Associated kernel discriminant analysis for multivariate discrete data

Sobom Matthieu SOMÉ[1], C.C. KOKONENDJI[2] & Ibrahim MONA[3]

[1]Université Ouaga 2, Burkina Faso

[2]Université de Franche-Comté
Laboratoire de Mathématiques de Besançon - UMR 6623 CNRS-UFC

[2]University of Surrey, Guildford, United Kingdom

LmB Conferences "Multivariate Count Analysis"

Besançon : 04-06 July 2018

Outline

1. Introduction
2. Associated kernels for discriminant analysis
3. Simulations and application
4. Concluding remarks

Problem: Classify $\mathbf{x} \in \mathbb{T}_d (\subseteq \mathbb{R}^d)$ into one of J predefined classes ($d > 1$)

Optimal Bayes rule

Allocate $\mathbf{x}$ (target) to group j_0 where $j_0 = \arg \max_{j \in \{1,\dots,J\}} \pi_j f_j(\mathbf{x})$,

π_j : prior probabilities; f_j: probability "density" functions with support $\mathbb{T}_d$.

Problem: Classify $\mathbf{x} \in \mathbb{T}_d (\subseteq \mathbb{R}^d)$ into one of J predefined classes ($d > 1$)

Optimal Bayes rule

Allocate $\mathbf{x}$ (target) to group j_0 where $j_0 = \arg \max_{j \in \{1,\dots,J\}} \pi_j f_j(\mathbf{x})$,

π_j : prior probabilities; f_j: probability "density" functions with support $\mathbb{T}_d$.

Examples of $\mathbb{T}_d$:

$$\mathbb{T}_{k_1}^{[1]} \times \cdots \times \mathbb{T}_{k_L}^{[L]} \text{ with } \sum_{\ell=1}^{L} k_\ell = d; \qquad \mathbb{T}_{k_\ell}^{[\ell]} = [0,1]^\ell \, , \mathbb{N}^\ell \, , \mathbb{Z}^\ell \; [0,\infty)^\ell \, .$$

Let $\mathcal{X}_j = \{\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j}\}$ a sample of random vectors on $\mathbb{T}_d (\subseteq \mathbb{R}^d)$; $j = 1, \ldots, J$.

- Classical kernel estimator of f_j:

$$\widehat{f_j}(\mathbf{x}; \mathcal{K}, \mathbf{H}_j) = \frac{1}{n_j \det \mathbf{H}_j} \sum_{i=1}^{n_j} \mathcal{K}\left\{\mathbf{H}_j^{-1}(\mathbf{x} - \mathbf{X}_{ji})\right\}, \quad \forall \mathbf{x} \in \underline{\mathbb{T}_d := \mathbb{R}^d},$$

$\mathcal{K}$: smoother/kernel on $\mathbb{S}_d \subseteq \mathbb{R}^d$ $(\mu_{\mathcal{K}} = 0, \Sigma_{\mathcal{K}} = \mathbf{I}_d)$; $\mathbf{H}_j$: bandwidth matrix; n_j: known and not-random.

- Classical kernel discriminant analysis:

$$\text{KDR: Allocate } \mathbf{x} \text{ to group } \widehat{j_0} \text{ where } \widehat{j_0} = \arg\max_{j \in \{1,\ldots,J\}} \widehat{\pi_j} \widehat{f_j}(\mathbf{x}; \mathbf{H}_j)$$

$\widehat{\pi}_j = n_j/n$ with $\sum_{j=1}^{J} n_j = n$.

[i] Bouezmarni, T. & Rombouts, J.V. (2010). *J. Statist. Plann. Inference* **140**, 139–152

[ii] Kokonendji, C.C. & Somé, S.M. (2018). *J. Korean Statist. Soc.* **47**, 112–126.

Let $\mathcal{X}_j = \{\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j}\}$ a sample of random vectors on $\mathbb{T}_d (\subseteq \mathbb{R}^d)$; $j = 1, \ldots, J$.

- Classical kernel estimator of f_j:

$$\widehat{f_j}(\mathbf{x}; \mathcal{K}, \mathbf{H}_j) = \frac{1}{n_j \det \mathbf{H}_j} \sum_{i=1}^{n_j} \mathcal{K}\left\{\mathbf{H}_j^{-1}(\mathbf{x} - \mathbf{X}_{ji})\right\}, \quad \forall \mathbf{x} \in \underline{\mathbb{T}_d := \mathbb{R}^d},$$

$\mathcal{K}$: smoother/kernel on $\mathbb{S}_d \subseteq \mathbb{R}^d$ $(\mu_{\mathcal{K}} = 0, \Sigma_{\mathcal{K}} = \mathbf{I}_d)$; $\mathbf{H}_j$: bandwidth matrix; n_j: known and not-random.

- Classical kernel discriminant analysis:

KDR: Allocate $\mathbf{x}$ to group $\widehat{j_0}$ where $\widehat{j_0} = \arg \max_{j \in \{1,\ldots,J\}} \widehat{\pi}_j \widehat{f_j}(\mathbf{x}; \mathbf{H}_j)$

$\widehat{\pi}_j = n_j / n$ with $\sum_{j=1}^{J} n_j = n$.

- ?¿ Associated kernels ? $\longrightarrow$ Respect of the support $\mathbb{T}_d$.

  multivariate (associated) kernel[i], product of univariate (e.g. Senga Kiéssé, 2008; Libengué, 2013).

Correlation structure$\rightarrow$(full) bandwidth matrix[ii]

[i] Bouezmarni, T. & Rombouts, J.V. (2010). *J. Statist. Plann. Inference* **140**, 139–152

[ii] Kokonendji, C.C. & Somé, S.M. (2018). *J. Korean Statist. Soc.* **47**, 112–126.

