

Conditional limit laws for goodness-of-fit tests

RICHARD A. LOCKHART

*Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada.
E-mail: lockhart@sfu.ca; url: www.stat.sfu.ca/~lockhart*

We study the conditional distribution of goodness of fit statistics of the Cramér–von Mises type given the complete sufficient statistics in testing for exponential family models. We show that this distribution is close, in large samples, to that given by parametric bootstrapping, namely, the unconditional distribution of the statistic under the value of the parameter given by the maximum likelihood estimate. As part of the proof, we give uniform Edgeworth expansions of Rao–Blackwell estimates in these models.

Keywords: empirical distribution function; goodness-of-fit; local central limit theorem; parametric bootstrap; Rao–Blackwell

1. Introduction

In this paper, we compare conditional and unconditional goodness-of-fit tests and give conditions under which the two give essentially identical results in large samples. Our results apply in testing fit for exponential family models for independent and identically distributed (i.i.d.) data, X_1, \dots, X_n . Our interest is to test the null hypothesis that the distribution of the individual X_i belongs to a natural exponential family with density, relative to some σ -finite measure, $\mu(dx)$, on some sample space Ω , of the form

$$f(x; \theta) \equiv c(x) \exp\{\theta' T(x) - \kappa(\theta)\} \quad (1)$$

with natural parameter space $\Theta \subset \mathbb{R}^k$; we assume that Θ has non-empty interior which we denote $\text{int}(\Theta)$. In (1), T takes values in \mathbb{R}^k and superscript $'$ denotes transposition. A complete and sufficient statistic for the parameter θ is then

$$\mathbf{T}_n \equiv \mathbf{T}_n(X_1, \dots, X_n) = \sum_{i=1}^n T(X_i).$$

To apply classical hypothesis testing ideas, we regard this model as a null hypothesis. We consider the omnibus alternative hypothesis that the sample is drawn from a distribution which is not in the parametric model. One common approach to this hypothesis testing problem is to define some statistic $S(X_1, \dots, X_n; \theta)$ which measures in some way departure of the sample from what is expected if θ is the true value. Since θ is unknown, it is replaced in this measure by $\hat{\theta}_n$, the maximum likelihood estimate of the parameter vector, leading to the statistic $S_n \equiv S(X_1, \dots, X_n, \hat{\theta}_n)$.

Common examples include empirical distribution function statistics such as Cramér–von Mises, Kolmogorov–Smirnov, Anderson–Darling and many chi-squared statistics. The usual situation is that the test statistic has a distribution which depends, even in large samples, on the unknown parameter value (exceptions arise in the normal and other families which have only location and/or scale parameters). Thus, to implement the tests in practice it is necessary to specify how to compute critical points for the tests or how to compute appropriate P -values corresponding to the test statistics. A method long in use is to derive large sample theory for the statistic S_n , establishing the convergence in distribution of S_n to some limiting distribution which depends on the true value of θ . If $C_\alpha(\theta)$ is the upper α critical point of this limiting distribution and C_α depends continuously on θ , then the test which rejects if $S_n > C_\alpha(\hat{\theta}_n)$ has asymptotic level α . See Lockhart and Stephens [9] for a discussion of this method in testing fit for the von Mises distribution for directional data; this testing problem is discussed below in more detail.

A more modern method which achieves the same asymptotic behaviour is the parametric bootstrap. Let $H_n(\cdot; \theta)$ denote the cumulative distribution function of S_n when the true parameter value is θ . Then

$$P_b = 1 - H_n(S_n; \hat{\theta}_n)$$

is the parametric bootstrap P -value. This P -value is usually computed approximately by generating some number, B , of bootstrap samples drawn from the density $f(\cdot, \hat{\theta}_n)$, computing the statistic S_n for each of these B samples and then counting the fraction of these bootstrap statistic values which exceed the value of S_n for the data set at hand.

These two methods for goodness-of-fit testing both depend on asymptotic theory to justify their performance. They do not have, except in the location-scale situation mentioned, exact level α and thus no exact finite sample optimality properties. Conditional tests, which we discuss next, offer at least the potential for such optimality. (See Remark 9 in the Discussion section for some comments.)

One standard approach (discussed in detail in [5]) to optimality theory is to search for powerful unbiased level α tests: tests whose power never falls below α on the alternative. Such tests will generally have Neyman structure; that is, their level will be α everywhere on the boundary of the null hypothesis. For the omnibus alternative, this boundary is generally the entire model.

Now suppose \mathbf{T}_n is a complete sufficient statistic for this model. Then the requirement that the level of the test be α everywhere in the parametric model and completeness guarantee that the test must have conditional level α . That is, an unbiased level α test must have the property that the conditional probability of rejection given \mathbf{T}_n is identically α . This is precisely the argument used in Lehmann and Romano [5] to show that Student's t test is uniformly most powerful unbiased.

By a conditional test, then, we mean a test whose level, given the sufficient statistic \mathbf{T}_n , is identically α . Two recent papers on goodness-of-fit, Lockhart, O'Reilly and Stephens [7,8], have compared such conditional tests with parametric bootstrap tests. They implemented their conditional tests as follows. For a test statistic S_n , let $G_n(\cdot|\cdot)$ denote the conditional distribution function, when the true distribution of the data comes from the exponential family, of S_n given \mathbf{T}_n . This function G_n does not depend on θ . If this conditional distribution function is continuous then

$$P_c = 1 - G_n(S_n|\mathbf{T}_n)$$

has a uniform distribution under H_0 ; it is therefore an exact P -value. These P -values are often computed by Monte Carlo or Markov Chain Monte Carlo; see Lockhart, O'Reilly and Stephens [7,8] for examples and references.

In Lockhart, O'Reilly and Stephens [7], for instance, the authors considered an i.i.d. sample from the von Mises distribution. Observations X_i are points on the unit circle; see Section 4.2 below for details of the density. The complete sufficient statistic is $\mathbf{T}_n = \sum X_i$ and the authors use Watson's U^2 statistic for S_n . They use Markov Chain Monte Carlo methods to generate a sequence of samples from the conditional distribution of X_1, \dots, X_n given \mathbf{T}_n ; all the generated samples have the same value of \mathbf{T}_n . The authors evaluate P_c by computing U^2 for each data set and estimating P_c by the fraction of samples giving larger values of U^2 than the original data sample.

These authors also compute the parametric bootstrap value, P_b , for the same statistic by generating i.i.d. samples from the von Mises distribution using, for the parameter value, the estimate of the parameter derived from the original data. Of course the values of the sufficient statistic \mathbf{T}_n vary from one bootstrap sample to another. Again U^2 is computed for each bootstrap sample and a P value is computed as the fraction of bootstrap U^2 values which are larger than the observed value of U^2 .

Very high correlations between the P -values computed using these two methods were observed in Lockhart, O'Reilly and Stephens [7]. For example, they considered a test that a sample of size 34 comes from a von Mises distribution. Using Watson's U^2 and generating samples from the null hypothesis they observed a correlation of 0.997 between the two P -values. For a sample of 55 observations, the correlation observed was 0.9997.

Here we show that for statistics S_n of the Cramér–von Mises type these two methods must give similar P -values because, when the null hypothesis is true,

$$\sup_s \{|G_n(s|\mathbf{T}_n) - H_n(s; \hat{\theta}_n)|\} \rightarrow 0$$

in probability, at least when the model being tested is an exponential family. In fact, the convergence is almost sure for samples from any distribution for which $\hat{\theta}_n/n$ converges almost surely to an interior point of the parameter space. For statistics S_n which are sums of the form $\sum_i u_n(X_i, \hat{\theta})$ this result is established by Holst [4]. Our results extend his to statistics which we now describe.

When Ω is the real line, many goodness-of-fit tests are based on statistics which are functionals of the estimated empirical process

$$W_n(s) = \sqrt{n}\{F_n(x) - F(x, \hat{\theta}_n)\},$$

where we now use $F(x, \theta)$ for the cumulative distribution function, s is related to x by $s = F(x, \hat{\theta}_n)$ and F_n is the usual empirical distribution function:

$$F_n(x) = n^{-1} \sum_{i=1}^n 1(X_i \leq x).$$

Common choices for statistics include:

- Cramér–von Mises type:

$$S_n = \int_0^1 \psi^2(s) W_n^2(s) ds; \quad (2)$$

- Watson type:

$$\int_0^1 \left\{ W_n(s) - \int_0^1 \psi(u) W_n(u) du \right\}^2 \psi^2(s) ds;$$

- Kolmogorov–Smirnov type:

$$\sup_{0 < s < 1} |\psi(s) W_n(s)|.$$

In each case, ψ is some weight function defined on $(0, 1)$.

The large sample analysis of the unconditional distribution of such statistics comes from the well known weak convergence, in $D[0, 1]$, of the process W_n to a Gaussian process, W , which we now describe. Let $\mathcal{I}(\theta)$ be the Fisher information matrix and define the column vector

$$\xi(s, \theta) = \frac{\partial F(x; \theta)}{\partial \theta},$$

where x is defined as a function of s by $F(x, \theta) \equiv s$. Then the limit process W has mean 0 and covariance function

$$\rho_\theta(s, t) = \min\{s, t\} - st - \xi(s, \theta)' \{\mathcal{I}(\theta)\}^{-1} \xi(t, \theta).$$

The statistics indicated above are all continuous functionals of W_n (under mild conditions on the weight functions involved) and as such converge in distribution to the same functional applied to the limit process W . See Stephens [14] for a detailed discussion of the resulting tests and Shorack and Wellner [12] for mathematical details.

