Testing One Hypothesis Multiple Times: The Multidimensional Case - Supplementary Material

Sara Algeri[1] and David A. van Dyk[2]

[1] School of Statistics, University of Minnesota,

0461 Church St SE, Minneapolis, MN 55455, USA.

Email: salgeri@umn.edu

[2] Department of Mathematics, Imperial College London,

180 Queen's Gate, London SW7 2AZ, UK

In this supplementary document we assess the regularity conditions in Appendix A for our Examples 1 and 2. Below we recall the models being tested.

- **Model in Example 1:**

$$(1-\eta)\frac{1}{\lambda(\boldsymbol{\Theta})}+\eta\frac{1}{k_{\theta_1\theta_2}}\exp\left\{-0.5\left[\left(\frac{x-\theta_1}{0.5}\right)^2+\left(\frac{y-\theta_2}{0.5}\right)^2\right]\right\} \tag{1}$$

  where $\eta \in [0,1]$, $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is the location of the signal over the search region $\boldsymbol{\Theta}$, which in this case corresponds to a disc in the sky of 30° radius and centered at (195,28). Its area is given by $\lambda(\boldsymbol{\Theta})$, and $k_{\theta_1\theta_2}$ is a normalizing constant. The support of Y coincides with the search region $\boldsymbol{\Theta}$.

- **Model in Example 2:**

$$(1-\eta)\frac{e^{-y/\tau}y^{\gamma-1}}{k_{\tau\gamma}}+\eta\frac{\exp\left\{-\frac{\ln y-\mu}{2\sigma^2}\right\}}{yk_{\mu\sigma}}, \tag{2}$$

  where $\eta \in [0,1]$, $\boldsymbol{\gamma} = (\gamma, \tau)$, $\boldsymbol{\Gamma} = [1,10] \times [70,90]$, $k_{\tau\gamma}$ and $k_{\mu\sigma}$ are normalizing constants. In this setting, $\boldsymbol{\theta} = (\mu, \sigma)$ is constrained over the range $\boldsymbol{\Theta} = [1,8] \times [0.2,2]$ and we consider $y \in (0,1000]$.

It is immediate to see that assumption A0 is verified by both (1) and (2) since the densities involved in the mixture models considered belong to different parametric families (i.e., uniform and bivariate normal in Example 1, gamma and log-normal in Example 2). A0 would not necessarily be valid if the components of the mixture belonged to the same parametric family (e.g., when working with a mixture of two normal distributions). The reader is directed to Dacunha-Castelle et al. (1999), among others, for extensions of Ghosh and Sen (1985) to the case of identifiable mixture.

Conditions A1(i)-(iv) are the equivalent of the classical regularity conditions which guarantee consistency and normality of the MLE, and in our setting are required to hold for each $\boldsymbol{\theta}$ fixed. They can be easily assessed by writing explicitly the score vector and the Hessian matrix of $\log h(\boldsymbol{y}, \eta, \boldsymbol{\gamma}, \boldsymbol{\theta})$. Condition A1(v) implies stochastic equicontinuity of the Fisher information at $(0, \boldsymbol{\gamma}_0, \boldsymbol{\theta})$. For our Example 1, A1(v) naturally follows from the fact that no nuisance parameter $\boldsymbol{\gamma}$ is present under the null model, the support of η is compact and $H_{11}(\boldsymbol{\theta})$ is continuous over $[0, 1] \times \boldsymbol{\Theta}$. A similar reasoning can be done for Example 2. Specifically, since the goal here is to perform a non-nested models comparison, the hypotheses $H_0 : \eta = 1$ and $H_1 : \eta \neq 1$ are also tested and thus both $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ are required to be compact. Further, it can be shown that all $H_{jk}(\boldsymbol{\theta})$ (see the integrand functions in (7)-(12)) are continuous over $[0, 1] \times \boldsymbol{\Gamma} \times \boldsymbol{\Theta}$. Thus, for both examples, the uniform continuity of $H_{jk}(\boldsymbol{\theta})$ over the compact $[0, 1] \times \boldsymbol{\Gamma} \times \boldsymbol{\Theta}$ guarantees that for each $\boldsymbol{\theta}$ and for each $j, k = 1, \ldots, p + 1$, $H_{jk}(\boldsymbol{\theta})$ is bounded, and thus

$$\lim_{\delta \to 0} \sup_{||(\eta, \boldsymbol{\gamma}, \boldsymbol{\theta}) - (0, \boldsymbol{\gamma}_0, \boldsymbol{\theta})|| < \delta} |H_{jk}(\boldsymbol{\theta}) - H^0_{jk}(\boldsymbol{\theta})| = 0.$$

Whereas

$$E[\sup_{||(\eta,\boldsymbol{\gamma},\boldsymbol{\theta})-(0,\boldsymbol{\gamma}_0,\boldsymbol{\theta})||<\delta} |H_{jk}(\boldsymbol{\theta}) - H^0_{jk}(\boldsymbol{\theta})|] \leq E[\sup_{\delta} \sup_{||(\eta,\boldsymbol{\gamma},\boldsymbol{\theta})-(0,\boldsymbol{\gamma}_0,\boldsymbol{\theta})||<\delta} |H_{jk}(\boldsymbol{\theta}) - H^0_{jk}(\boldsymbol{\theta})|] \tag{3}$$

$$\leq 2E[\sup_{[0,1]\times\boldsymbol{\Gamma}\times\boldsymbol{\Theta}} |H_{jk}(\boldsymbol{\theta})|] \leq \infty \tag{4}$$

and thus A1(v) follows by dominated convergence.

