STATISTICS–THEORY AND PRACTICE

Richard A. Johnson Conference, University of Wisconsin

Revisiting Local Asymptotic Normality (LAN) and Passing on to Local Asymptotically Mixed Normality (LAMN)

G.G. Roussas
University of California, Davis
(formerly, University of Wisconsin, Madison)

May, 2008

1. Basic Problem

On the basis of a random sample, whose probability law depends on a parameter θ, discriminate between two values θ and θ^* ($\theta \neq \theta^*$).

2. Some Notation

$X_0, X_1, \ldots, X_n$ i.i.d., $(\mathcal{X}, \mathcal{A}, P_\theta)$, $\theta \in \Theta$ open $\subseteq \mathbb{R}^k$,
$k \geq 1$, $\mathcal{A}_n = \sigma(X_0, \ldots, X_n)$, $P_{n,\theta} = P_\theta \mid \mathcal{A}_n$,
$P_{0,\theta} \approx P_{0,\theta^*}$, $\theta, \theta^* \in \Theta$ $(\theta \neq \theta^*)$,
$q(X_0; \theta, \theta^*) = \dfrac{dP_{0,\theta^*}}{dP_{0,\theta}}$,
$\varphi_j(\theta, \theta^*) = \varphi(X_j; \theta, \theta^*) = \left[q(X_j; \theta, \theta^*)\right]^{1/2}$,
$L_n(\theta, \theta^*) = \dfrac{dP_{n,\theta^*}}{dP_{n,\theta}} = \displaystyle\prod_{j=0}^{n} \varphi_j^2(\theta, \theta^*)$,
$\Lambda_n(\theta, \theta^*) = \log L_n(\theta, \theta^*)$;
restrict to θ^*'s close to θ; i.e., $\theta_n = \theta + \frac{h_n}{\sqrt{n}}$,
$h_n \to h \in \mathbb{R}^k$ (all limits taken as $n \to \infty$).

3. Basic Assumption

$\varphi_0(\theta, \theta^*)$ is differentiable in q.m. with respect to θ^* at θ – under P_θ – with q.m. derivative $\dot{\varphi}_0(\theta)$ (a $k \times 1$ random vector). Set

$$L_n(\theta, \theta_n) = \frac{\mathrm{d}P_{n,\theta_n}}{\mathrm{d}P_{n,\theta}}, \quad \Lambda_n(\theta, \theta_n) = \log L_n(\theta, \theta_n),$$

$$\Delta_n(\theta) = \frac{2}{\sqrt{n}} \sum_{j=0}^{n} \dot{\varphi}_j(\theta),$$

$$\Gamma(\theta) : h'\Gamma(\theta)h = 4\mathcal{E}_\theta \left[h'\dot{\varphi}_0(\theta)\right]^2, \quad h \in \mathbb{R}^k,$$

$$A(h, \theta) = \frac{1}{2} h'\Gamma(\theta)h.$$

4. First Installment of Results

Theorem 1: $\Lambda_n(\theta, \theta_n) - h'\Delta_n(\theta) \xrightarrow{P_{n,\theta}} -A(h, \theta)$

Theorem 2: $\mathcal{L}\left[\Delta_n(\theta) \mid P_{n,\theta}\right] \Rightarrow N(0, \Gamma(\theta))$

Theorem 3: $$\mathcal{L}\left[\Lambda_n(\theta, \theta_n) \mid P_{n,\theta}\right] \Rightarrow N\left(-\tfrac{1}{2}h'\Gamma(\theta)h,\ h'\Gamma(\theta)h\right).$$

5. Reminder

$\{P_n\}$ and $\{Q_n\}$, defined on $(\mathcal{X}_n, \mathcal{A}_n)$, are *contiguous* if whenever $P_n(A_n) \to 0$, $A_n \in \mathcal{A}_n$, then $Q_n(A_n) \to 0$, and vice versa.

