CHAPITRE 10 : Echantillonnage - Estimation

[1 Effectifs et fréquences avec la loi binomiale 2](#_Toc460274435)

[2 La valeur de la proportion *p* dans la population est connue. On prédit un intervalle de fluctuation *In* contenant la fréquence *Fn* 5](#_Toc460274436)

[2.1 Définition de l’intervalle de fluctuation à 95 % avec la loi binomiale 5](#_Toc460274437)

[2.2 Définition de *In* l’intervalle de fluctuation à 95 % avec la loi normale 5](#_Toc460274438)

[2.3 Intervalle *In* de fluctuation asymptotique au seuil de (1-α)% 8](#_Toc460274439)

[RÉSUMÉ : Échantillonnage dans une population où *p* est connue 10](#_Toc460274440)

[3 La valeur de la proportion *p* dans la population est une hypothèse qu’on veut tester 11](#_Toc460274441)

[4 La valeur de la proportion *p* dans la population est une inconnue qu’on estime par un intervalle de confiance *Ic* 13](#_Toc460274442)

[RÉSUMÉ : Estimation par un intervalle de confiance de la proportion *p* qui reste inconnue 17](#_Toc460274443)

CHAPITRE 10 : Echantillonnage - Estimation

# Effectifs et fréquences avec la loi binomiale

***Notation :***

Prélever un échantillon de taille dans une population très nombreuse d’individus possédant un caractère considéré comme « succès » avec la probabilité , revient à effectuer épreuves de Bernoulli identiques et indépendantes.

* On note la variable aléatoire qui donne le nombre de succèsdans l’échantillon :
, avec et .
* On note la variable aléatoire qui donne la fréquence de succès dans l’échantillon :
 avec et .

***Exemple 1 : Fréquence dans un échantillon***

Soit une population d’individus telle que personnes sur aient la grippe.

1. On prend un échantillon de taille personnes dans cette population. Déterminer la probabilité d’observer un nombre de malades dans l’échantillon tel que .

*Réponse :*

* Choisir personnes dans une population importante revient à répéter **épreuves de Bernoulli identiques et indépendantes**. On considère comme « succès » l’évènement « la personne a la grippe » (probabilité et comme « échec » l’évènement « la personne n’a pas la grippe » (probabilité ).
* Donc le nombre observé dans l’échantillon de taille peut **varier entre et**  et  **suit la loi binomiale** .

Le calcul de se fait à la calculatrice avec la fonction **binomFRép** :

 à près.



1. Toujours dans l’échantillon de personnes, on pose la fréquence observée des malades de la grippe. Déterminer la probabilité de trouver une fréquence observée de malades dans l’échantillon telle que .

*Réponse :*

On a .

équivaut successivement à

D’où :

ou encore, avec des valeurs approchées : et donc

 à près.



***Propriété :***

Si est la fréquence d’observation d’un caractère binaire (chaque individu possède ou ne possède pas une certaine caractéristique) sur un échantillon de taille , alors :

où suit la loi binomiale et où et sont des entiers de l’intervalle .

***Exemple 2 : Intervalle de fluctuation à 95 % de la fréquence observée***

1. En reprenant la même population que dans l’exemple 1, déterminer les entiers et tels que :

 soit le plus petit entier tel que .

 soit le plus petit entier tel que .

*Réponse :*

On cherche avec les probabilités cumulées de la loi  :

* Soit le premier entier tel que
* Soit le premier entier tel que

A la calculatrice, on utilise les listes L1 et L2 pour calculer les **probabilités cumulées** de la loi binomiale à l’aide de la fonction binomFRép(220, 0.05 ,….)

* Appuyer sur la touche Stats, puis dans le menu EDIT choisir EffListe L1, L2
* Appuyer sur Stats, puis dans le menu EDIT choisir Edite.

**Remplissage de la liste L1 avec les entiers compris entre et  :**

* Sélectionner le titre de la colonne L1, entrée, et saisir la formule

Pour trouver la fonction suite, appuyer sur  et aller dans le menu OPS.

**Remplissage de la liste L2 avec les probabilités :**

* Sélectionner le titre de la colonne L2, entrée, et saisir

Pour trouver la fonction binomFRép, appuyer sur  et descendre dans le menu DISTRIB.

Après quelques dizaines de secondes de calcul, les listes sont prêtes :

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |
|  |  **est la 1ère valeur supérieure à 2,5%** |
|  | … |
|  | … |
|  |  |
|  |  **est la 1ère valeur supérieure à 97,5%** |
|  | … |

 est le premier entier tel que
et est le premier entier tel que .

