Feuille 1 de travaux dirigés

Exercice 1. [*Estimateur des moindres carrés en grande dimension*]

On observe
$$Y_i = f^\star(i/n) + \varepsilon_i, \quad 1 \leq i \leq n$$
où les ε_i sont i.i.d. $\mathcal{N}(0, \sigma^2)$ (σ connu) et $f^\star : [0,1] \to \mathbb{R}$ est un “signal" inconnu.

On cherche à estimer $\mu^\star = (f^\star(i/n))_{1 \leq i \leq n} \in \mathbb{R}^n$. Pour cela, on suppose que $f^\star$ s'écrit comme le début d'un développement en série de Fourier : pour tout $x \in [0,1]$,

$$f^\star(x) = a_0 + \sum_{k=1}^{K} \big(a_k \cos(2\pi k x) + b_k \sin(2\pi k x)\big),$$

pour a_0, a_k, b_k, $1 \leq k \leq K$ des réels inconnus.

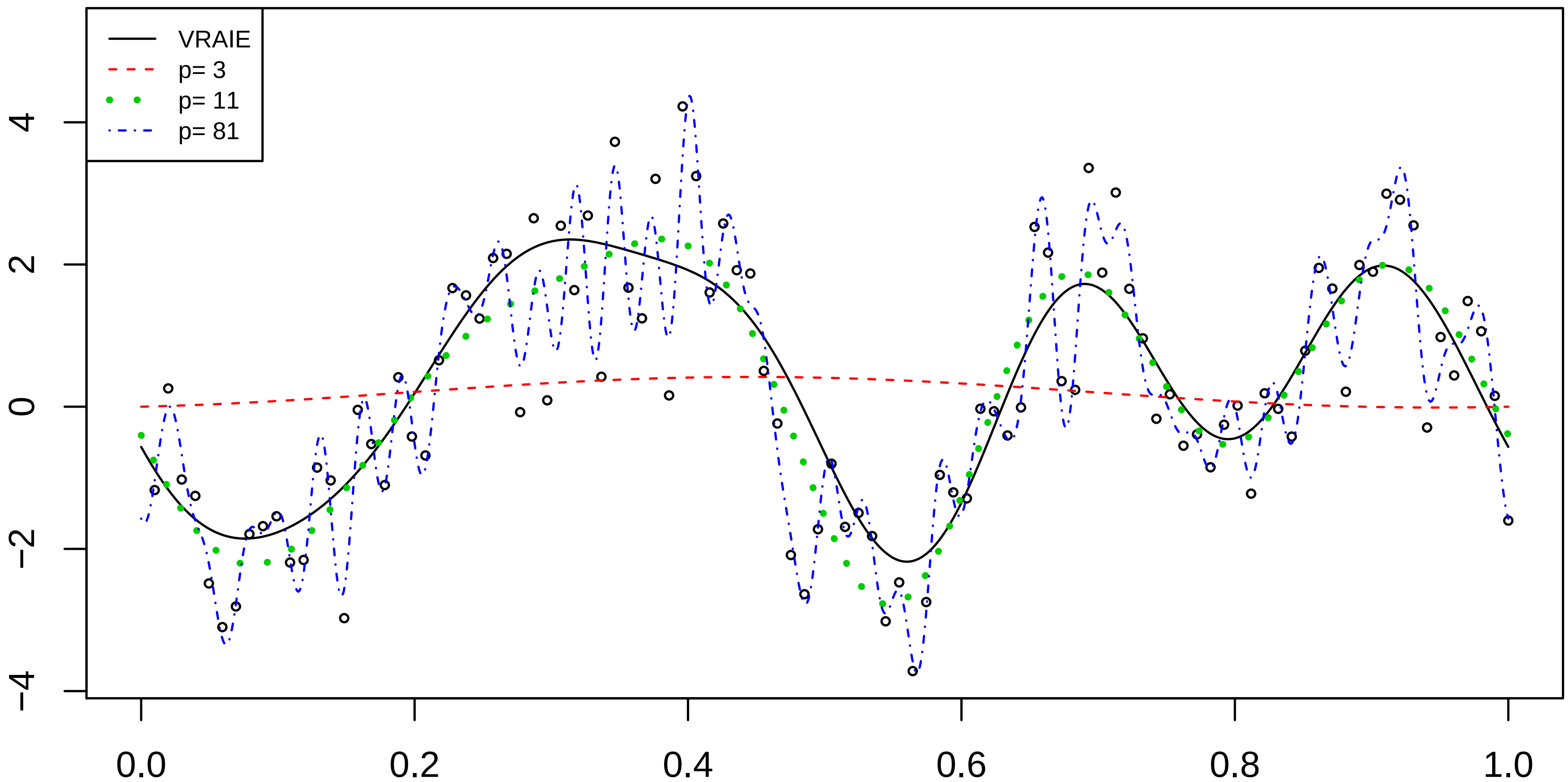
1. Écrire le modèle comme un modèle linéaire gaussien et préciser $X, \beta^\star$ et p.

2. On suppose à présent $p \leq n$. Calculer d =Vect$(X_1, \ldots, X_p)$. Donner l'estimateur des moindres carrés $\widehat{\beta}$ de $\beta^\star$. En déduire un estimateur $\widehat{\mu}$ de $\mu^\star$. Proposer également un estimateur $\widehat{f}$ de la *fonction* $f^\star$.

3. Donner la forme du risque de prédiction de l'estimateur des moindres carrés $R(\beta^\star, \widehat{\beta})$. Quelle est sa valeur ? Quel est l'ordre de ce risque lorsque p est fixe et n converge vers l'infini ?

4. On suppose maintenant $p = n$. Préciser la valeur de $\widehat{\mu}$ et de la somme des carrés résiduels dans ce cas. Que penser de la qualité de $\widehat{\mu}$ lorsque $a_0 = 0$, $a_k = 0$ pour $2 \leq k \leq K$ et $b_k = 0$ pour $1 \leq k \leq K$.

5. Commenter la figure au dos de la feuille, pour laquelle $\beta^\star_j \neq 0$ pour $j \leq 11$, $\beta^\star_j = 0$ pour $j > 11$ et $n = 101$. Quel phénomène observe t-on ? Comment choisir un modèle toujours bien ajusté ?


VRAIE
p= 3
p= 11
p= 81
−4
−2
0
2
4
0.0
0.2
0.4
0.6
0.8
1.0