Étude à distance finie d'un R-estimateur pour la matrice de forme des distributions symétriques elliptiques complexes

Stefano Fortunati [1,2], *Alexandre Renaux* [1], *Frédéric Pascal* [1]

[1]Université Paris-Saclay, CNRS, CentraleSupeléc, L2S,

[2]DR2I-IPSA.

GRETSI 2022
Nancy, 06-09 Septembre 2022.

Modèles semi-paramétriques

- Un modèle semi-paramétrique[1] $\mathcal{P}_{\boldsymbol{\theta},h}$ est un ensemble de densités de probabilité (dp) caractérisées par un paramètre de dimension fini $\boldsymbol{\theta} \in \Theta$ ainsi que d'une *fonction* $h \in \mathcal{L}$ qui est un paramètre de dimension infini :

$$\mathcal{P}_{\boldsymbol{\theta},g} \triangleq \left\{ p_X(\mathbf{x}_1, \ldots, \mathbf{x}_M | \boldsymbol{\theta}, h), \boldsymbol{\theta} \in \Theta \subseteq \mathbb{C}^q, h \in \mathcal{L} \right\}.$$

- Généralement, $\boldsymbol{\theta}$ est le paramètre à estimer alors que h est une fonction de *nuisance*.

- **Avantage**: Un modèle semi-paramétrique combine la *concision* d'un modèle paramétrique avec la *généralité* d'un modéle non-paramétrique.

- **Inconvénient**: Des outils d'analyse fonctionnelle sont nécessaires.

[1] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

Exemples: données manquantes.

- Soit $\mathbf{z} \triangleq (\mathbf{x}^T, \mathbf{y}^T)^T$ un ensemble *complet* de données, où:
  - $\mathbf{x}$ est l'ensemble de données observées (disponibles).
  - $\mathbf{y}$ est l'ensemble de données non observable (manquantes).

- **Problème**: Estimer $\boldsymbol{\theta} \in \Theta$ à partir des données observées $\mathbf{x}$ lorsque la dp p_Y des donnèes manquantes $\mathbf{y}$ est inconnue.

- La dp p_X des données observées peut être exprimée comme :

$$p_X(\mathbf{x}|\boldsymbol{\theta}) = \int_{\mathcal{Y}} p_{X,Y}(\mathbf{x}, \mathbf{y}|\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathcal{Y}} p_{X|Y}(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}) p_Y(\mathbf{y}) d\mathbf{y}.$$

- L'ensemble des dp de $\mathbf{x}$ est un *modèle semi-paramétrique* :

$$\mathcal{P}_{\boldsymbol{\theta}, p_Y} \triangleq \{p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, p_Y), \boldsymbol{\theta} \in \Theta, p_Y \in \mathcal{K}\}.$$

Exemples: régression non-linéaire

- Considérons le modèle générale de régression non-linéaire :

$$\mathbf{x} = f(\mathbf{z}, \boldsymbol{\theta}) + \boldsymbol{\epsilon},$$

  - $\boldsymbol{\theta} \in \Theta$: vecteur de paramètres à estimer,
  - $f \in \mathcal{F}$: fonction éventuellement inconnue,
  - $\mathbf{z}$: vecteur aléatoire de dp $p_Z \in \mathcal{K}$ éventuellement inconnue,
  - $\boldsymbol{\epsilon}$: bruit de dp $p_\epsilon \in \mathcal{E}$ éventuellement inconnue.

- L'ensemble de dp de $\mathbf{x}$ est un *modèle semi-paramétrique* :

$$\mathcal{P}_{\boldsymbol{\theta}, f, p_Z, p_\epsilon} \triangleq \left\{ p_X(\mathbf{x}|\boldsymbol{\theta}, f, p_Z, p_\epsilon), \boldsymbol{\theta} \in \Theta, f \in \mathcal{F}, p_Z \in \mathcal{K}, p_\epsilon \in \mathcal{E} \right\}.$$

- Si on a une connaissance "a priori", nous pouvons réduire les inconnues.

