Cumulants and Bartlett Identities in Cox Regression

Per Aslak Mykland and Jianming Ye

Abstract Expressions are found for the cumulants needed to analyze and improve the accuracy of Cox regression up to order $O(n^{-3/2})$, and consistent estimators for these are given. In addition, the Bartlett identities are shown to hold for the Cox partial likelihood (and, in a broad sense, for partial likelihood in general), so that the cumulant estimators found can be used to adjust statistics in accordance with formulas from likelihood theory. Numerical results are also given.

1 Introduction

In statistics, *accuracy* refers to the the closeness of an (often asymptotic) approximation to the true distribution of a statistic. Frequently, accuracy can be improved by adjusting normal approximations with the help of expansions, saddlepoint approximations, and ancillary statistics. Good accuracy is crucial for correct inference.[1]

The study of accuracy is one of the main contributions of Ole Barndorff-Nielsen. His work stands out with creative approaches like the p^* formula [3] and the r and r^* statistics [4, 5], and with thoughtful views on the philosophy of statistical inference, as in, for example, [6].

Per Aslak Mykland
Department of Statistics, The University of Chicago, 5734 University Avenue, Chicago, IL 60637, USA; e-mail: mykland@pascal.uchicago.edu

Jianming Ye
Zicklin School of Business, Baruch College, One Bernard Baruch Way, New York, NY 10010, USA; and School of Management, Xiamen University, Baoxinliying Building, Postcode 361005, China; e-mail: jianmng.ye@baruch.cuny.edu

[1] Accuracy is sometimes confused with efficiency. The two are not the same, and both are important. In the case of unconditional inference, bootstrapping [11, 15] is also a popular approach to the accuracy problem, see [21] and the references therein.

One path to improved accuracy is to start with the an asymptotic (limiting) distribution, and to modify it with the help of Edgeworth-type correction terms or factors. In the case of likelihood inference, this approach involves the Bartlett identities for log likelihood derivatives [9, 10], and the ensuing Bartlett correction factor for the likelihood ratio statistic and its signed square root [7, 8], [28] (see Section 2).

The purpose of the current paper is to show that the same approach works for partial likelihood inference [14, 40], and specifically in Cox regression, and to operationalize the finding with an estimator for the correction factor.

Inference in Cox regression [13, 14], [39], [1], see also [16] and [2], is usually based on asymptotic (limiting) distributions. There is thus a *prima facie* case for trying the correction methods that work in ordinary likelihood.

For the Cox model, this issue has been addressed in [22], [18], [19] and [30, 32]. Our aim in this paper is to continue this research by finding expressions for the cumulants needed to describe the distributions of estimators and test statistics up to order $O(n^{-3/2})$. This provides a formula for the Bartlett correction factor. The cumulants will be given in a form that lends itself to estimation by empirical substitution, and we also give the resulting estimators.

We also show that the critical Bartlett identities [9, 10] hold quite generally for the Cox model (Theorem 1), thus enabling also other likelihood properties, such as those discussed in [33].

In the following, we begin by explaining the Bartlett identities for (ordinary) likelihood, and their implications for the Bartlett correction the likelihood ratio statistic (Section 2). We then discuss the question of Bartlett identities for partial likelihoods (Section 3). The Cox model and partial likelihood is reviewed in Section 4, with theorems about about Bartlett identities and correction for this criterion function. We then present expressions and estimators for the cumulants (Sections 5 and 6). A few technical arguments are gathered in an appendix.

2 The Bartlett Identities, the Bartlett Factor, and other likelihood quantities

Let $L(\theta,t)$ be a log likelihood function, and set

$$L_{\alpha_1,\dots,\alpha_p}(\theta,t) = \partial^p L(\theta,t)/\partial\theta^{\alpha_1}\cdots\partial\theta^{\alpha_p} \tag{1}$$

to the extent that the derivatives exist. We drop θ as an argument when dealing with the true value (in the notation only, the log likelihood derivatives depend on θ).

Since $\exp\{L(\theta,T)-L(\theta_0,T)\}$ is a likelihood ratio (Radon-Nikodym derivative), it follows that $E\exp\{L(\theta,T)-L(\theta_0,T)\}=1$. Assuming that one can interchange differentiation and expectation, one can thus take derivatives with respect to θ_α, obtaining $E_\theta L_\alpha(\theta,T)\exp\{L(\theta,T)-L(\theta_0,T)\}=0$, $E_\theta(L_{\alpha_1,\alpha_2}(\theta,T)+L_{\alpha_1}L_{\alpha_2})\exp\{L(\theta,T)-L(\theta_0,T)\}=0$, and so on. Setting $\theta=\theta_0$, we obtain $EL_\alpha(T)=0$ (meaning $E_{\theta_0}L_{\alpha_1}(\theta_0,T)=0$), $EL_{\alpha_1,\alpha_2}(T)+EL_{\alpha_1}(T)L_{\alpha_2}(T)=0$. The these two identities are crucial for the central limit theorem for the maximum likelihood estimator (MLE). One can, however, continue the process, obtaining

