

Examen médian : OS02 - Théorie de la décision - 2 h

I.Nikiforov

11 novembre 2012

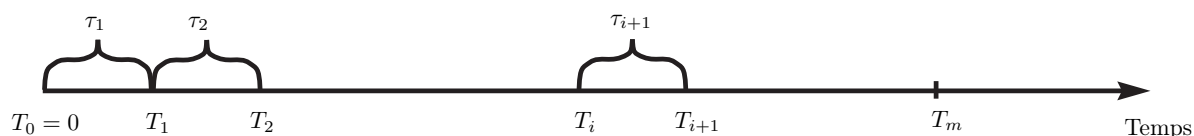
Documents autorisés : Polycopiés distribués, formulaires et notes de cours.**Sujet 1 :** [≈ 8 points] En physique ou en biologie, il nous faut très souvent calculer le nombre de particules

FIGURE 1 – Les instants d'arrivée de particules.

radioactives passant par un domaine pre-défini ou le nombre de bactéries apparaissant dans un volume donné (dans un réacteur bio-chimique, par exemple). Par fois, les expérimentateurs adoptent la stratégie suivante : ils attendent le premier instant T_m quand m premières particules ou bactéries sont comptées. L'expérimentation, donc, consiste à mesurer les instants T_1, \dots, T_m , où T_i est l'instant d'arrivée de i -ème particule/bactérié (voir la figure 1). Supposons que le temps d'attente $\tau_i = T_i - T_{i-1}$, $T_0 = 0$, est distribué selon une loi exponentielle $\Gamma(\lambda, 1)$ avec la densité définie par :

$$p(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases},$$

On considère que le paramètre $\lambda > 0$ est inconnu.

1. On considère que les «périodes d'attente» sont indépendantes. Écrire la densité commune des m premières «périodes d'attente» τ_1, \dots, τ_m .
2. Montrer que la famille de lois exponentielles $\mathcal{P} = \{\Gamma(\lambda, 1)\}_{\lambda > 0}$ admet un rapport de vraisemblance monotone pour les observations τ_1, \dots, τ_m ¹. Déterminer $T(\tau_1, \dots, \tau_m)$. Dessiner la fonction $T \mapsto g(T)$.
3. Soit τ_1, \dots, τ_m un échantillon de distribution exponentielle P_λ . Déterminer un test δ *UPP* (Uniformément le Plus Puissant) dans la classe \mathcal{K}_α pour choisir entre $\mathcal{H}_1 = \{\lambda \leq \lambda_0\}$ contre $\mathcal{H}_2 = \{\lambda > \lambda_0\}$.
4. Calculer les risques $\alpha_1 = \mathbb{P}_{\lambda_1}(\delta = \mathcal{H}_2)$ et $\alpha_2 = \mathbb{P}_{\lambda_2}(\delta = \mathcal{H}_1)$ où $\lambda_1 < \lambda_0 < \lambda_2$.

Indication : Soit $T(\tau_1, \dots, \tau_m) = \sum_{i=1}^m \tau_i$, où $\tau_i \sim \Gamma(\lambda, 1)$. Alors, $2\lambda T \sim \Gamma(\frac{1}{2}, m)$ (distribution $\chi^2(2m)$ à $2m$ degrés de liberté).

Réponses :

1. Voir la page 15 de polycopiés «OS02/OS12» pour la définition d'un rapport de vraisemblance monotone.

- 1) La densité commune des m premières «périodes d'attente» τ_1, \dots, τ_m est

$$p(x_1, \dots, x_m; \lambda) = \prod_{i=1}^m p(x_i; \lambda) = \begin{cases} \lambda^m e^{-\lambda \sum_{i=1}^m x_i} & \text{si } x_i \geq 0 \text{ pour } i \in [1, m] \\ 0 & \text{si } \exists i \in [1, m] \text{ t.q. } x_i < 0 \end{cases}$$

