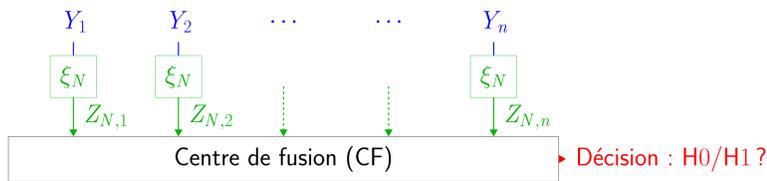


Modèle



- ▶ $(Y_k)_{k \in \mathbb{Z}}$: processus stationnaire ergodique, dans un convexe borné $Y \subset \mathbb{R}^d \dots$
- ▶ ... de loi \mathbb{P}_0 sous H_0 , \mathbb{P}_1 sous H_1
- ▶ $Z_{N,k} \triangleq \xi_N(Y_k)$: observation quantifiée sur $\log_2(N)$ bits
- ▶ Test au CF sur les observations quantifiées $Z_{N,1:n}$ pour décider H_0 ou H_1

Question

Quel est l'impact du quantificateur ξ_N sur la détection au CF ?

Hypothèse 1 (Densités)

Pour tout $i \in \{0, 1\}$, tout $n \geq 1$,

- ▶ $\mathbb{P}_i[Y_{1:n} \in \cdot]$ admet une densité p_i par rapport à la mesure de Lebesgue
- ▶ $p_i(y_{1:n}) > 0$, pour tout $y_{1:n} \in Y^n$
- ▶ $p_{i,N}(z_{1:n})$ est la loi des observations quantifiées $Z_{N,1:n}$ sous H_i

Test de Neyman-Pearson (de niveau α)

$$L_{N,n} = \log \frac{p_{1,N}(Z_{N,1:n})}{p_{0,N}(Z_{N,1:n})} \underset{H_0}{\overset{H_1}{\geq}} \gamma$$

avec γ tel que $\mathbb{P}_0[L_{N,n} > \gamma] \leq \alpha$.

- ▶ La probabilité de manque de ce test, $\beta_{N,n}(\alpha) = \inf_{\gamma} \mathbb{P}_1[L_{N,n} < \gamma]$, est une bonne mesure de performance...
- ▶ ... mais ne possède pas d'expression simple dans le cas général

Exposant d'erreur

Sous de bonnes hypothèses de convergence :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \beta_{N,n}(\alpha) = -K_N$$

avec

$$K_N = \lim_{m \rightarrow \infty} \mathbb{E}_0 \left[\log \frac{p_{0,N}(Z_{N,0} | Z_{N,-m:-1})}{p_{1,N}(Z_{N,0} | Z_{N,-m:-1})} \right]$$

- ▶ L'impact du quantificateur n'est pas directement quantifiable
- ▶ En particulier, $K_N \stackrel{?}{\simeq} K = \lim_{m \rightarrow \infty} \mathbb{E}_0 \left[\log \frac{p_0(Y_0 | Y_{-m:-1})}{p_1(Y_0 | Y_{-m:-1})} \right]$

Quantification

- Un quantificateur à N niveaux est un triplet $(C_N, \Xi_N, \xi_N)_{N \geq 1}$ composé de
- ▶ un ensemble de N cellules $C_{N,j}$, de volumes resp. $V_{N,j}$, qui partitionnent Y
 - ▶ un ensemble arbitraire d'éléments distincts $\xi_{N,j}$ (ici, les centres des $C_{N,j}$)
 - ▶ une fonction $\xi_N : Y \rightarrow \Xi_N$ telle que $\xi_N(y) = \xi_{N,j}$ si $y \in C_{N,j}$

et caractérisé par 2 fonctions sur Y , la densité spécifique ζ_N , et le profil spécifique M_N , définies pour $y \in C_{N,j}$ par :

$$\zeta_N(y) = \frac{1}{NV_{N,j}} \quad M_N(y) = \frac{1}{V_{N,j}^{1+2/d}} \int_{C_{N,j}} (s - \xi_{N,j})(s - \xi_{N,j})^T ds$$

Hypothèse 2 (Quantification haute résolution $N \rightarrow \infty$)

$$\zeta_N \xrightarrow[N \rightarrow \infty]{\text{unif.}} \zeta \quad M_N \xrightarrow[N \rightarrow \infty]{\text{unif.}} M$$