$$
\begin{aligned}
&EL_{\alpha_1,\alpha_2,\alpha_3}(T)+EL_{\alpha_1,\alpha_2}(T)L_{\alpha_3}(t)[3]+EL_{\alpha_1}(T)L_{\alpha_2}(T)L_{\alpha_3}(T)=0,\\
&EL_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}(T)+EL_{\alpha_1,\alpha_2,\alpha_3}(T)L_{\alpha_4}(T)[4]+EL_{\alpha_1,\alpha_2}L_{\alpha_3,\alpha_4}[3]\\
&\quad+EL_{\alpha_1,\alpha_2}L_{\alpha_3}(t)L_{\alpha_4}(T)[6]+EL_{\alpha_1}(T)L_{\alpha_2}(T)L_{\alpha_3}(T)L_{\alpha_4}(T)=0,
\end{aligned}
\tag{2}
$$

and so on by taking further derivatives of $E\exp\{L(\theta,T)-L(\theta_0,T)\}=1$ with respect to θ_αs. the "[3]" in the first line of (2) means that we sum over the three terms that arise when recombining α_1, α_2 and α_3, and similarly for "[3]", "[4]", and "[6]" in the second and third line. These are the four first Bartlett identities [9, 10] The '$[\cdot]$" notation, and the identities are further discussed in [28]. The Bartlett identities extend to martingales and can heuristically be derived from the likelihood identities [29].

By a combinatorial argument, the Bartlett identities also hold for *cumulants* [28]. Cumulants are moment-like objects that are additive for independent sums. See Section 2 of [28] for a description and discussion. The second cumulant, for example, is the covariance. The cumulant concept seems to have been invented independently by [38] (translated to English in the Appendix of [20]) and [17], see, for example, [20], [27], [37].

To write the Bartlett identities on cumulant form, one simply replaces moments by their corresponding cumulants. If we let "cum" denote cumulants (under probability P_{θ_0}), then cumulants κ of $L_{\alpha_1,\ldots,\alpha_p}(T)$ the are defined as in [28], Ch. 7.2; for example, $\kappa_{\alpha,\beta}=\mathrm{Cov}(L_\alpha\ (T),L_\beta(T))$, $\kappa_{\alpha\beta,\gamma\delta}=\mathrm{Cov}(L_{\alpha\beta}(T),L_{\gamma\delta}(T))$ and $\kappa_{\alpha,\beta,\gamma\delta}=\mathrm{cum}(L_\alpha(T),L_\beta(T),L_{\gamma\delta}(T))$. The cumulant form of the Bartlett identities then becomes

$$
\begin{aligned}
\kappa_\alpha &= 0,\\
\kappa_{\alpha_1\alpha_2}+\kappa_{\alpha_1,\alpha_2}&=0,\\
\kappa_{\alpha_1\alpha_2\alpha_3}+\kappa_{\alpha_1\alpha_2,\alpha_3}[3]+\kappa_{\alpha_1,\alpha_2,\alpha_3}&=0,\\
\kappa_{\alpha_1\alpha_2\alpha_3\alpha_4}+\kappa_{\alpha_1\alpha_2\alpha_3,\alpha_4}[4]+\kappa_{\alpha_1\alpha_2,\alpha_3\alpha_4}[3]+\kappa_{\alpha_1\alpha_2,\alpha_3,\alpha_4}[6]+\kappa_{\alpha_1,\alpha_2,\alpha_3,\alpha_4}&=0,
\end{aligned}
\tag{3}
$$

and so on.

The first four Bartlett identities yield the following. In the absence of nuisance parameters, the likelihood ratio statistic $w=2(L(\hat{\theta},T)-L(\theta,T))$ has a Bartlett correction factor ([7], [8], [28]) $b(\theta)$ given by (using the summation convention)

$$
\begin{aligned}
npb(\theta) = {} & (\kappa_{\alpha\beta\gamma\delta}+4\kappa_{\alpha,\beta\gamma\delta}+4(\kappa_{\alpha,\gamma,\beta\delta}+\kappa_{\alpha\gamma,\beta\delta}))\kappa^{\alpha,\beta}\kappa^{\gamma,\delta}/4\\
&+\lambda_\alpha\lambda_\beta\kappa^{\alpha,\beta}/4\\
&+(\kappa_{\alpha\beta\gamma}\kappa_{\delta\varepsilon\zeta}+6\kappa_{\alpha\beta,\gamma}\kappa_{\delta\varepsilon,\zeta})\kappa^{\alpha,\delta}\kappa^{\beta,\varepsilon}\kappa^{\gamma,\zeta}/6
\end{aligned}
$$

where $\lambda_\alpha=-(2\kappa_{\alpha\beta,\gamma}+\kappa_{\alpha\beta\gamma})\kappa^{\beta,\gamma}$ (cf. Ch. 7 of [28]). We have here put the expression in a form where it can immediately be estimated by the methods developed in the previous section.

