Correction de la fiche 9

La correction comporte 18 pages

1 Estimation ponctuelle et risque quadratique

a Nous avons par définition $T_n = \dfrac{2}{n}\displaystyle\sum_{k=1}^{n} X_k$ et comme les variables X_k admettent chacune une espérance et une variance, T_n admet à son tour une espérance et une variance. Il vient par linéarité de l'espérance :

$$\begin{aligned}
\mathbf{E}(T_n) &= \frac{2}{n}\sum_{k=1}^{n}\mathbf{E}(X_k)\\
&= \frac{2}{n}\sum_{k=1}^{n}\frac{a}{2}\\
&= a
\end{aligned}$$

Conclusion :

$$\boxed{B(T_n, a) = 0 \text{ et } T_n \text{ est un estimateur sans biais}}$$

Comme T_n est un estimateur sans biais, son risque quadratique est égal à sa variance soit :

$$\begin{aligned}
EQ(T_n, a) &= \mathbf{V}(T_n)\\
&= 4\mathbf{V}\left(\overline{X_n}\right)\\
&= \frac{4}{n^2}\sum_{k=1}^{n}\mathbf{V}(X_k)\\
&\quad \text{par indépendance des } X_k\\
&= \frac{4}{n^2}\sum_{k=1}^{n}\frac{a^2}{12}
\end{aligned}$$

Conclusion :

$$\boxed{EQ(T_n, a) = \frac{a^2}{3n}}$$

b On sait qu'en notant respectivement $F_{T'_n}$ et F_X les fonctions de répartition de T'_n et de X nous avons clairement $F_{T'_n} = F_X^n$ qui donne :

$$\boxed{F_{T'_n}(x) = \begin{cases} 0 & \text{si } x < 0 \\ \left(\dfrac{x}{a}\right)^n & \text{si } x \in [0, a] \\ 1 & \text{si } x > a \end{cases}}$$

Nous voyons à ce niveau que :

- $F_{T'_n}$ est continue sur $]-\infty, 0[$ (fonction nulle) ;
- $F_{T'_n}$ est continue sur $]0, a[$ (fonction constante égale à 1) ;
- $F_{T'_n}$ est continue sur $[0, a]$ (monôme) ;
- $\lim\limits_{0^-} F_{T'_n} = 0 = \lim\limits_{0^+} F_{T'_n} = F_{T'_n}(0)$;
- $\lim\limits_{a^-} F_{T'_n} = 0 = \lim\limits_{a^+} F_{T'_n} = F_{T'_n}(a)$
  donc $F_{T'_n}$ est continue sur $\mathbb{R}$.

- D'autre part $F_{T'_n}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ sauf peut être en 0 et a.
  Tout ceci fait que $F_{T'_n}$ possède toutes les propriétés requises pour affirmer que T'_n est une variable à densité dont une densité $f_{T'_n}$ est obtenue par dérivation de $F_{T'_n}$ là où c'est possible, et en complétant la définition de $f_{T'_n}$ en donnant des valeurs arbitraires et positives en 0 et en a. Ce qui donne par exemple :

$$\boxed{f_{T'_n}(x) = \begin{cases} 0 & \text{si} \quad x \in]-\infty, 0[\cup]a, +\infty[\\ \dfrac{nx^{n-1}}{a^n} & \text{si} \quad x \in [0, a] \end{cases}}$$

- T'_n étant une variable à **support fini**, elle admet une espérance et une variance avec :

$$\begin{aligned} \mathbf{E}(T'_n) &= \int_0^a \frac{nx^n}{a^n} dx \\ &= \left[\frac{nx^{n+1}}{(n+1)\, a^n} \right]_0^a \\ &= \frac{na}{n+1} \end{aligned}$$

  Par conséquent :

$$\begin{aligned} B(T'_n, a) &= \mathbf{E}(T'_n - a) \\ &= \mathbf{E}(T'_n) - \mathbf{E}(a) \\ &= \frac{na}{n+1} - a \end{aligned}$$

  Conclusion :

$$\boxed{B(T'_n, a) = -\frac{a}{n+1}}$$

- Le moment d'ordre deux de T'_n existe et vaut :

$$\begin{aligned} \mathbf{E}\left(T'^2_n\right) &= \int_0^a \frac{nx^{n+1}}{a^n} dx \\ &= \left[\frac{nx^{n+2}}{(n+2)\, a^n} \right]_0^a \\ &= \frac{na^2}{n+2} \end{aligned}$$

  T'_n admet une variance donnée par le **théorème de Huygens-Koenig** égale à :

$$\begin{aligned} \mathbf{V}(T'_n) &= \mathbf{E}\left(T'^2_n\right) - (\mathbf{E}(T'_n))^2 \\ &= \frac{na^2}{n+2} - \left(\frac{na}{n+1} \right)^2 \\ &= \frac{na^2}{(n+1)^2 (n+2)} \end{aligned}$$

  Dans ce cas l'erreur quadratique de T'_n par rapport à a vaut :

$$\begin{aligned} EQ(T'_n, a) &= \mathbf{V}(T'_n) + (B(T'_n, a))^2 \\ &= \frac{na^2}{(n+1)^2 (n+2)} + \left(-\frac{a}{n+1} \right)^2 \end{aligned}$$

$$\boxed{EQ(T'_n, a) = \frac{2a^2}{(n+1)(n+2)} \sim \frac{2a^2}{n^2}}$$

c Comme la variable T'_n admet une espérance et une variance, T''_n admet à son tour une espérance et une variance, car obtenue par **transformation affine**, avec :

$$\begin{aligned}\mathbf{E}\left(T''_n\right) &= \left(\frac{n+1}{n}\right)\mathbf{E}\left(T'_n\right)\\ &= \left(\frac{n+1}{n}\right)\times\left(\frac{na}{n+1}\right)\\ &= a\end{aligned}$$

Conclusion :

$$\boxed{B\left(T''_n, a\right) = 0 \text{ ce qui fait que } T''_n \text{ est un estimateur sans biais de } a}$$

Comme T''_n est un estimateur sans biais, son risque quadratique est égal à sa variance soit :

$$\begin{aligned}EQ\left(T''_n, a\right) &= \left(\frac{n+1}{n}\right)^2\mathbf{V}\left(T'_n\right)\\ &= \frac{na^2}{(n+1)^2(n+2)}\left(\frac{n+1}{n}\right)^2\end{aligned}$$

$$\boxed{EQ\left(T''_n, a\right) = \frac{a^2}{n(n+2)} \sim \frac{a^2}{n^2}}$$

d Pour de grandes valeurs de n :

$$\boxed{T''_n \text{ est le meilleur estimateur de } a}$$

2 Estimation ponctuelle

a Par théorème, nous avons :

$$\boxed{\forall n \in \mathbb{N}^*, \quad S_n \hookrightarrow \mathcal{P}(n\lambda)}$$

