Feuille 3 exo VII

Population A   population B (de grandes tailles les 2)

echantillon $n = 100$ $n = 100$

variables aleatoires X_i^A X_i^B

$\begin{cases} X_i^A = 1 & \text{si le } i^{eme} \text{ individu de l'echantillon issu de A possède le caractère} \\ X_i^A = 0 & \text{sinon.} \end{cases}$

$\begin{cases} Pr(X_i^A = 1) = \frac{1}{5} \\ Pr(X_i^A = 0) = \frac{4}{5} \end{cases}$ $Pr(X_i^B = 1) = \frac{2}{5}$ $Pr(X_i^B = 0) = \frac{3}{5}$

$F_n^A = \frac{\sum_{i=1}^{n} X_i^A}{n} = \frac{T^A}{n}$ $F_n^B = \frac{\sum_{i=1}^{n} X_i^B}{n} = \frac{T^B}{n}$

X_i^A suit une loi de Bernoulli de paramètre $\frac{1}{5}$: $\mathcal{B}\left(\frac{1}{5}\right)$

X_i^B suit $B\left(\frac{2}{5}\right)$

$T^A = \sum_{i=1}^{n} X_i^A = \mathcal{B}\left(n, \frac{1}{5}\right)$ loi Binomiale de paramètre $n, \frac{1}{5}$ car T^A est une somme de variables de Bernoulli independantes.

Pour un n assez grand $n \geq 30$ $np \geq 15$ $nq \geq 15$ p pas trop proche de 0,

on peut approximer $\mathcal{B}(n, \frac{1}{5})$ par la loi normale: $N(np, npq)$,

or ici $n = 100$ ~~100~~ $np = \frac{100}{5}$ $nq = \frac{400}{5}$ pour ~~l~~'echantillon A

donc $T^A \rightsquigarrow N\left(\frac{100}{5} ; \frac{100}{5} \times \frac{4}{5}\right)$

$T^A \rightsquigarrow N(20, 16)$

or $F_n^A = \frac{T^A}{n}$ donc F_n^A combinaison linéaire d'une variable suivant une loi normale alors

F_n^A suivra une ~~suit~~ loi normale de paramètre $E\left(\frac{T^A}{n}\right)$ et de variance

$Var\left(\frac{T^A}{n}\right) = \frac{1}{n^2} Var(T^A)$ ainsi la v.a $F_n^A \rightsquigarrow N\left(\frac{E(T^A)}{n}, \frac{Var(T^A)}{n^2}\right)$

Par conséquent $F_n^A \rightsquigarrow N\left(0{,}2 ; \frac{16}{100^2}\right)$

Conclusion: la loi de F_n^A peut être approximée par $N\left(0{,}2 ; \frac{16}{100^2}\right) = N\left(\frac{1}{5} ; \frac{1}{n} \times \frac{4}{25}\right)$

d'une manière analogue

La loi de F_n^B peut être approximée par la loi $N\left(\frac{2}{5} ; \frac{1}{n}\times\frac{2}{5}\times\frac{3}{5}\right)$

$$F_n^B \rightsquigarrow N\left(\frac{2}{5} ; \frac{6}{25n}\right)$$

3/ Calculs de $\alpha_1 = \text{Prob}\left(0,2 \leq F_n^A \leq 0,3\right)$

$$\alpha_1 \simeq \text{Prob}\left(0,2 \leq N\left(\frac{1}{5} ; \frac{4}{25n}\right) \leq 0,3\right)$$

$$\alpha_1 \simeq \text{Prob}\left(\frac{0,2 - \frac{1}{5}}{\frac{2}{5}\times\frac{1}{\sqrt{n}}} \leq N(0,1) \leq \frac{0,3 - \frac{1}{5}}{\frac{2}{5}\times\frac{1}{\sqrt{n}}}\right)$$ après centrage et réduction.

$$\alpha_1 \simeq \text{Prob}\left(0 \leq N(0,1) \leq \frac{0,1\times 5\sqrt{n}}{2}\right)$$ avec $n = 100$

$$\alpha_1 \simeq \text{Prob}\left(0 \leq N(0,1) \leq 2,5\right)$$

$$\alpha_1 \simeq \Pi(2,5) - \Pi(0) = 0,9938 - 0,5 \simeq 0,4938$$

$$\alpha_1 \simeq 49,4\ \%$$

de la même manière on peut calculer $\alpha_2 = \text{Prob}\left(0,2 \leq F_n^B \leq 0,3\right)$

$$\alpha_2 \simeq \text{Prob}\left(0,2 \leq N\left(\frac{2}{5} ; \frac{6}{25n}\right) \leq 0,3\right)$$

et $$\alpha_2 \simeq \text{Prob}\left(\frac{0,2 - \frac{2}{5}}{\frac{\sqrt{6}}{5\sqrt{n}}} \leq N(0,1) \leq \frac{0,3 - \frac{2}{5}}{\frac{\sqrt{6}}{5\sqrt{n}}}\right)$$

soit $$\alpha_2 \simeq \text{Prob}\left(\frac{-0,2\times 5\times\sqrt{n}}{\sqrt{6}} \leq N(0,1) \leq \frac{-0,1\times 5\times\sqrt{n}}{\sqrt{6}}\right)$$ avec $n = 100$

$$\alpha_2 \simeq \text{Prob}\left(-\frac{10}{\sqrt{6}} \leq N(0,1) \leq -\frac{5}{\sqrt{6}}\right)$$

