The Annals of Statistics
1994, Vol. 22, No. 1, 21–38

BARTLETT TYPE IDENTITIES FOR MARTINGALES[1]

BY PER ASLAK MYKLAND

The University of Chicago

Bartlett type identities are shown to exist for martingales. As applications, we give a cumulant-based proof of the martingale central limit theorem, and we give an algorithm for calculating approximate cumulants of the least squares estimator in the AR(1) process.

1. Introduction. The Bartlett identities for moments and cumulants of log likelihood derivatives [Bartlett (1953a, b), Skovgaard (1986), McCullagh (1987)] are a very powerful tool in likelihood inference, leading to some quite general results in that area. For a good description of some of their consequences, see McCullagh (1987), Chapters 7 and 8.

This paper shows that these identities also apply to martingales. The martingale takes the place of the score function, and higher-order derivatives of the log likelihood are replaced by measures of variation of the martingale.

Although the martingale identities are unlikely to yield results which are as powerful as the likelihood identities, we believe that they are a useful tool for both theoretical and computational purposes. As an example of the former, we use them to prove the martingale central limit theorem (Section 4). As an example of the latter, we present an algorithm for calculating cumulants of the least squares estimator in the AR(1) process (Section 5). We also believe the identities will be useful for computations in survival analysis; in fact, special cases of the third Bartlett identity for martingales are given in Hjort [(1985), Lemma A.2] and Gu [(1992), page 411] for use in Cox regression.

The identities themselves are presented in Sections 2 and 6, and in Section 3 we give a heuristic proof based on likelihood theory. The real proof is in Section 7.

2. The Bartlett identities for discrete time martingales. In likelihood inference, these identities concern derivatives of the log likelihood ratio $L_t(\theta)$ with respect to θ (t denotes number of observations or, more generally, time). If the parameter θ is scalar, the two first such identities are $E\dot{L} = 0$ and

Received December 1991; revised March 1993.

[1]Research supported in part by NSF Grants DMS-89-02667 and DMS-92-04504. This manuscript was prepared using computer facilities supported in part by NSF 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.

AMS 1991 *subject classifications*. Primary 60G42, 60G44, 62E99; secondary 60F99, 62E20, 62M09, 62M10.

Key words and phrases. Bartlett identities, central limit theorem, Edgeworth expansions, martingales, time series.

$E\dot{L}^2 + E\ddot{L} = 0$. It then continues

$$E\dot{L}^3 + 3E\dot{L}\ddot{L} + E\dddot{L} = 0, \tag{2.1}$$

$$E\dot{L}^4 + 4E\dot{L}\dddot{L} + 3E\ddot{L}^2 + 6E\dot{L}^2\ddot{L} + E\ddddot{L} = 0 \tag{2.2}$$

and so on by taking further derivatives of the equation $E \exp(L) = 1$. Similar identities hold for cumulants; the fourth Bartlett identity, for example, becomes

$$\operatorname{cum}_4(\dot{L}) + 4 \operatorname{cov}(\dot{L}, \dddot{L}) + 3 \operatorname{var}(\ddot{L}) + 6 \operatorname{cum}(\dot{L}, \dot{L}, \ddot{L}) + E\ddddot{L} = 0. \tag{2.3}$$

In the case of several parameters, one can set

$$U_t(\{\alpha_1, \ldots, \alpha_p\}) = \frac{\partial^p L_t}{\partial\theta_{\alpha_1} \cdots \partial\theta_{\alpha_p}}, \tag{2.4}$$

The Bartlett identities are now as follows: for any set Υ of indices, and subject to regularity conditions,

$$\sum_{\Upsilon} EU_t(v_1) \cdots U_t(v_p) = 0 \tag{2.5}$$

and

$$\sum_{\Upsilon} \operatorname{cum}\left(U_t(v_1), \ldots, U_t(v_p)\right) = 0, \tag{2.6}$$

where the sum extends over all partitions $v_1 | \cdots | v_p$ of the set Υ. Our notation is roughly as in McCullagh (1987); see also Speed (1983) and McCullagh (1984). As can be seen from these references, cumulants can be defined in terms of moments in a similar fashion: If $W_1, \ldots, W_q$ are random variables, then

$$\operatorname{cum}(W_1, \ldots, W_q) = \sum_{\{1, \ldots, q\}} (-1)^{p-1}(p-1)! \prod_{i=1}^{p} E\left(\prod_{j \in v_i} W_j\right). \tag{2.7}$$

Conversely, moments are given from cumulants by

$$E(W_1 \cdots W_q) = \sum_{\{1, \ldots, q\}} \prod_{i=1}^{p} \operatorname{cum}(W_j, j \in v_i). \tag{2.8}$$

Note that if the elements of Υ are not distinct, one has to pretend that they are in order to get the correct coefficients in the sum. A nice explicit notation for displaying low-order Bartlett identities is given on page 202 in McCullagh (1987).

As far as the identities for martingales are concerned, we first discuss the discrete time d-dimensional zero-mean martingale $\ell_t = (\ell_t^1, \ldots, \ell_t^d)$, where

$$\ell_t^\alpha = \sum_{n=1}^{t} X_n^\alpha. \tag{2.9}$$

The *optional variation* of this martingale for an index set $v = \{\alpha_1, \dots, \alpha_p\}$ is

$$[\ell^{\alpha_1}, \dots, \ell^{\alpha_p}]_t = \sum_{n=1}^{t} X_n^{\alpha_1} \cdots X_n^{\alpha_p}. \tag{2.10}$$

Similarly defined is the *cumulant variation*

$$\kappa(\ell^{\alpha_1}, \dots, \ell^{\alpha_p})_t = \sum_{n=1}^{t} \operatorname{cum}(X_n^{\alpha_1}, \dots, X_n^{\alpha_p} | \mathcal{F}_{n-1}), \tag{2.11}$$

where $(\mathcal{F}_t)$ is a filtration with respect to which ℓ_t is a martingale.

