Approximate conditional inference using projected scores

Richard P. Waterman
University of Pennsylvania

Bruce G. Lindsay
Pennsylvania State University

July 14, 1994

Overview

- Nuisance parameters example – Poisson regression.
- What goes wrong with maximum likelihood?
- Conditional score functions.
- Approximating the conditional score using the B-Basis.
- How good an approximation is it? - example 2 x 2 table.
- Exponential family models.
- Rasch model, CPH with ties.
- Square arrays.

CONTEXT

- ∘ Neyman-Scott problems.
- ∘ General inferential methodology.
- ∘ Approximate conditional inference.

Options for conditional inference:

- ⋆ Exact: Mehta et.al.
- ⋆ Saddlepoint approximations: Davison, Tanner.

Second order inference:

- Barndorff-Nielsen.
- Cox and Reid.
- Efron and DiCiccio.
- McCullagh and Tibshirani.

Poisson Regression with Heterogeneity

Motivation. **Paired Poisson**

$$Y_1 \sim Po(\lambda_1) \qquad Y_2 \sim Po(\lambda_2)$$

$$\lambda_1 = \exp(\alpha) \qquad \lambda_2 = \exp(\alpha + \beta)$$

$\beta = \log(\lambda_2) - \log(\lambda_1)$: contrast in log means.

$$Y_2 \,|\, (Y_1 + Y_2) \quad \sim \quad Binomial\left(y_1 + y_2, \frac{\exp(\beta)}{1 + \exp(\beta)}\right)$$

<u>Reparameterize:</u>

$$\beta = \tfrac{1}{2}\left(\log \lambda_2 - \log \lambda_1\right) \qquad \omega = \tfrac{1}{2}\left(\log \lambda_2 + \log \lambda_1\right)$$

$$\log(\lambda_1) = \omega - \beta \qquad \log(\lambda_2) = \omega + \beta$$

$$\boldsymbol{Y} \sim Po\left(\exp(Z\beta + \Gamma\omega)\right)$$

$$Z = \begin{pmatrix} -1 \\ +1 \end{pmatrix} \qquad \Gamma = \begin{pmatrix} +1 \\ +1 \end{pmatrix} \qquad Z \perp \Gamma$$

In general:

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \sim Po(\boldsymbol{\lambda}) = Po\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

$$\log(\boldsymbol{\lambda}) = (Z, \Gamma)\begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\omega} \end{pmatrix} = Z\boldsymbol{\beta} + \Gamma\boldsymbol{\omega}$$

$$rank(Z) = r, \quad rank(\Gamma) = s, \qquad Z \perp \Gamma.$$

- $\boldsymbol{\omega}$ is the nuisance parameter.
- Contingency table analysis, $\Gamma \equiv$ sampling constraints.
- Heterogeneity: each observation has its own mean.
- CSS, $Z^T\mathbf{Y}$ for $\boldsymbol{\beta}$ and $\Gamma^T\mathbf{Y}$ for $\boldsymbol{\omega}$.
- Ideally inference conditional on $\Gamma^T\mathbf{Y}$. **Ugly!**
- Use an <u>"approximate conditional approach"</u>.

The trouble with maximum likelihood

Example 1.

Model: n i.i.d. observations from $Y \sim N(\lambda, \psi)$.

The variance, ψ, is the parameter of interest.

Objective – maximize the loglikelihood:

$$l(\psi, \lambda; x) = -\frac{n}{2}\ln(\psi) - \frac{1}{2\psi}\sum_{i=1}^{n}(x_i - \lambda)^2.$$

Problem: Curvature of likelihood surface.

