INTERVALLES DE FLUCTUATION ET DE CONFIANCE POUR UNE PROPORTION

Aspects mathématiques et statistiques

Séminaire IREM,
Besançon, 29 juin 2012

Yves DUCEL, Bruno SAUSSEREAU
Institut de recherche sur l'enseignement des mathématiques
Laboratoire de Mathématiques de Besançon

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.
- Large exploitation du travail effectué dans le groupe de rédaction du documents ressource "Probab-stat" de Première.
- Le diaporama sera disponible.

Programmes 2009-2012

Sur le programme de Seconde

- Pratiquement la seule définition exploitable en classe est celle donnée par $\left[p - \dfrac{1}{\sqrt{n}}\,,\, p + \dfrac{1}{\sqrt{n}}\right]$.
- La formulation de l'IF est admise ; éventuellement justifiée par simulation.
- La formule de l'IF n'est donnée que sous les conditions d'utilisation : $n > 25$ et $0,2 < p < 0,8$.
- Le seuil 95% de l'IF ne peut pas être changé car lié à l'expression de l'IF admise.
- La probabilité $\mathbb{P}\left(F_n \in \left[p - \dfrac{1}{\sqrt{n}}\,,\, p + \dfrac{1}{\sqrt{n}}\right]\right)$ joue un rôle important et doit être étudiée de plus près.

Sur le programme de Première

- On parle de l'intervalle de fluctuation sans donner une expression analytique.
- L'IF de Première est construit en classe ;
- cette construction s'appuie sur les acquis de Seconde.
- L'IF de Première se détermine à l'aide d'un tableur.
- L'intervalle de confiance et son application à l'estimation ne sont pas au programme.

Sur le programme de Terminale

- L'IF asymptotique est donné par son expression analytique.
- Lien avec l'IF de Seconde explicitement mentionné, mais pas de référence explicite à l'IF de Première.
- On se limite aux conditions $n \geq 30$ avec $np \geq 5$ et $n(1-p) \geq 5$ (formulation différentes de la Seconde).
- L'intervalle de confiance et son application à l'estimation sont au programme.
- L'IC préconisé en Terminale est l'IC de Seconde, mais un autre IC (IC standard) est mentionné comme important.
- Les conditions de validité pour l'IC proposées ici sont énoncées avec p inconnu.

Problématique et conventions

Problématique de l'exposé

1. Étudier la continuité des programmes de lycée sur les intervalles de fluctuation et de confiance, et apporter un éclairage sur les choix effectués :
   * en comparant les diverses définitions introduites de la Seconde à la Terminale ;
   * en explicitant le cadre mathématique qu'elles supposent, notamment en vue du nouveau programme de Terminale ;
   * 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.

Problématique de l'exposé

1. Étudier la continuité des programmes de lycée sur les intervalles de fluctuation et de confiance, et apporter un éclairage sur les choix effectués :
   * en comparant les diverses définitions introduites de la Seconde à la Terminale ;
   * en explicitant le cadre mathématique qu'elles supposent, notamment en vue du nouveau programme de Terminale ;
   * 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.

Problématique de l'exposé

1. Étudier la continuité des programmes de lycée sur les intervalles de fluctuation et de confiance, et apporter un éclairage sur les choix effectués :
   * en comparant les diverses définitions introduites de la Seconde à la Terminale ;
   * en explicitant le cadre mathématique qu'elles supposent, notamment en vue du nouveau programme de Terminale ;
   * 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.

Problématique de l'exposé

1. Étudier la continuité des programmes de lycée sur les intervalles de fluctuation et de confiance, et apporter un éclairage sur les choix effectués :
   * en comparant les diverses définitions introduites de la Seconde à la Terminale ;
   * en explicitant le cadre mathématique qu'elles supposent, notamment en vue du nouveau programme de Terminale ;
   * 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.

Problématique de l'exposé

1. Étudier la continuité des programmes de lycée sur les intervalles de fluctuation et de confiance, et apporter un éclairage sur les choix effectués :
   * en comparant les diverses définitions introduites de la Seconde à la Terminale ;
   * en explicitant le cadre mathématique qu'elles supposent, notamment en vue du nouveau programme de Terminale ;
   * 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.

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 et IC :
Seconde et Première

IF et IC : Seconde

IF de Seconde

L'intervalle $IF_2(n,p) = \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]$ est appelé **intervalle de fluctuation pour la fréquence empirique** F_n **au seuil** 95%

IF et IC en Seconde

On a l'équivalence :

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

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.