The crux of our results is then that (2.5) and (2.6) hold, where the U's can be either

$$U_t(\{\alpha_1, \dots, \alpha_p\}) = (-1)^{p-1}(p-1)![\ell^{\alpha_1}, \dots, \ell^{\alpha_p}]_t \tag{2.12}$$

[note that $U(\{\alpha\}) = \ell_t^\alpha$], or

$$\begin{aligned} U_t(\{\alpha\}) &= \ell_t^\alpha, \\ U_t(\{\alpha_1, \dots, \alpha_p\}) &= -\kappa(\ell^{\alpha_1}, \dots, \ell^{\alpha_p})_t, \quad \text{for } p > 1. \end{aligned} \tag{2.13}$$

Before going into general results, it is worth considering the low-order identities for a scalar martingale. The two first identities are not exactly news, being $E\ell_t = 0$ and (four different variations over the theme) $\operatorname{var}(\ell_t) = E[\ell, \ell]_t$. The third and fourth identities, however, are more interesting. For example, the cumulant identities for cumulant variations are

$$\operatorname{cum}_3(\ell_t) - 3 \operatorname{cov}(\ell_t, \kappa(\ell, \ell)_t) - E\kappa(\ell, \ell, \ell)_t = 0 \tag{2.14}$$

and

$$\begin{aligned} &\operatorname{cum}_4(\ell_t) - 4 \operatorname{cov}(\ell_t, \kappa(\ell, \ell, \ell)_t) + 3 \operatorname{var}(\kappa(\ell, \ell)_t) \\ &\quad - 6 \operatorname{cum}(\ell_t, \ell_t, \kappa(\ell, \ell)_t) - E\kappa(\ell, \ell, \ell, \ell)_t = 0, \end{aligned} \tag{2.15}$$

whereas the cumulant identities for optional variations are

$$\operatorname{cum}_3(\ell_t) - 3 \operatorname{cov}(\ell_t, [\ell, \ell]_t) + 2E[\ell, \ell, \ell]_t = 0 \tag{2.16}$$

and

$$\begin{aligned} &\operatorname{cum}_4(\ell_t) + 8 \operatorname{cov}(\ell_t, [\ell, \ell, \ell]_t) + 3 \operatorname{var}([\ell, \ell]_t) \\ &\quad - 6 \operatorname{cum}(\ell_t, \ell_t, [\ell, \ell]_t) - 6E[\ell, \ell, \ell, \ell]_t = 0. \end{aligned} \tag{2.17}$$

There are, of course, regularity conditions for this to hold. We shall explore these in Section 6. In the process, the results will be related to general càdlàg martingales, thus covering, for example, the martingales occurring in survival analysis.

The main result of the paper is Theorem 3 in Section 6. Specialized to discrete time martingales, it yields the following result.

COROLLARY. *Let ℓ_t be the zero-mean martingale given in* (2.9), *and let U be defined by* (2.12) *or* (2.13). *Suppose, for all $n \leq t$, that $E|X_n^{\alpha_1} \cdots X_n^{\alpha_q}| < \infty$ for all $\{\alpha_1, \dots, \alpha_q\} \subseteq \Upsilon$, and that $E|U_n(v_1) \cdots U_n(v_p)| < \infty$ for all partitions v of Υ. Then* (2.5) *holds. If, in addition, $E|U_n(v_1) \cdots U_n(v_p)| < \infty$ for all partitions v of all subsets of Υ, then* (2.6) *also holds.*

3. Relationship to the likelihood identities. The proof which we are giving for the martingale identities uses stochastic calculus and has little to do with likelihoods. To provide an explanation of why we still refer to them as Bartlett identities, we give a heuristic derivation of them based on likelihood theory. This is done by setting up artificial inference problems.

We consider a one-dimensional discrete time martingale with mean zero,

$$\ell_t = \sum_{n=1}^{t} X_n, \tag{3.1}$$

the argument being similar in the d-dimensional case.

Consider first the identities for optional variations. A derived martingale [which can also be used in proving the CLT, cf. Hall and Heyde (1980), Chapter 3] is

$$m_t(\theta) = \prod_{n=1}^{t} (1 + \theta X_n). \tag{3.2}$$

If we suppose that the X_n's are bounded, $1 + \theta X_n$ is positive for small θ. Since $m_t(\theta)$ has mean 1, it can therefore be viewed as a likelihood function for θ. Also,

$$\begin{aligned} \ln m_t(\theta) &= \sum_{n=1}^{t} \ln(1 + \theta X_n) \\ &= \sum_{n=1}^{t} \sum_{p=1}^{\infty} (-1)^{p-1} \frac{1}{p} \theta^p X_n^p \\ &= \sum_{p=1}^{\infty} (-1)^{p-1} \frac{1}{p} \theta^p \underbrace{[\ell, \dots, \ell]_t}_{p \text{ times}}, \end{aligned} \tag{3.3}$$

where θ^p in this instance denotes the pth power. If we are estimating θ on the basis of the likelihood (3.2), (3.3) then gives that the pth derivative of the log likelihood at $\theta = 0$ is

$$\frac{\partial^p}{\partial \theta^p} \ln m_t(\theta)\bigg|_{\theta=0} = (-1)^{p-1}(p-1)!\,\underbrace{[\ell, \ldots, \ell]_t}_{p \text{ times}}, \tag{3.4}$$

which is the same as (2.12). One then gets (2.5) and (2.6) from the Bartlett identities for likelihoods.

The argument for the cumulant variations is similar. Let $K_n(\theta)$ be the conditional cumulant generating function of X_n given $\mathcal{F}_{n-1}$, that is,

$$K_n(\theta) = \frac{\theta^2}{2!}\operatorname{cum}(X_n, X_n | \mathcal{F}_{n-1}) + \frac{\theta^3}{3!}\operatorname{cum}(X_n, X_n, X_n | \mathcal{F}_{n-1}) + \cdots. \tag{3.5}$$

Since (again supposing that X_n is bounded)

$$E\big[\exp(\theta X_n - K_n(\theta)) | \mathcal{F}_{n-1}\big] = 1 \tag{3.6}$$

for θ in a neighborhood of 0, it follows that

$$m_t(\theta) = \prod_{n=1}^{t} \exp(\theta X_n - K_n(\theta)) \tag{3.7}$$

is a likelihood in such a neighborhood. Hence

$$\ln m_t(\theta) = \theta \ell_t - \frac{1}{2!}\theta^2 \kappa(\ell, \ell)_t - \frac{1}{3!}\theta^3 \kappa(\ell, \ell, \ell)_t - \cdots \tag{3.8}$$

is a log likelihood, and the derivatives at $\theta = 0$ are ℓ_t, $-\kappa(\ell, \ell)_t$, $-\kappa(\ell, \ell, \ell)_t, \ldots$. This corresponds to (2.13), whence the Bartlett identities for likelihoods yield (2.5) and (2.6) for the cumulant variations.

4. A proof of the martingale central limit theorem. As an example of an application of the Bartlett identities, we show how we can use them to derive central limit results.

Consider a triangular array of discrete time martingales

$$\ell_t^N = \sum_{n=1}^{t} X_n^N, \quad 1 \le t \le t_N. \tag{4.1}$$

We are proposing to give a proof of the following well-known theorem [see, e.g., Hall and Heyde (1980), Theorem 3.2, page 58; and Helland (1982), Theorem 2.5, page 82].

MARTINGALE CENTRAL LIMIT THEOREM (Asymptotically ergodic case). *Suppose, as $N \to \infty$,*

$$[\ell^N, \ell^N]_{t_N} \to_P \sigma^2, \tag{4.2}$$

σ^2 being nonrandom, and that the following asymptotic negligibility condition holds:

$$E \max_{1 \le n \le t_N} |X_n^N| \to 0. \tag{4.3}$$

Then $\ell_{t_N}^N$ converges in law to $N(0, \sigma^2)$.

