

# Cumulants and Bartlett Identities in Cox Regression

by

*Per Aslak Mykland*<sup>1</sup> and *Jianming Ye*<sup>2</sup>

TECHNICAL REPORT NO. 332

Department of Statistics  
The University of Chicago  
Chicago, Illinois 60637

December, 1991

*Revised June, 1995*

---

<sup>1</sup> Per A. Mykland is Assistant Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. Research was supported in part by National Science Foundation grant DMS 93-05601.

<sup>2</sup> Jianming Ye is Assistant Professor, Graduate School of Business, University of Chicago, Chicago, IL 60637. Research was supported in part by The University of Chicago Graduate School of Business Research Fund.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, DMS 87-03942, and DMS 86-01732 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

*Key words and phrases:* Bartlett identities, Bartlett correction, partial likelihood, proportional hazards model, survival analysis.

*Running head:* Cox regression.

# Cumulants and Bartlett Identities in Cox Regression

by

*Per Aslak Mykland*<sup>1</sup> and *Jianming Ye*<sup>2</sup>

## Abstract

Expressions are found for the cumulants needed to analyze 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.

---

<sup>1</sup> Per A. Mykland is Assistant Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. Research was supported in part by National Science Foundation grant DMS 93-05601.

<sup>2</sup> Jianming Ye is Assistant Professor, Graduate School of Business, University of Chicago, Chicago, IL 60637. Research was supported in part by The University of Chicago Graduate School of Business Research Fund.

This manuscript was prepared using computer facilities supported in part by the National Science Foundation grants DMS 89-05292, DMS 87-03942, and DMS 86-01732 awarded to the Department of Statistics at The University of Chicago, and by The University of Chicago Block Fund.

*Key words and phrases:* Bartlett identities, Bartlett correction, partial likelihood, proportional hazards model, survival analysis.

*Running head:* Cox regression.

**1. Introduction.** Inference in Cox regression (Cox (1972, 1975), Tsiadis (1981), Andersen & Gill (1982); see also Fleming & Harrington (1991) and Andersen, Borgan, Gill & Keiding (1993)) is usually based on asymptotic distributions. This raises the question of whether we can find Edgeworth-type correction terms or factors which can improve on the asymptotic approximation. This issue has begun to be addressed in Hjort (1985), Gu (1992), Gu & Zheng (1993) and Mykland (1995a, 1995c). Our purpose 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  $O(n^{-3/2})$ . The cumulants will be given in a form that lends itself to estimation by empirical substitution, and we also give the resulting estimators.

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 Andersen & Gill (1982).

Expressions and estimators for the cumulants are particularly useful because of the following fact (explained in Section 2):

*Theorem 1.* 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. These now generalize to the Cox partial likelihood. Important instances of such results are the validity of correction factors and terms for the log likelihood ratio statistic and the signed square root of this statistic, see, e.g., McCullagh (1987).

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 Gu & Zheng (1993). Alternatively, we can use the results in Mykland (1995b), which can be extended to likelihood ratio statistics in much the same way as the proof of Theorem 2 in Mykland (1995a). The conditions in Gu & Zheng (1993) are stronger than in Mykland (1995a), but so are the conclusions (in terms of topology of convergence).

To state a specific result, we shall borrow the expansion of Gu and Zheng (1993).

*Theorem 2.* Assume the conditions stated in Section 3; 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 6) is  $\chi_1^2 + O(n^{-3/2})$ . ■

Note that the additional conditions imposed relative to those in Section 3 is what is needed to make Gu and Zheng's (1993) result hold, cf. Section 1 of that paper. The result follows from Gu and Zheng (1993) and (our) Theorem 1 by the same method as used in Cordeiro (1987) and Jensen (1993).

In a sense, the results in this paper are orthogonal to those of Gu & Zheng (1993) and Mykland (1995b). 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  $\kappa_4$  in Gu & Zheng (1993) is zero (cf. the proof of Theorem 2).

In the following, we begin by elaborating on the question of Bartlett identities for partial likelihoods (Section 2). The Cox model is then reviewed (Section 3), whereupon we present expressions and estimators for the cumulants (Sections 4 and 5). Finally, we discuss the Bartlett factor (Section 6). A few technical arguments are gathered in an appendix.

**2. 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  $\theta_0$  be the true value of  $\theta$ . 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  $\theta$  and  $\theta_0$ . The Bartlett identities are then satisfied (subject to regularity conditions) for the derivatives of the log partial likelihood provided  $Q_{\theta}$  is a probability measure, i.e., if the partial likelihood ratio integrates to 1.

