II - Théorèmes

Soit X_1, X_2, ..., X_n une suite de n variables aléatoires indépendantes de même loi de probabilités admettant une espérance mathématique μ. On pose $\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$.

1°) - Loi faible des grands nombres

Pour tout $\varepsilon > 0$, $\mathrm{P}(|\overline{X} - \mu| < \varepsilon)$ tend vers 1 quand n tend vers l'infini.

Application aux les échantillons de type "sondage" constitués avec remise

Considérons dans une population $\mathcal{P}$, une sous-population $\mathcal{A}$ contenant une proportion p des individus de $\mathcal{P}$.

On note $\mathcal{E}_n(\mathcal{P})$ l'ensemble des échantillons de taille n, constitués avec remise dans $\mathcal{P}$. $\mathcal{E}_n(\mathcal{P})$ peut être assimilé à $\mathcal{P}^n$. On considère les variables aléatoires X_i pour $1 \leq i \leq n$:

$$X_i : \quad \begin{array}{ccl} \mathcal{E}_n(\mathcal{P}) & \longrightarrow & \{0, 1\} \\ (a_1, \ldots, a_n) & \longmapsto & \begin{cases} 0 & \text{si } a_i \text{ n'appartient pas à } \mathcal{A} \\ 1 & \text{si } a_i \text{ appartient à } \mathcal{A} \end{cases} \end{array}$$

Les X_i suivent la loi de Bernoulli de paramètre p. Elles ont pour espérance $\mathrm{E}(X_i) = p$.

Alors la variable aléatoire qui à chaque échantillon de taille n associe la fréquence d'individus de $\mathcal{A}$ est $F = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ et par application de la loi des grands nombres :

pour tout $\varepsilon > 0$, $\mathrm{P}(|F - p| < \varepsilon)$ tend vers 1 quand n tend vers l'infini.

Lors du tirage d'un échantillon, plus n est grand, plus la probabilité d'obtenir un échantillon avec une fréquence observée d'éléments de $\mathcal{A}$ proche de p est forte.

Des résultats :

La loi de nF est la loi binomiale $\mathcal{B}(n\,;p)$. Ainsi, nF a pour espérance np et pour écart-type $\sqrt{n\,p\,(1-p)}$.

F a pour espérance $\frac{np}{n} = p$ et pour écart-type $\frac{\sqrt{n\,p\,(1-p)}}{n} = \sqrt{\frac{p\,(1-p)}{n}}$. L'écart-type qui quantifie la dispersion des résultats (fréquences observées de $\mathcal{A}$) autour de p (l'espérance) tend vers 0 quand n tend vers l'infini.

Application aux les échantillons de type "expérimental"

Considérons une expérience aléatoire à deux issues : *succès* ($\mathcal{S}$) et *échec* ($\mathcal{E}$), $\mathcal{S}$ étant obtenu avec la probabilité p (et donc $\mathcal{E}$ avec la probabilité $1 - p$).

On répète cette expérience n fois. À la i-ème répétition de l'expérience aléatoire ($1 \leq i \leq n$), on associe la variable aléatoire X_i prenant la valeur 1 si le résultat de l'expérience aléatoire est $\mathcal{S}$ et 0 sinon.

Les X_i suivent la loi de Bernoulli de paramètre p. Elles ont pour espérance $\mathrm{E}(X_i) = p$.

Alors la fréquence d'observations de $\mathcal{S}$ est $F = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ et par application de la loi des grands nombres : pour tout $\varepsilon > 0$, $\mathrm{P}(|F - p| < \varepsilon)$ tend vers 1 quand n tend vers l'infini.

Lors de la répétition d'une expérience aléatoire, plus n est grand, plus la probabilité d'obtenir un fréquence d'observation de $\mathcal{S}$ proche de p est forte.

Remarque

La loi des grands nombres permet d'expliquer le fait que l'on peut attribuer comme probabilité à un événement, une valeur autour de laquelle la fréquence d'apparition de cet événement semble se stabiliser lorsqu'on répète l'expérience aléatoire un grand nombre de fois (approche fréquentiste de la probabilité).

2°) - Théorème limite central

On suppose de plus que les X_i ont un (même) écart-type σ.

Alors pour n grand, la loi de la moyenne $\overline{X}$ peut être approchée par la loi normale de paramètres μ et $\frac{\sigma}{\sqrt{n}}$.

Remarque :

Ce théorème est remarquable à plusieurs égards :

- Il s'applique quelle que soit la loi de probabilités commune des X_i pourvu qu'elle possède une variance.
- Il contribue à expliquer le rôle privilégié que joue la loi normale en statistique (en tant que loi limite d'une conjugaison de phénomènes aléatoires).
- Il prouve qu'une suite de variables aléatoires discrètes peut converger en loi vers une variable aléatoire continue.

Application à la distribution de la fréquence d'échantillonnage

D'après le théorème limite central appliqué aux suites précédentes, la loi de F est approchée par la loi normale $\mathcal{N}\left(p\,;\sqrt{\frac{p\,(1-p)}{n}}\right)$ pour $n \geq 30$, $n\mathrm{p} \geq 15$ et $np(1-p) > 5$.