Definition

Let $\mathbf{x} \in \mathbb{T}_d$ *and* $\mathbf{H} = (h_{ij})_{i,j=1,\ldots,d}$ *a bandwidth matrix. A parametrized probability "density" function* $K_{\mathbf{x},\mathbf{H}}$ *of support* $\mathbb{S}_{\mathbf{x},\mathbf{H}} \left(\subseteq \mathbb{R}^d\right)$ *is called multivariate (or general) associated kernel if:*

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \quad \mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} + \mathbf{a}(\mathbf{x},\mathbf{H}) \ \ \textit{and} \ \ Cov\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{B}(\mathbf{x},\mathbf{H}),$$

with $\mathcal{Z}_{\mathbf{x},\mathbf{H}} \sim K_{\mathbf{x},\mathbf{H}}$, $\mathbf{a}(\mathbf{x},\mathbf{H}) = (a_1(\mathbf{x},\mathbf{H}),\ldots,a_d(\mathbf{x},\mathbf{H}))^\top \to \mathbf{0}$ *and* $\mathbf{B}(\mathbf{x},\mathbf{H}) = \left(b_{ij}(\mathbf{x},\mathbf{H})\right)_{i,j=1,\ldots,d} \to \mathbf{0}_d$ *as* $\mathbf{H} \to \mathbf{0}_d$.

Definition

Let $\mathbf{x} \in \mathbb{T}_d$ and $\mathbf{H} = (h_{ij})_{i,j=1,\dots,d}$ a bandwidth matrix. A parametrized probability "density" function $K_{\mathbf{x},\mathbf{H}}$ of support $\mathbb{S}_{\mathbf{x},\mathbf{H}} \left(\subseteq \mathbb{R}^d\right)$ is called multivariate (or general) associated kernel if:

$$\mathbf{x} \in \mathbb{S}_{\mathbf{x},\mathbf{H}}, \quad \mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} + \mathbf{a}(\mathbf{x},\mathbf{H}) \ \text{ and } \ Cov\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{B}(\mathbf{x},\mathbf{H}),$$

with $\mathcal{Z}_{\mathbf{x},\mathbf{H}} \sim K_{\mathbf{x},\mathbf{H}}$, $\mathbf{a}(\mathbf{x},\mathbf{H}) = (a_1(\mathbf{x},\mathbf{H}),\dots,a_d(\mathbf{x},\mathbf{H}))^\top \to \mathbf{0}$ and $\mathbf{B}(\mathbf{x},\mathbf{H}) = \left(b_{ij}(\mathbf{x},\mathbf{H})\right)_{i,j=1,\dots,d} \to \mathbf{0}_d$ as $\mathbf{H} \to \mathbf{0}_d$.

- Classical (continuous) associated kernel (i.e. $\mathbb{T}_d = \mathbb{R}^d$) :

$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \frac{1}{\det \mathbf{H}} \mathcal{K}\left\{\mathbf{H}^{-1}(\mathbf{x} - \cdot)\right\}, \quad \mathbb{S}_{\mathbf{x},\mathbf{H}} = \mathbf{x} - \mathbf{H}\mathbb{S}_d;$$

$$\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{x} \ \text{ and } \ \mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{H}\Sigma_{\mathcal{K}}\mathbf{H}.$$

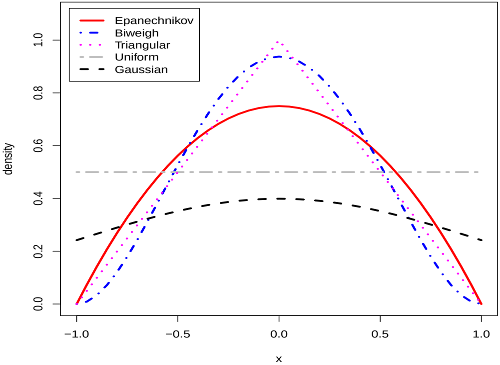


Figure: Some classical (continuous) associated kernels

[iii] Belaid et al., 2015. Bayesian bandwidth selection … *J. Korean Statist. Soc.* **45**, 557–567.

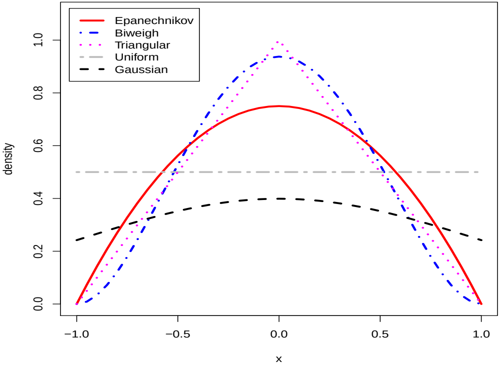


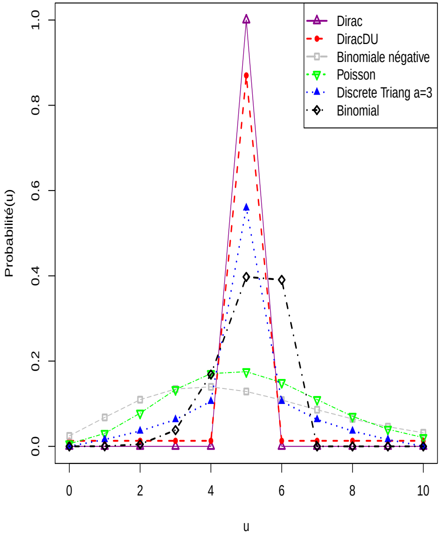
Figure: Some classical (continuous) associated kernels

- Multiple associated kernel[iii] (i.e. $\mathbb{T}_d = \times_{j=1}^{d} \mathbb{T}_1^{[j]}$, continuous or discrete) :

