Parametric estimation through divergences. Duality approach:good and bad points

M.Broniatowski, LSTA, Université Paris 6. Joint work and reflections with colleagues in Paris, Bucarest and in Prague

October 14, 2009

For any signed finite measure Q in $\mathcal{M}(P)$, the ϕ-divergence between Q and P is defined by

$$D_\phi(Q,P) := \int_{\mathcal{X}} \phi\left(\frac{dQ}{dP}(x)\right)\, dP(x). \qquad (1)$$

When Q is not a.c. w.r.t. P, we set $D_\phi(Q,P) = +\infty$. The ϕ-divergences were introduced by I.Csiszar as "f-divergences". For all p.m. P, the mappings $Q \in \mathcal{M} \mapsto D_\phi(Q,P)$ are convex and take nonnegative values. When $Q = P$ then $D_\phi(Q,P) = 0$. Furthermore, if the function $x \mapsto \phi(x)$ is strictly convex on a neighborhood of $x = 1$, then the following fundamental property holds

$$D_\phi(Q,P) = 0 \text{ if and only if } Q = P. \qquad (2)$$

All these properties are presented in Csiszar (1967), Liese and Vajda (1987) for ϕ-divergences defined on the set of all p.m.'s $\mathcal{M}^1$. When the ϕ-divergences are defined on $\mathcal{M}$, then the same properties hold. In general $D_\phi(Q,P)$ and $D_\phi(P,Q)$ are not equal. Hence, ϕ-divergences usually are not distances, but they merely measure some difference between two measures. A main feature of divergences between distributions of random variables X and

Y is the invariance property with respect to common smooth change of variables.

0.1 Examples of ϕ-divergences.

When defined on $\mathcal{M}^1$, the Kullback-Leibler (KL),

$$\phi(x) = x \log x - x + 1,$$

modified Kullback-Leibler (KL_m)

$$\phi(x) = -\log x + x - 1$$

, χ^2

$$\phi(x) = \frac{1}{2}(x-1)^2$$

, modified χ^2 (χ^2_m)

$$\phi(x) = \frac{1}{2}(x-1)^2/x$$

, Hellinger (H)

$$\phi(x) = 2(\sqrt{x}-1)^2$$

, and L_1

$$\phi(x) = |x-1|$$

power divergences" (Rényi, Cressie Read). They are defined through the class of convex functions

$$x \in]0,+\infty[\mapsto \phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)} \quad (3)$$

The χ^2-divergence: the corresponding ϕ function $\phi_2(x) := \frac{1}{2}(x-1)^2$ is defined and convex on whole $\mathbb{R}$.

Estimation and test using ϕ-divergences. An i.i.d. sample $X_1, \ldots, X_n$ with common unknown distribution

P is observed and some p.m. Q is given. We aim to estimate $D_\phi(Q,P)$ and, more generally, $\inf_{Q\in\Omega} D_\phi(Q,P)$ where Ω is some set of measures, as well as the measure Q^* achieving the infimum on Ω. In the parametric context, these problems can be well defined and lead to new results in estimation and tests, extending classical notions.

0.1.1 Parametric estimation and tests

Let $\{P_\theta; \theta \in \Theta\}$ be some parametric model with Θ an open set in $\mathbb{R}^d$. On the basis of an i.i.d. sample $X_1, \ldots, X_n$ with distribution P_{θ_T}, we want to estimate θ_T, the unknown true value of the parameter and perform statistical tests on the parameter using ϕ-divergences. When all p.m.'s P_θ share the same finite support S,

$$\widetilde{\theta}_\phi := \arg\inf_{\theta\in\Theta} D_\phi(P_\theta, P_n). \tag{4}$$

Various parametric tests can be performed based on the previous estimates of ϕ-divergences; Lindsay 1994, Morales

Pardo Vajda 1995nd The class of estimates (4) contains the maximum likelihood estimate (MLE). Indeed, when $\phi(x) = \phi_0(x) = -\log x + x - 1$, we obtain

$$\widetilde{\theta}_{KL_m} := \arg \inf_{\theta\in\Theta} KL_m(P_\theta, P_n) = \arg \inf_{\theta\in\Theta} \sum_{j\in S} -\log(P_\theta(j))P_n$$

Robustness and divergences? Lindsay 199, Basu-Lindsay 1994 Jimenez Shao (2001)

When the support S is continuous, the plug-in estimates are not well defined;

Smmothing (kernel) .example: Hellinger distance: Beran (1977). However: does not include MLE.