Maximizing the log-likelihood

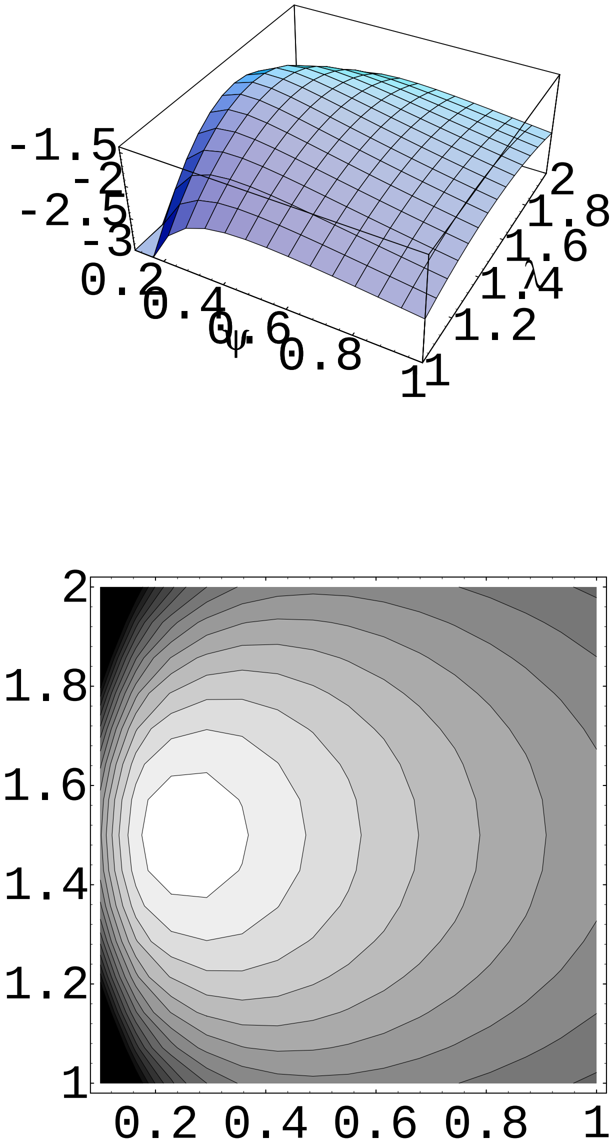

Figure 1: Likelihood surface for $N(\lambda, \psi)$

Conditional score functions

Desirability:

- Optimality theory: Godambe (1976), Lindsay (1982,83)
- No score bias: Lindsay (1982)

Let $T(\psi)$ be the complete and sufficient statistic for λ.

U_0 is the score function for ψ, $U_0 = \partial/\partial\psi \, \{l(\psi, \lambda; x)\}$.

U_c is the conditional score function.

$$\boxed{U_c = U_0 - E(U_0 \,|\, T)}$$

The Conditional Score as a Projection.

Define U_0 as the ψ score.

$$U_c = U_0 - \Pi_\infty U_0,$$

where $\Pi_\infty U_0$ is the projection of U_0 onto the closed linear space generated by

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

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_∞

Assuming sufficient regularity

$$\lim_{h\to 0}\frac{1}{h}\left[\left\{\frac{f(Y;\psi_0,\lambda_0+h)}{f(Y;\psi_0,\lambda_0)}-1\right\}-\left\{\frac{f(Y;\psi_0,\lambda_0)}{f(Y;\psi_0,\lambda_0)}-1\right\}\right]$$

is in the space, and equals

$$V_1=\frac{\frac{\partial}{\partial\lambda}f(\lambda)}{f(\lambda)}.$$

Similarly

$$\begin{aligned}
V_2 &= \frac{\partial^2}{\partial\lambda^2}f(\lambda)\Big/f(\lambda)\\
V_3 &= \frac{\partial^3}{\partial\lambda^3}f(\lambda)\Big/f(\lambda)\\
V_4 &= \frac{\partial^4}{\partial\lambda^4}f(\lambda)\Big/f(\lambda)\\
\vdots &= \vdots \qquad \vdots \quad \vdots
\end{aligned}$$

are all in the space and form (ideally) a basis for Λ_∞.

Call the functions $V_1, V_2, \cdots$ the **Bhattacharyya Basis**.

Finding Projections.

$$U_0 = \frac{\partial}{\partial \psi} l(\psi, \lambda)$$
$$U_1 = U_0 - a_1 V_1$$
$$U_2 = U_0 - (b_1 V_1 + b_2 V_2) = U_1 - b V_2^*$$

Find b_1, b_2 so that $E[(U_0 - b_1 V_1 - b_2 V_2)^2]$ is minimized.