The weak convergence result can be proved in two steps: prove convergence in distribution of the finite dimensional distributions of W_n and then prove tightness of the sequence of processes in $D[0, 1]$. We believe a similar result holds, in exponential families, conditional on the sufficient statistic. Results in Holst [4] can be used to establish convergence of the conditional finite dimensional distributions but we are unable to extend the calculations to prove conditional tightness. Instead we use Holst's results and a truncation argument to deal directly with statistics of the Cramér–von Mises or Watson types. Without tightness we cannot handle statistics of the Kolmogorov–Smirnov type.

Our truncation argument uses an accurate approximation to the conditional expectation, given \mathbf{T}_n , of the statistic in question. This approximation is based on an expansion of the difference between a Rao–Blackwell estimate and the corresponding maximum likelihood estimate. Our results here extend the work of Portnoy [11].

Section 2 gives precise statements of our results for the case of Cramér–von Mises statistics. Section 3 gives the expansion of the Rao–Blackwell estimate. Section 4 applies the calculations to two examples showing how to verify the main condition, Condition D below, and illustrating the expansions of Section 3. Section 5 provides some discussion and indicates the extension

to Watson's statistic and other statistics which are quadratic functionals of the empirical distribution. In that section, we consider power and discuss various rephrasings of our main result. Details of some proofs are in Section 6.

2. Main results

2.1. Absolutely continuous distributions

We seek to test the hypothesis that the distribution of each X_i belongs to a natural exponential family with density, relative to some σ -finite measure $\mu(dx)$ on Ω , of the form (1) and complete sufficient statistic \mathbf{T}_n as described in the [Introduction](#). We will need a number of well known facts about exponential families which we gather here in the form of a lemma.

Lemma 1. *The random vector \mathbf{T}_n has moment generating function*

$$E_\theta[\exp\{\phi'\mathbf{T}_n\}] = \exp[n\{\kappa(\phi + \theta) - \kappa(\theta)\}]$$

which is finite whenever $\theta + \phi \in \Theta_0$, and cumulants $n\kappa_{i_1, \dots, i_r}$ where

$$\kappa_{i_1, \dots, i_r} = \frac{\partial^r \kappa(\theta)}{\partial \theta_{i_1} \dots \partial \theta_{i_r}}.$$

In particular, the mean of \mathbf{T}_n is

$$E_\theta(\mathbf{T}_n) = n\mu(\theta) \equiv n\nabla\kappa(\theta),$$

where ∇ is the gradient operator. The covariance matrix is

$$\text{Var}_\theta(\mathbf{T}_n) \equiv nV(\theta) = n\nabla^2\kappa(\theta),$$

where ∇^2 denotes the Hessian operator. Thus, $V(\theta)$ has entries

$$V_{ij}(\theta) = \frac{\partial^2 \kappa(\theta)}{\partial \theta_i \partial \theta_j}.$$

Moreover, all moments and cumulants of \mathbf{T}_n depend smoothly on θ on the interior of Θ .

Our results apply to exponential families where \mathbf{T}_n has a density relative to Lebesgue measure. We assume the following condition.

Condition D. *For every compact subset Γ of $\text{int}(\Theta)$, there is an integer r such that the characteristic function*

$$\eta_\theta(\phi) \equiv E_\theta\{\exp(i\phi'\mathbf{T}_r)\} = \exp[r\{\kappa(\theta + i\phi) - \kappa(\theta)\}]$$

is integrable for all $\theta \in \Gamma$ and

$$\sup_{\theta \in \Gamma} \int_{\mathbb{R}^k} |\eta_{\theta}(\phi)| \, d\phi < \infty.$$

Condition **D** has two consequences we need. First, it means the matrix $\text{Var}_{\theta}(\mathbf{T}_1) = \nabla^2 \kappa(\theta)$ is positive definite for each $\theta \in \text{int}(\Theta)$. This implies the map $\theta \mapsto \mu(\theta) = \nabla \kappa(\theta)$ is an open bijective mapping of $\text{int}(\Theta)$ to $\mu(\text{int}(\Theta))$. A second consequence is that \mathbf{T}_n has bounded continuous density for each $\theta \in \Gamma$ and $n \geq r$. In the examples it will be useful to know the converse is also true. The following lemma is essentially Theorem 19.1 in Bhattacharya and Ranga Rao [2], page 180; see also Lemma 6 in Section 6 below.

Lemma 2. *Condition **D** is equivalent to Condition **D**^{*}.*

Condition **D^{*}.** *For every compact subset Γ of $\text{int}(\Theta)$ there is an integer r such that \mathbf{T}_r has continuous (Lebesgue) density $f_r(t; \theta)$ for each $\theta \in \Gamma$ and*

$$\sup_{\theta \in \Gamma} \sup_{t \in \mathbb{R}^k} f_r(t, \theta) < \infty.$$

As in the **Introduction**, we let $G_n(\cdot|t)$ denote the conditional cumulative distribution function of S_n given $\mathbf{T}_n = t$. Also let $H_n(\cdot; \theta)$ denote the unconditional cumulative distribution function of S_n when θ is the value of the parameter.

We will show that for statistics which are sums as in (3) below or of the Cramér–von Mises type these two cumulative distributions are uniformly close provided that t and θ are related properly, that is, $t = n\mu = n\nabla \kappa(\theta)$.

Our results use a minor modification of Corollary 3.6 of Holst [4] which establishes this uniform closeness for statistics which are sums over the data as described below. We use the following notation. By $\mathcal{L}(S_n; \theta)$ we mean the unconditional distribution of S_n under the model with true parameter θ . By $\mathcal{L}(S_n|\mathbf{T}_n = t)$ we mean the conditional distribution of S_n given $\mathbf{T}_n = t$. We use the symbol \Rightarrow to denote convergence in distribution (weak convergence) and $\mathcal{L}(W)$ and similar notation for limiting distributions. Our version of Holst's results is:

Lemma 3. *Assume Condition **D**. Suppose that $u_n(\cdot; \cdot)$ is a sequence of measurable functions mapping $\Omega \times \Theta$ to \mathbb{R}^m . Let*

$$S_n(\theta) = n^{-1/2} \sum_{i=1}^n [u_n(X_i, \theta) - E_{\theta}\{u_n(X_i, \theta)\}]. \quad (3)$$

Assume that for any deterministic sequence θ_n of parameter values converging to some $\theta \in \text{int}(\Theta)$ the joint law

$$\mathcal{L}_{\theta_n}(S_n(\theta_n), n^{-1/2}\{\mathbf{T}_n - n\mu(\theta_n)\})$$

converges to multivariate normal with mean 0 and variance–covariance matrix of the form

$$\begin{bmatrix} A(\theta) & B(\theta) \\ B'(\theta) & V(\theta) \end{bmatrix}$$

which may depend on θ but not on the specific sequence θ_n . Then with S_n denoting $S_n(\hat{\theta}_n)$ we have for every such sequence θ_n

$$\mathcal{L}(S_n|T_n = t_n) \equiv G_n(\cdot|n\mu(\theta_n)) \Rightarrow \text{MVN}(0, A(\theta) - B(\theta)V^{-1}(\theta)B'(\theta)),$$

where $t_n = n\mu(\theta_n)$. Moreover, for every compact subset Γ of $\text{int}(\Theta)$ we have

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} \sup_{\theta \in \Gamma} |G_n(x|n\mu) - H_n(x|\theta)| = 0.$$

The condition involving the sequence θ_n amounts to requiring that the central limit theorem apply uniformly on compact subsets of Θ . Our main result extends the last conclusion of the lemma to statistics of the Cramér–von Mises type for the case where Ω is the real line; see Remark 8 in Section 5 for discussion of more general sample spaces.

Theorem 1. Suppose S_n is as defined in (2). Suppose the weight ψ is continuous on $[0, 1]$. Assume Condition D. Then for every compact subset Γ of $\text{int}(\Theta)$ we have

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} \sup_{\theta \in \Gamma} |G_n(x|n\mu) - H_n(x|\theta)| = 0. \quad (4)$$

The theorem asserts that two distribution functions, one conditional, the other unconditional, are close together everywhere and simultaneously for all θ belonging to some compact set. In the Introduction, we described our results in terms of P -values; we now recast the theorem in those terms. The conditional P value, now denoted $P_{c,n}$, is

$$P_{c,n} \equiv 1 - G_n(S_n|\mathbf{T}_n).$$

The unconditional P value, $P_{u,n}$, is

$$P_{u,n} \equiv 1 - H_n(S_n; \hat{\theta}_n).$$

We then have the following result which also clarifies the sampling properties of the distributions G_n and H_n evaluated at sample estimates.

Theorem 2. Assume the conditions of Theorem 1.

