Structural Inference and Asymptotics

Nancy Reid

August 11 2022

AUGUST 6–11
2022
WASHINGTON, DC
JSM
STATISTICS: A FOUNDATION FOR INNOVATION

The Canadian Journal of Statistics
Vol. xx, No. yy, 20??, Pages 1–25
La revue canadienne de statistique

D.A.S. Fraser: From structural inference to asymptotics

Nancy Reid[1]*

[1]*Department of Statistical Sciences, University of Toronto*

Key words and phrases: inference; foundations; likelihood; saddlepoint approximation; tangent exponential model.

MSC 2010: Primary 6202; secondary 62F99

Abstract: Don Fraser was my collaborator and life partner, so I had a uniquely close view of his life in research. This note describes how his early work in the structure of models informed our work in asymptotic theory. *The Canadian Journal of Statistics* xx: 1–25; 20??

Résumé: Insérer votre résumé ici. We will supply a French abstract for those authors who can't prepare it themselves. *La revue canadienne de statistique* xx: 1–25; 20??

ANDROMEDA trial

	Died	Lived	
New	74	138	212
Old	92	120	212
Total	166	258	424

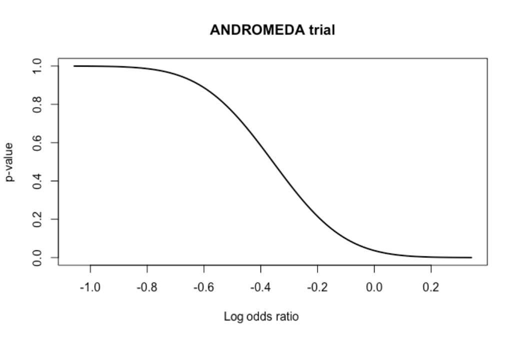
2-sided p-value = 0.07

$$\mathrm{pr}(|T| \geq t^o; \theta = 0)$$

probability of a result as or more extreme than observed, under the model

T: likelihood ratio stat

no adjustment for covariates



$$p(\theta) = \mathrm{pr}\{T \leq t^o; \theta\}$$

percentile point of the observed value t^o as a function of θ

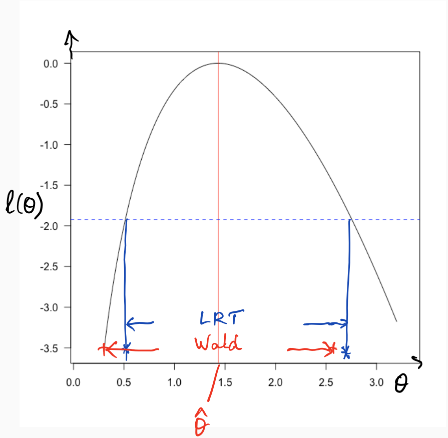
... a p-value function

Fraser 1991

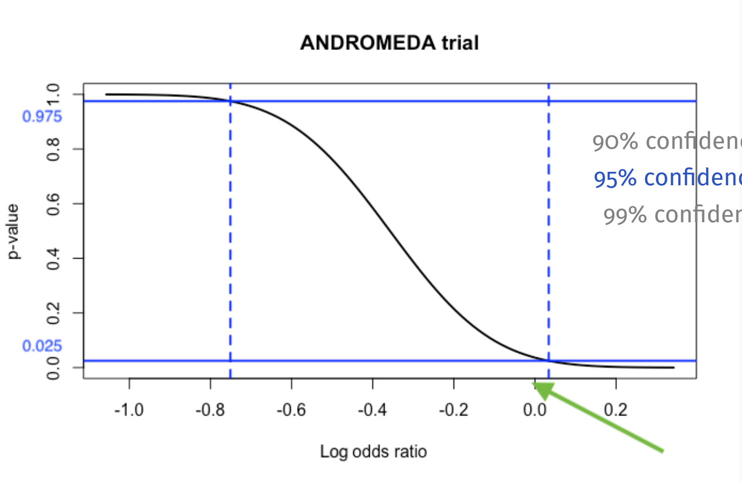
- in "large samples" $\hat{\theta} \dot{\sim} N(\theta, \hat{\sigma}_\theta)$

$$2\{\ell(\hat{\theta}) - \ell(\theta)\} \dot{\sim} \chi^2_1$$

- p-value function $p(\psi) \doteq \Phi(q_\theta) \doteq \Phi(r_\theta)$
- Wald test $q_\theta = (\hat{\theta} - \theta)/\hat{\sigma}_\theta$
- likelihood ratio test
  $r_\theta = \pm\surd[2\{\ell(\hat{\theta}) - \ell(\theta)\}]$
- use profile likelihood if
  $\theta = (\psi, \boldsymbol{\lambda}); \theta \longleftarrow \psi$



