A Directional Look at F-tests

Andrew McCormack[a], Nancy Reid[a], Nicola Sartori[b], Sri-Amirthan Theivendran[a]

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

[b]*Dipartimento di Scienze Statistiche, Università di Padova*

June 23, 2019

Abstract

Directional testing of vector parameters, based on higher order approximations of likelihood theory, can ensure extremely accurate inference, even in high-dimensional settings where standard first order likelihood results can perform poorly. Here we explore examples of directional inference where the calculations can be simplified, and prove that in several classical situations the directional test reproduces exact results based on F-tests. These findings give a new interpretation of some classical results and support the use of directional testing in general models, where exact solutions are typically not available.

Keywords: directional test; exponential family; Hotelling's T^2; linear regression

1 Introduction

In many statistical settings we are interested in hypotheses about vector parameters. Examples include testing sets of dummy variables indexing levels of a factor in a regression model, testing interactions in loglinear models for multi-way contingency tables, and tests in models for multivariate responses, such as testing hypotheses for the mean vector or the covariance matrix of a multivariate normal distribution.

To fix notation, we assume a model for a response y_i with parametric density function $f_i(y_i;\theta)$. We write $f(y;\theta)$ for the joint density for a sample $y = (y_1, \ldots, y_n)$; the maximum likelihood estimator from this sample is $\hat{\theta} = \hat{\theta}(y) = \arg\sup_\theta f(y;\theta)$.

One general approach to testing a given θ value is to construct the quadratic form $q(\theta) = (\hat{\theta} - \theta)^{\mathrm{T}} V^{-1} (\hat{\theta} - \theta)$, where V is an estimate of the covariance matrix of $\hat{\theta}$. Under some regularity conditions ensuring that $\hat{\theta}$ is consistent and asymptotically normally distributed, and that V is a consistent estimator of the covariance matrix of $\hat{\theta}$, $q(\theta)$ has a limiting χ^2_p distribution under the model $f(y;\theta)$, as $n \to \infty$, where p is the dimension of θ. An asymptotically equivalent test for θ is that based on the log-likelihood ratio

$$w(\theta) = 2\{\log f(y;\hat{\theta}) - \log f(y;\theta)\}; \tag{1}$$

this also has a limiting χ^2_p distribution, but the distributions of q and w will differ in finite samples.

Tests about a subvector of θ are similarly constructed. Suppose $\theta = (\psi, \lambda)$ where ψ, of dimension d, is the parameter of interest. The analogous expressions for inference are

$$q(\psi) = (\hat{\psi} - \psi)^{\mathrm{T}} V_1^{-1} (\hat{\psi} - \psi), \tag{2}$$

$$w(\psi) = 2\{\log f(y;\hat{\theta}) - \log f(y;\hat{\theta}_\psi)\}, \tag{3}$$

where V_1 is an estimate of the covariance matrix of $\hat{\psi}$, $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$, and $\hat{\lambda}_\psi$ is the constrained maximum likelihood estimator obtained by maximizing $f(y;\theta)$ over λ with ψ fixed; these have limiting χ^2_d distributions.

In the context of linear regression, Fraser and Massam (1985) proposed tests that measure departure of $\hat{\theta}$ from θ, or $\hat{\psi}$ from ψ, in a particular direction on the parameter space. Skovgaard (1988) derived a saddlepoint-type expansion for directional tests in exponential family models. Davison et al. (2014) and Fraser et al. (2016) showed how to calculate directional p-values via one-dimensional integrals, and illustrated this in a number of models. While the theory was developed in a standard asymptotic scenario, with n increasing and p fixed, empirical results have shown that the method is extremely accurate even in cases where the dimensions p and d are rather large relative to n, when standard first order methods, and some higher order improvements generally fail.

In this paper we show that in some notable examples the directional tests simplify to very well-known omnibus tests. This sheds light on directional testing and gives a new look at some familiar test statistics, as well as providing additional support for the adoption of this approach in general models.

In Section 2 we briefly review the directional testing approach. In Sections 3 and 4 we show that directional tests coincide with exact well-known solutions in some examples where inference is focused on scale parameters and location parameters respectively. Technical details are given in the Supplementary Material.