This is obviously true for the partial likelihood formulation in Cox (1975) and Wong (1986). 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 of Andersen, Borgan, Gill and Keiding (1993, p. 93), see also Jacod (1975).

It should be emphasized that there is no need for  $Q_\theta$  to be the actual distribution associated with regression coefficients  $\theta$ . 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 Fleming and Harrington (1991)). 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,  $\dot{L}$ ,  $[\dot{L}, \dot{L}] + \ddot{L}$ ,  $[\dot{L}, \dot{L}, \dot{L}] + 3[\ddot{L}, \ddot{L}] + \dddot{L}$ , etc., are martingales (the  $[\cdot, \cdot]$ -notation is as in stochastic calculus (cf. Jacod & Shiryaev (1987), for example); the higher order  $[\cdot]$ -notation is discussed in Mykland (1994)). 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 (Cox (1972, 1975), Wong (1986)). Since  $\text{Var } \dot{L} = E[\dot{L}, \dot{L}]$ , 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 (Mykland (1994)). We have not explored this matter further, however, as the current Theorem 1 follows without any technicalities whatsoever.

**3. The Cox Model.** Following Andersen & Gill (1982), the model can be described by point processes  $N^1(t), \dots, N^n(t)$ ,  $0 \leq t \leq T < \infty$ , 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

$$\lambda^i(\theta, t) = Y_i(t) \lambda_0(t) \exp \left( \sum_{\alpha} \theta^{\alpha} Z_{\alpha, i}(t) \right),$$

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,  $\theta^{\alpha}$  is an unknown parameter, and  $\lambda_0(t)$  is a baseline intensity. All remaining patients are assumed to be censored at the nonrandom time  $T$ .

ASSUMPTIONS. Apart from in Theorem 1, we shall assume that  $\lambda_0(t)$  is nonrandom and that if  $\mathcal{F}_i = \sigma(Z_{\alpha, i}(t), Y_i(t), N^i(t))$ , all  $\alpha, t$ , then the  $\mathcal{F}_i$  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 covariates are bounded, and that  $\int_0^T \lambda_0(t) dt < \infty$ . ■

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  $\bar{N} = N^1 + \dots + N^n$ ,  $\bar{\lambda} = \lambda^1 + \dots + \lambda^n$  and  $f(\theta, t) = \ln \bar{\lambda}(\theta, t)$  when  $\bar{\lambda}(\theta, t) \neq 0$ , and  $= 0$  otherwise.

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  $\lambda_0$  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 (3)).

**4. Cumulants.** Let  $L_{\alpha_1, \dots, \alpha_p}(\theta, t) = \partial^p L(\theta, t) / \partial \theta^{\alpha_1} \dots \partial \theta^{\alpha_p}$ , and similarly for  $f_{\alpha_1, \dots, \alpha_p}(\theta, t)$  and  $\bar{\lambda}_{\alpha_1, \dots, \alpha_p}(\theta, t)$ . We drop  $\theta$  as an argument when dealing with the true value. Null cumulants  $\kappa$  are defined as in McCullagh (1987), Ch. 7.2; for example,  $\kappa_{\alpha, \beta} = \text{Cov}(L_\alpha(T), L_\beta(T))$ ,  $\kappa_{\alpha, \beta, \gamma, \delta} = \text{Cov}(L_{\alpha\beta}(T), L_{\gamma\delta}(T))$  and  $\kappa_{\alpha, \beta, \gamma, \delta} = \text{cum}(L_\alpha(T), L_\beta(T), L_\gamma(T), L_\delta(T))$ .

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  $f_\alpha$  calculated on the basis of  $n - 1$  patients.

*Theorem 3.* Under the assumptions stated,

$$\begin{aligned}\kappa_v &= -E \int_0^T f_v(t) \bar{\lambda}(t) dt, \\ \kappa_{\alpha,v} &= -\sum_i E \int_0^T f_v^{(i)}(t) \bar{\lambda}^{(i)}(t) dt \int_0^t (Z_{\alpha,i}(s) - f_\alpha(s)) \lambda^i(s) ds, \\ \kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta} &= \sum_i E \int_0^T f_{\gamma\delta}^{(i)}(t) \bar{\lambda}^{(i)}(t) dt \int_0^t [f_{\alpha\beta}(s) - (Z_{\alpha,i}(s) - f_\alpha(s))(Z_{\beta,i}(s) - f_\beta(s))] \lambda^i(s) ds\end{aligned}$$

and

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

■

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 Skovgaard (1986) or McCullagh (1987), ex. 7.1 (p. 222)).