Ainsi $\forall n \in \mathbb{N}^*$, $\mathbf{E}(S_n) = \mathbf{V}(S_n) = n\lambda$ donc $\forall n \in \mathbb{N}^*$:

$$\begin{aligned}\mathbf{E}\left(\overline{X_n}\right) &= \frac{1}{n}\mathbf{E}(S_n)\\ &\quad \text{par propiété élémentaire de } \mathbf{E}\\ &= \frac{1}{n}n\lambda\\ &= \lambda \qquad (1)\end{aligned}$$

et

$$\begin{aligned}\mathbf{V}\left(\overline{X_n}\right) &= \frac{1}{n^2}\mathbf{V}(S_n)\\ &\quad \text{par propiété élémentaire de } \mathbf{V}\\ &= \frac{1}{n^2}n\lambda\\ &= \frac{\lambda}{n}\end{aligned}$$

ce qui entraîne que :

$$\lim_{n\to+\infty}\mathbf{V}\left(\overline{X_n}\right) = 0 \qquad (2)$$

Selon (1) et (2) nous obtenons une **condition suffisante** pour dire que :

$$\boxed{\overline{X_n} \text{ est un estimateur sans biais et convergent de } \lambda}$$

b.i Par le **théorème de transfert** :

$$\begin{aligned}
& T_n \text{ admet une espérance} \\
\iff & \sum_{k\geq 0} e^{-\frac{k}{n}} \mathbf{P}\left([S_n = k]\right) \text{ est convergente} \\
\iff & \sum_{k\geq 0} e^{-\frac{k}{n}} \frac{e^{-n\lambda}\left(n\lambda\right)^k}{k!} \text{ est convergente} \\
\iff & \sum_{k\geq 0} e^{-n\lambda} \frac{\left(n\lambda e^{-1/n}\right)^k}{k!} \text{ est convergente}
\end{aligned}$$

A ce niveau nous constatons que nous sommes en présence d'*une* série convergente en tant que **série exponentielle** avec $n\lambda e^{-\frac{1}{n}} \in \mathbb{R}$. Dans ce cas $\mathbf{E}\left(T_n\right)$ existe et est égale à :

$$\begin{aligned}
\mathbf{E}\left(T_n\right) &= e^{-n\lambda}\sum_{k=0}^{+\infty}\frac{\left(n\lambda e^{-1/n}\right)^k}{k!} \\
&= e^{-n\lambda}\exp\left(n\lambda e^{-1/n}\right) \\
&= \exp\left(n\lambda\left(e^{-1/n}-1\right)\right)
\end{aligned}$$

Comme $\mathbf{E}\left(T_n\right) \neq e^{-\lambda}$ mais que $\lim\limits_{n\to+\infty}\mathbf{E}\left(T_n\right) = e^{-\lambda}$ du fait que $e^{-\frac{1}{n}} - 1 \underset{+\infty}{\sim} -\frac{1}{n}$ cela entraîne que $n\lambda\left(e^{-1/n}-1\right) \underset{+\infty}{\sim} -\lambda$ et par **continuité** de la fonction exp, nous pouvons conclure que T_n n'est pas un estimateur sans biais certes, mais **un estimateur asymptotiquement sans biais.**

b.ii Les valeurs de X_1 lorsque l'événement $[S_n = s]$ est réalisé appartiennent à l'ensemble $[\![0, s]\!]$, et $\forall k \in [\![0, s]\!]$:

$$\begin{aligned}
\mathbf{P}_{[S_n=s]}\left([X_1 = k]\right) &= \frac{\mathbf{P}\left([X_1 = k]\cap[S_n = s]\right)}{\mathbf{P}\left([S_n = s]\right)} \\
&= \frac{\mathbf{P}\left([X_1 = k]\cap[X_1+\cdots+X_n = s]\right)}{\mathbf{P}\left([S_n = s]\right)} \\
&= \frac{\mathbf{P}\left([X_1 = k]\cap[X_2+\cdots+X_n = s-k]\right)}{\mathbf{P}\left([S_n = s]\right)} \\
&= \frac{\mathbf{P}\left([X_1 = k]\right)\mathbf{P}\left([X_2+\cdots+X_n = s-k]\right)}{\mathbf{P}\left([S_n = s]\right)} \\
& \quad \text{par indépendance des } X_k \\
&= \frac{e^{-\lambda}\lambda^k}{k!}\times\frac{e^{-(n-1)\lambda}\left(\left(n-1\right)\lambda\right)^{s-k}}{(s-k)!}\times\frac{s!}{e^{-n\lambda}\left(n\lambda\right)^s} \\
&= \frac{s!}{k!\left(s-k\right)!}\times\frac{\lambda^k\left(\left(n-1\right)\lambda\right)^{s-k}}{\left(n\lambda\right)^s} \\
&= \binom{s}{k}\left(\frac{\lambda}{n\lambda}\right)^k\left(\frac{\left(n-1\right)\lambda}{n\lambda}\right)^{s-k} \\
&= \binom{s}{k}\left(\frac{1}{n}\right)^k\left(\frac{n-1}{n}\right)^{s-k} \qquad (3)
\end{aligned}$$

en notant que $X_2+\cdots+X_n \hookrightarrow \mathcal{P}\left(\left(n-1\right)\lambda\right)$. Selon (3) :

$$\boxed{\text{La loi conditionnelle de } X_1 \text{ sachant que } [S_n = s] \text{ est la loi } \mathcal{B}\left(s, \frac{1}{n}\right)}$$

En particulier :

$$\boxed{\mathbf{P}_{[S_n=s]}\left([X_1 = 0]\right) = \left(\frac{n-1}{n}\right)^s}$$

b.iii Calculons pour commencer, en cas d'existence, $\mathbf{E}\left(\widehat{\theta}\right)$. Par le **théorème de transfert** :

$$\begin{aligned}
& \widehat{\theta} \text{ admet une espérance} \\
\iff & \sum_{k\geq 0}\left(1-\frac{1}{n}\right)^k \mathbf{P}\left([S_n=k]\right) \text{ est convergente} \\
\iff & \sum_{k\geq 0}\left(1-\frac{1}{n}\right)^k \frac{e^{-n\lambda}\left(n\lambda\right)^k}{k!} \text{ est convergente} \\
\iff & \sum_{k\geq 0} e^{-n\lambda}\frac{\left(n\lambda\left(1-\frac{1}{n}\right)\right)^k}{k!} \text{ est convergente}
\end{aligned}$$

A ce niveau nous constatons que nous sommes en présence d'une série convergente en tant que **série proportionnelle à la série exponentielle** de paramètre $n\lambda\left(1-\dfrac{1}{n}\right)\in\mathbb{R}$. Dans ce cas $\mathbf{E}\left(\widehat{\theta}\right)$ existe et vaut :

$$\begin{aligned}
\mathbf{E}\left(\widehat{\theta}\right) &= e^{-n\lambda}\sum_{k=0}^{+\infty}\frac{\left(n\lambda\left(1-\frac{1}{n}\right)\right)^k}{k!} \\
&= e^{-n\lambda}\exp\left(n\lambda-\lambda\right) \\
&= \exp\left(n\lambda-n\lambda-\lambda\right) \\
&= e^{-\lambda} \\
&= \theta
\end{aligned} \tag{4}$$

