Partie II

INTERVALLES DE FLUCTUATION ET DE CONFIANCE EN TERMINALE

Aspects mathématiques et statistiques

Journée académique "Terminale"
Besançon, octobre 2012

Yves DUCEL & Bruno SAUSSEREAU
Institut de recherche sur l'enseignement des mathématiques
Université de Franche-Comté

Travail IREM - Université de Franche-Comté

- **Yves DUCEL** et **Bruno SAUSSEREAU**,
  enseignants-chercheurs à l'Université de Franche-Comté.
- Cet exposé est le fruit d'une réflexion menée dans le cadre du groupe de travail *Probabilités et statistique* de l'IREM de Franche-Comté.
- Le diaporama sera disponible.

Problématique et conventions

Problématique de la partie II

1. Apporter un peu de recul et un éclairage nouveau sur les IF et IC mentionnés dans le programme de Terminale pour faire ressortir la continuité des programmes sur les trois années du lycée :
   * en explicitant le cadre mathématique sous-jacent ;
   * en détaillant la relation d'inversion qu'entretiennent la notion d'IF et celle d'IC.
2. De façon générale, dégager des idées directrices dans la construction mathématique des intervalles de fluctuation et de confiance, ainsi que dans la discussion de leur pertinence en statistique.

Notations

Dans toute la suite de l'exposé :

- les échantillons sont obtenus par répétition d'une même expérience de Bernoulli (Succès/Échec) ;
- n est la taille de l'échantillon et p $(0 < p < 1)$ le paramètre de l'expérience de Bernoulli ;
- la variable aléatoire F_n désigne **la fréquence empirique** ;
- la variable aléatoire $X_n = nF_n$, de loi binomiale $\mathcal{B}(n, p)$, représente **l'effectif empirique** ;
- Abréviations : IF est mis pour "intervalle de fluctuation" et IC pour "intervalle de confiance" ;
- sauf précision contraire, les IF et IC sont considérés au seuil $1 - \alpha$, où $\alpha \in]0, 1[$.

Convention : conditions des grandes binomiales

- On dira que **les conditions des grandes binomiales** sont vérifiées si n est "grand" et si p est "voisin" de $\frac{1}{2}$.
- L'interprétation des termes "grand" et de "voisin" varie suivant les auteurs. Par exemple, on peut trouver l'une des conditions suivantes :
  * $n \geq 25$ et $0,2 \leq p \leq 0,8$ (Seconde) ;
  * $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ (Terminale) ;
  * $np \geq 5$ et $n(1-p) \geq 5$ (ou 10 ou 20) ;
  * $np(1-p) \geq 5$ (ou 10) ;
  * $n \geq 30$, $nf \geq 20$ et $n(1-f) \geq 20$, où f est la fréquence observée dans l'échantillon.

Convention : conditions des grandes binomiales

- On dira que **les conditions des grandes binomiales** sont vérifiées si n est "grand" et si p est "voisin" de $\frac{1}{2}$.
- L'interprétation des termes "grand" et de "voisin" varie suivant les auteurs. Par exemple, on peut trouver l'une des conditions suivantes :
  * $n \geq 25$ et $0,2 \leq p \leq 0,8$ (Seconde) ;
  * $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$ (Terminale) ;
  * $np \geq 5$ et $n(1-p) \geq 5$ (ou 10 ou 20) ;
  * $np(1-p) \geq 5$ (ou 10) ;
  * $n \geq 30$, $nf \geq 20$ et $n(1-f) \geq 20$, où f est la fréquence observée dans l'échantillon.

IF asymptotique (standard)

Le problème

Le programme de Terminale contient l'énoncé suivant :

- **Théorème** : *Pour tout* $\alpha \in]0,1[$, *notons* u_α *l'unique réel positif vérifiant* $\mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$, *où* Z *est une variable aléatoire normale centrée-réduite. Alors, pour tout* $p \in]0,1[$, *on a*

$$\lim_{n \to +\infty} \mathbb{P}\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1 - \alpha\,.$$

- D'où sort l'encadrement de F_n dans le théorème ?
- Quel lien y a t-il avec les IF de Seconde et de Première ?

Rappel des prérequis

Construction de l'IF de Première

En Première, on prend pour intervalle de fluctuation au seuil $1-\alpha$ de F_n **le plus petit intervalle fermé** $[f_1\,,\,f_2]$ vérifiant les deux relations :

* $\mathbb{P}(F_n \in [0\,,f_1[) = \mathbb{P}(X_n \in [0\,,nf_1[) \leq \dfrac{\alpha}{2}$, où $[0\,,f_1[$ est la partie extérieure gauche de l'IF ;
* $\mathbb{P}(F_n \in]f_2\,,1]) = \mathbb{P}(X_n \in]nf_2\,,1]) \leq \dfrac{\alpha}{2}$, où $]f_2\,,1]$ est la partie extérieure droite de l'IF.

- Représentation graphique pour le seuil 95% : Voir le graphique

♣ L'IF de Première est utilisable en statistique pour n'importe quelles valeurs de n, p et α.