$$M_2 = E \begin{pmatrix} U_0 U_0 & U_0 V_1 & U_0 V_2 \\ U_0 V_1 & V_1 V_1 & V_1 V_2 \\ U_0 V_2 & V_1 V_2 & V_2 V_2 \end{pmatrix} = \begin{pmatrix} m_{00} & m_{01} & m_{02} \\ m_{01} & m_{11} & m_{12} \\ m_{02} & m_{12} & m_{22} \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}^{-1} \begin{pmatrix} m_{01} \\ m_{02} \end{pmatrix}$$

$$E[U_2^2] = E[U_0^2] - \begin{pmatrix} m_{01} & m_{02} \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix}^{-1} \begin{pmatrix} m_{01} \\ m_{02} \end{pmatrix}$$

Extended Information matrix, M_2.

$$M_2 = \begin{pmatrix} \frac{n}{2\psi^2} & 0 & \frac{n}{\psi^2} \\ 0 & \frac{n}{\psi} & 0 \\ \frac{n}{\psi^2} & 0 & \frac{2n^2}{\psi^2} \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{n}{\psi^2}\frac{\psi^2}{2\,n^2} \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2\,n} \end{pmatrix}$$

So

$$E[U_2(\hat{\lambda})] = E[U_1(\hat{\lambda}) - \hat{b}_2 V_2^*(\hat{\lambda})] = -\frac{1}{2\,\psi} - (-\frac{1}{2\,\psi}) = 0$$

$U_2(\hat{\lambda})$ has no score bias, and in fact $U_2 = U_c$ in this case.

Using U_t to approximate U_c.

Example: 2×2 Table, paired Binomials.

$$X \sim Bin\left(r, \frac{exp(\psi+\lambda)}{1+exp(\psi+\lambda)}\right)$$

$$Y \sim Bin\left(s, \frac{exp(\lambda)}{1+exp(\lambda)}\right)$$

The log odds ratio ψ is the parameter of interest, and a complete sufficient statistic for the nuisance parameter λ is $X+Y$.

$$U_0 = -r\frac{exp(\psi+\lambda)}{1+exp(\psi+\lambda)} + x$$

$$V_1 = -r\frac{exp(\psi+\lambda)}{1+exp(\psi+\lambda)} + x - s\frac{exp(\lambda)}{1+exp(\lambda)} + y$$

$$V_1' = r\frac{exp(2\lambda+2\psi)}{(1+exp(\lambda+\psi))^2} - r\frac{exp(\lambda+\psi)}{1+exp(\lambda+\psi)} + s\frac{exp(2\lambda)}{(1+exp(\lambda))^2} - s\frac{exp(\lambda)}{1+exp(\lambda)}$$

$$V_2 = V_1' + V_1^2$$

Projective and Conditional Information Comparison

$\psi = 0.5$ and $\lambda = -0.25$.

Table 1: Comparison of Informations

$r = s$	U_1	U_2	U_c	$\sqrt{\frac{U_1 Info}{U_c Info}}$	$\sqrt{\frac{U_2 Info}{U_c Info}}$
1	0.123	0.119	0.119	1.015585102	1.000000000
5	0.615	0.613	0.613	1.001743723	1.000000071
10	1.231	1.229	1.229	1.000826326	1.000000007
15	1.846	1.844	1.844	1.000541461	1.000000002
20	2.461	2.459	2.459	1.000402652	1.000000001

What's so good about the B-basis?

- We get close very quickly.
- The basis $\{V_1, V_2\}$ is enough to repair score bias.

Exponential Family

$$f(x;\psi,\lambda) = \exp\{\psi x + \lambda S - a(\psi,\lambda)\} \tag{1}$$

A complete sufficient statistic for λ with ψ fixed is $S = S(x)$. A Type I problem.

Magic

$$\begin{aligned} U_t' &= -a_t' V_t^*, \qquad & t = 0, 1 \ldots \\ V_t^{*\prime} &= c_t V_{t-1}^*, \qquad & t = 1, 2 \ldots \end{aligned}$$

Case: t = 1, then $V_1' = -a_1 V_0 = -a_1$; a constant. But $E[V_1^2] = -E[V_1'] =$ constant. Expected information = Observed information..

Bottom Line:

$$U_t(\hat{\lambda}) - U_t(\lambda) = O_p(n^{-\frac{1}{2}(t-1)}).$$

Case t = 2, then $U_2(\hat{\lambda}) - U_2(\lambda) = O_p(n^{-\frac{1}{2}})$.
Usually $U_2(\hat{\lambda}) - U_2(\lambda) = O_p(1)$.

