TD20-Estimations

Exercice 1.

1. (a) Il s'agit bien d'un estimateur car son expression ne dépend pas de θ. Par stabilité par somme des lois de Poisson, on sait que $X_1 + \cdots + X_n$ suit une loi de Poisson de paramètre $n\theta$.

D'après le théorème de transfert, $e^{-\overline{X}_n}$ possède donc une espérance si et seulement si la série $\sum_{k\geq 0} e^{\frac{-k}{n}} \frac{e^{-n\theta}(n\theta)^k}{k!}$ converge absolument.

Or :

$$\sum_{k\geq 0} e^{\frac{-k}{n}} \frac{e^{-n\theta}(n\theta)^k}{k!} = e^{-n\theta} \sum_{k\geq 0} \frac{(n\theta e^{-\frac{1}{n}})^k}{k!}.$$

Donc cette série converge (absolument car à termes positifs). On en déduit que, pour tout $\theta > 0$, $e^{-\overline{X}_n}$ possède une espérance donnée par :

$$E_\theta\left(e^{-\overline{X}_n}\right) = e^{-n\theta} \sum_{k=0}^{+\infty} \frac{(n\theta e^{-\frac{1}{n}})^k}{k!} = e^{-n\theta} e^{n\theta e^{-\frac{1}{n}}} = e^{n\theta(e^{-\frac{1}{n}}-1)}.$$

Or, on sait que :

$$e^{-\frac{1}{n}} - 1 \sim -\frac{1}{n}$$

d'où :

$$\lim_{n\to+\infty} E_\theta\left(e^{-\overline{X}_n}\right) = e^{-\theta}.$$

Ainsi $e^{-\overline{X}_n}$ est un estimateur asymptotiquement sans biais de $e^{-\theta}$.

(b) On cherche à calculer :

$$E_\theta\left(\left(e^{-\overline{X}_n} - e^{-\theta}\right)^2\right).$$

Pour tout $\theta > 0$ on a (sous réserve d'existence) :

$$\begin{aligned} E_\theta\left(\left(e^{-\overline{X}_n} - e^{-\theta}\right)^2\right) &= E_\theta\left(e^{-2\overline{X}_n} - 2e^{-\theta}e^{-\overline{X}_n} + e^{-2\theta}\right) \\ &= E_\theta\left(e^{-2\overline{X}_n}\right) - 2e^{-\theta}E_\theta\left(e^{-\overline{X}_n}\right) + e^{-2\theta}. \end{aligned}$$

Un calcul similaire à celui de la question précédente donne :

$$E_\theta\left(e^{-2\overline{X}_n}\right) = e^{n\theta(e^{-\frac{2}{n}}-1)}.$$

Ainsi :

$$E_\theta\left(\left(e^{-\overline{X}_n} - e^{-\theta}\right)^2\right) = e^{n\theta(e^{-\frac{2}{n}}-1)} - 2e^{-\theta}e^{n\theta(e^{-\frac{1}{n}}-1)} + e^{-2\theta}.$$

(c) Soit $\varepsilon > 0$. D'après l'inégalité de Markov, on a :

$$\begin{aligned} 0 \leq P_\theta(|e^{-\overline{X}_n} - e^{-\theta}| > \varepsilon) &= P_\theta((e^{-\overline{X}_n} - e^{-\theta})^2 > \varepsilon^2) \\ &\leq \frac{E_\theta\left(\left(e^{-\overline{X}_n} - e^{-\theta}\right)^2\right)}{\varepsilon^2}. \end{aligned}$$

Or, d'après la question précédente on sait que :

$$E_\theta\left(\left(e^{-\overline{X}_n} - e^{-\theta}\right)^2\right) = e^{n\theta(e^{-\frac{2}{n}}-1)} - 2e^{-\theta}e^{n\theta(e^{-\frac{1}{n}}-1)} + e^{-2\theta}$$

avec $\lim_{n\to+\infty} e^{n\theta(e^{-\frac{2}{n}}-1)} - 2e^{-\theta}e^{n\theta(e^{-\frac{1}{n}}-1)} + e^{-2\theta} = e^{-2\theta} - 2e^{-\theta}e^{-\theta} + e^{-\theta} = 0.$

Par encadrement on en déduit donc pour tout $\theta > 0$:

$$\lim_{n\to+\infty} P_\theta(|e^{-\overline{X}_n} - e^{-\theta}| > \varepsilon) = 0$$

C'est estimateur convergent de $e^{-\theta}$.?

