Estimation.

I. Compléments sur les lois normale et binomiale.

1) Loi normale : intervalles centrés sur l'espérance.

Théorème

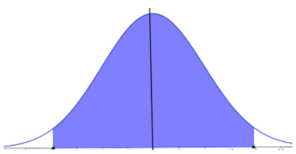
Soit Z une variable aléatoire qui suit la loi $N(\mu\,;\,\sigma^2)$. *Pour tout nombre* α *de* $]0\,;1[$ $P(\mu-\sigma u_\alpha \leq Z \leq \mu+\sigma u_\alpha) = 1-\alpha$
Avec $u_\propto$ tel que $P(-u_\alpha \leq \frac{Z-\mu}{\sigma} \leq u_\alpha) = 1 - \propto$

Démonstration :

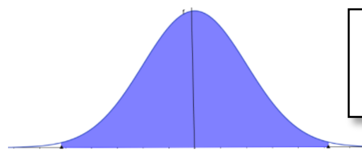
....

....

....



$u \quad =$
$P(\mu - 1{,}96\sigma \leq Z \leq \mu + 1{,}96\sigma) \approx$



$u \quad =$
$P(\mu - 2{,}58\sigma \leq Z \leq \mu + 2{,}58\sigma) \approx$

2) Loi de la fréquence de succès F_n.

X_n est une variable aléatoire qui suit la loi binomiale $B(n;p)$. La variable aléatoire $F_n = \frac{Xn}{n}$ indique la fréquence de succès lors des n épreuves.

Théorème

L'espérance et l'écart type de la variable sont $E(F_n) =$ ………… et $\sigma(F_n) =$ …………

Démonstration :

....

....

3) Approximation de la loi de F_n par une loi normale.

Règle

Sous les conditions $n \geq 30, np \geq 5$ *et* $n(1-p) \geq 5$.La loi des fréquences F_n peut-être approchée par la loi normale $N($… … … … ; … … … … $)$.

Exemple :

....

....

II. Prise de décision.

1) Intervalle de fluctuation asymptotique

Dans ce paragraphe, on suppose que la proportion *p* du caractère étudié est connue.

Définition

X_n est une variable aléatoire qui suit une loi binomiale $B(n\,;p)$. La variable aléatoire $F_n = \frac{X_n}{n}$ s'appelle la variable aléatoire fréquence de succès pour un schéma de Bernoulli de paramètres *n* et *p*.

Propriété

Soit $\alpha \in]0\,;1[$. On pose $I_n = \left[p - u_\propto \frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\propto \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ alors $\lim_{n\to+\infty} P(F_n \in I_n) = 1 - \propto$.

Démonstration ROC (exigible BAC) :

X_n suit la loi binomiale $B(n;p)$ donc la suite de variables aléatoires $Z_n = \frac{X_n - E(X_n)}{\sigma(X_n)} = \frac{X_n - np}{\sqrt{np(1-p)}}$ suit une loi normale centrée réduite $N(0;1)$ et d'après le théorème de Moivre-Laplace, on a :

$\lim_{n\to+\infty} P(a \leq Z_n \leq b) = \int_a^b \varphi(t)dt$, pour tous réels a et b avec $a < b$.

Soit $\alpha \in]0;1[$, il existe un unique réel positif u_α tel que $\lim_{n\to+\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = 1-\alpha$

Or $-u_\alpha \leq Z_n \leq u_\alpha \leftrightarrow -u_\alpha \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq u_\alpha \leftrightarrow p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq \frac{X_n}{n} \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}$

Donc $\lim_{n\to+\infty} P(F_n \in I_n) = \lim_{n\to+\infty} P(-u_\alpha \leq Z_n \leq u_\alpha) = 1-\alpha$

2) Intervalle de fluctuation au seuil de 95%

Définition

On appelle intervalle de fluctuation au seuil 0,95 de la variable aléatoire fréquence l'intervalle :

$$I = \left[p - 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \; ; \; p + 1{,}96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

Exemple :

On dispose d'une urne contenant un grand nombre de boules blanches et noires. La proportion de boules blanches contenues dans l'urne est p = 0,3. On tire successivement avec remise n = 50 boules. Soit X_{50} la variable aléatoire dénombrant le nombre de boules blanches tirées. X_{50} suit la loi binomiale $B(50 ; 0{,}3)$.