La probabilité que l’on observe un nombre de malades entre et dans l’échantillon est :

 soit environ

1. Que peut-on en déduire pour la fréquence de malades dans l’échantillon de taille  ?

*Réponse :*

D’après la propriété précédente :

Avec des valeurs approchées, on a :

La probabilité que l’on observe une fréquence entre et de malades dans un échantillon de personnes est . L’intervalle est l’**intervalle de fluctuation de au seuil .**

# La valeur de la proportion *p* dans la population est connue. On prédit un intervalle de fluctuation *In* contenant la fréquence *Fn*

On suppose que la taille de la population est très grande, de sorte que les prélèvements des individus soient considérés comme une succession d’épreuves de Bernoulli identiques et indépendantes à 2 issues : l’individu a le caractère avec la probabilité « succès » ou il ne l’a pas avec la probabilité « échec ». On **prédit** un intervalle de fluctuation dans lequel **il est probable** que se situe la fréquence des individus qui présentent ce caractère observée sur cet échantillon de taille .

## Définition de l’intervalle de fluctuation à 95 % avec la loi binomiale

Soit une variable aléatoire qui suit une loi binomiale et la variable aléatoire qui représente la fréquence observée des « succès » dans un échantillon de taille .

Un intervalle de fluctuation de au seuil de est un intervalle :

* est de la forme où et sont des entiers compris entre et
* est tel que ce qui équivaut à

En pratique[[1]](#footnote-1), on cherche avec **binomFRép** le plus petit entier tel que et le plus petit entier tel que .

## Définition de *In* l’intervalle de fluctuation à 95 % avec la loi normale

**1. Théorème de Moivre-Laplace pour la suite définie pour tout par**

Soit la variable aléatoire qui donne le nombre d’individus possédant le caractère étudié dans un échantillon de taille . suit la loi Soit la variable centrée réduite.

Pour connaitre l’intervalle de fluctuation avec la loi normale au seuil 95 %, on écrit le théorème de Moivre-Laplace avec les bornes et  :

|  |  |
| --- | --- |
|  |  |
|  |

FracNormale(0,975) donne

Conclusion :

**2. On écrit une équivalence de qui fasse intervenir**

Equivaut successivement à :

Conclusion :

**3. On conclut sur**

Conclusion :

ou encore : en posant.

On appelle  **l’intervalle de fluctuation asymptotique à**  car, quand la taille de l’échantillon tend vers , la probabilité que se situe dedans tend vers .

En pratique, si , avec et , la probabilité que soit dedans est proche de .

***Exemple :***

Sur une population d’individus, personnes sur ont la grippe. On prend un échantillon de personnes. Déterminer l’intervalle de fluctuation au seuil de avec la loi normale.

*Réponse :*

La population présente une proportion d’individus atteints par la grippe. On a comme taille d’échantillon . L’intervalle de fluctuation asymptotique est :

On a

D’oùen arrondissant les bornes de l’intervalle de fluctuation à près.

***Remarques :***

* On avait trouvé l’intervalle de fluctuation en utilisant la loi binomiale . On n’a donc pas retrouvé la même chose exactement car n’est pas l’infini. Cependant, l’intervalle de fluctuation obtenu avec la loi normale est assez proche.
* Si est très grand alors l’approximation est très bonne. Cela est dû au fait qu’on assimile où est une variable aléatoire discrète, à l’intégrale de la densité de la loi normale centrée réduite qui est égale à .
*  En pratique, on considère que si les conditions suivantes sont vraies :

alors on peut utiliser l’intervalle de fluctuation asymptotique donné par la loi normale centrée réduite. Cet intervalle de fluctuation est d’autant plus précis que la taille de l’échantillon tend vers .

* On donne l’intervalle de fluctuation en arrondissant la borne inférieure par défaut et la borne supérieure par excès, de façon à garantir au moins de probabilité pour que fréquence observée soit dans cet intervalle.
* On peut définir l’intervalle de fluctuation asymptotique des valeurs de autres que . Par exemple  ; etc.

