Estimation:

intervalle de fluctuation et de confiance

Mars 2012

IREM: groupe Proba-Stat

connu :
probabilité p, taille de l'échantillon n

but : estimer une fréquence f à partir d'une probabilité

construire un intervalle *à l'aide de la probabilité p*

- centré
- contenant les fréquences observées à 95 %

($0,95$ *est appellé le seuil ; parfois* $s = 1 - \alpha$ *avec* α *appellé le risque*)

exemple : *Un joueur qui doit choisir au hasard une carte dans un jeu de 32 cartes obtient certains avantages s'il découvre un roi.*
On constate qu'il a retourné 11 fois un roi sur 50 essais.
Peut-on présumer, au risque de 5%, que ce joueur est un tricheur ?

notion de test

Intervalle de fluctuation : dans les programmes

1. ***en Seconde :*** si $n \geq 25$ et $0,2 \leq p \leq 0,8$, la fréquence observée

$$f \in \left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$$

avec une probabilité d'au moins 0,95

2. ***en Terminale*** si $n \geq 30$, $np > 5$, $n(1-p) > 5$,

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

avec une probabilité de 0,95

Intervalle de fluctuation : lien 2nd - Terminale

On approxime

- $p \in [0,1] \quad \Rightarrow \quad p(1-p) \leq \dfrac{1}{4} \quad \Rightarrow \quad \sqrt{p(1-p)} \leq \dfrac{1}{2}$
  avec l'étude de la fonction $f(x) = x(1-x)$
- $1,96 \lesssim 2$

Pour simplifier, en seconde, $1,96\sqrt{p(1-p)} \leq 1$

Est ce vraiment valable ?

p	0,1	**0,2**	**0,3**	**0,4**	**0,5**	**0,6**	**0,7**	**0,8**	0,9
d	0,59	0,78	0,9	0,96	0,98	0,96	0,9	0,78	0,59

avec $d = 1,96\sqrt{p(1-p)}$

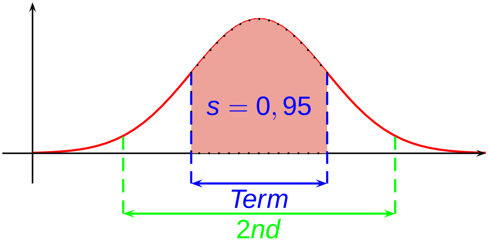
Intervalle de fluctuation : lien 2nd - Terminale

On approxime

- $p \in [0,1] \quad \Rightarrow \quad p(1-p) \leq \dfrac{1}{4} \quad \Rightarrow \quad \sqrt{p(1-p)} \leq \dfrac{1}{2}$
  avec l'étude de la fonction $f(x) = x(1-x)$
- $1,96 \lesssim 2$

Pour simplifier, en seconde, $\quad 1,96\sqrt{p(1-p)} \leq 1$

En seconde, l'intervalle est <u>plus large</u> que celui de terminale
qui est l'intervalle théorique



c'est pour cela qu'apparaît "au moins" en seconde

Intervalle de confiance

connu :
fréquence f, taille de l'échantillon n

but : estimer une probablité à partir des observations

construire un intervalle *à l'aide la fréquence f* contenant la probabilité ou la proportion inconnue p à 95 %
(*avec un niveau de confiance de 95%*)

en Terminale

$p \in \left[f - \frac{1}{\sqrt{n}} ; f + \frac{1}{\sqrt{n}}\right]$ avec une probabilité d'au moins 0,95

on fait les même approximations qu'en classe de seconde

exemple : *Un sondage dans une commune révèle que sur les 500 personnes interrogées, 42% sont mécontentes de l'organisation des transport.*
On veut déterminer, au seuil d'au moins 95%, un intervalle de confiance du pourcentage p de personnes mécontentes dans la commune.