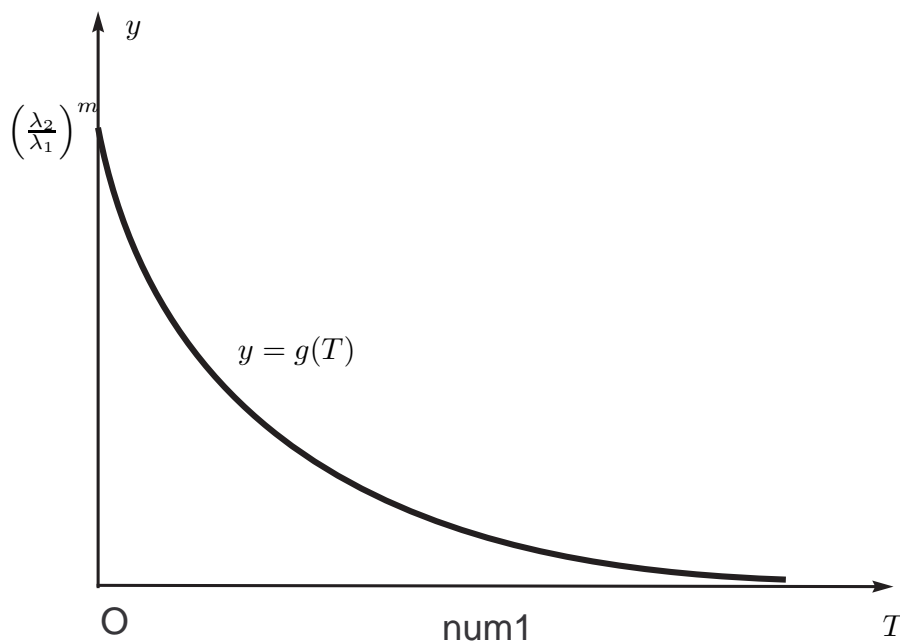


FIGURE 2 – La fonction $T \mapsto g(T)$.

- 2) Soit $T(\tau_1, \dots, \tau_m) = \sum_{i=1}^m \tau_i$. Le rapport de vraisemblance est

$$\Lambda(x_1, \dots, x_m) = \frac{p(x_1, \dots, x_m; \lambda_2)}{p(x_1, \dots, x_m; \lambda_1)} = g(T(x_1, \dots, x_m))$$

où

$$g(T) = \left(\frac{\lambda_2}{\lambda_1}\right)^m e^{-(\lambda_2 - \lambda_1)T}.$$

On peut constater que la fonction $T \mapsto g(T)$ est décroissante pour tous les $\lambda_1 > 0$ et $\lambda_2 > 0$, $\lambda_2 > \lambda_1$. La fonction $T \mapsto g(T)$ est représentée sur la figure 2. Donc, la famille de lois exponentielles admet un rapport de vraisemblance monotone pour les observations τ_1, \dots, τ_m .

- 3) Puisque la fonction $T \mapsto g(T)$ est décroissante, le test δ UPP dans la classe \mathcal{K}_α pour choisir entre $\mathcal{H}_1 = \{\lambda \leq \lambda_0\}$ contre $\mathcal{H}_2 = \{\lambda > \lambda_0\}$ est

$$\delta(\xi) = \begin{cases} \mathcal{H}_1 & \text{si } T(\tau_1, \dots, \tau_m) = \sum_{i=1}^m \tau_i > h(\alpha) \\ \mathcal{H}_2 & \text{si } T(\tau_1, \dots, \tau_m) = \sum_{i=1}^m \tau_i \leq h(\alpha) \end{cases}$$