2 Directional testing

The theory and methods for directional tests are given in Davison et al. (2014) and Fraser et al. (2016), and provided for completeness in the Supplementary Material. Here we introduce the necessary notation and key concepts.

Suppose we have an exponential family model with sufficient statistic $u = u(y)$ and canonical parameter φ, with density

$$f(y; \varphi) = \exp\{\varphi^{\mathrm{T}} u - \kappa(\varphi)\} h(y), \tag{4}$$

and we are interested in the hypothesis $H_\psi : \psi(\varphi) = \psi$. We write $\hat{\varphi}_\psi$ for the constrained maximum likelihood estimate of φ under H_ψ. The directional test of H_ψ restricts attention to the line in the sample space joining u_ψ with u^0, where u_ψ is the value of the sufficient statistic for which $\hat{\varphi}_\psi$ is the maximum likelihood estimate, and u^0 is the observed value of the sufficient statistic, for which $\hat{\varphi}$ is the maximum likelihood estimate. The directional p-value is the tail area probability for the length $||u_\psi - u^0||$, conditional on the direction $(u_\psi - u^0)/||u_\psi - u^0||$. It is sometimes convenient to define $s = u - u^0$, so that $s^0 = 0$ and $s_\psi = u_\psi - u^0$.

The p-value for this directional approach is defined, as in Davison et al. (2014, Section 3.2), by a ratio of two integrals,

$$p(\psi) = \frac{\int_1^{t_{\max}} t^{d-1} h(t;\psi)\,dt}{\int_0^{t_{\max}} t^{d-1} h(t;\psi)\,dt}, \tag{5}$$

where d is the dimension of ψ, and t indexes points along the line, with $t = 0$ corresponding to the value u_ψ, and $t = 1$ corresponding to the observed value u^0. The upper bound, $t_{\max}$, of these integrals is the largest value of t where the corresponding sufficient statistic on the line between u_ψ and u^0 still lies in the support of its distribution. The two one-dimensional integrals in (5) can be easily and accurately computed numerically. The factor t^{d-1} comes from the Jacobian computed in transforming the joint density to the conditional density of the length, given the direction, which is essentially a transformation to spherical coordinates.

The ingredients needed for the calculation of $h(t;\psi)$ in (5) are the log-likelihood function, the maximum likelihood estimate, the observed Fisher information, as functions of t, as well as the observed value of the constrained maximum likelihood estimate. An expression for $h(t;\psi)$ when the hypothesis is linear in φ is given in Davison et al. (2014, Eq.(8)), and when the hypothesis is nonlinear in Fraser et al. (2016, Eq.(4)). For completeness the general expression for $h(t;\psi)$ is presented here and described in detail in the Supplementary Material:

$$h(t;\psi) \propto \exp\left[\ell\{\hat{\varphi}_\psi; s(t)\} - \ell\{\hat{\varphi}; s(t)\}\right] |\hat{J}_{\varphi\varphi}|^{-1/2} |\tilde{J}_{(\lambda\lambda)}|^{1/2}, \tag{6}$$

where $\ell(\varphi; s) = \varphi^{\mathrm{T}}(\theta)s + \ell^0\{\theta(\varphi)\}$, $\ell^0(\theta) = \ell(\theta; y^0)$ is the observed value of the log-likelihood function from (4), and the score variable s is constrained to the line $s(t), t > 0$. When the hypothesis is linear in the canonical parameter, the factor $|\tilde{J}_{(\lambda\lambda)}|$ does not depend on t and therefore is not needed in (5). It may also be indpendent of t in other cases, such as that in §4.2.

If the underlying model is not an exponential family model, an initial approximation to that model, called the tangent exponential model, is constructed first, and the arguments apply again within this model (Fraser et al., 2016, Ex. 4.3).

We show in this paper that the ratio of integrals in (5) can be calculated explicitly in some simple models, which helps to explain the accuracy of the approximation as evidenced in Davison et al. (2014) and Fraser et al. (2016). We present the calculation for simple examples first in Section 3, and then use similar techniques in Section 4 to provide a new view of classical tests in multivariate normal models.

