Correction du sujet d'examen 2008-2009

Centrale - Statistiques avancées

February, 2010

1 Exerice 1

La torsion est positive et importante (queue de distribution nettement plus importante à droite qu'à gauche) et la kurtosis est positive (forme plus "piquée" que la densité gaussienne). L'écart avec la distribution gaussienne est trop marqué pour qu'on puisse utiliser le modèle linéaire : on peut penser à utiliser le bootstrap.

2 Exercice 2

correction faite en cours

3 Exerice 3

1. Soit Z_i la variable de Bernoulli égale à 1 quand la i-ème pièce inspectée est défectueuse. Soit n un entier strictement positif, et soit $(x_1, \dots, x_n)$ un n-uplet d'entiers. Pour $1 \leq k \leq n$, notons $s_k = x = 1 + \cdots + x_k$; on a alors :

$$\begin{aligned}
\mathbb{P}_\theta\left(\bigcap_{k=1}^{n}\{X_k = x_k\}\right) &= \mathbb{P}_\theta\left(\bigcap_{k=1}^{n}\bigcap_{j=s_k+1}^{s_{k+1}-1}\{X_k = 0\} \cap \{X_{s_k} = 1\}\right) \\
&= \prod_{k=1}^{n}(1-\theta)^{s_k - s_{k-1} - 1}\theta \\
&= \prod_{k=1}^{n}(1-\theta)^{x_k - 1}\theta \qquad (1) \\
&= (1-\theta)^{s_n - n}\theta^n \qquad (2)
\end{aligned}$$

   On voit sur (1) que, sous $\mathbb{P}_\theta$, les X_i sont i.i.d. de loi géométrique de paramètre θ. Cette loi a pour espérance $1/\theta$ et pour variance $(1-\theta)/\theta^2$. Dans la suite, pour l'approche bayésienne, on note $p(x_1, \dots, x_n|\theta) = \mathbb{P}_\theta\left(\bigcap_{k=1}^{n}\{X_k = x_k\}\right)$ la densité conditionnelle des observations sous le paramètre θ.

2. La loi beta$(2, 100)$ a pour espérance $2/102 \approx 0.02$ et pour variance $2 \times 100/(102^2 \times 103) \approx 2e-4$. On peut choisir comme estimateur a priori l'espérance de la loi a priori (proche de 0.02).

3. La loi jointe a pour densité :

$$\pi(\theta)p(x_1, \dots, x_n|\theta) = \frac{\theta^{n+1}(1-\theta)^{99+s_n-n}}{\beta(2,100)}\mathbb{1}_{[0,1]}(\theta).$$

   La loi de θ a posteriori est a donc pour densité :

$$\Pi(\theta|X_1, \dots, X_n) = \frac{\theta^{n+1}(1-\theta)^{99+s_n-n}}{\int_0^1 \theta^{n+1}(1-\theta)^{99+s_n-n}d\theta} = \frac{\theta^{n+2-1}(1-\theta)^{100+s_n-n-1}}{\beta(2+n, 100+s_n-n)},$$

   c'est donc la loi beta$(2+n, 100+s_n-n)$.

4. Si on choisit comme estimateur la moyenne a posteriori, on trouve

$$\hat{\theta}_\pi = \frac{2+n}{102+s_n}.$$

On a vu en (2) que la log-vraisemblance s'écrit

$$l(\theta) = \log(1-\theta)^{s_n-n}\theta^n = (s_n-n)\log(1-\theta) + n\log(\theta).$$

Il est facile de maximiser l (c'est le même calcul que pour la binômiale) : on trouve $\hat{\theta}_{ML} = n/s_n = 1/\bar{X}_n$. Par la loi forte des grands nombres, $\bar{X}_n \to \theta$ et donc $\hat{\theta}_{ML} \to \theta$. La différence vaut donc :

$$\hat{\theta}_\pi - \hat{\theta}_{ML} = \frac{2s_n - 102n}{s_n(102+s_n)} = \frac{2-102n/s_n}{102+s_n} \to 0$$

presque sûrement quand n tend vers l'infini. $\hat{\theta}_\pi$ est donc une variante légèrement biaisé, mais consistante de $\hat{\theta}_{ML}$.

4 Exercice 4

–

$$\Omega_2 = \left\{\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\right\}$$
$$\Omega_4 = \left\{\frac{1}{2}\begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\-1\\0\end{pmatrix} \frac{1}{\sqrt{2}}\begin{pmatrix}0\\1\\0\\1\end{pmatrix} \frac{1}{2}\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}\right\}$$

- Calcul très classique : $\langle \tilde{c}_k, \tilde{c}_l \rangle = 0$ et $\langle \tilde{s}_k, \tilde{s}_l \rangle = 0$ si $k \neq l$, et $\langle \tilde{c}_k, \tilde{s}_l \rangle = 0$.
- Se déduit immédiatement de la question précédente que Ω_n est une famille d'éléments normés deux à deux orthogonaux, et donc une base orthonormée de $\mathbb{R}^n$.
- Pour chaque entier $k \in \{0, \ldots, n/2-1\}$, on définit $M_k = \mathrm{Vect}(c_0, c_1, s_1, \ldots, c_k, s_k$. La projection de Y su M_k définit un estimateur $\hat{Y}_k$ qui est d'autant plus régulier (mais aussi d'autant plus biaisé) que k est petit.
  Pour trouver le bon compromis, on peut chercher à minimiser le risque quadratique moyen $\mathbb{E}[(\hat{Y}_k - f(/n))^2]$ - le critère de Mallows suggère de choisir l'indice k qui minimise $\left\|Y - \hat{Y}_k\right\|^2 + 2\dim(M_k) = \left\|Y - \hat{Y}_k\right\|^2 + 2(2k+1)$. On peut aussi penser au critère AIC ou BIC.
  Le calcul de $\hat{Y}_k$ se fait efficacement en utilisant la transformée de Fourier rapide, en mettant à 0 les coefficients correspondants aux vecteurs c_l et s_l pour $l > k$, et en prenant la transformée de Fourier inverse (c'est-à-dire, à une constante près, la transformée de Fourier rapide) du résultat.