IC de Seconde

L'intervalle $IC_2(n, F_n) = \left[F_n - \frac{1}{\sqrt{n}}\, , \, F_n + \frac{1}{\sqrt{n}}\right]$ est appelé **intervalle de confiance pour la proportion** p **au seuil** 95% ;

IF et IC : utilisation

Étant donné une pièce, on note p la probabilité (a priori inconnue) d'obtenir "Pile". On a deux cas de figure :

- **soit je pense que la pièce est équilibrée ;**
- soit je pense que la pièce est truquée.

IF et IC : utilisation

Étant donné une pièce, on note p la probabilité (a priori inconnue) d'obtenir "Pile". On a deux cas de figure :

- **soit je pense que la pièce est équilibrée** ;
- **soit je pense que la pièce est truquée**.

IF et IC : utilisation

Si je pense que la pièce est **équilibrée** :

- est-ce que "$p = 0,5$" ?
- appel à une démarche de prise de décision sur le rejet ou non de l'hypothèse "$p = 0,5$" à partir de l'observation d'un échantillon de taille n ;
- travail avec $p = 0,5$ connu ;
- intervention de l'intervalle de fluctuation $IF_2(n;\ 0,5)$ pour F_n avec $p = 0,5$.

IF et IC : utilisation

Si je pense que la pièce est **truquée** :

- quelle valeur attribuer à p ?
- appel à une démarche d'estimation de p à partir de l'observation d'un échantillon de taille n ;
- travail avec p inconnu ;
- intervention de l'intervalle de confiance $IC_2(n, F_n)$ pour p.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n,p)$ et $f_2(n,p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \dfrac{1}{\sqrt{n}}$ et $f_2 = p + \dfrac{1}{\sqrt{n}}$;
- on notera $P_1(n,F_n)$ et $P_2(n,F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \dfrac{1}{\sqrt{n}}$ et $P_2 = F_n + \dfrac{1}{\sqrt{n}}$.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n,p)$ et $f_2(n,p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \dfrac{1}{\sqrt{n}}$ et $f_2 = p + \dfrac{1}{\sqrt{n}}$;
- on notera $P_1(n,F_n)$ et $P_2(n,F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \dfrac{1}{\sqrt{n}}$ et $P_2 = F_n + \dfrac{1}{\sqrt{n}}$.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n, p)$ et $f_2(n, p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \dfrac{1}{\sqrt{n}}$ et $f_2 = p + \dfrac{1}{\sqrt{n}}$;
- on notera $P_1(n, F_n)$ et $P_2(n, F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \dfrac{1}{\sqrt{n}}$ et $P_2 = F_n + \dfrac{1}{\sqrt{n}}$.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n,p)$ et $f_2(n,p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \dfrac{1}{\sqrt{n}}$ et $f_2 = p + \dfrac{1}{\sqrt{n}}$;
- on notera $P_1(n,F_n)$ et $P_2(n,F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \dfrac{1}{\sqrt{n}}$ et $P_2 = F_n + \dfrac{1}{\sqrt{n}}$.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n,p)$ et $f_2(n,p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \dfrac{1}{\sqrt{n}}$ et $f_2 = p + \dfrac{1}{\sqrt{n}}$;
- on notera $P_1(n,F_n)$ et $P_2(n,F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \dfrac{1}{\sqrt{n}}$ et $P_2 = F_n + \dfrac{1}{\sqrt{n}}$.

Notations

Dans la suite de cet exposé,

- on notera $f_1(n, p)$ et $f_2(n, p)$ (et plus simplement f_1 et f_2) les bornes d'un intervalle de fluctuation pour F_n :
  - les bornes f_1 et f_2 sont des nombres réels déterministes, fonctions de n et de p ;
  - en Seconde : $f_1 = p - \frac{1}{\sqrt{n}}$ et $f_2 = p + \frac{1}{\sqrt{n}}$;
- on notera $P_1(n, F_n)$ et $P_2(n, F_n)$ (et plus simplement P_1 et P_2) les bornes d'un intervalle de confiance pour p :
  - les bornes P_1 et P_2 sont des variables aléatoires, fonctions de n et de la variable aléatoire F_n ;
  - en Seconde : $P_1 = F_n - \frac{1}{\sqrt{n}}$ et $P_2 = F_n + \frac{1}{\sqrt{n}}$.

Abaque Seconde

- Représentation graphique, pour n fixé, des fonctions numériques $p \in [0,1] \mapsto f_1(n,p)$ et $p \in [0,1] \mapsto f_2(n,p)$.

abaque seconde

Résumé

- On passe de l'IF à l'IC en "inversant" les bornes de l'IF.
- Les bornes de l'IF et celles de l'IC sont données par des expressions analytiques.
- Leurs représentations graphiques donnent une famille de courbes, indexée par n, appelée **abaque**.
- On peut déterminer les bornes de l'IC graphiquement avec l'abaque.
- Les conditions de grandes binomiales ne sont pas respectées dans les zones hors carré-unité.