IF de Première pour les grandes binomiales

Pour les grandes binomiales :

- Le diagramme en bâtons de la loi de X_n est pratiquement symétrique et centré sur l'espérance np.
- L'IF de Première est très voisin de l'IF de Seconde.
- Pour les grandes binomiales, l'IF exact de Première au seuil $1-\alpha$ de F_n est approximativement décrit comme **le plus petit intervalle fermé** $[f_1\,,f_2]$ **centré sur** p tel que

$$\mathbb{P}(F_n \in [f_1\,,f_2]) = \mathbb{P}(X_n \in [nf_1\,,nf_2]) \simeq 1-\alpha.$$

Conséquence sur l'IF de Première

Sous les conditions de grandes binomiales, l'IF de Première (et donc aussi celui de Seconde) est pratiquement de la forme

$$[p - e\,,p + e]$$

où e est une expression (fonction de n, p et α) à préciser.

u_α : fractile normal d'ordre $1 - \frac{\alpha}{2}$

- *Si Z est une variable aléatoire de loi normale centrée-réduite, pour tout $\alpha \in]0,1[$, il existe un unique réel positif u_α tel que* $\mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$.
- Deux valeurs classiques à connaître : $u_{0,05} \simeq 1,96$ et $u_{0,01} \simeq 2,58$.
- Dans la pratique, $0 \leq u_\alpha \leq 3$.

Théorème de De Moivre-Laplace

- Posons, pour tout entier $n \geq 1$, $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$.
- **Théorème de De Moivre-Laplace** : *Pour tous réels a et b avec $a < b$, on a*

$$\lim_{n \to +\infty} \mathbb{P}(a \leq Z_n \leq b) = \mathbb{P}(a \leq Z \leq b)$$

où Z est une variable aléatoire normale centrée-réduite.

- En pratique, dans les conditions des grandes binomiales, on accepte l'approximation $\mathbb{P}(a \leq Z_n \leq b) \simeq \mathbb{P}(a \leq Z \leq b)$.

Démarche heuristique

Heuristique de l'IF asymptotique - 1/3

Pour p fixé :

- des équivalences
  $$F_n \in [p-e\,,p+e] \quad \Leftrightarrow \quad np-ne \leq X_n \leq np+ne$$
  $$\Leftrightarrow \frac{-ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\ ;$$
- on déduit
  $$\mathbb{P}(F_n \in [p-e\,,p+e]) =$$
  $$= \quad \mathbb{P}\left(\frac{-ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\right)\ ;$$

Heuristique de l'IF asymptotique - 1/3

Pour p fixé :

- des équivalences
  $$F_n \in [p-e\,,p+e] \quad \Leftrightarrow \quad np-ne \leq X_n \leq np+ne$$
  $$\Leftrightarrow \frac{-\,ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\ ;$$
- on déduit
  $$\mathbb{P}(F_n \in [p-e\,,p+e]) =$$
  $$= \quad \mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\right)\ ;$$

Heuristique de l'IF asymptotique - 1/3

Pour p fixé :

- des équivalences
  $$F_n \in [p-e\,,p+e] \quad \Leftrightarrow \quad np-ne \leq X_n \leq np+ne$$
  $$\Leftrightarrow \frac{-\,ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\ ;$$
- on déduit
  $$\mathbb{P}(F_n \in [p-e\,,p+e]) =$$
  $$= \quad \mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq \frac{X_n-np}{\sqrt{np(1-p)}} \leq \frac{ne}{\sqrt{np(1-p)}}\right)\ ;$$

Heuristique de l'IF asymptotique - 2/3

- C'est-à-dire

$$\mathbb{P}(F_n \in [p-e\,,p+e]) = \mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right).$$

- Pour α donné, prenons n assez grand de sorte que

$$\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) \simeq \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$$

où Z est une variable aléatoire normale centrée-réduite.

- Dans ce cas, la condition $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$ peut s'écrire :

$$\mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right) \simeq \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha).$$

Heuristique de l'IF asymptotique - 2/3

- C'est-à-dire

$$\mathbb{P}(F_n \in [p-e\,,p+e]) = \mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right).$$

- Pour α donné, prenons n assez grand de sorte que

$$\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) \simeq \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$$

où Z est une variable aléatoire normale centrée-réduite.

- Dans ce cas, la condition $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$ peut s'écrire :

$$\mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right) \simeq \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha).$$

Heuristique de l'IF asymptotique - 2/3

- C'est-à-dire

$$\mathbb{P}(F_n \in [p-e\,,p+e]) = \mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right).$$

- Pour α donné, prenons n assez grand de sorte que

$$\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) \simeq \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$$

où Z est une variable aléatoire normale centrée-réduite.

- Dans ce cas, la condition $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$ peut s'écrire :

$$\mathbb{P}\left(\frac{-\,ne}{\sqrt{np(1-p)}} \leq Z_n \leq \frac{ne}{\sqrt{np(1-p)}}\right) \simeq \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha).$$

Heuristique de l'IF asymptotique - 3/3

- Cette condition sera satisfaite si on choisit e tel que
$$\frac{ne}{\sqrt{np(1-p)}} = u_\alpha,$$
- c'est-à-dire $e = u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}$.
- D'où le choix possible pour l'intervalle $[f_1\,,f_2]$:

$$[f_1\,,f_2] = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\,, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

Heuristique de l'IF asymptotique - 3/3

- Cette condition sera satisfaite si on choisit e tel que
$$\frac{ne}{\sqrt{np(1-p)}} = u_\alpha,$$
- c'est-à-dire $e = u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$.
- D'où le choix possible pour l'intervalle $[f_1\,,\,f_2]$:
$$[f_1\,,\,f_2] = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\,,\,p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

Heuristique de l'IF asymptotique - 3/3

- Cette condition sera satisfaite si on choisit e tel que
$$\frac{ne}{\sqrt{np(1-p)}} = u_\alpha,$$
- c'est-à-dire $e = u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$.
- D'où le choix possible pour l'intervalle $[f_1\,,\,f_2]$:

$$[f_1\,,\,f_2] = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

IF asymptotique

Théorème de l'IF asymptotique

- **Théorème** : *Pour tout* $\alpha \in]0,1[$, *notons* u_α *l'unique réel positif vérifiant* $\mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1-\alpha$, *où* Z *est une variable aléatoire normale centrée-réduite. Alors, pour tout* $p \in]0,1[$, *on a*

$$\lim_{n \to +\infty} \mathbb{P}\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 1-\alpha\,.$$

♣ L'intervalle

$$IF_{A,n} = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$$

est appelé **IF asymptotique (standard) pour F_n au seuil** $1-\alpha$.