Ainsi selon (4) :

$$\boxed{\widehat{\theta} \text{ est un estimateur sans biais}}$$

b.iv Calculons la variance de $\widehat{\theta}$ et montrons que $\lim_{n\to+\infty}\mathbf{V}\left(\widehat{\theta}\right)=0$.

$$\begin{aligned}
& \widehat{\theta} \text{ admet une variance} \iff \widehat{\theta}^2 \text{ admet une espérance} \\
& \text{et par le théorème de transfert } \widehat{\theta}^2 \text{ admet une espérance} \\
\iff & \sum_{k\geq 0}\left(1-\frac{1}{n}\right)^{2k} \mathbf{P}\left([S_n=k]\right) \text{ est convergente} \\
\iff & \sum_{k\geq 0}\left(1-\frac{1}{n}\right)^{2k} \frac{e^{-n\lambda}\left(n\lambda\right)^k}{k!} \text{ est convergente} \\
\iff & \sum_{k\geq 0} e^{-n\lambda}\frac{\left(n\lambda\left(1-\frac{1}{n}\right)^2\right)^k}{k!} \text{ est convergente}
\end{aligned}$$

A ce niveau nous constatons que nous sommes en présence d'*une* série convergente en tant que **série proportionnelle à la série exponentielle** de paramètre $n\lambda\left(1-\frac{1}{n}\right)^2\in\mathbb{R}$. Dans ce cas $\mathbf{E}\left(\widehat{\theta}^2\right)$ existe et vaut :

$$\begin{aligned}
\mathbf{E}\left(\widehat{\theta}^2\right) &= e^{-n\lambda}\sum_{k=0}^{+\infty}\frac{\left(n\lambda\left(1-\frac{1}{n}\right)^2\right)^k}{k!} \\
&= e^{-n\lambda}\exp\left(n\lambda\left(1-\frac{1}{n}\right)^2\right) \\
&= \exp\left(n\lambda\left(\left(1-\frac{1}{n}\right)^2-1\right)\right) \\
&= \exp\left(\frac{\lambda}{n}-2\lambda\right)
\end{aligned}$$

$\widehat{\theta}$ admet une variance égale, par le **théorème de Huygens-Koenig**, à :

$$\begin{aligned}
\mathbf{V}\left(\widehat{\theta}\right) &= \mathbf{E}\left(\widehat{\theta}^2\right) - \left(\mathbf{E}\left(\widehat{\theta}\right)\right)^2 \\
&= \exp\left(\frac{\lambda}{n} - 2\lambda\right) - \left(e^{-\lambda}\right)^2 \\
&= \exp\left(\frac{\lambda}{n} - 2\lambda\right) - e^{-2\lambda} \\
&= e^{-2\lambda}\left(e^{\frac{\lambda}{n}} - 1\right)
\end{aligned}$$

Comme :

$$\begin{aligned}
& e^{\frac{\lambda}{n}} - 1 \underset{+\infty}{\sim} \frac{\lambda}{n} \\
\Longrightarrow\ & e^{-2\lambda}\left(e^{\frac{\lambda}{n}} - 1\right) \underset{+\infty}{\sim} \frac{\lambda e^{-2\lambda}}{n} \\
\Longrightarrow\ & \boxed{\lim_{n\to+\infty} \mathbf{V}\left(\widehat{\theta}\right) = 0}
\end{aligned} \tag{5}$$

Nota bene : selon (5), $\widehat{\theta}$ est un estimateur convergent.

3 Estimation ponctuelle et par intervalle

a Pour déterminer une densité de X nous déterminons sa fonction de répartition F_X à l'aide de la **formule des probabilités totales** associée au système complet d'événements de probabilités non nulles $([X \leq a], [X > a])$, ce qui donne pour tout réel x :

$$\begin{aligned}
F_X(x) &= \mathbf{P}([X \leq x]) \\
&= \mathbf{P}_{[X\leq a]}([X \leq x])\,\mathbf{P}([X \leq a]) + \mathbf{P}_{[X>a]}([X \leq x])\,\mathbf{P}([X > a]) \\
&= \begin{cases} 0 & \text{si } x < 0 \\ \left(\dfrac{x}{a}\right) \times \mathbf{P}([X \leq a]) + 0 \times \mathbf{P}([X > a]) & \text{si } x \in [0, a] \\ 1 \times \mathbf{P}([X \leq a]) + \left(\dfrac{x-a}{1-a}\right) \times \mathbf{P}([X > a]) & \text{si } x \in [a, 1] \\ 1 & \text{si } x > 1 \end{cases} \\
&= \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x}{2a} & \text{si } x \in [0, a] \\ \dfrac{1}{2} + \dfrac{1}{2}\left(\dfrac{x-a}{1-a}\right) & \text{si } x \in [a, 1] \\ 1 & \text{si } x > 1 \end{cases} \\
&= \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x}{2a} & \text{si } x \in [0, a] \\ \dfrac{x-2a+1}{2(1-a)} & \text{si } x \in [a, 1] \\ 1 & \text{si } x > 1 \end{cases}
\end{aligned}$$

Je vous laisse démontrer que F_X possède les propriétés recquises pour confirmer que X est une variable à densité dont une densité f_X est obtenue par dérivation de F_X ce qui nous donne, par exemple :

$$\boxed{f_X(x) = \begin{cases} \dfrac{1}{2a} & \text{si} \quad x \in [0, a] \\ \dfrac{1}{2(1-a)} & \text{si} \quad x \in [a, 1] \\ 0 & \text{sinon} \end{cases}}$$

Comme X est une **variable bornée**, elle des moments à tous ordre, donc en particulier une espérance et une variance. Après un petit calcul sans prétention utilisant la **relation de Chasles**, nous obtenons que :

$$\boxed{\mathbf{E}(X) = \frac{2a+1}{4} \quad \text{et} \quad \mathbf{V}(X) = \frac{4+(2a-1)^2}{48}}$$

b La variable M_n admet une espérance et une variance en tant que somme de telles variables avec :

$$\begin{aligned} \mathbf{E}(M_n) &= \frac{n\mathbf{E}(X)}{n} \\ &= \mathbf{E}(X) \\ &= \frac{2a+1}{4} \end{aligned}$$

et :

$$\begin{aligned} \mathbf{V}(M_n) &= \frac{n\mathbf{V}(X)}{n^2} \\ &= \frac{\mathbf{V}(X)}{n} \\ &= \frac{4+(2a-1)^2}{48n} \end{aligned}$$

