Echantillonnage
Corrigés de quelques exemples
Terminale S 734

Frédéric Junier[1]

Lycée du Parc, Lyon

1. http://frederic-junier.org/

Table des matières

Exemple 1 : Intervalle de fluctuation exact

Soit une variable aléatoire X_n suivant une loi binomiale $\mathcal{B}(n\ ;\ p)$ mesurant le nombre de succès sur un échantillon de taille n c'est-à-dire le nombre d'apparition d'un caractère de proportion p dans la population totale.

L'**intervalle de fluctuation exacte** au seuil de 95% pour la variable aléatoire fréquence $\dfrac{X_n}{n}$ mesurant la fréquence de succès dans l'échantillon de taille n est $\left[\dfrac{a}{n}\,;\ \dfrac{b}{n}\right]$ où :

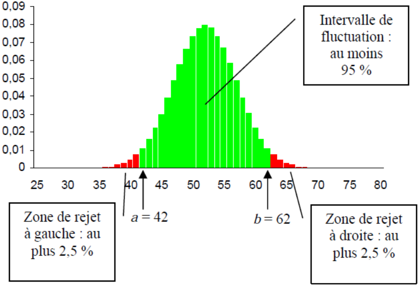
- a est le plus petit entier tel que $\mathrm{P}(X_n \leqslant a) > 0,025$;
- b est le plus petit entier tel que $\mathrm{P}(X_n \leqslant b) \geqslant 0,975$.

On détermine a et b avec un algorithme dont on donne ci-après une implémentation pour calculatrices TI ou Casio. Une autre implémentation est donnée dans l'Annexe du cours, elle est plus rapide car le calcul de $\mathrm{P}(X_n \leqslant k)$ n'est pas reprise à chaque itération.

Exemple d'utilisation d'un intervalle de fluctuation exact au seuil de 95% dans un test d'hypothèse.

4. On considère que l'affirmation de Monsieur Z est exacte.

X suit la loi $\mathscr{B}(100 ; 0,52)$



Remarque : la recherche de l'intervalle de fluctuation peut-être illustrée par le diagramme en bâtons de la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

Soit X une variable aléatoire suivant la loi binomiale $\mathcal{B}\,(100\ ;\ 0,3)$.
Pour la variable aléatoire fréquence $\dfrac{X}{100}$, le programme ci-après donne l'intervalle de fluctuation exact au seuil de 95% :

$$\left[\frac{21}{100}\,;\,\frac{39}{100}\right]$$

Intervalle de fluctuation exact, programme TI

Programme TEXAS

```
Prompt N
Prompt P
0→ K
While binomFRep(N,P,K)⩽ 0.025
K+1→K
End
Disp K
While binomFRep(N,P,K)< 0.975
K+1→K
End
Disp K
```

Intervalle de fluctuation exact, programme CASIO

```
Programme CASIO
?→ N
?→ P
0→ K
While binominalCD(K,N,P)⩽ 0.025
K+1→K
WhileEnd
K ◢
While binominalCD(K,N,P)< 0.975
K+1→K
WhileEnd
K ◢
```

Intervalle de fluctuation exact, programme Python

```python
from scipy.stats import binom

def binomFrep(n, p, k):
    """P(X <= k) si X suit la loi B(n,p)"""
    return binom.cdf(k, n, p)


def if_exact(n, p):
    k = 0
    while binomFrep(n, p, k) <= 0.025:
        k = k + 1
    binf = k
    while binomFrep(n, p, k) < 0.975:
        k = k + 1
    bsup = k
    return [binf, bsup]
```

Exemple 2 : Intervalle de fluctuation exact

Dans une maternité, on admet qu'il naît en moyenne 51 % des garçons. On fait le point sur la proportion de garçons toutes les 100 naissances.

La variable aléatoire X donnant le nombre de garçons dans un échantillon de 100 naissances, suit une loi binomiale de paramètres $n = 100$ et $p = 0,51$, car nous sommes en présence d'une répétition de 100 épreuves de Bernoulli identiques et indépendantes dont la probabilité de succès (naissance d'un garçon) est de $= 0,51$.

Le programme réalisé à l'exemple 1 retourne l'intervalle de fluctuation exact au seuil de 95% pour la fréquence $\dfrac{X}{100}$ de garçons dans un échantillon de taille $n = 100$:

$$\left[\frac{41}{100} ; \frac{61}{100}\right]$$

Exemple 2 : Intervalle de fluctuation asymptotique (0/5)

```python
from scipy.stats import norm
from math import sqrt

def invNorm(s):
    """Retourne u tel que que P(X <= u)= s
        si X suit la loi N(0,1)"""
    return norm.ppf(s)


def if_asymptotique(n, p, s=0.95):
    u = invNorm((1 + s) / 2)
    binf = p - u * np.sqrt(p * (1 - p) / n)
    bsup = p + u * np.sqrt(p * (1 - p) / n)
    return [binf, bsup]
```

Exemple 2 : Intervalle de fluctuation asymptotique (1/5)

Soit une variable aléatoire X_n suivant la loi binomiale $\mathcal{B}\,(n\ ;\ p)$ avec p dans l'intervalle $]0\,;\,1[$ et soit

$\mathrm{I}_n = \left[p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ **un intervalle de fluctuation « asymptotique au seuil de** $1-\alpha$ **»** de la variable aléatoire fréquence $F_n = \dfrac{X_n}{n}$.

