# Leçon 448 : Corrigé.

## Exercice 1 : Maximum de vraisemblance pour le paramètre $p$ d'une loi $(N, p)$ ([DUP] n°5.1 p.243)

On a $\Theta = ]0;1[$ et la fonction de vraisemblance est donc :

$$L_{\mathbf{x}} : \begin{array}{ccl} \Theta & \to & \mathbb{R}_+ \\ \theta & \mapsto & \displaystyle\prod_{i=1}^{n} \binom{N}{x_i} \theta^{x_i}(1-\theta)^{N-x_i} \end{array}$$

On pose ensuite $H_{\mathbf{x}} = \ln \circ L_{\mathbf{x}}$ et on a donc par stricte croissance de $\ln$ que $L_{\mathbf{x}}$ et $H_{\mathbf{x}}$ atteignent leurs extrema aux mêmes points. On a donc $\forall \theta \in \Theta$ :

$$H_{\mathbf{x}}(\theta) = \sum_{i=1}^{n} \ln \binom{N}{x_i} + \ln(\theta) \sum_{i=1}^{n} x_i + nN \ln(1-\theta) - \ln(1-\theta) \sum_{i=1}^{n} x_i$$

$H_{\mathbf{x}}$ est deux fois dérivable et on peut donc calculer $\forall \theta \in \Theta$ :

$$H'_{\mathbf{x}}(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{nN}{1-\theta} + \frac{\sum_{i=1}^{n} x_i}{1-\theta} \quad \text{d'où} \quad H'_{\mathbf{x}}(\theta) = \frac{\sum_{i=1}^{n} x_i - nN\theta}{\theta(1-\theta)}$$

$$H''_{\mathbf{x}}(\theta) = \frac{-nN\theta(1-\theta) + (nN\theta - \sum_{i=1}^{n} x_i) \times (1-2\theta)}{\theta^2(1-\theta)^2} = \frac{-nN\theta + nN\theta^2 + nN\theta - 2nN\theta^2 - \sum_{i=1}^{n} x_i + 2\theta \sum_{i=1}^{n} x_i}{\theta^2(1-\theta)^2}$$

$$\text{Soit : } H''_{\mathbf{x}}(\theta) = \frac{P(\theta)}{\theta^2(1-\theta)^2} \quad \text{avec} \quad P(\theta) = -nN\theta^2 + 2\theta \left( \sum_{i=1}^{n} x_i \right) - \sum_{i=1}^{n} x_i$$

On calcule le discriminant de la fonction polynomiale $P$ :

$$\Delta = 4 \left( \sum_{i=1}^{n} x_i \right)^2 - 4nN \sum_{i=1}^{n} x_i = 4 \left( \sum_{i=1}^{n} x_i \right) \left[ \sum_{i=1}^{n} x_i - nN \right]$$

Or $\forall i \in [\![1;n]\!], 0 \leqslant x_i \leqslant N$, donc $\left[ \sum_{i=1}^{n} x_i - nN \right] \leqslant 0$ et donc $\Delta \leqslant 0$. Donc $P(\theta)$ est de signe constant sur $\mathbb{R}$. Or son terme de plus haut degré est $-nN\theta^2 < 0$, donc $P(\theta) \leqslant 0$. Donc $H_{\mathbf{x}}$ est concave sur $]0;1[$, et il suffit donc de trouver un point critique (où sa dérivée s'annule) pour trouver ses maxima globaux :

$$X'_{\mathbf{x}}(\theta) = 0 \Leftrightarrow \sum_{i=1}^{n} x_i = nN\theta \Leftrightarrow \theta = \frac{\sum_{i=1}^{n} x_i}{nN} = \frac{\overline{\mathbf{x}}_n}{N}$$

L'estimation de $p$ par le maximum de vraisemblance est donc $\boxed{\widehat{p}_{MV} = \frac{\overline{\mathbf{x}}_n}{N}}$.

Ce résultat est cohérent avec l'intuition car la loi des grands nombres assure que pour $n$ assez grand, $\frac{\overline{\mathbf{x}}_n}{N} \approx \frac{\mathbb{E}[X]}{N} = \frac{Np}{N} = p$.