- (a) If X_1, X_2, \dots is an i.i.d. sequence generated from the model with true parameter value $\theta \in \text{int}(\Theta)$ (i.e., if the null hypothesis is true and the true parameter value is not on the boundary of the parameter space), then

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} \sup_{\theta \in \Gamma} |G_n(x|\mathbf{T}_n) - H_n(x|\hat{\theta}_n)| = 0 \quad \text{almost surely}$$

and

$$P_{c,n} - P_{u,n} \rightarrow 0 \quad \text{almost surely.}$$

- (b) Suppose X_1, X_2, \dots is an i.i.d. sequence generated from some fixed alternative distribution. Suppose that for this alternative $E(\mathbf{T}_1) = \mu_a$ exists and is in the open set $\mu(\text{int}(\Theta))$, that is, the image of the interior of Θ under the map $\theta \mapsto \mu$. Then both conclusions of part (a) still hold. In particular, if one test is consistent against the alternative then so is the other.

Details of proofs are in Section 6 but here we outline the strategy of proof for our Theorem 1. Fix a complete orthonormal system of functions g_j defined on $[0, 1]$; for definiteness we take $g_j(s) = \sqrt{2} \sin(\pi j s)$. Define

$$U_{n,j} = \int_0^1 \psi(s) W_n(s) g_j(s) ds.$$

Then by Parseval's identity

$$S_n = \sum_{j=1}^{\infty} U_{n,j}^2.$$

The proof then has the following steps:

1. The sequence of distribution functions $H_n(\cdot|\theta)$ converges weakly to a limiting distribution function $H_{\infty}(\cdot|\theta)$; the convergence is uniform on compact subsets of Θ . The distribution in question is the law of

$$\int_0^1 \psi^2(s) W^2(s) ds = \sum_{j=1}^{\infty} U_{\infty,j}^2,$$

where we define

$$U_{\infty,j} = \int_0^1 \psi(s) W(s) g_j(s) ds.$$

This reduces the problem to proving that the sequence $G_n(\cdot|n\mu)$ converges uniformly on compact subsets of $\text{int}(\Theta)$ to $H_{\infty}(\cdot|\theta)$ where $\mu = \nabla\kappa(\theta)$.

2. Uniform convergence is established by considering an arbitrary sequence θ_n of parameter values converging to some $\theta \in \text{int}(\Theta)$ and showing that, with $\mu_n = \nabla\kappa(\theta_n)$,

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} |G_n(x|n\mu_n) - H_{\infty}(x; \theta)| = 0. \quad (5)$$

3. Apply standard weak convergence ideas to see that for each K fixed

$$\mathcal{L}((U_{n,1}, \dots, U_{n,K}); \theta_n) \Rightarrow \mathcal{L}((U_{\infty,1}, \dots, U_{\infty,K})).$$

4. Use Holst's results to prove that

$$\mathcal{L}((U_{n,1}, \dots, U_{n,K}) | \mathbf{T}_n = n\mu_n) \Rightarrow \mathcal{L}((U_{\infty,1}, \dots, U_{\infty,K}));$$

this is the same joint limit law as in the previous step.

5. Prove the sequence $\mathcal{L}(S_n | \mathbf{T}_n = n\mu_n)$ of conditional distributions of S_n is tight.
 6. Prove that there is a sequence K_n tending to infinity sufficiently slowly that

$$\mathcal{L}\left(\sum_{j=1}^{K_n} U_{n,j}^2; \theta_n\right) \Rightarrow \mathcal{L}\left(\int_0^1 \psi^2(s) W^2(s) ds\right).$$

7. Prove the corresponding conditional result given $\mathbf{T}_n = n\mu_n$.
 8. Prove that for any sequence K_n tending to infinity

$$\sum_{j=K_n}^{\infty} U_{n,j}^2$$

converges to 0 in probability given $\mathbf{T}_n = n\mu_n$.

9. Apply Slutsky's theorem to 6, 7 and 8 and use $S_n = \sum_{j=1}^{\infty} U_{n,j}^2$ to see

$$\mathcal{L}(S_n | \mathbf{T}_n = n\mu_n) \Rightarrow \mathcal{L}\left(\int_0^1 \psi^2(s) W^2(s) ds\right)$$

which establishes (5) and completes the proof.

2.2. Unconditional limits

We now consider the random function $Y_n(t) = \psi(t)W_n(t)$ and review some well known facts about the unconditional limiting distributions of the processes Y_n ; see Shorack and Wellner [12], for example. If θ_n converges to θ , then the unconditional laws of Y_n converge weakly in $D[0, 1]$ to the law of a Gaussian process Y with mean 0 and covariance

$$\zeta_{\theta}(s, t) = \psi(s)\rho_{\theta}(s, t)\psi(t).$$

The covariance ζ_{θ} is square integrable over the unit square; it is convenient to suppress θ in the notation for what follows. There is a sequence of bounded continuous orthonormal eigenfunctions $\chi_j(t)$, $j = 1, 2, \dots$, with corresponding eigenvalues λ_j such that

$$\int_0^1 \zeta(s, t) \chi_j(t) dt \equiv \lambda_j \chi_j(s).$$

Then

$$\int_0^1 Y^2(t) dt = \sum \lambda_j Z_j^2, \tag{6}$$

where

$$Z_j = \lambda_j^{-1/2} \int_0^1 Y(t) \chi_j(t) dt.$$

The Z_j are independent standard normal. Let $H_\infty(\cdot; \theta)$ denote the cumulative distribution of (6). It is then standard that

$$\lim_{n \rightarrow \infty} \sup_{-\infty < x < \infty} \sup_{\theta \in \Gamma} |H_n(x; \theta) - H_\infty(x; \theta)| = 0.$$

Our main result will therefore follow if we establish (5).

Next, recall that W_n converges weakly to the Gaussian process W with covariance function ρ_θ . The map

$$f \mapsto \left(\int_0^1 f(s) \psi(s) g_1(s) ds, \dots, \int_0^1 f(s) \psi(s) g_K(s) ds \right)$$

is continuous from $D[0, 1]$ to \mathbb{R}^K so that

$$(U_{n,1}, \dots, U_{n,K}) \Rightarrow (U_{\infty,1}, \dots, U_{\infty,K}).$$

This limit vector has a multivariate normal distribution with mean 0 and covariance

$$\text{Cov}(U_{\infty,i}, U_{\infty,j}) = \int_0^1 \int_0^1 g_i(s) g_j(t) \zeta_\theta(s, t) ds dt. \quad (7)$$

It follows that

$$\sum_{j=1}^K U_{n,j}^2 \Rightarrow \sum_{j=1}^K U_{\infty,j}^2.$$

Since

$$\int_0^1 \psi^2(s) W^2(s) ds = \sum_{j=1}^{\infty} U_{\infty,j}^2$$

almost surely we have, for any sequence K_n tending to infinity, that

$$\sum_{j=1}^{K_n} U_{\infty,j}^2 \Rightarrow \int_0^1 \psi^2(s) W^2(s) ds.$$

This completes the analysis of the unconditional limit behaviour of S_n . The next subsection considers the conditional limit behaviour.

2.3. Convergence of finite dimensional distributions – conditional case

In the following, all distributional assertions are statements about the conditional distribution of the objects involved given $\mathbf{T}_n = n\mu_n$ for a specific sequence θ_n converging to some $\theta \in \Gamma$ and $\mu_n = \nabla K(\theta_n)$. We apply Lemma 3 as follows. We have

$$U_{n,j} \equiv \int_0^1 \psi(t) W_n(t) g_j(t) dt = n^{-1/2} \sum_{i=1}^n \Phi_{jn}(X_i),$$

where

$$\Phi_{jn}(x) = \int_0^1 [1\{F(x; \theta_n) \leq t\} - t] \psi(t) g_j(t) dt.$$

It follows from Lemma 3 that

$$\mathcal{L}((U_{n,1}, \dots, U_{n,K}) | \mathbf{T}_n = n\mu_n) \Rightarrow \mathcal{L}((U_{\infty,1}, \dots, U_{\infty,K})).$$

The vector $(U_{\infty,1}, \dots, U_{\infty,K})$ has a multivariate normal distribution with mean 0 and variance covariance matrix with entries as at (7). This is the same limit behaviour as in the unconditional case. Thus,

$$\mathcal{L}\left(\sum_{j=1}^K U_{n,j}^2 | \mathbf{T}_n = n\mu_n\right) \Rightarrow \sum_{j=1}^K U_{\infty,j}^2.$$

Again this is the same weak limit as in the previous section. Finally, since convergence in distribution is metrizable there is a sequence K_n tending to infinity so slowly that

$$\mathcal{L}\left(\sum_{j=1}^{K_n} U_{n,j}^2 | \mathbf{T}_n = n\mu_n\right) \Rightarrow \sum_{j=1}^{\infty} U_{\infty,j}^2.$$

We need only show, therefore, that for any sequence K_n tending to infinity we have, conditionally on $\mathbf{T}_n = n\mu_n$,

$$\sum_{j=K_n+1}^{\infty} U_{n,j}^2 \rightarrow 0$$

in probability. It suffices to show that

$$\mathbb{E}\left(\sum_{j=K_n+1}^{\infty} U_{n,j}^2 | \mathbf{T}_n = n\mu_n\right) \rightarrow 0. \quad (8)$$

We will prove this from the following statements. First, we will show that for each fixed j

$$\mathbb{E}(U_{n,j}^2 | \mathbf{T}_n = n\mu_n) \rightarrow \mathbb{E}\{U_{\infty,j}^2\}. \quad (9)$$

This shows that for each fixed K we have

$$\mathbb{E}\left(\sum_{j=1}^K U_{n,j}^2 \mid \mathbf{T}_n = n\mu_n\right) \rightarrow \mathbb{E}\left(\sum_{j=1}^K U_{\infty,j}^2\right). \quad (10)$$

Finally, we will show that

$$\mathbb{E}\left(\sum_{j=1}^{\infty} U_{n,j}^2 \mid \mathbf{T}_n = n\mu_n\right) \rightarrow \mathbb{E}\left(\sum_{j=1}^{\infty} U_{\infty,j}^2\right). \quad (11)$$

Assertion (8) is a straightforward consequence of (9) and (11). It is now straightforward to apply Slutsky's theorem to complete the proof of the main theorem.

