Accurate directional inference for vector parameters

Nancy Reid

October 28, 2016

with Don Fraser, Nicola Sartori, Anthony Davison

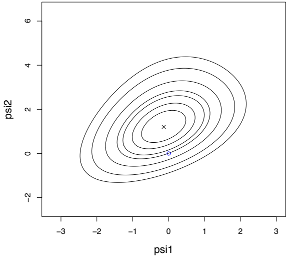
Imperial College London

Parametric models and likelihood

- model $f(y;\theta)$, $\theta \in \mathbb{R}^p$
- data $y = (y_1, \ldots, y_n)$ independent observations
- log-likelihood function $\ell(\theta; y) = \log f(y;\theta)$
- parameter of interest $\theta = (\psi, \lambda)$, $\psi \in \mathbb{R}^d$

- likelihood inference $w(\psi) = 2\{\ell(\hat\psi, \hat\lambda) - \ell(\psi, \hat\lambda_\psi)\} \; \dot\sim \; \chi^2_d$

Likelihood Contours



$$w_B(\psi) = \frac{w(\psi)}{1 + B(\psi)} \; \dot\sim \; \chi^2_d \quad O_p(n^{-2})$$

$$B(\psi) = \mathrm{E}\{w(\psi)\}/d$$

$$w^*(\psi) = w(\psi)\left\{1 - \frac{\log\gamma(\psi)}{w(\psi)}\right\}$$

Skovgaard, 2001

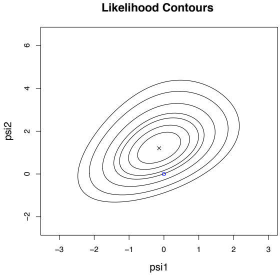
Example: 2×3 contingency table

- activity amongst psychiatric patients — Everitt, 1992

	Affective disorders	Schizophrenics	Neurotics
Retarded	12	13	5
Not retarded	18	17	25

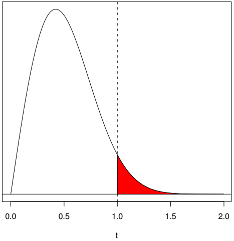
- model: log-linear $y \sim$ Poisson, $\log\{\mathrm{E}(y)\} = X\theta, \quad \theta \in \mathbb{R}^6$
- log-likelihood $\ell(\theta; y) = \theta^{\mathrm{T}} X^{\mathrm{T}} y - 1^{\mathrm{T}} e^{X\theta} = \theta^{\mathrm{T}} s - c(\theta)$
- $\theta = (\psi, \lambda) \qquad \psi \in \mathbb{R}^2, \lambda \in \mathbb{R}^4$
- (ψ_1, ψ_2) interaction parameters
- $H_0 : \psi = \psi_0 = (0, 0)$ independence
- log-likelihood $\ell(\psi, \lambda; y) = \psi^{\mathrm{T}} s_1 + \lambda^{\mathrm{T}} s_2 - c(\psi, \lambda)$

Testing $\psi = 0$



$w(\psi_0) \dot{\sim} \chi^2_2$ p-value 0.047

$w^*(\psi_0)$ p-value 0.048



directional p-value 0.050

exact p-value 0.051

Linear exponential families

- model $f(y;\theta) = \exp[\varphi(\theta)^{\mathrm{T}} s(y) - c\{\varphi(\theta)\} - d(y)]$ $\quad y_1, \ldots, y_n$ i.i.d

- sufficient statistic $s(y)$

$$f(s;\theta) = \exp[\varphi(\theta)^{\mathrm{T}} s - nc\{\varphi(\theta)\} - \tilde{d}(s)] \qquad = \int_{\{y:s(y)=s\}} f(y;\theta)dy$$

- reduce dimension from n to p by marginalization

- linear parameter of interest $\varphi(\theta) = \theta = (\psi, \lambda)$

- model $f(s_1, s_2; \psi, \lambda) = \exp\{\psi^{\mathrm{T}} s_1 + \lambda^{\mathrm{T}} s_2 - nc(\psi, \lambda) - \tilde{d}(s)\}$