Pratique de l'approximation

Dans la pratique, sous les conditions des grandes binomiales, on s'autorise l'approximation

$$\mathbb{P}\left(p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) \simeq 1 - \alpha\,.$$

Démonstration

- On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}} = \dfrac{F_n - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.
- D'après le théorème de De Moivre-Laplace, on a $\lim\limits_{n \to +\infty} \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$.
- La conclusion résulte de l'égalité : $\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}\left(p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$. $\square$

Démonstration

- On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}} = \dfrac{F_n - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.
- D'après le théorème de De Moivre-Laplace, on a $\lim\limits_{n \to +\infty} \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$.
- La conclusion résulte de l'égalité : $\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}\left(p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$. $\square$

Démonstration

- On pose $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}} = \dfrac{F_n - p}{\sqrt{\dfrac{p(1-p)}{n}}}$.
- D'après le théorème de De Moivre-Laplace, on a $\lim\limits_{n \to +\infty} \mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}(-u_\alpha \leq Z \leq u_\alpha) = 1 - \alpha$.
- La conclusion résulte de l'égalité : $\mathbb{P}(-u_\alpha \leq Z_n \leq u_\alpha) = \mathbb{P}\left(p - u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + u_\alpha \dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right)$. $\square$

Notion de pivot

- Cette démonstration prend appui, pour n fixé, sur une variable aléatoire (ici $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$) :
  - * qui fait intervenir p dans son expression ;
  - * dont la loi (fonction de p) est asymptotiquement une loi (ici la loi normale centrée-réduite) qui, elle, ne dépend pas de p.
- ♣ On dit que cette variable aléatoire est un **pivot asymptotique pour** p.
- L'intérêt des variables-pivots est de permettre le calcul des probabilités les impliquant, même si la valeur du paramètre est inconnue.

Compléments sur l'IF de Seconde

Justification de la formulation de IF_2

- **Proposition** : *Pour tous* $p \in]0,1[$ *et* n *fixés, l'intervalle de fluctuation asymptotique* $IF_{A,n}$ *au seuil* $0,95\%$ *pour* F_n *est inclus dans l'intervalle* $\left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]$ *donné en Seconde.*

- Ces deux intervalles sont d'autant plus voisins que n est grand.

Démonstration

- Au seuil 95%, on a $u_{0,05} \approx 1,96$. L'intervalle asymptotique s'écrit : $IF_{A,n} = \left[p - 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right]$.
- La fonction, $x \in [0,1] \mapsto f(x) = x(1-x)$, admet un maximum en $x = 0,5$ égal à $f(0,5) = 0,25$.
- Les inégalités $1,96 \leq 2$ et $\sqrt{p(1-p)} \leq 0,5$, permettent de majorer $1,96 \frac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\frac{1}{\sqrt{n}}$.
- Par suite, $IF_{A,n}$ est inclus dans $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}} \right]$, et ils sont d'autant plus voisins que n est grand. $\square$

Démonstration

- Au seuil 95%, on a $u_{0,05} \approx 1,96$. L'intervalle asymptotique s'écrit : $IF_{A,n} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
- La fonction, $x \in [0,1] \mapsto f(x) = x(1-x)$, admet un maximum en $x = 0,5$ égal à $f(0,5) = 0,25$.
- Les inégalités $1,96 \leq 2$ et $\sqrt{p(1-p)} \leq 0,5$, permettent de majorer $1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\frac{1}{\sqrt{n}}$.
- Par suite, $IF_{A,n}$ est inclus dans $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$, et ils sont d'autant plus voisins que n est grand. $\square$

Démonstration

- Au seuil 95%, on a $u_{0,05} \approx 1,96$. L'intervalle asymptotique s'écrit : $IF_{A,n} = \left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
- La fonction, $x \in [0,1] \mapsto f(x) = x(1-x)$, admet un maximum en $x = 0,5$ égal à $f(0,5) = 0,25$.
- Les inégalités $1,96 \leq 2$ et $\sqrt{p(1-p)} \leq 0,5$, permettent de majorer $1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\frac{1}{\sqrt{n}}$.
- Par suite, $IF_{A,n}$ est inclus dans $\left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]$, et ils sont d'autant plus voisins que n est grand. $\square$

Démonstration

- Au seuil 95%, on a $u_{0,05} \approx 1,96$. L'intervalle asymptotique s'écrit : $IF_{A,n} = \left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
- La fonction, $x \in [0,1] \mapsto f(x) = x(1-x)$, admet un maximum en $x = 0,5$ égal à $f(0,5) = 0,25$.
- Les inégalités $1,96 \leq 2$ et $\sqrt{p(1-p)} \leq 0,5$, permettent de majorer $1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}$ par $\dfrac{1}{\sqrt{n}}$.
- Par suite, $IF_{A,n}$ est inclus dans $\left[p - \dfrac{1}{\sqrt{n}}, p + \dfrac{1}{\sqrt{n}}\right]$, et ils sont d'autant plus voisins que n est grand. $\square$

Conséquences sur $\mathbb{P}(F_n \in IF_{A,n})$ et $\mathbb{P}(F_n \in IF_2)$

- On a $\mathbb{P}(F_n \in IF_{A,n}) \leq \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right)$ et, pour n grand, $\mathbb{P}(F_n \in IF_{A,n})$ est proche de $0,95$.
- $\lim_{n\to+\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, mais comme la suite de terme général $\mathbb{P}(F_n \in IF_{A,n})$ n'est pas monotone, on ne peut pas affirmer que $\mathbb{P}(F_n \in IF_{A,n}) \geq 0,95$.
- On n'a pas nécessairement $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$.