## Intervalle *In* de fluctuation asymptotique au seuil de (1-α)%

***Théorème :***

Si la variable aléatoire suit la loi , alors pour tout réel , on a :

Autre écriture :

***Démonstration (exigible) :***

* Soit la variable aléatoire qui donne le nombre d’observations dans un intervalle de taille . suit la loi binomiale a comme valeurs les entiers

Soit la variable centrée réduite. a comme valeurs les réels

* Alors, d’après **le théorème de Moivre-Laplace**, pour tous réels et tels que , on a :
* Pour connaitre l’intervalle de fluctuation asymptotique au seuil de , on écrit le théorème de Moivre-Laplace avec et  :

équivaut successivement à :

* Conclusion :

Dans un échantillon de taille prélevé dans une population avec une proportion d’individus présentant un certain caractère, on observera que la fréquence de ce caractère est dans l’intervalle :

avec une probabilité de .

**C’est l’intervalle de fluctuation asymptotique au seuil de**  d’une fréquence

***Exemple :***

Sur une population d’individus, personnes sur ont la grippe.

On prend un échantillon de personnes. Déterminer l’intervalle de fluctuation asymptotique au seuil de .

*Réponse :*

La population présente une proportion d’individus atteints par la grippe. On a comme taille d’échantillon . L’intervalle de fluctuation avec la loi normale est :

On a donné par FracNormale(0.995)

D’oùen arrondissant les bornes de l’intervalle de fluctuation à près.

Ainsi, dans un échantillon de 220 personnes, on peut prédire avec une probabilité de que la fréquence de personnes atteintes par la grippe sera entre et .

RÉSUMÉ : Échantillonnage dans une population où *p* est connue

Echantillon de taille

Population

***Position du problème :***

* La proportion d’individus présentant une certaine caractéristique dans la population est connue.
* On prélève un échantillon de individus. On cherche à prédire par un intervalle de fluctuation, les valeurs probables de la fréquence observée de cette caractéristique dans un échantillon de taille connue.

|  |  |
| --- | --- |
|  | où est le nombre d’individus observés présentant la caractéristique dans l’échantillon de taille . |

On note la proportion d’individus dans la population qui ne présentent pas la caractéristique. On a .

**Gros échantillon et probabilités et moyennes**

**Petit échantillon ou probabilités et extrêmes**

non

oui

 Ou bien on peut utiliser l’intervalle de fluctuation donné par la loi binomiale comme pour les petits échantillons

 On doit utiliser la loi binomiale pour calculer l’intervalle de fluctuation de la fréquence observée

 On peut approcher l’intervalle de fluctuation de par l’intervalle de fluctuation asymptotique

La fréquence dans l'échantillon suit approximativement la loi

L’intervalle de fluctuation, au seuil de la fréquence correspondant à la réalisation sur un échantillon de taille d’une variable aléatoire suivant une loi binomiale, est approximativement :

 avec une probabilité d’au moins .

 est le réel positif tel que

Pour trouver on utilise FracNormale sur la calculatrice.

**Exemple :** Avec , et les conditions pour utiliser l’intervalle de fluctuation asymptotique sont vérifiées.

.

L’intervalle de fluctuation, au seuil 95% de la fréquence correspondant à la réalisation sur un échantillon de taille d’une variable aléatoire suivant une loi binomiale, est :

 avec une probabilité d’au moins .

 sont les plus petits entiers tels que

Pour trouver et on utilise binomFRép sur la calculatrice.

**Exemple :** Dans une population, la proportion d’un caractère est . Donner L’intervalle de fluctuation, au coefficient 95% de la fréquence du caractère sur un échantillon de taille .

 donne et .

# La valeur de la proportion *p* dans la population est une hypothèse qu’on veut tester

Dans ce paragraphe, on ne connait pas la proportion d’individus dans la population à présenter un caractère donné.

* Mais on fait **une hypothèse (conjecture) sur la valeur de** .

Après avoir conjecturé la valeur de , on calcule l’intervalle de fluctuation , ou bien par la loi binomiale, ou bien l’intervalle de fluctuation asymptotique, d’après la taille connue de l’échantillon et d’après la valeur du risque qu’on a choisie (par exemple c'est-à-dire ).

* On prélève un échantillon de taille dans la population et on observe la fréquence d’individus dans l’échantillon qui présentent le caractère.
* **Règle de décision :**

Le test de l’hypothèse sur la valeur de repose sur la règle de décision suivante :

* La fréquence observée appartient à l’intervalle de fluctuation. Donc cela veut dire que l’observation est compatible avec l’hypothèse faite sur la valeur de . On accepte donc l’hypothèse ce qui veut dire qu’elle reste une hypothèse plausible.
* La fréquence observée n’appartient pas à l’intervalle de fluctuation. Donc cela veut dire que l’observation est incompatible avec l’hypothèse faite sur la valeur de . Donc on rejettera l’hypothèse.

