Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

R.D. Gill

A note on two papers in dependent central limit theory

Department of Mathematical Statistics Report MS-R8410 May

A NOTE ON TWO PAPERS IN DEPENDENT CENTRAL LIMIT THEORY

R.D. GILL
Centre for Mathematics and Computer Science, Amsterdam

In this note a discussion is given of two papers on dependent central limit theory which were presented at the 44th meeting of the I.S.I. in Madrid, 1983. The papers are by I.S. Helland, "Applications of Central Limit Theorems for Martingales with Continuous Time", and by T.G. Kurtz, "Gaussian Approximations for Markov Chains and Counting Processes". The note addresses itself to the main differences between the approaches described by Helland and Kurtz, called the martingale approach and the Poisson process approach respectively, and discusses their application to statistical problems in survival analysis, life-testing, demography, epidemiology etc.

1980 MATHEMATICS SUBJECT CLASSIFICATION: 60F17, 62Mxx, 62P10.
KEY WORDS & PHRASES: martingale central limit theorems, Poisson random measures, approximation of population processes, Markov processes, semi-Markov processes.
NOTE: This paper will appear in *Proc. 44th. Session I.S.I. (Madrid), Bull. Int. Stat. Inst.* 50 *(3).*

Report MS-R8410
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

In this discussion of the papers of Helland (1983) and Kurtz (1983) my aim is to bring out the differences and similarities, advantages and disadvantages, of the two approaches described by these authors to deriving dependent central limit theorems. The discussion will be limited to their application in proving central limit theorems for statistical quantities (estimators, test-statistics) in counting process models such as those used in demography, epidemiology, actuarial science, survival analysis and life testing. Models of the occurrence in time of the events of interest in these fields are often "dynamic" in the sense that they describe the instantaneous development of the whole system conditioned on its past development. There is emphasis on hazard rates and intensities; there are often various forms of censoring present (loss to observation as time develops); and models are often non- or semi-parametric. All these factors make a counting process approach very natural.

Considering a univariate counting process $N(t)$ the "martingale approach" exploits the fact that $M(t) = N(t)-\int_0^t \Lambda(s)ds$ is a matingale, where $\Lambda(t)$ times h is approximately $\mathrm{E}(N(t+h) - N(t)|$ past up to time $t)$. We can often model $\Lambda(t)$ as a simple function of parameters and observables. The "Poisson approach" on the other hand uses the representation $N(t) = Y(\int_0^t \Lambda(s)ds)$ where Y is a standard Poisson process. This representation carries exactly the same intuitive meaning as the martingale representation. However, to use it we need to add the assumption that N is "self-exciting"; i.e. $\Lambda(s) = \lambda(s,N_{s-})$ for some function λ, where N_{s-} is the whole path of N up to time s. This representation now defines N as an implicit function of Y.

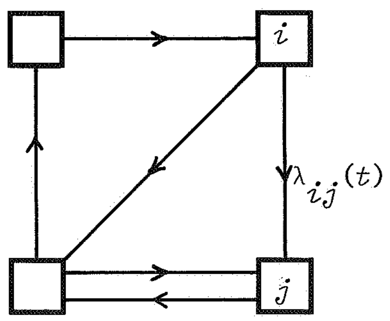
What are the advantages of the martingale approach? When it is applicable it is often extraordinarily easy to apply. Also one need not explicitly model the whole system in order to apply it, i.e. no "self-exciting" assumption is needed. Another advantage is that the method is applicable so often in statistical applications, but this phenomenon calls for explanation. Two facts are relevant here.

Firstly, suppose Z is some test-statistic which we hope will be asymptotically normally distributed with zero mean under the null hypothesis. We will often construct Z so that it is unbiased in the sense that (under the hypothesis) $E(Z) = 0$. We will also be able to compute the test-statistic at a stopping time T, call the result $Z(T)$, and hopefully this will be unbiased in the same sense. But if $EZ(T) = 0$ for all T, $Z(t)$ is automatically a martingale (A similar argument would apply to some estimators).

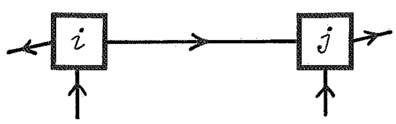
A second source of martingales is the use of likelihood or partial likelihood based methods. Exploiting (a conditional version of) the fact that $E(\partial/\partial\theta \log f(X;\theta)) = 0$ when X has smooth density $f(x,\theta)$, one can show that the derivative with respect to some parameter of the log likelihood (or log partial likelihood) based on the data available at time t (and at the true parameter value) is a martingale in t.