3 Inference in scale models

3.1 Scalar parameter of interest

In a one-dimensional sub-model, the directional test gives two-sided p-values as it reduces to the probability of the right (or left) tail, conditional on being in that tail. In the two examples in this section we demonstrate this, as the calculations can be carried out analytically, and the arguments motivate exact calculations in the regression setting of Section 4.

3.2 Comparison of exponential rates

We consider first the example of Davison et al. (2014, §5.2) in the case of just two groups ($g = 2$). Suppose that y_{ij} are independent random variables following an exponential distribution with rates θ_i for $i = 1, 2$ and $j = 1, \ldots, n_i$, and the null hypothesis is $H_\psi : \theta_1/\theta_2 = \psi$ for some $\psi \in (0, \infty)$. An exact test is available, since $W_\psi = \psi \bar{y}_1/\bar{y}_2$ follows an $F(2n_1, 2n_2)$ distribution under H_ψ.

The log-likelihood of the full model is

$$\ell(\theta; y) = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (\log \theta_i - \theta_i y_{ij}). \tag{7}$$

The canonical parameter is $\varphi(\theta) = (-\theta_1, -\theta_2)$, the sufficient statistic is $u = (u_1, u_2) = (\Sigma_j y_{1j}, \Sigma_j y_{2j})$, the maximum likelihood estimate of θ is $\hat{\theta} = (n_1/u_1, n_2/u_2)$, and the constrained maximum likelihood estimate is

$$\hat{\theta}_\psi = \left(\frac{n\,\psi}{\psi u_1 + u_2}, \frac{n}{\psi u_1, +u_2} \right),$$

giving

$$u_\psi = (u_{1\psi}, u_{2\psi}) = \frac{1}{n} \left(u_1 n_1 + \frac{u_2 n_1}{\psi}, \psi u_1 n_2 + u_2 n_2 \right).$$

By a standard property of exponential families u_ψ is the expected value of u under H_ψ. With

$s_\psi = u_\psi - u^0$, the line $s(t)$, $t \geq 0$, to be $(1-t)s_\psi$. At $t = 1$, $s(t) = s^0 = 0$. The saddlepoint approximation to the density of s on the ray $s(t)$ is

$$t^{d-1}h(t;\psi) = (1 - t/a_1)^{n_1-1}(1 - t/a_2)^{n_2-1}, \tag{8}$$

with $a_i = u_{i\psi}/(u_{i\psi} - u_i)$ for $i = 1, 2$. The support of the density is $s(t) > -u^0$, or equivalently $\hat{\theta}\{s(t)\} \geq 0$, so $t_{\max} = \max(a_1, a_2)$. Although the hypothesis is not linear in the canonical parameter, the nuisance parameter adjustment term $|\tilde{J}_{(\lambda\lambda)}|$ is independent of t because $\varphi(\theta)$ is linear in the nuisance parameter θ_2.

If $t_{\max} = a_2$, $\psi\bar{y}_1 \geq \bar{y}_2$, and the directional p-value is

$$p(\psi) = \frac{\int_1^{a_2}(1 - t/a_1)^{n_1-1}(1 - t/a_2)^{n_2-1}\,dt}{\int_0^{a_2}(1 - t/a_1)^{n_1-1}(1 - t/a_2)^{n_2-1}\,dt}. \tag{9}$$

Let p_{num} and p_{denom} be the numerator and denominator of (9) respectively. Taking $x = (1 - t/a_1)/(1 - t/a_2)$ gives

$$p_{\text{num}} = c\int_{\psi\bar{y}_1/\bar{y}_2}^{\infty} x^{n_1-1}\left(1 + \frac{n_1}{n_2}x\right)^{-(n_1+n_2)} dx, \quad p_{\text{den}} = c\int_1^{\infty} x^{n_1-1}\left(1 + \frac{n_1}{n_2}x\right)^{-(n_1+n_2)} dx.$$