4) Les risques α_1 et α_2 sont calculés de la façon suivante :

$$\alpha_1 = \mathbb{P}_{\lambda_1}(\delta = \mathcal{H}_2) = \mathbb{P}_{\lambda_1}(T = \sum_{i=1}^m \tau_i \leq h) = \mathbb{P}_{\lambda_1}(2\lambda_1 T \leq 2\lambda_1 h) = \int_0^{2\lambda_1 h} p(x) dx,$$

où la densité $p(x)$ de la loi $\chi^2(2m)$ est $p(x) = \frac{x^{m-1} e^{-\frac{x}{2}}}{2^m \Gamma(m)}$, $x \in \mathbb{R}_+^*$, et

$$\alpha_2 = \mathbb{P}_{\lambda_2}(\delta = \mathcal{H}_1) = \mathbb{P}_{\lambda_2}(T = \sum_{i=1}^m \tau_i > h) = \mathbb{P}_{\lambda_2}(2\lambda_2 T > 2\lambda_2 h) = 1 - \int_0^{2\lambda_2 h} p(x) dx,$$

où $\lambda_1 < \lambda_0 < \lambda_2$.

Sujet 2 : [$\simeq 6$ points]

Une loi binomiale $B(n, p)$ est définie par :

$$\mathbb{P}(\xi = k) = C_n^k p^k (1-p)^{n-k}, \quad C_n^k = \frac{n!}{k!(n-k)!},$$

où $0 \leq k \leq n$ sont entiers et $0 < p < 1$. Soit ξ un échantillon de distribution $B(n, p)$. On considère l'hypothèse de base $\mathcal{H}_1 = \{\xi \sim B(2, \frac{1}{2})\}$ et l'hypothèse alternative $\mathcal{H}_2 = \{\xi \sim B(2, \frac{2}{3})\}$.

1. Déterminer un test δ du rapport de vraisemblance (RV) pour choisir entre \mathcal{H}_1 et \mathcal{H}_2 .
2. Calculer les risques α_1, α_2 en fonction du seuil h .
3. Dessiner la courbe $\alpha_1 \mapsto \alpha_2(\alpha_1)$ en coordonnées $\alpha_1 O \alpha_2$.

Réponses :

1) Le rapport de vraisemblance est

$$\Lambda(\xi) = \frac{p_2^\xi (1-p_2)^{n-\xi}}{p_1^\xi (1-p_1)^{n-\xi}} = \frac{\left(\frac{2}{3}\right)^\xi \left(\frac{1}{3}\right)^{2-\xi}}{\left(\frac{1}{2}\right)^\xi \left(\frac{1}{2}\right)^{2-\xi}} = 4 \left(\frac{2}{3}\right)^\xi \left(\frac{1}{3}\right)^{2-\xi}$$

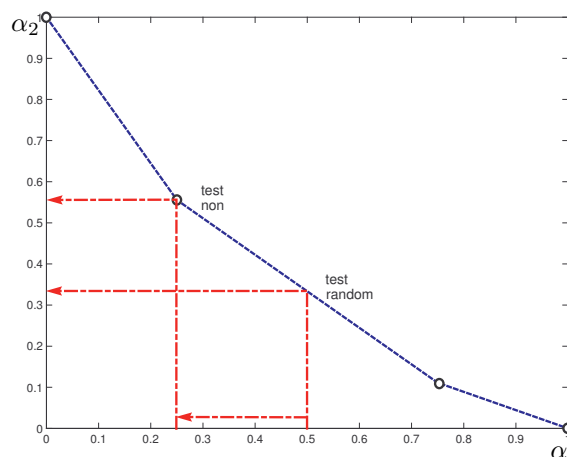
Le test δ du RV

$$\delta(\xi) = \begin{cases} \mathcal{H}_1 & \text{si } \Lambda(\xi) = 4 \left(\frac{2}{3}\right)^\xi \left(\frac{1}{3}\right)^{2-\xi} < h \\ \mathcal{H}_2 & \text{si } \Lambda(\xi) = 4 \left(\frac{2}{3}\right)^\xi \left(\frac{1}{3}\right)^{2-\xi} > h \end{cases}$$