**5. Approximation and Estimation of the Cumulants.** We focus on the three first quantities in Theorem 2, 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 McCullagh (1987)). Consider first  $\kappa_{\alpha,v}$  and  $\kappa_{\alpha,\beta,\gamma\delta} + \kappa_{\alpha\beta,\gamma\delta}$ . These are both on the form  $EK$ , where

$$K = \sum_i \int_0^T (f_v(t) \bar{\lambda}(t) - f_v^{(i)}(t) \bar{\lambda}^{(i)}(t)) dt \int_0^t g_i(s) \lambda^i(s) ds,$$

(where  $g_i$  is  $Z_{\alpha,i} - f_\alpha$  or  $f_{\alpha\beta} - (Z_{\alpha,i} - f_\alpha)(Z_{\beta,i} - f_\beta)$ ). This is because  $\sum_i g_i(s) \lambda^i(s) = 0$  in all the cases considered. If we set, for  $w \subseteq v$ ,

$$u^w(\theta, t) = \partial f_v(\theta, t) \bar{\lambda}(\theta, t) / \partial \lambda_w(\theta, t),$$

then

$$K = - \sum_{w,i} \int_0^T u^w(t) \lambda_w^i(t) dt \int_0^t g_i(s) \lambda^i(s) ds + o_p(n) \quad (1)$$

by Taylor expansion. Hence  $K$  is, to first order, and independent sum 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

$$K = EK + o(n). \quad (2)$$

Also, this representation shows that consistent estimation is possible. A natural estimator is

$$\hat{K} = - \sum_i \int_0^T f_v^{(i)}(\tilde{\theta}, t) \left( \bar{\lambda}^{(i)}(\tilde{\theta}, t) / \lambda_0(t) \right) d\hat{\Lambda}_0(t) \int_0^t g_i(\tilde{\theta}, s) \left( \lambda^i(\tilde{\theta}, s) / \lambda_0(s) \right) d\hat{\Lambda}_0(s), \quad (3)$$

where  $\tilde{\theta}$  and  $\hat{\Lambda}_0$  are consistent estimators of  $\theta$  and the baseline hazard, respectively. By the same type of reasoning as in (1),

$$\hat{K} = K + o_p(n), \quad (4)$$

whence consistent estimation of  $EK/n$  can be carried out. Note that we do not assume that  $EK/n$  has a limit as  $n \rightarrow \infty$ , 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 by

$$\hat{\Lambda}_0(t) = \int_0^t \left( \bar{\lambda}(\tilde{\theta}, s) / \lambda_0(s) \right)^{-1} d\bar{N}(s)$$

(cf. Andersen & Gill (1982), p. 1103).

We also suppose that  $\liminf_{n \rightarrow \infty} \sum_{i=1}^n EY_i(t)/n$  is nonzero in an interval in  $t$  containing 0, which is sufficient (in view of our earlier assumptions) to ensure that Conditions A–D on p. 1105 of Andersen & Gill (1982) are satisfied, whence  $\tilde{\theta}$  is consistent and asymptotically normal, and similarly for  $\hat{\Lambda}_0$ .

*Theorem 4.* Under the assumptions stated above and in Section 3, (2) and (4) 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,

$$\hat{\kappa}_v = \kappa_v + o_p(n). \quad (5)$$

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.

**6. The Bartlett Factor, and other likelihood quantities.** In the absence of nuisance parameters, the likelihood ratio statistic  $w = 2(L(\hat{\theta}, T) - L(\theta, T))$  has a Bartlett correction factor  $b(\theta)$  given by (using the summation convention)

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

where  $\lambda_\alpha = -(2\kappa_{\alpha\beta,\gamma} + \kappa_{\alpha\beta\gamma})\kappa^{\beta,\gamma}$  (cf. Ch. 7 of McCullagh (1987)). We have here put the expression in a form where it can immediately be estimated by the methods developed in the previous section.

Does this give good results in practice? Bartlett correction typically works well in small samples, but one might worry about the effect of estimating the Bartlett factor on the performance of the correction. To investigate this, we performed a few numerical experiments, the outcomes of which are reported in tables 1–3. It was assumed in the simulations that there was one covariate (with a standard normal distribution), and that the censoring and the covariate were independent. The null hypothesis tested is  $H_0 : \theta = 0$ . As can be seen from the tables, the correction yielded improvements over the uncorrected  $\chi^2$  test in almost all cases. It should be emphasized that the simulations are done with estimated Bartlett factors (thus mimicking the actual situation of use).

The improvement is particularly strong in the tails. This is presumably due to the connection between Bartlett correction and the saddlepoint approximation, cf. Barndorff-Nielsen and Cox (1984) and Reid (1988).