6. Second Installment of Results

Theorem 4: $\Lambda_n(\theta, \theta_n) - h'\Delta_n(\theta) \xrightarrow{P_{n,\theta_n}} -A(h, \theta)$

Theorem 5: $\mathcal{L}\left[\Delta_n(\theta) \mid P_{n,\theta_n}\right] \Rightarrow N(\Gamma(\theta)h, \Gamma(\theta))$

Theorem 6: $$\mathcal{L}\left[\Lambda_n(\theta, \theta_n) \mid P_{n,\theta_n}\right] \Rightarrow N\left(\frac{1}{2}h'\Gamma(\theta)h,\ h'\Gamma(\theta)h\right).$$

7. Loose Interpretation of Theorem 1

$$\Lambda_n(\theta, \theta_n) \simeq h'\Delta_n(\theta) - A(h, \theta)$$

or

$$L_n(\theta, \theta_n) \simeq \mathrm{e}^{h'\Delta_n(\theta) - A(h,\theta)}.$$

8. Precise Formulation of Above Result

Theorem 7: There exists a (suitably) truncated version $\Delta_n^*(\theta)$ of $\Delta_n(\theta)$ such that:

$$\mathcal{E}_\theta \mathrm{e}^{h'\Delta_n^*(\theta)} \stackrel{def}{=} \mathrm{e}^{B_n(h)} < \infty,$$

$$P_{n,\theta}\left[\Delta_n^*(\theta) \neq \Delta_n(\theta)\right] \longrightarrow 0,$$

$$P_{n,\theta_n}\left[\Delta_n^*(\theta) \neq \Delta_n(\theta)\right] \longrightarrow 0,$$

and if

$$R_{n,h}(A) = \mathrm{e}^{-B_n(h)} \int_A \mathrm{e}^{h'\Delta_n^*(\theta)} \mathrm{d}P_{n,\theta}, \quad A \in \mathcal{A}_n,$$

$$\left(\text{so that } \frac{\mathrm{d}R_{n,h}}{\mathrm{d}P_{n,\theta}} = \mathrm{e}^{h'\Delta_n^*(\theta) - B_n(h)}, \quad h \in \mathbb{R}^k\right),$$

then

$$\|P_{n,\theta_n} - R_{n,h_n}\| \to 0 \quad \text{or}$$

$$\sup\left\{\left\|P_{n,\theta+\frac{h}{\sqrt{n}}} - R_{n,h}\right\|; h \in B \text{ bounded} \subset \mathbb{R}^k\right\} \to 0.$$

9. Statistical Significance of Theorem 7

Exploitation of the local approximation (in the L_1-norm or total variation norm) of the given family of probability measures by an exponential family of probability measures.

10. Statistical Applications of the Theorems

(i) $\Theta \subseteq \mathbb{R}$

(a) For hypotheses (alternatives) for which there are UMP tests in the exponential family, we can construct tests φ_n – based on $\Delta(\theta_j), \theta_j,\ j = 0, 1, 2$ boundary points – which are AUMP
(i.e., $\limsup \left[\sup \left(\mathcal{E}_\theta \omega_n - \mathcal{E}_\theta \varphi_n; \theta \in A\right)\right] \leq 0$) among all tests ω_n such that $\limsup \left[\sup \left(\mathcal{E}_\theta \omega_n; \theta \in H\right)\right] \leq \alpha$. For example: $H : \theta \leq \theta_0$, $A > \theta_0$ at level α,

$$\varphi_n = \varphi_n \left(\Delta_n(\theta_0)\right) = \begin{cases} 1, & \Delta_n(\theta_0) \geq c_n \\ \gamma_n, & \Delta_n(\theta_0) = c_n \\ 0, & \Delta_n(\theta_0) < c_n \end{cases},$$

$c_n, \gamma_n : \mathcal{E}_{\theta_0} \varphi_n = \alpha$.