Statements (9) and (11) are proved in Section 6. The proofs relate $\mathcal{L}(U_{n,j}^2 \mid \mathbf{T}_n = n\mu_n)$ to an integral involving

$$\mathbb{E}(1(X_i \leq x)1(X_k \leq y) \mid \mathbf{T}_n = n\mu_n)$$

and other similar Rao–Blackwell estimates. They then use a conditional Edgeworth expansion of Rao–Blackwell estimates which is of some interest in its own right. We describe these expansions in the next section.

3. Conditional Edgeworth expansions

In this section, we compute the first term in an Edgeworth expansion of the conditional expectation of a function of X_1, \dots, X_m given \mathbf{T}_n . We will focus on uniformity, extending the work of Portnoy [11]. The calculations may be interpreted as a computation of the difference, to order $1/n$, between a Rao–Blackwell estimate of a parameter and the maximum likelihood estimate.

Our results use the Edgeworth expansion of the density of \mathbf{T}_n . Assuming Condition D, for $n \geq r$ the quantity $\{\mathbf{T}_n - n\mu(\theta)\}/\sqrt{n}$ has a density $q_n(\cdot; \theta)$. The following lemma is essentially a uniform version of Theorem 19.2 in Bhattacharya and Ranga Rao [2]; see Holst [4], Yuan and Clarke [15]. It extends a lemma appearing in Lockhart and O'Reilly [6]. Let \mathbf{u} denote a k vector with entries u_1, \dots, u_k .

Lemma 4. *Assume Condition D. Then there are functions*

$$\psi_j(\mathbf{u}; \theta), \quad j = 1, 2, \dots,$$

and

$$\psi_{jk}(\mathbf{u}; \theta), \quad k = 0, \dots, j + 2,$$

such that

1. ψ_{jk} is homogeneous of degree k as a function of u_1, \dots, u_k . That is

$$\psi_{jk}(u_1, \dots, u_k; \theta) = \sum_{i_1 \dots i_k} a_{jk; i_1 \dots i_k}(\theta) u_{i_1} \dots u_{i_k}$$

for some coefficients $a_{jk;i_1 \dots i_k}(\theta)$ not depending on \mathbf{u} .

2. If $j - k$ is odd, then $\psi_{jk} \equiv 0$.

3. ψ_j is a polynomial of degree $j + 2$ as a function of \mathbf{u} given by

$$\psi_j = \sum_{k=0}^{j+2} \psi_{jk}.$$

4. The coefficients $a_{jk;i_1 \dots i_k}(\theta)$ in these polynomials are smooth functions of θ .

5. Fix an integer $s \geq 0$ and a compact subset Γ of $\text{int}(\Theta)$. Let $\phi(\mathbf{u}, V)$ be the multivariate normal density with mean 0 and covariance matrix V . Then

$$\varepsilon_n \equiv \sup_{\theta \in \Gamma} \sup_{\mathbf{u}} \left| q_n(\mathbf{u}; \theta) - \phi\{\mathbf{u}, V(\theta)\} \left\{ 1 + \sum_{j=1}^s \frac{\psi_j(\mathbf{u}, \theta)}{n^{j/2}} \right\} \right| = O(n^{-(s+1)/2}).$$

We will use this lemma with $s = 3$ to get an error rate on our 1 term expansion. We need the following notation. Define

$$B_m(x_1, \dots, x_m) = \sum_{i=1}^m \{T(x_i) - \mu\}$$

and let \mathbf{B}_m denote the random vector

$$B_m(X_1, \dots, X_m) = \mathbf{T}_m - m\mu.$$

Let $D = V^{-1}$ be the inverse of the variance covariance matrix V . The lowest degree term in the polynomial ψ_1 has the form

$$\psi_{1,1}(\mathbf{u}) = - \sum_1^k a_{1,1;\ell}(\theta) u_\ell$$

where, from Bhattacharya and Ranga Rao [2], page 55, we have

$$\begin{aligned} a_{1,1;\ell} &= \sum_i \kappa_{iii} D_{ii} D_{i\ell} / 2 + \sum_{i \neq j} \kappa_{ijj} (2D_{ij} D_{i\ell} + D_{ii} D_{j\ell}) / 2 \\ &\quad + \sum_{i < j < k} \kappa_{ijk} (D_{ij} D_{k\ell} + D_{ik} D_{j\ell} + D_{jk} D_{i\ell}). \end{aligned}$$

If $J(x_1, \dots, x_m)$ is a real valued measurable function on Ω^m ; we let $\mathbf{J} = J(X_1, \dots, X_m)$. Remember in the following that μ and θ are related through $E_\theta(\mathbf{T}_n) = n\mu$.

Theorem 3. Fix an integer $m > 0$. Suppose $\bar{J} \geq 0$ is a real valued measurable function on Ω^m such that

$$E_\theta\{\bar{J}(X_1, \dots, X_m)\} < \infty$$

for all $\theta \in \text{int}(\Theta)$. Then for each compact subset Γ of $\text{int}(\Theta)$ we have

$$\limsup_{n \rightarrow \infty} n^2 \sup_{\theta \in \Gamma} \sup_J |\mathbb{E}\{\mathbf{J}|\mathbf{T}_n = n\mu\} - A(n, J, \theta)| < \infty, \quad (12)$$

where

$$A(n, J, \theta) \equiv \mathbb{E}_\theta(\mathbf{J}) + \frac{R(J, \theta)}{n}$$

and

$$R(J, \theta) \equiv \frac{mk}{2} \mathbb{E}_\theta(\mathbf{J}) - \frac{1}{2} \mathbb{E}_\theta\{\mathbf{J}\mathbf{B}_m' V^{-1}(\theta)\mathbf{B}_m\} - \mathbb{E}_\theta\{\mathbf{J}\psi_{1,1}(\mathbf{B}_m)\} \quad (13)$$

$$= \nabla^2 \mathbb{E}_\theta(\mathbf{J}) + \psi_{1,1}\{\nabla \mathbb{E}_\theta(\mathbf{J})\}. \quad (14)$$

The supremum over J is over all measurable J defined on Ω^m with $|J| \leq \bar{J}$ (almost everywhere). Moreover,

$$\sup_{\theta \in \Gamma} \sup_{J: |J| \leq \bar{J}} R(J, \theta) = O(1). \quad (15)$$

In (14), the symbols ∇ and ∇^2 are as in Lemma 1. It is part of the theorem that the quantities on the right in (13) and (14) are equal.

4. Examples

In this section, we consider the Gamma and von Mises models and show that the theory of the previous sections applies. These two models were considered in Lockhart, O'Reilly and Stephens [7, 8] where Gibbs sampling was used to implement the conditional tests discussed here via Markov Chain Monte Carlo. In the case of the Gamma distribution, we also illustrate the use of the expansion of the Rao–Blackwell estimate by giving a formula for an approximate Rao–Blackwell estimate of the shape parameter.

4.1. The Gamma distribution

Suppose X_1, X_2, \dots are i.i.d. with density

$$f(x; \alpha, \beta) = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp(-x/\beta) 1(x > 0).$$

We take $\theta_1 = \alpha$, $\theta_2 = 1/\beta$ and $\Theta = \{\theta : \theta_1 > 0, \theta_2 > 0\}$. We then have

$$T(x) = (\log(x), -x)$$

and

$$\kappa(\phi_1, \phi_2) = \log \left\{ \frac{\Gamma(\theta_1 + \phi_1)}{\Gamma(\theta_1)} \frac{\theta_2^{\theta_1}}{(\theta_2 + \phi_2)^{\theta_1 + \phi_1}} \right\}.$$

The characteristic function of \mathbf{T} is

$$\Psi(\phi_1, \phi_2) = \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \frac{(\theta_2 + i\phi_2)^{\theta_1 + i\phi_1}}{\theta_2^{\theta_1}}.$$

Fix a compact set Γ in the parameter space and let

$$\varepsilon = \inf\{\theta_1: \exists \theta_2: (\theta_1, \theta_2) \in \Gamma\}.$$

In Section 6, we use properties of the Gamma function in the complex plane to show that for r so large that $r\varepsilon > 2$ and $r > 4$ we have

$$\sup_{\theta \in \Gamma} \int |\Psi(\phi_1, \phi_2)|^r d\phi_1 d\phi_2 < \infty. \quad (16)$$

This establishes Condition D in this case.

