Neyman-Scott, the Square Array and efficient estimation.

RICHARD P. WATERMAN
Department of Statistics, University of Pennsylvania

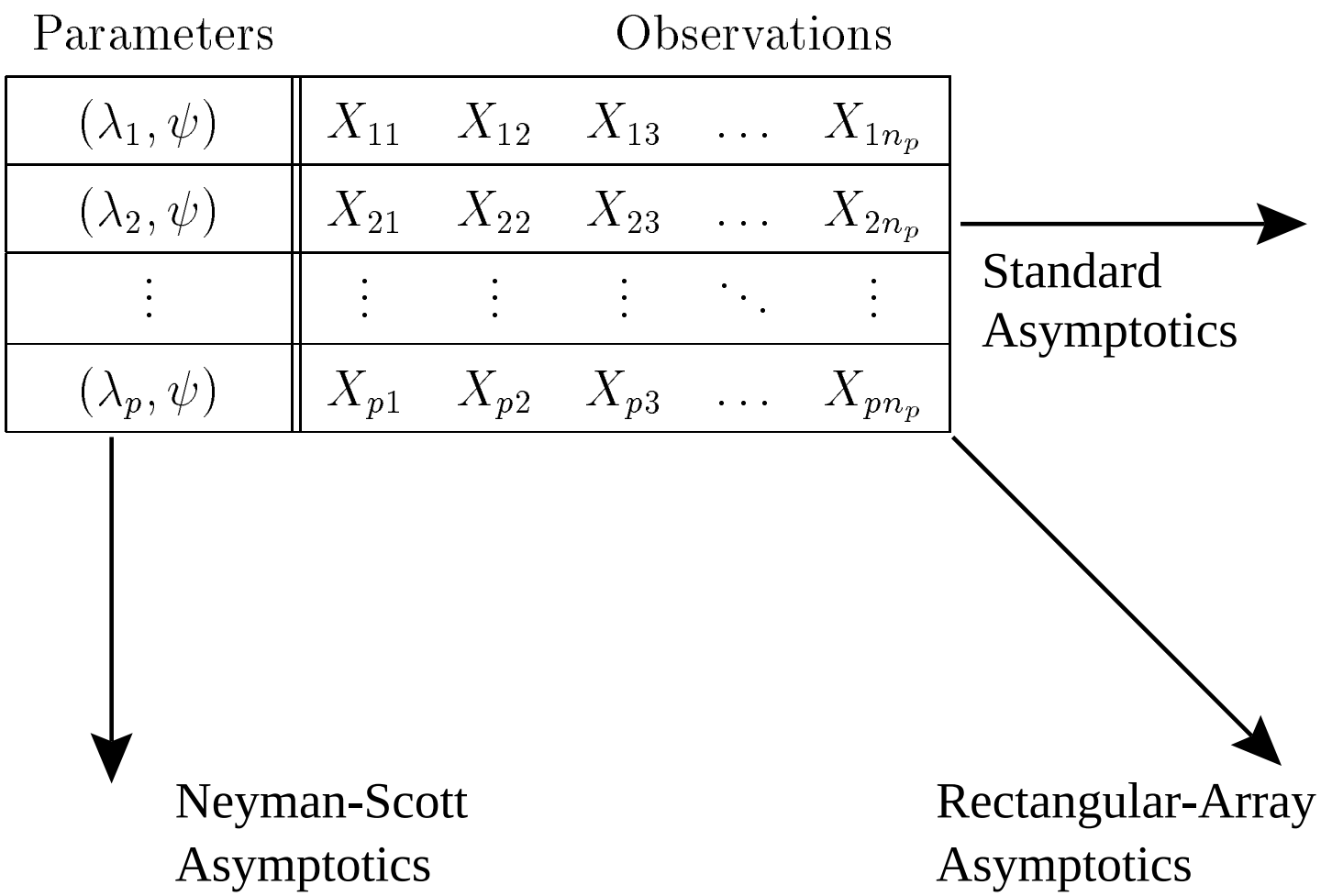
BRUCE G. LINDSAY
Department of Statistics, Pennsylvania State University

June 1994

Three Embeddings

Number of strata $= p$.
Stratum density $f_i = f(x_{ij}; \psi, \lambda_i)$.

- Independent identically distributed within strata
- Independent between strata



Why?