(b) For hypotheses (alternatives) for which there are UMPU tests in the exponential family, we can construct tests φ_n – based on $\Delta(\theta_j), \theta_j,\ j = 0, 1, 2$ boundary points – which are AUMP
(i.e., $\limsup\,[\sup\,(\mathcal{E}_\theta \omega_n - \mathcal{E}_\theta \varphi_n; \theta \in A)] \leq 0$) among all asymptotically unbiased tests ω_n
(i.e., $\liminf\,[\inf\,(\mathcal{E}_\theta \omega_n; \theta \in H)] \geq \alpha$). For example: $H : \theta = \theta_0,\ A : \theta \neq \theta_0$ at level α,

$$\begin{aligned} \varphi_n &= \varphi_n(\Delta_n(\theta_0)) \\ &= \begin{cases} 1, & \Delta_n(\theta_0) < a_n \text{ or } \Delta_n(\theta_0) > b_n \\ 0, & a_n \leq \Delta_n(\theta_0) \leq b_n \end{cases}, \end{aligned}$$

$(a_n < b_n)$ with $a_n \to -\xi_{\alpha/2},\ b_n \to \xi_{\alpha/2}$ ($\xi_p = p$-th quantile of $N(0,\ \Gamma(\theta_0))$) is AUMPU.

Remark: Actually, the tests are locally AUMP or AUMPU, but they become globally so under the additional assumption:

$$\Delta_n(\theta_j) \xrightarrow{P_{n,\theta_n}} \pm\infty \text{ if } \sqrt{n}(\theta_n - \theta_j) \to \pm\infty, \quad j = 0, 1, 2.$$

(ii) $\Theta \subseteq \mathbb{R}^k, k \geq 1$

$$H : \theta = \theta_0, \quad A : \theta \neq \theta_0 \text{ at level } \alpha.$$

Simple tests φ_n are constructed – based on $\Delta_n(\theta_0)$ – which enjoy Wald-type asymptotically optimal properties:

Weighted average power over certain surfaces is largest – within a class of competing tests.

The sup of the difference of the sup and the inf of the power over certain surfaces $\to 0$.

The sup of the difference between the envelope power and the power over certain surfaces – of the test φ_n – compared to the same sup of any other competing test have a difference, whose limsup is ≤ 0.

11. Set $\theta_n = \theta + \frac{h}{\sqrt{n}}$. Then the estimate T_n is *regular* if

$$\mathcal{L}\left[\sqrt{n}(T_n - \theta_n) \mid P_{n,\theta_n}\right] \Rightarrow \mathcal{L}(\theta),$$

a probability measure.

Theorem 8: For regular estimates,

$$\mathcal{L}(\theta) = N(0, \Gamma^{-1}(\theta)) * \mathcal{L}^*(\theta),$$

$\mathcal{L}^*(\theta)$ *a specific probability measure*.

12. Application to Asymptotic Efficiency of Estimates via Theorem 8: The Weiss-Wolfowitz Approach

$\underline{\Theta \subseteq \mathbb{R}}$

$P_{n,\theta}\left[\sqrt{n}(T_n - \theta) \le x\right] \to F_T(x;\theta)$, a d.f., continuously in Θ for each fixed $x \in \mathbb{R}$ and continuously on $\mathbb{R}$ for each $\theta \in \Theta$. Let $\ell_T(\theta)$ and $u_T(\theta)$ be the "smallest" and the "largest" median of the (continuous) d.f. $F_T(\cdot;\theta)$. Then

$$\lim P_{n,\theta}\left(\theta - \frac{t_1}{\sqrt{n}} + \frac{\ell_T(\theta)}{\sqrt{n}} \le T_n \le \theta + \frac{t_2}{\sqrt{n}} + \frac{u_T(\theta)}{\sqrt{n}}\right) \le B(\theta; t_1, t_2),$$

$B(\theta; t_1, t_2) = \Phi[t_2\sigma(\theta)] - \Phi[t_1\sigma(\theta)]$, $t_1, t_2 > 0$,
Φ is the d.f. of $N(0, \sigma^2(\theta))$, $\sigma^2(\theta) = 4\mathcal{E}_\theta\left[\dot{\varphi}_0(\theta)\right]^2$.