For completeness, we record here the functions needed to apply Theorem 3 to this family. Let $\psi(\theta) = d \log \Gamma(\theta) / d\theta$ denote the digamma function and let ψ' and ψ'' denote its first and second derivatives. Let $\delta = \theta_1 \psi'(\theta_1) - 1$. Then we find

$$\begin{aligned} \mu_1 &= \psi(\theta_1) - \log(\theta_2), & \mu_2 &= -\theta_1/\theta_2, \\ V_{11} &= \psi'(\theta_1), & V_{12} &= V_{21} = 1/\theta_2, \\ V_{22} &= \theta_1/\theta_2^2, & D_{12} &= D_{21} = \theta_2/\delta, \\ D_{11} &= \theta_1/\delta, & D_{22} &= \theta_2^2 \psi'(\theta_1)/\delta, \\ \kappa_{111} &= \psi''(\theta_1), & \kappa_{112} &= \kappa_{121} = \kappa_{211} = 0, \\ \kappa_{222} &= -2\theta_1/\theta_2^3, & \kappa_{122} &= \kappa_{221} = \kappa_{212} = 1/\theta_2^2, \\ a_{11;1} &= \frac{\theta_1^2 \psi''(\theta_1) + 2\theta_2 \psi'(\theta_1) + 2}{2\delta^2}, & a_{11;2} &= \frac{\theta_1 \psi''(\theta_1) + 2\theta_2 \{\psi'(\theta_1)\}^2 + 2\theta_2 \psi'(\theta_1)}{2\delta^2}. \end{aligned}$$

These formulas may be used to give approximations in terms of the maximum likelihood estimate $\hat{\theta}$ to order $1/n$ of the Rao–Blackwell estimate of a parameter. As an example, we consider the approximation to the Rao–Blackwell estimate of the shape parameter θ_1 . In this case $E_\theta(\mathbf{J}) = \theta_1$ so the Hessian matrix in $R(J, \theta)$ is 0 and the gradient is simply $(1, 0)'$. Our approximation from (14) is then

$$\tilde{\theta}_1 = \hat{\theta}_1 - \frac{\hat{\psi}_{1,1}(1, 0)}{n} = \hat{\theta}_1 + \frac{\hat{\theta}_1^2 \psi''(\hat{\theta}_1) + 2\hat{\theta}_2 \psi''(\hat{\theta}_1) + 2}{2n\{\hat{\theta}_1 \psi'(\hat{\theta}_1) - 1\}^2}.$$

Remark. I do not know if there is, for some value of m , an unbiased estimate of θ_1 . That is, I do not know if \mathbf{J} exists in the calculation just given. It seems worth noting that the expansion can be computed anyway since the terms therein depend only on the function of the parameters which is being estimated and the derivatives of that function.

4.2. The von Mises distribution

Suppose X_1, X_2, \dots are i.i.d. with density

$$f(x; \alpha, x_0) = \frac{1}{2\pi I_0(\alpha)} \exp\{\alpha \cos(x - x_0)\} 1(0 < x < 2\pi),$$

where I_0 is the modified Bessel function of the first kind of order 0. We take $\theta_1 = \alpha \cos(x_0)$, $\theta_2 = \alpha \sin(x_0)$ and $\Theta = \mathbb{R}^2$. We then have

$$T(x) = (\cos(x), \sin(x)).$$

Here we find it easier to verify Condition \mathbf{D}^* . For a sample of size m the density of the sufficient statistics is known analytically in the case $\theta_1 = \theta_2 = 0$, that is, when the distribution is uniform on the interval $(0, 2\pi)$. Write \mathbf{T}_m in polar coordinates as $(R \cos \delta, R \sin \delta)$ with the angle δ in $[0, 2\pi)$ and $R = \|\mathbf{T}_m\|$; then R and δ are independent. The distribution of δ is uniform on $[0, 2\pi)$. From Stephens [13], we find R has the density

$$f_m(u) = u \int_0^\infty J_0(ut) J_0^m(t) t \, dt,$$

where J_0 is the Bessel function of the first kind of order 0. The function $J_0(t)$ is bounded and decays at infinity like $t^{-1/2}$. So for all $m > 4$ there is a constant C_m such that

$$f_m(u) \leq C_m u$$

for all $u > 0$. The density f_m vanishes for negative u and for $u > m$. Change variables to see that for all $m \geq 5$ the density of \mathbf{T}_m is bounded by $C_m/(2\pi)$. For $\theta = (\theta_1, \theta_2)$ not 0 the likelihood ratio of θ to 0 is $\exp(\theta' \mathbf{T}_m)/I_0^m(\|\theta\|)$. Since the density of \mathbf{T}_m for θ is the density for 0 multiplied by the likelihood ratio Condition \mathbf{D}^* holds with $r \equiv 5$.

5. Discussion

We conclude with a series of remarks.

Remark 1. For a given goodness-of-fit test statistic we may compute P -values in several ways. The parametric bootstrap technique proceeds by estimating the unknown parameters and then generating a large number of samples from the hypothesized distribution using the estimated value of the parameters. Except in location-scale models the resulting tests are approximate; that is, the distribution of the P -value is not exactly uniform though it becomes more so as the sample size increases.

An alternative technique is to compute a conditional P value using

$$P(S_n > s | \mathbf{T}_n)$$

evaluated at s equal to the observed value of S_n . This P -value must generally be evaluated by Monte Carlo methods. For some distributions, such as the Inverse Gaussian, there is a direct way to simulate samples from the conditional distribution of the data given \mathbf{T}_n . See O'Reilly and Gracia-Medrano [10]. For other distributions, Markov Chain Monte Carlo may be used; see Lockhart, O'Reilly and Stephens [7,8].

If the null hypothesis is true and the true value of θ is in $\text{int}(\Theta)$, then we have shown that the difference between these two P -values converges almost surely to 0. In our experience, these two P -values are usually extremely close together suggesting the agreement extends to some higher order expansion; I do not know how to show such a thing.

Remark 2. Indeed this equivalence of P -values requires only a large sample size and an estimate $\hat{\theta}$ not too close to the boundary of Θ . It is not at all necessary that the null hypothesis be true. Of course if the null hypothesis is not true the estimate $\hat{\theta}_n$ could converge to the boundary of the parameter space and then our results permit the P -values to be different even in large samples.

Remark 3. For fixed alternatives, our results imply that the difference in powers between the two tests tends to 0 except when \mathbf{T}_n/n does not have a limit in $\mu(\text{int}(\Theta))$. The conclusions in Theorem 2 can be extended to contiguous sequences of alternatives yielding conclusions that the two tests have identical limiting powers along such sequences.

Remark 4. The local central limit theorem for lattice distributions may be used to prove the equivalent of Theorem 1 if $T(x)$ takes values in a lattice and the data are discrete.

Remark 5. The result also extends to a variety of other statistics such as

$$\int_0^1 \left\{ W_n(t) - \int_0^1 \psi(u) W_n(u) du \right\}^2 \psi^2(t) dt$$

or

$$\int_0^1 \int_0^1 K(s, t) W_n(s) W_n(t) ds dt$$

or any other suitable quadratic form in the process W_n , under regularity conditions on the weight functions ψ , the kernel K , or the quadratic form.

Remark 6. One important case not covered by our proof is the Anderson–Darling test which is of the Cramér–von Mises type but with weight function

$$\psi(s) = 1/\sqrt{s(1-s)}$$

which is not square integrable. It may be possible to verify our assertions (9) and (11) by more careful analysis of the conditional moments of W_n near the ends of the unit interval.

Remark 7. Our proofs show that the Edgeworth expansion to order $2s$ given in Lemma 4 may be used to provide an expansion of any Rao–Blackwell estimate about the maximum likelihood estimate of $E_\theta(\mathbf{J})$ in inverse powers of n out to terms of order n^{-s} with a remainder which is $O(n^{-(s+1)})$ uniformly on compact subsets of $\text{int}(\Theta)$. We have not done the algebra for any $s > 1$ but we can state the following theorem.

Theorem 4. *Under the conditions of Theorem 3, there are functions $R_j(J, \theta)$ for $j = 1, 2, \dots$ such that for any integer $s \geq 1$ we have*

$$\limsup_{n \rightarrow \infty} n^{1+s} \sup_{\theta \in \Gamma} \sup_J |E\{\mathbf{J}|\mathbf{T}_n = n\mu\} - A_s(n, J, \theta)| < \infty, \quad (17)$$

where

$$A_s(n, J, \theta) \equiv E_\theta(\mathbf{J}) + \sum_{j=1}^s \frac{R_j(J, \theta)}{n^j}.$$

The functions R_j are computed using Taylor expansions as in Theorem 3 and collecting terms in inverse powers of n . Each R_i is bounded uniformly over $\theta \in \Gamma$ and $|J| \leq \bar{J}$.