***Exemple :***

Un laboratoire annonce qu’un médicament sauve des patients atteints d’une maladie. Pour contrôler cette affirmation, on le teste sur patients atteints de cette maladie.

Soit le nombre de malades sauvés par ce médicament dans cet échantillon aléatoire de malades et assimilé à un tirage avec remise de taille .

On fait l’hypothèse : « La proportion de malades sauvés dans la population de malades est  »

1. Quelle est la loi suivie par  ?
2. Déterminer les plus petits entiers et tel que et .
3. Énoncer la règle de décision permettant de rejeter ou non l’hypothèse , selon la valeur de la fréquence des malades sauvés dans l’échantillon.
4. Sur les malades auxquels on a administré ce médicament, on en a sauvé . Au seuil de risque , que peut-on dire de l’annonce faite par le laboratoire ?

*Réponse :*

1. Il s’agit d’une succession de 100 épreuves de Bernoulli identiques et indépendantes (car on assimile le prélèvement de l’échantillon à un tirage avec remise). On appelle « succès » l’évènement « Le malade a été sauvé » avec la probabilité .

Donc suit la loi binomiale .

1. A la calculatrice, on utilise les listes L1 et L2 pour calculer les **probabilités cumulées** de la loi binomiale à l’aide de la fonction binomFRép(100, 0.4 ,….)
* Appuyer sur la touche Stats, puis dans le menu EDIT choisir EffListe[[2]](#footnote-2) L1, L2
* Appuyer sur Stats, puis dans le menu EDIT choisir Edite.
* Sélectionner le titre de la colonne L1, entrée, et saisir la formule

Pour trouver la fonction suite, appuyer sur  et aller dans le menu OPS.

* Sélectionner le titre de la colonne L2, entrée, et saisir

Pour trouver la fonction binomFRép, appuyer sur  et descendre dans le menu DISTRIB.

Après quelques secondes de calcul, les listes sont prêtes. On peut alors déterminer :

* Le plus petit entier tel que . On trouve .
* Le plus petit entier tel que . On trouve .
1. Règle de décision :

L’intervalle de fluctuation à de la fréquence est

* Si la fréquence observée dans l’échantillon de malades sauvés appartient à l’intervalle alors on accepte l’hypothèse :
* Si la fréquence observée dans l’échantillon de malades sauvés n’appartient pas à l’intervalle alors on rejette l’hypothèse :
1. Sur les malades auxquels on a administré ce médicament, on en a sauvé . Donc la fréquence observée est .

Donc

Donc on rejette l’hypothèse .

Mais le rejet se fait **au risque de se tromper** car il y a un une probabilité de que soit à l’extérieur de l’intervalle **du seul fait de la fluctuation d’échantillonnage**.

***Remarque :***

Dans cet exemple , on a :

 ;  ;

Donc les trois conditions : sont vérifiées.

Donc, à la place d’utiliser l’intervalle de fluctuation donné par et de la loi , on aurait pu utiliser l’intervalle de fluctuation asymptotique .

# La valeur de la proportion *p* dans la population est une inconnue qu’on estime par un intervalle de confiance *Ic*

Population

Echantillon de taille

Le problème de l’estimation est le problème « inverse » de celui de la recherche d’un intervalle de fluctuation de la fréquence observée étudié au paragraphe 2.

A partir de la fréquence observée sur un échantillon de taille , dans quel intervalle de confiance peut se situer la proportion des individus d’une population (de taille très grande) présentant un certain caractère ? L’estimation se fait toujours à un niveau de confiance donné.

Dans ce paragraphe 4, dans un souci de simplification :

* On se place toujours au niveau de confiance ce qui signifie que la proportion inconnue a une probabilité de de se trouver dans l’intervalle de confiance qu’on va calculer.
* Plutôt que d’utiliser l’intervalle de fluctuation de au seuil , démontré au §2.3 : , on utilisera l’intervalle de fluctuation « simplifié » :

***Propriété 1 : est dans un intervalle centré sur***

Soit une variable aléatoire de loi binomiale et la fréquence observée d’individus présentant un certain caractère sur un échantillon de taille .