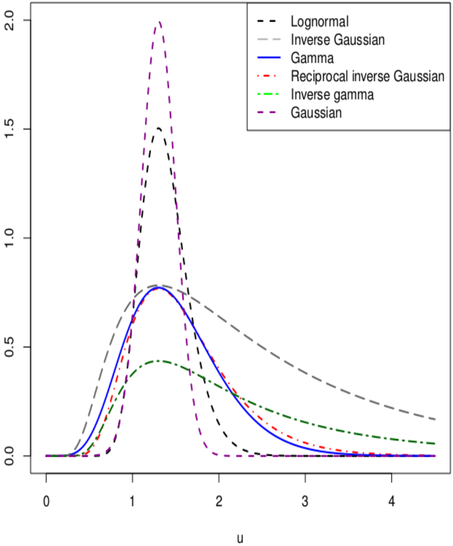
$$K_{\mathbf{x},\mathbf{H}}(\cdot) = \prod_{j=1}^{d} K^{[j]}_{x_j,h_{jj}}(\cdot), \qquad \mathbb{S}_{\mathbf{x},\mathbf{H}} = \times_{j=1}^{d} \mathbb{S}_{x_j,h_{jj}};$$

$$\mathbb{E}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \left(x_j + a_{jj}(x_j,h_{jj})\right)^{\top}_{j=1,\dots,d} \text{ and } \mathrm{Cov}\left(\mathcal{Z}_{\mathbf{x},\mathbf{H}}\right) = \mathbf{Diag}\left(b_{jj}(x_j,h_{jj})\right)_{j=1,\dots,d}.$$

[iii] Belaid et al., 2015. Bayesian bandwidth selection … *J. Korean Statist. Soc.* **45**, 557–567.



(a) Discrete[iv]



(b) Continuous[v]

[iv] Senga Kiéssé, T. (2008). Thèse, Univ. Pau et Pays de l'Adour.
[v] Libengué, F.G. (2013). Thèse, Univ. Franche-Comté & Univ. Ouagadougou.

- Constructed associated kernel ($\mathbb{T}_d \subseteq \mathbb{R}^d$):
  - $K_{x,\mathbf{H}}$ linked to both $x \in \mathbb{T}_d$ and **H** with $\leq d(d+1)/2$ parameters.
  - *Type of kernel* K_θ, $\theta \in \Theta \subseteq \mathbb{R}^{k_d}$: probab. density function, with support $\mathbb{S}_\theta \subseteq \mathbb{R}^d$,$k_d \geq d(d+3)/2$, unimodal with mode **m** and dispersion matrix **D**.

- Constructed associated kernel ($\mathbb{T}_d \subseteq \mathbb{R}^d$):
  - $K_{x,\mathbf{H}}$ linked to both $x \in \mathbb{T}_d$ and $\mathbf{H}$ with $\leq d(d+1)/2$ parameters.
  - *Type of kernel* K_θ, $\theta \in \Theta \subseteq \mathbb{R}^{k_d}$: probab. density function, with support $\mathbb{S}_\theta \subseteq \mathbb{R}^d$,$k_d \geq d(d+3)/2$, unimodal with mode $\mathbf{m}$ and dispersion matrix $\mathbf{D}$.
  - *Construction* (continuous case) :
    Type of kernel $K_\theta \longrightarrow$ associated kernel $K_{\theta(\mathbf{x},\mathbf{H})}$
    Multivariate mode-dispersion method ($\theta = \theta(\mathbf{m},\mathbf{D})$):

$$(\theta(\mathbf{m},\mathbf{D}))^\top = (\mathbf{x}, \mathrm{vech}\mathbf{H})^\top,$$

    vech$\mathbf{H}$: stacking the columns of the lower triangular of $\mathbf{H}$.

    Multivariate associated kernel $K_{\theta(\mathbf{x},\mathbf{H})}$, of support $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})}$, $\mathbf{a}_\theta(\mathbf{x},\mathbf{H}) \to \mathbf{0}$ and $\mathbf{B}_\theta(\mathbf{x},\mathbf{H}) \to \mathbf{0}_d$ as $\mathbf{H} \to \mathbf{0}_d$.
  - *Dimension* $d = 2$: Distribution K_θ, $\theta \in \Theta \subseteq \mathbb{R}^5$.

Univariate beta density :

$$g_i(t) = \frac{1}{\mathscr{B}(a_i, b_i)} t^{a_i - 1}(1-t)^{b_i - 1}\mathbb{I}_{[0,1]}(t), \quad i = 1, 2,$$

$a_i > 0$, $b_i > 0$, $\mathscr{B}(a_i, b_i) = \Gamma(a_i + b_i)/\{\Gamma(a_i)\Gamma(b_i)\}$.

Bivariate beta with correlation :

$$g_\theta(v) = g_1(v_1)g_2(v_2)\left[1 + \left\{\frac{v_1 - \mu_1(a_1, b_1)}{\sigma_1(a_1, b_1)}\right\}\left\{\frac{v_2 - \mu_2(a_2, b_2)}{\sigma_2(a_2, b_2)}\right\}\rho\right]\mathbb{I}_{[0,1]^2}(v),$$

with $v = (v_1, v_2)^T$, $\theta := \theta(a_1, b_1, a_2, b_2, \rho) \in \Theta \subseteq \mathbb{R}^5$ and, according to Sarmanov (1966), $\rho = \rho(a_1, b_1, a_2, b_2) \in \left[-\varepsilon, \varepsilon'\right] \subset [-1, 1]$, $\varepsilon, \varepsilon' \geqslant 0$:

$$\varepsilon = \left(\max\left\{\frac{v_1 - \mu_1(a_1, b_1)}{\sigma_1(a_1, b_1)}\right\}\left\{\frac{v_2 - \mu_2(a_2, b_2)}{\sigma_2(a_2, b_2)}\right\}\right)^{-1}, \varepsilon' = \left|\left(\min\left\{\frac{v_1 - \mu_1(a_1, b_1)}{\sigma_1(a_1, b_1)}\right\}\left\{\frac{v_2 - \mu_2(a_2, b_2)}{\sigma_2(a_2, b_2)}\right\}\right)^{-1}\right|.$$