Advantages of the Poisson approach are that, since we consider the objects of interest as a hopefully nice function of a single Poisson process, much extra information can be derived. We could for instance derive rates of convergence; or, by considering the *same* statistic under a *range* of models (parameter values) as different functions of the *same* Poisson process, we could derive results on uniform convergence. Disadvantages are that the functions considered soon become very complicated. A disadvantage of the martingale approach is that it often breaks down when different time scales are involved simultaneously, whereas the Poisson random measure extension of the Poisson approach has much promise here.

Some of these points will hopefully be illustrated by considering a simple example. Consider a finite state space, time inhomogeneous Markov process with intensities $\lambda_{ij}(t)$. Such models are extremely important in the applied fields mentioned above. So we have a number of states i, and a number of individuals or particles moving independently from one state to another, such that h times $\lambda_{ij}(t)$ is approximately the probability that a particle in state i at time t moves to j in the small time interval $(t,t+h)$. The system is started at time 0 with n particles distributed about the system, n_i in state i.



Let $N_{ij}(t)$ (later denoted $N_{ij}^{n}(t)$) denote the number of transitions made from i to j at or before time t. Then $N_{ij}(t)$ has intensity $Y_i(t)\lambda_{ij}(t)$ where $Y_i(t) = n_i + N_{\cdot i}(t-) - N_{i\cdot}(t-)$ is the number of individuals in state i just before time t; a dot denotes summation. Let M_{ij} denote N minus its integrated intensity. Important statistical problems concern e.g. estimation of $\int_0^t \lambda_{ij}(s)ds$ for t in some interval; or testing equality of the λ_{ij}'s corresponding to several different possible transitions. In the estimation problem, the natural estimator is $\int_0(Y_i(s))^{-1}dN_{ij}(s)$. Since in fact $\int_0^t(Y_i(s))^{-1}dN_{ij}(s) - \int_0^t \mathrm{I}\{Y_i(s) > 0\}\ \lambda_{ij}(s)ds$ turns out to be a martingale, weak convergence of $n^{1/2}$ times estimator minus estimand to a Gaussian martingale is proved extremely simply with a martingale CLT one it has been shown that $Y_i(s)/n \xrightarrow{P} y_i(s) > 0$ for each s plus some kind of uniformity condition (see Helland's paper). Note that we need only to consider a perhaps very small part of the model here:



The rest of the model doesn't need to be explicitly considered at all; it doesn't need to have the Markov structure just assumed. There may be entry into i and censoring from i according to almost any mechanism. We just need the martingale property of M_{ij} and the predictability of Y_i.

For some applications we may be interested in very complicated functions of $\{N_{ij}(t)\}$ which are not closely related to any particular martingales and the most feasible approach is to prove weak convergence of $\{N_{ij}(t)\}$ (suitably centred and normalized) and then use a δ-method or von Mises expansion type approach to approximate the object of interest with a linear function of these processes. For instance a recent paper by Borgan and Ramlau-Hansen (1983) on so-called demographic incidence rates uses such an approach. The Poisson approach comes into its own now. Suppose for simplicity that (for all n) all particles start in the same state at time zero. We first consider the model for $n = 1$ and then rescale. Let N be the vector of N_{ij}'s (with n=1). N has (vector) intensity $\lambda(s;N_{s-})$. Rescaling means that we *define* N^n as a process with intensity $n(s;N_{s-}/n)$. We should check that this does yield the process we're interested in (which is not always the case !). Here we have

$$\lambda(s;N_{s-}) = A(s)N(s-) + B(s)$$

where $A(s)$ is a matrix with entries $\pm\lambda_{ij}(s)$ and 0 in appropriate places, and $B(s)$ is a similar looking vector (depending on the initial state). So rescaling gives a process N^n with intensity

$$n(A(s)\,\frac{N^n(s-)}{n} + B(s)) = A(s)\,N^n(s-) + nB(s)$$

exactly as we want. Since $\lambda(s;N_{s-})$ (which is essentially Kurtz's function F, with components corresponding to his functions β_ℓ) is linear in $N(s-)$, all the conditions of Kurtz's theorems 2.1 and 2.2 are trivially satisfied if the λ_{ij} are continuous.