The constant c is the same in p_{num} and p_{den} because the same change of variables is used in both integrals. Recognizing these integrands as the density of a $F(2n_1, 2n_2)$ random variable with cumulative distribution function $G_{2n_1,2n_2}(x)$, (9) becomes

$$p(\psi) = \frac{1 - G_{2n_1,2n_2}(W_\psi)}{1 - G_{2n_1,2n_2}(1)},$$

showing how the directional p-value is related to that based on the F-test of $W_\psi = \psi\bar{y}_1/\bar{y}_2$. Similarly, when $\psi\bar{y}_1 < \bar{y}_2$ the directional p-value is $p(\psi) = G_{2n_1,2n_2}(W_\psi)/G_{2n_1,2n_2}(1)$, so that

$$p(\psi) = \mathrm{I}(W_\psi \geq 1)\frac{1 - G_{2n_1,2n_2}(W_\psi)}{1 - G_{2n_1,2n_2}(1)} + \mathrm{I}(W_\psi < 1)\frac{G_{2n_1,2n_2}(W_\psi)}{G_{2n_1,2n_2}(1)}. \tag{10}$$

The terms that are multiplied by the indicator functions in (10) are uniformly distributed over $[0, 1]$ when conditioned on W_ψ being in the appropriate region. It follows directly that the directional

p-value is also uniformly distributed on $[0, 1]$.

A two-tailed F-test has p-value $2 \min\{G_{2n_1,2n_2}(W_\psi), 1 - G_{2n_1,2n_2}(W_\psi)\}$, which can be expressed as

$$\mathrm{I}(W_\psi \geq \mathcal{G}_{1/2}) \frac{1 - G_{2n_1,2n_2}(W_\psi)}{1 - G_{2n_1,2n_2}(\mathcal{G}_{1/2})} + \mathrm{I}(W_\psi < \mathcal{G}_{1/2}) \frac{G_{2n_1,2n_2}(W_\psi)}{G_{2n_1,2n_2}(\mathcal{G}_{1/2})}, \tag{11}$$

where $\mathcal{G}_{1/2}$ is the median. The tail region here is slightly different from that in (10), although for practical purposes the difference is slight. Even in a highly asymmetric setting where $n_1 = 5$ and $n_2 = 10,000$, $G_{10,20000}(1) = 0.559$.

3.3 Comparison of normal variances

Suppose that $y_{ij} \sim N(\mu_i, \sigma_i^2)$ are independent random variables for $i = 1, 2$ and $j = 1 \ldots, n_i$, and the null hypothesis is $H_\psi : \sigma_1^2/\sigma_2^2 = \psi$. Under the hypothesis H_ψ, $W_\psi = \psi s_2^2/s_1^2$ follows an $F(\nu_2, \nu_1)$ distribution, where $s_i^2 = n_i(\nu_i)^{-1} v_i^2$ is the unbiased sample variance estimate for group i, with $v_i^2 = n_i^{-1} \Sigma_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ and $\nu_i = n_i - 1$.

Following a derivation like that in Davison et al. (2014, §5.1), the directional p-value is

$$p(\psi) = \frac{\int_1^{1/a_1} (1 - ta_1)^{(n_1-3)/2} (1 - ta_2)^{(n_2-3)/2}\, dt}{\int_0^{1/a_1} (1 - ta_1)^{(n_1-3)/2} (1 - ta_2)^{(n_2-3)/2}\, dt}, \tag{12}$$

where $a_i = (\hat{\sigma}_{i\psi}^2 - v_i^2)/\hat{\sigma}_{i\psi}^2$, $i = 1, 2$, and $\hat{\sigma}_{i\psi}^2$ is the constrained maximum likelihood estimator for σ_i^2. The same change of variable as in (9) gives

$$p(\psi) = \{1 - G_{\nu_2,\nu_1}(W_\psi)\}/[1 - G_{\nu_2,\nu_1}\{n_2\nu_1/(n_1\nu_2)\}],$$

when $v_1^2 \leq \psi v_2^2$, or equivalently, $\nu_1 n_2/(\nu_2 n_1) \leq \psi s_2^2/s_1^2$. Combining this with the case $v_1^2 > \psi v_2^2$ gives

