Université de Montpellier — Année 2020-2021
M2 Biostatistique - HMMA310

**Examen terminal - vendredi 6 novembre 2020**
**Durée 2h00 - Documents autorisés**

**Exercice 1 (6 pts)**

Nous souhaitons estimer
$$\theta = \int_0^1 \exp(-x)\sin^2(x)\mathrm{d}x\,.$$

**1 (1 pt)** Proposer deux méthodes de Monte Carlo.

**2 (1 pt)** Comparer ces deux stratégies.

**3 (2 pts)** Déterminer le nombre de simulation nécessaire pour que l'erreur absolue relative correspondant à l'une des deux méthodes soit inférieure à 1% avec une probabilité supérieure ou égale à 99%.

**4 (2 pts)** Donner un intervalle de confiance asymptotique de niveau 95% pour $\theta$.

**Exercice 2 (4 pts)**

Nous considérons la variable aléatoire $X$ admettant pour densité
$$f_X(x) \propto x(1-x)^2\mathbb{I}_{[1/4,3/4]}(x)\,.$$
Proposer deux méthodes de simulation suivant $X$ et donner les codes R associés.

**Exercice 3 (5 pts)**

Nous observons un $n$-échantillon $(x_1,\ldots,x_n)$. Nous souhaitons déterminer si ces réalisations, discrètes strictement positives et idd, proviennent d'une loi de Poisson de paramètre $\theta_1 > 0$ [le modèle 1] ou d'une loi géométrique de paramètre $\theta_2 \in [0,1]$ [le modèle 2].

Nous nous plaçons dans le paradigme bayésien et supposons que
- $\pi(\theta_1)$ est une loi gamma de paramètre $(1, 0.1)$,
- $\pi(\theta_2)$ est une loi uniforme sur $[0,1]$,
- $\mathbb{P}(\mathcal{M}_1) = \mathbb{P}(\mathcal{M}_2)$.

Donner la procédure bayésienne de discrimination entre ces deux modèles et le code R associé.

**Exercice 5 (5 pts)**

Nous considérons le modèle de régression suivant

$$\mathbf{y}|\boldsymbol{\beta}, \sigma^2 \sim \mathcal{N}_n\left(\mathbf{X}\boldsymbol{\beta}, \sigma^2\,\mathbf{I}_n\right)$$

où $\mathbf{y} = (y_1, \ldots, y_n)$ est le vecteur contenant les réalisations de la variable à expliquer et $\mathbf{X}$ est la matrice des variables explicatives :

$$\mathbf{X} = [\mathbf{x}_1 \quad \ldots \quad \mathbf{x}_p] = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ x_{31} & x_{32} & \ldots & x_{3p} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{bmatrix}.$$

Nous supposons que $p > 0$ et que $\mathbf{X}$ est de plein rang.

Nous considérons la loi a priori suivante

$$\boldsymbol{\beta}|\sigma^2 \sim \mathcal{N}_p\left(0_p, \sigma^2 I_n\right) \quad ; \quad \sigma^2 \sim \mathcal{IG}(a,b) \quad a > 0 \quad \text{et} \quad b > 0\,.$$

Donner l'estimateur bayésien du vecteur $\boldsymbol{\beta} \in \mathbb{R}^p$ pour la fonction de perte quadratique.

1

$$\theta = \int_0^1 e^{-x} \sin^2(x)\, dx$$

1] $$\widehat{\theta}_1 = \frac{1}{N} \sum_{i=1}^{N} e^{-X_i} \sin^2(X_i)$$

où $X_1, \ldots, X_N \overset{iid}{\sim} \mathcal{U}_{[0,1]}$

$$\widehat{\theta}_2 = \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{[0,1]}(X_i) \sin^2(X_i)$$

où $X_1, \ldots, X_N \overset{iid}{\sim} \mathcal{E}xp(1)$

2] $$\mathbb{E}(\widehat{\theta}_1) = \mathbb{E}(\widehat{\theta}_2) = \theta$$

$$\mathbb{V}(\widehat{\theta}_1) = \frac{1}{N} \left( \mathbb{E}_{\mathcal{U}_{[0,1]}}\left[ e^{-2X} \sin^4(X) \right] - \theta^2 \right)$$

$$\mathbb{V}(\widehat{\theta}_2) = \frac{1}{N} \left( \mathbb{E}_{\mathcal{E}xp(1)}\left( \mathbb{1}_{[0,1]}(X) \sin^4(X) \right) - \theta^2 \right)$$

Il s'agit de comparer

$$a = \int_0^1 e^{-2x} \sin^4(x)\, dx \quad \text{et}$$

$$b = \int_0^1 e^{-x} \sin^4(x)\, dx$$

Clairement, $a \leq b$ et donc

$$\mathbb{V}(\widehat{\theta_1}) \leq \mathbb{V}(\widehat{\theta_2}).$$

3) On prend $\widehat{\theta_1}$.
On cherche la plus petite valeur de N telle que

$$\mathbb{P}\left(\frac{|\widehat{\theta_1} - \theta|}{|\theta|} \leq 0.01\right) \geq 0.99$$

3

D'après l'inégalité de Bienaymé-Tchebyshev nous avons

$$\mathbb{P}\left( \frac{|\hat{\theta}_1 - \theta|}{|\theta|} \leq 10^{-2} \right) \geq 1 - \frac{\mathbb{V}(\hat{\theta}_1)}{\theta^2 10^{-4}}$$

Et,

$$1 - \frac{\mathbb{V}(\hat{\theta}_1)}{\theta^2 10^{-4}} \geq 1 - 10^{-2}$$

$$\Longleftrightarrow \frac{\mathbb{V}(\hat{\theta}_1)}{\theta^2} \leq 10^{-6}$$

$$\Longleftrightarrow N \geq 10^6 \left[ \frac{\int_0^1 e^{-2x} \sin^4(x)\, dx}{\theta^2} - 1 \right]$$

$$\theta = \int_0^1 e^{-x} \sin^2(x)\, dx \qquad \forall \varepsilon > 0,\ \varepsilon < 1$$

$$\geq \int_\varepsilon^1 e^{-x} \sin^2(x)\, dx \geq \sin^2(\varepsilon)\left[e^{-\varepsilon} - e^{-1}\right]$$

4

On pourrait optimiser en $\varepsilon$ mais on prend $\varepsilon = \frac{1}{2}$ qui est une valeur raisonnable (voir contre)

Finalement,

$$\frac{1}{\theta^2} < \sin^{-4}\left(\frac{1}{2}\right)\left(e^{-1/2} - e^{-1}\right)^{-2}$$

Et,

$$\left(\frac{\int_0^1 \sin^4(x)\, e^{-2x}\, dx}{\theta^2} - 1\right)$$

$$\leq \left(\frac{1}{2}\right)\left(1 - e^{-2}\right)\sin^{-4}\left(\frac{1}{2}\right)\left(e^{-1/2} - e^{-1}\right)^{-2} - 1$$

$$\# \quad 142,68$$

$$\boxed{N \geq 10^6 \times 143}$$