

1. Intervalle de fluctuation.

Définition.

Soit X une variable aléatoire suivant une loi binomiale, $\alpha \in]0; 1[$ et a et b deux réels. Un intervalle $[a; b]$ tel que $\mathbb{P}(a \leq X \leq b) \geq 1 - \alpha$ est appelé **intervalle de fluctuation** au seuil de $1 - \alpha$ (ou au risque α) associé à X .

Propriété.

L'intervalle $[a; b]$ tel que a et b soient les plus petits entiers vérifiant respectivement $\mathbb{P}(X \leq a) > \frac{\alpha}{2}$ et $\mathbb{P}(X \leq b) \geq 1 - \frac{\alpha}{2}$ est un **intervalle de fluctuation** centré au seuil de $1 - \alpha$ associé à X .

Remarque. En Statistiques, on utilise souvent comme seuil 95 % ou 99%. En pratique, si l'effectif de la population est très grand par rapport à l'effectif n de l'échantillon on considère que le tirage des éléments de l'échantillon s'effectue avec remise.

Un exemple au seuil de 95 %.

L'intervalle de fluctuation à 95 % correspondant à une variable aléatoire X suivant une loi binomiale de paramètres n et p , est l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ où :

- a est le plus petit entier tel que $\mathbb{P}(X \leq a) > \dots\dots\dots$;
- b est le plus petit entier tel que $\mathbb{P}(X \leq b) \geq \dots\dots\dots$

Exemple :

On considère une population pour laquelle la proportion d'un caractère C est $p = 0,28$.

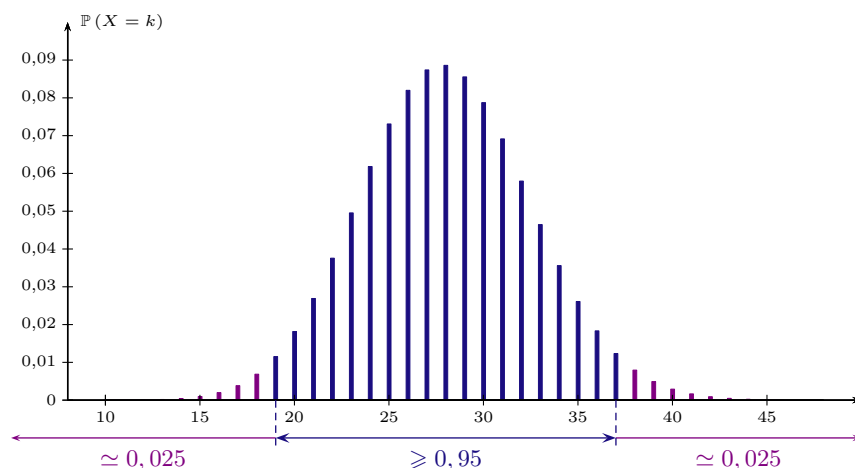
On prélève au hasard et avec remise un échantillon de taille $n = 100$. La variable aléatoire X associée au nombre d'individus ayant le caractère C au sein de l'échantillon, suit la loi binomiale $\mathcal{B}(100; 0,28)$ de paramètres $n = 100$ et $p = 0,28$.

On a ci-dessous, un extrait du tableau des probabilités cumulées $\mathbb{P}(X \leq k)$ de la loi binomiale $\mathcal{B}(100; 0,28)$.

k	$\mathbb{P}(X \leq k)$	k	$\mathbb{P}(X \leq k)$	k	$\mathbb{P}(X \leq k)$	k	$\mathbb{P}(X \leq k)$	k	$\mathbb{P}(X \leq k)$
14	0,000 7	20	0,044	26	0,374 8	32	0,842	38	0,988 7
15	0,001 7	21	0,070 9	27	0,462 2	33	0,888 4	39	0,993 6
16	0,003 7	22	0,108 5	28	0,550 7	34	0,924	40	0,996 5
17	0,007 5	23	0,158	29	0,636 2	35	0,950 1	41	0,998 2
18	0,014 4	24	0,219 8	30	0,714 9	36	0,968 4	42	0,999 1
19	0,025 9	25	0,292 9	31	0,784	37	0,980 7	43	0,999 5

- Le plus petit entier a tel que $\mathbb{P}(X \leq a) > \dots\dots\dots$ est $a = \dots\dots\dots$
- Le plus petit entier b tel que $\mathbb{P}(X \leq b) \geq \dots\dots\dots$ est $b = \dots\dots\dots$

Loi binomiale de paramètres $n = 100$ et $p = 0,28$



L'intervalle de fluctuation à 95 % de la fréquence du nombre d'individus ayant le caractère C au sein d'un échantillon de taille 100 correspondant à la variable aléatoire est donc l'intervalle $\left[\frac{19}{100}; \frac{37}{100}\right] = [0, 19; 0, 37]$. Cet intervalle est centré en 0,28 qui est la proportion du caractère C dans la population.