- Mean vector and covariance matrix of g_θ:

$$\mu = (\mu_1, \mu_2)^T \quad \text{et} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

$$\mu_i = \frac{a_i}{a_i + b_i} = \mu_i(a_i, b_i), \sigma_i^2 = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)} = \sigma_i^2(a_i, b_i), \rho = \rho(a_1, b_1, a_2, b_2).$$

- Mean vector and covariance matrix of g_θ:

$$\mu = (\mu_1, \mu_2)^T \quad \text{et} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

$$\mu_i = \frac{a_i}{a_i + b_i} = \mu_i(a_i, b_i), \sigma_i^2 = \frac{a_i b_i}{(a_i + b_i)^2(a_i + b_i + 1)} = \sigma_i^2(a_i, b_i), \rho = \rho(a_1, b_1, a_2, b_2).$$

- Mode of g_θ:

$$\mathbf{m}(a_1, a_2, b_1, b_2, \rho) := \mathbf{m}_\rho \neq \mathbf{m}_0 =: (m_1(a_1, b_1), m_2(a_2, b_2))^T,$$

$m_i(a_i, b_i) = (a_i - 1)/(a_i + b_i - 2)$, with $a_i \geqslant 1$, $b_i \geqslant 1$ and $(a_i, b_i) \neq (1, 1)$.

- Mean vector and covariance matrix of g_θ:

$$\mu = (\mu_1, \mu_2)^T \quad \text{et} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

$$\mu_i = \frac{a_i}{a_i+b_i} = \mu_i(a_i,b_i), \sigma_i^2 = \frac{a_ib_i}{(a_i+b_i)^2(a_i+b_i+1)} = \sigma_i^2(a_i,b_i), \rho = \rho(a_1,b_1,a_2,b_2).$$

- Mode of g_θ:

$$\mathbf{m}(a_1,a_2,b_1,b_2,\rho) := \mathbf{m}_\rho \neq \mathbf{m}_0 =: (m_1(a_1,b_1), m_2(a_2,b_2))^T,$$

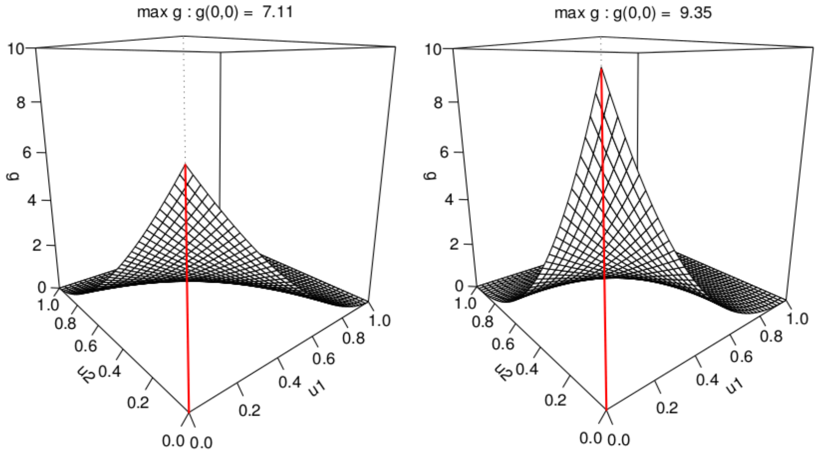
$m_i(a_i,b_i) = (a_i-1)/(a_i+b_i-2)$, with $a_i \geqslant 1$, $b_i \geqslant 1$ and $(a_i,b_i) \neq (1,1)$.

- Dispersion matrix of g_θ:

$$\mathbf{D}_\rho = \begin{pmatrix} d_1 & (d_1d_2)^{1/2}\rho \\ (d_1d_2)^{1/2}\rho & d_2 \end{pmatrix},$$

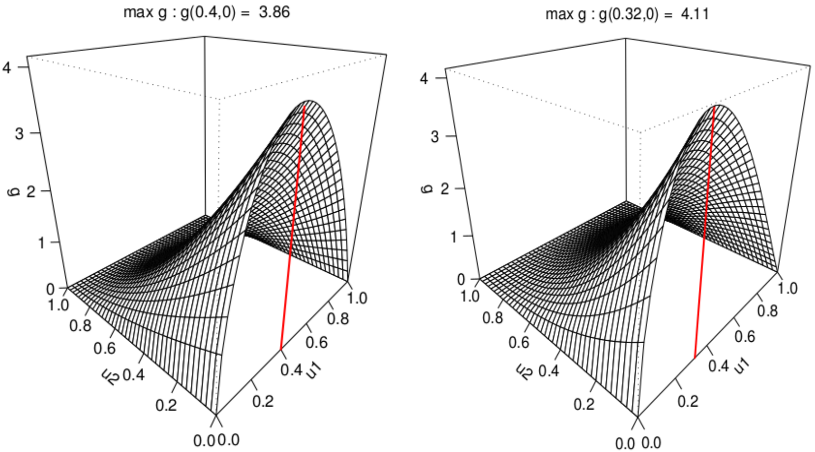
$d_i = 1/(a_i+b_i-2) = d_i(a_i,b_i)$, with $a_i \geqslant 1$, $b_i \geqslant 1$ and $(a_i,b_i) \neq (1,1)$.