Conséquences sur $\mathbb{P}(F_n \in IF_{A,n})$ et $\mathbb{P}(F_n \in IF_2)$

- On a $\mathbb{P}(F_n \in IF_{A,n}) \leq \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right)$ et, pour n grand, $\mathbb{P}(F_n \in IF_{A,n})$ est proche de $0,95$.
- $\lim_{n\to+\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, mais comme la suite de terme général $\mathbb{P}(F_n \in IF_{A,n})$ n'est pas monotone, on ne peut pas affirmer que $\mathbb{P}(F_n \in IF_{A,n}) \geq 0,95$.
- On n'a pas nécessairement $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$.

Conséquences sur $\mathbb{P}(F_n \in IF_{A,n})$ et $\mathbb{P}(F_n \in IF_2)$

- On a $\mathbb{P}(F_n \in IF_{A,n}) \leq \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right)$ et, pour n grand, $\mathbb{P}(F_n \in IF_{A,n})$ est proche de $0,95$.
- $\lim_{n \to +\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, mais comme la suite de terme général $\mathbb{P}(F_n \in IF_{A,n})$ n'est pas monotone, on ne peut pas affirmer que $\mathbb{P}(F_n \in IF_{A,n}) \geq 0,95$.
- On n'a pas nécessairement $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$.

Contre-exemple

Les outils de Première permettent de calculer la probabilité
$$\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right).$$

- Pour $n = 30$ et $p = 0,55$, l'intervalle de fluctuation de Seconde est $[0,3674; 0,7326]$ au seuil 95%.
- Un calcul direct avec la loi binomiale conduit à $\mathbb{P}(0,3674 \leq F_n \leq 0,7326) = \mathbb{P}(11,02 \leq X_n \leq 21,98) = 0,935$.
- Bien que dans les conditions de grandes binomiales, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}\,,\, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Comportement asymptotique de $\mathcal{C}_2(n,p) = \mathbb{P}(F_n \in \ IF_2)$

Proposition : *Pour tout* $p \in]0,1[$,

$$\begin{aligned}
\lim_{n\to\infty} \quad & \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right) \\
= \quad & \mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}} \leq Z \leq \frac{1}{\sqrt{p(1-p)}}\right) \\
\geq \quad & \mathbb{P}\left(-2 \leq Z \leq 2\right) > 0,95\,,
\end{aligned}$$

où Z *est une v.a. normale centrée-réduite.*

Comportement asymptotique de $\mathcal{C}_2(n,p)$: démonstration-1/2

Pour tout n et tout p,

$$\begin{aligned}
\mathcal{C}_2(n,p) &= \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \\
&= \mathbb{P}\left(X_n \in \left[np - \sqrt{n}, np + \sqrt{n}\right]\right) \\
&= \mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}} \leq \frac{X_n - np}{\sqrt{np(1-p)}} \leq \frac{1}{\sqrt{p(1-p)}}\right) \\
&= \mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}} \leq Z_n \leq \frac{1}{\sqrt{p(1-p)}}\right),
\end{aligned}$$

où on a posé $Z_n = \dfrac{X_n - np}{\sqrt{np(1-p)}}$.

Comportement asymptotique de $\mathcal{C}_2(n,p)$: démonstration-2/2

- On conclut par le théorème de De Moivre-Laplace :

$$\begin{aligned}&\lim_{n\to+\infty}\mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}}\leq Z_n\leq\frac{1}{\sqrt{p(1-p)}}\right)\\ =\ &\mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}}\leq Z\leq\frac{1}{\sqrt{p(1-p)}}\right).\end{aligned}$$

- De plus, pour tout $p\in]0,1[$, $\frac{1}{\sqrt{p(1-p)}}\geq 2$; ce qui implique

$$\begin{aligned}\mathbb{P}\left(-\frac{1}{\sqrt{p(1-p)}}\leq Z\leq\frac{1}{\sqrt{p(1-p)}}\right)\ &\geq\ \mathbb{P}(-2\leq Z\leq 2)\\ &\simeq\ 0,954>0,95.\ \square\end{aligned}$$

Conséquences

♠ L'IF (resp. IC) de Seconde n'est donc pas un IF (resp. IC) asymptotique au seuil 95%.

- **Corollaire** : *Pour tout* $p \in]0,1[$, *il existe un entier naturel* n_0, *tel que , pour tout entier* $n \geq n_0$,
$$\mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}, F_n + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95.$$
- Représentation graphique de $p \mapsto \lim_{n\to\infty} \mathcal{C}_2(n,p)$: Voir le graphique

Inversion de l'IF asymptotique

Problème de l'inversion

Inversion de l'IF de Seconde

- On utilise en Seconde l'équivalence :

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

de sorte que :

$$\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right) = \mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}\,,\, F_n + \frac{1}{\sqrt{n}}\right]\right).$$

♣ On dit qu'on a **inversé** l'intervalle de fluctuation.