$I_{50} = \left[0{,}3 - 1{,}96 \times \frac{\sqrt{0{,}3\times 0{,}7}}{\sqrt{50}} ; 0{,}3 + 1{,}96\times\frac{\sqrt{0{,}3\times 0{,}7}}{\sqrt{50}}\right]$ Soit $I_{50} = [0{,}173 ; 0{,}427]$.

En effectuant 50 tirages dans cette urne, la fréquence d'apparition d'une boule blanche est comprise dans l'intervalle [0,173 ; 0,427] avec une probabilité de 0,95.

Pour 500 tirages, on obtient : $I_{500} = \left[0{,}3 - 1{,}96 \times \frac{\sqrt{0{,}3\times 0{,}7}}{\sqrt{500}} ; 0{,}3 + 1{,}96\times\frac{\sqrt{0{,}3\times 0{,}7}}{\sqrt{500}}\right] = [0{,}26 ; 0{,}34]$

On constate que l'intervalle, pour un même seuil, se resserre fortement lorsqu'on augmente le nombre de tirages.

3) Prise de décision

Dans ce paragraphe, la proportion du caractère étudié n'est pas connue mais est supposée être égale à p. La prise de décision consiste à valider ou invalider l'hypothèse faite sur la proportion p.

Règle de décision

Soit f la fréquence du caractère étudié d'un échantillon de taille n.
Soit l'hypothèse : "La proportion de ce caractère dans la population est p."
Soit I l'intervalle de fluctuation asymptotique au seuil 0,95.
- Si $f \in I$, alors on accepte l'hypothèse faite sur la proportion p.
- Si $f \notin I$, alors on rejette l'hypothèse faite sur la proportion p.

Remarque :
On peut interpréter cette propriété par le fait que la probabilité qu'on rejette à tort l'hypothèse sur p sachant qu'elle est vraie (le risque d'erreur) est approximativement égale à 5%.

☑*Savoir-faire : Savoir prendre une décision à l'aide d'un intervalle de fluctuation:*

Un fabricant d'alarme commande auprès de son fournisseur deux types de composants électroniques : R1 et P4. Il demande 900 composants de chaque sorte. Au moment de la livraison, le service de contrôle retire 50 composants et constate que 19 sont des modèles R1.
Peut-on affirmer que la commande est respectée par le fournisseur ?

……
……
……
……
……
……

III. Estimation.

Dans ce paragraphe, on suppose que la proportion *p* du caractère étudié est inconnue. C'est le problème inverse de celui de l'échantillonnage. A partir de la fréquence observée sur un échantillon de taille *n*, on va estimer la proportion *p* d'un caractère dans la population tout entière. On suppose que $n \geq 30, np \geq 5$ *et* $n(1-p) \geq 5$

Propriété

X_n est une variable aléatoire qui suit une loi binomiale $B(n\,;p)$ et F_n est la fréquence associée à X_n. Pour n suffisamment grand, $P\left(F_n - \frac{1}{\sqrt{n}} \leq p \leq F_n + \frac{1}{\sqrt{n}}\right) \approx 0{,}95$

Définition

Soit f une fréquence observée du caractère étudié sur un échantillon de taille n. On appelle <u>intervalle de confiance</u> de la proportion *p* au <u>niveau de confiance</u> 0,95, l'intervalle $\left[f - \frac{1}{\sqrt{n}}\,;f + \frac{1}{\sqrt{n}}\right]$.

<u>Remarque :</u>

- Un niveau de confiance 0,95 signifie que dans 95 cas sur 100, on affirme à juste titre que *p* appartient à l'intervalle de confiance.
- Il n'est pas vrai d'affirmer que *p* est égal au centre de l'intervalle de confiance. Il n'est pas possible d'évaluer la position de *p* dans l'intervalle de confiance.