Exemples: processus autorégressifs

- Considérons le processus $\mathrm{AR}(p)$:

$$x_n = \sum_{i=1}^{p} \theta_i x_{n-i} + w_n, \quad n \in \mathbb{Z}$$

  - $\boldsymbol{\theta} \triangleq [\theta_1, \ldots, \theta_p]$: vecteur de paramètres à estimer,
  - w_n: i.i.d. innovations de dp $p_w \in \mathcal{W}$ inconnue,

- Soit $\mathbf{x} \in \mathbb{R}^N$ un vecteur de N observations issus d'un $\mathrm{AR}(p)$.

- L'ensemble des dp de $\mathbf{x}$ est un *modèle semi-paramétrique* :

$$\mathcal{P}_{\boldsymbol{\theta},p_w} \triangleq \left\{p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, p_w), \boldsymbol{\theta} \in \Theta, p_w \in \mathcal{W}\right\}.$$

Distributions CES

- Un vecteur symétrique elliptique complexe $\mathbf{x} \in \mathbb{C}^N$ a une dp :

$$p_X(\mathbf{x}) = |\boldsymbol{\Sigma}|^{-1} h((\mathbf{x}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})) \triangleq CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h).$$ [2]

  - $\mathcal{L} \ni h : \mathbb{R}^+ \to \mathbb{R}^+$ est le fonction dite *génératrice*,
  - $\boldsymbol{\mu} \in \mathbb{C}^N$ est un paramètre de position,
  - $\boldsymbol{\Sigma} \in \mathcal{M}_N$ est une matrice de dispersion.

- Notez que $\boldsymbol{\Sigma}$ et h ne sont pas identifiables conjointement:

$$CES_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}, h(t)) \equiv CES_N(\boldsymbol{\mu}, c\boldsymbol{\Sigma}, h(ct)), \ \forall c > 0.$$

- Pour éviter cela, on définit la *matrice de forme* :

$$\mathbf{V}_1 \triangleq \boldsymbol{\Sigma}/[\boldsymbol{\Sigma}]_{1,1}.$$

[2] E. Ollila, D. E. Tyler, V. Koivunen and H. V. Poor, "Complex Elliptically Symmetric Distributions: Survey, New Results and Applications", *IEEE Trans. on Signal Processing*, vol. 60, no. 11, pp. 5597-5625, Nov. 2012.

Estimation semi-paramétrique de $\mathbf{V}_1$

- Sans perte de généralité, on suppose que $\boldsymbol{\mu} \equiv \mathbf{0}$.

- L'ensemble des dp CES est un *modèle semi-paramétrique* :

$$\mathcal{P}_{\boldsymbol{\theta},h} = \left\{ p_X | p_X(\mathbf{x}|\boldsymbol{\theta}, h) = |\mathbf{V}_1|^{-1} h(\mathbf{x}^H \mathbf{V}_1^{-1} \mathbf{x}); \boldsymbol{\theta} \in \Theta, h \in \mathcal{L} \right\},$$

- h joue le rôle d'un paramètre de nuisance *de dimension infini*.

- Le paramètre d'intérêt à estimer est : [3]

$$\boldsymbol{\theta} \triangleq \underline{\mathrm{vec}}(\mathbf{V}_1) \in \Theta.$$

- Notez que l'estimateur du *maximum de vraisemblance* **n'est pas une option** puisque $h \in \mathcal{L}$ est inconnue !

[3] L'operateur $\underline{\mathrm{vec}}(\mathbf{A})$ définit le vecteur de dimension $N^2 - 1$ obtenu à partir de $\mathrm{vec}\,(\mathbf{A})$ en supprimant son premier élément, c'est-à-dire $\mathrm{vec}\,(\mathbf{A}) \triangleq [a_{11}, \underline{\mathrm{vec}}(\mathbf{A})^T]^T$.