- Theoretical: insight
- Practical: more appoximation theorems
- Efficiency result: a baseline

Other work

- Andersen
- Mallows
- Haberman
- Tak
- Portnoy
- Pfanzagl

Example

Oneway ANOVA.

Parameter of interest is σ^2.

Nuisance parameter is μ_i, taken as FIXED.

The stratum density is $f(x_{ij}; \sigma^2, \mu_i)$, $i = 1, \ldots, p$, $j = 1, \ldots, n_p$.

Maximum likelihood limiting distributions

$$\begin{aligned}
\text{Standard:} \quad & \hat{\sigma}^2 - \sigma^2 \xrightarrow{p} 0 \\
\text{Neyman-S:} \quad & \hat{\sigma}^2 - \sigma^2 \xrightarrow{p} \{-1/n\}\sigma^2 \\
\text{SA, } n = p\text{:} \quad & \hat{\sigma}^2 - \sigma^2 \xrightarrow{p} 0
\end{aligned}$$

$$\begin{aligned}
\text{Standard:} \quad & \sqrt{n\,p}\left(\hat{\sigma}^2 - \sigma^2\right) \xrightarrow{\ell} N(0, 2\sigma^4) \\
\text{SA, } n = p\text{:} \quad & \sqrt{n\,p}\left(\hat{\sigma}^2 - \sigma^2\right) \xrightarrow{\ell} N(-\sigma^2, 2\sigma^4)
\end{aligned}$$

Table 1: Properties of the *mle*, $\hat{\psi}$, in the three asymptotic embeddings.

Asymptotics	Consistency	Efficiency
Standard	✓	✓
Square Array	✓	✗
Neyman-Scott	✗	✗

Asymptotic Efficiency in the presence of nuisance parameters.

Standard:

$$\sqrt{n}\,(\psi_n - \psi) \xrightarrow{\ell} N(0, c^{-1}),$$

where $c = (i^{\psi\psi})^{-1} = i_{\psi\psi} - i_{\psi\lambda} i_{\lambda\lambda}^{-1} i_{\lambda\psi}$.

$$M_1 = E\left\{ \begin{pmatrix} U_0 \\ V_1 \end{pmatrix} (U_0, V_1) \right\} = \begin{pmatrix} i_{\psi\psi} & i_{\psi\lambda} \\ i_{\lambda\psi} & i_{\lambda\lambda} \end{pmatrix}.$$

Infinite nuisance parameter

$$\sqrt{n\,p}\,(\psi_n - \psi) \xrightarrow{\ell} N(0, \overline{c}^{-1}),$$

where $\overline{c} = \lim_{p\to\infty} \frac{1}{p} \sum_{k=1}^{p} (i_k^{\psi\psi})^{-1}$ and $(i_k^{\psi\psi})^{-1}$ is the information in the k^{th} stratum.

Anderson 1970. Suggestion for NS asymptotics, but not attainable in general.

Is it a sensible here in the Square Array?

Since the λ_i is an infinite sequence we need some structure if we are going to say something useful.

- Bounded set.
- Lindenberg-Feller type conditions.
- Come from a distribution.

Bias removal

MLE comes from score equations. Score equations are mean zero but substitution versions are NOT.

Example: In the i^{th} stratum .

$$U_0 = \frac{\partial}{\partial \psi} ln\, f(\psi, \lambda_i, \mathbf{x_i}) = -\frac{n}{2\,\psi} + \frac{1}{2\,\psi^2}\sum_{j=1}^{n}(x_{ij} - \lambda_i)^2 \quad (1)$$

$$V_1 = \frac{\partial}{\partial \lambda_i} ln\, f(\psi, \lambda_i, \mathbf{x_i}) = \frac{1}{\psi}\sum_{j=1}^{n}(x_{ij} - \lambda_i) \quad (2)$$

The solution to (2) is $\hat{\lambda}_i = \overline{\mathbf{x}}_i$. If this estimate for λ_i is "plugged" back into (1), then

$$E\left[-\frac{n}{2\,\psi} + \frac{1}{2\,\psi^2}\sum_{j=1}^{n}(x_{ij} - \hat{\lambda}_i)^2\right] = -\frac{1}{2\,\psi}. \quad (3)$$

The estimator $\hat{\psi}$ is obtained from solving

$$-\frac{n}{2\,\psi} + \frac{1}{2\,\psi^2}\sum_{j=1}^{n}(x_{ij} - \hat{\lambda}_i)^2 = 0,$$

but (3) shows that this doesn't even equal 0 on average! The consequence is an inconsistent estimator of ψ.