Exercice 2.

1. Soit $\theta \in]0,1[$. D'après le théorème de transfert, X possède une espérance si et seulement si la série $\sum_{k\geq 0} kP_\theta(X = k)$ converge absolument. Comme il s'agit d'une série à

termes positifs, il suffit de montrer la convergence. Pour tout $n \in \mathbb{N}$, on a :

$$\begin{aligned}\sum_{k=0}^{n} kP_\theta(X=k) &= (1-\theta)^2 \sum_{k=0}^{n} k(k+1)\theta^k \\ &= (1-\theta)^2 \sum_{i=1}^{n+1} i(i-1)\theta^{i-1} \\ &= (1-\theta)^2\theta \sum_{i=0}^{n+1} i(i-1)\theta^{i-2}.\end{aligned}$$

On reconnaît la suite des sommes partielles d'une série géométrique dérivée d'ordre 2 de raison $\theta \in]0,1[$, donc convergente. Ainsi, la série converge et X possède une espérance donnée par :

$$E_\theta(X) = (1-\theta)^2\theta \sum_{i=0}^{+\infty} i(i-1)\theta^{i-2} = (1-\theta)^2\theta \frac{2}{(1-\theta)^3} = \frac{2\theta}{1-\theta}.$$

2. Soit $(x_1, \ldots, x_n)$ une réalisation d'un n-échantillon $(X_1, \ldots, X_n)$ de la loi de X.
   (a) Soit $\theta \in]0,1[$. On a :

$$\begin{aligned}\mathcal{L}_n(\theta) &= P_\theta(X_1 = x_1, \ldots, X_n = x_n) \\ &= P_\theta(X_1 = x_1) \times \cdots \times P(X_n = x_n) \quad \text{(indépendance)} \\ &= (1-\theta)^{2n}\theta^s \prod_{i=1}^{n}(x_i+1)\end{aligned}$$

   où $s = x_1 + \cdots + x_n$.

   (b) On pose $\ell_n = \ln \circ \mathcal{L}_n$:

$$\forall \theta \in]0,1[, \quad \ell_n(\theta) = 2n\ln(1-\theta) + s\ln(\theta) + \sum_{k=1}^{n} \ln(x_k+1).$$

   Cette fonction est dérivable sur $]0,1[$ et on a :

$$\ell_n'(\theta) = \frac{-2n}{1-\theta} + \frac{s}{\theta} = \frac{-\theta(s+2n)+s}{\theta(1-\theta)}.$$

   On en déduit
   - si $s \neq 0$:

θ	0		$\frac{s}{s+2n}$		1
Signe de $\ell_n'(\theta)$		$+$	0	$-$	
Variations de ℓ_n		$\nearrow$		$\searrow$	

   - si $s = 0$, ℓ_n est strictement décroissante sur $]0,1[$ et ne possède pas de maximum.

   Comme la fonction ln est strictement croissante, les maxima de ℓ_n sont les maxima de $\mathcal{L}_n$. On en déduit :
   - si $s \neq 0$: l'estimateur du maximum de vraisemblance est :

$$\frac{s}{s+2n} = \frac{\overline{X}_n}{\overline{X}_n + 2};$$

   - si $s = 0$, il n'y a pas d'estimateur du maximum de vraisemblance.

Exercice 3. Soit X une variable aléatoire suivant la loi uniforme sur $[0,\theta]$ où $\theta \in \mathbb{R}_+^*$ est un paramètre inconnu. Soit $(X_1, \ldots, X_n)$ un n-échantillon de la loi de X. On considère les estimateurs suivants :

$$T_n = 2\overline{X}_n \quad ; \quad T_n' = \max(X_1, \ldots, X_n) \quad ; \quad T_n'' = \frac{n+1}{n} T_n'.$$

1. (a) Montrer que T_n est un estimateur sans biais de θ.
   (b) Déterminer son risque quadratique.
   (c) L'estimateur T_n est-il un estimateur convergent de θ ?
2. (a) Déterminer le biais et le risque quadratique de T_n'.
   (b) Donner un équivalent simple de $r_\theta(T_n')$ quand n tend vers $+\infty$.
3. Mêmes questions pour T_n''.
4. Quel est le meilleur estimateur de θ ?

