Appendices of the paper entitled

Testing for a Global Maximum of the Likelihood

APPENDIX A: PROOFS OF BOTH THEOREMS

A.1 Proof of Theorem 1

This proof is strongly inspired from Theorem 1 of GJ99's paper.

Suppose that $\hat{\theta}_n \xrightarrow{P} \theta_0$. By condition (a), we have $P(\hat{\theta}_n \in \Theta) \to 1$. Then, using condition (c), Taylor expansion gives:

$$\phi_2(\hat{\theta}_n) = \phi_2(\theta_0) + \nabla\phi_2(\theta_0)'(\hat{\theta}_n - \theta_0) + o_P(|\hat{\theta}_n - \theta_0|). \tag{A.1}$$

When condition (d) is verified, the strong law of large number (SLLN) allows to obtain $\phi_2(\theta_0) \xrightarrow{a.s.} d_2(\theta_0)$ and $\nabla\phi_2(\theta_0) \xrightarrow{a.s.} \nabla d_2(\theta_0)$. We deduce that $\nabla\phi_2(\theta_0)'(\hat{\theta}_n - \theta_0) = o_P(1)$, therefore (5) holds.

Now consider that both $P(\hat{\theta}_n \in \Theta) \to 1$ and (5) hold. We suppose that $\hat{\theta}_n \xrightarrow{P} \theta_0$ is false. So, there are $\delta_0 > 0$ and $\varepsilon_0 > 0$ such that, for all n,

$$P(|\hat{\theta}_n - \theta_0)| \geq \delta_0) \geq \varepsilon_0. \tag{A.2}$$

We define $\Theta_1 = \{\theta \in \Theta : |\theta - \theta_0| \geq \delta_0\}$. Since $d(\cdot) = E_t\phi(\cdot)$ is continuous (Proposition 2 in Appendix C) and $\theta_0 \notin \Theta_1$, we have by (4) that $\rho = \inf_{\theta\in\Theta_1} |d(\theta) - d(\theta_0)| > 0$. Let $\eta = \rho/[4(B_1 \vee B_2)]$. Then, there are an integer m and points $\theta_1, \theta_2, \ldots, \theta_m \in \Theta_1$ such that for any $\theta \in \Theta_1$ there is $1 \leq l \leq m$ such that $|\theta_l - \theta| < \eta$. Suppose that $\hat{\theta}_n \in \Theta$ and $|\hat{\theta}_n - \theta_0| \geq \delta_0$. Then $\hat{\theta}_n \in \Theta_1$ and hence there is $\tilde{\theta} \in \{\theta_l, 1 \leq l \leq m\}$ such that $|\tilde{\theta} - \hat{\theta}_n| < \eta$. Now, we seek a lower bound of $|\phi_2(\hat{\theta}_n) - d_2(\theta_0)|$:

$$\begin{aligned} |\phi_2(\hat{\theta}_n) - d_2(\theta_0)| &= |\phi(\hat{\theta}_n) - d(\theta_0)| \quad (\text{since } \phi_1(\hat{\theta}_n) = 0 \text{ and } d_1(\theta_0) = 0) && \text{(A.3)} \\ &\geq |\phi(\tilde{\theta}) - d(\theta_0)| - |\phi(\tilde{\theta}) - \phi(\hat{\theta}_n)| && \text{(A.4)} \\ &\geq \min_{1\leq l\leq m} |\phi(\theta_l) - d(\theta_0)| - |\phi_1(\tilde{\theta}) - \phi_1(\hat{\theta}_n)| - |\phi_2(\tilde{\theta}) - \phi_2(\hat{\theta}_n)|. && \text{(A.5)} \end{aligned}$$

In addition, we have

$$\min_{1\leq l\leq m} |\phi(\theta_l) - d(\theta_0)| \geq \min_{1\leq l\leq m} |d(\theta_l) - d(\theta_0)| - \max_{1\leq l\leq m} |\phi(\theta_l) - d(\theta_l)|, \tag{A.6}$$

where

$$\begin{aligned}
\min_{1\le l\le m}|d(\theta_l)-d(\theta_0)| &= \frac{1}{2}\min_{1\le l\le m}|d(\theta_l)-d(\theta_0)|+\frac{1}{2}\min_{1\le l\le m}|d(\theta_l)-d(\theta_0)| && \text{(A.7)}\\
&\ge \frac{\rho}{2}+\frac{\rho}{2} && \text{(A.8)}\\
&\ge \frac{\rho}{2}+\frac{\rho}{2}\frac{B_1+B_2}{2(B_1\vee B_2)} && \text{(A.9)}\\
&= \frac{\rho}{2}+\eta(B_1+B_2). && \text{(A.10)}
\end{aligned}$$