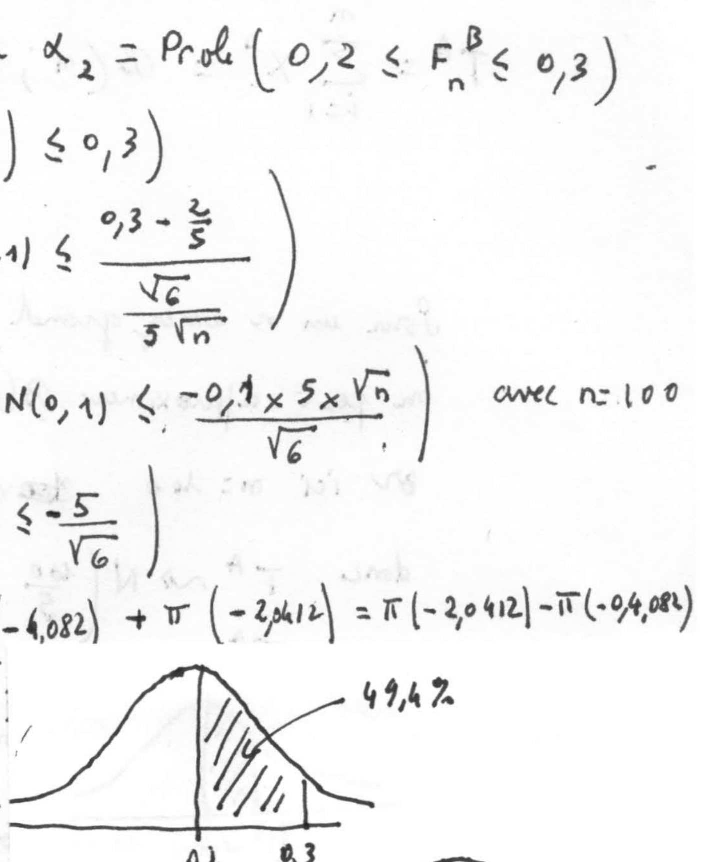
$$\alpha_2 \simeq \Pi\left(\frac{-5}{\sqrt{6}}\right) - \Pi\left(-\frac{10}{\sqrt{6}}\right) = -\Pi(-4,082) + \Pi(-2,0412) = \Pi(-2,0412) - \Pi(-4,082)$$

$$\alpha_2 \simeq 0,0206\,1335 - 0,00022$$

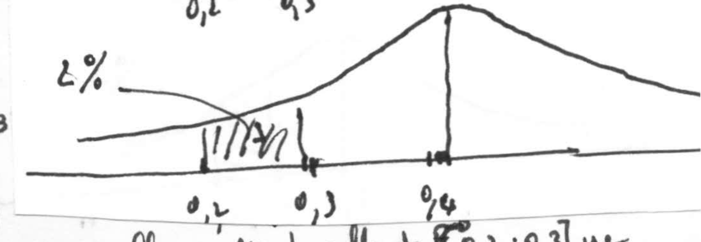
$$\alpha_2 \simeq 0,0205$$

$$\alpha_2 \simeq 2\ \%$$

F_n^A



F_n^B



les deux graphiques illustrent mal les proportions mais le passage de 49% à 2% lorsque on passe de F_n^A à F_n^B s'explique par la double raison suivante: La variance de F_n^A : $\left(\frac{4}{25n}\right)$ est nettement plus petite que celle de F_n^B $\left(\frac{6}{25n}\right)$ et par ailleurs l'intervalle de $[0,2 ; 0,3]$ ne contient pas l'espérance de F_n^B $0,4 \notin [0,2 ; 0,3]$ ce qui n'est pas le cas de F_n^A.

feuille 3 exo VII suite

4/ On recherche β tel que $\beta = Prob\left(F_n^A \geq F_n^B\right) = Prob\left(F_n^B - F_n^A \leq 0\right)$

