# Asymptotic theory and practice

Nancy Reid

University of Toronto

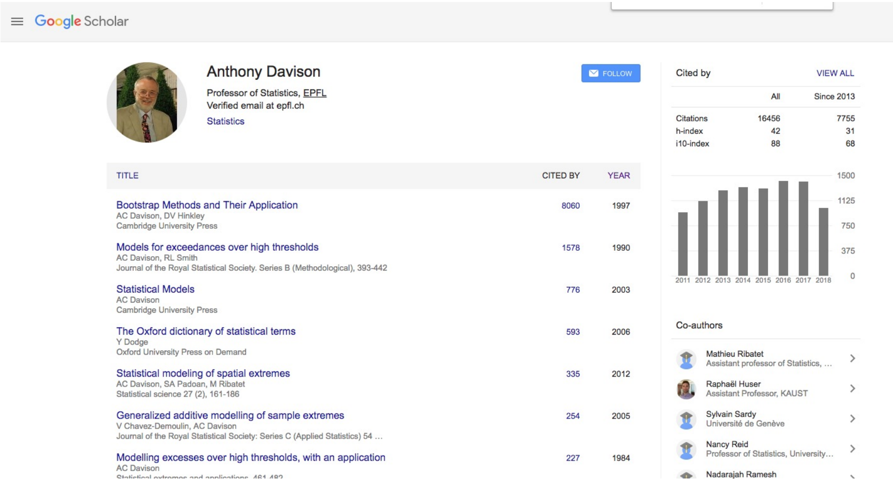
September 15, 2018

DATA, MODELS AND STATISTICAL INFERENCE

A one day workshop in honour of Anthony C. Davison

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

# Practical asymptotics



Google Scholar

Anthony Davison

Professor of Statistics, EPFL

Verified email at epfl.ch

Statistics

FOLLOW

| TITLE | CITED BY | YEAR |
|---|---|---|
| Bootstrap Methods and Their Application<br>AC Davison, DV Hinkley<br>Cambridge University Press | 8060 | 1997 |
| Models for exceedances over high thresholds<br>AC Davison, RL Smith<br>Journal of the Royal Statistical Society. Series B (Methodological), 393-442 | 1578 | 1990 |
| Statistical Models<br>AC Davison<br>Cambridge University Press | 776 | 2003 |
| The Oxford dictionary of statistical terms<br>Y Dodge<br>Oxford University Press on Demand | 593 | 2006 |
| Statistical modeling of spatial extremes<br>AC Davison, SA Padoan, M Ribatet<br>Statistical science 27 (2), 161-186 | 335 | 2012 |
| Generalized additive modelling of sample extremes<br>V Chavez-Demoulin, AC Davison<br>Journal of the Royal Statistical Society: Series C (Applied Statistics) 54 ... | 254 | 2005 |
| Modelling excesses over high thresholds, with an application<br>AC Davison | 227 | 1984 |

Cited by VIEW ALL

| | All | Since 2013 |
|---|---|---|
| Citations | 16456 | 7755 |
| h-index | 42 | 31 |
| i10-index | 88 | 68 |



Co-authors

Mathieu Ribatet — Assistant professor of Statistics, ...

Raphaël Huser — Assistant Professor, KAUST

Sylvain Sardy — Université de Genève

Nancy Reid — Professor of Statistics, University...

Nadarajah Ramesh

*J. R. Statist. Soc.* B (1988)
**50**, *No.* 3, *pp.* 445–461

## Approximate Conditional Inference in Generalized Linear Models

By A. C. DAVISON†

*Imperial College, London, UK*

[Received February 1987. Revised May 1988]

SUMMARY

Easily calculated and accurate approximations are developed to the conditional densities and distributions of sufficient statistics in generalized linear models with canonical link functions. They enable conditional inferences based on modified profile log-likelihoods and tail probabilities to be made using only the deviance and the variance matrix estimate based on fitted models. Examples are given in binary logistic regression and a log-linear model, and the results are applied to added variable tests of model adequacy.