Exercice 4. La durée de vie d'une lampe est modélisée par une variable aléatoire réelle suivant une loi exponentielle de paramètre λ.
On cherche à estimer la durée de vie moyenne $\frac{1}{\lambda}$ de la lampe.
On prélève un échantillon de n lampes et on note $X_1, \ldots, X_n$ leurs durées de vie.

1. D'après le cours, $\overline{X}_n$ est un estimateur sans biais de l'espérance (donc de $\frac{1}{\lambda}$) dont le risque quadratique est :

$$r(\overline{X}_n) = \frac{1}{n\lambda^2}.$$

2. On pose $Y_n = \min(X_1, \ldots, X_n)$.
   (a) La variable Y_n suit la loi $\mathcal{E}(n\lambda)$ (voir l'exercice 15 du TD 15).
   (b) Le biais de nY_n est :

$$E(nY_n) - \frac{1}{\lambda} = n \times \frac{1}{n\lambda} - \frac{1}{\lambda} = 0.$$

   Ainsi nY_n est un estimateur sans biais de $\frac{1}{\lambda}$.

(c) Son risque quadratique est :

$$r(nY_n) = E\left(\left(nY_n - \frac{1}{\lambda}\right)^2\right) = V(nY_n) = n^2V(Y_n) = n^2\frac{1}{n^2\lambda^2} = \frac{1}{\lambda^2}.$$

3. Le premier est meilleurs car son risque quadratique est plus faible.

Exercice 5 (Estimateur sans biais du paramètre d'une loi géométrique)**.**

1. Comme pour tout $i \in \{1,\dots,n\}$, $X_i(\Omega) = \mathbb{N}^*$ donc $S_n(\Omega) = \{n, n+1, \dots\}$. Soit k un entier supérieur ou égal à n. L'événement $[S_n = k]$ est réalisé si et seulement si la somme des n valeurs de X_1,…,X_n vaut k : une issue correspond donc à une partition de k en n éléments. Il y a donc $\binom{k-1}{n-1}$ issues. Chacune de ces issues $[T_1 = x_1, \dots, T_n = x_n]$ (avec donc $x_1 + \cdots + x_n = k$) se réalise, par indépendance, avec la probabilité :

$$p(1-p)^{x_1-1}\dots p(1-p)^{x_n-1} = p^n(1-p)^{k-n}.$$

Donc finalement :

$$P(S_n = k) = \binom{k-1}{n-1}p^nq^{k-n}.$$

2. (a) La variable T_n suit une loi géométrique de paramètre p donc $P(T_n = 1) = p$. Soit $k \geq n$.

$$\begin{aligned}
P_{[S_n=k]}(T_n = 1) &= \frac{P(S_n = k, T_n = 1)}{P(S_n = k)} \\
&= \frac{P(T_2 + \cdots + T_n = k-1, T_n = 1)}{P(S_n = k)} \\
&= \frac{P(T_2 + \cdots + T_n = k-1)P(T_n = 1)}{P(S_n = k)} \quad \text{par indépendance} \\
&= p\frac{P(T_2 + \cdots + T_n = k-1)}{P(S_n = k)} \\
&= p\frac{P(S_{n-1} = k-1)}{P(S_n = k)}
\end{aligned}$$

car $T_2 + \cdots + T_n$ suit la loi de S_{n-1}. Ainsi :

$$\begin{aligned}
P_{[S_n=k]}(T_n = 1) &= p\frac{P(S_{n-1} = k-1)}{P(S_n = k)} \\
&= p\frac{\binom{k-2}{n-2}p^{n-1}q^{k-n}}{\binom{k-1}{n-1}p^nq^{k-n}} \\
&= \frac{(k-2)!(k-n)!(n-1)!}{(k-1)!(n-2)!(k-n)!} \\
&= \frac{n-1}{k-1}.
\end{aligned}$$