1 Dual representation for ϕ-divergences

Liese and Vajda 2006, Broniatowski and Kéziou 2006, on different proofs. Here the French approach

. The Fenchel-Legendre transform of ϕ will be denoted ϕ^*, i.e.,

$$t \in \mathbb{R} \mapsto \phi^*(t) := \sup_{x \in \mathbb{R}} \{tx - \phi(x)\},$$

By the closedness of ϕ, applying the duality principle, the conjugate ϕ^{**} of ϕ^* coincides with ϕ, i.e.,

$$\phi^{**}(t) := \sup_{x \in \mathbb{R}} \{tx - \phi^*(x)\} = \phi(t), \quad \text{for all } t \in \mathbb{R}.$$

If additionally ϕ is strictly convex, then for all $t \in \text{Im}\phi'$ we have

$$\phi^*(t) = t{\phi'}^{-1}(t) - \phi\left({\phi'}^{-1}(t)\right) \quad \text{ and } \quad {\phi^*}'(t) = {\phi'}^{-1}(t).$$

.

Let $\mathcal{F}$ be some class of $\mathcal{B}$-measurable real valued functions f defined on $\mathcal{X}$, and denote $\mathcal{M}_{\mathcal{F}}$, the real vector subspace of $\mathcal{M}$, defined by

$$\mathcal{M}_{\mathcal{F}} := \left\{ Q \in \mathcal{M} \text{ such that } \int |f| \; d|Q| < \infty, \text{ for all } f \in \mathcal{F} \right\}.$$

Theorem 1 *Assume that ϕ is differentiable. Then, for all $Q \in \mathcal{M}_{\mathcal{F}}$ such that $D_{\phi}(Q, P)$ is finite and $\phi'\left(\frac{dQ}{dP}\right)$ belongs to $\mathcal{F}$, the ϕ-divergence $D_{\phi}(Q, P)$ admits the dual representation*

$$D_{\phi}(Q, P) = \sup_{f \in \mathcal{F}} \left\{ \int f \; dQ - \int \phi^{*}(f) \; dP \right\}, \qquad (5)$$

and the function $f := \phi'\left(\frac{dQ}{dP}\right)$ is an optimal solution. Furthermore, if ϕ is essentially smooth then $f := \phi'(dQ/dP)$ is the unique optimal solution (P-a.e.).

Any class F containing $f := \phi'(dQ/dP)$ and such that $\int |f| \, d|Q| < \infty$, for all $f \in F$ holds can be used.

2 Parametric estimation and tests through minimum ϕ-divergence approach

For a given **escort parameter** $\theta \in \Theta$, consider the class of functions **(the smallest when θ_T is unknown and under the model)**

$$\mathcal{F} = \mathcal{F}_\theta := \left\{ x \mapsto \phi'\left(\frac{p_\theta(x)}{p_\alpha(x)}\right);\ \alpha \in \Theta \right\}.$$

Then

$$D_\phi(\theta, \theta_T) = \sup_{f \in \mathcal{F}_\theta} \left\{ \int f\ dP_\theta - \int \phi^*(f)\ dP_{\theta_T} \right\},$$

which, by (**??**), can be written as

$$D_\phi(\theta, \theta_T) = \sup_{\alpha \in \Theta} \left\{ \begin{array}{c} \int \phi'\left(\frac{p_\theta}{p_\alpha}\right)\ dP_\theta \\ -\int \left[\frac{p_\theta}{p_\alpha}\phi'\left(\frac{p_\theta}{p_\alpha}\right) - \phi\left(\frac{p_\theta}{p_\alpha}\right)\right]\ dP_{\theta_T} \end{array} \right\}. \tag{6}$$