*Keywords*: ADDED VARIABLE; CANONICAL LINK FUNCTION; CONDITIONAL INFERENCE; DOUBLE-SADDLEPOINT APPROXIMATION; GENERALIZED LINEAR MODEL; MODIFIED PROFILE LIKELIHOOD; NATURAL EXPONENTIAL FAMILY; ORTHOGONAL PARAMETERS; PROFILE LIKELIHOOD; SADDLEPOINT METHODS

### 1. INTRODUCTION

### 6.2. *Texas Ozone*

The data in Table 2 are taken from a larger set of observations concerning levels of ambient ozone in East Texas for the years 1981–84. The full data set is analysed by Davison and Hemphill (1987). The data considered here consist of counts of the

# The sandboxes

- J Fluid Mechanics 2008
- Quantitative Finance 2005
- Pain 2006
- Physiological Entomology 2004
- Plant Physiology 2007
- Nonlinear and Nonstationary Signal Processing 2000
- Animal Biology 2006
- Water Resources Research 2013 2001
- Atmospheric Chemistry and Physics 2013 2010 2012
- Forensic Science International 2005
- Agricultural and Forest Meteorology 2010
- Methods in Ecology and Evolution 2014
- Tellus B: Chemical and Physical Meteorology 2010
- Physical Review E 2006
- Ecological Modelling 2006
- Journal of Insect Behavior 2007
- Ecological Entomology 2001
- Biomedical Optics Express 2007
- Atmospheric Environment 1989 1987
- The European Physical Journal 2004
- American J of Sports Medicine 1994
- Atmospheric Environment 1986
- J of Theoretical Biology 2009
- Theoretical and Applied Genetics 2007
- Journal of Hydrology 2018
- Stochastic Transport and Emergent Scaling 2007
- Radiation-risk-protection 1984
- Air Pollution Modeling and its Application 1984
- Mycologia 1959

# ...The sandboxes

| | | |
|---|---|---|
| Implementation of saddlepoint approximations in resampling problems<br>AJ Canty, AC Davison<br>Statistics and Computing 9 (1), 9-15 | 2 | 1999 |
| Implementation of saddlepoint approximations to bootstrap distribution<br>AJ Canty, AC Davison<br>Proceedings of the 28th Symposium on the Interface, Computer Science and ... | 2 | 1996 |
| An exact conditional test for covariance selection models<br>AC Davison, PWF Smith, J Whittaker<br>Australian Journal of Statistics 33 (3), 313-318 | 2 | 1991 |
| The probability distribution of individual exposure due to hypothetical accidental releases of various radionuclides to the atmosphere<br>HM ApSimon, A Davison, AJH Goddard<br>Radiological protection-advances in theory and practice: Proceedings of the ... | 2 | 1982 |
| Myxomycetes from Iraq<br>Y Al-Doory<br>Mycologia, 299-300 | 2 | 1959 |
| Parameter estimation for discretely-observed linear birth-and-death processes<br>AC Davison, S Hautphenne, A Kraus<br>arXiv preprint arXiv:1802.05015 | 1 | 2018 |
| Optimal regionalization of extreme value distributions for flood estimation<br>P Asadi, S Engelke, AC Davison<br>Journal of Hydrology 556, 182-193 | 1 | 2018 |
| Meta-analysis of incomplete microarray studies<br>A Zollinger, AC Davison, DR Goldstein<br>Biostatistics 16 (4), 686-700 | 1 | |

# …Practical asymptotics

*J. R. Statist. Soc.* B (1988)
**50**, *No.* 3, *pp.* 445–461

## Approximate Conditional Inference in Generalized Linear Models

By A. C. DAVISON†

*Imperial College, London, UK*

[Received February 1987. Revised May 1988]

SUMMARY

Easily calculated and accurate approximations are developed to the conditional densities and distributions of sufficient statistics in generalized linear models with canonical link functions. They enable conditional inferences based on modified profile log-likelihoods and tail probabilities to be made using only the deviance and the variance matrix estimate based on fitted models. Examples are given in binary logistic regression and a log-linear model, and the results are applied to added variable tests of model adequacy.

*Keywords*: ADDED VARIABLE; CANONICAL LINK FUNCTION; CONDITIONAL INFERENCE; DOUBLE-SADDLEPOINT APPROXIMATION; GENERALIZED LINEAR MODEL; MODIFIED PROFILE LIKELIHOOD; NATURAL EXPONENTIAL FAMILY; ORTHOGONAL PARAMETERS; PROFILE LIKELIHOOD; SADDLEPOINT METHODS

### 1. INTRODUCTION

$\beta_1$ will in general be lost if inference is based on the conditional likelihood $l_1$ alone. See Barnard (1984), for a simple example. Further, it may be desirable to base a test of a hypothesis about the value of $\beta_1$ on the conditional distribution of $S_1$ given $S_2 = s_2$, which is independent of $\beta_2$ by sufficiency. However, even in cases where these simplifications can in principle be made, it may be difficult to make the calculations required.