$$p(\psi) = I\left\{W_\psi \geq \frac{n_2\nu_1}{n_1\nu_2}\right\} \frac{1 - G_{\nu_2,\nu_1}(W_\psi)}{1 - G_{\nu_2,\nu_1}\{n_2\nu_1/(n_1\nu_2)\}} + I\left\{W_\psi < \frac{n_2\nu_1}{n_1\nu_2}\right\} \frac{G_{\nu_2,\nu_1}(W_\psi)}{G_{\nu_2,\nu_1}\{n_2\nu_1/(n_1\nu_2)\}}. \tag{13}$$

As noted in Davison et al. (2014, Ex. 5.1), this expression simplifies to the two-sided F-test if

$n_1 = n_2$ since $v_1^2/v_2^2 = s_1^2/s_2^2$ in this case. When $n_1 \neq n_2$ (13) is the p-value for the exact F-test based on the tail probabilities of ψ times the ratio of the biased maximum likelihood estimators v_i^2. With more than two groups there is no exact test for comparison, but simulations in Davison et al. (2014) show that the directional test is very accurate even with a very large number of groups, and hence large numbers of nuisance parameters and a large dimension of ψ, whereas the usual likelihood ratio test (3) breaks down, as does the modified likelihood ratio version proposed in **?**.

4 Inference for location parameters

4.1 Linear regression

We now consider testing the the null hypothesis $H_\psi : A\beta = \psi$ in a linear regression model, $y_i = x_i^{\mathrm{T}}\beta + \epsilon_i$, $i = 1, ..., n$, where both x_i and β are vectors of length p, and ϵ_i are independently distributed as $N(0, \sigma^2)$ with an unknown variance. We assume A is a given $d \times p$ matrix with maximal rank, so the dimension of the parameter of interest is d and that of the implicit nuisance parameter is $p + 1 - d$. This null hypothesis encompasses many other hypotheses of interest, such as testing for the equality of group means when the group variances are equal. If X is taken to be the matrix with rows x_i^{T}, the log-likelihood function for $\theta^{\mathrm{T}} = (\beta^{\mathrm{T}}, \sigma^2)$ is

$$\ell(\theta; y) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\big(y^{\mathrm{T}}y - 2y^{\mathrm{T}}X\beta + \beta^{\mathrm{T}}X^{\mathrm{T}}X\beta\big).$$

This can be re-expressed in exponential family form with canonical parameter $\varphi(\theta)^{\mathrm{T}} = \sigma^{-2}(\beta^{\mathrm{T}}, -1/2)$ and sufficient statistic $u^{\mathrm{T}} = (y^{\mathrm{T}}X, y^{\mathrm{T}}y) = (u_1^{\mathrm{T}}, u_2)$. The unconstrained and constrained maximum likelihood estimates for β are

$$\widehat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}y, \quad \widehat{\beta}_\psi = \widehat{\beta} - (X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\big\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\big\}^{-1}(A\widehat{\beta} - \psi),$$

and those of σ^2 are $\hat{\sigma}^2 = n^{-1}\Sigma(y_i - x_i^{\mathrm{T}}\widehat{\beta})^2$ and $\hat{\sigma}_\psi^2 = n^{-1}\Sigma(y_i - x_i^{\mathrm{T}}\widehat{\beta}_\psi)^2$. The value of s that has $\hat{\theta}_\psi$ as the maximum likelihood estimate of θ is

$$s_\psi^{\mathrm{T}} = (\widehat{\beta}_\psi^{\mathrm{T}} X^{\mathrm{T}} X - y^{\mathrm{T}} X, n\hat{\sigma}_\psi^2 + \widehat{\beta}_\psi^{\mathrm{T}} X^{\mathrm{T}} X \widehat{\beta}_\psi - y^{\mathrm{T}} y).$$