PROOF. Assume first, for all nonnegative integers, p, q and r, the uniform integrability of products of the form

$$(\ell_{t_N}^N)^p \left(\max_{1 \le n \le t_N} |X_n^N| \right)^q [\ell^N, \ell^N]_{t_N}^r. \tag{4.4}$$

We shall undo this condition afterwards.

We first establish the limiting behavior of joint cumulants of $\ell_{t_N}^N$ and optional variations $[\ell^N, \dots, \ell^N]_{t_N}$. Since, for $k \ge 2$,

$$\Bigl| \underbrace{[\ell^N, \dots, \ell^N]_{t_N}}_{k \text{ times}} \Bigr| \le \left(\max_{1 \le n \le t_N} |X_n^N| \right)^{k-2} [\ell^N, \ell^N]_{t_N}, \tag{4.5}$$

it follows from the uniform integrability of terms of the form (4.4) that such cumulants are always well-defined and that the limit of any such cumulant is the cumulant of the limit. Since (4.2)–(4.3) and (4.5) also imply that $[\ell^N, \dots, \ell^N]_{t_N} \to_P 0$ for $k \ge 3$, one can conclude that all cumulants involving optional variations of order at least 2 converge to zero, with the exception of $E[\ell^N, \ell^N]_{t_N}$, which converges to σ^2.

Turning next to the cumulants of $\ell_{t_N}^N$, consider first $\mathrm{cum}_p(\ell_{t_N}^N)$ for $p \ge 3$. The identity (2.6) for optional variations reexpresses this cumulant as a sum of cumulants, each of which involves at least one optional variation. Also, no term is of the form $E[\ell^N, \ell^N]_{t_N}$. Hence, by the above, all these terms tend to zero, and so $\mathrm{cum}_p(\ell_{t_N}^N) \to 0$. On the other hand, $\mathrm{var}(\ell_{t_N}^N) = E[\ell^N, \ell^N]_{t_N} \to \sigma^2$ and $E\ell_{t_N}^N = 0$. Since $(\ell_{t_N}^N)^p$ is uniformly integrable, it follows that $\ell_{t_N}^N$ converges to $N(0, \sigma^2)$.

It then only remains to deal with the assumption of uniform integrability for terms of the form (4.4). Assume first that

$$\sup_N \max_{1 \le n \le t_N} |X_n^N| \le C, \tag{4.6}$$

C being nonrandom. One can then, without loss of generality, also assume that $[\ell^N, \ell^N]_{t_N}$ is bounded by a nonrandom quantity, since $\ell_{t_N}^N$ can be replaced by $\ell_{\tau_N}^N$, where

$$\tau_N = \inf\{t\colon [\ell^N, \ell^N]_t > \sigma^2 + 1\} \wedge t_N. \tag{4.7}$$

This is because (4.2) implies that $P(\tau_N \neq t_N) \to 0$. The uniform integrability of (4.4) then follows from Burkholder's inequality [see, e.g., Hall and Heyde (1980), Theorem 2.10 (page 23)].

Finally, (4.6) can be assumed without loss of generality by embedding $\ell_1^N, \ldots, \ell_{t_N}^N$ in a martingale $\bar{\ell}_t^N, 0 \leq t \leq t_N$, which lives in continuous time and has continuous sample paths [cf. Heath (1977)]. This martingale is stopped at the time σ_N when $|\bar{\ell}_t^N - \bar{\ell}_{\lfloor t \rfloor}^N|$ exceeds C (σ_N being t_N if this never happens), $\lfloor t \rfloor$ being the integer part of t. One then replaces the martingale $\ell_1^N, \ldots, \ell_{t_N}^N$ by $\bar{\ell}_{1\wedge\sigma_N}^N, \ldots, \bar{\ell}_{t_N\wedge\sigma_N}^N$. This can be done since

$$\begin{aligned} P(\sigma_N \neq t_N) &= P\left(|\bar{\ell}_{\sigma_N}^N - \bar{\ell}_{\lfloor\sigma_N\rfloor}^N| \geq C\right) \\ &\leq \frac{1}{C} E\,|\bar{\ell}_{\sigma_N}^N - \bar{\ell}_{\lfloor\sigma_N\rfloor}^N| \\ &\leq \frac{1}{C} E\,|\bar{\ell}_{\lceil\sigma_N\rceil}^N - \bar{\ell}_{\lfloor\sigma_N\rfloor}^N| \\ &\leq \frac{1}{C} E \max_{1\leq n\leq t_N} |X_n^N| \\ &\to 0, \end{aligned} \tag{4.8}$$

$\lceil\sigma_N\rceil$ being the smallest integer exceeding σ_N. Here we have used that since $\lfloor\sigma_N\rfloor$ is measurable with respect to $\bar{\mathcal{F}}_{\sigma_N}^N$ [$(\bar{\mathcal{F}}_t^N)$ being the filtration generated by $(\bar{\ell}_t^N)$],

$$E\left(|\bar{\ell}_{\lceil\sigma_N\rceil}^N - \bar{\ell}_{\lfloor\sigma_N\rfloor}^N| \;\middle|\; \bar{\mathcal{F}}_{\sigma_N}^N\right) \geq |\bar{\ell}_{\sigma_N}^N - \bar{\ell}_{\lfloor\sigma_N\rfloor}^N|. \tag{4.9}$$

This completes the proof. □

5. Inference in the AR(1) process. Consider the process

$$X_{t+1} = \theta X_t + \varepsilon_{t+1}, \quad t = 0, 1, 2, \ldots, \tag{5.1}$$

where $|\theta| < 1$ and the ε's are i.i.d. with mean zero. The least squares estimator of θ is given by

$$\hat{\theta}_t = \frac{\sum_{n=0}^{t-1} X_n X_{n+1}}{\sum_{n=0}^{t-1} X_n^2} \quad \text{and} \quad \hat{\theta}_t - \theta = \frac{\ell_t}{\sum_{n=0}^{t-1} X_n^2},$$

where ℓ_t is the martingale given by $\ell_t = \sum_{n=0}^{t-1} X_n \epsilon_{n+1}$.

Suppose one wants to find approximate cumulants of $\sqrt{t}(\hat{\theta}_t - \theta)$. This reduces to finding cumulants of ℓ_t and $\sum_{n=0}^{t-1} X_n^2$. We shall describe how this can be done using the martingale Bartlett identities.

The calculations involved in finding these cumulants are of a nontrivial complexity. The third cumulants of ℓ_t and $\sqrt{t}(\hat{\theta}_t - \theta)$ are known [cf. Phillips (1978) and McCullagh (1987), Example 3.13 (page 83), for Gaussian ε and Mykland (1993) for general ε], but the fourth cumulant is only known in the Gaussian case [Phillips (1978)]. Higher-order cumulants are not known.