Unvarying mode if "angle(s)" but changing maximum (1/3)



(c) $\rho = 0$ (d) $\rho = 0.18$

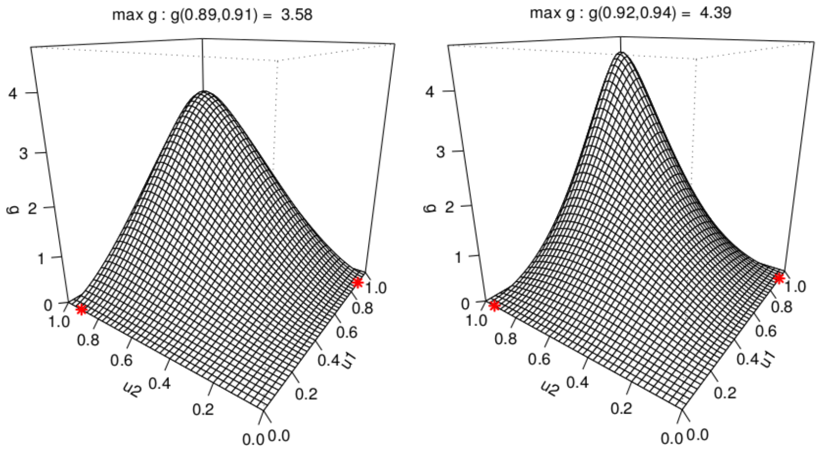
Figure: $\rho \in [-0.100, 0.210]$, $a_1 = a_2 = 1$, $b_1 = b_2 = 8/3$.



(a) $\rho = 0$ (b) $\rho = 0.19$

Figure: $\rho \in [-0.120, 0.143]$, $a_1 = 5/3, a_2 = 1$, $b_1 = 2, b_2 = 8/3$.

Unvarying mode if “angle(s)” but changing maximum (3/3)



(a) $\rho = 0$ (b) $\rho = 0.12$

Figure: $\rho \in [-0.008, 0.214]$ $a_1 = 149/60, a_2 = 151/60$, $b_1 = 71/60, b_2 = 23/20$.

- Density $g_\theta \longrightarrow$ Bivariate kernel $BS_{\theta(x,\mathbf{H})}$ by variant mode-mispersion method.

Let $\mathbf{x} = (x_1, x_2)^T$ et $\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$,

$$(\mathbf{m}_0, \mathbf{D}_\rho) = (\mathbf{x}, \mathbf{H}) \iff \left(m_i = x_i, \quad d_i = h_{ii}, \quad (d_1 d_2)^{1/2}\rho = h_{12}\right).$$

- Density $g_\theta \longrightarrow$ Bivariate kernel $BS_{\theta(x,\mathbf{H})}$ by variant mode-mispersion method.

Let $\mathbf{x} = (x_1, x_2)^T$ et $\mathbf{H} = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$,

$$(\mathbf{m}_0, \mathbf{D}_\rho) = (\mathbf{x}, \mathbf{H}) \iff \left(m_i = x_i, \quad d_i = h_{ii}, \quad (d_1 d_2)^{1/2}\rho = h_{12}\right).$$

- New parametrization:

$$\theta(\mathbf{x}, \mathbf{H}) = \left(\frac{x_1}{h_{11}} + 1, \frac{1-x_1}{h_{11}} + 1, \frac{x_2}{h_{22}} + 1, \frac{1-x_2}{h_{22}} + 1, \frac{h_{12}}{(h_{11}h_{22})^{1/2}}\right)^\top, \quad \forall \mathbf{x}, \mathbf{H}.$$

$$\begin{aligned} BS_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{v}) &= Be\left(v_1, 1 + \frac{x_1}{h_{11}}, 1 + \frac{1-x_1}{h_{11}}\right) \times Be\left(v_2, 1 + \frac{x_2}{h_{22}}, 1 + \frac{1-x_2}{h_{22}}\right) \\ &\times \left\{1 + \left(\frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1, h_{11})}\right)\left(\frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2, h_{22})}\right) h_{12}\right\} \mathbb{I}_{[0,1]^2}(v), \end{aligned}$$

$h_{12} \in \left[-\beta, \beta'\right] \cap [-(h_{11}h_{22})^{1/2}, (h_{11}h_{22})^{1/2}]$, $\beta, \beta' \geqslant 0$,

$$\beta = \left(\max\left\{\frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1, h_{11})}\right\}\left\{\frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2, h_{22})}\right\}\right)^{-1}, \beta' = \left|\left(\min\left\{\frac{v_1 - \widetilde{\mu}_1(x_1, h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1, h_{11})}\right\}\left\{\frac{v_2 - \widetilde{\mu}_2(x_2, h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2, h_{22})}\right\}\right)^{-1}\right|.$$

- Bivariate beta kernel with correlation (Kokonendji & Somé, 2018):

$$BS_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{v}) = Be\left(v_1, 1+\frac{x_1}{h_{11}}, 1+\frac{1-x_1}{h_{11}}\right)\times Be\left(v_2, 1+\frac{x_2}{h_{22}}, 1+\frac{1-x_2}{h_{22}}\right)$$
$$\times\left\{1+\left(\frac{v_1-\widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1,h_{11})}\right)\left(\frac{v_2-\widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2,h_{22})}\right)h_{12}\right\}\mathbb{I}_{[0,1]^2}(v),$$

$h_{12} \in \left[-\beta,\beta'\right] \cap [-(h_{11}h_{22})^{1/2},(h_{11}h_{22})^{1/2}],\ \beta,\ \beta' \geqslant 0,$

$$\beta = \left(\max\left\{\frac{v_1-\widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1,h_{11})}\right\}\left\{\frac{v_2-\widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2,h_{22})}\right\}\right)^{-1}, \beta' = \left|\left(\min\left\{\frac{v_1-\widetilde{\mu}_1(x_1,h_{11})}{h_{11}^{1/2}\,\widetilde{\sigma}_1(x_1,h_{11})}\right\}\left\{\frac{v_2-\widetilde{\mu}_2(x_2,h_{22})}{h_{22}^{1/2}\,\widetilde{\sigma}_2(x_2,h_{22})}\right\}\right)^{-1}\right|.$$