Résumé

- On passe de l'IF à l'IC en "inversant" les bornes de l'IF.
- Les bornes de l'IF et celles de l'IC sont données par des expressions analytiques.
- Leurs représentations graphiques donnent une famille de courbes, indexée par n, appelée **abaque**.
- On peut déterminer les bornes de l'IC graphiquement avec l'abaque.
- Les conditions de grandes binomiales ne sont pas respectées dans les zones hors carré-unité.

Résumé

- On passe de l'IF à l'IC en "inversant" les bornes de l'IF.
- Les bornes de l'IF et celles de l'IC sont données par des expressions analytiques.
- Leurs représentations graphiques donnent une famille de courbes, indexée par n, appelée **abaque**.
- On peut déterminer les bornes de l'IC graphiquement avec l'abaque.
- Les conditions de grandes binomiales ne sont pas respectées dans les zones hors carré-unité.

Résumé

- On passe de l'IF à l'IC en "inversant" les bornes de l'IF.
- Les bornes de l'IF et celles de l'IC sont données par des expressions analytiques.
- Leurs représentations graphiques donnent une famille de courbes, indexée par n, appelée **abaque**.
- On peut déterminer les bornes de l'IC graphiquement avec l'abaque.
- Les conditions de grandes binomiales ne sont pas respectées dans les zones hors carré-unité.

Résumé

- On passe de l'IF à l'IC en "inversant" les bornes de l'IF.
- Les bornes de l'IF et celles de l'IC sont données par des expressions analytiques.
- Leurs représentations graphiques donnent une famille de courbes, indexée par n, appelée **abaque**.
- On peut déterminer les bornes de l'IC graphiquement avec l'abaque.
- Les conditions de grandes binomiales ne sont pas respectées dans les zones hors carré-unité.

IF et IC : Première

Propriété de l'IF de Seconde

Sous les conditions de validité $n \geq 25$ et $0,2 < p < 0,8$, l'IF de Seconde (au seuil $1 - \alpha = 95\%$) est *approximativement* le *plus petit intervalle fermé* $[f_1\,,\,f_2]$ vérifiant les deux relations :

* $\mathbb{P}(F_n \in [0\,,\,f_1[) \leq 0,025 = \dfrac{\alpha}{2}$, où $[0\,,\,f_1[$ est la partie extérieure gauche de l'IF ;
* $\mathbb{P}(F_n \in]f_2\,,\,1]) \leq 0,025 = \dfrac{\alpha}{2}$, où $]f_2\,,\,1]$ est la partie extérieure droite de l'IF.

Conséquence de la propriété de IF_2

- Le fait que IF_2 soit centré en p est une conséquence de la propriété précédente.

  ◇ Cela résulte, sous les conditions des grandes binomiales, de la quasi-symétrie du diagramme en bâtons de F_n et de la symétrisation du risque α.

- Cette remarque sera également exploitée en Terminale.

Passage à 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.

♣ L'IF de Première est aussi appelé **IF exact de F_n au seuil** $1-\alpha$. On le note $IF_E(n,p)$.

Commentaires sur l'IF exact

- La détermination de l'IF exact,
  * est conduite avec des calculs exacts ;
  * est possible pour n et p quelconques ;
  * est possible pour un seuil $1-\alpha$ quelconque.
- Les bornes de l'IF exact ne sont pas explicitées par une "formule".
- Elles dépendent de n et de p et sont déterminées à l'aide d'un outil scientifique.
- Pour tous n et p, on a $\mathbb{P}[F_n \in IF_E(n,p)] \geq 1-\alpha$.
- L'IF exact n'est pas nécessairement centré en p.
- Pour les grandes binomiales, IF de Seconde et IF exact sont très proches.
- La probabilité $\mathbb{P}[F_n \in IF_E(n,p)]$ est d'autant plus proche de $1-\alpha$ que n est grand.

Problème de l'inversion

Inversion de l'IF de Seconde

On a déjà noté 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 :

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

Problème de l'inversion d'un IF

- **Problème** : Étant donné $[f_1(n,p)\,,\ f_2(n,p)]$, on cherche $P_1(n,F_n)$ 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 $[f_1, f_2]$.

Problème de l'inversion d'un IF

- **Problème** : Étant donné $[f_1(n,p)\,,\ f_2(n,p)]$, on cherche $P_1(n,F_n)$ 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 $[f_1, f_2]$.