Problème général de l'inversion

- **Problème** : Étant donné un IF $[f_1(n,p)\,,\, f_2(n,p)]$, on cherche deux v.a. $P_1(n,F_n)$ et $P_2(n,F_n)$ de façon à avoir l'équivalence :

$$F_n \in [f_1(n,p)\,,\, f_2(n,p)] \;\Leftrightarrow p \;\in [P_1(n,F_n)\,,\, P_2(n,F_n)]\,.$$

- En conséquence :

$$\mathbb{P}\left(p \in [P_1(n,F_n)\,,\, P_2(n,F_n)]\right) = \mathbb{P}\left(F_n \in [f_1(n,p)\,,\, f_2(n,p)]\right).$$

♣ On dira qu'on a résolu **le problème de l'inversion** de l'intervalle de fluctuation $[f_1, f_2]$.

Application à l'IF asymptotique

Inversion de l'IF asymptotique - 1/2

- On pose pour simplifier les écritures : $F = F_n$, $u = u_\alpha$ et

$$I_n = IF_A = \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}} \right].$$

- Il s'agit de trouver $P_1 = P_1(n, F)$ et $P_2 = P_2(n, F)$ de façon à avoir l'équivalence $F \in I_n \Leftrightarrow p \in [P_1\,,\,P_2]$.
- On aura donc $\lim\limits_{n \to +\infty} \mathbb{P}\left(p \in [P_1(n, F_n), P_2(n, F_n)]\right) = \lim\limits_{n \to +\infty} \mathbb{P}\left(F_n \in I_n\right) = 1 - \alpha$.

Inversion de l'IF asymptotique - 2/2

On peut écrire :

$$
\begin{aligned}
F \in I_n \Leftrightarrow |F - p| \leq u\frac{\sqrt{p(1-p)}}{\sqrt{n}} \\
\Leftrightarrow \quad & (F-p)^2 \leq u^2\,\frac{p(1-p)}{n} \\
\Leftrightarrow \quad & F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} \leq 0 \\
\Leftrightarrow \quad & p^2\left(1+\frac{u^2}{n}\right) - 2p\left(F + \frac{u^2}{2n}\right) + F^2 \leq 0.
\end{aligned}
$$

Représentation graphique

- Dans un repère orthonormé $(O; p, F)$, l'équation $F^2 + p^2 \left(1 + \frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} = 0$ est celle d'une ellipse passant par l'origine et le point de coordonnées $(1, 1)$, points où elle a une tangente verticale.
- Représentation graphique de l'ellipse : Voir l'ellipse

Représentation graphique

- Dans un repère orthonormé $(O; p, F)$, l'équation $F^2 + p^2\left(1+\dfrac{u^2}{n}\right) - 2pF - \dfrac{u^2 p}{n} = 0$ est celle d'une ellipse passant par l'origine et le point de coordonnées $(1, 1)$, points où elle a une tangente verticale.
- Représentation graphique de l'ellipse : Voir l'ellipse

Signification des parties hors carré-unité

- Les parties de l'ellipse extérieures au carré-unité sont sans signification.
- Elles correspondent aux valeurs de n et de p pour lesquelles l'approximation normale n'est pas acceptée :
  - ***"Partie gauche"*** :
    $p - u\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \leq 0 \Rightarrow np \leq (1-p)u^2 \leq u^2$. Au niveau de confiance 95%, $u < 2$, d'où $np \leq 4 < 5$.
  - ***"Partie droite"*** :
    $p + u\dfrac{\sqrt{p(1-p)}}{\sqrt{n}} \geq 1 \Rightarrow n(1-p) \leq pu^2 \leq u^2$. Au niveau de confiance 95%, $u < 2$, d'où $n(1-p) \leq 4 < 5$.

Expression analytique de l'IC, inverse de IF_A

- $F \in I_n \Leftrightarrow F^2 + p^2\left(1 + \frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} \leq 0 \Leftrightarrow p \in [P_1, P_2]$, où P_1 et P_2 avec $P_1 \leq P_2$, sont les racines du trinôme de second degré en p : $F^2 + p^2\left(1 + \frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} = 0$.
- Les bornes de l'IC sont donc les solutions de l'équation du second degré : $p^2\left(1 + \frac{u^2}{n}\right) - 2p\left(F + \frac{u^2}{2n}\right) + F^2 = 0$.
- Son discriminant réduit s'écrit :

$$\Delta' = \left(F + \frac{u^2}{2n}\right)^2 - \left(1 + \frac{u^2}{n}\right)F^2 = \frac{u^2}{n}\left[\frac{u^2}{4n} + F(1-F)\right] > 0.$$

Expression analytique de l'IC, inverse de IF_A

- $F \in I_n \Leftrightarrow F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} \leq 0 \Leftrightarrow p \in [P_1, P_2]$, où P_1 et P_2 avec $P_1 \leq P_2$, sont les racines du trinôme de second degré en p : $F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} = 0$.
- Les bornes de l'IC sont donc les solutions de l'équation du second degré : $p^2\left(1+\frac{u^2}{n}\right) - 2p\left(F + \frac{u^2}{2n}\right) + F^2 = 0$.
- Son discriminant réduit s'écrit :

$$\Delta' = \left(F + \frac{u^2}{2n}\right)^2 - \left(1+\frac{u^2}{n}\right)F^2 = \frac{u^2}{n}\left[\frac{u^2}{4n} + F(1-F)\right] > 0.$$