Of course R_1 is just R of Theorem 3 and the point is that the arguments in the proof of that theorem can be applied to all remainder terms occurring here.

Remark 8. In Theorem 1, the X_i are real valued; this is needed only for the weak convergence results. In the von Mises case, for instance, it is useful to regard the observation X_i not as an angle but as a unit vector X_i as was suggested in the introduction. This makes $\mathbf{T}_n = \sum X_i$. In many examples, the X_i can usefully be taken to be multivariate. Our results may be expected to extend to any statistic admitting a sum of squares expansion like that of Cramér–von Mises statistics.

Remark 9. The conditional tests described here have level identically equal to α . In the introduction, we noted that this is a necessary condition for an unbiased level α test in models with a complete sufficient statistic. Though necessary, the condition is not sufficient; we do not know how to check that a given conditional test is unbiased, nor how to establish any optimal power properties for the tests considered here.

6. Proofs

6.1. Proof of Lemma 3

The proof in Holst [4] of his Corollary 3.6 extends directly to prove this lemma. However, Holst's Corollary 3.6 assumes “the general conditions” of his Section 2. In particular, we must verify the integrability hypothesis of his Proposition 2.1 which we now describe in our notation. Let

$$\Psi_{r,\theta}(\zeta_1, \zeta_2) = E_\theta\{\exp(i\zeta_1' S_r(\theta) + i\zeta_2' \mathbf{T}_r)\}$$

be the joint characteristic function of $S_r(\theta)$, \mathbf{T}_r . Holst requires that for each ζ_1 and each compact subset Γ of $\text{int}(\Theta)$ there is an $r > 0$ such that for all $\theta \in \Gamma$

$$\int |\Psi_{r,\theta}(\zeta_1, \zeta_2)| d\zeta_2 < \infty. \quad (18)$$

Lemma 5. *Condition D implies (18). In fact, r can be chosen free of ζ_1 .*

This is an easy consequence of the following lemma.

Lemma 6. *Suppose $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^m$ have joint distribution $F(dx, dy)$ and joint characteristic function $\psi(u, v)$. Then*

1. *If Y has density f bounded by M and ψ is real valued and nonnegative, then*

$$\int_{\mathbb{R}^m} \psi(u, v) dv \leq M(2\pi)^m.$$

2. *If Y has density f bounded by M , then*

$$\int_{\mathbb{R}^m} |\psi(u, v)|^2 dv \leq M(2\pi)^m.$$

3. *If $M \equiv \int_{\mathbb{R}^m} |\psi(0, v)| dv < \infty$, then Y has a density f such that for all y*

$$f(y) \leq M/(2\pi)^m.$$

Proof. Statement 3 is a well-known consequence of the Fourier inversion formula. Statement 2 follows from Statement 1 by symmetrization: if the pair (X^*, Y^*) has the same joint distribution as (X, Y) and is independent of (X, Y) then the second statement is the first applied to $(X - X^*, Y - Y^*)$ noting that $Y - Y^*$ has a density also bounded by M .

To prove Statement 1, we follow Feller [3], pages 480ff. Let ξ denote the standard normal density in \mathbb{R}^m . Then for each $a > 0$ the function $a\xi(ax)$ is a density with characteristic function $(2\pi)^{m/2}\xi(u/a)$.

$$\begin{aligned} \int \exp\{-i\zeta'v\} a^m \xi(av) \psi(u, v) dv &= \int a^m \xi(av) e^{iu'x} \exp\{iv'(y - \zeta)\} F(dx, dy) dv \\ &= (2\pi)^{m/2} \int e^{iu'x} \xi\{(y - \zeta)/a\} F(dx, dy). \end{aligned}$$

At $\zeta = 0$, we get

$$\begin{aligned} 0 &\leq \int \psi(u, v) \exp(-a^2 v'v/2) dv \leq (2\pi)^m \int (1/a)^m \xi(y/a) F(dx, dy) \\ &= (2\pi)^m \int (1/a)^m \xi(y/a) f(y) dy \leq M(2\pi)^m. \end{aligned}$$

Now let $a \rightarrow 0$ to get Statement 1. □

6.2. Proof of Theorem 3

We use the shorthands \mathbf{x} for the vector (x_1, \dots, x_m) and $d\mathbf{x}$ for $\mu(dx_1) \cdots \mu(dx_m)$. Let f_m be the joint density of X_1, \dots, X_m ; we suppress the dependence of this density on θ . For $n \geq r$, we let q_n denote the density of $(\mathbf{T}_n - n\mu)/\sqrt{n}$ again suppressing the dependence on θ . (Densities of sufficient statistics are relative to Lebesgue measure while those of the data are relative to products of the carrier measure μ .) We adopt the useful notation

$$Q_m = B'_m V^{-1} B_m, \quad Q_{mn} = Q_m/n \quad \text{and} \quad q_n^*(x) = q_n(x)/\phi(0, V).$$

It is elementary that

$$E\{\mathbf{J}|\mathbf{T}_n = n\mu\} = \left(\frac{n}{n-m}\right)^{k/2} \int J(\mathbf{x}) f_m(\mathbf{x}) \frac{q_{n-m}(A_m)}{q_n(0)} d\mathbf{x},$$

where

$$A_m = A_m(\mathbf{x}) = -\frac{\sum_{i=1}^m \{T(x_i) - \mu\}}{\sqrt{n}} = -\frac{B_m}{\sqrt{n}}.$$

The quantity in (12) may be written as $|I_1 + \cdots + I_8|$ where $I_i = \int J(\mathbf{x}) f_m(\mathbf{x}) \tau_i(\mathbf{x}) d\mathbf{x}$ for suitable functions τ_1, \dots, τ_8 . We will argue below that each integral is $O(n^{-2})$ uniformly in θ over compact subsets Γ of $\text{int}(\Theta)$. The functions τ_i are given by

$$\begin{aligned} \tau_1(\mathbf{u}) &= \left(\frac{n}{n-m}\right)^{k/2} \frac{q_{n-m}(A_m) - \phi(A_m, V)\{1 + \sum_{j=1}^4 \psi_j(A_m)/(n-m)^{j/2}\}}{q_n(0)}, \\ \tau_2(\mathbf{u}) &= \left\{ \left(\frac{n}{n-m}\right)^{k/2} - \left(1 + \frac{mk}{2n}\right) \right\} \frac{\phi(A_m, V)\{1 + \sum_{j=1}^4 \psi_j(A_m)/(n-m)^{j/2}\}}{q_n(0)}, \\ \tau_3(\mathbf{u}) &= \left(1 + \frac{mk}{2n}\right) \left[\frac{1}{q_n^*(0)} - \left\{1 - \frac{\psi_{2,0}}{n}\right\} \right] e^{-Q_{mn}/2} \left\{ 1 + \sum_{j=1}^4 \frac{\psi_j(A_m)}{(n-m)^{j/2}} \right\}, \\ \tau_4(\mathbf{u}) &= \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) e^{-Q_{mn}/2} \left\{ \sum_{j+\ell \geq 4} \frac{\psi_{j\ell}(A_m)}{(n-m)^{j/2}} \right\} \\ &= \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) e^{-Q_{mn}/2} \left\{ \sum_{j+\ell \geq 4} \frac{(-1)^\ell \psi_{j\ell}(B_m)}{n^{\ell/2} (n-m)^{j/2}} \right\}, \\ \tau_5(\mathbf{u}) &= \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) e^{-Q_{mn}/2} \psi_{1,1}(B_m) \left\{ \frac{1}{n} - \frac{1}{\sqrt{n(n-m)}} \right\}, \\ \tau_6(\mathbf{u}) &= \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) e^{-Q_{mn}/2} \psi_{2,0} \left\{ \frac{1}{n-m} - \frac{1}{n} \right\}, \\ \tau_7(\mathbf{u}) &= \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) \left(e^{-Q_{mn}/2} - 1 + \frac{Q_m}{2n} \right) \left\{ 1 + \frac{\psi_{2,0} - \psi_{1,1}(B_m)}{n} \right\}, \end{aligned}$$

$$\tau_8(\mathbf{u}) = \left(1 + \frac{mk}{2n}\right) \left(1 - \frac{\psi_{2,0}}{n}\right) \left(1 - \frac{Q_m}{2n}\right) \left\{1 + \frac{\psi_{2,0} - \psi_{1,1}(B_m)}{n}\right\} - \tau_9(\mathbf{u}),$$

where

$$\tau_9(\mathbf{u}) = 1 + \frac{mk/2 - Q_m/2 - \psi_{1,1}(B_m)}{n}.$$

Theorem 1 will follow if we show for $i = 1, \dots, 8$ that

$$\sup_{|J| \leq \bar{J}} \sup_{\theta \in \Gamma} |I_i| = O(n^{-2}).$$

These 8 assertions may be established using several bounds. We do not give complete details since the arguments are routine but we illustrate some of the details. For instance, it is elementary that

$$\left(\frac{n}{n-m}\right)^{k/2} \leq (m+1)^{k/2} \quad \text{and} \quad \left(\frac{n}{n-m}\right)^{k/2} - \left(1 + \frac{mk}{2n}\right) = O(n^{-2}).$$