Assumption A2 guarantees that the log-likelihood is uniquely maximized over $[0,1] \times \boldsymbol{\Gamma} \times \boldsymbol{\Theta}$. Also in this case, for both Examples 1 and 2, continuity of the log-likelihood and compactness of $[0,1] \times \boldsymbol{\Gamma} \times \boldsymbol{\Theta}$ guarantee this result. Similar considerations can be made to ensure the validity of A3. It may worth noticing that for Example 2, we consider $\gamma \in [1, 10]$ and thus the parameter space of the shape parameter of the gamma distribution is bounded away from zero. If one was to consider $\gamma \in (0, \infty)$, A2 can be verified by choosing $\mathcal{N}^c = [\epsilon_\gamma, \infty]$, for some $\epsilon_\gamma > 0$.

One of the most important assumptions in our setting is A4. For both Examples 1 and 2, continuity of $\boldsymbol{I}(\boldsymbol{\theta})$ can be easily assessed by noticing that all its elements consists of sums, multiplication and ratios of continuous functions and thus their are also continuous. Specifically, for Example 1 we have

$$I(\boldsymbol{\theta}) = \frac{\lambda(\boldsymbol{\Theta}) k'_{\theta_1\theta_2}}{k_{\theta_1\theta_2}} - 1 \tag{5}$$

where $\lambda(\boldsymbol{\Theta}) = 2827.433$, $k_{\theta_1\theta_2}$ and $k'_{\theta_1\theta_2}$ are normalizing constants and take the form

$$k_{\theta_1\theta_2} = \int_{165}^{225} \int_{28-\sqrt{(30^2-(x-195)^2)}}^{28+\sqrt{(30^2-(x-195)^2}} \exp\left\{-0.5\left[\left(\frac{x-\theta_1}{0.5}\right)^2 + \left(\frac{y-\theta_2}{0.5}\right)^2\right]\right\}$$

$$k'_{\theta_1\theta_2} = \int_{165}^{225} \int_{28-\sqrt{(30^2-(x-195)^2)}}^{28+\sqrt{(30^2-(x-195)^2}} \exp\left\{-0.5\left[\left(\frac{x-\theta_1}{\sqrt{0.125}}\right)^2 + \left(\frac{y-\theta_2}{\sqrt{0.125}}\right)^2\right]\right\}.$$

Notice that (5) is continuous and positive for any (θ_1, θ_2) in $\boldsymbol{\Theta}$ and thus A4 holds. In Example 2 the Fisher information matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{bmatrix} I_{\eta\eta}(\boldsymbol{\theta}) & I_{\eta\gamma}(\boldsymbol{\theta}) & I_{\eta\tau}(\boldsymbol{\theta}) \\ & I_{\gamma\gamma} & I_{\gamma\tau} \\ & & I_{\tau\tau} \end{bmatrix} \tag{6}$$

has elements

$$I_{\eta\eta}(\boldsymbol{\theta}) = \int \left[\frac{g(y,\mu,\sigma)}{f(y,\gamma_0,\tau_0)} - 1\right]^2 f(y,\gamma_0,\tau_0)\partial y \tag{7}$$

$$I_{\eta\gamma}(\boldsymbol{\theta}) = \int \left[\frac{g(y,\mu,\sigma)}{f(y,\gamma_0,\tau_0)} - 1\right]\left[\log y - \frac{\int e^{-y/\tau_0} y^{\gamma_0-1}\log y \partial y}{k_{\tau_0\gamma_0}}\right] f(y,\gamma_0,\tau_0)\partial y \tag{8}$$

$$I_{\eta\tau}(\boldsymbol{\theta}) = \int \left[\frac{g(y,\mu,\sigma)}{f(y,\gamma_0,\tau_0)} - 1\right]\left[\frac{y}{\tau_0^2} - \frac{\int e^{-y/\tau_0} y^{\gamma}\partial y}{\tau_0^2 k_{\tau_0\gamma_0}}\right] f(y,\gamma_0,\tau_0)\partial y \tag{9}$$

$$I_{\gamma\gamma} = \int \left[\log y - \frac{\int e^{-y/\tau_0} y^{\gamma_0-1}\log y \partial y}{k_{\tau_0\gamma_0}}\right]^2 f(y,\gamma_0,\tau_0)\partial y \tag{10}$$