Expression analytique de l'IC, inverse de IF_A

- $F \in I_n \Leftrightarrow F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} \leq 0 \Leftrightarrow p \in [P_1, P_2]$, où P_1 et P_2 avec $P_1 \leq P_2$, sont les racines du trinôme de second degré en p : $F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} = 0$.
- Les bornes de l'IC sont donc les solutions de l'équation du second degré : $p^2\left(1+\frac{u^2}{n}\right) - 2p\left(F+\frac{u^2}{2n}\right) + F^2 = 0$.
- Son discriminant réduit s'écrit :

$$\Delta' = \left(F+\frac{u^2}{2n}\right)^2 - \left(1+\frac{u^2}{n}\right)F^2 = \frac{u^2}{n}\left[\frac{u^2}{4n} + F(1-F)\right] > 0.$$

Bornes de IC_W

- D'où les bornes de l'intervalle de confiance :

$$P_1 = \frac{\left(F_n + \frac{u_\alpha^2}{2n}\right) - \frac{u_\alpha}{\sqrt{n}}\sqrt{\frac{u_\alpha^2}{4n} + F_n(1 - F_n)}}{\left(1 + \frac{u_\alpha^2}{n}\right)},$$

$$P_2 = \frac{\left(F_n + \frac{u_\alpha^2}{2n}\right) + \frac{u_\alpha}{\sqrt{n}}\sqrt{\frac{u_\alpha^2}{4n} + F_n(1 - F_n)}}{\left(1 + \frac{u_\alpha^2}{n}\right)}.$$

♣ L'intervalle $[P_1, P_2]$, noté IC_W, est appelé **intervalle de confiance de Wilson** (1927).

IC standard

et

IC de Terminale

L'IC standard comme approximation de IC_W

- En général $0 < u_\alpha \leq 3$, d'où $u_\alpha^2 \leq 9$. Si n est grand, on peut alors négliger les termes en $\dfrac{1}{n}$ dans les expressions de P_1 et P_2, d'où :
$$P_1 \simeq F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \text{ et } P_2 \simeq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}.$$

♣ L'intervalle aléatoire
$$\left[F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}, F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \right],$$
noté IC_S est appelé **intervalle de confiance standard** ou **de Wald-Laplace** (1812).

L'IC de Terminale comme approximation de l'IC standard

- Par une majoration classique, on obtient que l'IC standard au seuil 95% est inclus dans $IC_2 = \left[F_n - \frac{1}{\sqrt{n}}\,,\,F_n + \frac{1}{\sqrt{n}}\right]$,
- et ces deux IC sont d'autant plus voisins que n est grand.
- On justifie ainsi IC_2, préconisé par les programmes de Terminale et de Seconde, comme approximation de l'IC standard.

Comparaison des IC pour une même observation

- Considérons un échantillon de taille $n = 1000$ pour lequel la fréquence observée est $f = 0,51$.
- La réalisation des divers IC pour la proportion p au niveau de confiance de 95% correspondant à cette observation donne :
  - IC de Seconde-Terminale : $IC_2 = [0,478377 \; ; \; 0,541623]$;
  - IC de Wilson : $IC_W = [0,479036 \; ; \; 0,540887]$;
  - IC standard : $IC_S = [0,479015 \; ; \; 0,540985]$.

Abaques IF de Seconde, de Première et de Terminale

- Si $[f_1(n,p)\ ,\ f_2(n,p)]$ est un IF pour F_n, on appelle **abaques** les courbes représentatives des fonctions numériques de paramètre n, $p \in [0,1] \mapsto f_1(n,p)$ et $p \in [0,1] \mapsto f_2(n,p)$.
- Par exemple, pour l'IF de Seconde :

  abaque seconde

- Abaques pour les IF de Seconde, de Première et de Terminale :
  Voir les graphiques

Approche directe de l'IC standard

Corollaire du théorème de De Moivre-Laplace

- Considérons, pour tout entier $n \geq 1$, la variable aléatoire
$$T_n = \frac{F_n - p}{\sqrt{\frac{F_n(1-F_n)}{n}}}.$$
- **Théorème** : *pour tous réels a et b avec $a < b$, on a*
$$\lim_{n\to+\infty} \mathbb{P}(a \leq T_n \leq b) = \mathbb{P}(a \leq Z \leq b),$$
où Z est une variable aléatoire normale centrée-réduite.
- On admettra ce résultat qui est une conséquence du théorème de De Moivre-Laplace par application d'un théorème, dit **de Slutsky**, sur les convergences en loi et en probabilité.
- La variable aléatoire T_n est un pivot asymptotique pour p.

Approche directe de l'IC standard

- En raisonnant comme pour l'IF asymptotique, on obtient que, pour tout $p \in]0,1[$,
$$\lim_{n\to+\infty} \mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right) = 1-\alpha\,.$$
- En pratique, dans les conditions des grandes binomiales, on accepte l'approximation
$$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right) \simeq 1-\alpha\,.$$

Qualité statistique d'un IC ou d'un IF

Quels indicateurs ?

Commentaire sur la construction des IC et IF

Dans la construction des IC et IF précédents, l'attention a constamment porté sur :

- la **probabilité que p soit dans l'IC, ou que F_n soit dans l'IF** : on la souhaite la plus proche et au-dessus du seuil fixé ;
- les **conditions d'utilisation de l'IC ou de l'IF** : on les souhaite les plus générales possibles ;
- la **longueur de l'IC ou de l'IF** : on la souhaite la plus petite possible.