Fix this up.

Good estimating equations

Some good score functions/ estimating equations:

If we are lucky can condition on a complete and sufficient statistic for the nuisance parameters:

$$\text{Conditional Score} \qquad U_c = U_0 - E(U_0 \,|\, T).$$

Optimality results from Godambe (1976) and Lindsay (1982).

More generally U_∞.

$$U_\infty = U_0 - \Pi_\infty U_0.$$

$$E_\lambda\{U_\infty(\lambda_0)\} = 0 \quad \text{for all} \quad \lambda, \lambda_0, \psi_0.$$

Π_∞ is projection onto

$$\left\{ \frac{f(Y;\psi_0,\lambda)}{f(Y;\psi_0,\lambda_0)} - 1 \right\}, \lambda \in \Lambda.$$

Space generated by the centered likelihood ratios, Λ_∞. (Barankin(1949)). Characterization: the space of Locally Minimum Variance Unbiased Estimators of functions of λ for ψ_0 fixed.

A basis for Λ_∞

Basis for this space is $\{V_i, V_{ij}, \ldots\}$, where

$$V_{i_1 \ldots i_t} = \frac{\partial^t f}{\partial \lambda_{i_1} \ldots \partial \lambda_{i_t}} \Big/ f.$$

Examples

$$V_i = \frac{\partial f}{\partial \lambda_i} \Big/ f \qquad V_{ij} = \frac{\partial^2 f}{\partial \lambda_i \partial \lambda_j} \Big/ f.$$

So use (univariate λ.)

$$\begin{aligned} U_1 &= U_0 - \rho_1 V_1 && \text{Neyman's } C(\alpha). \\ U_2 &= U_0 - \rho_{11} V_1 - \rho_{12} V_{11} && \text{Small and McLeish} \end{aligned}$$

Claim the root of U_2 is just right for obtaining full efficiency in the rectangular array.

Facts about the Bhattachrayya Basis

Recurrence relation:

$$V_{i_1 \ldots i_{t-1} i_t} = (\partial / \partial \lambda_t) V_{i_1 \ldots i_{t-1}} + V_{i_1 \ldots i_{t-1}} V_{i_t}.$$

Example

$$V_{ij} = (\partial / \partial \lambda_j) V_i + V_i V_j.$$

Expectation Identity: If $T(\lambda) \perp \{V_{i_1}, V_{i_1 i_2}, \ldots V_{i_1 \ldots i_{t-1}}\}$, υ_1 and υ_2 are partitions of the indices $\{i_1, \ldots, i_t\}$ and q is the number of indices in υ_1 then

$$E(T_{\upsilon_1} V_{\upsilon_2}) = (-1)^{t-q} E(T_{i_1 \ldots i_t}) \qquad 0 \leq q \leq t.$$

Example

$$E(U_1'') = -E(U_1' V_1) = E(U_1 V_{11}).$$

Idea of proof of efficiency of U_2 root.

Call the root $\tilde{\psi}$.

Want expansion of

$$\sqrt{n\,p}\left(\tilde{\psi}-\psi\right).$$

Technique: von-Mises expansion. For the i^{th} stratum define

$$F_i(\epsilon_i) = F_{i0} + \epsilon_i\sqrt{n_i}(\hat{F}_i - F_{i0}),$$

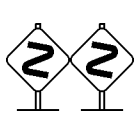
so that the derivative with respect to ϵ_i is

$$\sqrt{n_i}(\hat{F}_i - F_i) = \Delta_{in}.$$

The second order von-Mises expansion is defined by

$$\psi\left(\frac{1}{\sqrt{n_1}},\ldots,\frac{1}{\sqrt{n_p}}\right) - \psi(0,\ldots,0) = \sum_k \psi'_k\Big|_{\epsilon_k=0}\frac{1}{\sqrt{n_k}} +$$

$$\frac{1}{2}\sum_{kl}\psi''_{kl}\Big|_{\epsilon_k=\epsilon_l=0}\frac{1}{\sqrt{n_k}}\frac{1}{\sqrt{n_l}} + Rem(\hat{F}-F_0)$$