Moreover, Taylor inequality and the fact that $|\tilde{\theta}-\hat{\theta}_n|<\eta$ give (with $j=1,2$)

$$\begin{aligned}
|\phi_j(\tilde{\theta})-\phi_j(\hat{\theta}_n)| &\le |\tilde{\theta}-\hat{\theta}_n|\Big(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}|\nabla\varphi_j(X;\theta)|\Big) && \text{(A.11)}\\
&\le \eta\Big(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}|\nabla\varphi_j(X;\theta)|\Big). && \text{(A.12)}
\end{aligned}$$

Combining Equations (A.5), (A.6), (A.10) and (A.12), we obtain:

$$|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\frac{\rho}{2}-\max_{1\le l\le m}|\phi(\theta_l)-d(\theta_l)|-\eta\sum_{j=1}^{2}\Big(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}|\nabla\varphi_j(X;\theta)|-B_j\Big). \quad \text{(A.13)}$$

Then, condition (d) and the SLLN give (recall also that $d(\cdot)=E_t\phi(\cdot)$):

$$|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\frac{\rho}{2}-o_P(1). \quad \text{(A.14)}$$

Therefore, with the same $o_P(1)$, we have

$$\begin{aligned}
P(|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\rho/4) &\ge P(|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\rho/2-o_P(1),|o_P(1)|\le\rho/4) && \text{(A.15)}\\
&\ge P(|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\rho/2-o_P(1))-P(|o_P(1)|>\rho/4), && \text{(A.16)}
\end{aligned}$$

since for any events A and B, we know that $P(A,B)\ge P(A)-P(B^c)$, B^c being the complementary event of B. Moreover, Inequality (A.14) being the consequence of both $\hat{\theta}_n\in\Theta$ and $|\hat{\theta}_n-\theta_0|\ge\delta_0$, and for all events $A\subset B\Rightarrow P(A)\le P(B)$, we deduce that

$$\begin{aligned}
P(|\phi_2(\hat{\theta}_n)-d_2(\theta_0)|\ge\rho/4) &\ge P(\hat{\theta}_n\in\Theta,|\hat{\theta}_n-\theta_0|\ge\delta_0)-P(|o_P(1)|>\rho/4) && \text{(A.17)}\\
&\ge P(|\hat{\theta}_n-\theta_0|\ge\delta_0)-P(\hat{\theta}_n\notin\Theta)-P(|o_P(1)|>\rho/4) && \text{(A.18)}\\
&\ge \varepsilon_0-o(1)\to\varepsilon_0 && \text{(A.19)}
\end{aligned}$$

as $n\to\infty$, which contradicts (5). Therefore, we must have $\hat{\theta}_n\xrightarrow{P}\theta_0$.

A.2 Proof of Theorem 2

Using conditions (d) and (e), the central limit theorem leads to

$$\phi_2(\theta_0) \xrightarrow{D} N\left(d_2(\theta_0), \frac{Var_t \ln f(X;\theta_0)}{n}\right). \tag{A.20}$$

Suppose now that $\hat{\theta}_n \xrightarrow{P} \theta_0$. Since we proved in Theorem 1 that $\nabla\phi_2(\theta_0)'(\hat{\theta}_n - \theta_0) = o_P(1)$, then Taylor expansion (A.1) and Slutsky's theorem allow to conclude that (6) holds.

APPENDIX B: SELECTING AN ESTIMATOR OF THE VARIANCE

The problem is to study, in the case of a correct model, behaviour of two possible estimators of the variance $v(\theta_0) = Var_{\theta_0} \ln f(X;\theta_0)$ which are $v(\hat{\theta}_n) = Var_{\hat{\theta}_n} \ln f(X;\hat{\theta}_n)$ and $V_n(\hat{\theta}_n) = \sum_{i=1}^n (\ln f(X_i;\hat{\theta}_n) - \ell(\hat{\theta}_n)/n)^2/n$. First, we show consistency of both. Then, we compare them through the mean squared error criterion and also performance of the test in the context of the simple mixture case (Section 3.1).

B.1 Consistency of both estimators

Proposition 1. Suppose that conditions (a), (b), (c), (e) and (f) are satisfied.
If $\hat{\theta}_n \xrightarrow{P} \theta_0$, then $v(\hat{\theta}_n) \xrightarrow{P} v(\theta_0)$ and $V_n(\hat{\theta}_n) \xrightarrow{P} V_n(\theta_0)$.

Proof First, using condition (c), Taylor expansion of $v(\hat{\theta}_n)$ gives

$$v(\hat{\theta}_n) = v(\theta_0) + \nabla v(\theta_0)'(\hat{\theta}_n - \theta_0) + o_P(|\hat{\theta}_n - \theta_0|). \tag{A.21}$$

Then, condition (f) and convergence $\hat{\theta}_n \xrightarrow{P} \theta_0$ involve that $v(\hat{\theta}_n) \xrightarrow{P} v(\theta_0)$.

