Intervalle de fluctuation et intervalle de confiance

1 Prise de décision, intervalle de fluctuation

1.1 Présentation du problème

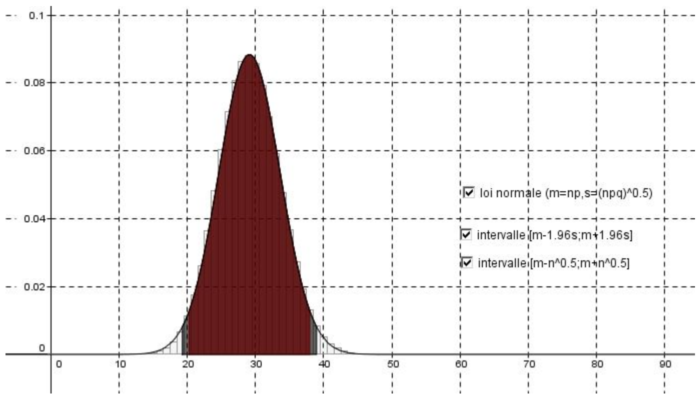
Soit X une variable qui suit une loi de Binomiale de paramètres $(n; p)$, où n est connue ($n \geqslant 30$) et p est connue. On prélève un échantillon de la population (une partie de la population observée), on souhaite savoir si la fréquence observée du succès est significative et on procédera alors à une prise de décision qui a 95% de chance d'être la bonne décision.

1. Dans une population d'un pays, il y a autant d'hommes que de femmes, on cherche si une entreprise de 350 salariés respecte la parité (autant d'hommes que de femmes) sachant que 162 femme y travaillent.
   Mise en place du problème :
   L'entreprise représente un échantillon de la population, X est la variable qui donne le nombre de femmes qui travaillent dans une entreprise (à chaque "tirage" d'un salarié, il y a bien deux issues, le salarié est une femme ou non, et on admet que chaque tirage est indépendant de l'autre et identique, les tirages sont assimilés à des tirages avec remise), X suit une loi Binomiale de paramètres $(350; 0,5)$, sachant qu'il y a 162 femmes.
2. On lance 100 fois un dé à six faces (il s'agit d'une simulation), on cherche à savoir si la simulation est fiable. Pour se faire on compte le nombre de fois que 6 est apparu. Mise en place du problème :
   L'échantillon de la simulation est de taille $n = 100$, X est la variable qui donne le nombre de fois que 6 peut apparaître (à chaque "tirage" d'un dé, il y a bien deux issues, 6 est apparu ou non, et on admet que chaque tirage est indépendant de l'autre et identique, les tirages sont assimilés à des tirages avec remise), X suit une loi Binomiale de paramètres $(100; \frac{1}{6})$.

1.2 Résolution du problème

Par approximation de la loi Binomiale par la loi Normale, on sait que 95% des valeurs de X sont dans l'intervalle $[m - 1,96\sigma; m + 1,96\sigma]$ avec $m = np$ et $\sigma = \sqrt{npq}$, $q = 1 - p$.
La fréquence observée de l'échantillon est f, la variable de la fréquence est $\frac{X}{n}$.
Ainsi 95% des valeurs de $\frac{X}{n}$ sont dans $[\frac{m-1,96\sigma}{n}; \frac{m+1,96\sigma}{n}]$, après simplification,
95% des valeurs de $\frac{X}{n}$ sont dans $[p - 1,96\frac{\sqrt{pq}}{\sqrt{n}}; p + 1,96\frac{\sqrt{pq}}{\sqrt{n}}]$.
On admet que l'intervalle $[p - 1,96\frac{\sqrt{pq}}{\sqrt{n}}; p + 1,96\frac{\sqrt{pq}}{\sqrt{n}}]$ est contenu dans l'intervalle $[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}]$.

Figure :



Propriété :
Par approximation de la loi Binomiale de paramètres $(n; p)$ par la loi Normale de paramètres $(np; \sqrt{npq})$, 95% des valeurs de $\frac{X}{n}$ sont dans l'intervalle

$$[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}] \quad (1)$$

Cet intervalle est appelé intervalle de fluctuation.

Exemples :

1. Dans une population d'un pays, il y a autant d'hommes que de femmes, on cherche si une entreprise de 350 salariés respecte la parité (autant d'hommes que de femmes) sachant que 162 femme y travaillent.
   Mise en place du problème :
   X est la variable qui donne le nombre de femmes qui travaille dans une entreprise, X suit une loi Binomiale de paramètres $(350; 0,5)$, sachant qu'il y a 162 femmes.
   (a) Calculer l'intervalle de fluctuation.
   (b) Quelle est la probabilité pour que la fréquence f du nombre femmes dans l'entreprise soit dans l'intervalle de fluctuation ?
   (c) Peut-on considérer que l'entreprise respecte la parité ?
2. On lance 100 fois un dé à six faces (il s'agit d'une simulation), on cherche à savoir si la simulation est fiable. Pour se faire on compte le nombre de fois que 6 est apparu. Mise en place du problème :
   L'échantillon de la simulation est de taille $n = 100$, X est la variable qui donne le nombre de fois que 6 peut apparaître, X suit une loi Binomiale de paramètres $(100; \frac{1}{6})$.
   (a) Calculer l'intervalle de fluctuation.
   (b) à l'aide du tableur, faire une simulation de l'échantillon. (=ENT(ALEA()*6+1) est la formule qui permet de simuler le lancé d'un dé.)
   (c) Quelle est la probabilité pour que la fréquence f du nombre de 6 dans l'échantillon soit dans l'intervalle de fluctuation ?
   (d) Peut-on considérer que l'échantillon observé est fiable ?
   (e) (avec la touche F9 du tableur, on peut obtenir un autre échantillon facilement).