On the line $s(t)$, $t \geq 0$, the log-likelihood function for $\varphi(\theta)$ is

$$\ell\{\varphi(\theta); s(t)\} = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\big\{u_2(t) - 2u_1(t)\beta + \beta^{\mathrm{T}} X^{\mathrm{T}} X \beta\big\}, \tag{14}$$

with $\{u_1(t), u_2(t)\} = u(t) = u^0 + s(t)$. From (14) we obtain the maximum likelihood estimates as functions of t: $\widehat{\beta}(t) = (X^{\mathrm{T}}X)^{-1}u_1(t)$ and $\hat{\sigma}^2(t) = n^{-1}\big\{u_2(t) - 2u_1(t)\widehat{\beta}(t) + \widehat{\beta}^{\mathrm{T}}(t)X^{\mathrm{T}}X\widehat{\beta}(t)\big\}$.

Evaluating (6) gives

$$h(t;\psi) = \{\hat{\sigma}^2(t)\}^{(n-p-2)/2} = \big\{\hat{\sigma}_\psi^2 - \frac{t^2}{n}(y - X\widehat{\beta}_\psi)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y - X^{\mathrm{T}}\widehat{\beta}_\psi)\big\}^{(n-p-2)/2}.$$

The density along the line $s(t)$, passing through s_ψ and s^0 is then

$$t^{d-1}h(t;\psi) = t^{d-1}\big\{\hat{\sigma}_\psi^2 - \frac{t^2}{n}(y - X\widehat{\beta}_\psi)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y - X^{\mathrm{T}}\widehat{\beta}_\psi)\big\}^{(n-p-2)/2} = t^{d-1}(a - bt^2)^{(n-p-2)/2}. \tag{15}$$

As the hypothesis can be expressed as a linear function of the canonical parameter, there is no need for the nuisance parameter adjustment term $|\tilde{J}_{(\lambda\lambda)}|$.

To compute the directional p-value (5) we need $t_{\max}$, which here is the largest value of t for which $\hat{\sigma}^2(t) \geq 0$,

$$t_{\max} = \Big[n\hat{\sigma}_\psi^2 / \{(y - X\hat{\beta}_\psi)^{\mathrm{T}} X(X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}(y - X\hat{\beta}_\psi)\}\Big]^{1/2} = (a/b)^{-1/2}.$$

With the the change of variable $x = \frac{n-p}{d}\left(\frac{a}{bt^2} - 1\right)$, a computation detailed in the Supplementary Material verifies that (5) simplifies to

$$p(\psi) = 1 - G_{d,n-p}\left[\frac{(A\hat{\beta} - \psi)^{\mathrm{T}}\{A(X^{\mathrm{T}}X)^{-1}A^{\mathrm{T}}\}^{-1}(A\hat{\beta} - \psi)/d}{(y - X\hat{\beta})^{\mathrm{T}}(y - X\hat{\beta})/(n-p)}\right], \tag{16}$$

which is the p-value based on the usual $F_{d,n-p}$ distribution (Rencher and Schaalje, 2008, Ch. 8).

This result gives a new interpretation of the F-statistic: it measures the magnitude of the sufficient statistic for $A\beta = \psi$, conditional on the direction indicated by the observed data. As the normal distribution is spherically symmetric, the magnitude is distributed independently of the direction.

4.2 Hotelling's T^2

The result in §4.1 suggests comparing the directional test for a multivariate normal mean with Hotelling's T^2 statistic. Suppose y_i, $i = 1, ..., n$ are independent observations from the multivariate normal distribution, $N_d(\mu, \Lambda^{-1})$, with unknown covariance matrix Λ^{-1}. The full parameter is $\theta = \{\mu, \text{vec}(\Lambda)\}$, where vec gives a vectorization of the columns of a matrix. Strictly speaking we need only $d(d+1)/2$ entries of Λ; the correction for dimension is made when determining the Hessian. We consider the hypothesis $H_\psi : \mu = \psi$. The distribution for $y = (y_1, \ldots, y_n)$ is an exponential family model, with canonical parameter $\varphi^{\mathrm{T}}(\theta) = \{\mu^{\mathrm{T}}\Lambda, \text{vec}^{\mathrm{T}}(\Lambda)\}$, and sufficient statistic $u^{\mathrm{T}} = \{n\bar{y}^{\mathrm{T}}, \text{vec}^{\mathrm{T}}(\Sigma_i y_i y_i^{\mathrm{T}})\}$. The unconstrained maximum likelihood estimates are $\hat{\mu} = \bar{y}$ and $\hat{\Lambda} = n\{\Sigma_i(y_i - \bar{y})(y_i - \bar{y})^{\mathrm{T}}\}^{-1}$ while the constrained maximum likelihood estimate for Λ is $\hat{\Lambda}_\psi = n\{\Sigma_i(y_i - \psi)(y_i - \psi)^{\mathrm{T}}\}^{-1}$. Under H_ψ, the expected value of the centered sufficient statistic s is