The supremum in this display is unique and it is achieved at $\alpha = \theta_T$ independently upon the value of θ. Hence, it is reasonable to estimate $D_\phi(\theta, \theta_T) := \int \phi(p_\theta / p_{\theta_T})\ dP_{\theta_T}$, the ϕ-divergence between P_θ and P_{θ_T}, by

$$\widehat{D_\phi}(\theta, \theta_T) := \sup_{\alpha \in \Theta} \left\{ \begin{array}{c} \int \phi'\left(\frac{p_\theta}{p_\alpha}\right)\ dP_\theta \\ - \int \left[\frac{p_\theta}{p_\alpha} \phi'\left(\frac{p_\theta}{p_\alpha}\right) - \phi\left(\frac{p_\theta}{p_\alpha}\right)\right]\ dP_n \end{array} \right\}, \tag{7}$$

in which we have replaced P_{θ_T} by its estimate P_n, the empirical measure associated to the data.

Since the supremum in (6) is unique and it is achieved at $\alpha = \theta_T$, define the following class of

M-estimates of θ_T

$$\widehat{\alpha}_\phi(\theta) := \arg \sup_{\alpha \in \Theta} \left\{ \begin{array}{c} \int \phi'\left(\frac{p_\theta}{p_\alpha}\right)\ dP_\theta \\ - \int \left[\frac{p_\theta}{p_\alpha} \phi'\left(\frac{p_\theta}{p_\alpha}\right) - \phi\left(\frac{p_\theta}{p_\alpha}\right)\right]\ dP_n \end{array} \right\} \tag{8}$$

. Further, we have **(natural estimator deduced from the previous one, minimize on the escort**)

$$\inf_{\theta \in \Theta} D_\phi(\theta, \theta_T) = D_\phi(\theta_T, \theta_T) = 0.$$

The infimum in this display is unique and it is achieved at $\theta = \theta_T$. It follows that a natural definition of a second estimate is

$$\widehat{\theta}_\phi := \arg \inf_{\theta \in \Theta} \sup_{\alpha \in \Theta} \left\{ \begin{array}{c} \int \phi' \left(\frac{p_\theta}{p_\alpha} \right) \, dP_\theta \\ - \int \left[\frac{p_\theta}{p_\alpha} \phi' \left(\frac{p_\theta}{p_\alpha} \right) - \phi \left(\frac{p_\theta}{p_\alpha} \right) \right] \, dP_n \end{array} \right\}. \tag{9}$$

3 Everything is OK under the model with any escort parameter, and solves old problems

Theorem 2 *Under standard conditions of integrability and domination (exchange $\int$ and $\frac{d}{d\theta}$)*

(a) *Let* $B(\theta_T, n^{-1/3}) := \left\{\alpha \in \Theta;\ \|\alpha - \theta_T\| \leq n^{-1/3}\right\}$*. Then, as* $n \to \infty$*, with probability one, the estimate* $\widehat{\alpha}_\phi(\theta)$ *is* $n^{1/3}$*-consistent and satisfies* $P_n(\partial/\partial\alpha)h(\theta, \widehat{\alpha}_\phi(\theta)) = 0$ *(is a regular M-estimator).*

(b) $\sqrt{n}\left(\widehat{\alpha}_\phi(\theta) - \theta_T\right)$ *converges in distribution to a centered multivariate normal random variable with covariance matrix*

$$V_\phi(\theta, \theta_T) = S^{-1}MS^{-1} \tag{10}$$

with $S := -P_{\theta_T}(\partial^2/\partial\alpha^2)h(\theta, \theta_T)$ *and* $M := P_{\theta_T}(\partial/\partial\alpha)h(\theta, \theta$
If $\theta_T = \theta$*, then* $V_\phi(\theta, \theta_T) = V(\theta_T) = I_{\theta_T}^{-1}$ *with* I_{θ_T} *the Fisher Information matrix .*