2 Sondage, intervalle de confiance

2.1 Présentation du problème

Soit X une variable qui suit une loi de Binomiale de paramètres $(n; p)$, où n est connue ($n \geqslant 30$) et p est inconnu. Le but est d'estimer la valeur de p.

Exemples :

- On souhaite connaitre la proportion p d'élèves de terminale, en France, qui fument régulièrement.
  Mise en place du problème :
  On peut faire une estimation de la valeur de p en choisissant un échantillon de la population (c'est à dire une partie de la population concernée, ici les élèves de seconde de France) d'une taille supérieure à 30 :
  En choisissant votre classe de terminale représentative des élèves de terminale, on a $n = 35$.
  X est le nombre d'élèves de seconde qui fument (à chaque "tirage" d'un élève, il y a bien deux issues, l'élève fume ou non, et on admet que chaque tirage est indépendant de l'autre et identique, les tirages sont assimilés à des tirages avec remise), X suit une loi Binomiale de paramètres $(35; p)$.
- On souhaite connaitre la proportion p de chance d'obtenir un 6 avant 3 coups lorsqu'on lance un dé à six faces (non truqué).
  Mise en place du problème :
  On peut faire une estimation de la valeur de p en choisissant un échantillon de la population d'une taille supérieure à 30 :
  On choisit une simulation de plus de 30 cas, par exemple $n = 100$ (la simulation se fait sur tableur). X est le nombre de cas où 6 est apparu avant les trois coups (à chaque "tirage" de cas, il y a bien deux issues, 6 apparaît avant 3 coups ou non, et on admet que chaque tirage est indépendant de l'autre et identique, les tirages sont assimilés à des tirages avec remise), X suit une loi Binomiale de paramètres $(100; p)$.

2.2 Résolution du problème

Pour estimer p, on va prendre un échantillon de la population observée, et observer la fréquence f du succès dans cet échantillon.

Propriété :
La probabilité pour que la probabilité p observée soit dans l'intervalle

$$[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}] \tag{2}$$

est de $0,95\%$. Cet intervalle est appelé intervalle de confiance.
Indication de démonstration :
$f \in [p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}}] \Leftrightarrow p - \frac{1}{\sqrt{n}} \leqslant f \leqslant p + \frac{1}{\sqrt{n}} \Leftrightarrow -f - \frac{1}{\sqrt{n}} \leqslant -p \leqslant -f + \frac{1}{\sqrt{n}}$
$\Leftrightarrow f - \frac{1}{\sqrt{n}} \leqslant p \leqslant f + \frac{1}{\sqrt{n}} \Leftrightarrow p \in [f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$
Exemples :

- On souhaite connaitre la proportion p d'élèves de terminale, en France, qui fument régulièrement.
  Mise en place du problème :
  On peut faire une estimation de la valeur de p en choisissant un échantillon de la population (c'est à dire une partie de la population concernée, ici les élèves de seconde de France) d'une taille supérieure à 30 :
  En choisissant votre classe de terminale représentative des élèves de terminale, on a $n = 35$.
  X est le nombre d'élèves de seconde qui fument, X suit une loi Binomiale de paramètres $(35; p)$.
  1. Compter le nombre de fumeurs dans la classe, en déduire la fréquence f de fumeurs.
  2. Quelle est la probabilité que p soit dans l'intervalle confiance $[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$?
  3. Calculer cet intervalle de confiance.
  4. Que pouvez dire de la valeur de p ?
- On souhaite connaitre la proportion p de chance d'obtenir un 6 avant 3 coups lorsqu'on lance un dé à six faces (non truqué).
  Mise en place du problème :
  On peut faire une estimation de la valeur de p en choisissant un échantillon de la population d'une taille supérieure à 30 :
  On choisit une simulation de plus de 30 cas, par exemple $n = 100$ (la simulation se fait sur tableur). X est le nombre de cas où 6 est apparu avant les trois coups (à chaque "tirage" de cas, il y a bien deux issues, 6 apparaît avant 3 coups ou non, et on admet que chaque tirage est indépendant de l'autre et identique, les tirages sont assimilés à des tirages avec remise), X suit une loi Binomiale de paramètres $(100; p)$.

1. Compter le nombre cas favorables (c'est à dire où le six est sorti avant les 3 coups) dans la simulation, en déduire la fréquence f de ces cas favorables.
2. Quelle est la probabilité que p soit dans l'intervalle confiance $[f - \frac{1}{\sqrt{n}}; f + \frac{1}{\sqrt{n}}]$?
3. Calculer cet intervalle de confiance.
4. Que pouvez dire de la valeur de p ?