentrate here on the one-sample case, so N_t is the observed number of deaths, Λ_t is the cumulative hazard, Y_t is the number under observation, and $M_t = N_t - \int_0^t Y_s d\Lambda_s$ is a martingale.

As contiguous alternatives in the nonparametric model, we consider cumulative hazards $\Lambda + \delta n^{-1/2}\Gamma$. It is easy to see from Corollary II.7.3 (p. 98) of Andersen, Borgan, Gill and Keiding (1993) that the log likelihood ratio between $\Lambda + \delta n^{-1/2}\Gamma$ and Λ is

$$\ell_{NP}(\delta) = \delta n^{-1/2} V_t - \tfrac{1}{2}\delta^2 n^{-1}[V, V]_t + \tfrac{1}{3}\delta^3 n^{-3/2}[V, V, V]_t + \dots \tag{A.4}$$

where V_t is any martingale on the form

$$V_t = \int_0^t g(s) dM_s. \tag{A.5}$$

This is obtained by setting $g(s) = d\Gamma(s)/d\Lambda(s)$. g can, of course, depend on n.

Our goal is to show that if A^r satisfies (A.1)–(A.3) under log likelihoods (A.4), *i.e.*, in the primal model, it must also satisfy (A.1)–(A.3) in the dual model, given by

$$\ell_{\mathrm{DUAL}}(\delta) = \delta n^{-1/2} U_t - \tfrac{1}{2}\delta^2 n^{-1}[U, U]_t + \tfrac{1}{3}\delta^3 n^{-3/2}[U, U, U]_t + \dots, \tag{A.6}$$

where

$$U_t = \int_0^t \left(\frac{Y_s}{n}\right)^{-1} dM_s. \tag{A.7}$$

This shows that nonparametric second order ancillarity implies dual second order ancillarity.

It should be noted that if g in (A.5) was not restricted to be nonrandom, the result would be more or less automatic.

To show the result, assume that A^r is second order ancillary in the nonparametric model. Define the martingales

$$A^r_s = E(A^r \mid \mathcal{F}_s) \tag{A.8}$$

where $\mathcal{F}_s$ is the (full) history of the process. We shall first deal with (A.1) and (A.2). Note that they can be rewritten, respectively, as

$$E\langle A^r, \dot{\ell}\rangle_t = o(n^{1/2}) \tag{A.9}$$

and

$$\mathrm{Cov}(\dot{\ell}_t, \langle A^r, \dot{\ell}\rangle_t) = o(n). \tag{A.10}$$

The latter follows from the Bartlett identities for martingales (Mykland (1994)) and from the fact that in both the likelihoods under consideration, $\ddot{\ell} = -[\dot{\ell}, \dot{\ell}]$.

It is clear from (A.9)–(A.10) that one can, without loss of generality, assume that

$$A^r_t = \int_0^t f^r_s dM_s. \tag{A.11}$$

If the censoring times are nonrandom, (A.11) must hold, anyway (see Section 7.2 of Mykland (1993)), but in general this need not be the case.

Since $\mathrm{Var}(A^r) = O(n)$ by assumption,

$$\begin{aligned} O(1) &= \frac{1}{n} E\langle A^r, A^r\rangle_t \\ &= E\int_0^t (f^r_s)^2 \frac{Y_s}{n} d\Lambda_s \\ &= \int_0^t E(f^r_s)^2 E\frac{Y_s}{n} d\Lambda_s + o(1) \end{aligned} \tag{A.12}$$

under asymptotic ergodicity assumptions on Y_t. Hence $E(f_s^r)^2$ is bounded, at least for *a.e.* s. Also, since A^r and M are assumed to be asymptotically jointly Gaussian, $\langle A^r, M\rangle_s/n$ must have a nonrandom limit. Since

$$\langle A^r, M\rangle_s/n = \int_0^s f_u^r \frac{Y_u}{n} d\Lambda_u, \tag{A.13}$$

it follows that f_s^r, $0 \le s \le t$ is nonrandom in the limit, for almost every s, and almost surely. In fact, this limit is zero, in view of (A.9) applied to (A.4).

Hence, under central limit assumptions on Y_t,

$$\int_0^t \mathrm{Cov}(f_s^r, Y_s) g_s d\Lambda_s = o(n^{1/2}) \tag{A.14}$$

for all g in (A.5). By setting $g_s = (EY_s/n)^{-1}$, it follows that

$$\begin{aligned}
\mathrm{Cov}(A^r, U) &= n \int_0^t f_s^r d\Lambda_s \\
&= E \int_0^t f_s^r E Y_s g_s d\Lambda_s \\
&= \int_0^t E(f_s^r Y_s) g_s d\Lambda_s + o(n^{1/2}) \\
&= \mathrm{Cov}(A^r, V) + o(n^{1/2}).
\end{aligned} \tag{A.15}$$

Hence if A^r satisfies (A.1) in the nonparametric model, it also satisfies (A.1) in the dual model.