Problème de l'inversion d'un IF

- **Problème** : Étant donné $[f_1(n,p)\,,\; f_2(n,p)]$, on cherche $P_1(n,F_n)$ 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 IF exact

- Absence d'expression analytique élémentaire pour les bornes de l'IF exact, d'où la difficulté de l'inverser ;
- ce qui peut expliquer qu'on ne parle pas d'intervalle de confiance en Première.
- cependant on peut avoir une démarche graphique pour définir un IC par inversion de l'IF exact ;
- Abaques pour IF de Seconde et IF exact : Voir les graphiques
- Abaques des formulaires de statistique : Voir les graphiques

Application à IF exact

- Absence d'expression analytique élémentaire pour les bornes de l'IF exact, d'où la difficulté de l'inverser ;
- ce qui peut expliquer qu'on ne parle pas d'intervalle de confiance en Première.
- cependant on peut avoir une démarche graphique pour définir un IC par inversion de l'IF exact ;
- Abaques pour IF de Seconde et IF exact : Voir les graphiques
- Abaques des formulaires de statistique : Voir les graphiques

Application à IF exact

- Absence d'expression analytique élémentaire pour les bornes de l'IF exact, d'où la difficulté de l'inverser ;
- ce qui peut expliquer qu'on ne parle pas d'intervalle de confiance en Première.
- cependant on peut avoir une démarche graphique pour définir un IC par inversion de l'IF exact ;
- Abaques pour IF de Seconde et IF exact : Voir les graphiques
- Abaques des formulaires de statistique : Voir les graphiques

Application à IF exact

- Absence d'expression analytique élémentaire pour les bornes de l'IF exact, d'où la difficulté de l'inverser ;
- ce qui peut expliquer qu'on ne parle pas d'intervalle de confiance en Première.
- cependant on peut avoir une démarche graphique pour définir un IC par inversion de l'IF exact ;
- Abaques pour IF de Seconde et IF exact : Voir les graphiques
- Abaques des formulaires de statistique : Voir les graphiques

Application à IF exact

- Absence d'expression analytique élémentaire pour les bornes de l'IF exact, d'où la difficulté de l'inverser ;
- ce qui peut expliquer qu'on ne parle pas d'intervalle de confiance en Première.
- cependant on peut avoir une démarche graphique pour définir un IC par inversion de l'IF exact ;
- Abaques pour IF de Seconde et IF exact : Voir les graphiques
- Abaques des formulaires de statistique : Voir les graphiques

Expression de l'IC inverse de IF_E

- Il existe des formules d'inversion, mais leur démonstration dépasse le niveau du lycée ;
- Ces formules font intervenir la loi de Fisher-Snedecor et la fonction eulérienne β incomplète.
- ♣ L'IC inverse de l'IF exact de Première, noté IC_E, est appelé **intervalle de confiance exact** ou de **Clopper-Pearson** (1934).

Expression de l'IC inverse de IF_E

- Il existe des formules d'inversion, mais leur démonstration dépasse le niveau du lycée ;
- Ces formules font intervenir la loi de Fisher-Snedecor et la fonction eulérienne β incomplète.
- L'IC inverse de l'IF exact de Première, noté IC_E, est appelé **intervalle de confiance exact** ou de **Clopper-Pearson** (1934).

Expression de l'IC inverse de IF_E

- Il existe des formules d'inversion, mais leur démonstration dépasse le niveau du lycée ;
- Ces formules font intervenir la loi de Fisher-Snedecor et la fonction eulérienne β incomplète.

♣ L'IC inverse de l'IF exact de Première, noté IC_E, est appelé **intervalle de confiance exact** ou de **Clopper-Pearson** (1934).

Probabilité de recouvrement de IC_E

- La probabilité $\mathcal{C}_E(n,p) = \mathbb{P}\left[p \in \; IC_E(n,F_n)\right] = \mathbb{P}\left(F_n \in \; \left[\frac{a_n}{n}, \frac{b_n}{n}\right]\right)$, où $\left[\frac{a_n}{n}, \frac{b_n}{n}\right]$ est l'IF exact de Première, est dite **de recouvrement** de IC_E.
- Notons que $\mathcal{C}_E(n,p) = \mathbb{P}\left(X_n \in [a_n, b_n]\right)$, d'où :

$$\mathcal{C}_E(n,p) = \sum_{k \in \mathbb{N} \cap [a_n, b_n]} \binom{n}{k} p^k (1-p)^{n-k} \; .$$

- Variations de $\mathcal{C}_E(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_E

- La probabilité $\mathcal{C}_E(n,p) = \mathbb{P}\left[p \in \; IC_E(n,F_n)\right] = \mathbb{P}\left(F_n \in \; \left[\frac{a_n}{n},\frac{b_n}{n}\right]\right)$, où $\left[\frac{a_n}{n},\frac{b_n}{n}\right]$ est l'IF exact de Première, est dite **de recouvrement** de IC_E.
- Notons que $\mathcal{C}_E(n,p) = \mathbb{P}\left(X_n \in [a_n,b_n]\right)$, d'où :

$$\mathcal{C}_E(n,p) = \sum_{k\in\mathbb{N}\cap[a_n,b_n]} \binom{n}{k} p^k (1-p)^{n-k} \; .$$