$\Theta \subseteq \mathbb{R}^k, k \geq 1$

Under similar assumptions as above,

$$\lim P_{n,\theta}\Big[-t_1 + \ell_T(\theta, h) \leq \sqrt{n}h'(T_n - \theta) \leq t_2 + u_T(\theta, h)\Big] \leq \Phi[t_2\sigma^{-1}(\theta, h)] - \Phi[-t_1\sigma^{-1}(\theta, h)],$$

$t_1, t_2 > 0$, Φ d.f. of $N(0, \sigma^2(\theta, h))$,
$\sigma^2(\theta, h) = h'\Gamma^{-1}(\theta)h$.

$\underline{\boldsymbol{\Theta} \subseteq \mathbb{R}}$

$\mathcal{L}\left[\sqrt{n}(T_n - \theta) \mid P_{n,\theta}\right] \Rightarrow \mathcal{L}_{T,\theta}$, a probability measure having 0 as its median. Then

$$\limsup \left(\theta - \frac{t_1}{\sqrt{n}} \le T_n \le \theta + \frac{t_2}{\sqrt{n}}\right) \le B(\theta; t_1, t_2),$$

$t_1, t_2 > 0$.

Consider median unbiased T_n's; i.e., $P_\theta(T_n \ge \theta) \ge \frac{1}{2}$, $P_\theta(T_n \le \theta) \ge \frac{1}{2}$. Then

$$\limsup P_\theta \left(\theta - \frac{t_1}{\sqrt{n}} \le T_n \le \theta + \frac{t_2}{\sqrt{n}}\right) \le B(\theta; t_1, t_2),$$

$t_1, t_2 > 0$.

13. Application to Asymptotic Efficiency of Estimates: The Classical Approach

(i) $\Theta \subseteq \mathbb{R}$

Consider estimates T_n such that

$$P_\theta \left[\sqrt{n}(T_n - \theta) \leq x\right] \to \Phi_T(x;\theta)$$

(or continuously so in Θ), where $\Phi_T(\cdot;\theta)$ is the d.f. of $N(0, \sigma_T^2(\theta))$. Then $\sigma_T^2(\theta) \geq 1/\sigma^2(\theta)$ $a.e.[\ell]$ (or pointwise, respectively).
Also, $\mathcal{E}_\theta \left[n\mathcal{E}_\theta(T_n - \theta)^2\right] \geq 1/\sigma^2(\theta)$ $a.e.[\ell]$ (or pointwise, respectively).

(ii) $\Theta \subseteq \mathbb{R}^k$, $k \geq 1$

Consider estimates T_n such that

$$P_\theta \left[\sqrt{n}(T_n - \theta) \leq x\right] \Rightarrow \Phi^{(k)}(x; C_T(\theta))$$

continuously in Θ, where $\Phi^{(k)}(\cdot; C_T(\theta))$ is the d.f. of $N(0, C_T(\theta))$ with $C_T(\theta)$ positive definite. Then, with $\Gamma(\theta) = \mathcal{E}_\theta \left[\dot{\varphi}_0(\theta)\dot{\varphi}_0'(\theta)\right]$, $C_T(\theta) - \Gamma^{-1}(\theta)$ is positive semi-definite $a.e.[\ell^k]$. Also, C_T is continuous in Θ. Furthermore, if Γ is also continuous, then $C_T(\theta) - \Gamma^{-1}(\theta)$ is positive semi-definite for all $\theta \in \Theta$.

14. Generalizations

Most of the above results have been generalized to:

(i) i.n.n.i.d. case

(ii) Markov processes

(iii) Semi-Markov processes

(iv) General stochastic processes, which need not even be stationary

(v) Continuous time Markov processes

(vi) Lévy processes of the discontinuous type

(vii) Continuous time diffusions and Gaussian processes with known covariance.
Also,

(viii) Certain generalizations when the sample size is a stopping time.