In the same manner, Taylor expansion of $V_n(\hat{\theta}_n)$ is

$$V_n(\hat{\theta}_n) = V_n(\theta_0) + \nabla V_n(\theta_0)'(\hat{\theta}_n - \theta_0) + o_P(|\hat{\theta}_n - \theta_0|). \tag{A.22}$$

By SLLN, conditions (e) and (f) been verified, we have $V_n(\theta_0) \xrightarrow{a.s.} v(\theta_0)$ and $\nabla V_n(\theta_0) \xrightarrow{a.s.} \nabla v(\theta_0)$. Consequently, $V_n(\hat{\theta}_n) \xrightarrow{P} v(\theta_0)$.

B.2 Empirical comparison of both estimators

We propose to compare by simulation the two candidates with (i) the mean squared criterion (mse) and (ii) their performance on the test. Remember that mse of an estimator H for a real value h is defined by

$$\text{mse}_h[H] = Var_h[H] + (E_h[H] - h)^2. \tag{A.23}$$

We consider the simple mixture case (Section 3.1) with the same experimental conditions . Figure 1 of the appendices displays both mse (in the case where $\hat{\theta}_n$ is the global maximizer) and the power of the test at significance level $\alpha = 0.05$. Clearly, mse is lower for $v(\hat{\theta}_n)$ and it leads also to a better power for small sample sizes.

[Figure 1 about here.]

APPENDIX C: OTHER PROPOSITIONS AND PROOFS

Proposition 2. If conditions (a)-(d) are satisfied, then $E_t\phi(\cdot)$ is continuous.

Proof For any sequence of parameters (θ_n) in Θ such that $\theta_n \to \theta$ as $n \to \infty$ with $\theta \in \Theta$, consider the corresponding sequence of functions $(\varphi(x;\theta_n))$. From condition (c), $\varphi(x;\cdot)$ is continuous and, so, $\varphi(x;\theta_n) \to \varphi(x;\theta)$ pointwise. Moreover, condition (d) gives

$$E_t \sup_n |\varphi(X;\theta_n)| \le E_t \sup_n |\varphi_1(X;\theta_n)| + E_t \sup_n |\varphi_2(X;\theta_n)| < \infty. \tag{A.24}$$

The dominated convergence theorem implies that $E_t\varphi(X;\theta_n) \to E_t\varphi(X;\theta)$. Recall that $E_t\varphi(X;\cdot) = E_t\phi(\cdot)$ to conclude that $E_t\phi(\cdot)$ is continuous for any $\theta \in \Theta$.

Proposition 3. If X has normal density $\psi(x;\mu,\sigma^2)$ with mean μ and variance σ^2 then

$$E_{\mu,\sigma^2} \ln\psi(X;\mu,\sigma^2) = -\frac{1}{2}\ln\sigma^2 - \frac{1}{2} - \frac{1}{2}\ln(2\pi) \tag{A.25}$$

and

$$Var_{\mu,\sigma^2} \ln\psi(X;\mu,\sigma^2) = \frac{1}{2}. \tag{A.26}$$

Proof From $\ln\psi(X;\mu,\sigma^2) = -\ln(\sigma^2)/2 - (X-\mu)^2/(2\sigma^2) - \ln(2\pi)/2$, (A.25) is obvious and (A.26) is proven in the following way:

$$\begin{aligned} Var_{\mu,\sigma^2} \ln\psi(X;\mu,\sigma^2) &= \frac{Var_{\mu,\sigma^2}(X-\mu)^2}{4\sigma^4} && \text{(A.27)} \\ &= \frac{E_{\mu,\sigma^2}(X-\mu)^4 - E^2_{\mu,\sigma^2}(X-\mu)^2}{4\sigma^4} && \text{(A.28)} \\ &= \frac{3\sigma^4 - \sigma^4}{4\sigma^4} = \frac{1}{2}. && \text{(A.29)} \end{aligned}$$

Proposition 4. Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \psi(x; \theta_0, \theta_0^2)$, $\bar{X} = \sum_{i=1}^n X_i/n$ and with $\bar{\chi} = \sum_{i=1}^n X_i^2/n$. If $\hat{\theta}_n$ is the consistent solution of the likelihood equation, then

$$Var_{\theta_0}\phi_2(\hat{\theta}_n) = Var_{\theta_0}\left\{\left[-1+\sqrt{1+4\frac{\bar{\chi}}{\bar{X}^2}}\right]^{-1}\right\}. \tag{A.30}$$