We now turn our attention to (A.2)/(A.10). Since f_s^r is zero in the limit, and by using the same g as above, $\langle A^r, U - V\rangle_t$ is $o_p(n^{1/2})$. (A.10) is then immediate in the dual model.

When discussing (A.3), one can no longer assume that A^r has the form (A.11). Instead

$$A_t^r = \widetilde{A}_t^r + \int_0^t f_s^r dM_s \tag{A.16}$$

where $\langle \widetilde{A}^r, M\rangle_t = 0$ and where f_s^r clearly retain the properties found above. In particular, $\langle A^r, U\rangle$ is $O_p(1)$ and converges to a nonrandom quantity, hence by the Bartlett identities for martingales,

$$\begin{aligned}
&\mathrm{cum}(A_t^r, A_t^s, U_t) \\
&\quad = E\langle A^r, A^s, U\rangle_t + \mathrm{Cov}(U_t, \langle A^r, A^s\rangle_t) + o(n^{1/2}).
\end{aligned} \tag{A.17}$$

Now expand

$$U_t = \int_0^t \left(\frac{EY_s}{n}\right)^{-1} dM_s + \widetilde{U}_t + O_p(n^{-1/2}) \tag{A.18}$$

where

$$\widetilde{U}_t = -\int_0^t \left(\frac{Y_s - EY_s}{n}\right)\left(\frac{EY_s}{n}\right)^{-2} dM_s. \tag{A.19}$$

The remainder term is $O_p(n^{-1/2})$ since

$$\left\langle \int_0^{\cdot} \left(\frac{Y_s - EY_s}{n}\right)^2 \left(\frac{EY_s}{n}\right)^{-3} dM_s \right\rangle_t = O_p(n^{-1}).$$

For similar reasons, $\widetilde{U}_t = O_p(1)$, and is easily seen to converge in law to $\int_0^t W_s^2 dW_s^1$, where W_s^1 is a Gaussian martingale, and W_s^2 is a Gaussian process. Similarly, $A_s^r/r_n \to W_s^r$ in law, where W_s^r is a Gaussian martingale.

It follows that $E\langle A^r, A^s, \widetilde{U}\rangle_t$ and $\mathrm{Cov}(U_t, \langle A^r, A\rangle_t)$ are both $o(n)$. Hence, under reasonable regularity conditions,

$$\begin{aligned} \mathrm{cum}(A^r, A^s, U) &= \mathrm{cum}\left(A^r, A^s, \int_0^t \left(\frac{EY_s}{n}\right)^{-1} dM_s\right) + o(n) \\ &= o(n) \end{aligned} \tag{A.20}$$

by the nonparametric (A.3), with $g = \left(\frac{EY_s}{n}\right)^{-1}$. ∎

References

AALEN, O. O. (1980). A model for non-parametric regression analysis of counting processes. *Springer Lect. Notes Statist.* **2**, 1–25.

AALEN, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **7**, 1121–1137.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D., & KEIDING, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

COX, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B* **34**, 187–220.

COX, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

COX, D. R. (1980). Local ancillarity. *Biometrika* **67**, 279–286.

JACOD, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Geb.* **31**, 235–253.

JACOD, J. & SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes.* Berlin: Springer-Verlag.

LAZAR, N., & MYKLAND, P. A. (1994). Empirical likelihood in the presence of nuisance parameters (in preparation).

MCCULLAGH, P. (1984). Local sufficiency. *Biometrika* **71**, 233–244.

MCCULLAGH, P. (1987). *Tensor Methods in Statistics.* London: Chapman and Hall.

MYKLAND, P. A. (1993). Dual likelihood. Technical Report No. 367, Department of Statistics, University of Chicago. To appear in *Ann. Statist.*

MYKLAND, P. A. (1994). Bartlett identities for martingales. *Ann. Statist.* **22**, 21–38.

MYKLAND, P. A., & YE, J. (1992). Cumulants and Bartlett identities in Cox regression. Technical Report No. 332, Department of Statistics, University of Chicago.

NÆS, T. (1982). The asymptotic distribution of the estimator for the regression parameter in Cox's regression model. *Scand. J. Statist.* **9**, 107–115.

QIN, J., & LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 300–325.

WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14**, 88–123.

Department of Statistics
University of Chicago
5734 University Avenue
Chicago, Illinois 60637