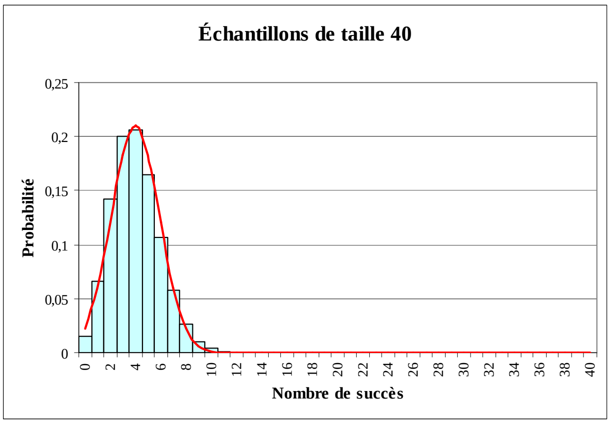
Dans ces conditions, la loi de $\frac{F-p}{\sqrt{\frac{p\,(1-p)}{n}}}$ est approchée par la loi normale $\mathcal{N}(0\ ;\ 1)$.

Application à la distribution de la fréquence dans les échantillons constitués avec remise

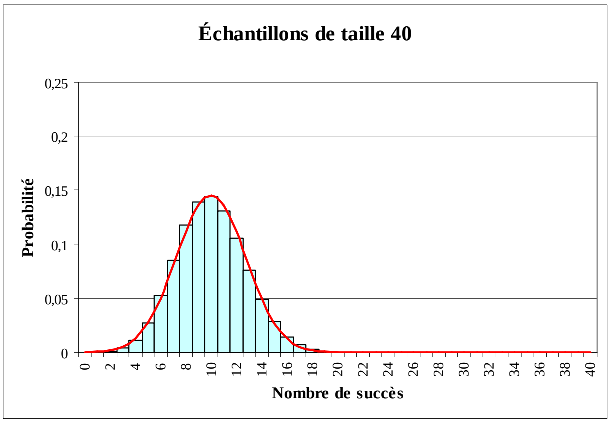
- La loi de nF est la loi binomiale $\mathcal{B}(n\,;p)$.
- D'après le théorème limite central appliqué à la suite précédente, la loi de F est approchée par la loi normale $\mathcal{N}\left(p\,;\sqrt{\frac{p\,(1-p)}{n}}\right)$ pour $n \geq 30$, $n\mathrm{p} \geq 15$ et $np(1-p) > 5$. Dans ces conditions, la loi de $\frac{F-p}{\sqrt{\frac{p\,(1-p)}{n}}}$ est approchée par la loi normale $\mathcal{N}(0\ ;\ 1)$.

$p = 0{,}25$

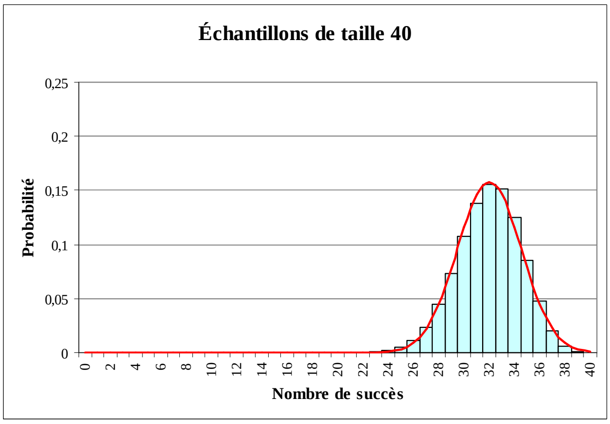
Si p est trop petit ou trop grandion est de mauvaise qualité : elle donnerait une probabilité non négligeable à $\mathbb{R}^-$ ou à $]1 ; +\infty[$.



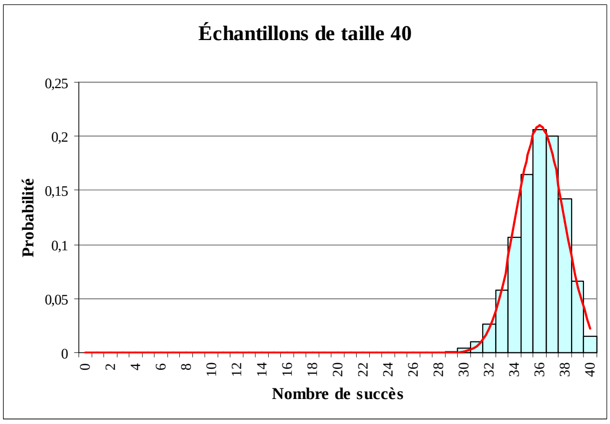
$p = 0{,}1$



$p = 0{,}25$



$p = 0{,}8$



$p = 0{,}9$

Plus n est grand, meilleure est la qualité de l'approximation.