Echantillonnage et Estimation

Vallon

12 avril 2016

- Dans une boîte il y a 100 cartons
- 40 ont une marque bleue
- On tire au hasard un carton de la boîte on note si oui ou non il a une marque bleue et on le remet dans la boîte
- On répète cela n fois
- Peut on prévoir dans quel plage de valeurs se trouvera la fréquence des cartons ayant une marque bleue ?

Quelques réalisations pour $n = 36$

- $f_1 = \dfrac{15}{36} \simeq 0,42$ à 10^{-2} près
- $f_2 = \dfrac{12}{36} \simeq 0,33$ à 10^{-2} près
- $f_3 = \dfrac{17}{36} \simeq 0,47$ à 10^{-2} près

Modélisation

- S_n est la variable aléatoire égale au nombre de cartons marqués obtenus au cours des n tirages
- S_n suit une loi binomiale de paramètre n et $p = 0,4$
- $F_n = \dfrac{S_n}{n}$ est la fréquence "prévue" du nombre de cartons marqués au cours des n tirages
- On pose $T = \dfrac{S_n - np}{\sqrt{npq}} = \dfrac{\frac{S_n}{n} - p}{\sqrt{\frac{pq}{n}}} = \dfrac{F_n - p}{\sqrt{\frac{pq}{n}}}$
- Si n est "suffisamment grand" d'après le théorème de De Moivre Laplace T suit la loi normale centrée réduite
- Il existe un unique réel $u_{0,05}$ tel que $P(-u_{0,05} \leqslant T \leqslant u_{0,05}) = 0,95$. Par convention $u_{0,05} \simeq 1,96$
- $P(1,96 \leqslant \dfrac{F_n - p}{\sqrt{\frac{pq}{n}}} \leqslant 1,96) = 0,95$
- $P(p - 1,96\dfrac{\sqrt{pq}}{\sqrt{n}} \leqslant F_n \leqslant p - 1,96\dfrac{\sqrt{pq}}{\sqrt{n}}) = 0,95$

Modélisation

- $P(p - 1,96\frac{\sqrt{pq}}{\sqrt{n}} \leqslant F_n \leqslant p - 1,96\frac{\sqrt{pq}}{\sqrt{n}}) = 0,95$
- On interprète cette probabilité : 95 fois sur 100 la fréquence sera dans l'intervalle $[0,4 - 1,96\sqrt{\frac{0{,}4\times 0{,}6}{36}}; 0,4 + 1,96\sqrt{\frac{0{,}4\times 0{,}6}{36}}] = [0,24; 0,56]$

Théorème

Soit X_n une variable aléatoire suivant une loi binomiale de paramètres n et p et $\alpha \in]0;1[$ $\lim_{n \to +\infty} P\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant \frac{X_n}{n} \leqslant p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha$

Démonstration.

Exercice $\square$

Définition

L'intervalle $[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}]$ est appelé intervalle de fluctuation asymptotique au seuil de $1 - \alpha$

Lorsque $\alpha = 0,05$, $[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} ; p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}]$ est appelé intervalle de fluctuation asymptotique au seuil de 95%

Problème : Dans une boîte il y a une proportion inconnue p de cartons bleus . Comment avoir une estimation de p ?

- Si on réalise un échantillon de taille n "suffisamment grand", de manière aléatoire alors la fréquence f des cartons bleus dans cet échantillon vérifiera dans 95% des cas

$$p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant f \leqslant p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

- Donc $f - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant p \leqslant f + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$
- Donc $f - 2\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leqslant p \leqslant f + 2\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ sera vérifié dans plus de 95% des cas
- Or pour tout $p \in [0;1]$ on a $p(1-p) \leqslant \frac{1}{4}$ donc $f - \frac{1}{\sqrt{n}} \leqslant p \leqslant f + \frac{1}{\sqrt{n}}$ sera vérifié dans plus de 95% des cas

Définition

L'intervalle $[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$ est appelé intervalle de confiance au niveau de confiance 0,95