Actually we can imitate Kurtz's proof of this theorem when proving the same result using the martingale approach; i.e. we consider N^n as being "driven" by the (vector) martingale M^n rather than by a collection of Poisson processes (I am indepted to O. Aalen for this idea). This way we can also include censoring in and out of the whole system without essentially changing the method of proof. The idea, using the previous notation (and replacing $N(s-)$ by $N(s)$ as we may), is to write

$$dN^n = AN^n\, dt + nB\, dt + dM^n.$$

Substracting expectations and multiplying by $n^{-1/2}$ gives the equation

$$dZ^n = AZ^n\, dt + dW^n$$

where $Z^n = n^{-1/2}\,(N^n - EN^n)$ and $W^n = n^{-1/2}\, M^n$ is a vector of martingales which is easily shown (if $Y_i^n\,(t)/n \xrightarrow{P} y_i\,(t)$ for all t plus some uniformity condition) to converge in distribution to a vector of independent Gaussian martingales W^∞ say. (In fact var $W_{ij}^\infty(t) = \int_0^t y_i\,(s)\,\lambda_{ij}\,(s)ds$). So taking account of the initial conditions $Z^n(0) = 0$ we would expect Z^n to converge in distribution to Z^∞ satisfying

$$dZ^\infty = AZ^\infty dt + dW^\infty\ ,\ Z^\infty(0) = 0$$

In fact (with all integrals ordinary Lebesgue-Stieltjes integrals) the equation $dZ = AZ\, dt + dW$, $Z(0) = W(0) = 0$ has the explicit solution $Z(t) = \int_{s=0}^{t} \prod_{u=t}^{s} (I+A(u)du)W(ds)$ or equivalently

$$Z(t) = W(t) + \int_{s=0}^{t} \prod_{u=t}^{s} (I + A(u)du)\ A(s)ds\ W(s).$$

so if A is bounded we see that Z is a continuous function of W (in the supremum norm over some finite interval). More information about the product integrals here can be found in Aalen & Johansen (1978). Thus weak convergence of Z^n to Z^∞ is easily established (using a Skorohod construction -cf. Kurtz's thm. 1.1- in order to pretend that $W^n \to W^\infty$ in the supremum norm almost surely).

Before discussing a final example, I would like to make a few comments on Helland's weaker but supposedly more easily verifiable conditions for the martingale CLT's. Certainly his formulation vastly improves the Andersen, Borgan, Gill & Keiding (1982) results he mentions. His discussions turn around the following idea. Suppose for some nonnegative process X^n we know that $X^n(t) \xrightarrow{P} f(t)$ for each t and would like to conclude that $\int_0^1 X^n(t)dt \xrightarrow{P} \int_0^1 f\,(t)dt$. More is needed to guarantee this conclusion so

Helland introduces the concept of "convergence boundedly in L_1", written $X^n \xrightarrow{\text{Lb}} f$, and shows that $X^n \xrightarrow{\text{Lb}} f$ implies $\int_0^1 X^n(t)dt \xrightarrow{\text{P}} \int_0^1 f(t)dt$. In fact $X^n \xrightarrow{\text{Lb}} f$ implies even *more* than $\text{E}(\int_0^1 X^n(t)dt - \int_0^1 f(t)dt) \longrightarrow 0$. So more is being proved that is needed and in fact $X^n \xrightarrow{\text{Lb}} f$ makes integrability conditions which are quite extraneous to Rebolledos theorem (the latter has been later shown to hold for "*local* square integrable martingales", for which M^n need not even be integrable). More importantly, in may applications $X^n \xrightarrow{\text{Lb}} f$ is very hard to verify or even untrue (e.g. in problems involving the Kaplan-Meier estimator when X^n often includes a factor $(1-\hat{F}^n)$ to some power, whose expectation would need to be bounded by a constant times the same power of $(1-F)$). Here is an alternative suggestion (a better one would be welcome!) involving a concept "convergence boundedly in probability".

For a nonnegative process X^n we say $X^n \xrightarrow{\text{Pb}} f$ if

(i) $X^n(s) \xrightarrow{\text{P}} f(s) \quad \forall \ s$

(ii) $\forall \ \delta > 0 \quad \exists \ k_\delta$ such that

$$\liminf_{n\to\infty} \text{P}\{X_n(s) \le k_\delta(s) \ \forall \ s\} \ge 1-\delta \text{ and } \int_0^1 k_\delta(s)ds < \infty$$

It is easily verified that $X^n \xrightarrow{\text{Pb}} f$ implies $\int_0^1 X^n(t)dt \xrightarrow{\text{P}} \int_0^1 f(t)dt$; also replacing $\xrightarrow{\text{Lb}}$ by $\xrightarrow{\text{Pb}}$ in Helland's theorems leaves the conclusions unaltered. (The new conditions are not actually weaker than Helland's; they overlap). My main point is that this new condition is satisfied in a number of situations where the other is not (cf. Andersen & Gill, 1982 or Gill, 1983).

Finally I should like to discuss another very important example where the martingale approach breaks down while the Poisson approach, in its random measures extension, should be applicable (though I have not succeeded in doing this yet). Another similar and important type of example concerns sequential analysis for staggered entry clinical trials (see Sellke & Siegmund, 1983, and Slud, 1984).