Conclusion :

$$\boxed{\mathbf{E}(M_n) = \frac{2a+1}{4} \quad \text{et} \quad \mathbf{V}(M_n) = \frac{4+(2a-1)^2}{48n}}$$

En prenant :

$$T_n = 2M_n - \frac{1}{2}$$

son espérance, qui existe, est égale à :

$$\begin{aligned} \mathbf{E}(T_n) &= 2\mathbf{E}(M_n) - \frac{1}{2} \\ &= 2\left(\frac{2a+1}{4}\right) - \frac{1}{2} \\ &= a \end{aligned} \tag{6}$$

et sa variance, qui existe aussi, vaut :

$$\begin{aligned} \mathbf{V}(T_n) &= 4\mathbf{V}(M_n) \\ &= \frac{4+(2a-1)^2}{12n} \end{aligned}$$

Ainsi comme $\mathbf{E}(T_n) = a$ et $\lim\limits_{n\to+\infty} \mathbf{V}(T_n) = 0$:

$$\boxed{\text{La suite d'estimateur } (T_n)_n \text{ est sans biais et convergent de } a}$$

c Comme $a \in]0,1[$ nous avons :

$$\begin{aligned}\mathbf{V}(T_n) &= \frac{4+(2a-1)^2}{12n} \\ &< \frac{5}{12n} \\ &\quad \text{car } (2a-1)^2 < 1\end{aligned} \tag{7}$$

Selon l'**inégalité de Bienaymé-Tchebychev**, nous écrivons :

$$\forall \varepsilon > 0, \quad \mathbf{P}([|T_n - \mathbf{E}(T_n)| \geq \varepsilon]) \leq \frac{\mathbf{V}(T_n)}{\varepsilon^2}$$

soit, selon (6) et (7) :

$$\forall \varepsilon > 0, \quad \mathbf{P}([|T_n - a| \geq \varepsilon]) \leq \frac{5}{12n\varepsilon^2}$$

ou encore :

$$\forall \varepsilon > 0, \quad \mathbf{P}([T_n - \varepsilon \leq a \leq T_n + \varepsilon]) \geq 1 - \frac{5}{12n\varepsilon^2}$$

Pour que l'intervalle $[T_n - \varepsilon, T_n + \varepsilon]$ soit un **intervalle de confiance** au risque α pour a, il **suffit** que ε soit tel que :

$$1 - \frac{5}{12n\varepsilon^2} \geq 1 - \alpha$$

soit tel que :

$$\boxed{\varepsilon \geq \sqrt{\frac{5}{12n\alpha}}}$$

d Comme M_n est une **somme** de **variables indépendantes et identiquement distribuées** alors le **théorème de la limite centrée** permet d'affirmer que la suite $(M_n)_{n\geq 1}$ est **convergente en loi** vers une variable normale. D'autre part vous devez savoir que **toute transformation affine appliquée à une variable normale redonne une variable normale**. Par conséquent comme la variable T_n est obtenue par **transformation affine** à partir de M_n, le **théorème de la limite centrée** permet de dire que :

$$\boxed{\text{La suite } \left(\frac{T_n - a}{\sqrt{\mathbf{V}(T_n)}}\right)_{n\geq 1} \text{ converge en loi vers une variable normale centrée et réduite}}$$

puisque, ne l'oublions pas $\mathbf{E}(T_n) = a$.

Comme vous connaissez par coeur la table donnant les valeurs de la fonction de répartition de la loi normale centrée et réduite, ce n'est donc pas à vous que vais apprendre que :

$$\mathbf{P}\left(\left[\left|\frac{T_n - a}{\sqrt{\mathbf{V}(T_n)}}\right| \leq 1.96\right]\right) = 0.95$$

soit donc :

$$\mathbf{P}\left(\left[-1.96 \leq \frac{T_n - a}{\sqrt{\mathbf{V}(T_n)}} \leq 1.96\right]\right) = 0.95$$

ou encore :

$$\mathbf{P}\left(\left[-1.96\sqrt{\mathbf{V}(T_n)} \leq T_n - a \leq 1.96\sqrt{\mathbf{V}(T_n)}\right]\right) = 0.95$$

et enfin :

$$\mathbf{P}\left(\left[T_n - 1.96\sqrt{\mathbf{V}(T_n)} \leq a \leq T_n + 1.96\sqrt{\mathbf{V}(T_n)}\right]\right) = 0.95$$

L'**intervalle de confiance pour *a*** à un risque inférieur à 5% est donc :

$$\left[\ T_n - 1.96\sqrt{\mathbf{V}(T_n)} \quad ; \quad T_n + 1.96\sqrt{\mathbf{V}(T_n)}\ \right]$$

qui est inclus dans l'intervalle :

$$\boxed{\left[\ T_n - 1.96\sqrt{\frac{5}{12n}}\ \ ;\ \ T_n + 1.96\sqrt{\frac{5}{12n}}\ \right]}$$

où le **risque pris est inférieur à 5%**.c'est à dire que l'intervalle de confiance obtenu, est observable avec un risque inférieur à $\alpha = 5\%$.

e La longueur de l'intervalle trouvé à la question **c.** est de 2ε supérieure ou égale à $2\sqrt{\dfrac{5}{12n\alpha}}$, et à la question **d.** elle est de $2 \times 1.96\sqrt{\dfrac{5}{12n}}$. En notant R le rapport des longueur nous obtenons :

$$\begin{aligned} R &\geq \frac{2\sqrt{\dfrac{5}{12n\alpha}}}{2 \times 1.96\sqrt{\dfrac{5}{12n}}} \\ &\geq 0.510\,2\sqrt{\frac{1}{\alpha}} \\ &\geq 0.510\,2\sqrt{\frac{1}{0.05}} \\ &\geq 2.\,281\,7 \end{aligned}$$

Conclusion :

La longueur de l'intervalle de confiance au risque de 5% de se tromper donnée par l'inégalité de Bienaymé-Tchebychev est de longueur au moins égale à 2.3 fois plus grande que celle donnée par le théorème de la limite centrée

4 Estimation ponctuelle

a.i La variable S_n est à valeurs dans $[a, b]$. Si $x \in [a, b]$ alors :

$$[S_n \leq x] = [X_1 \leq x] \cap [X_2 \leq x] \cap \ldots \cap [X_n \leq x]$$

On en déduit que :

$$\begin{aligned} F_{S_n}(x) &= \mathbf{P}([S_n \leq x]) \\ &= \begin{cases} 0 & \text{si} \quad x < a \\ \left(\dfrac{x-a}{b-a}\right)^n & \text{si} \quad x \in [a,b] \\ 1 & \text{si} \quad x > b \end{cases} \end{aligned}$$