$$I_{\gamma\tau} = \int \left[\log y - \frac{\int e^{-y/\tau_0} y^{\gamma_0-1}\log y \partial y}{k_{\tau_0\gamma_0}}\right]\left[\frac{y}{\tau_0^2} - \frac{\int e^{-y/\tau_0} y^{\gamma}\partial y}{\tau_0^2 k_{\tau_0\gamma_0}}\right] f(y,\gamma_0,\tau_0)\partial y \tag{11}$$

$$I_{\tau\tau} = \int \left[\frac{y}{\tau_0^2} - \frac{\int e^{-y/\tau_0} y^{\gamma}\partial y}{\tau_0^2 k_{\tau_0\gamma_0}}\right]^2 f(y,\gamma_0,\tau_0)\partial y \tag{12}$$

where $f(y,\gamma_0,\tau_0) = \frac{e^{-y/\tau_0} y^{\gamma_0-1}}{k_{\tau_0\gamma_0}}$, $k_{\tau_0\gamma_0} = \int e^{-y/\tau_0} y^{\gamma_0-1}\partial y$, $g(y,\mu,\sigma) = \frac{\exp\left\{-\frac{\ln y-\mu}{2\sigma^2}\right\}}{y k_{\mu\sigma}}$, and all the integrals are evaluated over $[0, 1000]$.

Notice that $\boldsymbol{I}(\boldsymbol{\theta})$ is positive definite uniformly over $\boldsymbol{\Theta}$, if its smallest eigenvalue is non-negative for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. We assess this with a numerical check. Specifically, we compute $\boldsymbol{I}(\boldsymbol{\theta})$ on a grid of 12851 values (μ, σ) over $[1, 8] \times [0.2, 2]$. For each of the 12851 points $\boldsymbol{I}(\boldsymbol{\theta})$ is evaluated at the MLE estimates of γ_0, τ_0 (i.e. $\hat{\gamma}_0 = 2.753$, $\hat{\tau}_0 = 83.379$)

and thus assuming $f(y, \hat{\gamma}_0, \hat{\tau}_0)$ as the true model. Finally, we compute the eigenvalues of each of the 12851 matrices obtained. Below we report a the summary of the distribution of the smallest eigenvalues.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01349 0.38326 0.41697 0.38100 0.43248 0.43604
```

Since the minimum is achieved at 0.01349, we expect A4 to hold for our Example 2 with $\epsilon \approx 0.0134$, and consequently $\boldsymbol{I}(\boldsymbol{\theta})$ is positive-definite for each $\boldsymbol{\theta}$ over the range $[1, 8] \times [0.2, 2]$.

Finally, A5 guarantees tightness of the score random field, $\{S_1^0(\boldsymbol{\theta})\}$, with components $S_1^0(\boldsymbol{\theta}) = \frac{g(\boldsymbol{y},\boldsymbol{\theta})}{f(\boldsymbol{y},\boldsymbol{\gamma}_0)} - 1$, for each value of $\boldsymbol{\theta}$. In both our examples the parameter space $[0, 1] \times \boldsymbol{\Gamma} \times \boldsymbol{\Theta}$ is compact and so is the support of $\boldsymbol{y}$. Furthermore, the derivatives of $\{S_1^0(\boldsymbol{\theta})\}$ are continuous and thus they are bounded for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. For instance, in Example 1, we have that

$$\left|\frac{\partial}{\partial\theta_1} g(x, y, \boldsymbol{\theta})\right| = \left|\frac{4(x - E[x])}{k_{\theta_1\theta_2}} \exp\left\{-0.5\left[\left(\frac{x-\theta_1}{0.5}\right)^2 + \left(\frac{y-\theta_2}{0.5}\right)^2\right]\right\}\right| \leq K_1 < \infty,$$

and similarly $\left|\frac{\partial}{\partial\theta_2} g(x, y, \boldsymbol{\theta})\right| \leq K_2 < \infty$, for some $K_1, K_2 > 0$, and for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$. In this setting $f(y, \boldsymbol{\gamma}_0) = \frac{1}{\lambda(\boldsymbol{\Theta})}$, and thus it is constant over $\boldsymbol{\Theta}$. Hence A5 follows by the mean value theorem, and for $\xi = 1$ and $\lambda = 1$. A numerical check has also been conducted and A5 is verified for $K > 50$. The same approach can be used to validate A5 for Example 2 which can be shown to hold for $\xi = 1$ and $\lambda = 1$. Also in this case, a numerical check has been conducted and A5 is verified for $K > 40$.

References

Dacunha-Castelle, D., Gassiat, E., et al. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary arma processes. *The Annals of Statistics*, 27(4):1178–1209.

Ghosh, J. and Sen, P. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer (L. LeCam and R. A. Olshen, eds.) 2 789–806. Wadsworth, Monterey, CA.*