The actual corrected likelihood ratio statistic is $w/(1+b(\theta)/n)$, which is well approximated by χ^2_p under a variety of regularity conditions.

The improvement remains strong in the tails (as compared to usual Edgeworth expansion based results). This is presumably due to the connection between Bartlett correction and the saddlepoint approximation, [8], [35]. Similar results (both additive and multiplicative) apply to the signed square root statistic r [7], [8], see also [28].

It should be emphasized that the Bartlett factor is not the only quantity that can be found from the Bartlett identities. Higher order identities (beyond four) give rise to improved approximation in the tails [33]. This can also be obtained by the elegant r^* approach developed by [4, 5], [26] and others. For the connection of this work to the Bartlett identities, see [34].

Also, for example, the expressions for bias, variance and third cumulant given on p. 209 in [28] carry over directly to any criterion function that satisfied the Bartlett identities, In our notation,

$$E\hat{\theta}^\alpha = \theta - \frac{1}{n}\sum_{\beta,\gamma,\delta}\kappa^{\alpha,\beta}\kappa^{\alpha,\delta}(\kappa_{\beta,\gamma,\delta}+\kappa_{\beta,\gamma\delta})/2+O(n^{-3/2}),$$

and so on ($\kappa^{\alpha,\beta}$ is the inverse of $\kappa_{\alpha,\beta}$).

3 Bartlett Identities for Partial Likelihood

The Bartlett identities hold quite generally in partial likelihood problems. To see this, let $\theta = (\theta^1, \theta^2, \dots)$ be the parameter appearing in the partial likelihood, and let θ_0 be the true value of θ. Also let P be the true probability distribution. Set up a dummy inference problem by $Q_{\theta_0} = P$, $dQ_\theta/dQ_{\theta_0} =$ the partial likelihood ratio between θ and θ_0. The Bartlett identities are then satisfied (subject to regularity conditions) for the derivatives of the log partial likelihood provided Q_θ is a probability measure, i.e., if the partial likelihood ratio integrates to 1 w.r.t. P.

This is obviously true for the partial likelihood formulation in [14] and [40]. It is also true for the counting process formulation of the Cox model with which we shall be working in view of Theorem II.7 (p. 93) of [2], see also [23].

It should be emphasized that there is no need for Q_θ to be the actual distribution associated with regression coefficients θ. Thus, for example, there is no need for any assumption that the censoring is noninformative (in the sense of, e.g., Definition 4.3.1 (p. 139) of [16]). If censoring is informative, one does, of course, lose efficiency by using the partial likelihood, but one does not lose the Bartlett identities.

A conceptually important point relating to the Bartlett identities is that there are two sets of them. In addition to the ordinary ones discussed above, there are also conditional variation identities. For example, if L is the partial log likelihood in a one parameter problem, $\dot{L}$, $[\dot{L},\dot{L}]+\ddot{L}$, $[\dot{L},\dot{L},\dot{L}]+3[\dot{L},\ddot{L}]+\dddot{L}$, etc., are martingales (the [,]-notation is as in stochastic calculus (cf. [24], for example); the higher order [

]-notation is discussed in [29]. These latter identities do not require the setting up of a dummy inference problem.

The two first conditional variation identities are, of course, at the heart of partial likelihood [13, 14], [40]. Since Var $\dot{L} = E[\dot{L},\dot{L}]$, they also relate closely to the two first ordinary Bartlett identities. We make the conjecture that a similar relationship also holds for higher order identities, by exploiting the Bartlett identities for martingales [29]. We have not explored this matter further, however, as the current Theorem 1 follows without any technicalities whatsoever.

4 The Cox Model

Following [1], the model can be described by point processes $N^1(t),\ldots,N^n(t)$, $0 \leq t \leq T < \infty$, where $N^i(t)$ jumps from 0 to 1 when patient no. i dies, provided he or she does so under observation. $N^i(t)$ has intensity $\lambda_0(t)\lambda^i(\theta,t)$, where

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

where $Y_i(t)$ is a predictable process which is 0 or 1 according to whether the i'th patient is under observation, $Z_{\alpha,i}(t)$ is a predictable covariate process, θ^α is an unknown parameter, and $\lambda_0(t)$ is a baseline intensity. All remaining patients are assumed to be censored at the nonrandom time T.

The partial log likelihood at time t is given by

$$L(\theta,t) = \sum_{\alpha,i}\int_0^t \theta^\alpha Z_{\alpha,i}(s)dN^i(s) - \int_0^t f(\theta,s)d\bar{N}(s)$$

where $\bar{N} = N^1 + \cdots + N^n$, $\bar{\lambda} = \lambda^1 + \cdots + \lambda^n$ and $f(\theta,t) = \ln\bar{\lambda}(\theta,t) + \ln\lambda_0(t)$ when $\bar{\lambda}(\theta,t) \neq 0$, and $= 0$ otherwise. Let $f_{\alpha_1,\ldots,\alpha_p}(\theta,t)$ and $\bar{\lambda}_{\alpha_1,\ldots,\alpha_p}(\theta,t)$ be derivatives with respect to components of θ, defined in analogy with (1).