… a p-value function



90% confidence interval: [−0.688, −0.030]

95% confidence interval: [−0.751, 0.034]

99% confidence interval: [−0.825, 0.107]

Confidence distribution function; Cox 1958

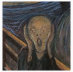
Higher-order asymptotics

- in "large samples" $r_\psi = \pm\surd[2\{\ell_p(\hat\psi) - \ell_p(\psi)\}] \mathrel{\dot\sim} N(0,1)$ $\qquad 2\{\ell_p(\hat\psi) - \ell_p(\psi)\} \mathrel{\dot\sim} \chi^2_1$
- p-value function $\quad p(\psi) \doteq \Phi(r_\psi)$ $\qquad \theta = (\psi, \boldsymbol\lambda), \quad \ell_p(\psi) = \ell(\psi, \hat\lambda_\psi)$ profile
- a better approximation $p(\psi) \doteq \Phi(r^*_\psi)$

$$r^*_\psi = r_\psi + \frac{1}{r_\psi}\log\left(\frac{Q_\psi}{r_\psi}\right)$$

$$Q_\psi = \frac{|\varphi(\hat\theta) - \varphi(\hat\theta_\psi) \quad \partial\varphi(\hat\theta_\psi)/\partial\lambda^{\mathrm{T}}|}{|\partial\varphi(\hat\theta)/\partial\theta^{\mathrm{T}}|}\left\{\frac{|j(\hat\theta)|}{|j_{\lambda\lambda}(\hat\theta_\psi)|}\right\}^{1/2}$$

$$= (\hat\chi - \hat\chi_\psi)/\hat\sigma_\chi$$

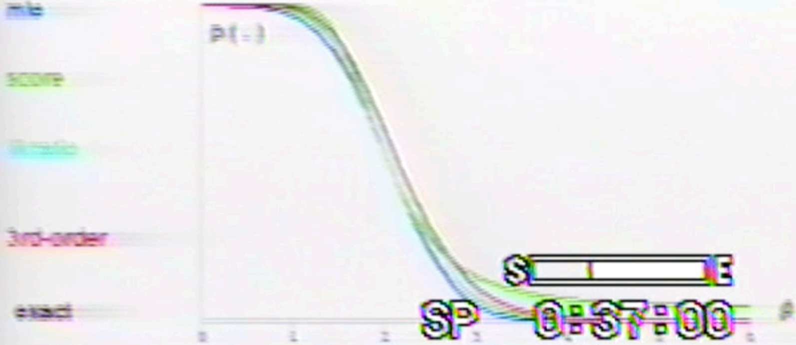


Example 2 ....again

Location log-gamma   log gamma(3; θ) $y^0 = 3.14$

The significance function plots again

3rd-order calculated via φ and plotted against θ



- location model $y = \theta + e, \quad e \sim f_0(e)$ $\qquad y \sim f_0(y - \theta)$

- y is observed to be y^o
- $\theta = y^o - e$
- probability statements about e can be converted to probability statements about θ

- if $y \to y + a$, and $\theta \to \theta - a$, any given quantile in the distribution of Y is unchanged
- perturbations in y are linked to perturbations in θ $\qquad \frac{dy}{d\theta} = 1$

- in structural inference, parameters are linked to observations using groups

Location models

- location model $y = \theta + e, \quad e \sim f_0(e)$ $\qquad y \sim f_0(y - \theta)$

- y is observed to be y^o
- $\theta = y^o - e$
- probability statements about e can be converted to probability statements about θ

- sample $y_1, \dots, y_n$ independently $\quad y_i = \theta + e_i,\ i = 1, \dots, n$ $\qquad y_i \sim f_0(y - \theta)$