There is less of a price to pay than expected.

Canonically Parameterized GLM's

(Dispersion parameter known.)

$$l(\theta; y) = \sum_{i=1}^{n} y_i \eta_i - \sum_{i=1}^{n} b_i(\eta_i), \tag{2}$$

where $b_i(\eta_i) = \exp(x_i\theta)$ for Poisson *e.t.c.* .

$$\mathbf{X} = (Z, \Gamma), \qquad \theta = (\psi, \lambda)^T.$$

A complete sufficient statistic for λ is $\Gamma^T Y$.

$$U_2(\hat{\lambda}_\psi) = Z^T(y - m) + Z^T \Xi \widetilde{M} \Delta^T (\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \Big[(\boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}^T) \Sigma$$
$$(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \Big]^+ (\boldsymbol{\Gamma}^T \otimes \boldsymbol{\Gamma}^T) \Delta M \mathbf{1}$$

- m = vector of means
- M = matrix of variances
- $\widetilde{M}$ = matrix of third cumulants
- $\bar{M}$ = matrix of fourth cumulants
- $\Xi = I - M\boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T M \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^T$
- $\Sigma = A + \Delta(\widetilde{M} M^{-1} \Xi \widetilde{M}) \Delta^T$
- $A = M \odot M + \Delta(\bar{M} - \widetilde{M} M^{-1} \widetilde{M}) \Delta^T$

Approximate Conditional Inference

Example **1** *Rasch Model:*

Source: Lindsay et.al.

Response pattern item m = 1 2 3 4	Score s	Observed frequencies
1 1 1 1	4	42
0 1 1 1	3	1
1 0 1 1	3	6
1 1 0 1	3	6
0 0 0 1	3	23
0 0 1 1	2	2
0 1 0 1	2	1
1 0 0 1	2	7
0 1 1 0	2	4
1 0 1 0	2	24
1 1 0 0	2	25
0 0 0 1	1	2
0 0 1 0	1	9
0 1 0 0	1	6
1 0 0 0	1	38
0 0 0 0	0	20
Total (N)		216

Model

Let y_{ij} be response of subject i for item j.

$$f(y_{ij};\psi_j,\lambda_i)=\frac{\exp\{(\lambda_i-\psi_j)y_{ij}\}}{1+\exp(\lambda_i-\psi_j)}$$

ψ_j: item difficulty parameter.
λ_i: subject ability parameter.

Note: Multivariate parameter of interest, and multivariate nuisance parameter.

+ No interaction between subject and item parameters. Independence of responses across items.

Why conditional inference??

Like semi-parametric model with an unspecified mixing distribution on the λ_i.

Model Matrices

$$Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ \vdots & \vdots & \vdots \\ 0 & 0 & 0 \end{pmatrix} \Gamma = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

+ A vector of weights.

```
              RESULTS FROM MAXIMUM LIKELIHOOD ESTIMATION

    Random Component is BINOMIAL
    Link Function is CANONICAL

    Number of parameters of interest:   3
    Number of nuisance parameters:      14

Convergence occured at iteration   6
Convergence criterion, epsilon =  0.100000D-04


Parameter Estimates and Estimated Standard Errors:
------------------------------------------------------------

Parameter       Estimate       Est. s.e.        p.val
 Beta_ 1         4.160451       0.373301        0.000
 Beta_ 2         1.716683       0.312971        0.000
 Beta_ 3         1.821934       0.313658        0.000

------------------------------------------------------------



Hypothesis test for all betas equal to zero

Score test    167.703
D.F.            3
p.val           0.000
```

```
                    RESULTS FROM U2 PROJECTIVE SCORE
                   (APPROXIMATE CONDITIONAL INFERENCE)

Full U2 score function was used

Convergence occured at iteration   8


Parameter Estimates and Estimated Standard Errors:
-----------------------------------------------------------

Parameter       Estimate       Est. s.e.       p.val
 Beta_ 1         2.943123       0.297330       0.000
 Beta_ 2         1.224695       0.260388       0.000
 Beta_ 3         1.301608       0.260745       0.000

-----------------------------------------------------------

Hypothesis test for all betas equal to zero

Score test    125.777
D.F.            3
p.val           0.000
```

Param.	Parameter estimate		Std. err. estimate	
	Un- conditional	Approx. conditional	Un- conditional	Approx. conditional
ψ_1	4.160451	2.943123	0.373301	0.297330
ψ_2	1.716683	1.224695	0.312971	0.260388
ψ_3	1.821934	1.301608	0.313658	0.260745

Table 2: Comparison of unconditional and approximate conditional inference in a Rasch model.