The model can be visualized exactly as the previous Markov model except that an individual or particle who entered into a state i at time $t-x$ and is still there at time t has intensity $\lambda_{ij}(x)$ (*not* $\lambda_{ij}(t)$) for leaving state i now and going to j. This is a semi-Markov or Markov renewal process. The intensity depends directly on *age* x not time t where age "starts anew" at zero after each transition into a new state. To fix ideas we consider again n individuals starting at time 0 and age 0 in, say, a fixed state, and suppose the system is observed up to a fixed time τ. Statistical procedures concerning the λ_{ij}'s will analogously (and sensibly) be based in identical fashion as in the Markov model on processes $N_{ij}^n(x)$ and $Y_i^n(x)$, where $N_{ij}^n(x)$ is the number of transitions observed in the *time* interval $[0,\tau]$ from state i to state j at an *age* less than or equal to x at transition; $Y_i^n(x)$ is defined as the number of times an individual is in state i at age x. A minor complication is that $Y_i^n(0)$ is now random, not fixed.

As mentioned before, the martingale approach fails here. A direct application of the Poisson approach (i.e. to processes counting transitions of various types at or before *time* t) fails too since the "rescaled" process isn't the one we are interested in. However the Poisson random measure approach is potentially applicable: cf. Kurtz's final examples which both include consideration of different "time" and "age" dimensions. Unfortunately I haven't succeeded in carrying out this yet.

It is possible to derive CLT's for the processes $\{N_{ij}^n(x)\}$ by brute force; i.e. by directly verifying tightness and using ordinary CLT for finite dimensional distributions, (Gill, 1980). Surprisingly the resulting limiting distribution has identical form to the one it has in the Markov case. If the Poisson approach works, it could give a probabilistic reason for this. A statistical explanation can also be given, as follows. Under a discrete time version of the two models it is clear that the likelihood functions in the two cases have identical form (as functions of N_{ij}^n, Y_i^n and λ_{ij}). The same holds (under the appropriate definitions, which I hope to present elsewhere) in the continuous time models; moreover it turns out that $M_{ij}^n = N_{ij}^n - \int Y_i^n \lambda_{ij}\, dt$ is the functional analogue of the derivative of the log likelihood with respect to the parameter $\int_0^{\cdot} \lambda_{ij}$. So this process, by the analogue of the usual property, has expectation zero in both models. Its covariance structure is also the same in both models by the analogue of the usual relation between the 1st. moments of 2nd. derivatives and 2nd. moments of 1st. derivatives. Thus in the semi-Markov case, $\{N^n_{ij}\}$ is "driven" by processes $\{M^n_{ij}\}$ which, though not martingales, have zero means and uncorrelated increments, and can be shown by standard (though not pretty) methods to be asymptotically Gaussian. So in the limit, the centred and normalized $\{N^n_{ij}\}$ process has the same limiting distribution (in terms of y_i and λ_{ij}) as in the Markov case.

REFERENCES

AALEN, O.O. & S. JOHANSEN (1978), *An empirical transition matrix for non-homogeneous Markov chains based on censored observations*, Scand. J. Statist. 5 141-150.

ANDERSEN, P.K. Ø. BORGAN, R.D. GILL & N. KEIDING (1982), *Linear nonparametric tests for comparison of counting processes, with application to censored survival data* (with discussion), Int. Statist. Rev. 50 219-258.

ANDERSEN, P.K. & R.D. GILL (1982), *Cox's regression model for counting processes: a large sample study*, Ann. Statist. 10 1100-1120.

BORGAN, Ø. & H. RAMLAU-HANSEN (1983), *Demographic incidence rates and estimation of intensities with incomplete information*, Stat. Res. Rep. 2, Inst. of Math., Univ. of Oslo (to appear in Ann. Statist.)

GILL, R.D. (1980), *Nonparametric estimation based on censored observations of a Markov renewal process*, Zeit. Wahrsch. u. verw. Geb. 53 97-116.

GILL, R.D. (1983), *Large sample behaviour of the product limit estimator on the whole line*, Ann. Statist. 11 49-58.

HELLAND, I.S. (1983), *Applications of central limit theorems for martingales with continuous time*, Proc. 44th Session I.S.I. (Madrid), Bull. Int. Stat. Inst. 50 (1) 346-360.

KURTZ, T.G. (1983), *Gaussian approximations for Markov chains and counting processes*, Proc. 44th Session I.S.I. (Madrid), Bull. Int. Stat. Inst. 50 (1) 361-376.

SELLKE, T. & D. SIEGMUND (1983), *Sequential analysis of the proportional hazards model*, Biometrika 70 315-326.

SLUD, E.V. (1984), *Sequential linear rank tests for two-sample censored survival data*, Ann. Statist. 12 (to appear).