Note that the way we define it, the partial log likelihood does depend on the baseline hazard $\lambda_0(t)$. This is done for notational convenience. However, the dependence on λ_0 washes out both in the likelihood ratio statistic and in the derivatives of $L(\theta,t)$. The same occurs in the estimators of the cumulants (formula (7) below).

Expressions and estimators for the cumulants are particularly useful because of the following fact (explained in Section 3):

Theorem 1. *Assume that the covariates are bounded, and that $\int_0^T \lambda_0(t)dt < \infty$. Then the Cox partial likelihood satisfies the Bartlett identities.*

The importance of this theorem is that there are a number of results relating to likelihood inference which depend only on the likelihood structure through the Bartlett identities, as discussed in Section 2.

In the following, the cumulants are found under the assumption that patients are independent of each other. For a given patient, however, censoring can be arbitrarily dependent on the history of the patient, and covariates can be random (predictable) and time varying. Specifically, we shall be using the framework of [1].

CONDITION 1 *We shall assume that* $\lambda_0(t)$ *is nonrandom and that if* $\mathscr{F}_i = \sigma(Z_{\alpha,i}(t), Y_i(t), N^i(t),$ *all* $\alpha, t)$, *then the* $\mathscr{F}_i$ *are independent for all i (i.e., the patients are independent). An implication of this assumption is that a decision to censor one patient cannot be dependent on what happens to another patient. A patient can, however, be censored on basis of his or her own medical history.*

In addition, we use the technical assumption that the (absolute value of the) covariates are bounded, and that there are nonrandom constants λ_0^- *and* λ_0^+ *so that* $0 < \lambda_0^- \leq \lambda_0(t) \leq \lambda_0^+ < \infty$.

In order to show that these corrections actually do improve the asymptotic accuracy of rejection and coverage probabilities, one also needs to show the existence of relevant Edgeworth expansions. We do not do this in this paper, but a result to this effect is stated in [19]. Alternatively, we can use the results in [31], which can be extended to likelihood ratio statistics in much the same way as the proof of Theorem 2 in [30]. The conditions in [19] are stronger than in [30], but so are the conclusions (in terms of topology of convergence).

To state a specific result, we shall borrow the expansion of [19].

Theorem 2. *Assume Condition 1, and also that there is only one covariate; that the patients (in terms of death, censoring and covariate) are i.i.d.; and that for each patient, death and censoring are conditionally independent given the covariate. Then the distribution of the Bartlett corrected Cox partial likelihood ratio statistic (the Bartlett factor being given in Section 2) is* $\chi_1^2 + O(n^{-3/2})$.

Note that the additional conditions imposed relative to those in Section 4 is what is needed to make the result in [19] hold, cf. Section 1 of that paper. The result follows from [19] and (our) Theorem 1 by the same method as used in [12] and [25].

In a sense, the results in this paper are orthogonal to those of [19] and [31]. Whereas these two papers provide conditions for expansions to exist, the issue here is what the expansions look like once they have been shown to exist. For example, it is clear from our results that the coefficient κ_4 in [19] is zero (cf. the proof of Theorem 3).

5 Cumulants for Partial Likelihood

Our basic result is the following theorem giving expressions for such cumulants of degree up to 4. In writing these expressions, we use v to denote any unpartitioned index set of more than one element (i.e., $\{\beta\gamma\}, \{\beta\gamma\delta\}$, etc.). Also, (i) means that patient no. i has been removed from the computation of the expression, so that, e.g., $f_\alpha^{(i)}$ denotes f_α calculated on the basis of $n-1$ patients with patient i excluded.

Theorem 3. *Under Condition 1,*

$$\kappa_v = -E\int_0^T f_v(t)\bar{\lambda}(t)\lambda_0(t)dt \text{ when } |v|\geq 2, \text{ and } = 0 \text{ when } |v| = 1,$$

$$\kappa_{\alpha,v} = -\sum_i E\int_0^T f_v^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(t)dt\int_0^t (Z_{\alpha,i}(s) - f_\alpha(s))\lambda^i(s)\lambda_0(s)ds,$$

$$\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} = \sum_i E\Bigg(\int_0^T f_{\gamma\delta}^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(t)dt \times\int_0^t \left[f_{\alpha\beta}(s) - (Z_{\alpha,i}(s) - f_\alpha(s))(Z_{\beta,i}(s) - f_\beta(s))\right]\lambda^i(s)\lambda_0(s)ds\Bigg),$$

and

$$\kappa_{\alpha\beta,\gamma\delta} = E\int_0^T f_{\alpha\beta}(t)f_{\gamma\delta}(t)\bar{\lambda}(t)\lambda_0(t)dt + \sum_i E\int_0^T\left[f_{\alpha\beta}^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(t)f_{\alpha\beta}(t)\bar{\lambda}(t)\right]dt\int_0^t f_{\gamma\delta}(s)\lambda^i(s)\lambda_0(s)ds \quad [2] + \int_0^T dt\int_0^T \mathrm{Cov}(f_{\alpha\beta}(t)\bar{\lambda}(t)\lambda_0(t), f_{\gamma\delta}(s)\bar{\lambda}(s)\lambda_0(s))ds.$$