(c) *If* $\theta_T = \theta$*, then the statistic* $\frac{2n}{\phi''(1)}\widehat{D_\phi}(\theta, \theta_T)$ *converges in distribution to a* χ^2 *random variable with* d *degrees of freedom.*

(d) *when* $\theta \neq \theta_T$*, we have* $\sqrt{n}\left(\widehat{D_\phi}(\theta, \theta_T) - D_\phi(\theta, \theta_T)\right)$ *converges in distribution* *to a centered normal random variable with variance*

$$\sigma_\phi^2(\theta,\theta_T) = P_{\theta_T} h(\theta,\theta_T)^2 - \left(P_{\theta_T} h(\theta,\theta_T)\right)^2.$$

Remark 3 *Using theorem* **??** *part (c), the estimate* $\widehat{D_\phi}(\theta_0,\theta_T)$ *can be used to perform statistical tests (asymptotically of level* ϵ*) of the null hypothesis* $\mathcal{H}_0 : \theta_T = \theta_0$ *against the alternative* $\mathcal{H}_1 : \theta_T \neq \theta_0$ *for a given value* θ_0*. Since* $D_\phi(\theta_0,\theta_T)$ *is nonnegative and takes value zero only when* $\theta_T = \theta_0$*, the tests are defined through the critical region*

$$C_\phi(\theta_0,\theta_T) := \left\{ \frac{2n}{\phi''(1)} \widehat{D_\phi}(\theta_0,\theta_T) > q_{d,\epsilon} \right\} \qquad (11)$$

where $q_{d,\epsilon}$ *is the* $(1-\epsilon)$*-quantile of the* χ^2 *distribution with* d *degrees of freedom. Note that these tests are all consistent, since* $\widehat{D_\phi}(\theta_0,\theta_T)$ *are* n*-consistent estimates of* $D_\phi(\theta_0,\theta_T) = 0$ *under* $\mathcal{H}_0$*, and* $\sqrt{n}$*-consistent estimate of* $D_\phi(\theta_0,\theta_T) > 0$ *under* $\mathcal{H}_1$*; see part (c) and (d) in theorem* **??** *above. Further, the asymptotic result (d) in theorem* **??** *above can be used to give approximation of* *the power function* $\theta_T \mapsto \beta(\theta_T) := P_{\theta_T}\left(C_\phi(\theta_0, \theta_T)\right)$. *We obtain then the following approximation*

$$\beta(\theta_T) \approx 1 - F_{\mathcal{N}}\left(\frac{\sqrt{n}}{\sigma_\phi(\theta_0, \theta_T)}\left[\frac{\phi''(1)}{2n} q_{d,\epsilon} - D_\phi(\theta_0, \theta_T)\right]\right) \tag{12}$$

where $F_{\mathcal{N}}$ *is the cumulative distribution function of a normal random variable with mean zero and variance one. An important application of this approximation is the approximate sample size (13) below that ensures a power* β *for a given alternative* $\theta_T \neq \theta_0$. *Let* n_0 *be the positive root of the equation*

$$\beta = 1 - F_{\mathcal{N}}\left(\frac{\sqrt{n}}{\sigma_\phi(\theta_0, \theta_T)}\left[\frac{\phi''(1)}{2n} q_{d,\epsilon} - D_\phi(\theta_0, \theta_T)\right]\right)$$

i.e., $n_0 = \frac{(a+b) - \sqrt{a(a+2b)}}{2 D_\phi(\theta_0, \theta_T)^2}$ *where* $a = \sigma_\phi^2(\theta_0, \theta_T)\left[F_{\mathcal{N}}^{-1}(1-\beta)\right]$ *and* $b = \phi''(1) q_{d,\epsilon} D_\phi(\theta_0, \theta_T)$. *The required sample size is then*

$$n^* = [n_0] + 1 \tag{13}$$

where [.] *is used here to denote "integer part of".*

3.1 A simple solution to the problem of testing the number of components in a mixture

Consider the following set of signed finite measures

$$p_\theta := (1-\theta)p_0 + \theta p_1 \text{ where } \theta \in R. \qquad (14)$$