- Characteristics of $BS_{\theta(\mathbf{x},\mathbf{H})}$:
  - $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} = [0,1]^2 = \mathbb{T}_2,$
  - $\mathbf{a}_\theta(\mathbf{x},\mathbf{H}) = (a_1,a_2)^T$ with $a_i = \frac{(1-2x_i)h_{ii}}{(1+2h_{ii})} = a_i(x_i,h_{ii}),$
  - $\mathbf{B}_\theta(\mathbf{x},\mathbf{H}) = (b_{ij})_{i,j=1,2} : b_{ii} = \widetilde{\sigma}_i^2(x_i,h_{ii})$ and $b_{12} = \frac{h_{12}}{(h_{11}h_{22})^{1/2}}\widetilde{\sigma}_1(x_1,h_{11})\widetilde{\sigma}_2(x_2,h_{22}).$

- Bivariate binomial kernel with correlation :

$$BnS_{\theta(\mathbf{x},\mathbf{H})}(\mathbf{v}) = Bn\left(v_1, 1+x_1, \frac{x_1+h_{11}}{x_1+1}\right) \times Bn\left(v_2, 1+x_2, \frac{x_2+h_{22}}{x_2+1}\right) \times \left\{1 + \left(\frac{h_{11}^{-1/2}\left(e^{-v_1} - L_1(1)\right)\widetilde{\sigma}_1}{L_1'(1) + \widetilde{\mu}_1 L_1(1)}\right)\left(\frac{h_{22}^{-1/2}\left(e^{-v_2} - L_2(1)\right)\widetilde{\sigma}_2}{L_1'(1) + \widetilde{\mu}_2 L_2(1)}\right) h_{12}\right\} \mathbb{I}_{\mathbb{S}_{x_1}\times\mathbb{S}_{x_2}}(\mathbf{v}),$$

$\mathbb{S}_{x_i} = \{0, 1, \ldots, x_i+1\}$, $h_{12} \in \left[-\beta, \beta'\right] \cap [-(h_{11}h_{22})^{1/2}, (h_{11}h_{22})^{1/2}]$, $\beta, \beta' \geqslant 0$
$L_i(t) = \mathbb{E}\{e^{-t\mathbf{X}_i}\}$ (Laplace transform of Bn),

$$\beta = \left(\max\left\{\frac{h_{11}^{-1/2}\left(e^{-v_1} - L_1(1)\right)\widetilde{\sigma}_1}{L_1'(1) + \widetilde{\mu}_1 L_1(1)}\right\}\left\{\frac{h_{22}^{-1/2}\left(e^{-v_2} - L_2(1)\right)\widetilde{\sigma}_2}{L_1'(1) + \widetilde{\mu}_2 L_2(1)}\right\}\right)^{-1}, \beta' = \left|\left(\min\left\{\frac{h_{11}^{-1/2}\left(e^{-v_1} - L_1(1)\right)\widetilde{\sigma}_1}{L_1'(1) + \widetilde{\mu}_1 L_1(1)}\right\}\left\{\frac{h_{22}^{-1/2}\left(e^{-v_2} - L_2(1)\right)\widetilde{\sigma}_2}{L_1'(1) + \widetilde{\mu}_2 L_2(1)}\right\}\right)^{-1}\right|.$$

- Characteristics of $BS_{\theta(\mathbf{x},\mathbf{H})}$:
  - $\mathbb{S}_{\theta(\mathbf{x},\mathbf{H})} = \mathbb{S}_{x_1} \times \mathbb{S}_{x_2}$; $\mathbb{T}_2 = \mathbb{N}^2$,
  - $\mathbf{a}_\theta(\mathbf{x},\mathbf{H}) = (a_1, a_2)^T$ with $a_i = h_{ii}$,
  - $\mathbf{B}_\theta(\mathbf{x},\mathbf{H}) = (b_{ij})_{i,j=1,2}$: $b_{ii} = \widetilde{\sigma}_i^2(x_i, h_{ii})$, $b_{12} = \frac{h_{12}}{(h_{11}h_{22})^{1/2}}\widetilde{\sigma}_1(x_1,h_{11})\widetilde{\sigma}_2(x_2,h_{22})$, and $\lim_{h_{ii}\to 0} b_{ii} \in [0, 1)$.

Associated kernel estimator $\widehat{f_j}$ of f_j:

$$\widetilde{f_j}(\mathbf{x}) = \frac{1}{n_j}\sum_{i=1}^{n_j} K_{\mathbf{x},\mathbf{H}_j}\left(\mathbf{X}_{ji}\right) = \widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa},\mathbf{H}_j),$$

Associated kernel discriminant rule:

AKDR: Allocate $\mathbf{x}$ to group $\widetilde{j_0}$ where $\widetilde{j_0} = \arg\max_{j\in\{1,\dots,J\}} \widehat{\pi_j}\widetilde{f_j}(\mathbf{x};\boldsymbol{\kappa})$.

Associated kernel estimator $\widehat{f_j}$ of f_j:

$$\widetilde{f_j}(\mathbf{x}) = \frac{1}{n_j} \sum_{i=1}^{n_j} K_{\mathbf{x},\mathbf{H}_j}\left(\mathbf{X}_{ji}\right) = \widetilde{f_j}(\mathbf{x}; \boldsymbol{\kappa}, \mathbf{H}_j),$$

Associated kernel discriminant rule:

AKDR: Allocate $\mathbf{x}$ to group $\widetilde{j_0}$ where $\widetilde{j_0} = \arg \max_{j \in \{1,\ldots,J\}} \widehat{\pi_j} \widetilde{f_j}(\mathbf{x}; \boldsymbol{\kappa})$.