15. Case Where the Underlying Family of Probability Measures is not LAN but Rather it is LAMN

Refer to Theorem 1:

$$\Lambda_n(\theta, \theta_n) - \left[h'\Delta_n(\theta) - A(h, \theta)\right] \xrightarrow{P_{n,\theta}} 0,$$

$A(h, \theta)$ non-random. Instead, it may happen that

$$\Lambda_n(\theta, \theta_n) - \left[h'\Delta_n(\theta) - \frac{1}{2}h'T_n(\theta)h\right] \xrightarrow{P_{n,\theta}} 0,$$

$T_n(\theta)$ $k \times k$ $\mathcal{A}_n$-measurable random matrices,
$\Delta_n(\theta) = T_n^{1/2}(\theta)W_n(\theta)$,
$W_n(\theta)$ k-dimensional $\mathcal{A}_n$-measurable random vectors

$$\mathcal{L}\left\{[W_n(\theta), T_n(\theta)] \mid P_{n,\theta}\right\} \Rightarrow \mathcal{L}\left\{[W, T(\theta)] \mid P_\theta\right\}$$

$W \sim N(0, I_k)$ independent of $T(\theta)$, or

$$\mathcal{L}\left\{[\Delta_n(\theta), T_n(\theta)] \mid P_{n,\theta}\right\} \Rightarrow \mathcal{L}\left\{[\Delta(\theta), T(\theta)] \mid P_\theta\right\}$$

$\Delta(\theta) = T^{1/2}(\theta)W$. It follows that

$$\mathcal{L}\left\{[\Delta(\theta) \mid T(\theta)] \mid P_\theta\right\} = N(T(\theta)h, T(\theta)).$$

Thus, Theorem 1 becomes here

Theorem $1'$:

$$\Lambda_n(\theta, \theta_n) - h'\Delta_n(\theta) + \frac{1}{2}h'T_n(\theta)h \xrightarrow{P_{n,\theta}} 0.$$

The underlying family of probability measures is referred to as *Locally Asymptotically Mixed Normal (LAMN)*.

From Theorem $1'$,

$$L_n(\theta, \theta_n) \simeq \mathrm{e}^{h'\Delta_n(\theta) + \frac{1}{2}h'T_n(\theta)h}$$

The expression on the RHS above is referred to as *curved exponential family*.

From the fact that

$$\mathcal{L}\{[\Delta(\theta) \mid T(\theta)] \mid P_\theta\} = N(T(\theta)h,\ T(\theta)),$$

it follows that, in the limit, inference can be carried out conditionally as in a normal distribution, and then revert to the original family of probability measures.

Examples

(i) Explosive autoregressive process of first order

$$X_j = \theta X_{j-1} + \epsilon_j, \quad X_0 = 0, \quad |\theta| > 1,$$

$$\epsilon_j,\ j \geq 1, \text{ i.i.d.} \sim N(0,1).$$

Here

$$\Delta_n(\theta) = \frac{\theta^2 - 1}{\theta^n} \sum_{j=1}^{n} X_{j-1}\epsilon_j$$

$$T_n(\theta) = \sum_{j=1}^{n} X_{j-1}\epsilon_j \Bigg/ \sum_{j=1}^{n} X_{j-1}^2$$

$\mathcal{L}[T(\theta) \mid P_\theta] = \chi_1^2$, $W \sim N(0,1)$, $T^{1/2}(\theta) \sim N(0,1)$, $T^{1/2}(\theta)$ and W independent.

(ii) Super-critical Galton-Watson branching process with geometric offspring distribution

Here

$$\Delta_n(\theta) = \frac{1}{\theta^{(n-1)/2}} \sum_{j=1}^{n} (X_j - \theta X_{j-1}),$$

$$T_n(\theta) = \frac{\theta - 1}{\theta^n} \sum_{j=1}^{n} X_{j-1},$$

$T(\theta)$ exponentially distributed with mean 1, $W \sim N(0,1)$, $T^{1/2}(\theta)$ and W independent.