$\beta = Prob(D \leq 0)$ avec $D = F_n^B - F_n^A$

loi de D   comme $F_n^A \leadsto N\left(0,2 \,;\, \frac{4}{25n}\right)$ et $F_n^B \leadsto N\left(0,4 \,;\, \frac{6}{25n}\right)$

donc $D = F_n^B - F_n^A$; F_n^A et F_n^B sont supposées indépendants

$D \leadsto N\left(E(F_n^B - F_n^A) \,;\, Var(F_n^B - F_n^A)\right)$

$E(D) = E(F_n^B - F_n^A) = E(F_n^B) - E(F_n^A)$ car Espérance linéaire

$= 0,4 - 0,2 = 0,2$

$Var(D) = Var(F_n^B - F_n^A) = Var(F_n^B) + Var(-F_n^A)$ car F_n^A, F_n^B indép

$= Var(F_n^B) + Var(F_n^A)$ car $Var(aX) = a^2 Var X$

$= \frac{6}{25n} + \frac{4}{25n}$

$= \frac{10}{25n}$

$Pr(D \leq 0) = Pr\left(\frac{D - 0,2}{\sqrt{\frac{10}{25n}}} \leq \frac{0 - 0,2}{\sqrt{\frac{10}{25n}}}\right) = Pr\left(N(0,1) \leq -\frac{0,2 \times 5 \times \sqrt{n}}{\sqrt{10}}\right)$

$= Pr\left(N(0,1) \leq -\frac{\sqrt{n}}{\sqrt{10}}\right) = Pr\left(N(0,1) \leq -\frac{10}{\sqrt{10}}\right)$

$= Pr\left(N(0,1) \leq -\sqrt{10}\right) = Prob\left(N(0,1) \leq -3,1623\right) \leq 0,0007828$

Conclusion : $\beta = Prob\left(F_n^A \geq F_n^B\right) = 0,0007828$

5/ On cherche n tel que $\beta \geq 0,05$ β étant $Prob\left(F_n^B \leq F_n^A\right) = Prob(D \leq 0)$

Ainsi on cherche n tel que $Prob\left(N(0,1) \leq -\frac{\sqrt{n}}{\sqrt{10}}\right) = 0,05$ en supposant n proche de 30 au moins

on lit sur la table de $N(0,1)$ $\frac{\sqrt{n}}{\sqrt{10}} = 1,6448548$ d'où $n = (1,6449)^2 \times 10$

soit $n \simeq 28$. A posteriori : n est inférieur à 30 donc ~~ce résultat~~ les approximations par la loi normale des lois de F_n^A et F_n^B sont à prendre avec précautions et il faut apprécier le résultat en connaissance de cause

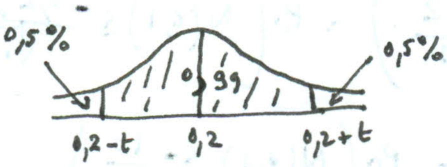
6/ Si on résume :

pour $n = 100$ Prob$\left(F_n^B \le F_n^A\right) \simeq 0{,}0008$

pour $n = 28$ Prob$\left(F_n^B \le F_n^A\right) \simeq 0{,}05$

lorsque on passe de $n=100$ à $n=28$ la probabilité de $\left(F_n^B \le F_n^A\right)$ est multipliée par plus de 50 fois. Ceci illustre le fait que plus l'échantillon est de grande taille, plus les résultats sur les fréquences F_n^B ou F_n^A sont proches des probabilités d'apparition des caractères pour les populations respectives, Prob$(X_i^A) = \frac{1}{5} = p_A$ et Prob$(X_i^B) = \frac{2}{5} = p_B$ et comme $p_B > p_A$ il devient de moins en moins probable, (soit de plus en plus rare) de trouver l'inverse $F_B \le F_A$ dans les échantillons qui ont des tailles assez grandes. Ce phénomène correspond bien à la loi faible des grands nombres.

3/ Intervalle de confiance de niveau 0,99 de F_n^A, $F_n^A \sim N\left(0{,}2 ; \frac{4}{25n}\right)$



cherchons t tq prob$\left(0{,}2 - t \le F_n^A \le 0{,}2 + t\right) = 0{,}99$

$E(F_n^A) = 0{,}2$

donc cherchons t tq Prob$\left(\frac{-t}{\sigma_{F_n^A}} \le N(0,1) \le \frac{t}{\sigma_{F_n^A}}\right) = 0{,}99$

après centrage et réduction.

soit cherchons t tq $\Pi\left(\frac{t}{\sigma_{F_n^A}}\right) - \Pi\left(-\frac{t}{\sigma_{F_n^A}}\right) = 2\Pi\left(\frac{t}{\sigma_{F_n^A}}\right) - 1 = 0{,}99$

soit $\Pi\left(\frac{t}{\sigma_{F_n^A}}\right) = \frac{1{,}99}{2} = 0{,}995$

on lit sur la table $\frac{t}{\sigma_{F_n^A}} = 2{,}57583$ soit $t = 2{,}57583\,\sigma_{F_n^A}$.

soit $t \simeq 2{,}58 \times \frac{2}{5\sqrt{n}} \simeq 2{,}58 \times \frac{2}{50}$ si $n = 100$

ainsi $t = 0{,}1030$

d'où l'intervalle de confiance de niveau 0,99 pour F_n^A.

$[0{,}2 - 0{,}1030 \; ; \; 0{,}2 + 0{,}1030]$

$[0{,}097 \; ; \; 0{,}303]$