(b) Soit $N \in \mathbb{N}$ et posons $n = N + 2$. D'après la formule des probabilités totales appliquées avec le système complet d'événements $([S_n = k])_{k \geq n}$ on a :

$$\begin{aligned}
p = P(T_n = 1) &= \sum_{k=n}^{+\infty} P_{[S_n=k]}(T_n = 1)P(S_n = k) \\
&= \sum_{k=n}^{+\infty} \frac{n-1}{k-1}\binom{k-1}{n-1}p^nq^{k-n} \\
&= \sum_{k=n}^{+\infty} \frac{n-1}{k-1}\frac{(k-1)!}{(n-1)!(k-n)!}p^nq^{k-n} \\
&= \sum_{k=n}^{+\infty} \frac{(k-2)!}{(n-2)!(k-n)!}p^nq^{k-n} \\
&= \sum_{k=n}^{+\infty} \binom{k-2}{n-2}p^nq^{k-n} \\
&= \sum_{i=n-2}^{+\infty} \binom{i}{n-2}p^nq^{i+2-n} \\
&= \sum_{i=N}^{+\infty} \binom{i}{N}p^nq^{i-N}
\end{aligned}$$

où on a poser $i = k - 2$. Ainsi :

$$\sum_{i=N}^{+\infty} \binom{i}{N}q^{i-N} = \frac{1}{(1-q)^{N+1}} = \frac{1}{p^{n-1}} = \frac{1}{p^{N+1}} = \frac{1}{(1-p)^{N+1}}.$$

3. D'après le théorème de transfert, U_n possède une espérance si et seulement si la série $\sum_{k \geq n} \frac{n-1}{k-1}P(S_n = k)$ converge absolument. Comme tout est positif, il suffit de vérifier qu'elle est convergente. Or, on a :

$$\sum_{k \geq n} \frac{n-1}{k-1}P(S_n = k) = \sum_{k \geq n} \frac{n-1}{k-1}\binom{k-1}{n-1}p^nq^{k-n}.$$

D'après un calcul précédent, cette série converge et sa somme vaut p. Ainsi U_n possède une espérance qui vaut p et donc $U_n = \dfrac{n-1}{S_n - 1}$ est un estimateur sans biais de p.

4. On suppose dans cette question que n est supérieur ou égal à 3.

(a) La variable aléatoire $V_n = \dfrac{(n-1)^2}{(S_n-1)(S_n-2)}$ possède une espérance si et seulement si la série $\sum_{k \geq n} \dfrac{(n-1)^2}{(k-1)(k-2)}P(S_n = k)$ converge (absolument mais tout est positif). Or on a :

$$\begin{aligned}
\sum_{k\geq n} \frac{(n-1)^2}{(k-1)(k-2)} P(S_n=k) &= \sum_{k\geq n} \frac{(n-1)^2}{(k-1)(k-2)} \binom{k-1}{n-1} p^n q^{k-n} \\
&= \sum_{k\geq n} \frac{(n-1)^2}{(k-1)(k-2)} \frac{(k-1)!}{(n-1)!(k-n)!} p^n q^{k-n} \\
&= \sum_{k\geq n} (n-1) \frac{(k-3)!}{(n-2)!(k-n)!} p^n q^{k-n} \\
&= \frac{n-1}{n-2} \sum_{k\geq n} \frac{(k-3)!}{(n-3)!(k-n)!} p^n q^{k-n} \\
&= \frac{n-1}{n-2} p^n \sum_{k\geq n} \binom{k-3}{n-3} q^{k-n}.
\end{aligned}$$

Or la série $\sum_{k\geq n-3} \binom{k-3}{n-3} q^{k-n}$ converge d'après une question précédente. Ainsi V_n possède une espérance donnée par :

$$\begin{aligned}
E(V_n) = \frac{n-1}{n-2} p^n \sum_{k=n}^{+\infty} \binom{k-3}{n-3} q^{k-n} &= \frac{n-1}{n-2} p^n \sum_{i=n-3}^{+\infty} \binom{i}{n-3} q^{i+3-n} \\
&= \frac{n-1}{n-2} p^n \times \frac{1}{p^{n-2}} \\
&= \frac{(n-1)p^2}{n-2}.
\end{aligned}$$

(b) On a : $U_n^2 \leq V_n$ presque sûrement. Comme les deux variables sont positives et que V_n possède une espérance, alors U_n^2 aussi et :

$$E(U_n^2) \leq E(V_n).$$

Comme $E(U_n) = p$ alors :

$$\begin{aligned}
r(U_n) = E((U_n-p)^2) &= V(U_n) \\
&= E(U_n^2) - E(U_n)^2 \\
&\leq \frac{n-1}{n-2} p^2 - p^2 \\
&\leq \frac{n-1-(n-2)}{n-2} p^2 \\
&\leq \frac{p^2}{n-2}.
\end{aligned}$$

Exercice 6.