We shall here demonstrate that one can use symbolic manipulation software to find formulas for all the relevant cumulants as a function of t, θ, X_0 and F (F being the distribution of ε). In other words, one can, for example, algorithmically find the symbolic expression for the function $\psi(t,\theta,X_0,F) = \mathrm{cum}_5(\ell_t)$. This is different from finding the expression for ψ at a fixed numerical value of the argument t [finding, say, $\psi(100,\theta,X_0,F)$ as a function of three symbolic arguments]. This latter task can presumably be accomplished by representing X_t as $\theta^t X_0 + \sum_{n=1}^{t} \theta^{t-n}\varepsilon_n$. It is not clear, however, how this approach could be used to algorithmically find ψ as a symbolic function of (t,θ,X_0,F).

To describe the algorithm for doing this, let $\theta \neq 0$, and set

$$\ell_t^{(\alpha,\beta,\gamma,\delta)} = \sum_{n=0}^{t-1} \theta^{\alpha(t-1-n)+\gamma}(t-n)^\beta X_n^\gamma\left(\varepsilon_{n+1}^\delta - E\varepsilon^\delta\right) \tag{5.2}$$

and

$$f_t^{(\alpha,\beta,\gamma)} = \sum_{n=0}^{t-1} \theta^{\alpha(t-1-n)+\gamma}(t-n)^\beta X_n^\gamma. \tag{5.3}$$

Cumulant variations are now given by (for $p > 1$)

$$\begin{aligned} &\kappa\left(\ell^{(\alpha_1,\beta_1,\gamma_1,\delta_1)},\ldots,\ell^{(\alpha_p,\beta_p,\gamma_p,\delta_p)}\right)_t \\ &\quad= \mathrm{cum}\left(\varepsilon^{\delta_1},\ldots,\varepsilon^{\delta_p}\right) f_t^{(\alpha_1+\cdots+\alpha_p,\,\beta_1+\cdots+\beta_p,\,\gamma_1+\cdots+\gamma_p)}. \end{aligned} \tag{5.4}$$

On the other hand, f_t's can be represented recursively in terms of ℓ_t's and other f_t's, as follows (see Section 8 for a derivation). For $\gamma \geq 1, \alpha \neq \gamma$,

$$\begin{aligned} f_t^{(\alpha,\beta,\gamma)} &= \theta^{\gamma-\alpha}\left(1-\theta^{\gamma-\alpha}\right)^{-1} \\ &\times \Bigg[\theta^{\alpha t}t^\beta X_0^\gamma - \theta^{\gamma t}X_0^\gamma \delta_{\beta,0} + \sum_{k=0}^{\beta-1}\binom{\beta}{k}(-1)^{\beta-k} f_t^{(\alpha,k,\gamma)} \\ &+ \sum_{k=0}^{\beta}\sum_{j=0}^{\gamma-1}\binom{\beta}{k}\binom{\gamma}{j}(-1)^{\beta-k}\left(\ell_t^{(\alpha,k,j,\gamma-j)} + f_t^{(\alpha,k,j)}E\varepsilon^{\gamma-j}\right) \\ &- \delta_{\beta,0}\sum_{j=0}^{\gamma-1}\binom{\gamma}{j}\left(\ell_t^{(\gamma,0,j,\gamma-j)} + f_t^{(\gamma,0,j)}E\varepsilon^{\gamma-j}\right)\Bigg], \end{aligned} \tag{5.5}$$

where $\delta_{\beta,0}$ is the Kronecker delta. For $\alpha = \gamma$,

$$\begin{aligned} f_t^{(\alpha,\beta,\alpha)} &= X_0^\alpha \theta^{\alpha t} f_t^{(0,\beta,0)} \\ &+ \sum_{j=0}^{\alpha-1}\sum_{i=1}^{\beta+1}\binom{\alpha}{j}\frac{1}{\beta+1}\binom{\beta+1}{i}B_{\beta+1-i}\left(\ell_t^{(\alpha,i,j,\alpha-j)} + f_t^{(\alpha,i,j)}E\varepsilon^{\alpha-j}\right) \\ &- \delta_{\beta,0}\sum_{j=0}^{\alpha-1}\binom{\alpha}{j}\left(\ell_t^{(\alpha,0,j,\alpha-j)} + f_t^{(\alpha,0,j)}E\varepsilon^{\alpha-j}\right), \end{aligned} \tag{5.6}$$

where B_k is the kth Bernoulli number [see, e.g., Gradshteyn and Ryzhik (1980), Sections 0.121, 9.61 and 9.71 (pages 1 and 1026–1080)]. Boundary conditions are given by the values of $f_t^{(\alpha,\beta,0)}$;

$$f_t^{(0,\beta,0)} = \frac{t^{\beta+1}}{\beta+1} + \frac{t^\beta}{2} + \sum_{k=1}^{\beta-1} \frac{1}{k+1}\binom{\beta}{k} B_{k+1} t^{\beta-k}, \tag{5.7}$$

with $f_t^{(0,0,0)} = t$, and, for $\alpha \neq 0$,

$$f_t^{(\alpha,\beta,0)} = \sum_{k=0}^{\beta} \mathcal{S}_\beta^{(k)} \left[k! \frac{\theta^{\alpha(k+1)}}{(1-\theta^\alpha)^{k+1}} - \sum_{j=0}^{k\wedge(t+1)} \frac{(t+1)!(k-j)!\theta^{\alpha(t+1+k-j+1)}}{(t+1-j)!(1-\theta^\alpha)^{k-j+1}} \right] - 1, \tag{5.8}$$

where the $\mathcal{S}_\beta^{(k)}$ are Stirling numbers of the second kind [see, e.g., Abramowitz and Stegun (1964), Section 24.1.4 (pages 824–825)].

Assume first that X_0 is nonrandom. The algorithm for computing cumulants is then roughly as follows:

ALGORITHM 1. This computes

$$\operatorname{cum}\left(\ell_t^{(\alpha_1,\beta_1,\gamma_1,\delta_1)}, \ldots, \ell_t^{(\alpha_p,\beta_p,\gamma_p,\delta_p)}, f_t^{(\zeta_1,\eta_1,\iota_1)}, \ldots, f_t^{(\zeta_q,\eta_q,\iota_q)}\right). \tag{5.9}$$

Step 1. If $(p,q) = (1,0)$ or if one of the ι's has the value 0, the algorithm terminates and returns the obvious value.

Step 2. If $q = 0$, use (5.4) and the Bartlett identities (2.6) with cumulant variations to reexpress the desired cumulant as a sum of terms of the form (5.9). Then call Algorithm 1 recursively for each term. The algorithm returns the sum of these values and terminates.

Step 3. Reexpress $f_t^{(\zeta_1,\eta_1,\iota_1)}$ with the help of (5.5) or (5.6). Expand the desired cumulant as a linear combination of cumulants. Cumulants then involving X_0 or $f_t^{(\zeta_1,\eta_1,0)}$ are 0, unless they are of order 1 [in the case of $f_t^{(\zeta_1,\eta_1,0)}$, (5.7) and (5.8) are then used]. For the rest, call Algorithm 1 recursively for each one. The combined results are returned, and the algorithm terminates.