2)

$$\mathbb{P}(\xi = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \frac{2}{k!(2-k)!} p^k (1-p)^{2-k}$$

k	$\left(\frac{2}{3}\right)^k$	$\left(\frac{1}{3}\right)^{2-k}$	$\Lambda(k)$	$\mathbb{P}_1(\xi = k)$	$\mathbb{P}_2(\xi = k)$
0	1	$\frac{1}{9}$	$\frac{4}{9}$	$\frac{1}{4}$	$\frac{1}{9}$
1	$\frac{2}{3}$	$\frac{1}{3}$	$\frac{8}{9}$	$\frac{1}{2}$	$\frac{4}{9}$
2	$\frac{4}{9}$	1	$\frac{16}{9}$	$\frac{1}{4}$	$\frac{4}{9}$

FIGURE 3 – La courbe $\alpha_1 \mapsto \alpha_2(\alpha_1)$ en coordonnées $\alpha_1 O \alpha_2$.

h	α_1	α_2
$\leq \frac{3}{9}$	1	0
$\frac{6}{9}$	$\frac{3}{4}$	$\frac{1}{9}$
$\frac{12}{9}$	$\frac{1}{4}$	$\frac{5}{9}$
$\geq \frac{18}{9}$	0	1

3) Finalement, on obtient, du dernier tableau, α_2 comme fonction de α_1 . Cette fonction est représentée sur la figure 3. La ligne brisée correspond au test randomisé et l'ensemble des points correspond au test non-randomisé.

Sujet 3 : [$\simeq 6$ points]

La densité d'une loi exponentielle double P_θ est définie par :

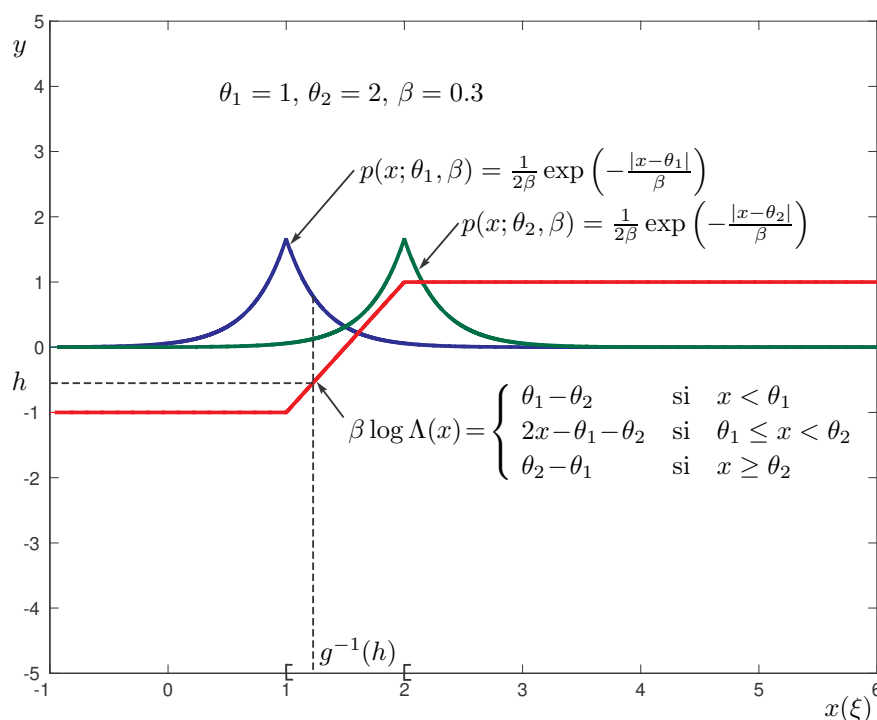
$$p(x; \theta, \beta) = \frac{1}{2\beta} \exp\left(-\frac{|x - \theta|}{\beta}\right)$$

On considère que le paramètre β est connu. On suppose maintenant que les hypothèses simples définies dans le **sujet 2** sont $\mathcal{H}_1 = \{\xi \sim P_{\theta_1}\}$ et $\mathcal{H}_2 = \{\xi \sim P_{\theta_2}\}$. Soit $\theta_1 < \theta_2$. Répondre aux questions 1-2 du **sujet 2**.

Bonus +[$\simeq 3$] **points** : Question 3 du **sujet 2**.