Misclassification rate for test data $\mathbf{Y}_1, \ldots, \mathbf{Y}_m \in \mathbb{T}_d$

$$\widehat{\mathrm{MR}} = 1 - m^{-1} \sum_{k=1}^{m} 1\{\mathbf{Y}_k \text{ is correctly classified using AKDR}\}$$

Bandwidth matrix selection

Classical cross-validation[vi,vii]

$$\widehat{\mathbf{H}_\mathbf{j}} = \arg \min_{\mathbf{H}_\mathbf{j} \in \mathcal{H}} \mathrm{LSCV}(\mathbf{H}_\mathbf{j}), \quad \mathrm{LSCV}(\mathbf{H}) = \int_{\mathbb{T}_d} \left\{\widehat{f}_n(\mathbf{x})\right\}^2 \nu(d\mathbf{x}) + \frac{2}{n_j} \sum_{i=1}^{n} \widehat{f}_{j,-i}(\mathbf{X}_{ji}),$$

$\widehat{f}_{j,-i}(\mathbf{X}_{ji}) = (n_j - 1^{-1} \sum_{k \neq i} K_{\mathbf{X}_{ji}, \mathbf{H}_j}(\mathbf{X}_{jk}),$ $\mathcal{H}$ = set of symmetric and positive definite matrices.

[vi] Chacón, J.E. & Duong, T. (2011). *Australian* & *New Zealand JS* **53**, 331–351.
[vii] Zougab, N., Adjabi, S. & Kokonendji, C.C. (2014). *Comput. Stat. Data Anal.* **75**, 28–38

Bandwidth matrices[viii,ix,x] : $\widehat{f}_{n,-i}(\mathbf{X}_i) = \frac{1}{n-1}\sum_{\ell=1,\ell\neq i}^{n} K_{\mathbf{x},\mathbf{H}}(\mathbf{X}_i)$

Very important to avoid over- and under-smoothing; several methods:

Cross-validation (global) : $\widehat{\mathbf{H}}_{cv} = \mathrm{argmin}_{\mathbf{H}\in\mathcal{H}}\left(\int_{\mathbb{T}_d}\left\{\widehat{f}_n(x)\right\}^2 dx + \frac{2}{n}\sum_{i=1}^{n}\widehat{f}_{n,-i}(\mathbf{X}_i)\right)$

Bayes global: $\widehat{\mathbf{H}}_{Bg} \propto \int \mathbf{H}\,\pi(\mathbf{H})\prod_{i=1}^{n}\widehat{f}_{n,\mathbf{H};-i}(\mathbf{X}_i)d\mathbf{H}$

Bayes local: $\widehat{\mathbf{H}}_{B\ell}(\mathbf{x}) \propto \int \mathbf{H}\,\pi(\mathbf{H})\widehat{f}_{n,\mathbf{H}}(\mathbf{x})d\mathbf{H}, \qquad \forall \mathbf{x}\in\mathbb{T}_d$

Bayes adaptive: $\widehat{\mathbf{H}}_{Ba,i} \propto \int \mathbf{H}_i\,\pi(\mathbf{H}_i)\widehat{f}_{n,\mathbf{H}_i;-i}(\mathbf{X}_i)d\mathbf{H}_i, \qquad \forall i = 1,\dots,n$

[viii] Zougab, N., Adjabi, S., Kokonendji, C.C., 2015. Comparison study to bandwidth selection ... *J. Statist. Theory and Practice*, **10**, 133–153. (d=1 and discrete)

[ix] Belaid, N., Adjabi, S., Zougab, N., Kokonendji, C.C., 2016a. Bayesian bandwidth selection ... *J. Korean Statist. Soc.* **45**, 557–567. (discrete multiple)

[x] Wansouwé, W.E., Somé, S.M. & Kokonendji, C.C. (2016). Ake : an R package ... The R Journal 8 (2), 258-276 (d=1 and continuous/discrete)

Algorithm for associated kernel discriminant analysis

1. For each training sample $\mathcal{X}_j = \{\mathbf{X}_{j1}, \ldots, \mathbf{X}_{jn_j}\}$, $j = 1, 2, \ldots, J$, compute $\widehat{f}_j$ using $\widehat{\mathbf{H}}_j$.
2. If $\widehat{\pi}_j$ are known then use these. Otherwise, use $\widehat{\pi}_j = n_j/n$ of training sample proportions.
3. (a) Allocate test data points $\mathbf{Y}_1, \ldots, \mathbf{Y}_m$ according to AKDR.
   (b) Allocate all points $\mathbf{x}$ from the sample space according to AKDR.
4. (a) If we have test data then the estimate of the misclassification rate is $\widehat{\mathrm{MR}}$.
   (b) If we do not have test data the cross-validation estimate of MR is

$$\widehat{\mathrm{MR}}_{\mathrm{cv}} = 1 - n^{-1} \sum_{j=1}^{J} \sum_{i=1}^{n_j} 1\{\mathbf{X}_{ji} \text{ is correctly classified using } \mathrm{AKDR}_{ji}\}$$

$\mathrm{AKDR}_{ji} \equiv AKDR$ except: $\widehat{\pi}_j \to \widehat{\pi}_{j,-i} = (n_j - 1)/n$;

$$\widetilde{f}_j(\mathbf{x};\boldsymbol{\kappa}) \to \widetilde{f}_{j,-i}(\mathbf{x};\boldsymbol{\kappa}) = \frac{1}{n_j - 1} \sum_{r \neq i}^{n_j} K_{\mathbf{x},\mathbf{H}_{j,-i}}\left(\mathbf{X}_{jr}\right).$$

Repeat step 3 to classify all $\mathbf{X}_{ji}$ using AKDR_{ji}.