To see that this algorithm converges, consider

$$\rho = p + q + \sum(\beta_i + 2\gamma_i) + \sum(\eta_i + 2\iota_i). \tag{5.10}$$

It is easy to see that Step 2 or 3 decrements ρ by at least 1 before recursively calling the algorithm. This shows the result since the algorithm will terminate at the latest when $\rho = 1$.

If one does not want to assume that X_0 is nonrandom, one can set up a dummy time -1, with $\mathcal{F}_{-1}$ as a 0–1 σ-field. One can then treat $0, X_0 - EX_0, X_0 - EX_0, \ldots$ as a martingale which does not evolve after time zero, and the algorithm extends easily to take care of this. Alternatively, of course, one can use the formula for computing cumulants from conditional cumulants [Brillinger (1969), Speed (1983)].

The algorithm can be carried out in MACSYMA or Mathematica [see, e.g., Heller (1991) and Wolfram (1991), respectively]. For example, we used it in MACSYMA to find the fourth cumulant of ℓ_t to first order:

$$
\begin{aligned}
\mathrm{cum}_4(\ell_t) = t\Big\{ & 9(E\varepsilon^2)^2 \text{ var } (\varepsilon^2)(1-\theta^2)^{-2} \\
& + 60(E\varepsilon^2)^4\theta^2(1-\theta^2)^{-3} \\
& + \mathrm{cum}_4(\varepsilon)(1-\theta^4)^{-1}(E\varepsilon^4 + 6\theta^2(1-\theta^2)^{-1}(E\varepsilon^2)^2) \\
& + 12(E\varepsilon^3)^2E\varepsilon^2(\theta^2(1-\theta^3)^{-2} + 2\theta(1-\theta^2)^{-1}(1-\theta^3)^{-1})\Big\} \\
& + o(t).
\end{aligned} \tag{5.11}
$$

A procedure similar to Algorithm 1 can be used to find moments. In Step 2, we then use (2.5) rather than (2.6), but still with the cumulant variations.

6. The Bartlett identities for general martingales.

6.1. *Variation measures.* Our discussion in the following draws on the "general theory of processes." The main reference that we shall use for this is Chapter 1 of Jacod and Shiryaev (1987); in particular, we shall use without further definition a number of concepts as they are defined in that work, such as *càdlàg* (page 3), *local martingale* (page 11), *semimartingale* (page 43) and *compensator* (pages 32–33). We also assume the "usual conditions" of right continuity and completeness (page 2, Definitions 1.2 and 1.3). These can often be dispensed with, however; see the remarks at the end of this section.

DEFINITION. Suppose that $Y_t = (Y_t^1, \dots, Y_t^d)$ is a càdlàg semimartingale. The *optional variation* for Y_t and an index set $v = \{\alpha_1, \dots, \alpha_p\}$ is the càdlàg version (modification) of the process defined by

$$
[Y^{\alpha_1}, \dots, Y^{\alpha_p}]_t = \lim_{\max(t_{i+1}-t_i)\downarrow 0} \sum_i \prod_{\alpha \in v} (Y^\alpha_{t_{i+1}} - Y^\alpha_{t_i}), \tag{6.1}
$$

where $0 = t_0, t_1, t_2, \dots$ are partitions of $[0, t]$. We shall at times refer to $[Y^{\alpha_1}, \dots, Y^{\alpha_p}]_t$ by the symbol $[Y; v]_t$.

The following result should clarify the structure of the optional variation.

PROPOSITION 1. *The optional variations are well defined for any càdlàg semimartingale Y_t, in the sense that the limit in* (6.1) *is independent of the sequence of partitions and has a càdlàg modification. The optional variations* *are semimartingales, and their form is*

$$p = 1\colon\ [Y^\alpha]_t = Y^\alpha_t \tag{6.2}$$

$$p = 2\colon\ [Y^\alpha, Y^\beta]_t = \langle Y^{\alpha,c}, Y^{\beta,c}\rangle_t + \sum_{s\le t} \Delta Y^\alpha_s\ \Delta Y^\beta_s \tag{6.3}$$

$$p \ge 3\colon\ [Y^{\alpha_1}, \dots, Y^{\alpha_p}]_t = \sum_{s\le t} \Delta Y^{\alpha_1}_s \cdots \Delta Y^{\alpha_p}_s, \tag{6.4}$$

$Y^{\alpha,c}_t$ *being the continuous martingale part of* Y^α_t.

For the notion of continuous martingale part, see Jacod and Shiryaev [(1987), Proposition 4.27 (page 45)]. Note that (6.2) is obvious and (6.3) is a standard result [Jacod and Shiryaev (1987), Theorems 4.47 (page 52) and 4.52 (page 55)]. The proof of (6.4) is similar; for completeness we have included the argument among the proofs in Section 7. Note that if the partitions are taken to be $0, t_1 \wedge t, t_2 \wedge t, \dots$, for all t, the convergence in (6.1) can be made uniform on compacts in probability (see the proof of Proposition 1).

There are two more variations to be defined.

DEFINITION. Let $\ell_t = (\ell^1_t, \dots, \ell^d_t)$ be a càdlàg local martingale. If $v = \{\alpha_1, \dots, \alpha_p\}$ is an index set, the *predictable variation* $\langle \ell^{\alpha_1}, \dots, \ell^{\alpha_p}\rangle_t$ or $\langle \ell; v\rangle_t$ is the compensator of $[\ell;v]_t$. The *cumulant variation* $\kappa(\ell^{\alpha_1}, \dots, \ell^{\alpha_p})_t$ or $\kappa(\ell;v)_t$ is given by

$$\kappa(\ell^{\alpha_1}, \dots, \ell^{\alpha_p})_t = \sum_v (-1)^{q-1}(q-1)!\,[\langle \ell;v_1\rangle, \dots, \langle \ell; v_q\rangle]_t, \tag{6.5}$$

where the sum extends over all partitions $v_1 | \cdots | v_q$ of v.

The cumulant variation $\kappa(\ell;v)$ relates to $\langle \ell;v\rangle$ the way cumulants relate to moments [cf. (2.7)].

The criteria for existence are as in Section 3b (pages 32–35) of Jacod and Shiryaev (1987), and we can invert (6.5) using the cumulant identity (2.8) [see, e.g., Speed (1983) and McCullagh (1984, 1987)].