$$s_\psi^{\mathrm{T}} = \left[n\psi - n\bar{y}, \text{vec}^{\mathrm{T}}\left\{\frac{n}{2}(\psi\bar{y}^{\mathrm{T}} + \bar{y}\psi^{\mathrm{T}} - 2\psi\psi^{\mathrm{T}})\right\}\right].$$

The maximum likelihood estimators on the ray $s(t)$ are

$$\begin{aligned}
\hat{\varphi}^{\mathrm{T}}(t) &= \left[\{\psi + t(\bar{y} - \psi)\}^{\mathrm{T}}\hat{\Lambda}(t), \text{vec}^{\mathrm{T}}\{\hat{\Lambda}(t)\}\right], \\
\hat{\Lambda}^{-1}(t) &= \left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \psi)(y_i - \psi)^{\mathrm{T}}\right\} - t^2(\bar{y} - \psi)(\bar{y} - \psi)^{\mathrm{T}} = \hat{\Lambda}_\psi^{-1} - t^2 vv^{\mathrm{T}},
\end{aligned} \tag{17}$$

where $v = \bar{y} - \psi$. These maximum likelihood estimators are valid if $\hat{\Lambda}^{-1}(t)$ is positive definite, and the largest t such that this is the case, $t_{\max}$, is found by solving for t in the equation $|\hat{\Lambda}^{-1}(t)| = 0$, where $|\hat{\Lambda}^{-1}(t)| = \det(\hat{\Lambda}_\psi^{-1})(1 - t^2 v^{\mathrm{T}}\hat{\Lambda}_\psi v)$. The components of (6) needed to compute the directional

p-value are

$$\exp\left[\ell\{\hat\varphi(0); s(t)\} - \ell\{\hat\varphi(t); s(t)\}\right] = |\hat\Lambda^{-1}(t)|^{n/2},$$
$$|J_{\varphi\varphi}\{\hat\varphi(t); s(t)\}|^{-1/2} = |\hat\Lambda^{-1}(t)|^{-(q+1)/2}.$$

As the canonical parameter is linear in Λ, the nuisance parameter information term does not depend on t and can be ignored. The directional p-value is

$$\frac{\int_1^{(v^{\mathrm T}\hat\Lambda_\psi v)^{-1/2}} t^{p-1}(1 - t^2 v^{\mathrm T}\hat\Lambda_\psi v)^{(n-p-2)/2}\, dt}{\int_0^{(v^{\mathrm T}\hat\Lambda_\psi v)^{-1/2}} t^{p-1}(1 - t^2 v^{\mathrm T}\hat\Lambda_\psi v)^{(n-p-2)/2}\, dt}. \tag{18}$$

There is a striking similarity between this and the directional p-value (9), and the same change of variables can be applied here. By the Sherman–Morrison formula

$$\frac{1 - v^{\mathrm T}\hat\Lambda_\psi v}{v^{\mathrm T}\hat\Lambda_\psi v} = (\bar y - \psi)^{\mathrm T}\left\{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar y)(y_i - \bar y)^{\mathrm T}\right\}^{-1}(\bar y - \psi) = v^{\mathrm T}\hat\Lambda v. \tag{19}$$

After the appropriate change of variables, (25) appears in the bound of the integrals in the directional p-value. After further simplification we have

$$p(\psi) = 1 - G\left(\frac{n-p}{p} v^{\mathrm T}\hat\Lambda v\right), \tag{20}$$

where G is the cumulative distribution function of an $F(p, n-p)$ random variable. As Hotelling's T^2 statistic is equal to $(n-1)v^{\mathrm T}\hat\Lambda v$ and $(n-p)/\{p(n-1)\}T^2 \sim F(p, n-p)$, this shows that the directional test is identical to Hotelling's T^2 test.