Je vous laisse voir rapidement que les propriétés de F_{S_n} nous permettant de dire que S_n reste une variable à densité de densité associée f_{S_n} telle que :

$$f_{S_n}(x) = \begin{cases} 0 & \text{si} \quad x < a \\ n\dfrac{(x-a)^{n-1}}{(b-a)^n} & \text{si} \quad x \in [a,b] \\ 0 & \text{si} \quad x > b \end{cases}$$

Comme S_n est à **support fini** elle admet une espérance égale à :

$$\begin{aligned} \mathbf{E}(S_n) &= \int_{-\infty}^{+\infty} t f(t)\, dt \\ &= \int_a^b n\frac{x(x-a)^{n-1}}{(b-a)^n} dx \\ &= \int_a^b n\frac{(x-a+a)(x-a)^{n-1}}{(b-a)^n} dx \\ &= \int_a^b n\frac{(x-a)^n}{(b-a)^n} dx + \int_a^b n\frac{a(x-a)^{n-1}}{(b-a)^n} dx \end{aligned}$$

$$= \frac{n}{(b-a)^n}\left(\left[\frac{(x-a)^{n+1}}{n+1}\right]_a^b + a\left[\frac{(x-a)^n}{n}\right]_a^b\right)$$

$$\boxed{\mathbf{E}(S_n) = \frac{1}{n+1}a + \frac{n}{n+1}b}$$

a.ii En calculant de même $\mathbf{E}\left(S_n^2\right)$ en supposant que $[a,b]=[0,1]$ on obtient facilement :

$$\begin{aligned}\mathbf{V}(S_n) &= \mathbf{E}\left(S_n^2\right) - (\mathbf{E}(S_n))^2\\ &= \frac{n}{(n+2)(n+1)^2}\end{aligned}$$

La transformation affine $x \longmapsto a+(b-a)x$ permet d'achever le calcul et donne :

$$\boxed{\mathbf{V}(S_n) = \frac{n(b-a)^2}{(n+2)(n+1)^2}}$$

a.iii Comme $\mathbf{E}(S_n) \neq b$:

$$\boxed{S_n \text{ n'est pas un estimateur sans biais de } b}$$

a.iv Comme $\lim\limits_{n\to+\infty} \mathbf{E}(S_n) = b$:

$$\boxed{S_n \text{ est asymptotiquement sans biais de } b}$$

a.v Comme $\mathbf{E}(S_n) < b$ on risque en moyenne de $\boxed{\text{sous-estimer } b}$ en ne corrigeant par S_n.

b Les mêmes calculs conduisent à :

$$\boxed{\mathbf{E}(I_n) = \frac{n}{n+1}a + \frac{1}{n+1}b}$$

On fait les $\boxed{\text{mêmes remarques}}$ qu'aux questions précédentes sur les qualités de I_n.

c On déduit :

$$\begin{cases}\mathbf{E}(S_n) = \dfrac{1}{n+1}a + \dfrac{n}{n+1}b\\[2mm] \mathbf{E}(I_n) = \dfrac{n}{n+1}a + \dfrac{1}{n+1}b\end{cases}$$

$$\Longleftrightarrow \boxed{\begin{cases}a = \dfrac{n\mathbf{E}(I_n) - \mathbf{E}(S_n)}{n-1}\\[2mm] b = \dfrac{n\mathbf{E}(S_n) - \mathbf{E}(I_n)}{n-1}\end{cases}}$$

Des estimateurs A_n et B_n sans biais de a et b sont :

$$\boxed{A_n = \frac{nI_n - S_n}{n-1} \text{ et } B_n = \frac{nS_n - I_n}{n-1} \text{ sont des estimateurs sans biais de } a \text{ et } b}$$

5 Estimation ponctuelle et par intervalle

a.i Nous avons :

$$\boxed{F_X(x) = \begin{cases} 0 & \text{si} \quad x \leq \theta \\ \displaystyle\int_{-\infty}^{x} f(t)\,dt = \int_0^x \frac{\lambda}{\theta}\left(\frac{\theta}{t}\right)^{\lambda+1} dt = 1 - \left(\frac{\theta}{x}\right)^{\lambda} & \text{si} \quad x > \theta \end{cases}}$$

a.ii Là encore c'est une question culturelle qui donne (et que vous devez retrouver en concours) :

$$\begin{aligned} & X \text{ admet une espérance} \\ \iff & \int_{-\infty}^{+\infty} x f(x)\,dx \text{ est absolument convergente} \\ \iff & \int_{\theta}^{+\infty} x\frac{\lambda}{\theta}\left(\frac{\theta}{x}\right)^{\lambda+1} dx \text{ est convergente} \\ \iff & \int_{\theta}^{+\infty} \lambda\left(\frac{\theta}{x}\right)^{\lambda} dx \text{ est convergente} \\ \iff & \lambda > 1 \end{aligned}$$

et

$$\begin{aligned} & X \text{ admet une variance} \\ \iff & X^2 \text{ admet une espérance} \\ \iff & \int_{-\infty}^{+\infty} x^2 f(x)\,dx \text{ est convergente} \\ \iff & \int_{\theta}^{+\infty} x^2\frac{\lambda}{\theta}\left(\frac{\theta}{x}\right)^{\lambda+1} dx \text{ est convergente} \\ \iff & \int_{\theta}^{+\infty} \frac{\lambda\theta^{\lambda}}{x^{\lambda-1}} dx \text{ est convergente} \\ \iff & \lambda > 2 \end{aligned}$$

Conclusion :

$$\boxed{\begin{array}{ll} \text{Si } \lambda > 1, & \mathbf{E}(X) = \dfrac{\lambda\theta}{\lambda - 1} \\ \text{Si } \lambda > 2, & \mathbf{V}(X) = \dfrac{\lambda\theta^2}{(\lambda-2)(\lambda-1)^2} \end{array}}$$

a.iii Les propriétés de $\varphi : x \longmapsto \ln\left(\frac{x}{\theta}\right)$ sur $\mathbb{R}_+^*$ nous permettent de dire que Y reste une variable à densité et nous noterons F_Y sa fonction de répartition. Nous avons $Y(\Omega) = \mathbb{R}_+^*$ et :