PROPOSITION 2. *If* $|v| \ge 2$, *then* $\langle \ell;v\rangle_t$ *is defined if* $[\ell;v]_t$ *is of locally integrable variation, while* $\kappa(\ell;v)_t$ *is defined if* $[\ell;w]_t$ *is of locally integrable variation for all* $w \subseteq v, |w| \ge 2$. *Under the conditions stated for their existence,* $\langle \ell;v\rangle_t$ *and* $\kappa(\ell;v)_t$ *are predictable processes of locally integrable variation. Also, under the conditions stated for the existence of* $\kappa(\ell;v)_t$,

$$\langle \ell;v\rangle_t = \sum_v [\kappa(\ell;v_1), \dots, \kappa(\ell;v_p)]_t, \tag{6.6}$$

where the sum *extends over all partitions* $v_1 | \cdots | v_p$ *of* v.

Using formulas (6.2)–(6.5) on a martingale of the form (2.9), the definitions clearly reduce to (2.10) and (2.11), in the latter case with the help of cumulant identities and since

$$\langle \ell^{\alpha_1}, \ldots, \ell^{\alpha_p} \rangle_t = \sum_{n=1}^{t} E\big(X_n^{\alpha_1} \cdots X_n^{\alpha_p} \,|\, \mathcal{F}_{n-1}\big). \tag{6.7}$$

The opposite extreme occurs when ℓ_t is quasi-left continuous, in which case $\kappa(\ell;v)_t = \langle \ell;v \rangle_t$. This happens, notably, in the survival analysis setup. For the Nelson–Aalen estimator, for example [see, e.g., Aalen (1978) or Fleming and Harrington (1991)], the "score function" is

$$\ell_t = \int_0^t Y_s^{-1}\, dN_s - \int_0^t \lambda_s\, ds, \tag{6.8}$$

N_t being a counting process with intensity $Y_t \lambda_t$ (Y_t being predictable and λ_t nonrandom). Up to the time when Y_t becomes zero, ℓ_t is a martingale, and its variations are given by

$$\underbrace{[\ell, \ldots, \ell]_t}_{p \text{ times}} = \int_0^t Y_s^{-p}\, dN_s \tag{6.9}$$

and

$$\underbrace{\langle \ell, \ldots, \ell \rangle_t}_{p \text{ times}} = \int_0^t Y_s^{1-p} \lambda_s\, ds, \tag{6.10}$$

the latter being also the value of $\kappa(\ell, \ldots, \ell)_t$.

In the special case where ℓ_t is a continuous martingale, $[\ell; v]_t = \kappa(\ell;v)_t = \langle \ell;v \rangle_t = 0$ for $|v| > 2$.

6.2. *The Bartlett identities.* Consider the process

$$M_t(\Upsilon) = \sum_{\Upsilon} U_t(v_1) \cdots U_t(v_p), \tag{6.11}$$

where U_t is defined either by (2.12) or (2.13). The basic result is now that $M_t(\Upsilon)$ is a local martingale.

THEOREM 1. *Suppose that ℓ_t is a càdlàg local martingale and that U_t is defined by* (2.12). *Then $M_t(\Upsilon)$ is a local martingale.*

THEOREM 2. *Suppose that ℓ_t is a càdlàg local martingale and that U_t is defined by* (2.13). *Assume that $[\ell;v]_t$ is of locally integrable variation for all $v \subseteq \Upsilon, |v| \geq 2$. Then $M_t(\Upsilon)$ is a local martingale.*

Getting (2.5) and (2.6) to hold now requires $\ell_0 = 0$ and is otherwise purely a matter of integrability conditions. First of all, $EM_t(\Upsilon) = 0$ if the set

$$A = \{M_\tau(\Upsilon), \tau \text{ stopping time, } \tau \leq t\} \tag{6.12}$$

is uniformly integrable [cf. Jacod and Shiryaev (1987), Proposition 1.47 (pages 11–12). The following is then obvious, the cumulant part following from the moment part with the help of the method outlined in Example 7.1 (page 222) of McCullagh (1987).

THEOREM 3. *Let* $\ell_0 = 0$. *Let* $M_s(\Upsilon), 0 \le s \le t$, *be defined by* (2.12) [*respectively*, (2.13)], *and assume that the conditions of Theorem* 1 (*respectively, Theorem* 2) *hold up to time t. Suppose that the set A in* (6.12) *is uniformly integrable. If, for all partitions v of* Υ,

$$E|U_t(v_1)\cdots U_t(v_p)| < \infty, \tag{6.13}$$

then (2.5) *holds. Similarly, if* (6.13) *holds for all partitions v of all subsets of* Υ, *then* (2.6) *holds.*

Finally, note that with respect to the "usual conditions" mentioned at the beginning of this section, these are clearly unnecessary in Theorems 1–3 provided ℓ_t and the (optional or cumulant) variations can be taken to be càdlàg anyway.

7. Proofs for Section 6. A main tool in the proofs will be Itô's formula. This result comes in several versions; we shall use the general semimartingale form. For reference, see Jacod and Shiryaev [(1987), Theorem 4.57 (page 57)]. Another feature which will be common to the proofs is that we shall assume that all index sets have only distinct elements. This assumption is notationally convenient and without loss of generality.

PROOF OF PROPOSITION 1. As mentioned just after the statement of this result, we only need to prove the proposition for $p \ge 3$. We shall prove the stronger statement of uniform convergence on compacts in probability, that is,

$$\sup_{0\le u\le t}\left|\sum_i \prod_{\alpha\in v} (Y^\alpha_{t_{i+1}\wedge u} - Y^\alpha_{t_i\wedge u}) - \sum_{s\le u}\prod_{\alpha\in v}\Delta Y^\alpha_s\right| \to_P 0. \tag{7.1}$$

Itô's formula yields that

$$\begin{aligned} d\prod_{\alpha\in v}(Y^\alpha_s - Y^\alpha_{t_i}) &= \prod_{\alpha\in v}\Delta Y^\alpha_s + \sum_{\beta\in v}\prod_{\substack{\alpha\in v\\ \alpha\ne\beta}}(Y^\alpha_{s-} - Y^\alpha_{t_i})\,dY^\beta_s \\ &\quad + \sum_{\substack{\beta,\gamma\in v\\ \beta\ne\gamma}}\prod_{\substack{\alpha\in v\\ \alpha\ne\beta,\gamma}}(Y^\alpha_{s-} - Y^\alpha_{t_i})\,d\langle Y^{\beta,c}, Y^{\gamma,c}\rangle_s \\ &\quad + \sum_{\substack{w\subseteq v\\ |w|\ge 2\\ |w|<|v|}}\prod_{\alpha\in w}\Delta Y^\alpha_s \prod_{\beta\in v-w}(Y^\beta_{s-} - Y^\beta_{t_i}). \end{aligned} \tag{7.2}$$

