Interrogation 11

Lundi 10 février 2025

NOM et Prénom ....

Cours

1. Inégalité de Bienaymé-Tchebychev.
   Soit X une variable aléatoire.
   Si X admet une variance, alors :
$$\forall a \in \mathbb{R}_*^+, \ \mathbb{P}\big([|X - \mathbb{E}(X)| \geqslant a]\big) \leqslant \frac{\mathbb{V}(X)}{a^2}$$

2. Loi faible des grands nombres.
   Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires sur $(\Omega, \mathcal{A}, \mathbb{P})$. Pour tout $n \in \mathbb{N}^*$, on note $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$.
   Si $(X_k)_{k\in\mathbb{N}^*}$ est une suite de variables aléatoires **indépendantes**, admettant toutes **la même espérance m** et **la même variance σ^2** (c'est le cas si elles ont toutes la même loi) alors :
$$\forall \varepsilon > 0, \ \lim_{n\to+\infty} \mathbb{P}\big([|\overline{X_n} - m| \geqslant \varepsilon]\big) = 0$$

3. Théorème central limite.
   Soit $(X_k)_{k\in\mathbb{N}^*}$ une suite de variables aléatoires **indépendantes**, **de même loi** (espérance notée m, variance $\sigma^2 \neq 0$)
   On pose, pour tout $n \in \mathbb{N}^*$:
$$\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k \quad ; \quad \overline{X_n}^* = \frac{\overline{X_n} - \mathbb{E}(\overline{X_n})}{\sigma(\overline{X_n})} = \sqrt{n}\,\frac{\overline{X_n} - m}{\sigma}$$
   Dans ce cas, la suite $\left(\overline{X_n}^*\right)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi $\mathcal{N}(0;1)$.

Exercice 1

On considère une variable aléatoire suivant une loi de Bernoulli de paramètre p inconnu ainsi qu'un n-échantillon $(X_1, \ldots, X_n)$ de X. On note $\overline{X_n}$ la moyenne empirique de cet échantillon.

1. Démontrer que $\overline{X_n}$ possède une espérance et une variance puis les déterminer.
   On sait que $X_1, \ldots, X_n$ admettent une espérance et une variance ; donc $\overline{X_n}$ également, comme combinaison linéaire de variables aléatoires admettant une espérance et une variance.
   Puis :
$$\begin{aligned}\mathbb{E}(\overline{X_n}) &= \mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)\\ &= \frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(X_k)\\ &= p\end{aligned}$$
   linéarité de l'espérance, licite car pour tout $k \in [\![1;n]\!]$, $\mathbb{E}(X_k)$ existe

   et :
$$\begin{aligned}\mathbb{V}(\overline{X_n}) &= \mathbb{V}\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right)\\ &= \frac{1}{n^2}\sum_{k=1}^{n}\mathbb{V}(X_k)\\ &= \frac{p(1-p)}{n}\end{aligned}$$
   indépendance de $X_1, \ldots, X_n$

   Conclusion : $\overline{X_n}$ admet une espérance et une variance et
$$\mathbb{E}(\overline{X_n}) = p \quad ; \quad \mathbb{V}(\overline{X_n}) = \frac{p(1-p)}{n}$$

2. Démontrer : $\forall \varepsilon > 0,\ \mathbb{P}\left([\overline{X_n} - \varepsilon \leqslant p \leqslant \overline{X_n} + \varepsilon]\right) \geqslant 1 - \dfrac{1}{4n\varepsilon^2}$.

   Soit $\varepsilon > 0$.
   D'après l'inégalité de Bienaymé-Tchebychev, licite car $\overline{X_n}$ admet une variance et que $\varepsilon > 0$:

$$\mathbb{P}\left([|\overline{X_n} - \mathbb{E}(\overline{X_n})| \geqslant \varepsilon]\right) \leqslant \frac{\mathbb{V}(\overline{X_n})}{\varepsilon^2}$$

   Autrement dit :

$$\mathbb{P}\left([|\overline{X_n} - p| \geqslant \varepsilon]\right) \leqslant \frac{p(1-p)}{n\varepsilon^2}$$

