Bachelor académique en Sciences et Ingénierie

Probabilités et Statistique 2

Corrigé TD 5 2021

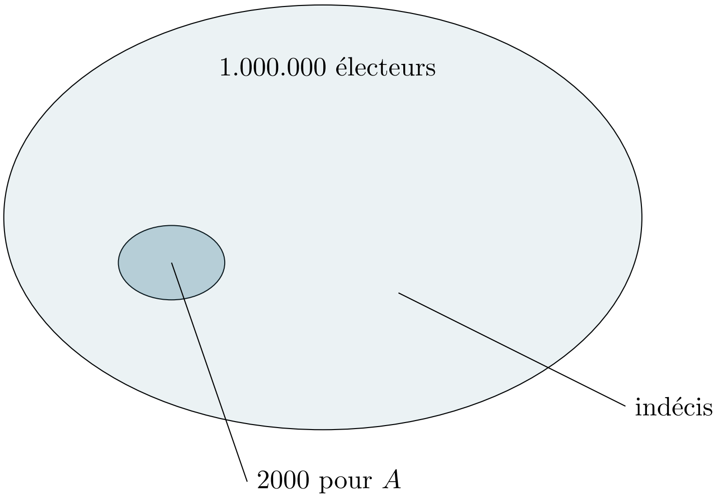
1. Comme les $X_1, X_2, \ldots$ sont indépendantes et de même loi, les transformées $g(X_1), g(X_2), \ldots$ sont aussi indépendates et suivent la même loi. Par la loi forte des grands nombres on a la convergence:

$$\frac{g(X_1) + \ldots + g(X_n)}{n} \longrightarrow \mathbb{E}[g(X_1)] \quad \text{presque sûrement, quand } n \to \infty.$$

Puisque X_1 suit la loi uniforme sur $[0, 1]$ on a

$$\mathbb{E}[g(X_1)] = \int_0^1 g(x)\, dx.$$

2. Soit p_A la probabilité d'une victoire du candidat A.



On note

$$S_n = X_1 + \ldots + X_n, \quad n = 998\,000,$$

avec $X_1, \ldots, X_n$ i.i.d. telles que

$$\mathbb{P}\{X_i = 1\} = \mathbb{P}\{X_i = 0\} = \frac{1}{2}.$$

On a l'interprétation $\{X_i = 1\} \equiv \{\text{"vote pour } A\text{"}\}$. La valeur exacte de p_A est

$$\begin{aligned} p_A &= \mathbb{P}\{S_n > 498\,000\} \\ &= 1 - \mathbb{P}\{S_n \leq 498\,000\} \\ &= 1 - \sum_{k=0}^{498\,000} \binom{998.000}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{998\,000-k}. \end{aligned}$$

On utilise le théorème de la limite centrale. On note que

$$\mathbb{E}[X_i] = \mathbb{E}[X_i^2] = 1/2 \quad \text{et} \quad \operatorname{var}(X_i) = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = \frac{1}{4} =: \sigma^2.$$

Alors

$$\begin{aligned}
\mathbb{P}\{0 \le S_n \le a\} &= \mathbb{P}\left\{ \frac{-\mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \le \frac{S_n - \mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \le \frac{a - \mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \right\} \\
&\approx N\left(\frac{a - \mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \right) - N\left(\frac{-\mathbb{E}[S_n]}{\sqrt{\operatorname{var}(S_n)}} \right) \\
&\approx N\left(\frac{a - \mathbb{E}[S_n]}{\sqrt{n}\sigma} \right) - N\left(\frac{-\mathbb{E}[S_n]}{\sqrt{n}\sigma} \right),
\end{aligned}$$

et avec $a = 498\,000$ on a

$$\begin{aligned}
p_A &= 1 - \mathbb{P}\{0 \le S_n \le a\} \\
&\approx 1 - N\left(\frac{a - \mathbb{E}[S_n]}{\sqrt{n}\sigma} \right) + N\left(\frac{-\mathbb{E}[S_n]}{\sqrt{n}\sigma} \right) \\
&= N\left(\frac{\mathbb{E}[S_n] - a}{\sqrt{n}\sigma} \right) + N\left(\frac{-\mathbb{E}[S_n]}{\sqrt{n}\sigma} \right), \quad \text{car } 1 - N(z) = N(-z), \\
&= N\left(\frac{499\,000 - 498\,000}{\sqrt{998\,000}} 2 \right) + N\left(-\frac{499\,000}{\sqrt{998\,000}} 2 \right) \\
&= N\left(\frac{2\,000}{\sqrt{998\,000}} \right) + N\left(-\frac{998\,000}{\sqrt{998\,000}} \right) \\
&\approx N(2) + N(-1) \approx N(2) \\
&= 0,9772.
\end{aligned}$$

3. Soit $X_1, \ldots, X_n$ un n-échantillon de la loi de Bernoulli de paramètre p inconnu, c.à.d. les variables $X_1, \ldots, X_n$ sont i.i.d. et chaque $X_i \sim \mathcal{B}(p)$ avec $p \in]0, 1[$. On a $n = 2\,645\,756$.