This paper is concerned with the construction of accurate approximations to conditional likelihoods and significance probabilities for conditional tests in generalized linear models with canonical link functions. An important feature of the approximations is that they use only standard output of regression packages, the deviance and variance–covariance matrix, so that they are calculable, for example, using GLIM. The results are based on saddlepoint approximations and extend slightly those of Barndorff-Nielsen and Cox (1979), but the development is different. Sections 2 and 3 respectively describe saddlepoint expansions and generalized linear models, which are then combined in Section 4. Section 5 discusses connections with

† *Address for correspondence*: Department of Mathematics, Imperial College of Science and Technology, Huxley Building, Queens Gate, London, SW7 2BZ, UK.

# Approximate conditional inference

- linear exponential family model

$$f(y;\theta) = \exp\{\psi^{\mathrm{T}} s_1(y) + \lambda^{\mathrm{T}} s_2(y) - c(\psi,\lambda)\} h(y), \quad y = (y_1, \ldots, y_n)$$

- sufficient statistic

$$f(s;\theta) = \exp\{\psi^{\mathrm{T}} s_1 + \lambda^{\mathrm{T}} s_2 - c(\psi,\lambda)\} \tilde{h}(s), \qquad n \downarrow p$$

- conditional inference

$$f(s_1 \mid s_2;\psi) = \exp\{\psi^{\mathrm{T}} s_1 - \tilde{c}(\psi)\} \tilde{h}_2(s_1), \qquad p \downarrow d$$

dim $\psi$

- saddlepoint approximation

reinterpreted

$$f(s_1 \mid s_2;\psi) \doteq c \underbrace{\exp\{\ell(\hat{\psi},\hat{\lambda}) - \ell(\psi,\hat{\lambda}_\psi)\}}_{\text{likelihood ratio}} \underbrace{|j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)|^{1/2}}_{\text{null model}} \underbrace{|j_{\theta\theta}(\hat{\psi},\hat{\lambda})|^{-1/2}}_{\text{full model}}$$

# Approximate conditional inference in generalized linear models

- saddlepoint approximation   reinterpreted

$$f(s_1 \mid s_2; \psi) \doteq c \underbrace{\exp\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi)\}}_{\text{likelihood ratio}} \underbrace{|j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|^{1/2}}_{\text{null model}} \underbrace{|j_{\theta\theta}(\hat{\psi}, \hat{\lambda})|^{-1/2}}_{\text{full model}}$$

- generalized linear model   canonical link

$$\theta = X\beta; \quad \psi = \beta_{(1)}; \quad s = X^{\mathrm{T}}y; \quad s_1 = X^{\mathrm{T}}y_{(1)}$$

- likelihood ratio $\longrightarrow$ `deviance full - deviance null`
- constrained mle $\longrightarrow$ `offset = x2`
- $j_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)$ $\longrightarrow$ `coef_null$ covariance`
- $j_{\theta\theta}(\hat{\psi}, \hat{\lambda})$ $\longrightarrow$ `coef_full$covariance matrix full`

$j_{\theta\theta}(\theta) = -\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}$

- Davison (1988) provides similar analysis for unknown scale parameter $\phi$
- It turns out that other parametric models can be similarly analysed

although nuisance parameter $\lambda$ is eliminated by marginalization

# Approximate $p$-values

- saddlepoint approximation — reinterpreted

$$f(s_1 \mid s_2; \psi) \doteq c \underbrace{\exp\{\ell(\hat\psi, \hat\lambda) - \ell(\psi, \hat\lambda_\psi)\}}_{\text{likelihood ratio}} \underbrace{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|^{1/2}}_{\text{null model}} \underbrace{|j_{\theta\theta}(\hat\psi, \hat\lambda)|^{-1/2}}_{\text{full model}}$$

- special case, scalar parameter of interest $\psi \in \mathbb{R}$ — $d = 1$
- find distribution function at $s_1^o$ — observed value

$$F(s_1^o; \psi) \doteq \Phi(r_\psi^*)$$

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

$r_\psi = r_\psi(s_1^o) \longrightarrow$ `log-likelihood root`
$q_\psi = q_\psi(s_1^o) \longrightarrow$ `standardized MLE`