This set (of signed finite measures with mass one) obviously contains the mixture model (**??**). In particular, the null value of θ_T (i.e., $\theta_T = 0$) is an **interior point** of the parameter space R. The likelihood ratio test (for a model of signed measures) cannot be used since the log-likelihood $l(\theta)$ may be infinite (when $\theta < 0$ or $\theta > 1$). In the context of divergences, this means that the estimate $\widehat{KL_m}(P_0, P_{\theta_T})$ may be infinite if we consider the model (14), which is due to the fact that the corresponding convex function $\phi(x) = -\log x + x - 1$ is infinite on $\mathbb{R}_-$. This suggests to use a divergence associated to a convex function ϕ which is finite on all $\mathbb{R}$, for instance, the

χ^2-divergence (which is associated to the convex function $\phi(x) = \frac{1}{2}(x-1)^2$). So, in order to perform a test asymptotically of level ϵ for (**??**), we propose to use the following estimate of the χ^2-divergence between P_0 and P_{θ_T}

$$\widetilde{\chi^2}(0,\theta_T) = \sup_{\alpha\in\Theta_e} \left\{ P_0 f(0,\alpha) - P_n g(0,\alpha) \right\}, \quad (15)$$

where $f(0,\alpha) = p_0/p_\alpha - 1$ and $g(0,\alpha) = 1/2(p_0/p_\alpha + 1)(p_0/p_\alpha - 1)$ as a consequence of definitions (**??**) and (**??**), and Θ_e is the new parameter space which we define as follows

$$\Theta_e := \left\{ \alpha \in R \text{ such that } \int |f(0,\alpha)| \ dP_0 \text{ is finite} \right\}.$$

The value of the parameter θ_T under the null hypothesis $\mathcal{H}_0$, i.e., $\theta_T = 0$, is in the interior of the new parameter space Θ_e which is generally non void. Hence, under conditions of theorem **??** where Θ is replaced by Θ_e and θ by zero, under $\mathcal{H}_0$ the statistic $2n\widetilde{\chi^2}(0,\theta_T)$ converges in distribution to a χ^2 random variable with one degree of freedom; the critical region takes then the form

$$CR := \left\{ 2n\widetilde{\chi^2}(0,\theta_T) > q_{1,\epsilon} \right\},$$

where $q_{1,\epsilon}$ is the $(1-\epsilon)$-quantile of the χ^2 distribution with one degree of freedom. Obviously other divergences which are associated to convex functions finite on all $\mathbb{R}$ can be used.

4 Semiparametrics

We consider a **nonlinear regression problem**, that is the observations $Y_i \in \mathbb{R}$, are given by

$$Y_i = r(\theta_T, X_i) + \epsilon_i \quad i = 1, \ldots, n \tag{16}$$

with $r(\theta, X_i)$ the regression function known up to some **finite dimensional parameter** $\theta \in \Theta$, where Θ is a subset of $\mathbb{R}^p$ with θ_T is the true parameter value.where $Y_i \in \mathbb{R}$, $X_i \in \mathfrak{X} \subset \mathbb{R}$ are observable, $\epsilon_i \in \mathbb{R}$ is unobservable. The samples $X_i, \quad i = 1 \cdots n$ are i.i.d. with common distribution G and density function g, ϵ_i, i.i.d $\sim f$

independent of X_i, where **the densities g and f are unknown**.

Let $p_{\theta_T}(x,y) = f(y - r(\theta_T, x))g(x)$ denote the true joint distribution of (X,Y). Under the assumptions of the regression model (16), the ϕ-divergence between a parametric model $\Big\{p_\theta(x,y) := f(y-r(\theta,x))g(x);\ \theta \in \Theta\Big\}$ and p_{θ_T} can be written as

$$D_\phi(\theta, \theta_T) = \sup_{\alpha \in \Theta} P_{\theta_T} h(\theta, \alpha), \tag{17}$$

where θ is the escort parameter and

$$h(\theta, \alpha, x, y) := \int \phi'\left(\frac{f(y - r(\theta,x))}{f(y - r(\alpha,x))}\right) f(y - r(\theta,x))g(x) - \left[\begin{array}{c} \frac{f(y-r(\theta,x))}{f(y-r(\alpha,x))}\phi'\left(\frac{f(y-r(\theta,x))}{f(y-r(\alpha,x))}\right) \\ -\phi\left(\frac{f(y-r(\theta,x))}{f(y-r(\alpha,x))}\right) \end{array}\right].$$