- $$\begin{pmatrix} y_2 - y_1 \\ \vdots \\ y_n - y_1 \end{pmatrix} = \begin{pmatrix} e_2 - e_1 \\ \vdots \\ e_n - e_1 \end{pmatrix}$$ these functions of $e = (e_1, \dots, e_n)$ are known

- for inference, use density of y_1 conditional on $(y_2 - y_1, \dots, y_n - y_1)$
- equivalently, density of MLE $\hat{\theta}$ conditional on $a = (y_1 - \hat{\theta}, \dots, y_n - \hat{\theta})$

a for "ancillary"

Local location models

Local Conditional Sufficiency

By D. A. S. Fraser

University of Toronto

[Received February 1963. Revised November 1963]

Summary

The concept of sufficiency can be examined locally for a neighbourhood of a parameter point θ_0 and it leads to an estimate that has minimum variance among locally-unbiased estimates (Fraser, 1964). This local estimate generates the natural global estimate for the case of the simple exponential model, but it does not do so for the important case of the general translation model. The global estimate, the Pitman estimate, for this translation model is rooted in ancillarity and conditional sufficiency. In this paper the concept of local conditional sufficiency is developed; the asymptotic distribution of the conditionally-sufficient statistic is investigated for large samples; and for the maximum-likelihood estimate a conditional large-sample variance is obtained which differs from that usually used.

- start with a general model $y \sim f(y;\theta), \quad \theta \in \mathbb{R}$
- define a transformation $x = \int^y -\frac{f(y;\theta_0)}{\partial F(y;\theta_0)/\partial\theta} dy$ — θ_0 an 'arbitrary' point
- model $g(x;\theta)$ is a location model 'near' θ_0 — $\left.\frac{\mathrm{d}x}{\mathrm{d}\theta}\right|_{\theta_0} = 1$
- sample $y_1, \ldots, y_n \quad \longrightarrow \quad x_1, \ldots, x_n$
- inference $\quad g(x_1 \mid a_1;\theta)$ or $g(\hat\theta \mid a)$ — x_1 or $\hat\theta$ is conditionally sufficient for θ
- local ancillary statistic is determined by a vector

$$v = (v_1, \ldots, v_n) = (\mathrm{d}y_1/\mathrm{d}\theta, \ldots, \mathrm{d}y_n/\mathrm{d}\theta)|_{\theta_0}$$

if $\theta \in \mathbb{R}^p$, $V = (\mathrm{d}y_1/\mathrm{d}\theta^{\mathrm{T}}, \ldots, \mathrm{d}y_n/\mathrm{d}\theta)^{\mathrm{T}}|_{\theta_0}$

- fix $q \in (0,1)$, find $y_q(\theta)$ s.t. $F\{y_q(\theta);\theta\} = q$. Then

F is cdf

$$0 = \frac{\partial F(y_q(\theta);\theta)}{\partial y_q}\frac{\partial y_q(\theta)}{\partial \theta} + \frac{\partial F(y_q;\theta)}{\partial \theta}$$

- q is arbitrary so

$$\frac{\partial y}{\partial \theta} = -\frac{\partial F(y;\theta)/\partial \theta}{f(y;\theta)}$$

- local ancillary statistic for a sample of size n is

$$v = (v_1,\ldots,v_n) = (\mathrm{d}y_1/\mathrm{d}\theta,\ldots,\mathrm{d}y_n/\mathrm{d}\theta)|_{\theta_0}$$

- classical statistics: $\sqrt{n}(\hat\theta - \theta_0) \xrightarrow{d} N\{0, i^{-1}(\theta_0)\}$ Wald, Rao (score), LRG
- Taylor series: $\ell(\theta) = \ell(\hat\theta) + (\theta - \hat\theta)\ell'(\hat\theta) + \frac{1}{2}(\theta - \hat\theta)^{\mathrm{T}}\ell''(\hat\theta)(\theta - \hat\theta) + \dots$ $\ell(\theta) = \log f(y^o; \theta)$
- ... much anguish