# ... approximate $p$-values



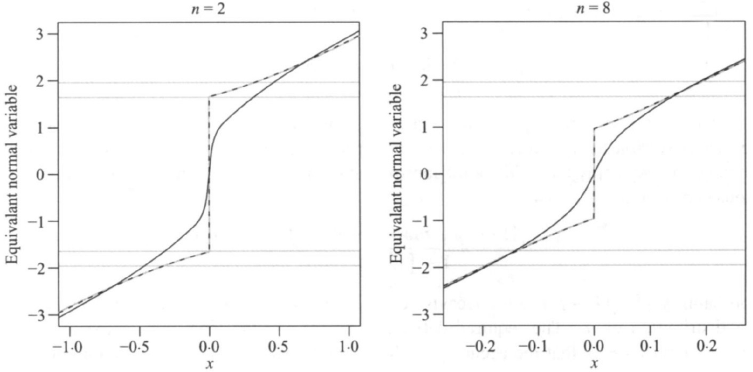
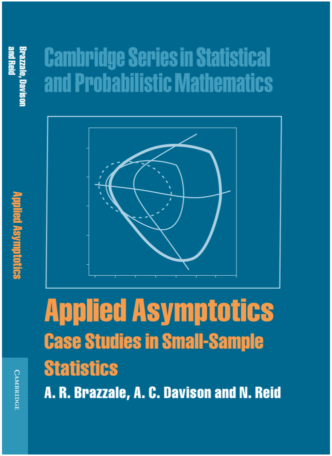



Fig. 1. Comparison of saddlepoint approximations $\tilde{F}$ (solid) and $\hat{F}$ (dashes) with exact distribution $F$ (solid, grey), which is identical to $\hat{F}$, for normal/point mass mixture, with $p = 0{\cdot}05$, $\mu = 0$, $\sigma = 1$ and $n = 2, 8$. For clarity, the vertical scale is the equivalent normal variable $\Phi^{-1}\{F(x)\}$. The horizontal lines correspond to probabilities 0·025, 0·05, 0·95 and 0·975.

Table 1. *Percent relative errors for estimation of quantiles using saddlepoint approximations $\tilde{F}$ and $\hat{F}$, based on the normal mixture distribution, with*




*Statistical Science*
2008, Vol. 23, No. 4, 465–484
DOI: 10.1214/08-STS273


# Accurate Parametric Inference for Small Samples

**Alessandra R. Brazzale and Anthony C. Davison**

# Practical saddlepoint approximations

| | | |
|---|---|---|
| Likelihood estimation for the *INAR*($p$) model | *JASA* | D & Pedeli, Fokianos |
| Saddlepoint approximation for mixture models | *Biometrika* | D & Mastropietro |
| The Banff challenge: statistical detection of a noisy signal | *Statistical Science* | D & Sartori |
| Three examples of accurate likelihood inference | *American Statistician* | D & Lozada-Can |
| Saddlepoint approximations as smoothers | *Biometrika* | D & Wang |
| Implementation of saddlepoint approximation in resampling problems | *JCGS* | D & Canty |

# Vector parameter of interest

- saddlepoint approximation — reinterpreted

$$f(s_1 \mid s_2; \psi) \doteq c \underbrace{\exp\{\ell(\hat\psi, \hat\lambda) - \ell(\psi, \hat\lambda_\psi)\}}_{\text{likelihood ratio}} \underbrace{|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|^{1/2}}_{\text{null model}} \underbrace{|j_{\theta\theta}(\hat\psi, \hat\lambda)|^{-1/2}}_{\text{full model}}$$

- a more general version

$$f_{SP}\{s(t); \psi_0\} = c \exp[\ell\{\hat\varphi_{\psi_0}; s(t)\} - \ell\{\hat\varphi; s(t)\}] |j_{(\lambda\lambda)}(\hat\varphi_{\psi_0})|^{1/2} |j_{\varphi\varphi}(\hat\varphi)|^{-1/2}$$

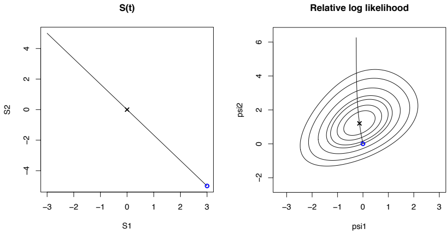
- $s$ is constrained to $\mathcal{L}_\psi$, where the nuisance parameter is fixed at $\hat\lambda_\psi$