Sous les conditions, $n \geqslant 30$, $np \geqslant 5$, $n(1-p) \geqslant 5$, on peut utiliser l'approximation $\mathrm{P}\big(F_n \in \mathrm{I}_n\big) \approx 1-\alpha$. Ainsi, l'intervalle de fluctuation asymptotique I_n peut être considéré comme un intervalle de fluctuation de F_n au seuil de $1-\alpha$.

Un cas particulier important est celui de l'intervalle de fluctuation asymptotique au seuil de $0,95$ qui

est $\left[\boldsymbol{p - 1,96 \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96 \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}}\right]$.

Exemple 2 : Intervalle de fluctuation asymptotique (2/5)

On considère ici une variable aléatoire suivant une loi binomiale $\mathcal{B}(100\ ;\ 0,51)$.
Les conditions d'approximation usuelles sont vérifiées :

- $n \geqslant 30$ car $n = 100$;
- $np \geqslant 5$ car $np = 100 \times 0,51 = 51$;
- $n(1-p) \geqslant 5$ car $n(1-p) = 100 \times 0,49 = 49$.

On peut donc utiliser l'intervalle de fluctuation asymptotique comme approximation d'un intervalle de fluctuation au seuil de 95% pour la variable aléatoire fréquence $\dfrac{X}{100}$:

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

Exemple 2 : Intervalle de fluctuation asymptotique (3/5)

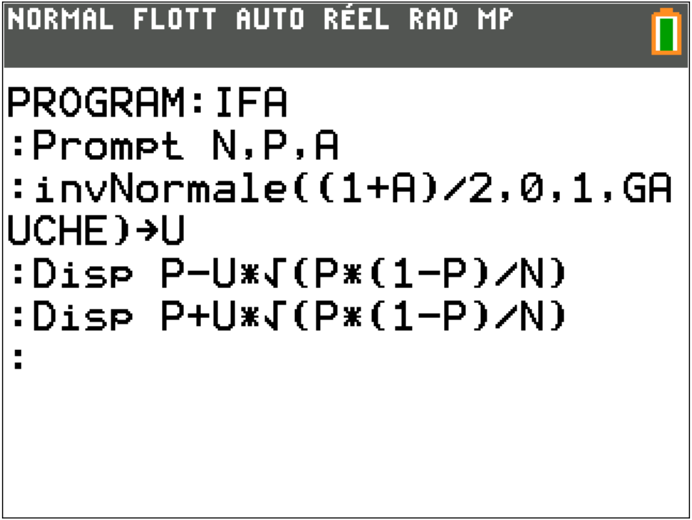
En pratique on calcule l'intervalle de fluctuation asymptotique avec le programme ci-après (saisir $N = 100$, $P = 0.51$ et $A = 0.95$) et on obtient :

$$\left[0,51 - 1,96\frac{\sqrt{0,51(1-0,51)}}{\sqrt{100}}\ ;\ 0,51 + 1,96\frac{\sqrt{0,51(1-0,51)}}{\sqrt{100}}\right]$$

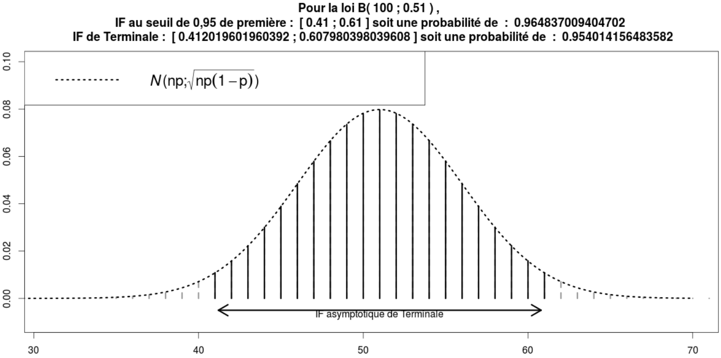
En arrondissant par défaut la borne inférieure et par excès la borne supérieure à $0,001$ près :

$$\boxed{[0,412\,;\ 0,608]}$$

Exemple 2 : Intervalle de fluctuation asymptotique (4/5)



Exemple 2 : Intervalle de fluctuation asymptotique (5/5)



Exemple 2 : Intervalles de fluctuation asymptotique avec d'autres seuils que $0,95$

La proportion d'un caractère dans une population est $p = 0,6$. Déterminons un intervalle de fluctuation asymptotique de la fréquence de ce caractère dans les échantillons de taille 100, prélevés au hasard et avec remise :

1. au seuil de $0,8$: $\boxed{[0,537\,;\,0,663]}$;
2. au seuil de $0,9$: $\boxed{[0,519\,;\,0,681]}$.