Continuity and compactness imply

$$\sup_{\theta \in \Gamma} \sup_x \left| \phi(x, V) \left\{ 1 + \sum_{j=1}^s \frac{\psi_j(x)}{n^{j/2}} \right\} \right| < \infty$$

and

$$\inf_{\theta \in \Gamma} \phi(0, V) > 0.$$

Lemma 1 guarantees that

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \Gamma} q_n(0) > 0$$

and so with ε_n as in Lemma 4 we have

$$|I_1| \leq (m+1)^{k/2} \varepsilon_n \sup_{\theta \in \Gamma} E_\theta(\bar{\mathbf{J}}) / \inf_{\theta \in \Gamma} q_n(0).$$

For I_2 , I_5 and I_6 use the elementary facts that

$$\frac{1}{n-m} - \frac{1}{n} = O(n^{-2}) \quad \text{and} \quad \frac{1}{\sqrt{n(n-m)}} - \frac{1}{n} = O(n^{-2}).$$

Integral I_3 is bounded using Lemma 1 again. Integral I_4 uses the powers of n in the displayed sum. For I_7 use the inequalities $0 < e^{-x} - 1 + x < x^2/2$ to see that

$$0 < \left(e^{-Q_m/2} - 1 + \frac{Q_m}{2n} \right) < \frac{Q_m^2}{4n^2}.$$

These bounds apply to the integrands; they are used to bound the integrals based on the following observation. The condition that $\bar{\mathbf{J}}$ have finite expectation for all θ in $\text{int}(\Theta)$ means that

$\bar{J}(\mathbf{x})f_m(\mathbf{x})/E_\theta(\bar{\mathbf{J}})$ defines another exponential family with natural parameter space including $\text{int}(\Theta)$. This permits differentiation under the integral sign with respect to θ as many times as desired. It is then easily established that for all $\alpha > 0$

$$\sup_{\theta \in \Gamma} E_\theta(\|\mathbf{T}_r\|^\alpha \mathbf{J}) < \infty.$$

This permits all the bounds derived above to be integrated against $J(\mathbf{x})f_m(\mathbf{x})$ to establish the desired conclusion.

Differentiation under the integral sign permits us to show for any J with $|J| \leq \bar{J}$ the following two identities:

$$\begin{aligned}\nabla E_\theta(\mathbf{J}) &= \text{Cov}_\theta(\mathbf{J}, \mathbf{T}_m), \\ \nabla^2 E_\theta(\mathbf{J}) &= \text{Cov}_\theta(\mathbf{J}, \mathbf{B}_m \mathbf{B}_m') \\ &= E_\theta(\mathbf{J} \mathbf{B}_m \mathbf{B}_m') - E_\theta(\mathbf{J}) V.\end{aligned}$$

From these two identities, we deduce

$$\begin{aligned}E_\theta(\mathbf{J} \mathbf{B}_m' V^{-1} \mathbf{B}_m) &= \text{trace}\{E_\theta(\mathbf{J} \mathbf{B}_m \mathbf{B}_m') V^{-1}\} \\ &= \text{trace}\{\nabla^2 E_\theta(\mathbf{J}) V^{-1}\} + E_\theta(\mathbf{J}) \text{trace}(V^{-1} V).\end{aligned}$$

This and the observation that $\psi_{1,1}$ is a linear function establish the equivalence of the two forms of $R(J, \theta)$ in (13) and (14).

6.3. Proof of assertions (9) and (11)

We must prove

$$E[U_{nj}^2 | \mathbf{T}_n = n\mu_n] \rightarrow \int_0^1 \int_0^1 \psi(s) \psi(t) g_j(s) g_j(t) \rho_\theta(s, t) ds dt$$

and

$$E[S_n | \mathbf{T}_n = n\mu_n] \rightarrow \int_0^1 \psi^2(s) \rho_\theta(s, s) ds.$$

To this end, define

$$\tilde{F}(u|\mu) = E\{1(X_1 \leq x) | \mathbf{T}_n = n\mu\},$$

where u is related to x by $u = F(x, \theta)$. Then $\tilde{F}(u|\mathbf{T}_n/n)$ is the Rao–Blackwell estimate of $F(x, \theta)$. Also define $u_i = F(x_i, \theta)$ for $i = 1, 2$ and

$$\tilde{F}(u_1, u_2|\mu) = E\{1(X_1 \leq x_1, X_2 \leq x_2) | \mathbf{T}_n = n\mu\}.$$

Then $\tilde{F}_2(u_1, u_2|\mathbf{T}_n/n)$ is the Rao–Blackwell estimate of $F(x_1, \theta)F(x_2, \theta)$ (the unconditional joint cumulative distribution function of X_1 and X_2).

Define

$$\rho_n(u_1, u_2|\mu) = E\{W_n(u_1)W_n(u_2)|\mathbf{T}_n = n\mu\}.$$

We then have

$$E\left[\left\{\int_0^1 Y_n(t)g_j(t)dt\right\}^2 \middle| \mathbf{T}_n = n\mu_n\right] = \int_0^1 \int_0^1 \psi(s)\psi(t)g_j(s)g_j(t)\rho_n(s, t|\mu)dsdt.$$

Direct calculation shows that

$$\begin{aligned} \rho_n(u_1, u_2|\mu) &= \tilde{F}(\min(u_1, u_2)|\mu) - \tilde{F}(u_1|\mu)u_2 - \tilde{F}(u_2|\mu)u_1 + u_1u_2 \\ &\quad + (n-1)\{\tilde{F}_2(u_1, u_2|\mu) - \tilde{F}(u_1|\mu)u_2 - \tilde{F}(u_2|\mu)u_1 + u_1u_2\} \\ &= \tilde{F}(\min(u_1, u_2)|\mu) - \tilde{F}(u_1|\mu)\tilde{F}(u_2|\mu) \end{aligned} \quad (19)$$

$$\begin{aligned} &\quad + (n-1)\{\tilde{F}_2(u_1, u_2|\mu) - \tilde{F}(u_1|\mu)\tilde{F}(u_2|\mu)\} \\ &\quad + n\{\tilde{F}(u_1|\mu) - u_1\}\{\tilde{F}(u_2|\mu) - u_2\}. \end{aligned} \quad (20)$$

We will establish (9) by proving

$$\rho_n(u_1, u_2|\mu) \rightarrow \rho_\theta(u_1, u_2) \quad (21)$$

uniformly in u_1 and u_2 . We apply Theorem 3. Take $\bar{J} \equiv 1$, $J_1(X_1, X_2) = 1(X_1 \leq x_1)$, $J_2(X_1) = 1(X_1 \leq x_2)$ and $J_3(X_1, X_2) = 1(X_1 \leq x_1, X_2 \leq x_2)$. (The odd looking indexes in J_2 are deliberate. The algebra involved in simplifying the remainder terms is easier if we take $m = 2$ for J_3 and $m = 1$ for J_1 and J_2 .) We find from (15) applied to J_1 and J_2 that the term (20) converges to 0 uniformly in u_1 and u_2 . Applying (15) to J_1 shows that the term (19) converges, uniformly in u_1 and u_2 , to

$$\min(u_1, u_2) - u_1u_2.$$

Finally from (12), we find that

$$(n-1)\{\tilde{F}_2(u_1, u_2|\mu) - \tilde{F}(u_1|\mu)\tilde{F}(u_2|\mu)\}$$

converges to

$$A(n, J_3, \theta) - A(n, J_1, \theta)A(n, J_2, \theta)$$

uniformly in u_1, u_2 . Adopt the temporary notation $R_i = R(J_i, \theta)$ and $A_i = A(n, J_i, \theta)$ for $i = 1, 2, 3$. Then

$$n(A_3 - A_1A_2) = R_3 - R_1E_\theta(\mathbf{J}_2) - R_2E_\theta(\mathbf{J}_1) + R_1R_2/n. \quad (22)$$

From (15), we see that R_1R_2/n converges to 0 uniformly in u_1, u_2, x_1 and x_2 .