En présence d'une fonction h inconnue

- Matrice de covariance empirique (ou *SCM*):
  - Non robuste,
  - Manque d'efficacité semi-paramétrique,

- M-estimateurs (*Maronna, Tyler*): [4,5]
  - Robustes,
  - Manque d'efficacité semi-paramétrique,

- R (rank-based)-estimateurs (*Hallin, Paindaveine*): [6,7]
  - Robustes,
  - Efficacité semi-paramétrique.

---

[4] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *The Annals of Statistics*, vol. 4, no. 1, pp. 51–67, 01 1976.

[5] D. E. Tyler, "A distribution-free M-estimator of multivariate scatter," *The Annals of Statistics*, vol. 15, no. 1, pp. 234–251, 1987.

[6] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal R-estimation of shape," *The Annals of Statistics*, vol. 34, no. 6, pp. 2757–2789, 2006.

[7] M. Hallin and B. J. M. Werker, "Semi-parametric efficiency, distribution-freeness and invariance," *Bernoulli*, vol. 9, no. 1, pp. 137–165, 2003.

Un R-estimateur efficace pour $\mathbf{V}_1$ (1/3)

- Soit $\{\mathbf{x}_l\}_{l=1}^{L}$ un ensemble de L observations i.i.d..

- Un R-estimateur peut être représenté comme : [8]

$$\hat{\boldsymbol{\theta}}_R = \hat{\boldsymbol{\theta}}^\star + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}_{\hat{\boldsymbol{\theta}}^\star}^{-1}\widetilde{\boldsymbol{\Delta}}_{\hat{\boldsymbol{\theta}}^\star}.$$

  - $\hat{\boldsymbol{\theta}}^\star$ est un *estimateur préliminaire* (consistent, non efficace) de $\boldsymbol{\theta} \triangleq \underline{\mathrm{vec}}(\mathbf{V}_1)$, par exemple la SCM or l'estimateur de Tyler,
  - $\widehat{\boldsymbol{\Upsilon}}_{\hat{\boldsymbol{\theta}}^\star}$ est un estimateur de la *matrice d'information semi-paramétrique*,
  - $\widetilde{\boldsymbol{\Delta}}_{\boldsymbol{\theta}^\star} \triangleq \sum_{l=1}^{L} \widetilde{\varphi}(\mathbf{x}_l, \hat{\boldsymbol{\theta}}^\star)$, où $\widetilde{\varphi}$ est une approximation du *vecteur de score semi-paramétrique*.

- L'estimation de $h \in \mathcal{L}$ **n'est pas nécessaire** !

[8] M. Hallin, H. Oja, and D. Paindaveine, "Semiparametrically efficient rank-based inference for shape II. optimal R-estimation of shape,"

Un R-estimateur efficace pour $\mathbf{V}_1$ (2/3)

La forme explicite du R-estimateur de la matrice de forme est :

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}_1^*) + \frac{1}{L\hat{\alpha}}\left[\mathbf{L}_{\widehat{\mathbf{V}}_1^*}\mathbf{L}_{\widehat{\mathbf{V}}_1^*}^H\right]^{-1}\mathbf{L}_{\widehat{\mathbf{V}}_1^*} \times$$
$$\times \sum_{l=1}^{L} K_h\left(\frac{r_l^\star}{L+1}\right)\mathrm{vec}(\hat{\mathbf{u}}_l^\star(\hat{\mathbf{u}}_l^\star)^H),$$

- $\{r_l^\star\}_{l=1}^L$: rangs de $\hat{Q}_l^\star \triangleq \mathbf{x}_l^H[\widehat{\mathbf{V}}_1^*]^{-1}\mathbf{x}_l$,
- $\hat{\mathbf{u}}_l^\star \triangleq (\hat{Q}_l^\star)^{-1/2}[\widehat{\mathbf{V}}_1^*]^{-1/2}\mathbf{z}_l$,
- $\hat{\alpha}$ est une quantité dépendant des données,
- $\widehat{\mathbf{V}}_1^*$ est un estimateur préliminaire: paramètre libre,
- $K_h(\cdot)$ est une fonction *score*: paramètre libre.