- conditional density $f(s_1 \mid s_2; \psi) = \exp\{\psi^{\mathrm{T}} s_1 - n\tilde{c}_2(\psi) - \tilde{d}_2(s_1)\}$

- reduce dimension from p to d by conditioning

... conditional density

- $f(s_1 \mid s_2; \psi) = \exp\{\psi^{\mathrm{T}} s_1 - n\tilde{c}_2(\psi) - \tilde{d}_2(s_1)\}$ $\quad \theta = (\psi, \lambda)$
- $f(s_1 \mid s_2; \psi) = \dfrac{f(s_1, s_2; \psi, \lambda)}{f(s_2; \psi, \lambda)} \propto f(s; \psi, \lambda)$ $\quad$ with s_2 held fixed
- s_2 held fixed $\iff$ $\hat{\lambda}_\psi$ held fixed $\quad \hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$
- saddlepoint approximation:

$$f(s_1 \mid s_2; \psi) \doteq c \exp[\ell(\hat{\theta}^0_\psi; s) - \ell\{\hat{\theta}(s); s\}] |j\{\hat{\theta}(s); s\}|^{-1/2}, \quad s \in L^0_\psi$$

- plane $L^0_\psi = \{s \in \mathbb{R}^p : \hat{\lambda}_\psi = \hat{\lambda}_{\psi^0}\}$ $\quad = \{s \in \mathbb{R}^p : s_2 = s_2^0\}$
- $\hat{\theta}(s) : \partial\ell(\theta; s)/\partial\theta = 0$ $\quad$ m.l.e. as a function of the variable
- log-likelihood function $\ell(\theta; s) = \psi^{\mathrm{T}} s_1 + \lambda^{\mathrm{T}} s_2 - c(\theta), \quad s \in L^0_\psi$

... conditional density

$$f(s_1 \mid s_2; \psi) \doteq c \exp[\ell(\hat{\theta}^0_\psi; s) - \ell\{\hat{\theta}(s); s\}] |j\{\hat{\theta}(s); s\}|^{-1/2}, \quad s \in L^0_\psi$$

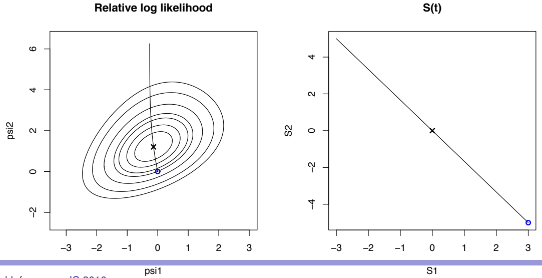
	Affective disorders	Schizophrenics	Neurotics
Retarded	12	13	5
Not retarded	18	17	25

$s_1 \in \mathbb{R}^2$, $s_2 \in \mathbb{R}^4$

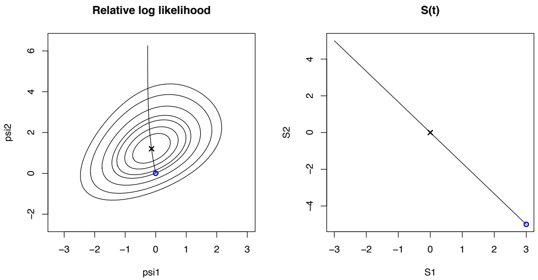
L^0_ψ: all 2×3 tables with the same row and column totals

Directional tests

- measure the directed departure from H_0 in L_ψ^0
- s_ψ: expected value of s_1, under H_0
- s^0: observed value of s_1 $\quad = 0$ from centering
- L_ψ^*: line through these two points
- $L_\psi^* = ts^0 + (1-t)(s_\psi - s^0), \quad t \in \mathbb{R}$