Computing we get

$$R_3 = kE_\theta(\mathbf{J}_1)E_\theta(\mathbf{J}_2) - \frac{1}{2}E_\theta(\mathbf{J}_3\mathbf{B}_2V^{-1}\mathbf{B}_2) + E_\theta(\mathbf{J}_3\psi_{1,1}(\mathbf{B}_2)),$$

$$\begin{aligned}
 R_1 E_\theta(\mathbf{J}_2) &= \frac{k}{2} E_\theta(\mathbf{J}_1) E_\theta(\mathbf{J}_2) - \frac{1}{2} E_\theta(\mathbf{J}_1 \mathbf{B}_1 V^{-1} \mathbf{B}_1) + E_\theta(\mathbf{J}_1 \psi_{1,1}(\mathbf{B}_1)), \\
 R_2 E_\theta(\mathbf{J}_1) &= \frac{k}{2} E_\theta(\mathbf{J}_1) E_\theta(\mathbf{J}_2) - \frac{1}{2} E_\theta(\mathbf{J}_2 \mathbf{B}_1 V^{-1} \mathbf{B}_1) + E_\theta(\mathbf{J}_2 \psi_{1,1}(\mathbf{B}_1)).
 \end{aligned}$$

Since \mathbf{B}_2 is a sum of two independent terms we expand the quadratic form in R_3 to see

$$\begin{aligned}
 E_\theta(\mathbf{J}_3 \mathbf{B}_2 V^{-1} \mathbf{B}_2) &= E_\theta(\mathbf{J}_1 \mathbf{B}_1 V^{-1} \mathbf{B}_1) E_\theta(\mathbf{J}_2) \\
 &\quad + E_\theta(\mathbf{J}_2 \mathbf{B}_1 V^{-1} \mathbf{B}_1) E_\theta(\mathbf{J}_1) + 2 E_\theta(\mathbf{J}_1 \mathbf{B}_1') V^{-1} E_\theta(\mathbf{B}_1 \mathbf{J}_2).
 \end{aligned}$$

We may also use the linearity of $\psi_{1,1}$ and the independence of X_1 and X_2 to see that

$$E_\theta(\mathbf{J}_3 \psi_{1,1}(\mathbf{B}_2)) = E_\theta(\mathbf{J}_1 \psi_{1,1}(\mathbf{B}_1)) E_\theta(\mathbf{J}_2) + E_\theta(\mathbf{J}_2 \psi_{1,1}(\mathbf{B}_1)) E_\theta(\mathbf{J}_1).$$

Thus, $R_3 - R_1 E_\theta(\mathbf{J}_2) - R_2 E_\theta(\mathbf{J}_1)$ simplifies to $-E_\theta(\mathbf{J}_1 \mathbf{B}_1') V^{-1} E_\theta(\mathbf{B}_1 \mathbf{J}_2)$. Since V is the Fisher information matrix in this problem, we have established (9). To check (11), we make a very similar calculation.

6.4. Verification of Condition D for the Gamma family

Here, we establish (16). Change variables via $u = \phi_2/\theta_2$ to show the integral in (16) is proportional to θ_2^r ; thus we take $\theta_2 = 1$ without loss. The integral becomes:

$$\sup_{\theta \in \Gamma} \int \int \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \frac{1}{(1 + \phi_2^2)^{r\theta_1/2}} \exp\{r\phi_1 \tan^{-1} \phi_2\} d\phi_1 d\phi_2.$$

The substitution $\phi_2 = \tan(u)$ reduces the integral to

$$\sup_{\theta \in \Gamma} \int_{-\infty}^{\infty} \int_{-\pi/2}^{\pi/2} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \cos^{r\theta_1-2}(u) \exp(r\phi_1 u) du d\phi_1.$$

We integrate separately over 4 ranges: $R_1 = \{-M \leq \phi_1 \leq M\}$, $R_2 = \{|\phi_1| > M, \phi_1 u < 0\}$, $R_3 = \{\phi_1 > M, u > 0\}$ and $R_4 = \{\phi_1 < -M, u < 0\}$. Since $|\Gamma(\theta_1 + i\phi_1)| = |\Gamma(\theta_1 - i\phi_1)|$ the integrals R_3 and R_4 are equal. Over R_1 we use the inequality

$$|\Gamma(\theta_1 + i\phi_1)/\Gamma(\theta_1)| \leq 1$$

(because the quantity inside the modulus signs is the characteristic function of $\log(X_1)$) to get the bound, for $\theta_1 \geq \varepsilon$ with $r\varepsilon > 2$

$$\begin{aligned}
 \int_{R_1} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \cos^{r\theta_1-2}(u) e^{r\phi_1 u} du d\phi_1 &\leq M \exp\{Mr\pi/2\} \int_{-\pi/2}^{\pi/2} \cos^{r\varepsilon-2}(u) du \\
 &\leq \pi M \exp\{Mr\pi/2\}.
 \end{aligned}$$

Over R_2 the term $\exp(r\phi_1 u)$ is bounded by 1. Thus,

$$\int_{R_2} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \cos^{r\theta_1-2}(u) e^{r\phi_1 u} du d\phi_1 \leq \pi \int_{-\infty}^{\infty} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r d\phi_1.$$

The integral is bounded by the supremum of the density of $\log(X_1)$ over the real line and the compact parameter set Γ .

Finally, we consider the integral over R_3 . From Section 6.1.45 of Abramowitz and Stegun [1], we find there is a constant C such that

$$\left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right| \leq C e^{-\pi\phi_1/2} \phi_1^{\theta_1-1/2}.$$

For $\theta_1 < 1/2$, $\phi_1 > M \geq 1$ and $r\varepsilon > 2$ we then get

$$\begin{aligned} \int_{R_3} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \cos^{r\theta_1-2}(u) e^{r\phi_1 u} du d\phi_1 &\leq C \int_0^{\pi/2} \int_M^{\infty} e^{-\phi_1 r(\pi/2-u)} \cos^{r\theta_1-2}(u) d\phi_1 du \\ &\leq C \int_0^{\pi/2} \frac{\cos^{r\theta_1-2}(u)}{r(\pi/2-u)} du \\ &= C \int_0^{\pi/2} \frac{\sin^{r\theta_1-2}(u)}{ru} du \\ &\leq C \int_0^{\pi/2} \frac{\sin^{r\varepsilon-2}(u)}{ru} du \\ &\leq \frac{C}{r} \int_0^{\pi/2} u^{r\varepsilon-3} du < \infty. \end{aligned}$$

For $\theta_1 \geq 1/2$ we get

$$\begin{aligned} \int_{R_3} \left| \frac{\Gamma(\theta_1 + i\phi_1)}{\Gamma(\theta_1)} \right|^r \cos^{r\theta_1-2}(u) \exp(r\phi_1 u) du d\phi_1 \\ \leq C \int_0^{\pi/2} \int_0^{\infty} e^{-\phi_1 r(\pi/2-u)} \phi_1^{r(\theta_1-1/2)} \cos^{r\theta_1-2}(u) d\phi_1 du \\ \leq C \frac{\Gamma(1+r(\theta_1-1/2))}{r^{1+r(\theta_1-1/2)}} \int_0^{\pi/2} \frac{\sin^{r\theta_1-2}(u)}{u^{r(\theta_1-1/2)+1}} du \\ \leq C \frac{\Gamma(1+r(\theta_1-1/2))}{r^{1+r(\theta_1-1/2)}} \int_0^{\pi/2} u^{r/2-3} du. \end{aligned}$$

For $r \geq 5$ the right hand side is uniformly bounded over $\Gamma \cap \{\theta_1 \geq 1/2\}$.

Acknowledgements

The author is grateful to anonymous reviewers whose advice lead to what he hopes is a much clearer paper. He also thanks Federico O'Reilly and Michael Stephens for many useful conversations on the topics discussed here. The author acknowledges grant support from the Natural Sciences and Engineering Research Council of Canada.

References

- [1] Abramowitz, M. and Stegun, I.A. (1965). *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*. New York: Dover.
- [2] Bhattacharya, R.N. and Ranga Rao, R. (1976). *Normal Approximations and Asymptotic Expansions*. Malabar, FL: R.B. Krieger.
- [3] Feller, W. (1966). *An Introduction to Probability Theory and Its Applications* 2. New York: Wiley.
- [4] Holst, L. (1981). Some conditional limit theorems in exponential families. *Ann. Probab.* **9** 818–830. [MR0628875](#)
- [5] Lehmann, E.L. and Romano, J.P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. New York: Springer. [MR2135927](#)
- [6] Lockhart, R. and O'Reilly, F. (2005). A note on Moore's conjecture. *Statist. Probab. Lett.* **74** 212–220. [MR2169379](#)
- [7] Lockhart, R.A., O'Reilly, F. and Stephens, M. (2009). Exact conditional tests and approximate bootstrap tests for the von Mises distribution. *J. Stat. Theory Pract.* **3** 543–554. [MR2750456](#)
- [8] Lockhart, R.A., O'Reilly, F.J. and Stephens, M.A. (2007). Use of the Gibbs sampler to obtain conditional tests, with applications. *Biometrika* **94** 992–998. [MR2416805](#)
- [9] Lockhart, R.A. and Stephens, M.A. (1985). Tests of fit for the von Mises distribution. *Biometrika* **72** 647–652. [MR0817579](#)
- [10] O'Reilly, F. and Gracia-Medrano, L. (2006). On the conditional distribution of goodness-of-fit tests. *Comm. Statist. Theory Methods* **35** 541–549. [MR2274070](#)
- [11] Portnoy, S. (1977). Asymptotic efficiency of minimum variance unbiased estimators. *Ann. Statist.* **5** 522–529. [MR0436434](#)
- [12] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. New York: Wiley. [MR0838963](#)
- [13] Stephens, M.A. (1962). Exact and approximate tests for directions. I. *Biometrika* **49** 463–477. [MR0159374](#)
- [14] Stephens, M.A. (1986). Tests based on EDF statistics. In *Goodness-of-fit Techniques* (R.B. D'Agostino and M.A. Stephens, eds.) Chapter 4, 97–193. New York: Marcel Dekker.
- [15] Yuan, A. and Clarke, B. (2004). Asymptotic normality of the posterior given a statistic. *Canad. J. Statist.* **32** 119–137. [MR2064396](#)

Received May 2010 and revised January 2011