Proof Direct calculus shows that

$$\frac{\ell(\theta)}{n} = -\frac{1}{2}\ln\theta^2 + \frac{\bar{X}}{\theta} - \frac{\bar{\chi}}{2\theta^2} - \frac{1}{2} - \frac{1}{2}\ln(2\pi) \tag{A.31}$$

and so, using (A.25) in Proposition 3, we obtain

$$\phi_2(\hat{\theta}_n) = \frac{\bar{X}}{\hat{\theta}_n} - \frac{\bar{\chi}}{2\hat{\theta}_n^2}. \tag{A.32}$$

Maximizing $\ell(\theta)$ leads to two roots

$$\hat{\theta}_n^{(1)} = \frac{1}{2}\left(-\bar{X}+\sqrt{\bar{X}^2+4\bar{\chi}}\right) \quad \text{and} \quad \hat{\theta}_n^{(2)} = \frac{1}{2}\left(-\bar{X}-\sqrt{\bar{X}^2+4\bar{\chi}}\right), \tag{A.33}$$

but only $\hat{\theta}_n^{(1)}$ is consistent since it is easy to see that $\hat{\theta}_n^{(1)} \xrightarrow{P} \theta_0$ whereas $\hat{\theta}_n^{(2)} \xrightarrow{P} -2\theta_0$. Then, after some algebra, conclusion follows.

Proposition 5. Consider the normal density $f(x;\theta) = \psi(x;\mu,\sigma^2)$ where $\theta = (\mu, \sigma^2)$. Denoting by $\hat{\theta}_n$ the (unique) solution of the likelihood equation, we have $\phi_2(\hat{\theta}_n) = 0$.

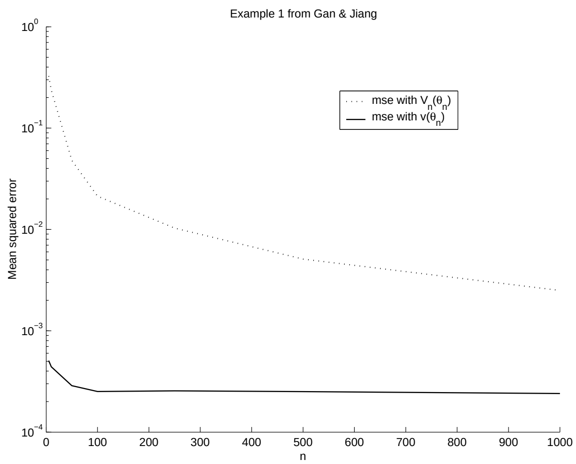
Proof Since $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ with $\hat{\mu}_n = \sum_{i=1}^n X_i/n$ and $\hat{\sigma}_n^2 = \sum_{i=1}^n (X_i - \hat{\mu}_n)^2/n$, we obtain

$$\begin{aligned}\frac{\ell(\hat{\theta}_n)}{n} &= -\frac{1}{2}\ln\hat{\sigma}_n^2 - \frac{1}{2n}\sum_{i=1}^n \frac{(X_i-\hat{\mu}_n)^2}{\sigma_n^2} - \frac{1}{2}\ln(2\pi) && \text{(A.34)}\\ &= -\frac{1}{2}\ln\hat{\sigma}_n^2 - \frac{1}{2} - \frac{1}{2}\ln(2\pi). && \text{(A.35)}\end{aligned}$$

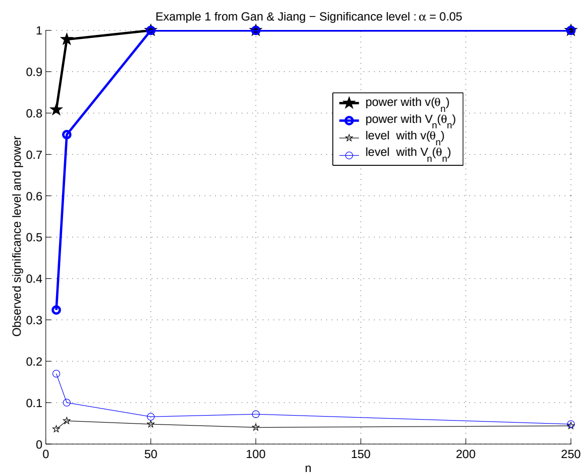
Then, if $Y \sim f(x;\hat{\theta}_n)$, Equation (A.25) in Proposition 3 allows to write

$$E_{\hat{\theta}_n}\ln f(Y;\hat{\theta}_n) = -\frac{1}{2}\ln\hat{\sigma}_n^2 - \frac{1}{2} - \frac{1}{2}\ln(2\pi). \tag{A.36}$$

We conclude by noting that, by definition, $\phi_2(\hat{\theta}_n) = \ell(\hat{\theta}_n)/n - E_{\hat{\theta}_n}\ln f(Y;\hat{\theta}_n)$.



(a)



(b)

Figure 1. Comparison of estimators $v(\hat{\theta}_n)$ and $V_n(\hat{\theta}_n)$ in the simple mixture case respectively with: (a) mse criterion and (b) power of the test.