- Variations de $\mathcal{C}_E(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_E

- La probabilité $\mathcal{C}_E(n,p) = \mathbb{P}\left[p \in \ IC_E(n,F_n)\right] = \mathbb{P}\left(F_n \in \ \left[\frac{a_n}{n}, \frac{b_n}{n}\right]\right)$, où $\left[\frac{a_n}{n}, \frac{b_n}{n}\right]$ est l'IF exact de Première, est dite **de recouvrement** de IC_E.
- Notons que $\mathcal{C}_E(n,p) = \mathbb{P}\left(X_n \in [a_n, b_n]\right)$, d'où :

$$\mathcal{C}_E(n,p) = \sum_{k \in \mathbb{N} \cap [a_n, b_n]} \binom{n}{k} p^k (1-p)^{n-k} \ .$$

- Variations de $\mathcal{C}_E(n,p)$ en fonction de n et p : Graphiques .

Compléments I sur l'IF de Seconde

Remarque sur l'IF de Seconde

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, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}\,,\, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$. Pourquoi ?
- Le programme de Terminale apportera un élément de réponse.
- On a donc $\mathbb{P}\left(0,55 \in \left[F_{30} - \frac{1}{\sqrt{30}}\,,\, F_{30} + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Remarque sur l'IF de Seconde

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, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$. Pourquoi ?
- Le programme de Terminale apportera un élément de réponse.
- On a donc $\mathbb{P}\left(0,55 \in \left[F_{30} - \frac{1}{\sqrt{30}}, F_{30} + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Remarque sur l'IF de Seconde

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, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}\,,\, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$. Pourquoi ?
- Le programme de Terminale apportera un élément de réponse.
- On a donc $\mathbb{P}\left(0,55 \in \left[F_{30} - \frac{1}{\sqrt{30}}\,,\, F_{30} + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Remarque sur l'IF de Seconde

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, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}\,,\, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$. Pourquoi ?
- Le programme de Terminale apportera un élément de réponse.
- On a donc $\mathbb{P}\left(0,55 \in \left[F_{30} - \frac{1}{\sqrt{30}}\,,\, F_{30} + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Remarque sur l'IF de Seconde

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, $\mathbb{P}\left(F_{30} \in \left[0.55 - \frac{1}{\sqrt{30}}, 0.55 + \frac{1}{\sqrt{30}}\right]\right) < 0,95$. Pourquoi ?
- Le programme de Terminale apportera un élément de réponse.
- On a donc $\mathbb{P}\left(0,55 \in \left[F_{30} - \frac{1}{\sqrt{30}}, F_{30} + \frac{1}{\sqrt{30}}\right]\right) < 0,95$.

Probabilité de recouvrement de IC_2

- La probabilité $\mathcal{C}_2(n,p) = \mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}\,,\, F_n + \frac{1}{\sqrt{n}}\right]\right)$ est dite **de recouvrement** de IC_2.
- Notons que $\mathcal{C}_2(n,p) = \mathbb{P}\left(X_n \in \left[np - \sqrt{n}, np + \sqrt{n}\right]\right)$, d'où :

$$\mathcal{C}_2(n,p) = \sum_{k \in \mathbb{N} \cap \left[np - \sqrt{n}, np + \sqrt{n}\right]} \binom{n}{k} p^k (1-p)^{n-k} .$$

- Variations de $\mathcal{C}_2(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_2

- La probabilité $\mathcal{C}_2(n,p) = \mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}\,,\, F_n + \frac{1}{\sqrt{n}}\right]\right)$ est dite **de recouvrement** de IC_2.
- Notons que $\mathcal{C}_2(n,p) = \mathbb{P}\left(X_n \in \left[np - \sqrt{n}, np + \sqrt{n}\right]\right)$, d'où :

$$\mathcal{C}_2(n,p) = \sum_{k \in \mathbb{N} \cap \left[np - \sqrt{n}, np + \sqrt{n}\right]} \binom{n}{k} p^k (1-p)^{n-k} \,.$$

- Variations de $\mathcal{C}_2(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_2

- La probabilité $\mathcal{C}_2(n,p) = \mathbb{P}\left(p \in \left[F_n - \frac{1}{\sqrt{n}}\,,\, F_n + \frac{1}{\sqrt{n}}\right]\right)$ est dite **de recouvrement** de IC_2.
- Notons que $\mathcal{C}_2(n,p) = \mathbb{P}\left(X_n \in \left[np - \sqrt{n}, np + \sqrt{n}\right]\right)$, d'où :

$$\mathcal{C}_2(n,p) = \sum_{k \in \mathbb{N} \cap \left[np - \sqrt{n}, np + \sqrt{n}\right]} \binom{n}{k} p^k (1-p)^{n-k} \ .$$

- Variations de $\mathcal{C}_2(n,p)$ en fonction de n et p : Graphiques .

