Exercice| [4370] | 1| Manipuler l'inégalité de Bienaymé-Tchebychev

Une usine confectionne des pièces dont une proportion p est défectueuse.
La valeur de p est inconnue mais on souhaiterait en connaître une estimation.
Pour cela, on prélève n pièces et on note Z_n le nombre de pièces défectueuses dans ce prélèvement.
On considère que le nombre total de pièces dans l'usine est assez grand pour que le prélèvement des n pièces soit considéré comme une suite de n tirages avec remise.
L'idée est d'approcher la valeur de p par la valeur de $\frac{Z_n}{n}$. On va donc chercher à partir de quelle valeur de n cette approximation sera « bonne ».

(1). Quelle est la loi de Z_n ?

(2). En déduire l'espérance et la variance de Z_n.

(3). (a). Montrer que pour tout $x \in [0;1]$, $x(1-x) \leq \frac{1}{4}$.

(b). Montrer que, pour tout $\varepsilon \geq 0$, on a : $\mathbb{P}\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\epsilon^2}$.

(c). En déduire une condition sur$n$ pour que $\frac{Z_n}{n}$ soit une valeur approchée de p à 10^{-2} près avec une probabilité supérieure ou égale à 95%.

Pistes de réflexion

(1). On pourra essayer de dégager un schéma de Bernoulli pour justifier que Z_n suit une loi binomiale dont on identifiera les paramètres.

(2). Il suffit d'expliciter dans ce cadre la formule donnant l'espérance d'une loi binomiale.

(a). Un étude rapide de la fonction $x \longmapsto x(1-x)$ sur $[0;1]$ permet d'obtenir cette inégalité.

(b). Il s'agit de mettre en oeuvre l'inégalité de Bienaymé-Tchébychev pour la variable aléatoire $\frac{Z_n}{n}$.

(c). La majoration précédente donne la valeur de n cherchée.

Éléments de correction

(1). L'hypothèse faite sur le nombre total de pièces produites par l'usine nous permet de considérer que la proportion de pièces défectueuses présentes dans le stock reste malgré tout quasiment constant lorsque l'on prélève n pièces dans le stock. Ainsi, le prélèvement de n pièces dans ce stock *a priori* considéré sans remise, pourra être supposé avec remise.
Le protocole de prélèvement retenu sera alors le suivant :

— On désigne alors D l'événement « la pièce prélevée est défectueuse » avec $\mathbb{P}(D) = p$.
— On prélève successivement et avec remise n pièces dans ce stock et on note leur caractère défectueux ou non.
— Comme les prélèvements sont supposés avec remise, les prélèvements peuvent donc être considérés comme étant réalisés de manière indépendante.

Ainsi, la variable aléatoire Z_n qui compte le nombre de fois que l'événement D s'est produit lors du prélèvement d'un échantillon de n pièces de ce stock suivant le protocole décrit ci-dessous, suit une loi binomiale de paramètres n et p.

(2). Par théorème, puisque $Z_n \hookrightarrow \mathcal{B}(n,p)$, on sait que Z_n admet une espérance et une variance qui sont $\mathbb{E}(Z_n) = n \times p$ et $\mathbb{V}(Z_n) = n \times p \times (1-p)$.

(3). (a). La fonction $x \longmapsto x(1-x)$ est une fonction polynôme de degré 2 dont le coefficient du terme de degré 2 est égal à -1. Par suite, elle admet un maximum sur $\mathbb{R}$ au point d'abscisse $\frac{0+1}{2} = \frac{1}{2}$ et qui a pour valeur $\frac{1}{4}$. Ainsi, on en déduit que : $\forall x \in [0;1]\,, \quad x(1-x) \leq \frac{1}{4}$.

(b). La variable aléatoire $\frac{Z_n}{n}$ est encore une variable aléatoire finie qui admet une espérance et une variance qui valent :

$$\begin{aligned} \mathbb{E}\left(\frac{Z_n}{n}\right) &= \frac{1}{n}\mathbb{E}(Z_n) \\ &= p \\ \text{et : } \quad \mathbb{V}\left(\frac{Z_n}{n}\right) &= \frac{1}{n^2}\mathbb{V}(Z_n) \\ &= p(1-p) \end{aligned}$$

Par suite, d'après l'inégalité de Bienaymé-Tchebychev appliquée à la variable aléatoire $\frac{Z_n}{n}$, on a :

$$\forall \varepsilon \geq 0, \quad \mathbb{P}\left(\left|\frac{Z_n}{n} - \mathbb{E}\left(\frac{Z_n}{n}\right)\right| \geq \varepsilon\right) \leq \frac{\mathbb{V}\left(\frac{Z_n}{n}\right)}{\varepsilon^2}$$

ce qui donne puisque $\mathbb{E}\left(\frac{Z_n}{n}\right) = p$ et $p(1-p) \leq \frac{1}{4}$:

$$\forall \varepsilon \geq 0, \quad \mathbb{P}\left(\left|\frac{Z_n}{n} - p\right| \geq \varepsilon\right) \leq \frac{1}{4n\varepsilon^2}$$

(c). On cherche donc à déterminer la(les) valeur(s) de n pour que $\mathbb{P}\left(\left|\frac{Z_n}{n} - p\right| < 10^{-2}\right) \geq 0,95$ ou ce qui revient au même en passant à l'événement contraire à $\mathbb{P}\left(\left|\frac{Z_n}{n} - p\right| \geq 10^{-2}\right) \leq 0,05$

Compte-tenu de l'inégalité précédente appliquée pour $\varepsilon = 10^{-2}$, il suffit de déterminer n tel que $0,05 \leq \frac{1}{4n \times 10^{-4}}$.

Il vient que :
$$\begin{aligned} \left(0,05 \leq \frac{1}{4n \times 10^{-4}}\right) &\Leftrightarrow \left(5 \times 10^{-6} \leq \frac{1}{4n}\right) \\ &\Leftrightarrow \left(20 \times 10^{-6} \leq \frac{1}{n}\right) \\ &\Leftrightarrow \left(n \geq \frac{1}{20} \times 10^{6}\right) \\ &\Leftrightarrow (n \geq 50\,000) \end{aligned}$$

Ainsi, dès lors que l'on testera 50 000 pièces, $\frac{Z_n}{n}$ sera une valeur approchée de p à 10^{-2} près avec une probabilité supérieure ou égale à 95%.