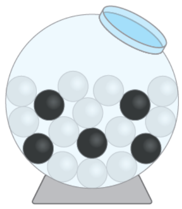
ECHANTILLONNAGE ET ESTIMATION

Le mathématicien d'origine russe *Jerzy Neyman* (1894 ; 1981), ci-contre, pose les fondements d'une approche nouvelle des statistiques. Avec l'anglais *Egon Pearson*, il développe la théorie de l'estimation et de la prise de décision sur un échantillon.
Ses travaux trouveront rapidement des applications dans de nombreux domaines concrets, tels la médecine, l'astronomie ou la météorologie.

Dans ce chapitre, on va étudier deux domaines des statistiques qu'il faut savoir distinguer :

Echantillonnage – Prise de décision	Estimation
- Une urne contient un très grand nombre de boules blanches et de boules noires dont **on connaît la proportion** *p* de boules blanches. On tire avec remise *n* boules (échantillon) et on observe la fréquence d'apparition des boules blanches. Cette fréquence observée appartient à un intervalle, appelé **intervalle de fluctuation** de centre *p*. - Dans le cas où on ne connaît pas la proportion *p* mais on est capable de faire une hypothèse sur sa valeur, on parle de **prise de décision**. On veut par exemple savoir si un dé est bien équilibré. On peut faire l'hypothèse que l'apparition de chaque face est égale à 1/6 et on va tester cette hypothèse à l'aide d'une expérience. Le résultat de l'expérience va nous permettre d'accepter ou rejeter l'hypothèse de départ.	Une urne contient un très grand nombre de boules blanches et de boules noires dont **on ignore la proportion** *p* de boules blanches. On tire avec remise *n* boules dans le but d'estimer la proportion *p* de boules blanches. On obtient ainsi une fréquence d'apparition qui va nous permettre d'estimer la proportion p à l'aide d'un **intervalle de confiance**. 

Conditions sur les paramètres : Dans tout le chapitre, sauf mention contraire, la taille de l'échantillon n et la proportion p vérifient :

$$n \geq 30,\ n \times p \geq 5 \text{ et } n \times (1-p) \geq 5.$$

I. Echantillonnage

1) Intervalle de fluctuation asymptotique

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

Méthode : Déterminer un intervalle de fluctuation

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.
1) On tire successivement avec remise n = 50 boules.
Déterminer l'intervalle de fluctuation à au moins 95% de la fréquence d'apparition d'une boule blanche.
2) Même question pour n = 500 boules.

Définition :
p est la proportion théorique.
L'intervalle de fluctuation à au moins 95% est :
$$I=\left[p-\frac{1}{\sqrt{n}}\,;\,p+\frac{1}{\sqrt{n}}\right]$$

1) On a : p = 0,3 et n = 50.
$$I_{50}=\left[0{,}3-\frac{1}{\sqrt{50}}\,;\,0{,}3+\frac{1}{\sqrt{50}}\right]$$
Soit $I_{50}=\left[0{,}159\,;\,0{,}441\right]$.
Cela signifie que pour 50 tirages, dans 95% des cas, la fréquence d'apparition de boules blanches se situe entre 15,9% et 44,1%.

2) Pour 500 tirages, on obtient :
$$I_{500}=\left[0{,}3-\frac{1}{\sqrt{500}};0{,}3+\frac{1}{\sqrt{500}}\right]=\left[0{,}255\,;\,0{,}345\right]$$
On constate que l'intervalle, pour un même seuil, se resserre fortement lorsqu'on augmente le nombre de tirages.

2) 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*.

Méthode : Prendre une décision à l'aide d'un intervalle de fluctuation

Vidéo https://youtu.be/QZ0YFthGI0Y

Un fabricant d'alarme commande auprès de son fournisseur deux types de composants électroniques : C1 et C2. 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 C1.
Peut-on affirmer que la commande est respectée par le fournisseur ?

1. Hypothèse : La commande est respectée par le fournisseur : $p=\frac{1}{2}$ = 0,5 (autant de composants de chaque sorte)

2. Intervalle de fluctuation : On a : p = 0,5 et $n=50$

$I=\left[0,5-\frac{1}{\sqrt{50}};0,5+\frac{1}{\sqrt{50}}\right]\approx[0,359;0,641]$.

3. Fréquence observée : $f=\frac{19}{50}=0,38$.

Règle de décision :
f la fréquence observée d'un échantillon de taille n.
I l'intervalle de fluctuation asymptotique à au moins 95%.
On fait l'hypothèse : "La proportion est *p*."
- Si $f\in I$, alors on accepte l'hypothèse.
- Si $f\notin I$, alors on rejette l'hypothèse.

4. Prise de décision : $f=0,38\in I$.
La fréquence observée appartient à l'intervalle de fluctuation. D'après la règle de décision, l'hypothèse faite est acceptable : la commande est respectée par le fournisseur.

II. 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, on va estimer la proportion *p* d'un caractère dans la population tout entière.

Méthode : Estimer une proportion inconnue par un intervalle de confiance

Vidéo https://youtu.be/cU5cJICVAM8

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 ?

La proportion *p* des électeurs de Martine Phinon est inconnue.

- La taille de l'échantillon est $n=1052$.

- La fréquence observée est $f=\frac{614}{1052}\approx0,5837$.

- L'intervalle de confiance de la proportion *p* au niveau de confiance 95% est :

$$J=\left[0{,}5837-\frac{1}{\sqrt{1052}}\ ;\ 0{,}5837+\frac{1}{\sqrt{1052}}\right].$$
$$\approx\left[0{,}553\ ;\ 0{,}615\right]$$

Définition :
f est la fréquence observée.
L'intervalle de confiance au niveau de confiance 95% est :
$$J=\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right].$$

- Pour être élue, la proportion *p* doit être strictement supérieure à 0,5. Selon ce sondage, il est envisageable que Martine Phinon soit élue.

Ne pas confondre :
Intervalle de FLUCTUATION V.S. Intervalle de CONFIANCE :

Vidéo https://youtu.be/97vzxWsyie8