As the Y_s^α's are càdlàg, the processes $g_s^\alpha = Y_{s-}^\alpha - Y_{t_i}^\alpha$ converge to zero and are almost surely bounded. Thus, if A_s is càdlàg and of finite variation,

$$\int_0^u g_s^{\alpha_1} \cdots g_s^{\alpha_q} \, dA_s \to 0 \quad \text{uniformly in } u \in [0,t] \tag{7.3}$$

almost surely. Now write

$$Y_s^\alpha = B_s^\alpha + k_s^\alpha, \tag{7.4}$$

where B_s^α is càdlàg, adapted and of finite variation and k_s^α is a local square integrable martingale [cf. the Corollary on page 104 in Protter (1990)]. By (7.3) and Lenglart's inequality [see Jacod and Shiryaev (1987), Lemma 3.30 (page 35)],

$$\sup_{0 \le u \le t} \left| \sum_{\beta \in v} \int_0^u \prod_{\substack{\alpha \in v \\ \alpha \neq \beta}} g_s^\alpha \, dk_s^\beta \right| \to_P 0. \tag{7.5}$$

In addition, the integrals of the third, fourth and remaining part of the second term on the right-hand side of (7.2) converge to zero uniformly in $u \in [0,t]$ a.s. by (7.3). Thus, (7.1) is proved. □

PROOF OF THEOREM 1. Itô's formula yields that $M_t(\Upsilon)$ is a semimartingale whose differential is given by

$$\begin{aligned} dM_t(\Upsilon) = \Delta M_t(\Upsilon) &+ \sum_{v \subseteq \Upsilon} M_{t-}(\Upsilon - v) \, dU_t(v) \\ &- \sum_{v \subseteq \Upsilon} M_{t-}(\Upsilon - v) \, \Delta U_t(v) \\ &+ \tfrac{1}{2} \sum_{\substack{\alpha, \beta \in \Upsilon \\ \alpha \neq \beta}} M_{t-}(\Upsilon - \{\alpha, \beta\}) \, d\langle \ell^{\alpha,c}, \ell^{\beta,c} \rangle_t, \end{aligned} \tag{7.6}$$

where $v \subseteq \Upsilon$ indicates the sum over all subsets of Υ except the empty set. To deal with these terms, note first that $dU_t(v) = \Delta U_t(v)$ for $|v| \geq 3$ and that

$$dU_t(\{\alpha, \beta\}) = \Delta U_t(\{\alpha, \beta\}) - d\langle \ell^{\alpha,c}, \ell^{\beta,c} \rangle_t. \tag{7.7}$$

Also,

$$\Delta M_t(\Upsilon) = \sum_{v \subseteq \Upsilon} M_{t-}(\Upsilon - v) \sum_{v} \Delta U_t(v_1) \cdots \Delta U_t(v_q), \tag{7.8}$$

Σ_v being the sum over all partitions of v. Hence, if $m(\cdot,\cdot)$ is the Möbius function for the relevant partition lattice and if 0_v is the finest partition of v, then

$$
\begin{aligned}
\Delta M_t(\Upsilon) &= \sum_{v\subseteq\Upsilon} M_{t-}(\Upsilon - v)\ \Delta[\ell;v]_t \\
&\quad \times \sum_{v}(-1)^{|v_1|-1}\cdots(-1)^{|v_q|-1}(|v_1|-1)!\cdots(|v_q|-1)! \\
&= \sum_{v\subseteq\Upsilon} M_{t-}(\Upsilon - v)\ \Delta[\ell;v]_t \sum_{w\leq v} m(0_v, w) \\
&= \sum_{\alpha\in\Upsilon} M_{t-}(\Upsilon - \{\alpha\})\ \Delta\ell_t^{\alpha}
\end{aligned}
\tag{7.9}
$$

by the Corollary on page 360 of Rota (1964). Combining all this yields that

$$
dM_t(\Upsilon) = \sum_{\alpha\in\Upsilon} M_{t-}(\Upsilon - \{\alpha\})\, d\ell_t^{\alpha},
\tag{7.10}
$$

whence $M_t(\Upsilon)$ is a local martingale [Jacod and Shiryaev (1987), Theorem 4.31 (page 46)]. □

PROOF OF THEOREM 2. Begin by noting that (7.6) and (7.8) remain valid with the new definitions of U_t. We begin by attacking (7.8). For $|v| \geq 2, U_t(v)$ is predictable, and since ℓ_t^{α} and $[\ell;v]_t - \langle\ell;v\rangle_t$ are local martingales, Proposition 4.49 in Jacod and Shiryaev [(1987), page 52)] yields that $[\ell^{\alpha}, U(v_1), \dots, U(v_q)]_t$ and $[[\ell;v] - \langle\ell;v\rangle, U(v_1), \dots, U(v_q)]_t$ are local martingales when $q \neq 0$ and $|v_1|, \dots, |v_q| \geq 2$. Setting

$$
\begin{aligned}
g_t(u) &= \sum_{\substack{u \\ |u_i|\geq 2}} [U_t(u_1), \dots, U_t(u_r)] \\
&= \sum_{u}(-1)^r[\kappa(\ell;u_1), \dots, \kappa(\ell;u_r)]_t,
\end{aligned}
\tag{7.11}
$$

it therefore follows that, for $|v| \geq 2$,

$$
\begin{aligned}
&\sum_{v} \Delta U_t(v_1)\cdots\Delta U_t(v_q) \\
&\quad = \sum_{\alpha\in v} \Delta\ell_t^{\alpha}\, \Delta g_t(v - \{\alpha\}) \\
&\qquad + \sum_{\substack{w\subseteq v \\ |w|\geq 2}} \Delta[\ell;w]_t\, \Delta g_t(v - w) + \Delta[\ell;v]_t + \Delta g_t(v) \\
&\quad = \sum_{w\subseteq v} \Delta\langle\ell;w\rangle_t\, \Delta g_t(v - w) \\
&\qquad + \Delta[\ell;v]_t + \Delta g_t(v) + \text{differential of local martingale.}
\end{aligned}
\tag{7.12}
$$

Substituting (7.12) into (7.8), and putting the resulting expression into (7.6) yields

$$dM_t(\Upsilon) = \sum_{v \subseteq \Upsilon} M_{t-}(\Upsilon - v)\, dZ_t(v), \tag{7.13}$$

where

$$\begin{aligned} dZ_t(v) &= \sum_{w \subseteq v} \Delta\langle \ell;w\rangle_t\, \Delta g_t(v-w) + d\langle \ell;v\rangle_t + \Delta g_t(v) \\ &\quad - d\kappa(\ell;v)_t + \Delta\kappa(\ell;v)_t + \text{differential of local martingale.} \end{aligned} \tag{7.14}$$

Here we have also used Proposition 1 to reexpress $\Delta[\ell;v]_t$ in terms of $d[\ell;v]_t$, together with the fact that $[\ell;v]_t - \langle \ell;v\rangle_t$ is a local martingale. Using Proposition 1 on (6.5) also yields that

$$d\kappa(\ell;v)_t - \Delta\kappa(\ell;v)_t = d\langle \ell;v\rangle_t - \Delta\langle \ell;v\rangle_t. \tag{7.15}$$