For LAMN families some results analogous to those associated with LAN families have been established.

Some Selected References

1. Akritas, M.G. (1978): *Contiguity of Probability Measures Associated with Continuous Time Stochastic Processes*. Ph.D. Thesis, Department of Statistics, University of Wisconsin, Madison.

2. Akritas, M.G. and Roussas, G.G. (1980): Asymptotic inference in continuous semi-Markov processes. *Scandinavian Journal of Statistics*, Vol.7, 73–79.

3. Bhattacharya, D. and Roussas, G.G. (2001): Exponential approximation for randomly stopped locally asymptotically mixture of normal experiments. *Stochastic Modeling and Applications*, Vol.4, No.2, 56–71.

4. Jeganathan, P. (1982): On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā*, Vol.44, Ser.A, 173–212.

5. Jeganathan, P. (1995): Some aspects of asymptotic theory with applications to time series model. *Economic Theory*, Vol.II, 818–887.

6. Johnson, R.A. and Roussas, G.G. (1969): Asymptotically most powerful tests in Markov processes. *Annals of Mathematical Statistics*, Vol.40, 1207–15.

7. Johnson, R.A. and Roussas, G.G. (1970): Asymptotically optimal tests in Markov processes. *Annals of Mathematical Statistics*, Vol.41, 918–38.

8. Johnson, R.A. and Roussas, G.G. (1972): Applications of contiguity to multiparameter hypothesis testing. *Proceedings of the Sixth Berkeley Symposium on Probability Theory and Mathematical Statistics*, Vol.1, 195–226.

9. Le Cam, L. (1960): *Locally asymptotically normal families of distributions.* Univ. of California Publications in Statistics, Vol.3, 37–98.

10. Le Cam, L. (1986): *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag, New York.

11. Le Cam, L. and Yang, G.L. (2000): *Asymptotics in Statistics: Some Basic Concepts* (2nd edition). Springer Series in Statistics, Springer, New York.

12. Lind, B. and Roussas, G.G. (1972): A remark on quadratic mean differentiability. *Annals of Mathematical Statistics*, Vol.43, 1030–34.

13. Roussas, G.G. (1972): *Contiguity of Probability Measures: Some applications in statistics*. Cambridge University Press, Cambridge.

14. Roussas, G.G. (1979): Asymptotic distribution of the log-likelihood function for stochastic processes. *Z. Wahrscheinlichkeitstheorie und Verwandte Gebiete*, Vol.47, 31–46.

15. Roussas, G.G. (2005): *An Introduction to Measure-Theoretic Probability*. Elsevier Academic Press, Burlington, Massachusetts.

16. Roussas, G.G. and Soms, A. (1972): On the exponential approximation of a family of probability measures and representation theorem of Hájek-Inagaki. *Annals of the Institute of Statistical Mathematics*, Vol.25, 27–39.

17. Roussas, G.G. and Bhattacharya, D. (1999): Asymptotic behavior of the log-likelihood function in stochastic processes when based on a random number of random variables. In *Semi-Markov Models and Applications*, Jacques Janssen and Nikolaos Limnios (eds), 119–147, Kluwer Academic Publishers, Dordrecht.

18. Roussas, G.G. and Bhattacharya, D. (2002): Exponential approximation of distributions. In *Teoriya Imovirnostey ta Matematichna Statystika*, Vol.66, 109–120. Also, in *Theory of Probability and Mathematical Statistics*, Vol.66, 119–132 (2003) (English version).

19. Roussas, G.G. and Bhattacharya, D. (2008): Hájek-Inagaki representation theorem, under general stochastic processes framework, based on stopping times. *Statistics & Probability Letters*.

20. Van der Vaart, A.W. (1998): *Asymptotic Statistics*, Cambridge series in Statistical and Probabilistic Mathematics. Cambridge Univ. Press.

21. Wald, A. (1943): Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, Vol.54, 426–482.