- Si $x \leq 0$, $F_Y(x) = 0$.
- Si $x > 0$:

$$\begin{aligned} F_Y(x) &= \mathbf{P}\left(\left[\ln\left(\frac{X}{\theta}\right) \leq x\right]\right) \\ &= \mathbf{P}\left(\left[\frac{X}{\theta} \leq e^x\right]\right) \\ &= \mathbf{P}([X \leq \theta e^x]) \\ &= F_X(\theta e^x) \\ &= 1 - \left(\frac{\theta}{\theta e^x}\right)^{\lambda} \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Conclusion :

$$\boxed{Y \hookrightarrow \varepsilon(\lambda) \quad \text{soit} \quad Y \hookrightarrow \Gamma\left(\frac{1}{\lambda}, 1\right)}$$

b Tout d'abord :

$$\begin{aligned} T &= \sum_{i=1}^{n} \ln\left(\frac{X_i}{\theta}\right) \\ &= \sum_{i=1}^{n} Y_i \end{aligned}$$

où les variables Y_i sont indépendantes car les variables X_i le sont. Alors par **stabilité de la loi gamma pour la somme de variables indépendantes** nous pouvons dire que :

$$\boxed{T \hookrightarrow \Gamma\left(\frac{1}{\lambda}, n\right)}$$

Une densité de T notée f_T est définie par :

$$\boxed{f_T(x) = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ \dfrac{e^{-\lambda x} x^{n-1}}{(1/\lambda)^n \Gamma(n)} & \text{si} \quad x > 0 \end{cases} = \begin{cases} 0 & \text{si} \quad x \leq 0 \\ \dfrac{\lambda^n e^{-\lambda x} x^{n-1}}{(n-1)!} & \text{si} \quad x > 0 \end{cases}}$$

c Nous avons $\widehat{\lambda} = \dfrac{n}{T}$ et donc d'après le **théorème de transfert :**

$$\begin{aligned} & \widehat{\lambda} \text{ admet une espérance} \\ \Longleftrightarrow \quad & \int_{-\infty}^{+\infty} \frac{n}{x} f_T(x)\, dx \text{ est absolument convergente} \\ \Longleftrightarrow \quad & \int_{0}^{+\infty} \frac{n\lambda^n e^{-\lambda x} x^{n-2}}{(n-1)!} dx \text{ est convergente} \\ \Longleftrightarrow \quad & \frac{n\lambda}{n-1} \int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-1)-1}}{(1/\lambda)^{n-1} (n-2)!} dx \text{ est convergente} \\ \Longleftrightarrow \quad & \frac{n\lambda}{n-1} \int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-1)-1}}{(1/\lambda)^{n-1} \Gamma(n-1)} dx \text{ est convergente} \end{aligned}$$

En introduisant G une variable aléatoire suivant la loi gamma $\Gamma\left(\dfrac{1}{\lambda}, n-1\right)$ nous pouvons donc affirmer que :

$$\int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-1)-1}}{(1/\lambda)^{n-1} \Gamma(n-1)} dx = 1$$

$\widehat{\lambda}$ admet une espérance égale à :

$$\boxed{\mathbf{E}\left(\widehat{\lambda}\right) = \frac{n\lambda}{n-1}}$$

d Cherchons un estimateur de λ. Pour cela il suffit de prendre :

$$\boxed{\widehat{\lambda_1} = \frac{n-1}{n} \lambda}$$

car dans ce cas $\mathbf{E}\left(\widehat{\lambda_1}\right)$ existe et vaut :

$$\begin{aligned} \mathbf{E}\left(\widehat{\lambda_1}\right) &= \frac{n-1}{n} \mathbf{E}\left(\widehat{\lambda}\right) \\ &= \lambda \end{aligned}$$

Pour savoir si l'estimateur $\widehat{\lambda_1}$ est convergent montrons que $\lim\limits_{n\to+\infty} \mathbf{V}\left(\widehat{\lambda_1}\right) = 0$.

$$\begin{aligned}
& \widehat{\lambda} \text{ admet une variance} \\
\iff\ & \widehat{\lambda}^2 \text{ admet une espérance} \\
\iff\ & \int_{-\infty}^{+\infty} \left(\frac{n}{x}\right)^2 f_T(x)\,dx \text{ est convergente} \\
\iff\ & \int_{0}^{+\infty} \frac{n^2\lambda^n e^{-\lambda x} x^{n-3}}{(n-1)!}\,dx \text{ est convergente} \\
\iff\ & \frac{(n\lambda)^2}{(n-1)(n-2)} \int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-2)-1}}{(1/\lambda)^{n-2}(n-3)!}\,dx \text{ est convergente} \\
\iff\ & \frac{(n\lambda)^2}{(n-1)(n-2)} \int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-2)-1}}{(1/\lambda)^{n-2}\Gamma(n-2)}\,dx \text{ est convergente}
\end{aligned}$$

En introduisant H une variable aléatoire suivant la loi gamma $\Gamma\left(\frac{1}{\lambda}, n-2\right)$, nous pouvons donc affirmer que :

$$\int_{0}^{+\infty} \frac{e^{-\frac{x}{1/\lambda}} x^{(n-2)-1}}{(1/\lambda)^{n-2}\Gamma(n-2)}\,dx = 1$$

et que $\mathbf{E}\left(\widehat{\lambda}^2\right)$ qui existe vaut :

$$\mathbf{E}\left(\widehat{\lambda}^2\right) = \frac{(n\lambda)^2}{(n-1)(n-2)}$$

$\widehat{\lambda}$ admet une variance égale selon le **théorème de Huygens-Koenig** à :

$$\begin{aligned}
\mathbf{V}\left(\widehat{\lambda}\right) &= \mathbf{E}\left(\widehat{\lambda}^2\right) - \left(\mathbf{E}\left(\widehat{\lambda}\right)\right) \\
&= \frac{(n\lambda)^2}{(n-1)(n-2)} - \left(\frac{n\lambda}{n-1}\right)^2 \\
&= \frac{\lambda^2 n^2}{(n-1)^2(n-2)}
\end{aligned}$$

et comme $\lim\limits_{n\to+\infty} \mathbf{V}\left(\widehat{\lambda_1}\right) = 0$ du fait que :

$$\frac{\lambda^2 n^2}{(n-1)^2(n-2)} \underset{+\infty}{\sim} \frac{\lambda^2}{n}$$

nous pouvons affirmer que :

$$\boxed{\text{L'estimateur } \widehat{\lambda_1} \text{ est convergent}}$$

e Nous avons :

$$\begin{aligned}
& \mathbf{P}\left([-1.96 \le Z_n \le 1.96]\right) = 0.95 \\
\iff\ & \mathbf{P}\left(\left[-1.96 \le \frac{\sqrt{n}\left(\widehat{\lambda_1}-\lambda\right)}{\lambda} \le 1.96\right]\right) = 0.95 \\
\iff\ & \mathbf{P}\left(\left[-1.96\lambda \le \sqrt{n}\left(\widehat{\lambda_1}-\lambda\right) \le 1.96\lambda\right]\right) = 0.95 \\
\iff\ & \mathbf{P}\left(\left[\frac{-1.96\lambda}{\sqrt{n}} \le \widehat{\lambda_1}-\lambda \le \frac{1.96\lambda}{\sqrt{n}}\right]\right) = 0.95
\end{aligned}$$

$$\begin{aligned}
\iff\ & \mathbf{P}\left(\left[\lambda - \frac{1.96\lambda}{\sqrt{n}} \le \widehat{\lambda_1} \le \lambda + \frac{1.96\lambda}{\sqrt{n}}\right]\right) = 0.95 \\
\iff\ & \mathbf{P}\left(\left[\frac{\sqrt{n}}{\sqrt{n}+1.96}\widehat{\lambda_1} \le \lambda \le \frac{\sqrt{n}}{\sqrt{n}-1.96}\widehat{\lambda_1}\right]\right) = 0.95
\end{aligned}$$