... directional tests $L^*_\psi = ts^0 + (1-t)s_\psi$



○ null hypothesis of independence $t = 0$

X observed value of s $t = 1$

p-value compares probability from $\times$ to ∞ to that from ○ to ∞
along the line in the sample space curve in the parameter space

like a 2-sided p-value Pr (response $>$ observed $|$ response > 0)

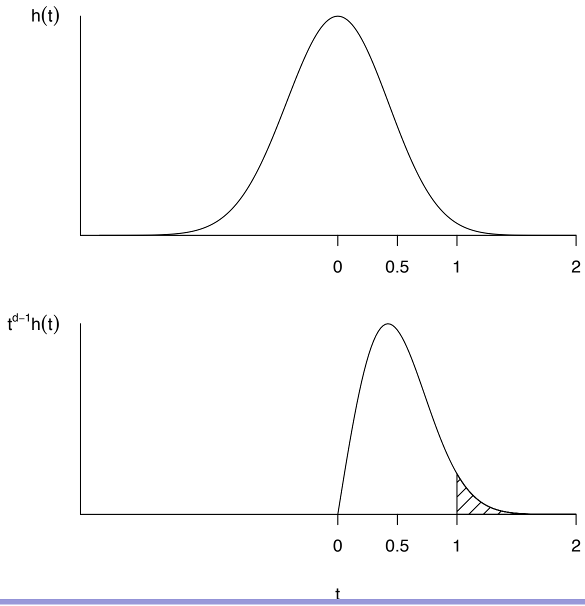
... directional p-value

- p-value $= \dfrac{\int_1^\infty t^{d-1} f\{s(t);\psi\}dt}{\int_0^\infty t^{d-1} f\{s(t);\psi\}dt}$ $\qquad$ $s(t)$ along the line ${L_\psi}^*$
- t^{d-1} from change to polar coordinates $\qquad$ $||s||$, conditional on $s/||s||$
- Simplifications:
  - 1: $L^*_\psi \subset L^0_\psi \subset \mathbb{R}^p$
  - 2: ratio of two integrals – drop any terms that don't depend on t
  - 3: saddlepoint approximation

$$f(s;\psi) \doteq c \exp[\ell(\hat\theta^0_\psi; s) - \ell\{\hat\theta(s); s\}]|j\{\hat\theta(s); s\}|^{-1/2}, \quad s \in L^0_\psi$$