- “Fraser series”:

$$\begin{aligned}
\ell(\theta; y) &= \ell(\hat\theta; y^o) + (\theta - \hat\theta)\ell_\theta(\hat\theta) + \frac{1}{2}(\theta - \hat\theta)^{\mathrm{T}}\ell_{\theta\theta}(\hat\theta; y^o)(\theta - \hat\theta) + \dots \\
&+ (y - y^o)\ell_{;y}(\hat\theta; y^o) + \frac{1}{2}(y - y^o)^{\mathrm{T}}\ell_{;yy}(\hat\theta; y^o)(y - y^o) + \dots \\
&+ (y - y^o)^{\mathrm{T}}\ell_{\theta;y}(\hat\theta; y^o)(\theta - \hat\theta) + \frac{1}{3}(y - y^o)^T\ell_{\theta\theta;y}(\hat\theta; y^o)(\theta - \hat\theta)^2 + \dots
\end{aligned}$$

- expand in both sample space and parameter space ???

... Asymptotics

Andrews et al 2005

$$\log\ f(y;\theta) = \ell(\theta; y) = \sum_{i,j\geqslant 0} a_{ij}(\theta-\theta^0)^i (y-y^0)^j / i!j!, \tag{2.1}$$

$$\bar{A} = \begin{pmatrix} \frac{a+3\alpha_4-5\alpha_3^2-12\gamma}{24n} & \frac{-\alpha_3}{2n^{1/2}} & -\left\{\frac{1+\alpha_4-2\alpha_3^2-5\gamma}{2n}\right\} & \frac{\alpha_3}{n^{1/2}} & \frac{\alpha_4-3\alpha_3^2-6\gamma}{n} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \gamma/n & - & - \\ \frac{-\alpha_3}{n^{1/2}} & 0 & - & - & - \\ \frac{-\alpha_4}{n} & - & - & - & - \end{pmatrix}, \tag{2.6}$$

where $a = -\frac{1}{2}\log(2\pi)$, and $\alpha_3/n^{1/2}, \alpha_4/n, \gamma/n$ are the standardized third and fourth cumulants and a measure of nonexponentiality, respectively; the asymptotic assumptions imply that various terms drop off as powers of n and this is indicated explicitly. The α parameters

$$\log f(y;\theta) = \ell(\theta; y) = \sum_{i,j \geqslant 0} a_{ij}(\theta - \theta^0)^i (y - y^0)^j / i!j!, \tag{2.1}$$

$$\bar{A} = \begin{pmatrix} \dfrac{a + 3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} & \dfrac{-\alpha_3}{2n^{1/2}} & -\left\{\dfrac{1 + \alpha_4 - 2\alpha_3^2 - 5\gamma}{2n}\right\} & \dfrac{\alpha_3}{n^{1/2}} & \dfrac{\alpha_4 - 3\alpha_3^2 - 6\gamma}{n} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \gamma/n & - & - \\ \dfrac{-\alpha_3}{n^{1/2}} & 0 & - & - & - \\ \dfrac{-\alpha_4}{n} & - & - & - & - \end{pmatrix}, \tag{2.6}$$

where $a = -\frac{1}{2}\log(2\pi)$, and $\alpha_3/n^{1/2}, \alpha_4/n, \gamma/n$ are the standardized third and fourth cumulants and a measure of nonexponentiality, respectively; the asymptotic assumptions imply that various terms drop off as powers of n and this is indicated explicitly. The α parameters

Exponential family models

- $f(y_i;\varphi) = \exp\{\varphi s(y_i) - \kappa(\varphi)\}h(y_i), \quad y_1,\dots,y_n, \quad$ i.i.d. $\qquad \varphi \in \mathbb{R}$
- Sufficient statistic $s = \Sigma s(y_i)$ — closed under sampling
- inference:

$$f(s;\varphi) = \exp\{\varphi s - n\kappa(\varphi)\}\tilde{h}(s), \qquad \ell'(\hat{\varphi}) = 0 \iff \kappa'(\hat{\varphi}) = s$$

$\hat{\varphi} \leftrightarrow s$

- saddlepoint approximation:

$$\hat{f}_n(s;\varphi) \doteq \frac{c}{\sqrt{(2\pi)}}|j(\hat{\varphi})|^{-1/2}\exp\left\{\ell(\varphi) - \ell(\hat{\varphi}) + s(\varphi - \hat{\varphi})\right\}$$