AKDR with multiple associated kernels with $j = 2$

- Mixture of bivariate dirichlet with $\pi_1 = 3/7$ and $\pi_2 = 4/7$:

m	n	Beta×Beta
200	250	0.042111111(0.01467407)
	500	0.03288889(0.009681806)

Table: $\widehat{\overline{\mathrm{MR}}}$ and $\sigma_{\widehat{\mathrm{MR}}}$ in parentheses with $N_{sim} = 100$.

- Mixture of bivariate Poisson $\pi_1 = 2/5$ and $\pi_2 = 3/5$:

m	n	Bin×Bin
	50	0.07487198(0.03347161)
100	100	0.06767442(0.02552657)
	200	0.05809524(0.02211258)

Table: $\widehat{\overline{\mathrm{MR}}}$ and $\sigma_{\widehat{\mathrm{MR}}}$ in parentheses with $N_{sim} = 100$.

- Mixture of Beta×Poisson $\pi_1 = 3/7$ and $\pi_2 = 4/7$:

m	n	Beta×Bin
	80	0.01822222(0.01163632)
150	250	0.01714286(0.008017837)
	500	0.01523252(0.00415842)

Table: $\widehat{\overline{\mathrm{MR}}}$ and $\sigma_{\widehat{\mathrm{MR}}}$ in parentheses with $N_{sim} = 100$.

- *Coronary heart disease* data (Rousseauw et al., 1983 ; Halvorsen, 2015[xi]) with $n = 476$
  - sbp: systolic blood pressure $\longrightarrow$ binomial kernel
  - tobacco: cumulative tobacco (kg) $\longrightarrow$ gamma kernel
  - ldl: low densiity lipoprotein cholesterol $\longrightarrow$ gamma kernel
  - adiposity $\longrightarrow$ gamma kernel
  - famhist: family history of heart disease (0,1) $\longrightarrow$ diracDU kernel
  - typea: type-A behavior $\longrightarrow$ binomial kernel
  - obesity $\longrightarrow$ gamma kernel
  - alcohol: current alcohol consumption $\longrightarrow$ gamma kernel
  - age: age at onset $\longrightarrow$ binomial kernel

  Class label: coronary heart disease: negative (0) or positive (1).

- AKDR with multiple assocated kernel
  $\mathbf{H}_1 = \mathbf{Diag}\,(0.01, 0.02, 0.03, 0.015, 0.004, 0.02, 0.03, 0.02, 0.02)$
  $\mathbf{H}_2 = \mathbf{Diag}\,(0.87, 0.3, 0.2, 0.15, 0.4, 0.2, 0.4, 0.2, 0.56, 0.6)$ $\mathrm{MR}_{cv} = 30.090\%$.

- KDR with $\widehat{\mathbf{H}}_{\mathrm{LSCV}}$ of Duong (2007): $\mathrm{MR}_{cv} = 30.952\%$.

[xi] Halvorsen, K. (2015). `ElemStatLearn`: Data sets, functions and examples from the book: “The Elements of Statistical Learning, Data Mining, Inference, and Prediction” by Trevor Hastie, Robert Tibshirani and Jerome Friedman, URL `http://cran.r-project.org/web/packages/ElemStatLearn/index.html`.

Conclusion (partial):

- Appropriateness of associated kernels
- Efficiency of AKD method

Forthcoming works:

- Improve bandwidth matrices choices (e.g. Bayesian approaches)
- Construction/illustration of discrete/mixed associated kernels with correlation

References

1. Belaid, N., Adjabi, S., Zougab, N., Kokonendji, C.C., 2016a. Bayesian bandwidth selection . . . *J. Korean Statist. Soc.* **45**, 557–567.
2. Belaid, N., Adjabi, S., Kokonendji, C.C., Zougab, N., 2016b. Bayesian local bandwidth selector . . . *J. Stat. Comput. Simul.* **86**, 3667–3681
3. Bouezmarni, T. & Rombouts, J.V.K. (2010). Nonparametric density estimation . . ., *J. Statist.Plann. Inference* **140**, 139–152.
4. Chacón, J.E. & Duong, T. (2011). Unconstrained pilot selectors . . ., *Aust. N. Z. J. Stat* **53**, 331–351.
5. Chen, S.X. (1999). A beta kernel estimation for density functions, *Comput. Statist. Data Anal.* **31**, 131–145.
6. Duong, T., 2007. ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R, *Journal of Statistical Software* **21**, 1–16.
7. Hayfield, T. & Racine, J.S. (2007). Nonparametric econometrics: the np package, *Journal of Statistical Software* **27**, 1–32.
8. Kokonendji, C.C. & Somé, S.M. (2018). On multivariate associated kernels to estimate general density functions, *J. Korean Statist. Soc.* **47**, 112–126.
9. Libengué, F.G. (2013). *Méthode Non-Paramétrique par Noyaux Associés Mixtes et Applications*. Thèse, Univ. Franche-Comté & Univ. Ouagadougou.
10. Rousseauw, J., du Plessis, J., Benade, A., Jordaan, P., Kotze, J., Ferreira, J. (1983). Coronary . . ., *S. Afr. Med. J.* **64**, 430–436.
11. Sarmanov, O.V. (1966). Generalized normal correlation and two-dimensionnal Frechet classes, *Doklady (Soviet Mathematics)* **168**, 596–599.
12. Senga Kiéssé, T. (2008). *Approche Non-Paramétrique* . . .. Thèse, Univ. Pau et Pays de l'Adour.
13. Somé, S.M. & Kokonendji, C.C. (2016). Effects of associated kernels in nonparametric multiple regressions, *J. Stat. Theory Pract.*, **10**, 456-471.
14. Somé, S.M., Kokonendji, C.C. & Ibrahim, M. (2016). Associated kernel discriminant analysis for mixed data, *Electronic Journal of Applied Statistical Analysis*, **09**, 385-399.

𝒯hank 𝒴ou ℱor 𝒜ttention.