Pour tout réel , il existe tel que si alors :

***Démonstration :***

* Etablissons un 1er résultat : **Pour tout , .**

Pour cela résolvons sur l’inéquation :

Un carré étant toujours positif,  **est vrai pour tout .**

* Etablissons un 2e résultat : **.**

 implique successivement :

|  |  |
| --- | --- |
|  | Rappel : équivaut à . |
|  |  est une valeur à connaitre. |

Or d’après le 1er résultat préalable, , d’où :

Conclusion :

|  |  |  |  |
| --- | --- | --- | --- |
| Si |  | alors |  |

* Reprenons le résultat du théorème du §2.3 :

On a montré que :

Cela implique donc, en utilisant le deuxième résultat préalable :

En utilisant la définition d’une limite finie d’une suite, on obtient **la propriété 1** :

Pour tout réel , il existe tel que si alors :

***Propriété 2 : est dans un intervalle centré sur***

On part de la ***propriété 1***

Pour tout réel , il existe tel que si alors :

On peut donc conclure :

***Propriété 2*** : Pour tout réel , il existe tel que si alors :

***Définition :***

Soit une population où **la proportion** des individus présentant un certain caractère **est inconnue**.

Soit  **la fréquence observée** de ce caractère dans un échantillon de taille .

Alors, dans au moins des cas, la proportion appartient à l’intervalle, dit **intervalle de confiance au niveau de confiance**

Cet intervalle de confiance peut être utilisé lorsque les conditions suivantes sont remplies :

***Exemple***

Dans une urne contenant des boules blanches et bleues en proportions inconnues, on effectue des tirages au hasard avec remise.

1. Après avoir effectué tirages, on compte boules blanches (et donc boules bleues). Donner l’intervalle de confiance à de la proportion de boules blanches dans l’urne.
2. Combien faudrait-il, au minimum, effectuer de tirages pour obtenir un intervalle de confiance au niveau de confiance d’amplitude inférieure ou égale à  ?

*Réponse :*

1. On a un échantillon de taille .

La fréquence du caractère « la boule est blanche » est .

|  |  |  |
| --- | --- | --- |
| Les conditions pour l’utilisation de l’intervalle de confiance |  | Sont vérifiées. |

Donc

 et

Donc, l’intervalle de confiance au niveau de confiance est .

Autrement dit, il y a au moins chances sur pour que la proportion de boules blanches dans l’urne soit comprise entre et .

1. On cherche tel que

En tirant boules, l’amplitude de l’intervalle de confiance au niveau de confiance est .

***Remarque :***

Dans certains domaines, on utilise un intervalle de confiance plus précis :
 .

RÉSUMÉ : Estimation par un intervalle de confiance de la proportion *p* qui reste inconnue

Population

Echantillon de taille

***Position du problème :*** Un échantillon est connu

Par sa taille et par la fréquence d'apparition d'un caractère dans l’échantillon.

On cherche à estimer, par un intervalle de confiance, les valeurs probables de la proportion des individus de la population qui présentent ce caractère.

**Gros échantillon et probabilités et moyennes**

**Petit échantillon ou fréquences et extrêmes**

non

oui

 Hors programme de terminale

 On peut utiliser une loi normale pour

La proportion de la population suit la loi :

L'intervalle de confiance au niveau **95%** de est :

 au risque qu'elle n'y soit pas.

Ou de façon moins précise mais simplifiée :

Dans au moins **95%** des cas, la proportion dans la population appartient à l’intervalle

 avec un risque inférieur à qu'elle n'y soit pas.

***Exemple :***

On a observé sur un échantillon de personnes que étaient cardiaques. Calculer l’intervalle de confiance au niveau de confiance dans lequel peut se situer la proportion de personnes cardiaques parmi toute la population.

*Réponse :*

Ici, on a : Donc les conditions sont vérifiées.

* Méthode précise :

La proportion de cardiaques dans la population avec un niveau de confiance de .

* Méthode simplifiée :

La proportion de cardiaques dans la population avec un niveau de confiance de (au moins)

1. Ces définitions de $a$ et $b$ viennent du fait qu’il faut au moins $95 \%$ des probabilités cumulées lorsque $a\leq X\_{n}\leq b$ et donc on a $P\left(X\_{n}\leq a-1\right)\leq 0,025$ et $P\left(X\_{n}\geq b+1\right)\leq 0,025$. [↑](#footnote-ref-1)
2. Pour toute nouvelle utilisation des fonctions statistiques, penser à effacer les listes précédentes. Ainsi l’ancien contenu ne sera pas pris en compte dans les nouveaux calculs. [↑](#footnote-ref-2)