1.
   - La fonction f est continue sur $\mathbb{R} - \{R\}$, positive ou nulle sur $\mathbb{R}$.
   - L'intégrale $\int_{-\infty}^{+\infty} f$ est impropre en $\pm\infty$ et on a :

   $$\int_{-\infty}^{0} f = \int_{R}^{+\infty} f = 0 \quad \text{et} \quad \int_0^R f = \int_0^R \frac{2t}{R^2} dt = \left[\frac{t^2}{R^2}\right]_0^R = 1 - 0 = 1.$$

   Donc $\int_{-\infty}^{+\infty} f$ converge et vaut 1.

   Ainsi, f est une densité de probabiltié.

2.
   - Pour $x < 0$ on a $F_X(x) = \int_{-\infty}^{x} 0 = 0$

     Pour $x > R$ on a $F_X(x) = \int_{-\infty}^{0} 0 + \int_0^R \frac{2t}{R^2} dt + \int_R^x 0 = 1$
   - Pour tout réel x de $[0\,;R]$, $F_X(x) = \int_{-\infty}^{0} 0 + \int_0^R \frac{2t}{R^2} dt = 0 + \left[\frac{t^2}{R^2}\right]_0^x = \frac{x^2}{R^2}$.

3. (a) La variable X admet une espérance si $\int_{-\infty}^{+\infty} t f(t)\, dt$ converge. Elle est impropre en $\pm\infty$ et on a :

   $$\int_{-\infty}^{0} t f(t)\, dt = 0 = \int_{R}^{+\infty} t f(t)\, dt$$

   et

   $$\int_0^R t f(t) = \int_0^R \frac{2t^2}{R^2} dt = \left[\frac{2t^3}{3R^2}\right]_0^R = \frac{2}{3} R.$$

   Ainsi $\int_{-\infty}^{+\infty} t f(t)\, dt$ converge et vaut $\frac{2}{3}R$. Finalement X a une espérance et $E(X) = \frac{2}{3}R$.

   (b) Comme X a une espérance, elle a une variance si et seulement si X^2 a une espérance.

   Donc si $\int_{-\infty}^{+\infty} t^2 f(t)\, dt$ converge absolument. (théorème de transfert) ce qui équivaut à la convergence simple, tout étant positif.

   Comme précédemment :

   $$\int_{-\infty}^{+\infty} t^2 f(t)\, dt = \int_0^R t^2 f(t)\, dt = \left[\frac{2t^4}{4R^2}\right]_0^R = \frac{R^2}{2}$$

   Donc $E(X^2) = \frac{R^2}{2}$ et ainsi :

   $$V(X) = E\left(X^2\right) - E(X)^2 = \frac{1}{2}R^2 - \frac{4}{9}R^2 = \frac{1}{18}R^2.$$

4. La fonction T_n est bien fonction du n-échantillon $(X_1,\ X_2 \dots X_n)$ dont l'expression est indépendante de R.
Comme les X_i ont une espérance, par linéarité, T_n aussi et on a :

$$E(T_n) = \frac{3}{2n}\sum_{k=1}^{n} E(X_k) = \frac{3}{2n} n \frac{2}{3} R = R.$$

Donc T_n est un estimateur sans biais de R. Son risque quadratique est :

$$\begin{aligned} r(T_n) = V(T_n) + b^2 = V(T_n) \\ &= V\left(\frac{3}{2n}\sum_{k=1}^{n} E(X_k)\right) \\ &= \frac{4}{9} V\left(\sum_{k=1}^{n} E(X_k)\right) \\ &= \frac{4}{9}\sum_{k=1}^{n} V(X_k) \quad \text{(indépendance)} \\ &= \frac{9}{4n^2} n \frac{R^2}{18} = \frac{R^2}{8n}. \end{aligned}$$