Approximation de la loi Binomiale par une loi normale

- f fréquence observée $\longrightarrow F_n$ variable aléatoire associée
- $F_n = \dfrac{X_n}{n}$ avec X_n suit une loi $\mathcal{B}(n, p)$

Avec le théorème de Moivre Laplace : <u>quand n devient grand</u>

la loi binomiale $\mathcal{B}(n, p)$ s'approche d'une loi normale $\mathcal{N}(m; \sigma)$ où
$m = E(X_n) = np$ et $\sigma = \sqrt{Var(X_n)} = \sqrt{np(1-p)}$

Dans la pratique, on considère que l'approximation est bonne lorsque

$n \geq 30$ $\qquad np \geq 5$ $\qquad np(1-p) > 5$

Approximation de la loi Binomiale par une loi normale

- f fréquence observée $\longrightarrow F_n$ variable aléatoire associée
- $F_n = \dfrac{X_n}{n}$ avec X_n suit une loi $\mathcal{B}(n,p)$

Avec le théorème de Moivre Laplace : quand n devient grand

la loi binomiale $\mathcal{B}(n,p)$ s'approche d'une loi normale $\mathcal{N}(m;\sigma)$ où

$$m = E(X_n) = np \text{ et } \sigma = \sqrt{Var(X_n)} = \sqrt{np(1-p)}$$

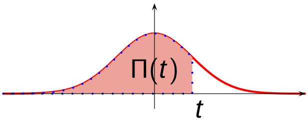
Après normalisation,

- $\dfrac{X_n - np}{\sqrt{np(1-p)}}$ suit une loi normale $\mathcal{N}(0;1)$
- $\dfrac{X_n - np}{\sqrt{np(1-p)}} = n\dfrac{\dfrac{X_n}{n} - p}{\sqrt{np(1-p)}} = \dfrac{\dfrac{X_n}{n} - p}{\sqrt{p(1-p)}/\sqrt{n}}$

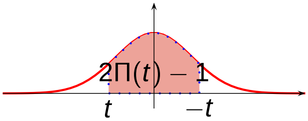
$$\frac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}} = Z_n \text{ suit une loi normale } \mathcal{N}(0;1)$$

Pourquoi 1,96 ?

$$P(-t \leq Z_n \leq t) = 1 - \alpha = s \text{ (ici } s = 0,95\text{)}$$



En notant $\Pi(t) = P(Z_n \leq t)$,
on a $P(-t \leq Z_n \leq t) = 1 - 2(1 - \Pi(t)) = 2\Pi(t) - 1$



d'où $\Pi(t) = \dfrac{s+1}{2} = \dfrac{0,95+1}{2} = 0,975$

<u>Par lecture sur la table ou avec la calculatrice</u>, $t = 1,96$

Détermination de l'intervalle

on remplace $Z_n = \dfrac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}}$ dans

$$\begin{aligned} P(-1,96 \leq Z_n \leq 1,96) &= 0,95 \\ P\left(-1,96 \leq \frac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq 1,96\right) &= 0,95 \end{aligned}$$

- si on connaît la probabilité p,
  on isole la (v.a associée) fréquence F_n :

  $$P(p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}) = 0,95]$$

  $\Longrightarrow$ intervalle de fluctuation

- si on connaît une fréquence de l'échantillon f,

  $\Longrightarrow$ intervalle de confiance

Détermination de l'intervalle

on remplace $Z_n = \dfrac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}}$ dans

$$\begin{aligned} P(-1,96 \leq Z_n \leq 1,96) &= 0,95 \\ P\left(-1,96 \leq \frac{F_n - p}{\sqrt{p(1-p)}/\sqrt{n}} \leq 1,96\right) &= 0,95 \end{aligned}$$

- si on connaît la probabilité p,

  $\Longrightarrow$ intervalle de fluctuation

- si on connaît une fréquence de l'échantillon f,
  - on remplace $\sqrt{p(1-p)}$ par $\sqrt{f(1-f)}$ *ce qui est valable théoriquement*
  - on isole la probabilité p :

$$P\left(-f - 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}} \leq -p \leq -f + 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right) = 0,95$$

$$P\left(f - 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}} \leq p \leq f + 1,96\frac{\sqrt{f(1-f)}}{\sqrt{n}}\right) = 0,95$$

$\Longrightarrow$ intervalle de confiance