Directional Testing and F-tests

Nancy Reid

University of Toronto

joint with **Andrew McCormack**, Nicola Sartori, Sri Theivendran
Don Fraser, Anthony Davison

Directional testing

- data $y \in \mathbb{R}^n$, parameter $\theta \in \mathbb{R}^p, p < n$, model $f(y;\theta)$

- parameter of interest $\psi \in \mathbb{R}^d, d < p$

- reduce dimension $n \downarrow p$: $s = s(y) \in \mathbb{R}^p$ measures θ
- reduce dimension $p \downarrow d$: constrain s to $\mathcal{L}_\psi$, where the nuisance parameter is fixed

- if $d = 1$, distribution on $\mathcal{L}_\psi$ gives p-value function or confidence distribution on $\mathbb{R}$

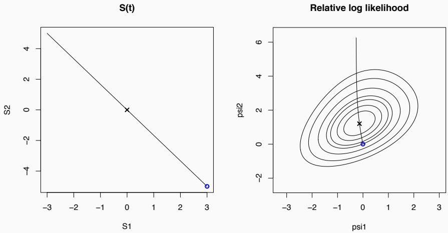
- if $d > 1$, find a line in $\mathcal{L}_\psi$ on which to measure from $H_0 : \psi = \psi_0$

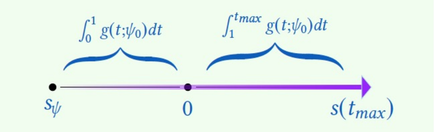
… directional testing

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

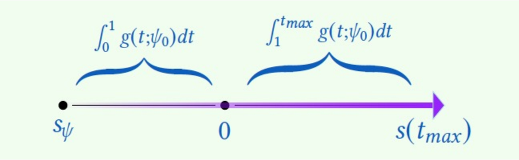
- $d \downarrow 1$: line on $\mathcal{L}_\psi$ between expected, s_ψ and observed s^o
- compute directional p-value on this line

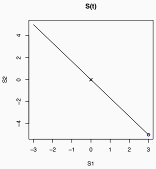
$p = 5, d = 2$





… directional testing





- 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

F-tests

- normal theory linear model $y = X\beta + \epsilon$
- linear constraint $A\beta = 0, \quad A_{d \times p}$

$\psi_0 = 0$

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

McCormack et al., 2018

- 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

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

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

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

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

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

Allen 2018; Hotelling's T^2 test

McCormack et al 2018

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_{ψ_0} and s^0

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

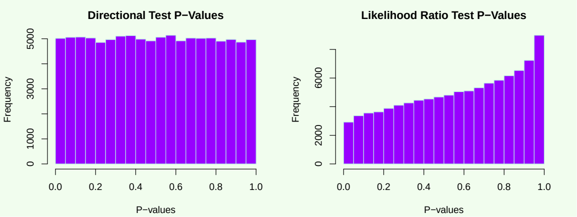
largest value of t

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

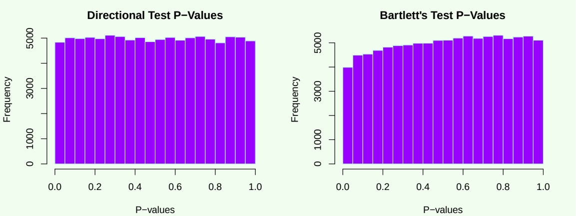


Moderate dimension

- 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

Davison et al. 2014

Sartori et al. 2016

McCormack et al. 2018

Moderate dimensional inference – asymptotics

- $f(y;\theta)$, $y \in \mathbb{R}^n$, $\theta \in \mathbb{R}^p$
- classical: p fixed, $n \to \infty$ — STA3000
- semi-classical: $p_n/n \to 0$, or $p_n^{3/2}/n \to 0$ — Huber, Portnoy; Sartori, Lunardon, ...
- moderate dimension $p_n/n \to \beta$ — Candes, Lei/Bickel/El Karoui, ...
- "high dimension" $p_n/n \to \infty$

- $y_{ij} \sim f(\cdot;\psi,\lambda_i), \quad i = 1,\ldots,q;\ j = 1,\ldots,m;\ \ n = mq$

  Neyman-Scott problems

- $q \to \infty$, m fixed: classical likelihood inference fails

- $q \to \infty$, $m \to \infty$: can recover if $q = o(n^{1/2})$

- using modified likelihood from HOA, can recover if $q = o(n^{3/4})$

  Sartori

- using bias-adjusted score equation , can recover if $q = o(n^{3/4})$

  Lunardon

- HOA elimination of nuisance parameters gives large improvements in asymptotic theory and finite-sample approximations