$$h(t) = f\{s(t);\psi\}$$

2 × 3 table



$t = 0$

10	10	10
20	20	20

$t = 0.5$

11.0	11.5	7.5
19.0	18.5	22.5

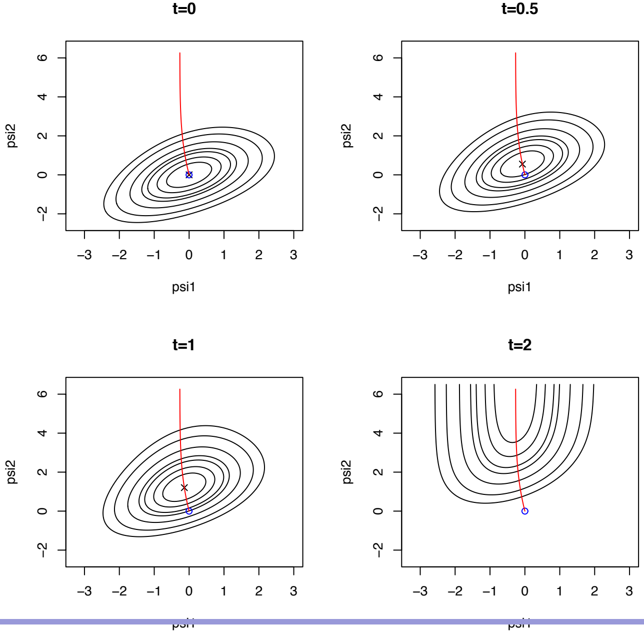
$t = 1$

12	13	5
18	17	25

$t = 2$

14	16	0
16	14	30

... 2 × 3 table



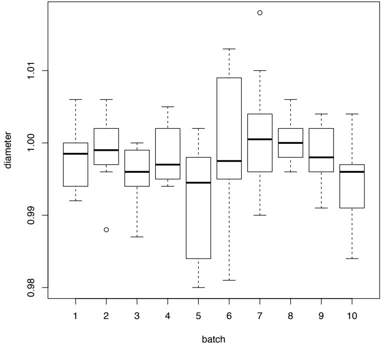
... 2 × 3 table

simulations

Nominal	0.010	0.025	0.050	0.100	0.250	0.500
LRT	0.011	0.028	0.055	0.107	0.260	0.510
Directional	0.010	0.024	0.050	0.100	0.250	0.501
Skovgaard, 2001	0.010	0.025	0.050	0.101	0.251	0.501

Nominal	0.750	0.900	0.950	0.975	0.990
LRT	0.757	0.905	0.952	0.974	0.992
Directional	0.752	0.902	0.950	0.973	0.992
Skovgaard, 2001	0.752	0.900	0.950	0.973	0.991

Example: comparison of normal variances

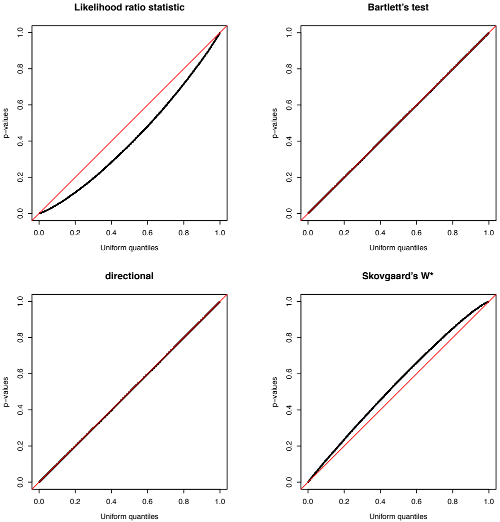


Likelihood ratio statistic	0.0042
Directional	0.0389
Skovgaard, 2001	0.0622
Bartlett's test	0.0136

F-test

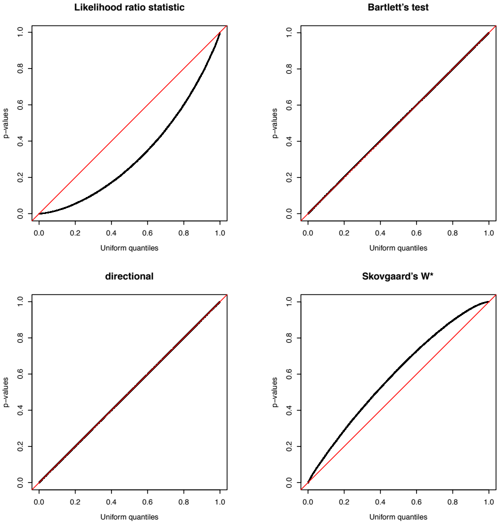
Simulations

3 groups, 10 observations per group



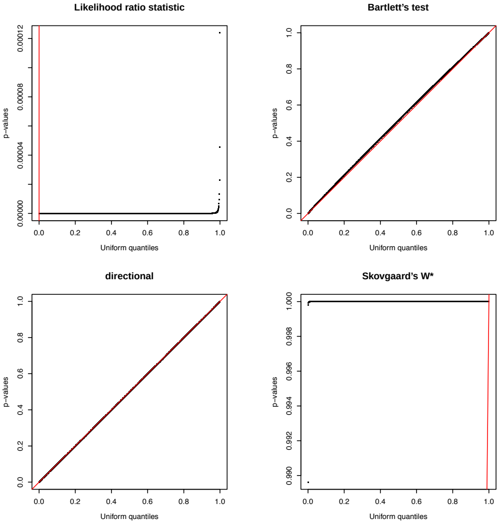
Simulations

3 groups, 5 observations per group



Simulations

1000 groups, 5 observations per group



dimension ψ 999; dimension nuisance 1001

Example: covariance selection

- model $y_i \sim N_q(\mu, \Lambda^{-1})$ inverse covariance matrix
- linear exponential family