2. Tests d'hypothèses.

On formule l'hypothèse que la proportion d'un caractère dans la population est p .

Pour valider cette hypothèse, on prélève au hasard dans la population un échantillon de taille n et on note f la fréquence observée du caractère étudié.

Si l'effectif de la population est suffisamment grand par rapport à l'effectif n de l'échantillon on considère que la variable aléatoire X associée au nombre d'éléments ayant le caractère étudié au sein de l'échantillon, suit la loi binomiale $\mathcal{B}(n; p)$.

On détermine l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ de fluctuation à 95 % d'une fréquence correspondant à la variable aléatoire X qui permet de fixer le seuil de décision.

- Si la fréquence observée f appartient à l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$, on accepte l'hypothèse selon laquelle la proportion est p dans la population au seuil de 95 %.
- Si la fréquence observée f n'appartient pas à l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$, on rejette l'hypothèse selon laquelle la proportion est p dans la population avec un risque d'erreur de 5 %.

Exemple. Selon une publication de l'INSEE, 28 % des ménages comprennent une famille avec au moins un enfant mineur.

On interroge un échantillon de 100 ménages choisis au hasard, et on constate que dans cet échantillon 35 % des ménages comprennent une famille avec au moins un enfant mineur.

Cet échantillon est-il représentatif de la population ?

L'intervalle de fluctuation à 95 % associé à la loi binomiale $\mathcal{B}(100; 0, 28)$ calculé précédemment est $I = [0, 19; 0, 37]$.

La fréquence observée des ménages comprenant une famille avec au moins un enfant mineur est $f = 0, 35$.

Donc f appartient à l'intervalle $[0, 19; 0, 37]$.

On considère que l'échantillon est représentatif de la population.

3. Seuil.

Dans une équipe de football, un défenseur discute d'une clause dans son contrat : il aura une prime s'il reçoit n cartons jaunes ou moins sur les 38 matches de la saison. Il a remarqué que la probabilité qu'il prenne un carton jaune lors d'un match est de 0,15.

En admettant que les cartons jaunes reçus soient indépendants, quelle doit-être la plus petite valeur de n pour qu'il soit sûr au seuil de 99 % de toucher cette prime ?

Pour tout entier i entre 1 et 38, X_i est la variable aléatoire prenant 1 si le joueur reçoit un carton jaune à la i - ème rencontre et 0 sinon.

1. Quelle loi suit la variable aléatoire X_i ?

.....

2. Que peut-on dire de la variable aléatoire $X = X_1 + X_2 + \dots + X_{38}$?

.....

.....

3. Répondre au problème posé à l'aide de la calculatrice.

.....

4. Étude de sondages.

Lorsque l'on réalise un sondage, son résultat est donné avec une marge d'erreur.

Cette marge d'erreur devient plus petite lorsque le nombre de personnes interrogées augmente mais **pas** dans la même mesure.

Exemple. Premier tour des élections présidentielles 2017 en France.

Lors des élections présidentielles de 2017, le candidat F.Fillon a obtenu 20 % des voix au premier tour, ce qui l'a placé en 3^e position.

1. On considère un sondage sur 1 000 personnes prises au hasard dans la population française avant cette élection.

- (a) On appelle X_i pour i entier entre 1 et 1 000 la variable aléatoire prenant la valeur 1 si i – ème personne interrogée déclare être favorable au candidat F.Fillon et 0 sinon.

Sous quelle(s) hypothèses peut-on dire que les variables aléatoires $X_1; \dots; X_{1\,000}$ forment un échantillon ?

.....

Quelle est la loi suivie par la variable aléatoire X_i ?

.....

Quelle est la loi suivie par la variable aléatoire $S = X_1 + X_2 + \dots + X_{1\,000}$?

.....

.....

- (b) Déterminer un intervalle de fluctuation centré au seuil de 95 %.

.....

.....

- (c) En déduire la marge d'erreur associée à ce sondage.

.....

.....

- (d) Quelques mois avant l'élection, la plupart des sondages donnaient plus de 25 % d'intentions de votes au candidat F.Fillon. Que peut penser de ces sondages ?

.....

.....

.....

- (e) Reprendre **1.b** et **1.c** dans le cas de 10 000 personnes.

.....

.....

.....

.....