The [2] denotes that we sum over the two permutations of $\alpha\beta$ and $\gamma\delta$. The cumulants $\kappa_{\alpha,\beta}$, $\kappa_{\alpha,\beta,\gamma}$ and $\kappa_{\alpha,\beta,\gamma,\delta}$ can be found from the above theorem with the help of the Bartlett identities for cumulants (which follow from the identities for moments in view of [36] or [28], ex. 7.1 (p. 222)).

6 Approximation and Estimation of the Cumulants

We focus on the three first quantities in Theorem 3, as all first order stable cumulant combinations of degree up to four can be expressed as a linear combination of these (in particular, this is true for the quantities appearing in the Bartlett correction factor, cf. Ch. 7 and 8 of [28]. Consider first $\kappa_{\alpha,v}$ and $\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta}$. These are both on the form $EK(\theta_0)$, where

$$K(\theta) = -\sum_i\int_0^T (f_v(\theta,t)\bar{\lambda}(\theta,t) - f_v^{(i)}(\theta,t)\bar{\lambda}^{(i)}(\theta,t))\lambda_0(t)dt\int_0^t g_{i,v}(\theta,s)\lambda^i(\theta,s)\lambda_0(s)ds \tag{4}$$

where $g_{i,v}$ is $Z_{\alpha,i} - f_\alpha$ or $f_{\alpha\beta} - (Z_{\alpha,i} - f_\alpha)(Z_{\beta,i} - f_\beta)$, respectively. This is because $\sum_i g_i(s)\lambda^i(s) = 0$ in both the cases considered.

Theorem 4. *Under Condition 1,*

$$\text{case } \kappa_{\alpha,v}: \quad K(\theta) = -\sum_i \int_0^T \lambda_v^i(\theta,t)\lambda_0(t)dt \int_0^t g_{i,v}(\theta,s)\lambda^i(\theta,s)\lambda_0(s)ds$$

$$\text{case } \kappa_{\alpha,\beta,\gamma\delta}+\kappa_{\alpha\beta,\gamma\delta}: \tag{5}$$

$$K(\theta) = \sum_i \int_0^T \frac{\bar{\lambda}_\gamma(\theta,t)\bar{\lambda}_\delta(\theta,t)}{\bar{\lambda}(\theta,t)}\lambda_0(t)dt \int_0^t g_{i,\gamma\delta}(\theta,s)\lambda^i(\theta,s)\lambda_0(s)ds + o_p(n)$$

uniformly on compact sets of θ*s.*

Hence $K(\theta)$ is, to first order, a functional of independent sums of bounded terms, and it is easy to see that the remainder term is $o(n)$ in expectation. By the law of large numbers and by uniform integrability, it follows that

$$K(\theta_0) = EK(\theta_0) + o(n). \tag{6}$$

Also, this representation, along with Theorem 4, shows that consistent estimation is possible. A natural estimator is

$$\hat{K} = \sum_i \int_0^T f_v^{(i)}(\tilde{\theta},t)\bar{\lambda}^{(i)}(\tilde{\theta},t)d\hat{\Lambda}_0(t) \int_0^t g_i(\tilde{\theta},s)\lambda^i(\tilde{\theta},s)d\hat{\Lambda}_0(s), \tag{7}$$

where $\tilde{\theta}$ and $\hat{\Lambda}_0$ are consistent estimators of θ and the baseline hazard, respectively. By the same type of reasoning as in Theorem 4,

$$\hat{K} = K + o_p(n), \tag{8}$$

whence consistent estimation of EK/n can be carried out. Note that we do not assume that EK/n has a limit as $n \to \infty$, and so do not require i.i.d. assumptions. (If EK/n only has cluster points, one can do the arguments with subsequences.)

To state a formal result summarizing the above, we assume that $\tilde{\theta}$ is the maximum partial likelihood estimator, either unrestricted or under a null hypothesis (in particular, it can be the true parameter). We also let $\hat{\Lambda}_0$ be the estimator of the baseline hazard given on p. 1103 of [1]:

$$\hat{\Lambda}_0(t) = \int_0^t \bar{\lambda}(\tilde{\theta},s)^{-1}d\bar{N}(s)$$

Theorem 5. *Assume Condition 1. Then (6) and (8) hold.*

As far as estimation of the $\kappa_v s$ is concerned, a natural choice is the observed derivatives of the log likelihood,

$$\hat{\kappa}_v = -\int_0^T f_v(\tilde{\theta},t)d\bar{N}(t),$$

and, obviously,

$$\hat{\kappa}_v = \kappa_v + o_p(n).$$

Estimating $\kappa_{\alpha\beta,\gamma\delta}$ is different from what is discussed above only in so far as the term involving the covariance is concerned. This covariance, however, can be asymptotically approximated by a sum of generalized cumulants of sums of independent random variables, and this approximation is easily used to specify a consistent estimator.