Solving for ψ'_k and ψ''_{kl}

Obtain by differentiating

$$n_i^{-1} U_{2i} = \int U_{1ij} dF(\epsilon_i) -$$

$$\rho_i(\psi, \lambda) \left[\left(\int V_{1ij} dF(\epsilon_i) \right)^2 + \epsilon_i^2 \left(\int V_{1ij}^{\lambda_i} dF(\epsilon_i) \right) + o(\epsilon_j^4) \right].$$

Working form (use $\hat{\lambda}_\psi$ as estimates for the nuisance parameters):

$$U_2(\boldsymbol{\epsilon}) = \sum_i \left\{ \int U_{1ij}(\psi(\boldsymbol{\epsilon}_1), \lambda_i(\psi(\boldsymbol{\epsilon}_1), \epsilon_{2i})) dF_{\epsilon_{3i}} - \right.$$

$$\underbrace{\left. \rho_i(\psi(\boldsymbol{\epsilon}_1), \lambda_i(\psi(\boldsymbol{\epsilon}_1), \epsilon_{2i})) \left(\epsilon_{4i}^2 \int V_{1ij}^{\lambda_i}(\psi(\boldsymbol{\epsilon}_1), \lambda_i(\psi(\boldsymbol{\epsilon}_1), \epsilon_{2i})) dF_{\epsilon_{3i}} \right) \right\}}_{\text{Correction term}}.$$

$$U_2(\boldsymbol{\epsilon}) = U_1(\boldsymbol{\epsilon}) + \text{Correction}.$$

First order terms

$$\sqrt{n\,p}\sum_k \psi_k'\Big|_{\epsilon_k=0}\frac{1}{\sqrt{n_k}} \xrightarrow{\ell} N(0,\overline{c}^{-1}).$$

Second order terms

$$\sqrt{n\,p}\,\frac{1}{2}\sum_{kl}\psi_{kl}''\Big|_{\epsilon_k=\epsilon_l=0}\frac{1}{\sqrt{n_k}}\frac{1}{\sqrt{n_l}}$$

has components ψ'' of the form

$$\sum_{kl}\psi_{kl}''O(p)+\sum_k\tilde{U}_k+O_p(\sqrt{p})=0,$$

where the $\tilde{U}_k$ are mean zero when U_2 is used but NOT mean zero if U_1 is used.

So

$$\sum_{kl}\psi_{kl}''=\frac{-\left(\sum_k\tilde{U}_k+O_p(\sqrt{p})\right)}{O(p)}\xrightarrow{p}\text{Const.}$$

The constant.

$$-\frac{1}{2}\frac{E(U_{1i}V_{2i}^*)}{\overline{c}E(V_{1i}^2)}.$$

$$M_2 = E\left\{\begin{pmatrix} U_0 \\ V_1 \\ V_2 \end{pmatrix}(U_0, V_1, V_2)\right\}.$$

Example: (for normal, many means).

$$\begin{pmatrix} \frac{1}{2\sigma^4} & 0 & \frac{1}{\sigma^4} \\ 0 & \frac{1}{\sigma^2} & 0 \\ \frac{1}{\sigma^4} & 0 & \frac{2}{\sigma^4} \end{pmatrix}.$$

$$E(U_{1i}V_{2i}^*) = \frac{1}{\sigma^4} \qquad \overline{c} = \frac{1}{2\sigma^4} \qquad E(V_{1i}^2) = \frac{1}{\sigma^2},$$

so

$$Const = \frac{-\frac{1}{\sigma^4}}{\frac{1}{2\sigma^4}\frac{1}{\sigma^2}} = -\sigma^2.$$

Simulations

Second order information correction:

True info

$$\begin{array}{rcl} InfoU_2(\hat{\lambda}) & = & InfoU_1 - \left(\boxed{\frac{1}{2}\kappa^2} + \boxed{\text{Term 2}} + \boxed{\text{Term 3}} \right). \\ & & O(n) \quad - \quad O(1) \end{array}$$

Oneway ANOVA

Four pivotals:

1.
$$\frac{\sqrt{n\,p}(\hat{\psi} - \psi)}{\sqrt{2}\psi}$$

2.
$$\frac{\sqrt{n\,p}(\hat{\psi} - \psi)}{\sqrt{2}\psi} - \frac{1}{\sqrt{2}}$$

3.
$$\frac{\sqrt{n\,p}(\tilde{\psi} - \psi)}{\sqrt{2}\psi}$$

4.
$$a(\kappa, i)\frac{\sqrt{n\,p}(\tilde{\psi} - \psi)}{\sqrt{2}\psi}$$

$$a(k, i) = 1 - \frac{\kappa^2}{4\,n\,i}$$

Exact 95% coverage rates

$n = p$	1	2	3	4
2	0.9915	0.9770	0.9080	0.9420
3	0.9866	0.9737	0.9254	0.9511
4	0.9533	0.9730	0.9271	0.9547
5	0.9355	0.9716	0.9298	0.9538
10	0.9089	0.9620	0.9388	0.9510
15	0.9021	0.9580	0.9424	0.9504
20	0.8991	0.9559	0.9443	0.9502

Interpretation

$$\hat{\psi} = \frac{1}{p}\frac{1}{n}\sum (x_{ij} - \overline{x}_i)^2$$

$$\tilde{\psi} = \frac{1}{p}\frac{1}{n-1}\sum (x_{ij} - \overline{x}_i)^2$$

If use information correction then get

$$a(\kappa, i)\sqrt{np}\ (\tilde{\psi} - \psi) = \sqrt{(n-1)p}\ (\tilde{\psi} - \psi).$$

Circular Normal problem

(X,Y) come from a bivariate normal with mean lying on a circle with unknown radius ρ.

Simulated 95% coverage rates

n = p	1	2	3	4
2	95052	95831	93715	94483
3	94693	95420	94287	94801
4	94597	95326	94568	94945
5	94586	95281	94668	94988
10	94348	95119	94834	94986
20	94322	95142	95007	95074

Maximizing the log-likelihood

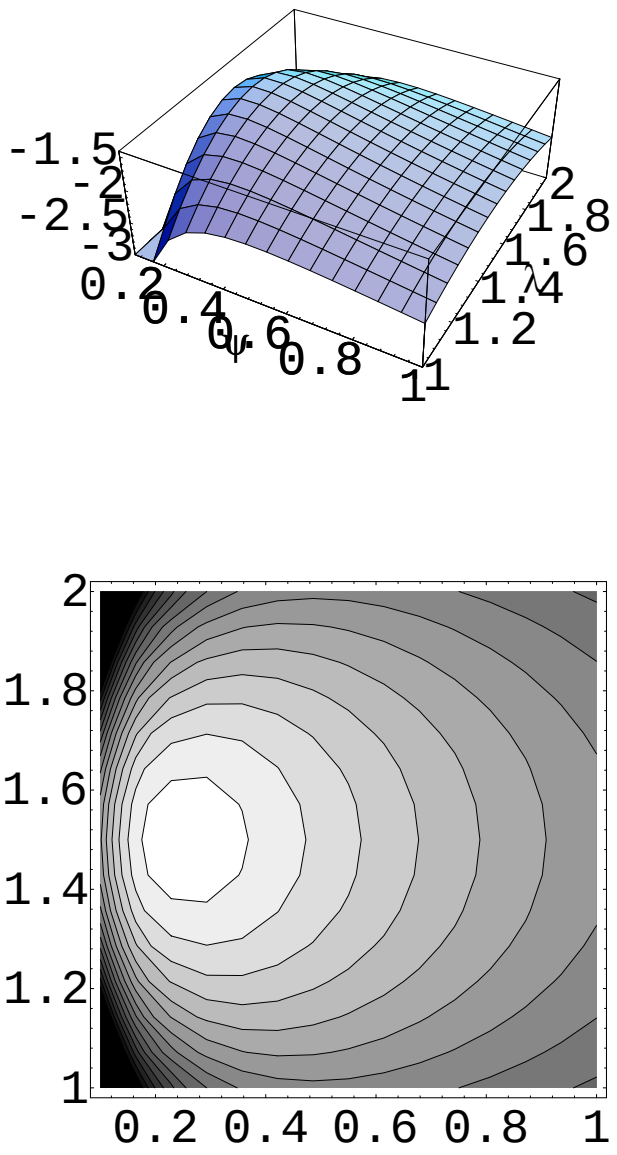


Figure 1: Normal variance and mean