1) Le test δ du RV

$$\delta(\xi) = \begin{cases} \mathcal{H}_1 & \text{si } \log \Lambda(\xi) = \frac{p(\xi; \theta_2, \beta)}{p(\xi; \theta_1, \beta)} = \log \frac{\exp\left(-\frac{|\xi - \theta_2|}{\beta}\right)}{\exp\left(-\frac{|\xi - \theta_1|}{\beta}\right)} < \frac{h}{\beta} \\ \mathcal{H}_2 & \text{si } \log \Lambda(\xi) = \frac{p(\xi; \theta_2, \beta)}{p(\xi; \theta_1, \beta)} = \log \frac{\exp\left(-\frac{|\xi - \theta_2|}{\beta}\right)}{\exp\left(-\frac{|\xi - \theta_1|}{\beta}\right)} \geq \frac{h}{\beta} \end{cases},$$

FIGURE 4 – Les fonctions $x \mapsto p(x; \theta_1, \beta)$, $x \mapsto p(x; \theta_2, \beta)$ et $x \mapsto \log \Lambda(x)$.

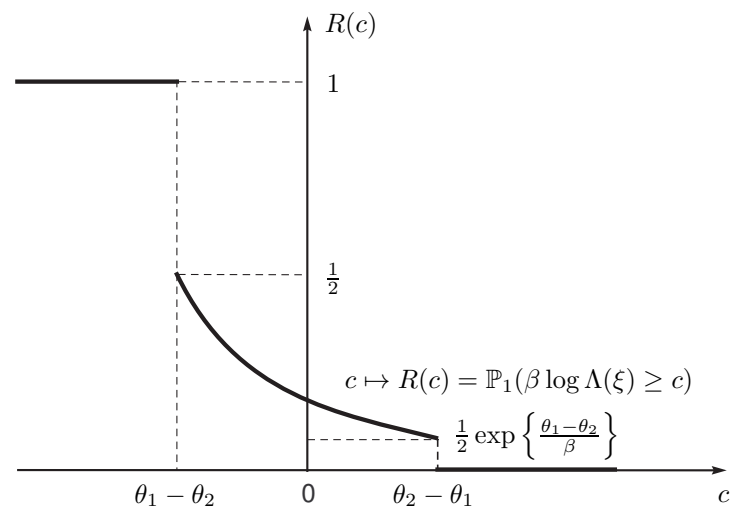
où

$$\beta \log \Lambda(\xi) = -|\xi - \theta_2| + |\xi - \theta_1| = \begin{cases} \theta_1 - \theta_2 & \text{si } \xi < \theta_1 \\ 2\xi - \theta_1 - \theta_2 & \text{si } \theta_1 \leq \xi < \theta_2 \\ \theta_2 - \theta_1 & \text{si } \xi \geq \theta_2 \end{cases}$$

est le plus puissant dans la classe \mathcal{K}_α si la fonction $c \mapsto R(c) = \mathbb{P}_1(\beta \log \Lambda(\xi) \geq c)$ est continue. Les fonctions $x \mapsto p(x; \theta_1, \beta)$, $x \mapsto p(x; \theta_2, \beta)$ et $x \mapsto \log \Lambda(x)$ sont représentées sur la figure 4.