It should be emphasized that the Bartlett factor is not the only quantity that can be found from the results in the preceding sections. For example, the expressions for bias, variance and third cumulant given on p. 209 in McCullagh (1987) carry over directly. 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  $\kappa_{\alpha, \beta}$ ).

**Appendix: Proofs of Theorems.** In addition to the definitions in Section 3, let  $M^i(t) = N^i(t) - \int_0^t \lambda^i(s) ds$ , with  $\bar{M}(t) = M^1(t) + M^2(t) + \dots + M^n(t)$ .

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

Next,

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

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

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

which yields the desired result since  $dM^i(s) = dN^i(s) - \lambda^i(s) ds$  is the differential of a martingale which is independent of  $f_v^{(i)}(t) \bar{\lambda}^{(i)}(t)$  (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  $L_\alpha(t)L_\beta(t) - [L_\alpha, L_\beta]_t = \int_0^t L_\alpha(s)dL_\beta(s)$  and  $L_{\alpha\beta}(t) + [L_\alpha, L_\beta]_t = \int_0^t [(Z_{\alpha,i}(s) - f_\alpha(s))(Z_{\beta,i}(s) - f_\beta(s)) - f_{\alpha\beta}(s)] dN^i(s)$  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  $\text{Cov}(L_\alpha(t)L_\beta(t) + L_{\alpha\beta}(t), L_{\gamma\delta}(t) + \langle L_\gamma, L_\delta \rangle_t) = 0$ , whence

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

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  $\kappa_{\alpha,v}$ .

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

$$\text{Cov}(\langle L_\alpha, L_\beta \rangle_T, L_{\gamma\delta}(T) + \langle L_\gamma, L_\delta \rangle_T) = - \sum_i E \int_0^T [f_{\alpha\beta}^{(i)}(t)\bar{\lambda}^{(i)}(t) - f_{\alpha\beta}(t)\bar{\lambda}(t)] dt \int_0^t f_{\gamma\delta}(s)\lambda^i(s)ds.$$

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

**Proof of Theorem 4.** (1)–(2) are obvious in view of our assumptions, while (4) follows by the same reasoning as in the proofs of Lemma 3.1 (p. 1105) and Theorem 3.4 (p. 1108) of Andersen & Gill (1982). ■

## REFERENCES

- ANDERSEN, P. K. & GILL, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **4**, 1100–1120.
- ANDERSEN, P. K., BORGAN, Ø., GILL, R. D., & KEIDING, N. (1993). *Statistical Models Based on Counting Processes* (Springer-Verlag, New York).
- BARNDORFF-NIELSEN, O. E., and COX, D. R. (1984). Bartlett adjustments to the likelihood ratio statistic and the distribution of the maximum likelihood estimator. *J. Roy. Statist. Soc. Ser. B.* **46** 483–495.
- CORDEIRO, G. M. (1987). On the corrections to the likelihood ratio statistics. *Biometrika* **74** 265-274.
- 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.
- FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. New York: Wiley.
- GU, M. (1992). On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20**, 399–414.
- GU, M., & ZHENG, Z. (1993). On the Bartlett adjustment for the partial likelihood ratio test in the Cox regression model. *Statistica Sinica* **3**, 543–555.
- HJORT, N. L. (1985). Bootstrapping Cox's regression model. Technical Report No. 241, Department of Statistics, Stanford University.
- Jensen, J. L. (1993). A historical sketch and some new results on the improved log likelihood ratio statistic. *Scand. J. Statist.* **20** 1-15.
- JACOD, J. (1975). Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Zeitschr. Wahr. verw. Geb.* **31** 235-253.
- JACOD, J. & SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Berlin: Springer-Verlag.
- MCCULLAGH, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall.

- MYKLAND, P. A. (1994). Bartlett identities for martingales. *Ann. Statist.* **22**, 21–38.
- MYKLAND, P. A. (1995a). Dual likelihood. To appear in *Ann. Statist.* **23**
- MYKLAND, P. A. (1995b). Embedding and asymptotic expansions for martingales. To appear in *Probab. Theory rel. Fields*
- MYKLAND, P. A. (1995c). Martingale expansions and second order inference. To appear in *Ann. Statist.* **23**
- REID, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **2** 213–227.
- SKOVGAARD, I. (1986). A note on the differentiation of cumulants of log likelihood derivatives. *Int. Statist. Rev.* **54**, 29–32.
- TSIADIS, A. A. (1981), A large sample study of Cox’s regression model. *Ann. Statist.* **9**, 93–108.
- WONG, W. H. (1986). Theory of partial likelihood. *Ann. Statist.* **14**, 88–123.

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

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

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

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

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

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