5. (a) Comme les $(X_k)_k$ sont indépendantes :

$$\mathrm{P}\left[(X_1 \le x) \cap (X_2 \le x) \cap \cdots \cap (X_n \le x)\right] = \mathrm{P}(X_1 \le x) \dots \mathrm{P}(X_n \le x).$$

Donc, pour tout x réel, $\mathrm{P}(M_n \le x) = (F_X(x))^n$.
La fonction de répartition de M_n est donc la fonction G donnée par :

$$G(x) = \begin{cases} 0 \text{ si } x < 0 \\ \dfrac{x^{2n}}{R^{2n}} \text{ si } x \in [0; R] \\ 1 \text{ si } x > R \end{cases}$$

On vérifie les critère de fonction de répartition de variable à densité. On a $G(x) = (F_X(x))^n$ pour tout x réel.
Comme F_X est une fonction de répartition de variable à densité, alors F_X est continue sur $\mathbb{R}$ et de classe C^1 sur $\mathbb{R} - \{R\}$ (là où f est continue) donc G est continue sur $\mathbb{R}$ et C^1 sur $\mathbb{R} - \{R\}$ comme composée de telles fonctions.
Donc G est bien une fonction de répartition de variable à densité.

(b) Une densité est alors donnée par g_n définie par

$$g_n(x) = G'(x) = \begin{cases} 0 & \text{si } x < 0 \\ 2n\dfrac{x^{2n-1}}{R^{2n}} & \text{si } x \in [0; R] \\ 0 & \text{si } x > R \end{cases}$$

(c) Comme $\displaystyle\int_{-\infty}^{+\infty} t\, g_n(t)\, dt$ est nulle en dehors de $[0, R]$ alors M_n a une espérance et :

$$\begin{aligned} E(M_n) &= \int_0^R t\, g_n(t)\, dt = \int_0^R 2n \frac{t^{2n}}{R^{2n}} dt \\ &= 2n\left[\frac{1}{2n+1}\frac{t^{2n+1}}{R^{2n}}\right]_0^R = \frac{2n}{2n+1}\frac{R^{2n+1}}{R^{2n}} \\ &= \frac{2n}{2n+1} R. \end{aligned}$$

De même M_n^2 a une espérance et

$$\begin{aligned} E\left(M_n^2\right) &= \int_0^R t^2 g_n(t)\, dt = \int_0^R 2n \frac{t^{2n+1}}{R^{2n}} dt \\ &= 2n\left[\frac{1}{2n+2}\frac{t^{2n+1}}{R^{2n}}\right]_0^R = \frac{n}{n+1} R^2 \end{aligned}$$

Ainsi M_n possède une variacen donnée par :

$$V(M_n) = E\left(M_n^2\right) - E(M_n)^2 = \frac{n}{n+1}R^2 - \left(\frac{2n}{2n+1}\right)^2 R^2 = \frac{n\,R^2}{(n+1)(2n+1)^2}.$$

(d) On cherche à estimer R avec M_n :
Le biais de M_n est

$$b(M_n) = E(M_n) - R = \frac{2n}{2n+1}R - R = -\frac{1}{2n+1}R$$

et son risque quadratique est

$$\begin{aligned} r(M_n) &= V(M_n) + b(M_n)^2 = \frac{n}{(n+1)(2n+1)^2}R^2 + \frac{1}{(2n+1)^2}R^2 \\ &= \frac{2n+1}{(n+1)(2n+1)^2}R^2 \\ &= \frac{R^2}{(n+1)(2n+1)}. \end{aligned}$$

6. (a) Quand $n \to +\infty$, on factorise par les prépondérants pour obtenir un équivalent :

$$b(M_n) = -\frac{1}{2n+1}R = -\frac{1}{2n}R\frac{1}{1+1/2n} \sim -\frac{1}{2n}R$$

et

$$r(M_n) = \frac{R^2}{(n+1)(2n+1)^2} = \frac{R^2}{4n^3(1+1/n)(1+1/2n)^2} \sim \frac{R^2}{4n^3}.$$

(b) Donc, en moyenne, T_n donne exactement R contrairement à M_n qui ne s'en approche en moyenne que lorsque n grandit.
N.B. on peut éliminer l'inconvénient du biais en considérant $M'_n = \frac{2n+1}{2n} M_n$ qui sera de biais nul et de risque quadratique du même ordre.
Mais pour n grand, l'écart quadratique sera plus faible avec M_n qu'avec T_n

Exercice 7.