## Exercice 2 : Estimation du nombre de Panzers ([DUP] n°5.5 p.243)

**1.a** Les tirages étant effectués avec remise, on a : $\forall k \in [\![1;n]\!], X_k \sim \mathcal{U}([\![1,N]\!])$ et donc $\mathbb{E}[\overline{X}_n] = \mathbb{E}[X_k] = \frac{N+1}{2}$.
On pase donc :

$$\boxed{T_n = 2\overline{X}_n - 1}$$

On vérifie qu'il s'agit bien d'un estimateur sans biais :

$$\mathbb{E}[T_n] = 2\mathbb{E}[\overline{X}_n] - 1 = 2\mathbb{E}[X_n] - 1 = 2 \times \frac{N+1}{2} - 1 = N$$

**1.b** Puisqu'il s'agit d'un estimateur sans biais, son RQM est égal à sa variance et donc :

$$r_N(T_n) = Var(T_n) = 4Var(\overline{X}_n) = \frac{4Var(X_k)}{n} = \frac{N^2-1}{3n} \xrightarrow[n \to +\infty]{} 0$$

Il s'agit bien d'un estimateur correct.

**2.a** $\forall k \in \mathbb{N}$ :

$$\mathbb{P}(M_n \leqslant k) = \mathbb{P}(\max(X_1, \dots, X_n) \leqslant k) = \mathbb{P}\left(\bigcap_{i=1}^{n}(X_i \leqslant k)\right) = \prod_{i=1}^{n}\mathbb{P}(X_i \leqslant k) = \mathbb{P}(X_1 \leqslant k)^n \text{ car les } X_i \text{ sont indépendantes}$$

$$\Longrightarrow \forall k \in \mathbb{N},\ \mathbb{P}(M_n \leqslant k) = \begin{cases} 0 & \text{si } k < 0 \\ \left(\dfrac{k}{N}\right)^n & \text{si } k \in [\![2; n]\!] \\ 1 & \text{si } k > N \end{cases}$$

**2.b** On a :

$$\mathbb{E}[Y] = \sum_{k=1}^{N} k\mathbb{P}(Y = k) = \sum_{k=1}^{N} k\left(\mathbb{P}(Y \geqslant k) - \mathbb{P}(Y \geqslant k+1)\right) = \sum_{k=1}^{N} k\mathbb{P}(Y \geqslant k) - \sum_{k=1}^{N} k\mathbb{P}(Y \geqslant k+1)$$

$$\mathbb{E}[Y] = \sum_{k=1}^{N} k\mathbb{P}(Y \geqslant k) - \sum_{k=2}^{N+1} (k-1)\mathbb{P}(Y \geqslant k) = \mathbb{P}(Y \geqslant 1) + \sum_{k=2}^{N} k\mathbb{P}(Y \geqslant k) - N\underbrace{\mathbb{P}(Y \geqslant N+1)}_{=0}$$

$$\Longrightarrow \mathbb{E}[Y] = \sum_{k=1}^{N} \mathbb{P}(Y \geqslant k)$$

**2.c** En appliquant le résultat précédent on a :

$$\mathbb{E}[M_n] = \sum_{k=1}^{N} \mathbb{P}(M_n \geqslant k) = \sum_{k=0}^{N-1} 1 - \mathbb{P}(M_n < k) = N - \sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n$$

Or dans $\displaystyle\sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n$ on reconnaît une somme de Riemann et donc $\displaystyle\sum_{k=0}^{N-1}\left(\frac{k}{N}\right)^n \leqslant \int_0^1 x^n\,dx = \frac{1}{n+1}$.
On obtient bien le résultat recherché :

$$\mathbb{E}[M_n] \geqslant N - \frac{N}{n+1}$$

**2.d** On a l'encadrement suivant :

$$N - \frac{N}{n+1} \leqslant \mathbb{E}[M_n] \leqslant N$$

Or $N - \dfrac{N}{n+1} \xrightarrow[n\to+\infty]{} N$ donc d'après le théorème des gendarmes on a $\mathbb{E}[M_n] \xrightarrow[n\to+\infty]{} N$, et donc $M_n$ est un estimateur asymptotiquement sans biais de $N$.