Indicateurs de la qualité statistique d'un IC (ou d'un IF)

Pour étudier la qualité statistique d'un IC $[P_1\ ,\ P_2]$ ou d'un IF $[f_1\ ,\ f_2]$, on s'intéressera naturellement à trois types d'indicateurs :

- sa **probabilité de recouvrement**, i.e. resp. $\mathbb{P}\,(p \in [P_1\ ,\ P_2])$ ou $\mathbb{P}\,(F_n \in [f_1\ ,\ f_2])$.
  En particulier, si IC et IF sont inverses,
  $$\mathbb{P}\,(p \in [P_1\ ,\ P_2]) = \mathbb{P}\,(F_n \in [f_1\ ,\ f_2])$$
- ses **conditions de validité**, i.e. les conditions de l'observation rendant l'application de l'IC pertinente en statistique,
- sa **longueur** $f_2 - f_1$ pour l'IF et sa **longueur espérée** $\mathbb{E}(P_2 - P_1)$ pour l'IC (on n'en parlera pas ici).

Probabilités de recouvrement

Proba. recouvrement : IC et IF de Seconde

- La probabilité de recouvrement de l'IC, inverse de l'IF, de Seconde est $\mathcal{C}_2(n,p) = \mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}\,,\, F_n + \frac{1}{\sqrt{n}}\right]\right) = \mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right)$.
- Graphe de la fonction $p \in]0,1[\mapsto \mathcal{C}_2(n,p)$ paramétrée par n : Graphiques .

Proba. recouvrement : IF de Première

- La probabilité de recouvrement de l'IF de Première $\left[\frac{a_n}{n}, \frac{b_n}{n}\right]$ est $\mathcal{C}_E(n,p) = \mathbb{P}\left(F_n \in \left[\frac{a_n}{n}, \frac{b_n}{n}\right]\right)$.
- Graphe de la fonction $p \in]0,1[\mapsto \mathcal{C}_E(n,p)$ paramétrée par n : Graphiques .

Proba. recouvrement : IF asymptotique et IC de Wilson

- La probabilité de recouvrement de l'IF asymptotique, inverse de l'IC de Wilson, s'écrit :
$$\mathcal{C}_W(n,p) = \mathbb{P}\left(p \in IC_W\right) =$$
$$\mathbb{P}\left(F_n \in \left[p - u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right).$$
- Graphe de la fonction $p \in]0,1[\mapsto \mathcal{C}_W(n,p)$ paramétrée par n : Graphiques .

Proba. recouvrement : IC standard

- La probabilité de recouvrement de l'IC standard s'écrit :
  $\mathcal{C}_S(n,p) =$
  $$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right).$$
- Graphe de la fonction $p \in]0,1[\mapsto \mathcal{C}_S(n,p)$ paramétrée par n :
  Graphiques .

Étude de $\mathcal{C}_S(n,p)$

- Le couple (n,p) est dit **heureux au seuil** $1-\alpha$ si $\mathcal{C}_S(n,p)$ est très proche ou supérieur à $1-\alpha$. Dans le cas contraire, le couple (n,p) est dit **malheureux**.
- Exemples de couples heureux ou malheureux étudiés dans [BROWN, 2001] : Voir les graphiques .

Proba. recouvrement pour divers IC

- De nombreux IC sont construits et étudiés par les statisticiens.
- En général, les expressions des probabilités de recouvrement en fonction de n et p sont très compliquées.
- Graphiques de $\mathcal{C}(n, p)$ pour différents IC étudiés dans [BROWN, 2001] pour $n = 50$: Voir les graphiques .

Conditions de validité pour l'IC standard

Conditions de validité pour IC_S

- **Théorème** (cf. [BROWN, 2002]) : *Soit $\gamma > 0$. Si on note $\mathcal{D}(n)$ l'ensemble des $p \in]0,1[$ tels que $np \geq \gamma$ et $n(1-p) \geq \gamma$, on obtient pour l'IC standard l'inégalité suivante :*

$$\lim_{n\to\infty} \inf_{p\in\mathcal{D}(n)} \mathcal{C}_S(n,p) \leq M_{\gamma,\alpha},$$

où $M_{\gamma,\alpha} \in [0,1]$.

- Quelques valeurs de $M_{\gamma,\alpha}$ pour $\alpha = 5\%$: $M_{5,\,5\%} \simeq 0,875$, $M_{7,\,5\%} \simeq 0,913$ et $M_{10,\,5\%} \simeq 0,926$.
- Représentation graphique de $\gamma \mapsto M_{\gamma,\,5\%}$: Voir le graphique .

Conditions de validité au lycée pour IC_S

- Au lycée, on a en général $\gamma = 5$ et $\alpha = 0,05$, donc $M_{5,\,5\%} \simeq 0,875$.
- $\lim\limits_{n \to \infty} \inf\limits_{p \in \mathcal{D}(n)} \mathcal{C}_S(n,p) \leq 0,875.$
- Il existe un entier n_0 tel que, pour tout entier $n \geq n_0$, il existe un p vérifiant $np \geq 5$ et $n(1-p) \geq 5$ avec $\mathcal{C}_S(n,p) \leq 0,90 < 0,95$.
- À partir d'un certain n_0, pour tout $n \geq n_0$, il existe un couple (n,p) vérifiant les conditions de validité de terminale et malheureux pour l'IC standard.

Quel IC choisir ?

Remarques

- l'IC standard est "simple à utiliser" mais pas "bon" ;
- l'IC de Wilson est "bon" mais pas "simple à utiliser" ;
- A. Agresti et B.A. Coull (1998) ont combiné ces deux IC pour obtenir un "bon" IC "simple à utiliser".