IF et IC : Terminale

Heuristique de l'IF asymptotique standard

u_α : le 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 : $u_{0,05} \simeq 1,96$ et $u_{0,01} \simeq 2,58$.
- On admettra que *l'intervalle* $[-u_\alpha, u_\alpha]$ *est le plus petit intervalle (au sens de la longueur) des intervalles* $[a,b]$*, où a et b avec $a \leq b$ sont des réels, tels que* $\mathbb{P}(a \leq Z \leq b) = 1 - \alpha$.

u_α : le 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 : $u_{0,05} \simeq 1,96$ et $u_{0,01} \simeq 2,58$.
- On admettra que *l'intervalle $[-u_\alpha, u_\alpha]$ est le plus petit intervalle (au sens de la longueur) des intervalles $[a,b]$, où a et b avec $a \leq b$ sont des réels, tels que $\mathbb{P}(a \leq Z \leq b) = 1 - \alpha$.*

u_α : le 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 : $u_{0,05} \simeq 1,96$ et $u_{0,01} \simeq 2,58$.
- On admettra que *l'intervalle $[-u_\alpha, u_\alpha]$ est le plus petit intervalle (au sens de la longueur) des intervalles $[a,b]$, où a et b avec $a \leq b$ sont des réels, tels que $\mathbb{P}(a \leq Z \leq b) = 1 - \alpha$.*

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 $\mathbb{P}(a \leq Z_n \leq b) \simeq \mathbb{P}(a \leq Z \leq b)$.

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 $\mathbb{P}(a \leq Z_n \leq b) \simeq \mathbb{P}(a \leq Z \leq b)$.

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 $\mathbb{P}(a \leq Z_n \leq b) \simeq \mathbb{P}(a \leq Z \leq b)$.

Rappel de Première

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

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

- L'intervalle $[f_1\,,f_2]$ peut alors s'écrire sous la forme $[p-e\,,p+e]$.

Rappel de Première

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

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

- L'intervalle $[f_1\,,f_2]$ peut alors s'écrire sous la forme $[p-e\,,p+e]$.

Rappel de Première

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

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

- L'intervalle $[f_1\,, f_2]$ peut alors s'écrire sous la forme $[p-e\,, p+e]$.

Heuristique de l'IF asymptotique standard

- 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 standard

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

Heuristique de l'IF asymptotique standard

- 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 standard

- C'est-à-dire

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

- Pour $\alpha \in]0,1[$ donné, prenons n assez grand de sorte que $\mathbb{P}(-u_\alpha \le Z_n \le u_\alpha) \simeq \mathbb{P}(-u_\alpha \le Z \le u_\alpha) = 1-\alpha$ où Z est une variable aléatoire normale centrée-réduite.
- Pour n assez grand, la condition souhaitée, $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$, peut s'écrire :

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

Heuristique de l'IF asymptotique standard

- 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 $\alpha \in]0,1[$ 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.
- Pour n assez grand, la condition souhaitée, $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$, 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 standard

- 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 $\alpha \in]0,1[$ 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.
- Pour n assez grand, la condition souhaitée, $\mathbb{P}(F_n \in [p-e\,,p+e]) \simeq 1-\alpha$, 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 standard

- 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 standard

- 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 standard

- 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 standard

Théorème de l'IF asymptotique standard

- **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** $1-\alpha$.

Théorème de l'IF asymptotique standard

- **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 pourra faire 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.

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.

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.

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.

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 II 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.

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 : $IF_{A,n} = \left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}, p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$.
- La fonction, $x \in [0,1] \mapsto f(x) = x(x-1)$, 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. □

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, $x \in [0,1] \mapsto f(x) = x(x-1)$, 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. □

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, $x \in [0,1] \mapsto f(x) = x(x-1)$, 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, $x \in [0,1] \mapsto f(x) = x(x-1)$, 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. $\square$

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 $0,95$.
- Comme $\lim\limits_{n \to +\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, et 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 $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ (voir contre-exemple plus haut).

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 $0,95$.
- Comme $\lim\limits_{n \to +\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, et 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 $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\,,\, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ (voir contre-exemple plus haut).