Moderate dimension

- $\hat{\beta}(\rho) = \arg\min \frac{1}{n}\sum_{i=1}^{n} \rho(y_i - x_i^{\mathrm{T}}\beta)$
- coordinate-wise asymptotic normality

$$max_j d_{TV}\left\{\mathcal{L}\left(\frac{\hat{\beta}_j - \beta_j^*}{\sqrt{\mathrm{var}(\hat{\beta}_j)}}\right), N(0,1)\right\} = o(1)$$

- “For instance for least-squares, standard degrees of freedom adjustments effectively take care of many dimensionality-related problems”
- ?perhaps HOA adjustments for nuisance parameters (= ‘standard degrees of freedom adjustments’) can be effective as using $p/n \to \kappa$ asymptotics? when? why not? Kosmidis

References i

Davison, A.C., Fraser, D.A.S., Reid, N. & Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *J. Am. Statist. Assoc.* **109**, 302 – 14.

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

McCormack, A., Theivendran, S., Reid, N. & Sartori, N. (2018). A New Look at F-tests. arXiv:1808.03946 [math.ST]

Cheah, P.K., Fraser, D.A.S. & Reid, N. (1994). *Biometrika* **81**, 271–278.
Fraser, D.A.S. & Massam, H. (1989). *Statistiche Hefte* **26**, 1–17.
Skovgaard, I.M. (1988). *J. R. Stat. Soc.* B, **50**, 3–32.

References ii

Sartori, N. (2003). *Biometrika* **90**, 533–549.
Lunardon, N. (2018). *Biometrika* **105**, 233–238.

Portnoy, S. (1984). *Ann. Statist.* **12**, 1298–1309.
Portnoy, S. (1985). *Ann. Statist.* **13**, 1403–1417.
Portnoy, S. (1988). *Ann. Statist.* **16**, 356–366.

Lei, L., Bickel, P.J., El Karoui, N. (2018). *Prob. Th. Related Fields*
https://doi.org/10.1007/s00440-017-0824-7

Check for updates

Empirical Article

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Advances in Methods and Practices in Psychological Science
1–20

DOI: 10.1177/2515245917747646
www.psychologicalscience.org/AMPPS
SAGE

Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results

R. Silberzahn[1], E. L. Uhlmann[2], D. P. Martin[3], P. Anselmi[4], F. Aust[5], E. Awtrey[6], Š. Bahník[7], F. Bai[8], C. Bannard[9], E. Bonnier[10], R. Carlsson[11], F. Cheung[12], G. Christensen[13], R. Clay[14], M. A. Craig[15], A. Dalla Rosa[4], L. Dam[16], M. H. Evans[17], I. Flores Cervantes[18], N. Fong[19], M. Gamez-Djokic[20], A. Glenz[21], S. Gordon-McKeon[22], T. J. Heaton[23], K. Hederos[24], M. Heene[25], A. J. Hofelich Mohr[26], F. Högden[5], K. Hui[27], M. Johannesson[10], J. Kalodimos[28], E. Kaszubowski[29], D. M. Kennedy[30], R. Lei[15], T. A. Lindsay[26], S. Liverani[31], C. R. Madan[32], D. Molden[33], E. Molleman[16], R. D. Morey[34], L. B. Mulder[16], B. R. Nijstad[16], N. G. Pope[35], B. Pope[36], J. M. Prenoveau[37], F. Rink[16], E. Robusto[4], H. Roderique[38], A. Sandberg[24], E. Schlüter[39], F. D. Schönbrodt[25], M. F. Sherman[37], S. A. Sommer[40],

"61 analysts used the same data set to address the same research question"

Reproducibility

- “Analytic approaches varied widely across the teams, and the estimated effect sizes ranged from 0.89 to 2.93
- Twenty teams (69%) found a statistically significant positive effect, and 9 teams (31%) did not observe a significant relationship
- Overall, the 29 different analyses used 21 unique combinations of covariates
- Peer ratings of the quality of the analyses also did not account for the variability
- The analytic techniques chosen ranged from simple linear regression to complex multilevel regression and Bayesian approaches
- Teams that employed logistic or Poisson models tended to report estimates that were larger than those of teams that used linear models”

Silberzahn et al., 2018, Association for Psychological Science

http://journals.sagepub.com/doi/pdf/10.1177/2515245917747646