$$\ell(\theta; y) = \frac{n}{2}\log|\Lambda| - \frac{1}{2}\mathrm{tr}(\Lambda y^{\mathrm{T}} y) + 1^{\mathrm{T}} y\xi - \frac{n}{2}\xi^{\mathrm{T}}\Lambda\xi \qquad s_1, s_2$$

- $\theta = (\xi, \Lambda) = (\Lambda\mu, \Lambda)$
- H_0: some off-diagonal elements of Λ are 0
  $\psi_1 = \cdots = \psi_d = 0$

  conditional independence
- need constrained m.l.e: use `fitConGraph` in `ggm`
- $\hat{\Lambda}^{-1}(t) = t\hat{\Lambda}^{-1} + (1-t)\hat{\Lambda}_0^{-1}$ m.l.e. along the line
- $f\{s(t); \psi\} \propto |t\hat{\Lambda}^{-1} + (1-t)\hat{\Lambda}_0^{-1}|^{(n-q-2)/2}$

... covariance selection

simulations

Nominal	0.010	0.025	0.050	0.100	0.250	0.500
LRT	0.055	0.105	0.170	0.270	0.487	0.730
Directional	0.011	0.026	0.050	0.101	0.248	0.498
Skovgaard, 2001	0.007	0.018	0.036	0.074	0.196	0.422

Nominal	0.750	0.900	0.950	0.975	0.990
LRT	0.895	0.967	0.985	0.994	0.998
Directional	0.749	0.899	0.949	0.974	0.990
Skovgaard, 2001	0.678	0.852	0.919	0.955	0.980

Covariance matrix 11×11; dimension of $\psi = 45$

first order Markov dependence

Tangent exponential model

$n \downarrow p$ (dimension of y to dimension of θ)

- Every model $f(y;\theta)$ on $\mathbb{R}^n$ can be approximated by an exponential family model:

$$f_{TEM}(s;\theta)ds = \exp\{\varphi(\theta)^{\mathrm{T}} s + \ell(\theta)\} f_0(s) ds$$

- s is a score variable on $\mathbb{R}^p$ — $s(y) = -\ell_{\varphi}(\hat{\theta}^0; y)$
- $\ell(\theta) = \ell(\theta; y^0)$ is the observed log-likelihood function
- $\varphi(\theta) = \varphi(\theta; y^0)$ is the canonical parameter $\in \mathbb{R}^p$

  to be described

- matches log-likelihood function at y^0, and its first derivative on the sample space, at y^0 — by construction
- implements conditioning on an approximate ancillary statistic

  contrast with exp fam

Aside: canonical parameter $\varphi(\theta)$

- if $f(y;\theta)$ is an exponential family, φ is sitting in the model

- if not find a pivotal quantity $z_i = z_i(y_i;\theta)$ with a fixed distribution

  example: $(y_i - \mu)/\sigma$

- define $V_i = -\left(\frac{\partial z_i}{\partial y_i}\right)^{-1} \frac{\partial z_i}{\partial \theta}\Big|_{y=y^0,\theta=\hat{\theta}^0}$ a vector of length p

$$\varphi(\theta) = \varphi(\theta; y^0) = \sum_{i=1}^{n} \frac{\partial \ell(\theta; y^0)}{\partial y_i} V_i = \ell_{;V}(\theta; y^0)$$