Data source: Lindsay et.al. (1991) JASA, p.100, Table 3.

	Parameter estimate			Std. err. estimate		
	Exact conditional	Approx. conditional	Ratio	Exact conditional	Approx. conditional	Ratio
ψ_1	2.9445	2.9431	0.9995	0.2988	0.2973	0.9950
ψ_2	1.2440	1.2245	0.9845	0.2631	0.2604	0.9897
ψ_3	1.3204	1.3016	0.9858	0.2633	0.2607	0.9901

Table 3: Comparison of exact conditional and approximate conditional inference in a Rasch model.

Data source: Lindsay et.al. (1991) JASA, p.100, Table 3.

Cox Proportional hazards with ties adjustment.

Hazard Rate: $\lambda(t) = \lim_{h \downarrow 0} P(T \leq t + h \,|\, T > t)$.

T is the survival time.

$$\lambda(t; z) = \lambda_0(t) \exp\{\beta_0^T z(t)\}, \qquad t \geq 0.$$

Survival times; $T_1, T_2, \ldots, T_n$, possibly right censored.

Partial likelihood: (No ties).

$$L(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta^T z_i(T_i)\}}{\sum_{j \in \mathcal{R}_i} \exp\{\beta^T z_j(T_i)\}} \right\}^{\delta_i}.$$

$\mathcal{R}_i = \{j : T_j \geq T_i\}$, $1 - \delta_i$ is the indicator for censoring.

Ties

Say three item at risk at time τ and items $i = 1, 2$ are seen to fail, then the contribution to likelihood is:

$$\frac{\exp(\beta^T z_1)}{\exp(\beta^T z_1) + \exp(\beta^T z_2) + \exp(\beta^T z_3)} \frac{\exp(\beta^T z_2)}{\exp(\beta^T z_2) + \exp(\beta^T z_3)} +$$

$$\frac{\exp(\beta^T z_2)}{\exp(\beta^T z_1) + \exp(\beta^T z_2) + \exp(\beta^T z_3)} \frac{\exp(\beta^T z_1)}{\exp(\beta^T z_1) + \exp(\beta^T z_3)}$$

Formal equivalence to exact conditional logistic regression.

Example **2** *Proportional hazards with ties.*

Data: Two groups. One with drug, other without.

Data Source: McCullagh and Nelder, p.424.

Group	
A	$6^*, 6, 6, 6, 7, 9^*, 10^*, 10, 11^*, 13, 16,$ $17^*, 19^*, 20^*, 22, 23, 25^*, 32^*, 32^*, 34^*, 35^*$
B	$1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11,$ $12, 12, 15, 17, 22, 23$

* denotes a censored event.

Results:

METHOD	$\hat{\beta}$	$\widehat{s.e.}(\hat{\beta})$
EXACT	1.6282	0.4332
PETO/BRESLOW	1.5088	0.4096
EFRON	1.5721	0.4124
WATERMAN/LINDSAY	1.6270	0.4325

Summary

- Simple adjustment.
- Multivariate ψ and λ.
- Natural generalization of standard asymptotics.
- Specially good in the Exponential family.

References-Links

- Lehmann (1981)
- Lindsay (1982, 1983)
- McCullagh and Tibshirani (1990)
- Efron, DiCiccio and Efron (1985, 1992)
- Small and McCleish (1988)
- Barankin (1949)
- Neyman and Scott (1948)

- Agresti (1990)
- McCullagh and Nelder (1989)
- Conditional Logistic, Hirji et al (1987)
- Mantel-Haenszel, Mehta et al (1985)
- Lindsay et.al (1991)
- Conaway (1989)
- Tjur (1982)

Difficulties with nuisance parameters.

Parametric model, parameter $\theta \in \Theta$. $\theta = (\psi, \lambda)$.
ψ is parameter of interest. λ is the nuisance.