Exemple 3 Partie 1

Exemple 3 Partie 1

- On s'intéresse au caractère « Favorable à la coupure de l'éclairage nocturne »

Exemple 3 Partie 1

- On s'intéresse au caractère « Favorable à la coupure de l'éclairage nocturne »
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,5$.

Exemple 3 Partie 1

- On s'intéresse au caractère « Favorable à la coupure de l'éclairage nocturne »
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,5$.
- La taille de l'échantillon est $n = 100$.

Exemple 3 Partie 1

- On s'intéresse au caractère « Favorable à la coupure de l'éclairage nocturne »
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,5$.
- La taille de l'échantillon est $n = 100$.
- Les conditions d'approximation usuelles $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont réunies et permettent d'utiliser $\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ comme intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n.

Exemple 3 Partie 2

Exemple 3 Partie 2

- À 10^{-3} près un intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n est $[0,402\,;\,0,598]$.

Exemple 3 Partie 2

- À 10^{-3} près un intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n est $[0,402\,;\,0,598]$.
- La fréquence mesurée sur l'échantillon est de $0,54$, elle est largement à l'intérieur de l'intervalle de fluctuation donc on peut accepter l'hypothèse que la proportion du caractère dans la population est $p = 0,5$.

Exemple 4 Partie 1

Exemple 4 Partie 1

- On s'intéresse au caractère « Conforme » d'un médicament produit.

Exemple 4 Partie 1

- On s'intéresse au caractère « Conforme » d'un médicament produit.
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,97$.

Exemple 4 Partie 1

- On s'intéresse au caractère « Conforme » d'un médicament produit.
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,97$.
- La taille de l'échantillon est $n = 1000$.

Exemple 4 Partie 1

- On s'intéresse au caractère « Conforme » d'un médicament produit.
- On fait l'hypothèse que la proportion de ce caractère dans la population totale est de $p = 0,97$.
- La taille de l'échantillon est $n = 1000$.
- Les conditions d'approximation usuelles $n \geqslant 30$, $np \geqslant 5$ et $n(1-p) \geqslant 5$ sont réunies et permettent d'utiliser $\left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\,;\, p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ comme intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n.

Exemple 4 Partie 2

Exemple 4 Partie 2

- À 10^{-3} près un intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n est $[0,959\,;\,0,981]$.

Exemple 4 Partie 2

- À 10^{-3} près un intervalle de fluctuation asymptotique au seuil de $0,95$ de la fréquence du caractère sur un échantillon de taille n est $[0,959\,;\,0,981]$.
- La fréquence mesurée sur l'échantillon est de $0,947$, elle est en dehors de l'intervalle de fluctuation donc on peut rejeter l'hypothèse que la proportion du caractère dans la population est $p = 0,97$ avec un risque d'erreur (faux positif) de $0,05$.

Exemple 5 Partie 1

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. On interroge un échantillon aléatoire de n personnes de cette population où n est un entier naturel supérieur à 50. Parmi ces personnes, une fréquence $f = 0,29$ est favorable au projet d'aménagement.

Exemple 5 Partie 1

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. On interroge un échantillon aléatoire de n personnes de cette population où n est un entier naturel supérieur à 50. Parmi ces personnes, une fréquence $f = 0,29$ est favorable au projet d'aménagement.

- OLes conditions d'approximation usuelle $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 4$ sont vérifiées et permettent d'utiliser l'intervalle $\left[f - \frac{1}{\sqrt{n}}\,;\, f + \frac{1}{\sqrt{n}}\right]$ pour estimer la probabilité du caractère dans la population totale au niveau de confiance $0,95$.

Exemple 5 Partie 1

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. On interroge un échantillon aléatoire de n personnes de cette population où n est un entier naturel supérieur à 50. Parmi ces personnes, une fréquence $f = 0,29$ est favorable au projet d'aménagement.

- OLes conditions d'approximation usuelle $n \geqslant 30$, $nf \geqslant 5$ et $n(1-f) \geqslant 4$ sont vérifiées et permettent d'utiliser l'intervalle $\left[f - \dfrac{1}{\sqrt{n}}\,;\, f + \dfrac{1}{\sqrt{n}}\right]$ pour estimer la probabilité du caractère dans la population totale au niveau de confiance $0,95$.
- L'amplitude de l'intervalle de confiance est $\dfrac{2}{\sqrt{n}}$. Le plus petit entier n tel que $\dfrac{2}{\sqrt{n}} \leqslant 0,04 \Leftrightarrow 50 \leqslant \sqrt{n} \Leftrightarrow 2500 \leqslant n$. Le seuil recherché est donc $n = 2500$.