7 Conclusion

The Bartlett identities and Bartlett correction are important tools for improving the accuracy of likelihood inference (see Section 2 for a brief review). We here argue that the technology carries over to partial likelihood (Section 3), and then provide specific conditions for the identities (Theorem 1) and the correction (Theorem 2) to hold in the case of Cox regression (Section 4). The identities and the correction involve cumulants of low order derivatives of the log partial likelihood, and these are given explicit expression in Theorem 3 in Section 5. In order to apply the Bartlett correction in practice, the relevant cumulants need to have consistent estimators, and these are found in Section 6. The proposed correction really does improve the accuracy, as documented in Tables 1-3.

8 Appendix: Proofs of Theorems

In addition to the definitions in Section 4, let $M^i(t) = N^i(t) - \int_0^t \lambda^i(s)ds$, with $\bar{M}(t) = M^1(t) + M^2(t) + \cdots + M^n(t)$.

Proof of Theorem 3. The formula for κ_v is obvious, as $L_v(T) = -\int_0^T f_v(t)d\bar{N}(t)$ for $|v| \geq 2$.

Next,

$$\begin{aligned}\kappa_{\alpha,v} &= -EL_\alpha(T)\int_0^T f_v(t)\bar{\lambda}(t)\lambda_0(t)dt \\ &\quad - EL_\alpha(T)\int_0^T f_v(t)(d\bar{N}(t) - \bar{\lambda}(t)\lambda_0(t)dt) \\ &= -\int_0^T EL_\alpha(t)f_v(t)\bar{\lambda}(t)\lambda_0(t)dt\end{aligned}$$

since $L_\alpha(t)$ is a martingale, and since $\langle L_\alpha, \int_0^\cdot f_v(t)(d\bar{N}(t) - \bar{\lambda}(t)\lambda_0(t)dt)\rangle_T$ $= \int_0^T (Z_{\alpha,i}(s) - f_\alpha(s))f_v(s)\lambda^i(s)\lambda_0(s)ds = 0$. On the other hand, $\lambda^i(t)$ is zero after the jump of $N^i(t)$, hence

$$\kappa_{\alpha,\nu} = -\sum_i \int_0^T E\left[\int_0^t (Z_{\alpha,i}(s) - f_\alpha(s))\, dN^i(s)\right] f_\nu^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(t)dt$$

which yields the desired result since $dM^i(s) = dN^i(s) - \lambda^i(s)\lambda_0(s)ds$ is the differential of a martingale which is independent of $f_\nu^{(i)}(t)\bar{\lambda}^{(i)}(t)\lambda_0(s)$ (so that the latter quantity can be moved inside an integral with respect to this martingale).

Turning to $\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta}$, note that both the martingales
$L_\alpha(t)L_\beta(t) - [L_\alpha, L_\beta]_t = \int_0^t L_\alpha(s)dL_\beta(s)[2]$ and
$L_{\alpha\beta}(t) + [L_\alpha, L_\beta]_t = \int_0^t \left[(Z_{\alpha,i}(s) - f_\alpha(s))\left(Z_\beta, i(s) - f_\beta(s)\right) - f_{\alpha\beta}(s)\right] dN^i(s)$
are orthogonal to $L_{\gamma\delta}(t) + \langle L_\gamma, L_\delta\rangle_t = -\int_0^t f_{\gamma\delta}(s)d\bar{M}(s)$, in the former case because $\langle L_\beta, \bar{M}\rangle_t \equiv 0$ and in the latter case for the same reason that makes $L_{\alpha\beta}(t) + [L_\alpha, L_\beta]_t$ a martingale. It follows that $\mathrm{Cov}\left(L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t), L_{\gamma\delta}(t) + \langle L_\gamma, L_\delta\rangle_t\right) = 0$, whence

$$\begin{aligned}\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} &= \mathrm{Cov}\left(L_\alpha(T)L_\beta(T) + L_{\alpha\beta}(T), \langle L_\gamma, L_\delta\rangle_T\right)\\ &= \int_0^T E\left[\left(L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t)\right) f_{\gamma\delta}(t)\bar{\lambda}(t)\lambda_0(t)\right] dt\end{aligned}$$

since $L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t)$ is a martingale. The stated formula then follows by the same reasoning as for $\kappa_{\alpha,\nu}$.