Neyman-Scott, (1948)

Oneway ANOVA. p treatments or strata, n observations in each treatment. $X_{ij} \sim N(\lambda_i, \psi)$.

Params	$\xrightarrow{n \to \infty}$				
(λ_1, ψ)	X_{11}	X_{12}	X_{13}	$\ldots$	X_{1n}
(λ_2, ψ)	X_{21}	X_{22}	X_{23}	$\ldots$	X_{2n}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
(λ_p, ψ)	X_{p1}	X_{p2}	X_{p3}	$\ldots$	X_{pn}

As $n \to \infty$ (p fixed) usual asymptotics hold.
As $p \to \infty$ (n fixed), $\hat{\psi} \xrightarrow{p} \frac{n-1}{n}\psi$, **inconsistent**.

Even when an maximum likelihood estimatoris consistent it may still be asymptotically **inefficient**.

Application: Square Array Asymptotics

For the i^{th} stratum, $i = 1, 2, \cdots, p$:

$$\Delta_{tin_p} = U_{ti}(\hat{\lambda}_{i\psi}) - U_{ti}(\lambda_i) \xrightarrow{\ell} Y_i$$

has a limiting distribution as $n_p \longrightarrow \infty$.

Using a triangular array central limit theorem, as $p \longrightarrow \infty$:

$$\tilde{\Delta}_{pn_p} = \frac{1}{\sqrt{p}} \left(\sum_{i=1}^{p} \left(\Delta_{tin_p} - \mu_i \right) \right) \xrightarrow{\ell} N(0, \sigma_\Delta^2)$$

$\mu_{in_p} = E[\Delta_{tin_p}]$.

So

$$\frac{1}{\sqrt{p}} \left(\sum_{i=1}^{p} \frac{U_{ti}(\hat{\lambda}_{i\psi}) - U_{ti}(\lambda_i)}{\sqrt{n_p}} \right) - \frac{\sqrt{p}}{\sqrt{n_p}} \overline{B}_p \xrightarrow{p} 0$$

$\overline{B}_p = \frac{1}{p} \sum_{i=1}^{p} \mu_{in_p}$

Take $n_p = p =$ "square array":

$$U_1 : \overline{B}_p \text{ is } O(1), \quad \text{and} \quad U_2 : \overline{B}_p \text{ is } O(n_p^{-\frac{1}{2}})$$

$$\boxed{\text{CONCLUSION: } U_2(\hat{\boldsymbol{\lambda}}) \text{ is fully efficient, } U_1(\hat{\boldsymbol{\lambda}}) \text{ is not.}}$$

Example: Neyman-Scott
Square Array

Consistency:

$$\hat{\sigma}^2_{MLE} \longrightarrow \sigma^2$$

Efficiency:

$$\sqrt{n\,p}\left(\hat{\sigma}^2_{MLE} - \sigma^2\right) \xrightarrow{\ell} N\left(\sigma^2, 2\sigma^4\right)$$

Curvature

Price for estimating λ is an asymptotic bias.

Reduction in information between U_1 and U_2.

Connection: Linked by model curvature.

Surface:

$$H(\lambda) = \left(\begin{array}{c} E_\lambda[U_1(\lambda_0)] \\ W(\lambda_0)^{-\frac{1}{2}} E_\lambda[\mathcal{V}_1(\lambda_0)] \end{array} \right).$$

Principle Curvatures of $H(\lambda)$ at λ_0, are $\kappa_1, \kappa_2 \cdots \kappa_s$.

Then

$$\boxed{\text{BIAS} = -\frac{1}{2} \sum_{i=1}^{s} \kappa_i.}$$

$$\boxed{\text{LOSS in INFORMATION} = \frac{1}{2} \sum_{i=1}^{n} \kappa_i^2.}$$

Where from??

The κ_i are roots of equation:

$$\det(\widetilde{T}_{ij} - \kappa \widetilde{W}_{i,j}) = 0.$$

Choose a parameterization in which $W = E[\mathcal{V}_1 \mathcal{V}_1^T]$ equals an *identity matrix*. Then an eigenvalue equation.

Also $T = E[U_0 \mathcal{V}_2^*]$.

Strong relationship to Efron (1985, 1987) DiCiccio and Efron (1992).