On souhaite tester l'hypothèse $\mathbf{H}_0$: $\quad p = \frac{1}{2} = \mathbb{P}\{\text{“garçon”}\}$.

D'abord on remarque que, sous l'hypothèse $\mathbf{H}_0$,

$$\mathbb{E}[X_i] = \frac{1}{2} \quad \text{et} \quad \sigma^2 = \operatorname{var}(X_i) = p(1-p) = \frac{1}{4}.$$

On note

$$\bar{p}_n = \frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n}$$

la fréquence empirique. D'après le théorème limite central, sous l'hypothèse $\mathbf{H}_0$, la suite

$$Z_n = \sqrt{n}\, \frac{\bar{p}_n - 1/2}{1/2} = \frac{S_n - \mathbb{E}[S_n]}{\sqrt{n}\,\sigma} = \frac{S_n - n/2}{\sqrt{n}/2}$$

converge en loi vers une variable aléatoire Z suivant la loi normale centrée réduite $N(0, 1)$. En particulier,

$$\mathbb{P}\{Z_n \ge z\} \to \mathbb{P}\{Z \ge z\} = 1 - N(z)$$

où $N(z)$ signifie la fonction de répartition de la loi normale standard $N(0,1)$. La valeur observée de la variable Z_n est

$$Z_n^{\text{obs}} = \frac{1\,359\,670 - 1\,322\,878}{\sqrt{2\,645\,756/2}} \geq 45,$$

mais

$$\mathbb{P}\{Z_n \geq 45\} \approx 1 - N(45) < \frac{1}{2}\, e^{-45^2/2} \approx 0.$$

L'hypothèse de départ $\mathbf{H}_0$, c.à.d. $p = 1/2$, est donc très certainement fausse. (On entend souvent dire qu'environ 51 % des naissances sont des garçons ...)

4. Une pièce tombe sur pile avec la probabilité $0,6$. À chaque coup, on mise 1 euro sur face. Soit G_n le gain après n parties.

(a) On a

$$G_n = X_1 + \ldots + X_n$$

avec $X_1, \ldots, X_n$ i.i.d. de même loi de Bernoulli de paramètre $p = 0,4$, c.à.d.

$$\forall i \in \{1, , \ldots, n\}, \quad \mathbb{P}\{X_i = 1\} = 1 - \mathbb{P}\{X_i = -1\} = p = 0,4$$

En particulier,

$$\begin{aligned} \mathbb{E}[X_i] &= 0,4 - 0,6 = -0,2 = -\frac{1}{5} \quad \text{et} \\ \operatorname{var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = 1 - \frac{1}{25} = \frac{24}{25}\,. \end{aligned}$$

D'où

$$\mathbb{E}[G_n] = n\,\mathbb{E}[X_1] = -\frac{n}{5} \quad \text{et} \quad \operatorname{var}(G_n) = n\, \operatorname{var}(X_i) = \frac{24}{25}\, n.$$

(b) La loi des grands nombres implique

$$\frac{G_n}{n} \to \mathbb{E}[X_1] \quad \text{p.s.}$$

Pour $n = 400$, donc

$$G_n \approx n\mathbb{E}[X_1] = -\frac{400}{5} = -80$$

et d'après l'inégalité de Tchebychev,

$$\begin{aligned} \mathbb{P}\{G_n \geq 0\} &= \mathbb{P}\{G_n - \mathbb{E}[G_n] \geq 80\} \leq \frac{1}{80^2}\, \mathbb{E}[|G_n - \mathbb{E}[G_n]|^2] \\ &= \frac{\operatorname{var}(G_n)}{80^2} = \frac{400}{80^2}\,\frac{24}{25} = \frac{3}{50}. \end{aligned}$$

(c) À l'aide du théorème de la limite central, on obtient

$$\begin{aligned} \mathbb{P}\{G_n \geq 0\} &= \mathbb{P}\left\{\frac{G_n - \mathbb{E}[G_n]}{\sqrt{n}\sigma} \geq \frac{-n\mathbb{E}[X_1]}{\sqrt{n}\sigma}\right\} \\ &\approx \mathbb{P}\left\{Z \geq \frac{1}{5}\,\frac{\sqrt{n}\,5}{\sqrt{24}}\right\} = \mathbb{P}\left\{Z \geq \frac{20}{\sqrt{24}}\right\} = \mathbb{P}\left\{Z \geq \frac{10}{\sqrt{6}}\right\} \end{aligned}$$

où Z désigne une variable normale standard et réduite. On lit dans la table $N(10/\sqrt{6}) = 0,999978$, d'où

$$\mathbb{P}\{G_n \geq 0\} \approx \mathbb{P}\left\{Z \geq \frac{10}{\sqrt{6}}\right\} = 1 - N(10/\sqrt{6}) = 0,000022.$$