- ▶ ζ , fonction continue telle que $\inf_Y \zeta > 0$, est la **densité limite**
- ▶ M , fonction continue telle que $\sup_{y \in Y} \|M(y)\| < \infty$, est le **profil limite**

Hypothèse 3 (Douceur et oubli $n \rightarrow \infty$)

Cette hypothèse impose des conditions

- ▶ de douceur des densités de probabilité p_i
- ▶ d'oubli des observations passées par les processus (Y_k) et $(Z_{N,k})$, par ex. : $\mathbb{E}_0 \left| \log p_i(Y_0 | Y_{-m:-1}) - \log p_i(Y_0 | Y_{-m-\ell:-1}) \right| = O(m^{-6-\epsilon})$

Perte asymptotique en exposant d'erreur

$$N^{2/d}(K - K_N) \xrightarrow[N \rightarrow \infty]{} D_e = \frac{1}{2} \int \frac{p_0(y)F(y)}{\zeta(y)^{2/d}} dy$$

où la fonction F est définie par

$$F(y) = \mathbb{E}_0 \left[\ell(Y_{\mathbb{Z}})^T M(Y_0) \ell(Y_{\mathbb{Z}}) \mid Y_0 = y \right],$$

et la variable aléatoire $\ell(Y_{\mathbb{Z}})$ est la limite dans $L^2(\mathbb{P}_0)$ de $\left(\nabla_{y_0} \log \frac{p_0}{p_1}(Y_{-k:k}) \right)_{k \geq 0}$.

Conséquence 1

L'approximation suivante est valable quand $n, N \rightarrow \infty$

- ▶ pour une large de processus (faible dépendance, Markov, Markov cachés...)
- ▶ pour les quantificateurs "classiques" (uniformes, produits par un compander ou l'algorithme LBG...)

$$\beta_{N,n}(\alpha) \simeq \exp \left[-n \left(K - \frac{D_e}{N^{2/d}} \right) \right]$$

Conséquence 2 : "Bons" quantificateurs

Quel quantificateur minimise la perte D_e ?

- ▶ Pour M fixé, la densité limite optimale est

$$\zeta^*(y) = \frac{[p_0(y)F(y)]^{\frac{d}{d+2}}}{\int [p_0(s)F(s)]^{\frac{d}{d+2}} ds}$$

- ▶ Cas scalaire ($d = 1$, $Y = [a; b]$)

▶ si les cellules sont des intervalles, $M(y) = \frac{1}{12}$

▶ réalisable en pratique via un compander : compresseur $\phi(y) = \int_a^y \zeta^*(s) ds$, suivi d'un quantificateur uniforme sur $[0; 1]$

- ▶ Cas vectoriel ($d \geq 2$)

▶ si les cellules sont congruentes à moment d'inertie minimal, $M(y) = \nu I_d$

▶ réalisable en pratique par l'algorithme LBG avec un jeu de données d'apprentissage de densité

$$q^*(y) = \frac{p_0(y)F(y)}{\int p_0(s)F(s) ds}$$

Exemple : Signaux gaussiens 2D-Détection d'une structure AR

État caché :

$$H_0 : X_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, 1)$$

$$H_1 : X_k = aX_{k-1} + \sqrt{1-a^2} U_k$$

$a \in (0, 1)$: coefficient de corrélation

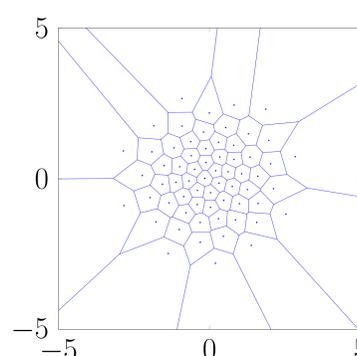
$U_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, 1)$: innovation

Observation :

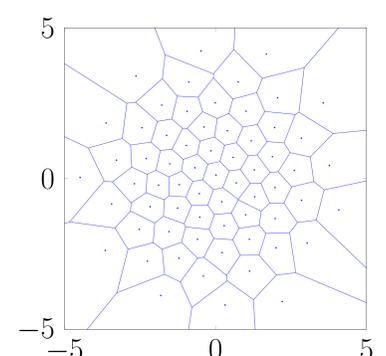
$$Y_k = X_k + W_k$$

$W_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{CN}(0, \sigma^2)$: bruit d'obs.

$a = 0,8$, $\sigma = 1$, $N = 64$, 20 000 échantillons fournis au LBG,



Quantificateur MSE-optimal



Quantificateur proposé