$j(\varphi) = -\ell_{\varphi\varphi}(\varphi)$

- canonical parameter $\varphi = \partial\ell(\varphi;s)/\partial s$ — up to linear transformations

So many models

- location models reduce $(y_1, \dots, y_n)$ to $(\hat{\theta} \mid a)$ by conditioning $\mathbb{R}^n \to \mathbb{R}$
- approximate location models do the same locally
- exponential models reduce $(y_1, \dots, y_n)$ to s by marginalizing $\mathbb{R}^n \to \mathbb{R}$
- the tangent exponential model does the same locally
- in an arbitrary model $y_1, \dots, y_n \sim f(\mathbf{y}; \theta), \quad \theta \in \mathbb{R}^p$ $\theta = (\psi, \lambda)$

  1. $\mathbb{R}^n \downarrow \mathbb{R}^p$ by conditioning on an approximate a
  2. $\mathbb{R}^p \downarrow \mathbb{R}$ by marginalizing to an approximate s

- use the saddlepoint approximation and the “Fraser series”

Tangent exponential model

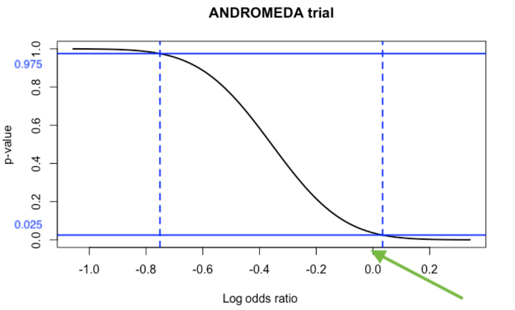
$$
\begin{aligned}
f_{\mathrm{TEM}}(s \mid a; \theta) &= \exp[s^{\mathrm{T}}\varphi(\theta) + \ell\{\theta(\varphi); y^o\}]h(s) \\
&\doteq c|j(\hat{\varphi})|^{-1/2} \exp\left[s^{\mathrm{T}}\{\varphi(\theta) - \varphi(\hat{\theta}^o)\} + \ell(\theta; y^o) - \ell(\hat{\theta}^o; y^o)\right],
\end{aligned}
$$

$$
\varphi(\theta) = \underbrace{\frac{\partial \ell(\theta; y)}{\partial V(y)}}_{\text{canonical parameter}}, \qquad V = \underbrace{\left.\frac{dy}{d\theta^T}\right|_{(y^o, \hat{\theta}^o)}}_{\text{ancillary directions}}, \qquad j(\hat{\varphi}) = -\frac{\partial^2 \ell(\theta; y)}{\partial \varphi \partial \varphi^T}
$$

Significance function

$$p(\psi) \doteq \Phi(r^*_\psi), \qquad r^*_\psi = r_\psi + \frac{1}{r_\psi}\log\left(\frac{Q_\psi}{r_\psi}\right), \qquad f_{\mathrm{TEM}}(s;\varphi) = \exp[s^{\mathrm{T}}\varphi(\theta) + \ell\{\theta(\varphi); y^o\}]h(s)$$

$$r_\psi = \pm\sqrt{[2\{\ell_p(\hat\psi) - \ell_p(\psi)\}]}, \qquad Q_\psi = \frac{|\varphi(\hat\theta) - \varphi(\hat\theta_\psi) \quad \partial\varphi(\hat\theta_\psi)/\partial\lambda^{\mathrm{T}}|}{|\partial\varphi(\hat\theta)/\partial\theta^{\mathrm{T}}|}\left\{\frac{|j(\hat\theta)|}{|j_{\lambda\lambda}(\hat\theta_\psi)|}\right\}^{1/2}$$



Special cases

$p(\psi) \doteq \Phi(r_\psi^*), \qquad r_\psi^* = r_\psi + \frac{1}{r_\psi}\log\left(\frac{Q_\psi}{r_\psi}\right)$ $\qquad \theta = (\psi, \boldsymbol{\lambda})$

1. exp fam $r_\psi = \pm\surd[2\{\ell_p(\hat\psi) - \ell_p(\psi)\}]$, $\qquad Q_\psi = \underbrace{(\hat\psi - \psi) j_p^{1/2}(\hat\psi)}_{\text{Wald}} \underbrace{\frac{|j_{\lambda\lambda}(\hat\psi, \hat\lambda)|^{1/2}}{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|^{1/2}}}_{\text{nuisance par}}$