Note that g cancels. Furthermore, the supremum in this display is unique and it is achieved at $\alpha = \theta_T$ independently upon the value of θ. Hence, it is reasonable to estimate $D_\phi(\theta, \theta_T) := \int \phi(p_\theta / p_{\theta_T}) \mathrm{d}P_{\theta_T}$, the

ϕ-divergence between P_θ and P_{θ_T}, by

$$\widehat{D_\phi}(\theta, \theta_T) := \sup_{\alpha \in \Theta} P_n h(\theta, \alpha), \tag{18}$$

in which we have replaced P_{θ_T} by its estimate P_n, the empirical measure associated to the data.

4.1 Construction of the estimators

Use a two-step method . We split the data sample $(Y_1, X_1) \ldots, (Y_n, X_n)$ into two parts

, $(Y_1, X_1) \ldots, (Y_N, X_N)$

and $(Y_{N+1}, X_{N+1}), \ldots, (Y_n, X_n)$, with $N \to \infty$ as $n \to \infty$ adaptive estimation (f. i.Bickel, Manski)

Initial estimate $\widehat{\theta}_n$ such that $|\widehat{\theta}_n - \theta_T| = O(n^{-1/2})$, **"estimate the residuals"** $\widetilde{\epsilon}_{iN} = Y_i - r(\widehat{\theta}_n, X_i)$ $i =$

$1, \cdots, N$. These residuals will be used to **construct a nonparametric density function estimator, say f_N.**

Introduce the following functions

$$m(\theta, \alpha) = \int \phi'\left(\frac{f(y - r(\theta, x))}{f(y - r(\alpha, x))}\right) f(y - r(\theta, x)) g(x) \mathrm{d}x \mathrm{d}y. \tag{19}$$

and

$$\begin{aligned} \varphi(\theta, \alpha, x, y) &= \frac{f(y - r(\theta, x))}{f(y - r(\alpha, x))} \phi'\left(\frac{f(y - r(\theta, x))}{f(y - r(\alpha, x))}\right) \qquad (20) \\ &\quad - \phi\left(\frac{f(y - r(\theta, x))}{f(y - r(\alpha, x))}\right). \qquad (21) \end{aligned}$$

That is the function h defined in (**??**), will be written as

$$h(\theta, \alpha, x, y) = m(\theta, \alpha) - \varphi(\theta, \alpha, x, y).$$

Since the functions f and g are unknown, we will give an estimate of the function h based on the first sample of the residuals. For given θ and α, we can estimate h by

$$\widehat{h}(\theta, \alpha, x, y) = \widehat{m}(\theta, \alpha) - \widehat{\varphi}(\theta, \alpha, x, y) \tag{22}$$

with f_N. in place of f

Now, for any θ in Θ, the divergence $D_\phi(\theta, \theta_T)$, can be estimated by

$$\widetilde{D_\phi}(\theta, \theta_T) := \sup_{\alpha \in \Theta} P_{\bar{n}} \widehat{h}(\theta, \alpha), \tag{23}$$

where $P_{\bar{n}}$ is the empirical distribution based on the second part of the sample $(Y_{N+1}, X_{N+1}), \ldots, (Y_n, X_n)$.