☑*<u>Savoir-faire : Savoir estimer une proportion inconnue par un intervalle de confiance:</u>*

Un institut de sondage interroge 1052 personnes entre les deux tours de l'élection présidentielle sur leur intention de vote.614 déclarent avoir l'intention de voter pour Martine Phinon. En supposant que les votes seront conformes aux intentions, la candidate a-t-elle raison de croire qu'elle sera élue ?

……
……
……
……
……
……
……
……

☑*<u>Savoir-faire : Savoir déterminer une taille d'échantillon suffisante pour obtenir une estimation d'une proportion:</u>*

Un constructeur automobile fait appel à un institut de sondage afin de mesurer le degré de satisfaction du service après-vente. L'institut souhaite estimer la proportion de clients satisfaits au niveau de confiance 0,95 avec une amplitude d'au plus 5 centièmes. Combien de personnes au minimum faut-il interroger ?

……
……
……
……
……

☑ Exercice Bac : Métropole 2016 / 6 points:

Partie A

Une usine fabrique un composant électronique. Deux chaînes de fabrication sont utilisées.
La chaîne A produit 40 % des composants et la chaîne B produit le reste.

Une partie des composants fabriqués présentent un défaut qui les empêche de fonctionner à la vitesse prévue par le constructeur. En sortie de chaîne A, 20 % des composants présentent ce défaut alors qu'en sortie de chaîne B, ils ne sont que 5 %.

On choisit au hasard un composant fabriqué dans cette usine.
On note :

A l'évènement « le composant provient de la chaîne A »

B l'évènement « le composant provient de la chaîne B »

S l'évènement « le composant est sans défaut »

1. Montrer que la probabilité de l'évènement S est $P(S) = 0,89$.
2. Sachant que le composant ne présente pas de défaut, déterminer la probabilité qu'il provienne de la chaîne A. On donnera le résultat à 10^{-2} près.

Partie B

Des améliorations apportées à la chaîne A ont eu pour effet d'augmenter la proportion p de composants sans défaut.

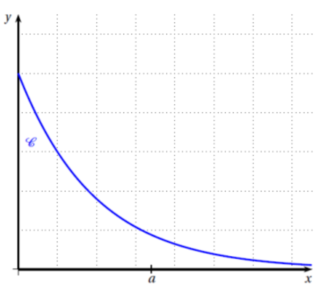
Afin d'estimer cette proportion, on prélève au hasard un échantillon de 400 composants parmi ceux fabriqués par la chaîne A.
Dans cet échantillon, la fréquence observée de composants sans défaut est de 0,92.

1. Déterminer un intervalle de confiance de la proportion p au niveau de confiance de 95 %.
2. Quelle devrait être la taille minimum de l'échantillon pour qu'un tel intervalle de confiance ait une amplitude maximum de 0,02 ?

Partie C

La durée de vie, en années, d'un composant électronique fabriqué dans cette usine est une variable aléatoire T qui suit la loi exponentielle de paramètre λ (où λ est un nombre réel strictement positif).
On note f la fonction densité associée à la variable aléatoire T. On rappelle que :
— pour tout nombre réel $x \geqslant 0$, $f(x) = \lambda e^{-\lambda x}$.
— pour tout nombre réel $a \geqslant 0$, $p(T \leqslant a) = \int_0^a f(x)\,dx$.

1. La courbe représentative $\mathscr{C}$ de la fonction f est donnée ci-dessous.



   a. Interpréter graphiquement $P(T \leqslant a)$ où $a > 0$.
   b. Montrer que pour tout nombre réel $t \geqslant 0$: $P(T \leqslant t) = 1 - e^{-\lambda t}$.
   c. En déduire que $\lim_{t \to +\infty} P(T \leqslant t) = 1$.
2. On suppose que $P(T \leqslant 7) = 0,5$. Déterminer λ à 10^{-3} près.
3. Dans cette question on prend $\lambda = 0,099$ et on arrondit les résultats des probabilités au centième.
   a. On choisit au hasard un composant fabriqué dans cette usine.
   Déterminer la probabilité que ce composant fonctionne au moins 5 ans.
   b. On choisit au hasard un composant parmi ceux qui fonctionnent encore au bout de 2 ans.
   Déterminer la probabilité que ce composant ait une durée de vie supérieure à 7 ans.
   c. Donner l'espérance mathématique $E(T)$ de la variable aléatoire T à l'unité près.
   Interpréter ce résultat.*

☑ Exercice Bac : Centre étrangers 2016 / 5 points:

Un institut effectue un sondage pour connaître, dans une population donnée, la proportion de personnes qui sont favorables à un projet d'aménagement du territoire. Pour cela, on interroge un échantillon aléatoire de personnes de cette population, et l'on pose une question à chaque personne.