Un R-estimateur efficace (3/3)

- Dans nos récents travaux [9,10], on a :
  - Généralisé le R-estimateur au cas complexe,
  - Étudier le réglage optimal de paramètres,
  - Étudier la robustesse aux données aberrantes,

  en utilisant la version matricielle suivante :

$$\widehat{\mathbf{V}}_{1,R} = \widehat{\mathbf{V}}_1^\star + \frac{1}{\hat{\alpha}} \left(\mathbf{W} - [\mathbf{W}]_{1,1} \widehat{\mathbf{V}}_1^\star \right)$$

  où $\mathbf{W}$ est une matrice dont l'expression explicite est connue.

- Quel est l'impact des deux paramètres libres
  1. l'estimateur préliminaire $\widehat{\mathbf{V}}_1^\star$,
  2. la fonction score $K_h(\cdot)$

  sur les performances à distance finie?

[9] S. Fortunati, A. Renaux, F. Pascal, "Robust semiparametric efficient estimators in complex elliptically symmetric distributions", *IEEE Transactions on Signal Processing*, vol. 68, pp. 5003-5015, 2020.

[10] S. Fortunati, A. Renaux, F. Pascal "Joint Estimation of Location and Scatter in Complex Elliptical Distributions: A robust semiparametric and computationally efficient R-estimator of the shape matrix," *MLSP Special Issue of the Journal of Signal Processing Systems*, 2021.

Données de test: t-distribution complexe

- Les données sont générées en utilisant une t-distribution complexe à moyenne nulle :

$$p_X(\mathbf{x}|\mathbf{V}_1, h) = |\mathbf{V}_1|^{-1} h\left(\mathbf{x}^H \mathbf{V}_1^{-1} \mathbf{x}\right),$$

$$h(t) = \frac{\Gamma(\lambda + N)}{\pi^N \Gamma(\lambda)} \left(\frac{\lambda}{\eta}\right)^{\lambda} \left(\frac{\lambda}{\eta} + t\right)^{-(\lambda+N)},$$

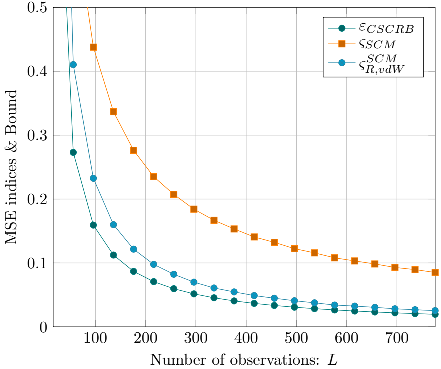
  - $\lambda \in]1, +\infty[$ est un paramètre de forme qui contrôle la queue de la distribution,
  - $\eta > 0$ est un paramètre d'échelle qui rend compte de la puissance des données.

Remark:
Grâce à la nature semi-paramétrique du R-estimateur considéré, les résultats obtenus pour des données t-distribuées sont valables pour toutes les autres distributions CES.

SCM comme estimateur préliminaire

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}_{1,R}^{SCM}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}^{\star}_{1,SCM}) + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}_{\widehat{\mathbf{V}}^{\star}_{1,SCM}}\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}^{\star}_{1,SCM}},$$

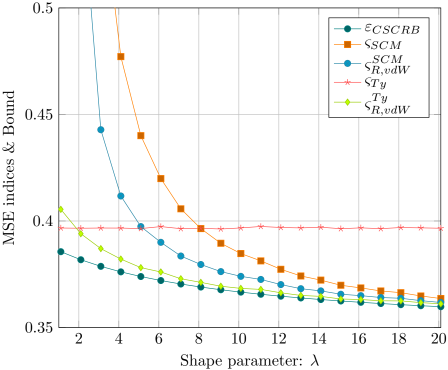
$$\varsigma_{\gamma}^{\varphi} \triangleq ||E\{\mathrm{vec}(\widehat{\mathbf{V}}_{\gamma}^{\varphi} - \mathbf{V}_0)\mathrm{vec}(\widehat{\mathbf{V}}_{\gamma}^{\varphi} - \mathbf{V}_0)^H\}||_F,$$



- Le terme de correction linéaire $L^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}_{\widehat{\mathbf{V}}^{\star}_{1,SCM}}\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}^{\star}_{1,SCM}}$ peut améliorer de manière significative l'efficacité de la SCM ($\lambda = 2$ et $N = 8$).