Remarques

- l'IC standard est "simple à utiliser" mais pas "bon" ;
- l'IC de Wilson est "bon" mais pas "simple à utiliser" ;
- A. Agresti et B.A. Coull (1998) ont combiné ces deux IC pour obtenir un "bon" IC "simple à utiliser".

Remarques

- l'IC standard est "simple à utiliser" mais pas "bon" ;
- l'IC de Wilson est "bon" mais pas "simple à utiliser" ;
- A. Agresti et B.A. Coull (1998) ont combiné ces deux IC pour obtenir un "bon" IC "simple à utiliser".

IC d'Agresti-Coull

- L'**intervalle de confiance d'Agresti-Coull** au niveau de confiance $1-\alpha$ est l'intervalle

$$IC_{AC} = \left[\widetilde{F_n} - u_\alpha \frac{\sqrt{\widetilde{F_n}(1-\widetilde{F_n})}}{\sqrt{\widetilde{n}}} \; ; \; \widetilde{F_n} + u_\alpha \frac{\sqrt{\widetilde{F_n}(1-\widetilde{F_n})}}{\sqrt{\widetilde{n}}} \right],$$

où $\widetilde{F_n} = \dfrac{F_n + \dfrac{u_\alpha^2}{2n}}{1 + \dfrac{u_\alpha^2}{n}}$ est le centre de l'IC de Wilson.

♣ cet intervalle est étudié dans [AGRESTI, 1998].

Relations entre ces IC

- On peut aussi écrire : $\widetilde{F_n} = \dfrac{\widetilde{X_n}}{\widetilde{n}}$, en posant $\widetilde{X_n} = X_n + \dfrac{u_\alpha^2}{2}$ et $\widetilde{n} = n + u_\alpha^2$.
- Au niveau de confiance 95%, avec l'approximation $u_\alpha \simeq 2$, l'intervalle d'Agresti-Coull $IC_{AC}(n, F_n)$ a la même formulation que l'intervalle standard $IC_S(\widetilde{n}, \widetilde{F_n})$ dans lequel on a "ajouté" 2 succès et 2 échecs à n.
- Les IC de Wilson et d'Agresti-Coull sont, tous les deux, centrés sur $\widetilde{F_n}$ au lieu de F_n.

IC d'Agresti-Coull : exemple numérique

Retour sur l'exemple numérique ($n = 1000$ et $f = 0,51$) vu plus haut :

- IC d'Agresti-Coull : $IC_{AC} = [0,478975\ ;\ 0,540945]$;
- qu'on peut comparer avec les réalisations des IC déjà vues :

  * IC de Seconde-Terminale : $IC_2 = [0,478377\ ;\ 0,541623]$;
  * IC de Wilson : $IC_W = [0,479036\ ;\ 0,540887]$;
  * IC standard : $IC_S = [0,479015\ ;\ 0,540985]$.

IC d'Agresti-Coull : exemple numérique

Retour sur l'exemple numérique ($n = 1000$ et $f = 0,51$) vu plus haut :

- IC d'Agresti-Coull : $IC_{AC} = [0,478975\ ;\ 0,540945]$;
- qu'on peut comparer avec les réalisations des IC déjà vues :

  * IC de Seconde-Terminale : $IC_2 = [0,478377\ ;\ 0,541623]$;
  * IC de Wilson : $IC_W = [0,479036\ ;\ 0,540887]$;
  * IC standard : $IC_S = [0,479015\ ;\ 0,540985]$.

Conclusion

Quel IC choisir ?

The performance is so erratic and the qualifications given in the influential texts are so defective that the standard Wald interval should be not used. [...] We recommended the Wilson interval for small n ($n \leq 40$) [...]. For larger n, the Wilson and the Agresti-Coull intervals are all comparable, and the Agresti-Coull interval is the simplest to present.

It is generally true in statistical practice that only those methods that are easy to describe, remember and compute are widely used. Keeping this in mind, we recommend the Agresti-Coull interval for practical use when $n > 40$. Even for small n the easy-to-present Agresti-Coull interval is much preferable to the standard one. (Cf. [BROWN, 2001])

Bibliographie

Bibliographie

Brown L.D., Cai T.T., Dasgupta A.,
Interval estimation for a binomial proportion
Statistical Science, Vol. 16, No. 2, pp. 101-133, 2001.

Brown L.D., Cai T.T., Dasgupta A.,
Confidence intervals for a binomial proportion and asymptotic expansions
The Annals of Statistics, Vol. 30, No. 1, pp. 160-201, 2002.

Bibliographie

Agresti A., Coull B.A.,
Approximate is better than "exact" for interval estimation of binomial proportions
The American Statistician, Vol. 52, No. 2, May 1998, pp. 119-126, American Statistical Association, 1998.

Clopper C.J., Pearson E.S.,
The use of confidence or fiducial limits illustrated in the case of the binomial
Biometrika, Vol. 26, No. 4, December 1934, pp. 404-413, Biometrika Trust, 1934.

Merci de votre attention

Pour nous contacter ...

- **Yves Ducel, Bruno Saussereau**
  Institut de recherche sur l'enseignement des mathématiques (IREM)
  YD : 03 81 66 62 32 / BS : 03 81 66 63 00
  yves.ducel@univ-fcomte.fr ; bruno.saussereau@univ-fcomte.fr

- **Site Web de l'IREM** : http ://www-irem.univ-fcomte.fr/

- **Site Web du groupe IREM "Proba-Stat"** :
  http ://bsauss.perso.math.cnrs.fr/IREM_FC_GrouProbaStat/grouprobastat.html