Testing a value for ψ

eliminating nuisance parameter: $p \downarrow d$

- $\theta = (\psi, \lambda)$; $\varphi = \varphi(\theta)$ is the canonical parameter of the TEM
- with ψ fixed by the hypothesis, we can integrate out the nuisance parameter by Laplace approximation
- define the same plane in the score space L^0_ψ, where $\hat\varphi_\psi$ is fixed at its observed value

$$f(s;\psi) = c\exp\{\ell(\hat\varphi^0_\psi; s) - \ell(\hat\varphi(s); s)\}|j_{\varphi\varphi}(\hat\varphi(s); s)|^{-1/2}|j_{(\lambda\lambda)}(\hat\varphi^0_\psi; s)|^{1/2}$$

$$s \in L^0_\psi$$

$$p\text{-value} = \frac{\int_1^\infty t^{d-1} f\{s(t);\psi\}dt}{\int_0^\infty t^{d-1} f\{s(t);\psi\}dt}$$

$s(t)$ along the line L^*_ψ

Example: marginal independence

- $y_i \sim N_q(\mu, \Sigma)$, $\quad H_0$: some entries of Σ are 0
- estimate Σ under H_0 using `fitCovGraph`
- $\ell(\Sigma) = \frac{n-1}{2}[\log|\varphi(\Sigma)| - \frac{n-1}{2}\mathrm{tr}\,\{\varphi(\Sigma)S\}], \quad \varphi(\Sigma) = \Sigma^{-1}$

  $S =$ sample covariance
- $f\{s(t);\psi) \propto |t\hat{\Sigma} + (1-t)\hat{\Sigma}_0|^{(n-q-2)/2}|j_{(\lambda\lambda)}\{\hat{\Sigma}_0; s(t)\}|^{1/2}$
- $j_{(\lambda\lambda)}$ needs $\partial\ell(\Sigma)/\partial(\sigma_{jk})$ for the non-zero elements
- $j_{\lambda_j\lambda_k}\{\hat{\Sigma}_0; s(t)\} = \frac{n-1}{2}\left(\mathrm{tr}(A_{kj} + t\left[\mathrm{tr}\{(A_{kj} + A_{jk})(\hat{\Sigma}_0^{-1}\hat{\Sigma} - I_q)\}\right]\right)$
- $A_{kj} = \Sigma_0^{-1}(\partial\Sigma/\partial\lambda_k)\Sigma_0^{-1}(\partial\Sigma/\partial\lambda_j)$

... independence

simulations; $n = 60; 600$, $d = 1000$

Nominal	0.010	0.025	0.050	0.100	0.250	0.500
LRT	1.00	1.00	1.00	1.00	1.00	1.00
Directional	0.007	0.024	0.049	0.097	0.250	0.497
Skovgaard, 2001	0.000	0.000	0.000	0.001	0.006	0.026

Nominal	0.750	0.900	0.950	0.975	0.990
LRT	1.00	1.00	1.00	1.00	1.00
Directional	0.749	0.904	0.951	0.975	0.990
Skovgaard, 2001	0.099	0.225	0.333	0.440	0.570

Nominal	0.010	0.025	0.050	0.100	0.250	0.500
LRT	0.065	0.122	0.199	0.311	0.538	0.786
Directional	0.010	0.024	0.049	0.098	0.251	0.496
Skovgaard, 2001	0.003	0.010	0.022	0.050	0.150	0.352