5 Discussion

Directional tests for a vector parameter create a one-dimensional sub-model by restricting attention to the line on the parameter space that is dictated by the observed data. The computation of these tests has recently been simplified by relying on saddlepoint approximations to the distributions, rather than computing them exactly.

This work concentrates on models for which saddlepoint approximations are exact or nearly exact, and shows that conventional F-tests emerge from the directional approach. This helps to explain the accuracy of the tests demonstrated in Davison et al. (2014) and Fraser et al. (2016).

All the directional p-value integrands appearing in this work share a common structure. Each integrand of the directional p-values has the form $t^{d-1}\hat{\sigma}^2(t)^{\alpha/2}$, where $\hat{\sigma}^2(t)$ is a measure of variability under H_ψ, and α depends on n, d and p. Small directional p-values correspond to observed data that have a relatively high weighted variability estimate under H_ψ.

The hypotheses considered in Section 4 constrain the mean vector to a linear subspace of the parameter space, and are also invariant under affine transformations of the parameter. The F-tests in Section 4 are derived as most powerful invariant tests in Lehmann and Romano (2005, Ch.7), and shown there to effectively test a scalar parameter, the noncentrality parameter of the related F distribution.

We are preparing an R package (R Core Team, 2018) to construct directional tests in generalized linear models, including gamma, Poisson, and logistic regression. With discrete probability functions, such as the binomial and Poisson, saddlepoint methods are not exact because they are implicitly continuous, but simulation results in Davison et al. (2014, §4.2) and in work in progress indicates that the directional p-values continue to be very accurate.

It is also straightforward to develop a directional test for normal theory non-linear regression models of the form $y_i \sim N\{\eta_i(\beta), \sigma^2\}, i = 1, \ldots, n$, but we have been unable to verify that these are the same as the conventional F-test based on the tangent model approximation to the mean surface.

By their nature, we might expect directional tests to have low power in regions of the parameter space that are not suggested by the data. This point was raised in work as yet unpublished by Jens Ledet Jensen. However for at least some settings the work here shows that the tests are the same as conventional F-tests for multivariate hypotheses, so share their power properties.

Acknowledgements

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Italian Ministry of Education under the PRIN 2015 grant 2015EASZF_003 and the University of Padova (PRAT 2015 CPDA153257). We would like to thank D.A.S. Fraser and I. Kosmidis for helpful discussion, A. Ruffato for help in the derivation of the result in Section 4.1, and A.C. Davison for helpful comments on an earlier version.

Supplementary Material

Additional information about the saddlepoint approximation and directional tests of §2, as well as detailed calculations to support the analytical results in §3 and §4, are provided in the Supplementary Material available online.

References

Davison, A. C., Fraser, D. A. S., Reid, N. & Sartori, N. (2014), "Accurate directional inference for vector parameters in linear exponential families," *Journal of the American Statistical Association*, 109, 302–314.

Fraser, D. A. S. & Massam, H. (1985), "Conical tests: observed levels of significance and confidence regions," *Statistiche Hefte*, 26, 1–17.

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

Lehmann, E. L. & Romano, J. P. (2005), *Testing Statistical Hypotheses*, New York: Springer.

R Core Team (2018), "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rencher, A. C. & Schaalje, G. B. (2008), *Linear Models in Statistics,* 2nd ed. New Jersey: Wiley.

Skovgaard, I. M. (1988), "Saddlepoint expansions for directional test probabilities," *J. R. Statist. Soc.* B, 50, 269–280.

Skovgaard, I. M. (2001), "Likelihood asymptotics," *Scand. J. Statist.*, 28, 3–32.