Nous obtenons donc que :

$$\boxed{IC_{5\%}(\lambda) = \left[\frac{\sqrt{n}}{\sqrt{n}+1.96}\lambda_0 \quad ; \quad \frac{\sqrt{n}}{\sqrt{n}-1.96}\lambda_0 \right] = [\ 4.18 \quad ; \quad 6.22\ ]}$$

6 Estimation ponctuelle

a D'après le cours, la **stabilité de la loi de Poisson pour la somme de variables indépendantes** nous permet d'écrire que :

$$\boxed{X + Y \hookrightarrow \mathcal{P}(\lambda + \mu)}$$

Pour les mêmes raisons on montre très facilement par récurrence que :

$$\boxed{S_n = \sum_{k=1}^{n} X_k \hookrightarrow \mathcal{P}(n\lambda)}$$

b Comme S_n admet une espérance, M_n en admet une aussi égale à :

$$\begin{aligned} \mathbf{E}(M_n) &= \frac{1}{n}\mathbf{E}(S_n) \\ &= \frac{n\lambda}{n} \\ &= \lambda \end{aligned}$$

Ainsi $B(M_n, \lambda) = 0$ et :

$$\boxed{M_n \text{ est un estimateur sans biais de } \lambda}$$

Comme S_n admet une variance, M_n en admet une aussi égale à :

$$\begin{aligned} \mathbf{V}(M_n) &= \frac{1}{n^2}\mathbf{V}(S_n) \\ &= \frac{n\lambda}{n^2} \end{aligned}$$

$$\boxed{\mathbf{V}(M_n) = \frac{\lambda}{n}}$$

c.i Par le **théorème de transfert :**

$$\begin{aligned} & T_n \text{ admet une espérance} \\ \Longleftrightarrow\ & \sum_{k\geq 0} \exp\left(-q\frac{k}{n}\right)\mathbf{P}([S_n = k]) \text{ est convergente} \\ \Longleftrightarrow\ & \sum_{k\geq 0} \exp\left(-q\frac{k}{n}\right)\frac{e^{-n\lambda}(n\lambda)^k}{k!} \text{ est convergente} \\ \Longleftrightarrow\ & \sum_{k\geq 0} e^{-n\lambda}\frac{\left(n\lambda e^{-\frac{q}{n}}\right)^k}{k!} \text{ est convergente} \end{aligned}$$

A ce niveau nous constatons que nous sommes en présence d'**une série convergente en tant que série proportionnelle à la série exponentielle** de paramètre $n\lambda e^{-\frac{q}{n}} \in \mathbb{R}$. Dans ce cas $\mathbf{E}(T_n)$ existe et est égale à :

$$\begin{aligned} \mathbf{E}(T_n) &= e^{-n\lambda}\sum_{k=0}^{+\infty}\frac{\left(n\lambda e^{-\frac{q}{n}}\right)^k}{k!} \\ &= e^{-n\lambda}\exp\left(n\lambda e^{-\frac{q}{n}}\right) \end{aligned}$$

$$\boxed{\mathbf{E}(T_n) = \exp\left(n\lambda\left(e^{-\frac{q}{n}} - 1\right)\right)}$$

c.ii Comme $\mathbf{E}(T_n) \neq e^{-q\lambda}$:

$$\boxed{T_n \text{ n'est pas un estimateur sans biais de } e^{-q\lambda}}$$

$$\boxed{\lim_{n\to+\infty} \mathbf{E}(T_n) = e^{-q\lambda}}$$

car :

$$n\lambda\left(e^{-\frac{q}{n}} - 1\right) \underset{+\infty}{\sim} -\frac{n\lambda q}{n} \underset{+\infty}{\sim} -\lambda q$$

et la fonction exp est continue sur $\mathbb{R}$.

c.iii Par le **théorème de transfert :**

$$\begin{aligned}
& T_n \text{ admet une variance} \\
\iff & T_n^2 \text{ admet une espérance} \\
\iff & \sum_{k\geq 0} \exp\left(-q\frac{2k}{n}\right) \mathbf{P}([S_n = k]) \text{ est convergente} \\
\iff & \sum_{k\geq 0} \exp\left(-q\frac{2k}{n}\right) \frac{e^{-n\lambda}(n\lambda)^k}{k!} \text{ est convergente} \\
\iff & \sum_{k\geq 0} e^{-n\lambda} \frac{\left(n\lambda e^{-\frac{2q}{n}}\right)^k}{k!} \text{ est convergente}
\end{aligned}$$

A ce niveau nous constatons que nous sommes en présence d'**une série convergente en tant que série proportionnelle à la série exponentielle** de paramètre $n\lambda e^{-\frac{2q}{n}} \in \mathbb{R}$. Dans ce cas $\mathbf{E}\left(T_n^2\right)$ existe et est égale à :

$$\begin{aligned}
\mathbf{E}\left(T_n^2\right) &= e^{-n\lambda} \sum_{k=0}^{+\infty} \frac{\left(n\lambda e^{-\frac{2q}{n}}\right)^k}{k!} \\
&= e^{-n\lambda} \exp\left(n\lambda e^{-\frac{2q}{n}}\right) \\
&= \exp\left(n\lambda\left(e^{-\frac{2q}{n}} - 1\right)\right)
\end{aligned}$$

$\mathbf{V}(T_n)$ existe et vaut par le **théorème de Hyugens-Koenig** :

$$\boxed{\mathbf{V}(T_n) = \exp\left(n\lambda\left(e^{-\frac{2q}{n}} - 1\right)\right) - \exp\left(2n\lambda\left(e^{-\frac{q}{n}} - 1\right)\right)}$$

$$\lim_{n\to+\infty} \mathbf{E}\left(T_n^2\right) = e^{-2q\lambda}$$

car :

$$n\lambda\left(e^{-\frac{2q}{n}} - 1\right) \underset{+\infty}{\sim} -\frac{2n\lambda q}{n} \underset{+\infty}{\sim} -2\lambda q$$

et la fonction exp est continue sur $\mathbb{R}$. On a aussi :

$$\lim_{n\to+\infty} \left(\mathbf{E}(T_n)\right)^2 = e^{-2q\lambda}$$

Donc :

$$\lim_{n\to+\infty} \mathbf{V}(T_n) = 0$$

T_n étant un estimateur asymptotiquement sans biais de $e^{-q\lambda}$ de variance tendant vers 0, ceci est une condition suffisante pour dire que :

$$\boxed{T_n \text{ est un estimateur convergent}}$$

d.i Comme $g(S_n)$ un estimateur sans biais de $e^{-q\lambda}$, $\mathbf{E}(g(S_n))$ exite et vaut :