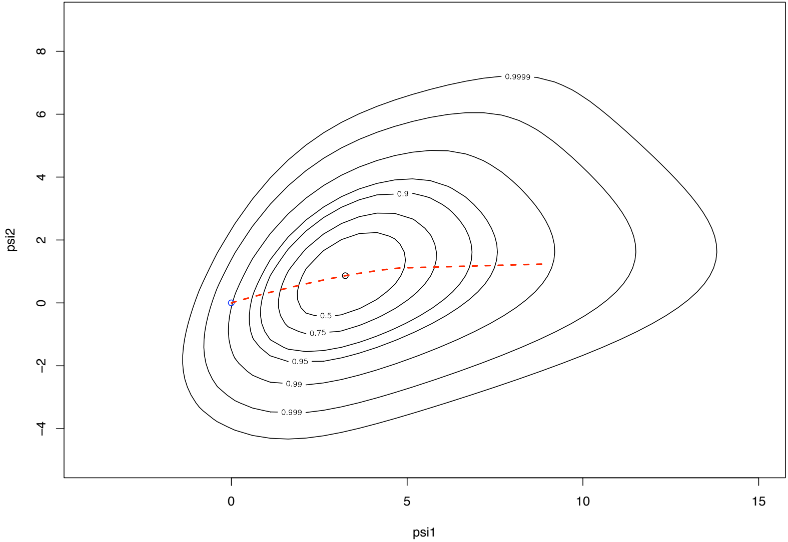
Nominal	0.750	0.900	0.950	0.975	0.990
LRT	0.927	0.979	0.991	0.997	0.999
Directional	0.751	0.898	0.950	0.975	0.989
Skovgaard, 2001	0.622	0.819	0.894	0.943	0.974

Conclusion

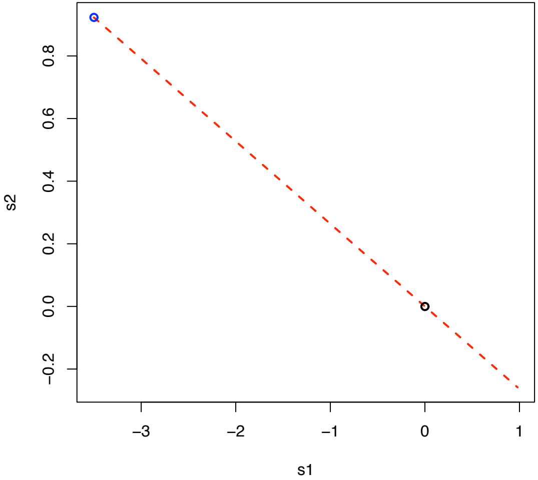
- different way to assess vector parameters
- incorporates information in the direction of departure
- easy to compute: two model fits, plus 1-d numerical integration
- accurate conditionally, by construction, and unconditionally

  simulations
- can be used in models of practical interest
- exponential family model not necessary – easily generalized using approximate exponential family model

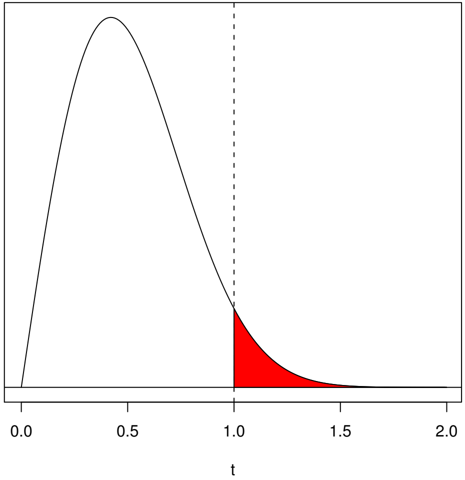
... conclusion



... conclusion



... conclusion



References

Fraser, D.A.S. and Massam, H. (1985). Conical tests. *Stat. Hefte*

Skovgaard, I.M. (1988). Saddlepoint expansions for directional test probabilities. *JRSS B*

Skovgaard, I.M. (2001). Likelihood asymptotics. *SJS*

Davison, A.C., Fraser, D.A.S., Reid, N., Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families.*JASA*

Fraser, D.A.S., Reid, N., Sartori, N. (2016). Accurate directional inference for vector parameters. *Biometrika* **103**, 625–639.

Jensen, J.L. (2016). On the use of saddlepoint approximations in high-dimensional inference. preprint

Lee, S.M.S. and Young, A. (2016). Distribution of likelihood-based p-values under a local alternative hypothesis. *Biometrika* **103**, 641–652.