Estim. Tyler comme estimateur préliminaire

$$\underline{\mathrm{vec}}(\widehat{\mathbf{V}}^{Ty}_{1,R}) = \underline{\mathrm{vec}}(\widehat{\mathbf{V}}^{\star}_{1,Ty}) + L^{-1/2}\widehat{\boldsymbol{\Upsilon}}^{-1}_{\widehat{\mathbf{V}}^{\star}_{1,Ty}}\widetilde{\boldsymbol{\Delta}}_{\widehat{\mathbf{V}}^{\star}_{1,Ty}},$$



- $\widehat{\mathbf{V}}^{Ty}_{1,R}$ est plus performant que $\widehat{\mathbf{V}}^{SCM}_{1,R}$ pour de données à codes lourdes ($L = 5N$ et $N = 8$).

Deux exemples utiles de fonctions score

- La fonction *score* de *van der Waerden* :

$$K_{vdW}(u) \triangleq \Phi_G^{-1}(u), \quad u \in]0,1[,$$

où Φ_G^{-1} indique la fonction inverse de la distribution cumulative (dc) d'une variable aléatoire Gamma avec paramètres $(N, 1)$.

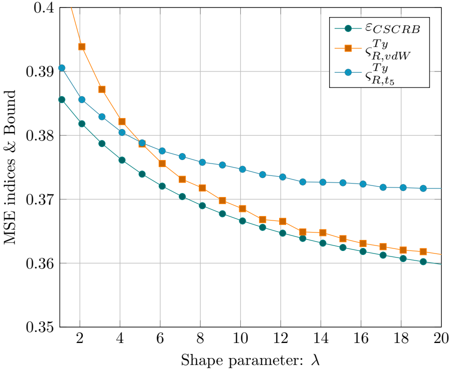
- La fonction t_ν-*score* :

$$K_{t_\nu}(u) = \frac{N(2N+\nu)F_{2N,\nu}^{-1}(u)}{\nu + 2NF_{2N,\nu}^{-1}(u)}, \quad u \in]0,1[,$$

où $F_{2N,\nu}(u)$ indique la dc de Fisher avec $2N$ et $\nu \in]0,\infty[$ degrés de liberté.

- La fonction t_ν-*score* dépend d'un paramètre supplémentaire ν qui doit être choisi avec soin.

van der Waerden et fonction t_ν-score



- La fonction t_ν*-score* fournit les meilleures performances pour les données à queue très lourde ($1 < \lambda < 5$).
- La fonction de *van der Waerden* offre de bonnes performances sans avoir besoin de paramètres de réglage supplémentaires.

Conclusions

- Une étude à distance finie d'un R-estimateur pour la matrice de forme des distributions symétriques elliptiques complexes a été développé.

- L'étude numérique proposée a montré que les paramètres suivant :
  - Estimateur préliminaire: *estimateur de Tyler*,
  - Function *score*: *van der Waerden*,

  conduisent à de bonne performances en termes d'erreur quadratique moyenne (EQM) pour quasiment toutes les valeurs de λ.

- Les travaux futurs porteront sur l'application de l'estimateur proposé aux problèmes de clustering robuste et de distance learning.

Laboratoire Signaux & Systèmes

Essayez-le et faites-nous savoir s'il fonctionne dans vos applications ;)

Le code est disponible sur GitHub.

Merci de votre attention !

Des questions?

Backup slides

A lower bound in semiparametric estimation

- Let $\{\mathbf{x}_l\}_{l=1}^{L}$ be a set of iid observations, such that $\mathbf{x}_l \sim p_0(\mathbf{x};\boldsymbol{\theta}_0, h_0) \in \mathcal{P}_{\boldsymbol{\theta},h}\ \forall l$.