   Mais, on sait que la fonction $x \longmapsto x(1-x)$ est majorée par $\dfrac{1}{4}$; d'où :

$$\mathbb{P}\left([|\overline{X_n} - m| \geqslant \varepsilon]\right) \leqslant \frac{1}{4n\varepsilon^2}$$

   Or :

$$\begin{aligned}\mathbb{P}\left([|\overline{X_n} - p| \geqslant \varepsilon]\right) &= 1 - \mathbb{P}\left([|\overline{X_n} - p| < \varepsilon]\right) \\ &= 1 - \mathbb{P}\left([\overline{X_n} - \varepsilon < p < \overline{X_n} + \varepsilon]\right) \qquad \varepsilon > 0\end{aligned}$$

   D'où :

$$\mathbb{P}\left([\overline{X_n} - \varepsilon < p < \overline{X_n} + \varepsilon]\right) \geqslant 1 - \frac{1}{4n\varepsilon^2}$$

   Enfin, on a $[\overline{X_n} - \varepsilon < p < \overline{X_n} + \varepsilon] \subset [\overline{X_n} - \varepsilon \leqslant p \leqslant \overline{X_n} + \varepsilon]$ et par croissance de $\mathbb{P}$, on obtient par transitivité :

$$\mathbb{P}\left([\overline{X_n} - \varepsilon \leqslant p \leqslant \overline{X_n} + \varepsilon]\right) \geqslant 1 - \frac{1}{4n\varepsilon^2}$$

   Conclusion : $\forall \varepsilon > 0,\ \mathbb{P}\left([\overline{X_n} - \varepsilon \leqslant p \leqslant \overline{X_n} + \varepsilon]\right) \geqslant 1 - \dfrac{1}{4n\varepsilon^2}$.

3. En déduire que $\left[\overline{X_n} - \sqrt{\dfrac{5}{n}}, \overline{X_n} + \sqrt{\dfrac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance 95%.

   ✓ En notant $U_n = \overline{X_n} - \sqrt{\dfrac{5}{n}}$ et $V_n = \overline{X_n} - \sqrt{\dfrac{5}{n}}$, on sait déjà que U_n et V_n sont des estimateurs de p (car fonctions d'un n-échantillon de X dont les expressions ne font pas apparaître p) et que $\mathbb{P}\left([U_n \leqslant V_n]\right) = 1$.

   ✓ Montrons ensuite que $\mathbb{P}\left(\left[p \in \left[\overline{X_n} - \sqrt{\dfrac{5}{n}}, \overline{X_n} + \sqrt{\dfrac{5}{n}}\right]\right]\right) \geqslant 0,95$.

   On a :

$$\begin{aligned}\mathbb{P}\left(\left[p \in \left[\overline{X_n} - \sqrt{\frac{5}{n}}, \overline{X_n} + \sqrt{\frac{5}{n}}\right]\right]\right) &= \mathbb{P}\left(\left[\overline{X_n} - \sqrt{\frac{5}{n}} \leqslant p \leqslant \overline{X_n} + \sqrt{\frac{5}{n}}\right]\right) \qquad \text{point précédent avec } \varepsilon = \sqrt{\frac{5}{n}} \\ &\geqslant 1 - \frac{1}{20} = 0,95\end{aligned}$$

   Conclusion : $\left[\overline{X_n} - \sqrt{\dfrac{5}{n}}, \overline{X_n} + \sqrt{\dfrac{5}{n}}\right]$ est un intervalle de confiance de p au niveau de confiance 95%.

Exercice 2

On considère une urne composée d'une boule blanche et d'une boule noire. On effectue une succession de tirages dans l'urne : si on tire la boule boule noire, on la remet dans l'urne en ajoutant une boule blanche ; si on tire une boule blanche, on s'arrête.
On note X la variable aléatoire égale au nombre de tirages effectués. Écrire une fonction **Python** permettant de simuler une réalisation de X.

```
import numpy.random as rd

def simule_X():
    b=1
    X=1
    while rd.random()<1/(b+1): #1/(b+1) = proba d'obtenir boule noire
        n=n+1
        X=X+1
    return X
```