**3** On note $t_5$ et $m_5$ les réalisations respectives de $T_5$ et $M_5$ pour l'échantillon d'observations $\mathbf{x} = (12, 19, 28, 52, 64)$. On a alors :

$$t_5 = 2 \times \frac{12+19+28+52+64}{5} - 1 = \mathbf{69} \quad \text{et} \quad m_5 = \max(12, 19, 28, 52, 64) = \mathbf{64}$$

## Exercice 3 : Approximation de $\pi$ par des méthodes de Monte-Carlo ([KAR] D8 p.358 & [DUP] n°5.3 p.243)

Voir développement séparé !

## Exercice 4 : Fiabilité d'un vaccin ([BDH] n°5 p.194)

On désigne par $p$ le pourcentage de sujets non immunisés dans la population vaccinée. $X_i \sim \mathcal{B}(p)$ vaut 1 si la $i$-ème personne n'est pas vaccinée, 0 sinon. On a avec le TCL l'intervalle de confiance suivant :

$$\left[\overline{X}_n - z_\gamma\sqrt{\frac{p(1-p)}{n}}; \overline{X}_n + z_\gamma\sqrt{\frac{p(1-p)}{n}}\right]$$

Notre condition devient donc :

$$z_\gamma\sqrt{\frac{p(1-p)}{n}} = 0{,}01$$

Avec $z_{0,95} \approx 1{,}96$ on a $n \approx 38\,416p(1-p)$.

Or $p \in [0,1;0,15]$ d'après l'énoncé et $p \mapsto p(1-p)$ est croissante sur cet intervalle, $\forall p \in [0,1;0,15]$, $p(1-p) \leqslant 0,15(1-0,15) = 0,1275$. Il suffit donc d'avoir :

$$\boxed{n \geqslant 38\ 416 \times 0,1275 = 4\ 899}$$

D'une manière générale pour une incertitude de $\pm\Delta$ et un niveau de confiance $\gamma$ on doit avoir $z_\gamma \sqrt{\dfrac{p(1-p)}{n}} = \Delta$, soit $n = \dfrac{z_\gamma^2\, p(1-p)}{\Delta^2}$.

Comme on ne dispose plus d'information *a priori* sur $p$, on peut simplement majorer $p(1-p)$ par $\frac{1}{2}$ (maximum de $p \mapsto p(1-p)$) d'où la relation :

$$\boxed{n \geqslant \frac{z_2}{4\Delta^2}}$$

## Exercice 5 : Test paramétrique pour une loi de Bernoulli ([DUP] n°7.1 p.316)

On veut tester l'hypothèse $(H_0)$ : « la molécule est efficace », ce qui revient à estimer le paramètre $p$ d'une loi $\mathcal{B}(p)$ et ainsi l'hypothèse nulle devient :

$$(H_0) : p \geqslant 0,99$$

Sur un échantillon d'observations $\mathbf{x}$ de taille $n = 10\ 000$, on a observé une fréquence de bactéries éliminées de $\overline{\mathbf{x}}_n = 0,995$. Si on utilise un intervalle de confiance simplifié on obtient donc avec un seuil de confiance supérieur à 95 % :

$$p \in \left[0,995 - \frac{1}{\sqrt{10\ 000}}; 0,995 + \frac{1}{\sqrt{10\ 000}}\right] = [0,985; 1,005]$$

On peut donc seulement affirmer que $p \geqslant 0,985$ avec un niveau de confiance supérieur à 95 %, ce qui ne permet pas d'accepter l'hypothèse $(H_0)$ en l'état.

On utilise donc plutôt un intervalle de confiance plus fin, toujours au niveau de confiance 95 % :

$$p \in \left[\overline{\mathbf{x}}_n - z_\gamma \sqrt{\frac{\overline{\mathbf{x}}_n(1-\overline{\mathbf{x}}_n)}{n}}; \overline{\mathbf{x}}_n + z_\gamma \sqrt{\frac{\overline{\mathbf{x}}_n(1-\overline{\mathbf{x}}_n)}{n}}\right] = [0,9936; 0,9964]$$

On peut donc accepter l'hypothèse $(H_0)$ avec un niveau de confiance de 95 %.