Chris Heyde's Contribution to Inference in Stochastic Processes

Ishwar Basawa[1]

This overview of Heyde's research on inference in stochastic processes is based on a subjective sampling of his extensive publications in this area.

The question of asymptotic efficiency of the maximum likelihood (ML) estimator was first addressed by R.A. Fisher (1925). Fisher's criterion of efficiency was further explored by C.R. Rao (1962). The Fisher-Rao approach seeks to establish the asymptotic optimality of the maximum likelihood estimator in the class of consistent and asymptotically normal estimators by showing that the ML estimator has the minimum asymptotic variance, under some broad regularity conditions. LeCam (1960) provided a general framework for asymptotics via the local asymptotic normal (LAN) family which formalizes the theoretical basis needed for the Fisher-Rao approach. Even though most of the early work on asymptotic inference was developed for independent and identically distributed (i.i.d.) observations, the theory has since been extended to stochastic processes. The extension to dependent observations mainly involves replacing the classical law of large numbers and central limit theorem for the likelihood score function by their analogues for martingales (Hall and Heyde (1980)).

In [M62], Heyde's early work on asymptotic efficiency in estimation for branching processes led to some unexpected foundational questions about the classical Fisher-Rao-LeCam approach. Suppose $L_n(\theta)$ is the likelihood function based on n dependent observations $(X_1, X_2, \ldots, X_n)$, and

$$\xi_n(\theta) = -\frac{d^2 \ln L_n(\theta)}{d\theta \, d\theta^T}.$$

The Fisher information matrix is denoted by $I_n(\theta) = E(\xi_n(\theta))$. For a large class of processes (including the classical i.i.d. model), typically one can verify that $I_n^{-1}(\theta)\xi_n(\theta)$ converges in probability to a non-random matrix (usually, the identity matrix). When this happens, the classical Fisher-Rao-LeCam asymptotics usually can be extended to dependent observations. However, for branching processes (and many other stochastic processes) the normalized information $I_n^{-1}(\theta)\xi_n(\theta)$ has a non-degenerate limit. For such processes, the Fisher-Rao-LeCam asymptotic optimality criterion no longer applies. In particular, the ML estimator is not asymptotically normal, and the LAN property fails. In [M62], Heyde provided a remarkably simple modification of Rao's (1962) asymptotic efficiency criterion by replacing $I_n(\theta)$ in Rao's criterion by $\xi_n(\theta)$. In [M62], an estimator T_n is defined to be asymptotically efficient if it satisfies:

$$\text{(H)} \qquad \xi_n^{\frac{1}{2}}(\theta)\left[(T_n - \theta) - \sigma_T(\theta)\xi_n^{-1}(\theta)\frac{d \ln L_n(\theta)}{d\theta}\right] \xrightarrow{P} 0,$$

[1]Department of Statistics, University of Georgia, Athens, Georgia 30602, USA

for some positive number σ_T. [M76] shows that the ML estimator $\hat{\theta}_{ML}$ satisfies (H) with σ_T equal to 1, and it maximizes the limiting probability of concentration

$$(W) \qquad \lim_{n \to \infty} P[I_n^{\frac{1}{2}}(\theta) \,|\, T_n - \theta \,|\leq a]$$

for all $a>0$ (see also, Wolfowitz (1965)). Note that the criterion (W) does not require the asymptotic normality of T_n, nor does it need the LAN framework. Moreover, in [M137], Heyde demonstrates the advantages of using the random norm $\xi_n(\theta)$ in constructing asymptotically optimal confidence intervals.

Heyde and Johnstone in [M81] established the asymptotic normality of the posterior density for stochastic processes. Interestingly, this result is applicable for branching processes, thus avoiding the problem of a threshold parameter (the offspring mean of a branching process) drastically affecting the asymptotic inference. Bayesian asymptotic inference can then be based on the asymptotic posterior normality.