1. Voir exemple 10 du cours
2. Voir exemple 12 du cours.

Exercice 8 (EDHEC 2000 voie S)**.**

1. Notons T_1 l'événement « la personne tire la carte numéro 1 ». La famille $(T_1, \overline{T_1})$ est un système complet d'énements. La formule des probabilités totales donne alors :

$$\theta = P(V) = P_{T_1}(V)\ P(T_1) + P_{\overline{T_1}}(V)\ P(\overline{T_1}).$$

Sachant que la personne a tiré la carte numéro 1, elle répond "vrai" si et seulement si elle est d'accord avec l'affirmation « A ». La proportion de gens de cette population qui sont réellement d'accord avec l'affirmation « A » étant p : $P_{T_1}(V) = p$.
Sachant que la personne n'a pas tiré la carte numéro 1, elle répond "vrai" si et seulement si elle n'est pas d'accord avec l'affirmation « A » ; alors : $P_{\overline{T_1}}(V) = 1 - p$.

De plus : $P(T_1) = \dfrac{1}{20}$ et $P(\overline{T_1}) = \dfrac{19}{20}$. Finalement :

$$\theta = P(V) = \frac{1}{20}p + \frac{19}{20}(1-p).$$

Ainsi : $\theta = \dfrac{19-18p}{20}$.
Ceci donne encore $18\,p = 19 - 20\,\theta$ et donc :

$$p = \frac{19-20\,\theta}{18}.$$

2. (a) Si l'échantillonnage est assimilable à un tirage avec remise, on peut alors affirmer que S_n suit une loi binìmiale de paramètres n et θ. En particulier :

$$E(S_n) = n\theta \quad \text{et} \quad V(S_n) = n\theta(1-\theta).$$

(b) Par linéarité :

$$E\left(\frac{S_n}{n}\right) = \frac{1}{n}E(S_n) = \theta.$$

Ainsi $\dfrac{S_n}{n}$ est un estimateur sans biais de θ.
On a de plus :

$$V\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\ V(S_n) = \frac{\theta(1-\theta)}{n}.$$

Par l'inégalité de Bienaymé-Tchebytchev on obtient alors :

$$\forall \varepsilon > 0, \quad P_\theta\left(\left|\frac{S_n}{n} - \theta\right| > \varepsilon\right) \leq \frac{\theta(1-\theta)}{n\varepsilon}.$$

Ainsi par encadrement :

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} P_\theta\left(\left|\frac{S_n}{n} - \theta\right| > \varepsilon\right) = 0.$$

L'estimateur $\dfrac{S_n}{n}$ est convergent.

3. (a) Comme 23 personnes ont répondu "vrai" sur un échantillon de 100 personnes alors une estimation ponctuelle de θ est : $0,23$.
Comme $p = \dfrac{19-20\,\theta}{18}$ une estimation ponctuelle de p est $0,8$.

(b) D'après le cours, si $t_{0,95}$ est un réel vérifiant $2\Phi(t_{0,95}) - 1 = 0,95$ alors

$$\left[\frac{23}{100} - \frac{t_{0,95}}{2\sqrt{100}}, \frac{23}{100} + \frac{t_{0,95}}{2\sqrt{100}}\right]$$

est un intervalle de confiance à 95% de θ.
On a

$$2\Phi(t_{0,95}) - 1 = 0,95 \quad \text{donc} \quad \Phi(t_{0,95}) = \frac{0,95+1}{2} = 0,975 = \Phi(1,96).$$

Nous poserons donc : $t_{0,95} = 1,96$.
Alors :

$$\frac{23}{100} - \frac{t_{0,95}}{2\sqrt{100}} = 0,23 - 0,098 = 0,132 \quad \text{et} \quad \frac{23}{100} + \frac{t_{0,95}}{2\sqrt{100}} = 0,23 + 0,098 = 0,328.$$

Finalement $[0,132 ; 0,328]$ est un intervalle de confiance à 95% de θ.
Par ailleurs, comme $p = \dfrac{19-20\,\theta}{18}$, alors :

$$\theta \in [0,132 ; 0,328] \Longleftrightarrow p \in \left[\frac{19 - 20\times 0,328}{18}, \frac{19-20\times 0,132}{18}\right].$$

Or $0,6911$ est une valeur approchée par défaut de $\dfrac{19-20\times 0,328}{18}$ et $0,9089$ est une valeur approchée par excès de $\dfrac{19-20\times 0,132}{18}$. Ainsi $[0,6911 ; 0,9089]$ est un intervalle de confiance à 95% de p.