- The class of *Regular and Asymptotically Linear (RAL) estimators* is defined as:
  1. $\sqrt{L}$-consistent: $\sqrt{L}\left(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_0\right) = O_P(1)$,
  2. Asymptotically normal: $\sqrt{L}(\hat{\boldsymbol{\theta}}_L - \boldsymbol{\theta}_0) \underset{L\to\infty}{\overset{d}{\sim}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Xi}(\boldsymbol{\theta}_0, h_0))$.

- The ML and all the robust estimators belong to this class.

Semiparametric CRB (SCRB): Any *RAL* estimator $\hat{\boldsymbol{\theta}}_L$ of $\boldsymbol{\theta}_0$, derived in $\mathcal{P}_{\boldsymbol{\theta},h}$ from $\{\mathbf{x}_l \sim\}_{l=1}^{L}$ iid observations, satisfies: [11]

$$\boldsymbol{\Xi}(\boldsymbol{\theta}_0, h_0) \geq \bar{\mathbf{I}}(\boldsymbol{\theta}_0|h_0)^{-1} \triangleq \mathrm{SCRB}(\boldsymbol{\theta}_0|h_0).$$

[11] P.J. Bickel, C.A.J Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, 1993.

Ranks (1/2)

- Let $\{x_l\}_{l=1}^{L}$ be a set of L continuous i.i.d. random variables with pdf p_X.

- Define the vector of the *order statistics* as

$$\mathbf{v}_X \triangleq [x_{L(1)}, x_{L(2)}, \ldots, x_{L(L)}]^T,$$

whose entries

$$x_{L(1)} < x_{L(2)} < \cdots < x_{L(L)}$$

are the values of $\{x_l\}_{l=1}^{L}$ ordered in an ascending way.[12]

- The rank $r_l \in \mathbb{N}$ of x_l is the position index of x_l in $\mathbf{v}_X$.

[12] Note that, since $x_l, \forall l$ are continuous random variable the equality occurs with probability 0.

Ranks (2/2)

- Let $\mathbf{r}_X \triangleq [r_1, \ldots, r_L]^T \in \mathbb{N}^L$ be the vector collecting the ranks.
- Let $\mathcal{K}$ be the family of score functions $K : (0,1) \to \mathbb{R}$ that are continuous, square integrable and that can be expressed as the difference of two monotone increasing functions.

Let $\{x_l\}_{l=1}^{L}$ be a set of i.i.d. random variables s.t. $x_l \sim p_X, \forall l$. Then, we have:

1. The vectors $\mathbf{v}_X$ and $\mathbf{r}_X$ are independent,
2. Regardless the actual pdf p_X, the rank vector $\mathbf{r}_X$ is uniformly distributed on the set of all $L!$ permutations on $\{1, 2, \ldots, L\}$,
3. For each $l = 1, \ldots, L$, $K\left(\frac{r_l}{L+1}\right) = K(u_l) + o_P(1)$, where $K \in \mathcal{K}$ and $u_l \sim \mathcal{U}[0,1]$ is a random variable uniformly distributed in $(0,1)$.

Tyler's M-estimator

- Tyler's M-estimator $\widehat{\mathbf{V}}_{Ty}$ is given by the convergence point ($k \to \infty$) of:

$$\begin{cases} \widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)} = \frac{N}{L} \sum_{l=1}^{L} [\hat{Q}_l^{(k)}]^{-1} \mathbf{x}_l \mathbf{x}_l^H \\ \widehat{\mathbf{V}}_{Ty}^{(k+1)} \triangleq N \widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)} / \mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{Ty}^{(k+1)}). \end{cases}$$

  where:

$$\hat{Q}_l^{(k)} = \mathbf{x}_l^H [\widehat{\mathbf{V}}_1^{(k)}]^{-1} \mathbf{x}_l,$$

- The estimator $\widehat{\mathbf{V}}_{Ty}$ is $\sqrt{L}$-consistent under any (unknown) density generator $h \in \mathcal{H}$.

- **It is not semiparametric efficient**.