In the context of stochastic processes, it is often the case that complete specification of the likelihood function may not be available. It is then not possible to use the ML estimation or Bayesian approach. [B3] gives a comprehensive review of the theory of quasi-likelihood estimation. Let $\mu_t(\theta) = E(X_t \,|\, \mathcal{F}_{t-1})$ and $V_t(\theta) = \mathrm{Var}(X_t \,|\, \mathcal{F}_{t-1})$ denote the conditional mean and variance respectively. A typical quasi-likelihood (QL) estimating equation is defined by

$$\sum_{t=1}^{n} V_t^{-1}(\theta) \left(\frac{d\mu_t(\theta)}{d(\theta)} \right)(X_t - \mu_t(\theta)) = 0.$$

This method of estimation requires specification only of the conditional mean and variance. In [M118] Godambe and Heyde, and in [M127], Heyde, discussed both finite sample and asymptotic optimality properties of the QL estimating functions. See also [M151] for quasi-likelihood estimation for diffusion processes.

Stochastic processes with long-range dependence are becoming increasingly important tools for modeling financial, meteorological and diverse other time series data. Heyde's joint work with Anh, Gao, et al. in [M171] and [M184], provides an excellent example of path-breaking research in this important area.

In [M171] estimation for a zero-mean stationary Gaussian process $\{Y(t)\}$ with spectral density

$$\phi(\omega,\theta) = \eta \,|\, \omega \,|^{-2\beta} \,(1+\omega^2)^{-\alpha}, \, \omega \in (-\infty, \infty),$$

where $\theta = (\alpha, \beta, \eta) \in [\frac{1}{2}, \infty) \times (0, \frac{1}{2}) \times (0, \infty)$ is a vector of unknown parameters, is discussed. This

process has long range dependence (LRD) and is known as the fractional Riesz-Bessel motion (fRBm). A continuous version of the Gauss-Whittle contrast function is defined as

$$L_N(\theta) = \frac{1}{4\pi} \int_{-\infty}^{\infty} \{\log \phi(\omega,\theta) + I_N(\omega)\phi^{-1}(\omega,\theta)\}(1+\omega^2)^{-1} d\omega$$

where

$$I_N(\omega) = \frac{1}{2\pi N} |\int_0^N e^{-i\omega t} Y(t) dt|^2$$

is the periodogram of $\phi(\omega, \theta)$, N being the upper bound of the interval $[0, N]$ on which $Y(t)$ is observed. The minimum contrast estimator of θ is defined by

$$\hat{\theta}_N = \arg\min_{\theta} L_N(\theta).$$

Consistency and asymptotic normality of the estimator $\hat{\theta}_N$ are established in [M171].
[M184] considers the nonstationary Gaussian process $\{Y(t)\}$ with spectral density given by

$$\phi^*(\omega,\theta) = \phi(\omega,\theta)\pi(\omega), \, \omega \in (-\infty, \infty),$$

where $\pi(\omega)$ is a positive continuous function satisfying a regularity condition, and

$\theta \in [0, \infty) \times \left(\dfrac{1}{2}, \dfrac{3}{2} \right) \times (0, \infty)$ is the unknown parameter. Under regularity conditions, the consistency and

asymptotic normality of the minimum contrast estimator $\hat{\theta}_N$ of θ are established.

The two papers [M171] and [M184] solve the problem of parameter estimation for both stationary and nonstationary Gaussian processes with long range dependence. The method of estimation used in both the papers is based on a natural Gauss-Whittle contrast function. These models are illustrated with an application to real data sets on time series of maximum daily wind speed [M171], and data on air pollution [M184].

References

[1] Fisher, R.A. (1925). Theory of statistical estimation. *Proc. Cambridge Phil. Soc.*, 11, 700–725.

[2] Hall, P.G. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press, NY.

[3] LeCam, L. (1960). Locally asymptotically normal families of distributions. *Univ. Calif. Pub. Prob. Statist.*, 3, 27–98.

[4] Rao, C.R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Roy. Statist. Soc. B*, 24, 46–72.

[5] Wolfowitz, J. (1965). Asymptotic efficiency of the maximum likelihood estimator. *Theor. Prob. Appl.*, 10, 247–260.