$$\sum_{k=0}^{+\infty} g(k)\,\mathbf{P}([S_n = k]) = e^{-q\lambda}$$

Soit :

$$\sum_{k=0}^{+\infty} g(k)\,\frac{e^{-n\lambda}(n\lambda)^k}{k!} = e^{-q\lambda}$$

soit encore :

$$\boxed{\sum_{k=0}^{+\infty} \frac{n^k g(k)}{k!}\lambda^k = e^{(n-q)\lambda}}$$

Nous avons :

$$\begin{aligned}
& \sum_{k=0}^{+\infty} \frac{n^k g(k)}{k!}\lambda^k = e^{(n-q)\lambda} \\
\iff & \sum_{k=0}^{+\infty} \frac{(n\lambda)^k g(k)}{k!} = \sum_{k=0}^{+\infty} \frac{((n-q)\lambda)^k}{k!} \\
\iff & \sum_{k=0}^{+\infty} \left(\frac{(n\lambda)^k g(k)}{k!} - \frac{((n-q)\lambda)^k}{k!}\right) = 0 \\
\iff & \sum_{k=0}^{+\infty} \left(\frac{n^k g(k)}{k!} - \frac{(n-q)^k}{k!}\right)\lambda^k = 0 \\
\iff & \sum_{k=0}^{+\infty} \frac{n^k g(k) - (n-q)^k}{k!}\lambda^k = 0 \\
\iff & \sum_{k=0}^{+\infty} \frac{g(k) - \left(\frac{n-q}{n}\right)^k}{k!}(n\lambda)^k = 0
\end{aligned}$$

et grâce au résultat admis :

$$\forall k \in \mathbb{N},\quad g(k) = \left(\frac{n-q}{n}\right)^k$$

Ainsi le seul estimateur sans biais, fonction de S_n de $e^{-q\lambda}$ est :

$$\boxed{Z_n = \left(1 - \frac{q}{n}\right)^{S_n}}$$

d.ii

$$\boxed{\text{Lorsque } q > n \text{ on risque d'estimer une probabilité par un nombre négatif si } S_n \text{ est impaire !}}$$

d.iii Nous avons :

$$\begin{aligned}
Z_n &= \left(1 - \frac{q}{n}\right)^{S_n} \\
&= \exp\left(S_n \ln\left(1 - \frac{q}{n}\right)\right) \\
&\simeq \exp\left(-\frac{S_n q}{n}\right)
\end{aligned}$$

et

$$\boxed{Z_n \text{ est très peu différent de } T_n}$$

7 Régression orthogonale

a **La droite de régression ou droite qui minimalise la somme des carrés des écarts de n points à une droite quelconque, dépend de l'écart utilisé.** Nous développons dans cet exercice les calculs dans le cas où l'écart retenu est tout simplement le **carré de la distance du point à la droite**. Nous avons :

L'angle $\widehat{B_iA_iA'_i}$ est égal à l'angle $\widehat{x'A'_iz}$, comme angles à côtés prependiculaires. Dans le triangle rectangle $A_iA'_iB_i$, on a $(A_iB_i)^2 = (A_iA'_i)^2 + (A'_iB_i)^2$ ce qui s'écrit :

$$\begin{aligned}(y_i - cx_i - d)^2 &= (A_iA'_i)^2 + (A_iA'_i)^2 \tan^2\alpha \\ &= (A_iA'_i)^2 \left(1 + \tan^2\alpha\right)\end{aligned}$$

En posant $c = \tan\alpha$, nous obtenons :

$$\boxed{(A_iA'_i)^2 = \frac{(y_i - cx_i - d)^2}{1 + c^2}}$$

b De $f(c,d) = \sum_{i=1}^{n} \frac{(y_i - cx_i - d)^2}{1+c^2}$, on déduit $f'_d(c,d) = -\frac{2}{1+c^2}\sum_{i=1}^{n}(y_i - cx_i - d)$. Pour $d = d_0$ avec :

$$\begin{aligned}d_0 &= \frac{1}{n}\left(\sum_{i=1}^{n} y_i - c\sum_{i=1}^{n} x_i\right) \\ &= \overline{y} - c\overline{x}\end{aligned}$$

on a $f'_d(c,d) = 0$. On obtient alors :

$$\begin{aligned}f(c,d_0) &= \sum_{i=1}^{n} \frac{(y_i - cx_i - \overline{y} + c\overline{x})^2}{1+c^2} \\ &= \frac{1}{1+c^2}\sum_{i=1}^{n}\left(\overset{\bullet}{y}_i - c\overset{\bullet}{x}_i\right)^2 \\ &= \frac{1}{1+c^2}\sum_{i=1}^{n}\left(c^2\overset{\bullet}{x}_i{}^2 - 2c\overset{\bullet}{x}_i\overset{\bullet}{y}_i + \overset{\bullet}{y}_i{}^2\right)\end{aligned}$$

$$\begin{aligned}&= \frac{1}{1+c^2}\left(c^2\sum_{i=1}^{n}\overset{\bullet}{x}_i{}^2 - 2c\sum_{i=1}^{n}\overset{\bullet}{x}_i\overset{\bullet}{y}_i + \sum_{i=1}^{n}\overset{\bullet}{y}_i{}^2\right) \\ &= \frac{n}{1+c^2}\left(c^2s_x^2 - 2cs_{xy} + s_y^2\right)\end{aligned}$$

D'où :

$$f'_c(c,d_0) = \frac{2n}{1+c^2}\left(c^2s_{xy} - c\left(s_y^2 - s_x^2\right) - s_{xy}\right)$$

L'équation $f'_d(c,d) = 0$ admet deux racines distinctes :

$$c_1 = \frac{1}{2}\left(\alpha - \left(\alpha^2+4\right)^{1/2}\right) \quad \text{et} \quad c_2 = \frac{1}{2}\left(\alpha + \left(\alpha^2+4\right)^{1/2}\right)$$

Le produit des racines vaut -1. Elles sont donc de signes contraires. Effectuons le tableau de variation :

c	$-\infty$		c_1	0	c_2		$+\infty$
$f'_c(c,d_0)$		$+$	0	$-$	0	$+$	
$f_c(c,d_0)$	ns_x^2	$\nearrow$	$f_c(c_1,d_0)$	$\searrow$	$f_c(c_2,d_0)$	$\nearrow$	ns_x^2

Il nous indique que le minimum global se situe en $f_c(c_2, d_0)$. Les valeurs de c et d qui répondent au problème de minimisation sont donc :

$$\boxed{\begin{array}{l} d_0 = \overline{y} - c_0\overline{x} \\ c_0 = \dfrac{1}{2}\left(\alpha + \left(\alpha^2+4\right)^{1/2}\right) \end{array}}$$

c L'équation cherchée est donc :

$$\boxed{y = \frac{1}{2}\left(\alpha + \left(\alpha^2+4\right)^{1/2}\right)x + d_0}$$

✠✠✠✠✠