On the other hand, however, (6.6) and the definition of g_t yield, by a combinatorial argument, that

$$\sum_{w \subseteq v} \Delta\langle \ell;w\rangle_t\, \Delta g_t(v-w) + \Delta\langle \ell;v\rangle_t + \Delta g_t(v) = 0. \tag{7.16}$$

Substituting (7.15) and (7.16) into (7.14) now yields that $dZ_t(v)$ is the differential of a local martingale. Theorem 2 then follows from (7.13) and from Theorem 4.31 of Jacod and Shiryaev [(1987), page 46]. □

8. Some notes on Section 5. Expanding $X_{n+1}^{\gamma} = (\theta X_n + \varepsilon_{n+1})^{\gamma}$ yields that, for $\gamma \geq 1$,

$$\begin{aligned} f_t^{(\alpha,\beta,\gamma)} = \theta^{\gamma-\alpha} \Bigg[& \theta^{\alpha t} t^{\beta} X_0^{\gamma} - X_t^{\gamma} \delta_{\beta,0} \\ & + \sum_{k=0}^{\beta} \sum_{j=0}^{\gamma} \binom{\beta}{k} \binom{\gamma}{j} (-1)^{\beta-k} \left(\ell_t^{(\alpha,k,j,\gamma-j)} + f_t^{(\alpha,k,j)} E\varepsilon^{\gamma-j} \right) \Bigg], \end{aligned} \tag{8.1}$$

where $\ell_t^{(\alpha,k,\gamma,0)} = 0$. For $\alpha \neq \gamma$, this yields

$$\begin{aligned} f_t^{(\alpha,\beta,\gamma)} &= \theta^{\gamma-\alpha} (1 - \theta^{\gamma-\alpha})^{-1} \\ &\quad \times \Bigg[\theta^{\alpha t} t^{\beta} X_0^{\gamma} - X_t^{\gamma} \delta_{\beta,0} + \sum_{k=0}^{\beta-1} \binom{\beta}{k} (-1)^{\beta-k} f_t^{(\alpha,k,\gamma)} \\ &\quad + \sum_{k=0}^{\beta} \sum_{j=0}^{\gamma-1} \binom{\beta}{k} \binom{\gamma}{j} (-1)^{\beta-k} \left(\ell_t^{(\alpha,k,j,\gamma-j)} + f_t^{(\alpha,k,j)} E\varepsilon^{\gamma-j} \right) \Bigg] \end{aligned} \tag{8.2}$$

while for $\alpha = \gamma$ and $\beta = 0$, we get

$$X_t^\alpha = \theta^{\alpha t} X_0^\alpha + \sum_{j=0}^{\alpha-1} \binom{\alpha}{j} \left(\ell_t^{(\alpha,0,j,\alpha-j)} + f_t^{(\alpha,0,j)} E\varepsilon^{\alpha-j} \right). \tag{8.3}$$

Substituting (8.3) into (8.2) gives (5.5), whereas summing over t in (8.3) gives (5.6).

An alternative way of solving the linear equations (5.5)–(5.8) would be to consider

$$g_t^{(\alpha,\beta,\gamma)} = f_t^{(\alpha,\beta,\gamma)} - Ef_t^{(\alpha,\beta,\gamma)} \tag{8.4}$$

and to calculate $g_t^{(\alpha,\beta,\gamma)}$ and $Ef_t^{(\alpha,\beta,\gamma)}$ separately; $g_t^{(\alpha,\beta,\gamma)}$ then satisfies equations (5.5) and (5.6) in place of $f_t^{(\alpha,\beta,\gamma)}$ and with X_0 set to 0. The boundary conditions become $g_t^{(\alpha,\beta,0)} = 0$. $Ef_t^{(\alpha,\beta,\gamma)}$ can be calculated directly from the values of the cumulants of X_n.

If one is only seeking to find approximate cumulants, in the sense of neglecting terms going to zero faster than polynomially, one can use this algorithm with the approximation

$$Ef_t^{(\alpha,\beta,\gamma)} \simeq \theta^\gamma E_\pi X^\gamma f_t^{(\alpha,\beta,0)},$$

where π is the stationary distribution. This is how (5.11) was actually computed.

Acknowledgments. The author would like to thank Siu-Kai Chan, Peter McCullagh, Michael Stein, David Wallace, the two referees and the Editor for helpful advice and comments, and Mitzi Nakatsuka for typing the manuscript.

REFERENCES

AALEN, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ABRAMOWITZ, M. and STEGUN, I. A. (1964). *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*. National Bureau of Standards, Washington, DC.

BARTLETT, M. S. (1953a). Approximate confidence intervals. *Biometrika* **40** 12–19.

BARTLETT, M. S. (1953b). Approximate confidence intervals. II. More than one unknown parameter. *Biometrika* **40** 306–317.

BRILLINGER, D. (1969). The calculation of cumulants via conditioning. *Ann. Inst. Statist. Math.* **21** 375–390.

FLEMING, T. R. and HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis*. Wiley, New York.

GRADSHTEYN, I. S. and RYZHIK, I. M. (1980). *Table of Integrals, Series and Products*. Academic, New York.

GU, M. (1992). On the Edgeworth expansion and bootstrap approximation for the Cox regression model under random censorship. *Canad. J. Statist.* **20** 399–414.

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic, New York.

HEATH, D. (1977). Interpolation of martingales. *Ann. Probab.* **5** 804–806.

HELLAND, I. S. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9** 79–94.

HELLER, B. (1991). *MACSYMA for Statisticians*. Wiley, New York.

HJORT, N. L. (1985). Bootstrapping Cox's regression model. Technical Report 241, Dept. Statistics, Stanford Univ.

JACOD, J. and SHIRYAEV, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer, Berlin.

MCCULLAGH, P. (1984). Tensor notation and cumulants of polynomials. *Biometrika* **71** 461–476.

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

MYKLAND, P. A. (1993). Asymptotic expansions for martingales. *Ann. Probab.* **21** 800–818.

PHILLIPS, P. C. B. (1978). Edgeworth and saddlepoint approximation in the first order noncircular autoregression. *Biometrika* **65** 91–98.

PROTTER, P. (1990). *Stochastic Integration and Differential Equations*. Springer, New York.

ROTA, G. C. (1964). On the foundations of combinatorial theory I. Theory of Möbius functions. *Z. Wahrsch. Verw. Gebiete* **2** 340–368.

SKOVGAARD, I. (1986). A note on the differentiation of cumulants of log likelihood derivatives. *Internat. Statist. Rev.* **54** 29–32.

SPEED, T. P. (1983). Cumulants and partition lattices. *Austral. J. Statist.* **25** 378–388.

WOLFRAM, S. (1991). *Mathematica. A System for Doing Mathematics by Computer*. Addison-Wesley, Reading, MA.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CHICAGO
5734 UNIVERSITY AVENUE
CHICAGO, ILLINOIS 60637