Les trois parties sont relatives à cette même situation, mais peuvent être traitées de manière indépendante.

Partie A : Nombre de personnes qui acceptent de répondre au sondage

On admet dans cette partie que la probabilité qu'une personne interrogée accepte de répondre à la question est égale à 0,6.

1. L'institut de sondage interroge 700 personnes. On note X la variable aléatoire correspondant au nombre de personnes interrogées qui acceptent de répondre à la question posée.
   - **a.** Quelle est la loi de la variable aléatoire X ? Justifier la réponse.
   - **b.** Quelle est la meilleure approximation de $P(X \geqslant 400)$ parmi les nombres suivants ?

   $0,92 \qquad 0,93 \qquad 0,94 \qquad 0,95.$

2. Combien de personnes l'institut doit-il interroger au minimum pour garantir, avec une probabilité supérieure à 0,9, que le nombre de personnes répondant au sondage soit supérieur ou égal à 400.

Partie B : Proportion de personnes favorables au projet dans la population

Dans cette partie, on suppose que n personnes ont répondu à la question, et on admet que ces personnes constituent un échantillon aléatoire de taille n (où n est un entier naturel supérieur à 50).
Parmi ces personnes, 29 % sont favorables au projet d'aménagement.

1. Donner un intervalle de confiance, au niveau de confiance de 95 %, de la proportion de personnes qui sont favorables au projet dans la population totale.
2. Déterminer la valeur minimale de l'entier n pour que l'intervalle de confiance, au niveau de confiance de 95 %, ait une amplitude inférieure ou égale à 0,04.

Partie C : Correction due à l'insincérité de certaines réponses

Dans cette partie, on suppose que, parmi les personnes sondées qui ont accepté de répondre à la question posée, 29 % affirment qu'elles sont favorables au projet.
L'institut de sondage sait par ailleurs que la question posée pouvant être gênante pour les personnes interrogées, certaines d'entre elles ne sont pas sincères et répondent le contraire de leur opinion véritable. Ainsi, une personne qui se dit favorable peut :

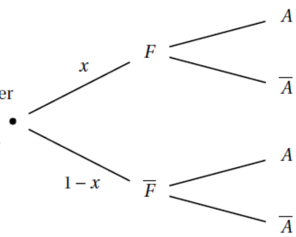
— soit être en réalité favorable au projet si elle est sincère.
— soit être en réalité défavorable au projet si elle n'est pas sincère.

Par expérience, l'institut estime à 15 % le taux de réponses non sincères parmi les personnes ayant répondu, et admet que ce taux est le même quelle que soit l'opinion de la personne interrogée.
Le but de cette partie est, à partir de ces données, de déterminer le taux réel de personnes favorables au projet, à l'aide d'un modèle probabiliste. On prélève au hasard la fiche d'une personne ayant répondu, et on définit :

- F l'évènement « la personne est en réalité favorable au projet » ;
- $\overline{F}$ l'évènement « la personne est en réalité défavorable au projet » ;
- A l'évènement « la personne affirme qu'elle est favorable au projet » ;
- $\overline{A}$ l'évènement « la personne affirme qu'elle est défavorable au projet ».

Ainsi, d'après les données, on a $p(A) = 0,29$.

1. En interprétant les données de l'énoncé, indiquer les valeurs de $P_F(A)$ et $P_{\overline{F}}(A)$.
2. On pose $x = P(F)$.
   - **a.** Reproduire sur la copie et compléter l'arbre de probabilité ci-contre.
   - **b.** En déduire une égalité vérifiée par x



3. Déterminer, parmi les personnes ayant répondu au sondage, la proportion de celles qui sont réellement favorables au projet.*