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 $0,95$.
- Comme $\lim_{n \to +\infty} \mathbb{P}(F_n \in IF_{A,n}) = 0,95$, et 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 $\mathbb{P}\left(F_n \in \left[p - \frac{1}{\sqrt{n}}\, ,\, p + \frac{1}{\sqrt{n}}\right]\right) \geq 0,95$ (voir contre-exemple plus haut).

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

- **Rappel** : Au seuil 95%, $\lim_{n\to\infty} \mathbb{P}\left(F_n \in \left[p - u_{0,05}\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_{0,05}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = 0,95\,.$
- **Proposition** : *Pour tout* $p \in]0,1[$,

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

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

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

- **Rappel** : Au seuil 95%, $\lim_{n\to\infty} \mathbb{P}\left(F_n \in \left[p - u_{0,05}\frac{\sqrt{p(1-p)}}{\sqrt{n}}, p + u_{0,05}\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right) = 0,95\,.$
- **Proposition** : *Pour tout* $p \in]0,1[$,

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

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

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

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

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

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

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

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

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

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

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

- 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

- Autre écriture :

$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}}\right]\right) = \mathbb{P}(-2\le Z\le 2) \simeq 0,954\,,$$

où Z est une v.a. $\mathcal{N}(0,1)$.

- ♠ 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* $n\ge n_0$,
$$\mathbb{P}\left(p\in\left[F_n-\frac{1}{\sqrt{n}},F_n+\frac{1}{\sqrt{n}}\right]\right)\ge 0,95.$$
- Représentation graphique de $p\mapsto\lim_{n\to\infty}\mathcal{C}_2(n,p)$: Voir le graphique

Conséquences

- Autre écriture :

$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}}\right]\right) = \mathbb{P}(-2\leq Z\leq 2) \simeq 0,954\,,$$

où Z est une v.a. $\mathcal{N}(0,1)$.

♠ 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

Conséquences

- Autre écriture :

$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}}\right]\right) = \mathbb{P}(-2\leq Z\leq 2) \simeq 0,954\,,$$

 où Z est une v.a. $\mathcal{N}(0,1)$.

- 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

Conséquences

- Autre écriture :

$$\begin{aligned}\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}},p+\frac{1}{\sqrt{n}}\right]\right) &= \mathbb{P}(-2\le Z\le 2)\\ &\simeq 0,954\,,\end{aligned}$$

où Z est une v.a. $\mathcal{N}(0,1)$.

♠ 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* $n\ge n_0$,
$$\mathbb{P}\left(p\in\left[F_n-\frac{1}{\sqrt{n}},F_n+\frac{1}{\sqrt{n}}\right]\right)\ge 0,95.$$
- Représentation graphique de $p\mapsto\lim_{n\to\infty}\mathcal{C}_2(n,p)$: Voir le graphique

Inversion de l'IF asymptotique standard

Inversion de l'IF asymptotique standard

- 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 standard

- 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 $\lim_{n \to +\infty} \mathbb{P}\left(p \in [P_1(n, F_n), P_2(n, F_n)]\right) = \lim_{n \to +\infty} \mathbb{P}\left(F_n \in I_n\right) = 1 - \alpha$.

Inversion de l'IF asymptotique standard

- 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 standard

On peut écrire :

$$F \in I_n \Leftrightarrow |F - p| \leq u \frac{\sqrt{p(1-p)}}{\sqrt{n}}$$

$$\Leftrightarrow \quad (F-p)^2 \leq u^2 \frac{p(1-p)}{n}$$

$$\Leftrightarrow \quad F^2 + p^2 \left(1 + \frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} \leq 0$$

$$\Leftrightarrow \quad p^2 \left(1 + \frac{u^2}{n}\right) - 2p \left(F + \frac{u^2}{2n}\right) + F^2 \leq 0.$$

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 $$F^2 + p^2\left(1+\frac{u^2}{n}\right) - 2pF - \frac{u^2 p}{n} = 0$$ 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

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"* : $p - u\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq 0 \Rightarrow np \leq (1-p)u^2 \leq u^2$. Au niveau de confiance 95%, $u < 2$, d'où $np \leq 4 < 5$.
  - *"Partie droite"* : $p + u\frac{\sqrt{p(1-p)}}{\sqrt{n}} \geq 1 \Rightarrow n(1-p) \leq pu^2 \leq u^2$. Au niveau de confiance 95%, $u < 2$, d'où $n(1-p) \leq 4 < 5$.

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 $n(1-p) \leq 4 < 5$.

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 $n(1-p) \leq 4 < 5$.