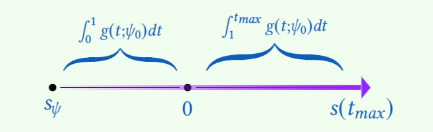
- $s$ is further constrained to a line in $\mathcal{L}_\psi$ on which we measure the discrepancy from $H_0 : \psi = \psi_0$

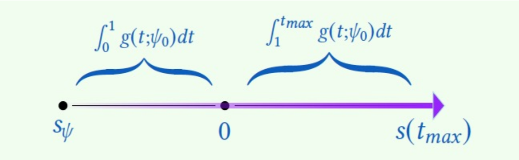
# … directional testing

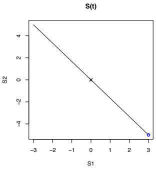
- $p \downarrow d : \theta = (\psi, \lambda)$, $\quad \hat{\lambda}_\psi$ constrained mle $\qquad \mathcal{L}_\psi = \{s \mid \hat{\lambda}_\psi = \hat{\lambda}^o_\psi\}$

- $d \downarrow 1$: line on $\mathcal{L}_\psi$ between expected, $s_\psi$ and observed $s^o$
- compute directional $p$-value on this line $\qquad p = 5, d = 2$






- need density on this line $s(t) = s_\psi + t(s^0 - s_\psi)$

$$p(\psi_0) = \frac{\int_1^{t_{max}} g(t;\psi_0)dt}{\int_0^{t_{max}} g(t;\psi_0)dt}$$

- use saddlepoint approximation to get density for $s \in \mathcal{L}_\psi$
- $g(t;\psi_0) = t^{d-1} f_{SP}\{s(t)\}$

implicitly creating a one-dimensional model

# The line $s(t)$

$t = 0$

| | | |
|---|---|---|
| 10 | 10 | 10 |
| 20 | 20 | 20 |

independence (null hypothesis)

$t = 0.5$

| | | |
|---|---|---|
| 11.0 | 11.5 | 7.5 |
| 19.0 | 18.5 | 22.5 |

a table on the line

$t = 1$

| | | |
|---|---|---|
| 12 | 13 | 5 |
| 18 | 17 | 25 |

observed data

$t = 2$

| | | |
|---|---|---|
| 14 | 16 | 0 |
| 16 | 14 | 30 |

largest value of $t$

# $F$-tests

- normal theory linear model $y = X\beta + \epsilon$

- linear constraint $A\beta = 0, \quad A_{d\times p}$ $\psi_0 = 0$

$$
\begin{aligned}
p &= \frac{\int_1^{t_{max}} g(t;\psi_0)dt}{\int_0^{t_{max}} g(t;\psi_0)dt} \\
g(t;\psi_0) &\propto t^{d-1}\{\hat\sigma^2(t)\}^{(n-p-2)/2} \\
n\hat\sigma^2(t) &= (Y - X\hat\beta)^{\mathrm{T}}\{I - t^2X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}\}(Y - X\hat\beta) \\
p &= \ldots = \Pr\{F_{d,n-p} \geq MSR/MSE\}
\end{aligned}
$$

McCormack et al., 2018; Sartori & Ruffato

# … $F$-tests

- ratio of exponential rates $y_{1j} \sim \theta_1 e^{-\theta_1 y_{1j}}, y_{2j} \sim \theta_2 e^{-\theta_2 y_{2j}}, \quad j = 1, \ldots, n$
- $H : \theta_1/\theta_2 = \psi$
- directional $p$-value

$$\frac{\text{pr}(F_{2n_1,2n_2} > \psi \bar{y}_2/\bar{y}_1)}{\text{pr}(F_{2n_1,2n_2} > 1)}$$

- ratio of normal variances $H : \sigma_1^2/\sigma_2^2 = 1$
- directional $p$-value

$$\frac{\text{pr}(F_{n_2-1,n_1-1} > \psi s_2^2/s_1^2)}{\text{pr}(F_{n_2-1,n_1-1} > \frac{n_2(n_1-1)}{n_1(n_2-1)})}$$