In analogy with the preceding section we will define estimates of the parameter θ_T based on the estimated divergence $\widetilde{D_\phi}(\theta, \theta_T)$. It follows that

$$\widetilde{\alpha}_\phi(\theta) := \arg \sup_{\alpha \in \Theta} P_{\bar{n}} \widehat{h}(\theta, \alpha) \tag{24}$$

and

$$\widetilde{\theta}_\phi := \arg \inf_{\theta \in \Theta} \sup_{\alpha \in \Theta} P_{\bar{n}} \widehat{h}(\theta, \alpha) \tag{25}$$

5 Consistency and asymptotic distribution

(C.1) the estimate $\widetilde{\alpha}_\phi(\theta)$ exists;

(C.2) for any positive δ, there exists some positive η such that for all $\alpha \in \Theta$ satisfying $\|\alpha - \theta_T\| > \delta$ we have

$$P_{\theta_T} h(\theta, \alpha) < P_{\theta_T} h(\theta, \theta_T) - \eta$$

(1) Under the assumptions (C.1), (H.0), (F.1-4), (R.1-8), (G.1-4), (K.0-2), the estimate $\widetilde{D_\phi}(\theta, \theta_T)$ converges in probability to $D_\phi(\theta, \theta_T)$.

(2) Assume that the assumptions (C.1-2), (H.0), (F.1-4), (R.1-8), (G.1-4), (K.0-2) hold then the estimate $\widetilde{\alpha}_\phi(\theta)$ converges in probability to θ_T.

The conditions under which the Proposition 5 are standards. Indeed conditions (C.1)-(C.3) are those used by [**?**] to prove consistency of $\widehat{\alpha}_\phi(\theta)$ and $\widehat{D_\phi}(\theta, \theta_T)$; Conditions (K.0)-(K.2) ensures the convergence of the kernel density estimates based on the residuals to the true disturbance density.

When all conditions (H.0-2), (F.1-6), (R.1-8), (G.1-4), (K.0-2) and (S.1) hold. Then, we have

(a) $\sqrt{n}\left(\widetilde{\alpha}_{\phi}(\theta)-\theta_T\right)$ converges in distribution to a centered multivariate normal random variable with covariance matrix

$$V_{\phi}(\theta,\theta_T)=S^{-1}MS^{-1} \tag{26}$$

If $\theta_T=\theta$, then $V_{\phi}(\theta,\theta_T)=I_{\theta_T}^{-1}$.

(b) If $\theta_T=\theta$, then the statistic $\frac{2n}{\phi''(1)}\widetilde{D_{\phi}}(\theta,\theta_T)$ converges in distribution to a χ^2 random variable with d degrees of freedom.

(c) If additionally assumption (S.2) holds, then when $\theta \neq \theta_T$, we have $\sqrt{n}\left(\widetilde{D_{\phi}}(\theta,\theta_T)-D_{\phi}(\theta,\theta_T)\right)$ converges in distribution to a centered normal random variable with variance

$$\sigma_{\phi}^2(\theta,\theta_T)=P_{\theta_T}h(\theta,\theta_T)^2-\left(P_{\theta_T}h(\theta,\theta_T)\right)^2. \tag{27}$$

5.0.1 Bias reduction through adaptive choice of the escort parameter

The above results prove that both estimates $\widetilde{D_\phi}(\theta, \theta_T)$ and $\widetilde{\alpha}_\phi(\theta)$ are asymptotically unbiased for all value of the escort parameter θ.

For small values of n these estimates have a significant bias.

A natural improvement of the present approach results in the plugging of an initial estimate of θ_T as an adaptive escort choice.

Let $\widehat{\theta}_{NLS}$ be the nonlinear least square estimate (NLS) of θ_T, the estimate of θ_T is easy to calculate. Therefore it converges to θ_T under the general assumptions of this paper.

By the continuity of of the derivatives of $h(\theta, \alpha)$ w.r.t α, adaptive procedure satisfies the following asymptotic result

When all conditions (H.0-2), (F.1-6), (R.1-8), (G.1-4), (K.0-2) and (S.1) hold and $\theta = \widehat{\theta}_{NLS}$. Then, we have

(a) $\sqrt{n}\left(\widetilde{\alpha}_{\phi}(\theta) - \theta_T\right)$ converges in distribution to a centered multivariate normal random variable with covariance matrix $I_{\theta_T}^{-1}$.

(b) The statistic $\frac{2n}{\phi''(1)}\widetilde{D_{\phi}}(\theta, \theta_T)$ converges in distribution to a χ^2 random variable with p degrees of freedom.