Finally, for $\kappa_{\alpha\beta,\gamma\delta}$, note that $\mathrm{Cov}\left(L_{\alpha\beta}(T) + \langle L_\alpha, L_\beta\rangle_T, L_{\gamma\delta}(T) + \langle L_\gamma, L_\delta\rangle_T\right) = E\int_0^T f_{\alpha\beta}(t)f_{\gamma\delta}(t)dt$, and that, by the method used for $\kappa_{\alpha,\nu}$,

$$\begin{aligned}&\mathrm{Cov}\left(\langle L_\alpha, L_\beta\rangle_T, L_{\gamma\delta}(T) + \langle L_\gamma, L_\delta\rangle_T\right)\\ &\quad = -\sum_i E\int_0^T \left[f_{\alpha\beta}^{(i)}(t)\bar{\lambda}^{(i)}(t) - f_{\alpha\beta}(t)\bar{\lambda}(t)\right]\lambda_0(t)dt \int_0^t f_{\gamma\delta}(s)\lambda^i(s)\lambda_0(s)ds.\end{aligned}$$

The expression for $\kappa_{\alpha\beta,\gamma\delta}$ then follows.

Proof of Theorem 4. The first case follows since $f_\nu(\theta,t)\bar{\lambda}(\theta,t) = \bar{\lambda}_\nu(\theta,t)$ and $f_\nu^{(i)}(\theta,t)\bar{\lambda}^{(i)}(\theta,t) = \bar{\lambda}_\nu^{(i)}(\theta,t)$, In the second case, $f_{\gamma\delta}(\theta,t)\bar{\lambda}(\theta,t) = (\bar{\lambda}_{\gamma\delta}(\theta,t) - \bar{\lambda}_\gamma(\theta,t)\bar{\lambda}_\delta(\theta,t))/\bar{\lambda}(\theta,t)$, and similarly when removing patient i. By Condition 1 and by Taylor expansion, it is then easy to see that
$\sum_i (f_\nu(\theta,t)\bar{\lambda}(\theta,t) - f_\nu^{(i)}(\theta,t)\bar{\lambda}^{(i)}(\theta,t)) = -\bar{\lambda}_\gamma(\theta,t))\bar{\lambda}_\delta(\theta,t)/\bar{\lambda}(\theta,t) + o_p(n)$, uniformly in $t \in [0,T]$ and in a compact set of θs.

Proof of Theorem 5. We use the result in Theorem 4. (6) is obvious in view of our assumptions, while (8) follows by the same reasoning as in the proofs of Lemma 3.1 (p. 1105) and Theorem 3.4 (p. 1108) of [1]. Note that in view of Condition 1, $\liminf_{n\to\infty} \sum_{i=1}^n EY_i(t)/n$ is nonzero in an interval in t containing 0, which is sufficient (in view of Condition 1) to ensure that Conditions A–D on p. 1105 of [1] are satisfied, whence $\tilde{\theta}$ is consistent and asymptotically normal, and similarly for $\hat{\Lambda}_0$.

Table 1 Baseline failure distribution: exponential (1); censoring: Uniform (0,1); # patients: 20; # simulations: 10000.

Classes	nominal	uncorrected LR	corrected LR
mean	1	1.081	1.045
variance	2	2.364	2.059
χ^2 tests, nominal and actual rejection probabilities	1% 5% 10% 50%	1.3% 6.0% 11.2% 52.0%	1.0% 5.4% 10.6% 52.0%

Table 2 Baseline failure distribution: exponential (1); censoring: Uniform (0,2); # patients: 10; # simulations: 15000.

Classes	nominal	uncorrected LR	corrected LR
mean	1	1.150	1.050
variance	2	2.635	2.099
χ^2 tests, nominal and actual rejection probabilities	1% 5% 10% 50%	1.5% 6.9% 12.8% 53.9%	0.8% 5.1% 10.8% 53.0%

Table 3 Baseline failure distribution: exponential (1); censoring: Uniform (0,=2); # patients: 10; # simulations: 20000.

Classes	nominal	uncorrected LR	corrected LR
mean	1	1.141	1.025
variance	2	2.298	1.813
χ^2 tests, nominal and actual rejection probabilities	1% 5% 10% 50%	1.3% 6.6% 12.6% 54.1%	0.8% 4.4% 9.8% 53.7%

Acknowledgements We would like to thank Ole Barndorff-Nielsen for your leadership and inspiring work, both on likelihood, and on high frequency data. We would also like to thank Almut Veraart and the referee for feedback that has substantially improved the paper. The authors are also grateful to the (U.S.) National Science Foundation for support under several grants, most recently DMS-14-07812.