- multivariate normal mean $y_i \sim N_q(\mu, \Sigma), \quad H : \mu = 0$
- directional $p$-value

$$\text{pr}\{F_{p,n-p} > (n-p)T^2/p(n-1)\}$$

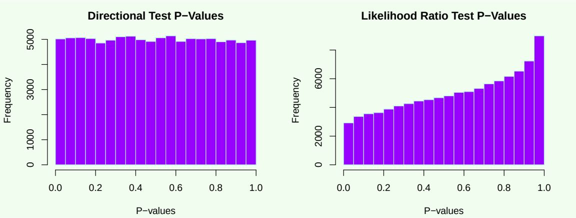
McCormack et al 2018

# Moderate dimension

- need $n > p$ ( $n \downarrow p \downarrow d$)
- seems to accommodate large number of parameters of interest and large number of nuisance parameters

- nuisance parameters eliminated using adjustment to log-likelihood
- this seems the most important aspect of HOA

Exponential rates
$n = 250, p = 50, d = 49$
1 nuisance par

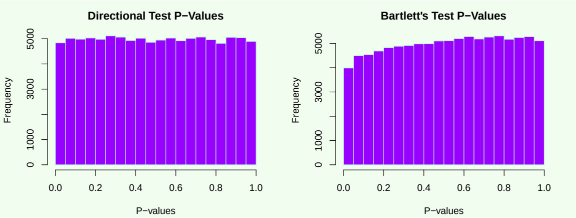


# Moderate dimension

- need $n > p$ ( $n \downarrow p \downarrow d$)
- seems to accommodate large number of parameters of interest and large number of nuisance parameters

- nuisance parameters eliminated using adjustment to log-likelihood
- this seems the most important aspect of HOA

Normal variances
$n = 250, p = 100, d = 49$
51 nuisance pars.



# … moderate dimension

| Example | data $n$ | parameter $p$ | par. of int. $d$ |
|---|---|---|---|
| contingency tables | 1000 | 36 | 10 |
| normal variances | 5000 | 2000 | 999 |
| exponential rates | 5000 | 1000 | 999 |
| covariance selection | 60 ($\mathcal{N}_q$) | 1275 | 1176 |
| normal means | 1000 | 400 | 199 |
| marginal independence | 60 ($\mathcal{N}_q$) | 1275 | 1000 |
| Box-Cox | 48 | 14 | 6 |

Improved asymptotics seems to adjust well for large numbers of nuisance parameters

Davison et al. 2014

Sartori et al. 2016

McCormack et al. 2018

New asymptotic theory being developed for $p/n \to \kappa \in (0, 1)$

# Thank You!

**DATA, MODELS AND STATISTICAL INFERENCE**

**A one day workshop in honour of Anthony C. Davison**

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

# Some technical details

- exponential model, linear hypothesis

$$\exp\{\psi^{\mathrm{T}} s_1 + \lambda^{\mathrm{T}} s_2 - \kappa(\psi, \lambda)\} h(s)$$

- based on conditional distribution of $s_1$, given $s_2$

- exponential model, nonlinear hypothesis

$$\exp\{\varphi(\theta)^{\mathrm{T}} s - \kappa(\varphi)\} h(s)$$

- uses a marginalization step to eliminate nuisance parameter $\lambda(\psi)$

- in a general model, use an approximating exponential family model as first step ($n \downarrow p$)
-

$$p = \frac{\int_1^{t_{max}} g(t; \psi_0) dt}{\int_0^{t_{max}} g(t; \psi_0) dt}$$

# ... some technical details

$$p = \frac{\int_1^{t_{max}} g(t;\psi_0)dt}{\int_0^{t_{max}} g(t;\psi_0)dt}$$

$$g(t;\psi_0) = t^{d-1} f_{SP}\{s(t);\psi_0\}$$

$$f_{SP}\{s(t);\psi_0\} = c\exp[\ell\{\hat{\varphi}_{\psi_0}; s(t)\} - \ell\{\hat{\varphi}; s(t)\}]|j_{\varphi\varphi}(\hat{\varphi})|^{-1/2}|j_{(\lambda\lambda)}(\hat{\varphi}_{\psi_0})|^{1/2}$$

$$\ell(\varphi;s) = \varphi^{\mathrm{T}}(\theta)s + \log f(y^0;\theta)$$

$$s \in \text{ plane } \mathcal{L}_{\psi_0} = \{s \mid \hat{\lambda}_{\psi_0} \text{ fixed } \}$$

$s$ on line in plane joining $s_{\psi_0}$ and $s^0$