Exercice 9 (ESSEC 2005 maths II)**.**

1. On a : $E(F_n) = E\left(\frac{S_n}{n}\right) = \dfrac{E(S_n)}{n} = \dfrac{\sum_{i=1}^n E(X_i)}{n} = \dfrac{nE(X)}{n} = p.$
Et F_n n'est fonction que des X_i donc F_n est un estimateur sans biais de p.

2. Le risque quadratique est :

$$\begin{aligned} r_n &= E\left((F_n-p)^2\right) = E\left[(F_n - E(F_n))^2\right] \\ &= V(F_n) \\ &= \frac{1}{n^2} V(S_n) \\ &= \frac{1}{n^2}\sum_{i=1}^n V(X_i) \text{ car les } X_i \text{ sont indépendantes} \\ &= \frac{1}{n} p(1-p). \end{aligned}$$

Aindi : $\lim_{n\to+\infty} r_n = 0$.

3. (a) Comme $p = E(X_i)$ et que $V(X_i) = p(1-p)$ alors la moyenne centrée réduite est

$$\frac{\frac{\sum_{i=1}^n X_i}{n} - p}{\sqrt{\frac{p(1-p)}{n}}} = \sqrt{n}\frac{F_n - p}{\sqrt{p(1-p)}}.$$

Les X_i étant indépendants, la loi limite de $\left(\sqrt{n}\frac{F_n-p}{\sqrt{p(1-p)}}\right)_{n\in\mathbb{N}^*}$ est une loi normale, centrée réduite d'après le TCL.

(b) Soit f_n la réalisation de F_n sur l'échantillon considéré. Soit t_α le réel défini par $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$, où Φ désigne la fonction de répartition de la loi normale centrée, réduite.
La défnition d'un intervalle de confiance de p au niveau $1-\alpha$ est donné par $[U_n, V_n]$ tel que

$$P(U_n \leq p \leq V_n) \geq 1 - \alpha$$

On vérifie cela avec

$$U_n = f_n - \frac{t_\alpha}{2\sqrt{n}} \quad , \quad V_n = f_n + \frac{t_\alpha}{2\sqrt{n}}$$

On a égalité des événements suivants :

$$\begin{aligned} (U_n \leq p \leq V_n) &= \left(f_n - \frac{t_\alpha}{2\sqrt{n}} \leq p \leq f_n + \frac{t_\alpha}{2\sqrt{n}}\right) \\ &= \left(-\frac{t_\alpha}{2\sqrt{n}} \leq f_n - p \leq \frac{t_\alpha}{2\sqrt{n}}\right) \\ &= \left(|f_n - p|\, 2\sqrt{n} \leq t_\alpha\right). \end{aligned}$$

Avec la majoration classique sur $p \in]0,1[$: $p(1-p) \leq \frac{1}{4}$ on a

$$\frac{1}{p(1-p)} \geq 4 \quad \text{et} \quad 2 \leq \frac{1}{\sqrt{p(1-p)}}$$

donc **si** $\frac{\sqrt{n}}{\sqrt{p(1-p)}}|F_n - p| \leq t_\alpha$ **alors** $|f_n - p|\, 2\sqrt{n} \leq t_\alpha$. Par conséquent :

$$\begin{aligned} \mathrm{P}\left(|f_n - p|\, 2\sqrt{n} \geq t_\alpha\right) &\geq \mathrm{P}\left(\frac{\sqrt{n}}{\sqrt{p(1-p)}}|F_n - p| \leq t_\alpha\right) = \Phi(t_\alpha) - \Phi(-t_\alpha) \\ &\geq 2\Phi(t_a) - 1 = \alpha \end{aligned}$$

Donc $[U_n, V_n]$ est bien un intervalle de confiance de niveau de confiance $1-\alpha$.