References

1. Andersen, P.K., Gill, R.D.: Cox's regression model for counting processes: a large sample study. Ann. Statist. **10**, 1100–1120 (1982)
2. Andersen, P.K., Borgan, Ø., Gill, R.D., Keiding, N. *Statistical Models Based on Counting Processes* (Springer-Verlag, New York, 1993)
3. Barndorff-Nielsen, O.E.: Conditionality resolutions. *Biometrika* **67** 293-310 (1980)
4. Barndorff-Nielsen, O.E.: Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73** 307-322 (1986)
5. Barndorff-Nielsen, O.E.: Modified signed log likelihood ratio. *Biometrika* **78** 557-563 (1991)
6. Barndorff-Nielsen, O.E.: Diversity of Evidence and Birnbaum's Theorem [with Comments and Reply]. *Scand. J. Statist.* **22** 513-522 (1995)
7. Barndorff-Nielsen, O.E., Cox, D.R.: Edgeworth and saddle-point approximations with statistical applications (with discussion). *J. Roy. Statist. Soc. Ser. B.* **41** 279312 (1979)
8. Barndorff-Nielsen, O.E., Cox, D.R.: Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. J. Roy. Statist. Soc. Ser. B. **46**, 483–495 (1984)
9. Bartlett, M.S.: Approximate Confidence Intervals. *Biometrika* **40** 12-19 (1953)
10. Bartlett, M.S.: Approximate Confidence Intervals. II. More than One Unknown Parameter. *Biometrika* **40** 306-317 (1953)
11. Beran, R.J: Estimated sampling distributions: The bootstrap and its competitors. Ann. Statist. **10**, 212-225 (1982)
12. Cordeiro, G.M.: On the corrections to the likelihood ratio statistics. *Biometrika* **74**, 265-274 (1987)
13. Cox, D.R.: Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B.* **34**, 187–220 (1972)
14. Cox, D.R.: Partial likelihood. *Biometrika* **62**, 269–276 (1975)
15. Efron, B. *The Jackknife, the Bootstrap and Other Resampling Plans* (CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadephia, 1982).
16. Fleming, T.R., Harrington, D.P. *Counting Processes and Survival Analysis*. (Wiley, New York, 1991)
17. Fisher, R.A.: Moments and product moments of sampling distributions. *Proceedings of the London Mathematical Society Ser. 2* **30** 199-238 (1929)
18. Gu, M.: On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20**, 399–414 (1992)
19. Gu, M., Zheng, Z.: On the Bartlett adjustment for the partial likelihood ratio test in the Cox regression model. *Statistica Sinica* **3**, 543–555 (1993)
20. Hald, A.: The early history of the cumulants and the Gram-Charlier series. *Inter. Statist. Rev.* **68** 137-153 (2000)
21. Hall, P. *The Bootstrap and Edgeworth Expansion* (Springer-Verlag, New York, 1992).
22. Hjort, N.L.: Bootstrapping Cox's regression model. Technical Report No. 241, Department of Statistics, Stanford University (1985)
23. Jacod, J.: Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschr. Wahr. verw. Geb.* **31** 235-253 (1975)
24. Jacod, J., Shiryaev, A.N. *Limit Theorems for Stochastic Processes.* 2nd ed. (Springer-Verlag, Berlin, 2003).
25. Jensen, J.L.: A historical sketch and some new results on the improved log likelihood ratio statistic. *Scand. J. Statist.* **20** 1-15 (1993)
26. Jensen, J.L.: A simple derivation of r^* for curved exponential families. Scand. J. Statist. *Scand. J. Statist.* **24** 33-46. (1997)
27. Lauritzen, S.L.: *Thiele - Pioneer in Statistics.* (OUP, Oxford, 2002).
28. McCullagh, P. *Tensor Methods in Statistics.* (Chapman and Hall, London, 1987).
29. Mykland, P.A.: Bartlett identities for martingales. *Ann. Statist.* **22**, 21–38 (1994)

30. Mykland, P.A.: Dual likelihood. *Ann. Statist.* **23**, 396-421 (1995)
31. Mykland, P.A.: Embedding and asymptotic expansions for martingales. *Probab. Theory rel. Fields* **103** 475-492 (1995)
32. Mykland, P.A.: Martingale expansions and second order inference. *Ann. Statist.* **23**, 707-731 (1995)
33. Mykland, P.A.: Bartlett identities and large deviations in likelihood theory. *Ann. Statist.* **27** 1105-1117 (1999)
34. Mykland, P.A.: Likelihood computations without Bartlett identities. *Bernoulli* **7** 473-485. (2001)
35. Reid, N.: Saddlepoint methods and statistical inference. *Statist. Sci.* **2** 213–227 (1988)
36. Skovgaard, I: A note on the differentiation of cumulants of log likelihood derivatives. *Int. Statist. Rev.* **54**, 29–32 (1986)
37. Stigler. S.M.: Discussion for ISI history session - IMP3 http://www.stat.fi/isi99/proceedings/arkisto/varasto/stig1042.pdf Cited 12 Aug 1999
38. Thiele, T.N.: Om iagttagelseslærens halvinvarianter. *Kgl. danske Videnskabernes Selskaps Forhandlinger* **3** 135-141 (1899)
39. Tsiadis, A.A.: A large sample study of Cox's regression model. *Ann. Statist.* **9**, 93-108 (1981)
40. Wong, W.H.: Theory of partial likelihood. *Ann. Statist.* **14**, 88–123 (1986)