2) La fonction $c \mapsto R(c) = \mathbb{P}_1(\beta \log \Lambda(\xi) \geq c)$ est représentée sur la figure 5. On constate que elle n'est pas continue. Donc, on ne peut pas toujours garantir l'obtention d'une valeur donnée a priori du risque $\alpha_1 \in [0; 1]$. Comme d'habitude, on reste dans la classe de tests non randomisés. D'abord nous allons traiter deux cas «pathologiques». Soit $h \leq \theta_1 - \theta_2 < 0$. Alors $\alpha_1(\delta) = \mathbb{P}_1(\beta \log \Lambda(\xi) \geq h) = 1$ et $\alpha_2(\delta) = \mathbb{P}_2(\beta \log \Lambda(\xi) < h) = 0$. Soit $h > \theta_2 - \theta_1 > 0$. Alors $\alpha_1(\delta) = \mathbb{P}_1(\beta \log \Lambda(\xi) \geq h) = 0$ et $\alpha_2(\delta) = \mathbb{P}_2(\beta \log \Lambda(\xi) < h) = 1$. Maintenant, passons au cas régulier qui correspond à l'intervalle de continuité $[\theta_1 - \theta_2; \theta_2 - \theta_1]$ de la fonction $c \mapsto R(c) = \mathbb{P}_1(\beta \log \Lambda(\xi) \geq c)$. Soit $\theta_1 - \theta_2 < h \leq \theta_2 - \theta_1$. On définit la fonction $g : x \mapsto g(x) = 2x - \theta_1 - \theta_2$ sur l'intervalle $[\theta_1; \theta_2]$. On a

$$\begin{aligned} \alpha_1(\delta) &= \mathbb{P}_1(\beta \log \Lambda(\xi) \geq h) = \int_{g^{-1}(h)}^{\infty} p(x; \theta_1, \beta) dx \\ &= \int_{\frac{h+\theta_1+\theta_2}{2}}^{\infty} \frac{1}{2\beta} \exp \left\{ \frac{-x+\theta_1}{\beta} \right\} dx = \frac{1}{2} \exp \left\{ \frac{\theta_1 - \theta_2 - h}{2\beta} \right\} \end{aligned}$$

FIGURE 5 – La fonction $c \mapsto R(c)$.

et

$$\begin{aligned}
 \alpha_2(\delta) &= \mathbb{P}_2(\beta \log \Lambda(\xi) < h) = \int_{-\infty}^{g^{-1}(h)} p(x; \theta_2, \beta) dx = \\
 &= \int_{-\infty}^{\frac{h + \theta_1 + \theta_2}{2}} \frac{1}{2\beta} \exp\left\{\frac{x - \theta_2}{\beta}\right\} dx = \frac{1}{2} \exp\left\{\frac{\theta_1 - \theta_2 + h}{2\beta}\right\}
 \end{aligned}$$

3) La fonction $\alpha_1 \mapsto \alpha_2(\alpha_1) = \frac{1}{4\alpha_1} e^{\frac{\theta_1 - \theta_2}{\beta}}$ est représentée sur la figure 6.

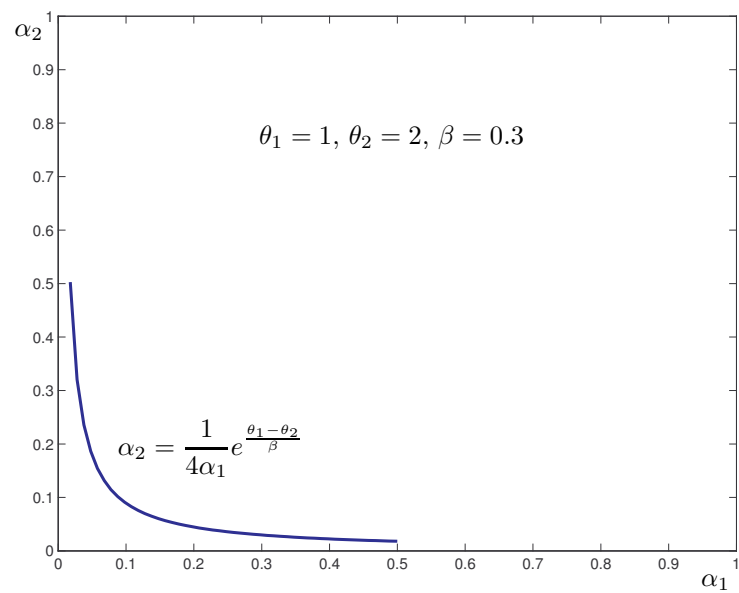


FIGURE 6 – La courbe $\alpha_1 \mapsto \alpha_2(\alpha_1)$ en coordonnées $\alpha_1 O \alpha_2$.