2. loc-scale $r_\psi = \pm\surd[2\{\ell_p(\hat\psi) - \ell_p(\psi)\}]$, $\qquad Q_\psi = \underbrace{\ell_p'(\psi) j_p^{-1/2}(\hat\psi)}_{\text{score}} \underbrace{\frac{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|^{1/2}}{|j_{\lambda\lambda}(\hat\psi, \hat\lambda)|^{1/2}}}_{\text{nuisance par}}$

profile log-likelihood: $\ell_p(\psi) = \ell(\psi, \hat\lambda_\psi)$

Thinking in pictures

3-6

$\ell(\theta, s) = \ell(\theta) + \varphi'(\theta) s$ $s^\circ = 0$ $\dim \theta, \varphi, s = p$ $H: \psi(\theta) = \psi_0$ $\dim \psi = d$

Rotate coordinates so L° lies on first d axes and tangents to $\psi(\varphi) = \psi_0$ align with last $p-d$ par. axes

(1) The following gives tgts to $\psi(\theta) = \psi_0$ at $\tilde{\theta}^\circ$

$$\chi(\theta) = \psi_\varphi'(\tilde{\theta}^\circ) \cdot \varphi(\theta) = \psi_\theta(\tilde{\theta}^\circ) \varphi_\theta(\tilde{\theta}^\circ) \cdot \varphi(\theta)$$

but they are not rotated coordinates

(2) To get rotated coords so that the first d align with tgt plane at $\tilde{\varphi}^\circ$ we could do as follows:

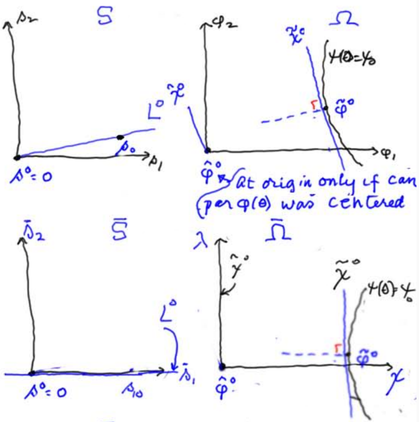
(a) We have given interest $\psi(\theta)$; let $\eta(\theta)$ be a convenient nuisance parameterization. And write $\bar{\theta} = (\psi' \; \eta')'$. This is a new θ defined for convenience.

(b) Let $B(\theta) = \theta_\varphi = \varphi_\theta^{-1}$ Let $\tilde{B}^\circ = B(\tilde{\theta}^\circ)$

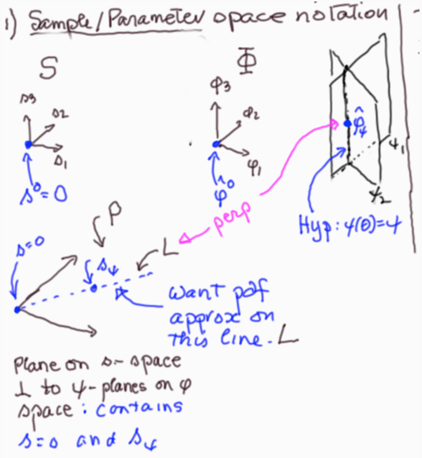
Do a "Pos Lower Triangular" x "Orthogonal" factoring:

$$\tilde{B}^\circ = PH$$

(c) Define $\begin{pmatrix} \chi \\ \lambda \end{pmatrix} = H \varphi(\theta)$ — This is a rotation of original $\varphi(\theta)$ but uses a modified $\bar{\theta} = \begin{pmatrix} \psi \\ \eta \end{pmatrix}$



Rotated coordinates


1) Sample / Parameter space notation
S
Φ
perp
Hyp: ψ(θ)=ψ
want pdf
approx on
this line. L
Plane on s- space
⊥ to ψ- planes on φ
space: contains
s=0 and s_ψ

Relaxing

THANK YOU

AUGUST 6–11
2022
WASHINGTON, DC
JSM
STATISTICS: A FOUNDATION FOR INNOVATION