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 :

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

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

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

Probabilité de recouvrement de IC_W

- La probabilité de recouvrement de IC_W, $\mathcal{C}_W(n,p) = \mathbb{P}\left(p \in [P_1, P_2]\right)$, s'écrit aussi :

$$\mathcal{C}_W(n,p) = \mathbb{P}\left(X_n \in \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right)$$

- ou encore : $\mathcal{C}_W(n,p) = \sum_{k \in \mathcal{N}(n,p)} \binom{n}{k} p^k (1-p)^{n-k}$, où

$$\mathcal{N}(n,p) = \mathbb{N} \cap \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

- Variations de $\mathcal{C}_W(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_W

- La probabilité de recouvrement de IC_W, $\mathcal{C}_W(n,p) = \mathbb{P}\left(p \in [P_1, P_2]\right)$, s'écrit aussi :

$$\mathcal{C}_W(n,p) = \mathbb{P}\left(X_n \in \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right)$$

- ou encore : $\mathcal{C}_W(n,p) = \sum_{k \in \mathcal{N}(n,p)} \binom{n}{k} p^k (1-p)^{n-k}$, où

$$\mathcal{N}(n,p) = \mathbb{N} \cap \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

- Variations de $\mathcal{C}_W(n,p)$ en fonction de n et p : Graphiques .

Probabilité de recouvrement de IC_W

- La probabilité de recouvrement de IC_W, $\mathcal{C}_W(n,p) = \mathbb{P}\left(p \in [P_1, P_2]\right)$, s'écrit aussi :

$$\mathcal{C}_W(n,p) = \mathbb{P}\left(X_n \in \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]\right)$$

- ou encore : $\mathcal{C}_W(n,p) = \displaystyle\sum_{k \in \mathcal{N}(n,p)} \binom{n}{k} p^k (1-p)^{n-k}$, où

$$\mathcal{N}(n,p) = \mathbb{N} \cap \left[np - nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}, np + nu_\alpha \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right].$$

- Variations de $\mathcal{C}_W(n,p)$ en fonction de n et p : Graphiques .

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 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 $\frac{1}{n}$ dans les expressions de P_1 et P_2, d'où :
$P_1 \simeq F_n - u_\alpha \dfrac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}$ et $P_2 \simeq F_n + u_\alpha \dfrac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}$.
- L'intervalle aléatoire $\left[F_n - u_\alpha \dfrac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}, F_n + u_\alpha \dfrac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right]$, noté IC_S est appelé **intervalle de confiance standard** ou **de Wald-Laplace** (1812).

L'IC 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.

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.

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.

Conditions de validité pour IC_2

- Rappel :
$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}}\,,\,p+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,954\simeq 0,95\,;$$
- d'où, grossièrement, sous les conditions de grandes binomiales,
$$\mathbb{P}\left(p\in\left[F_n-\frac{1}{\sqrt{n}}\,,\,F_n+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,95.$$
- En pratique : on acceptera cette approximation pour les échantillons de taille $n\geq 30$ dont la fréquence observée f vérifie $nf\geq 5$ et $n(1-f)\geq 5$.

Conditions de validité pour IC_2

- Rappel :
$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}}\,,\,p+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,954\simeq 0,95\,;$$
- d'où, grossièrement, sous les conditions de grandes binomiales,
$$\mathbb{P}\left(p\in\left[F_n-\frac{1}{\sqrt{n}}\,,\,F_n+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,95.$$
- En pratique : on acceptera cette approximation pour les échantillons de taille $n\geq 30$ dont la fréquence observée f vérifie $nf\geq 5$ et $n(1-f)\geq 5$.

Conditions de validité pour IC_2

- Rappel :
$$\inf_{p\in]0,1[}\lim_{n\to\infty}\mathbb{P}\left(F_n\in\left[p-\frac{1}{\sqrt{n}}\,,\,p+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,954\simeq 0,95\,;$$
- d'où, grossièrement, sous les conditions de grandes binomiales,
$$\mathbb{P}\left(p\in\left[F_n-\frac{1}{\sqrt{n}}\,,\,F_n+\frac{1}{\sqrt{n}}\right]\right)\simeq 0,95.$$
- En pratique : on acceptera cette approximation pour les échantillons de taille $n\geq 30$ dont la fréquence observée f vérifie $nf\geq 5$ et $n(1-f)\geq 5$.

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 : $IC_S = [0,479015\ ;\ 0,540985]$.

Comparaison : abaques IC standard, exact et Seconde

- Abaques pour les IF de Seconde, exact et standard :
  Voir les graphiques

Probabilité de recouvrement de l'IC standard

- Posons $\mathcal{C}_S(n,p) =$
$$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right).$$
- Notons que $\mathcal{C}_S(n,p) = \sum_{k\in\mathcal{N}_S} \binom{n}{k} p^k(1-p)^{n-k}$ où
$$\mathcal{N}_S = \left\{k \in \mathbb{N} \;/\; p \in \left[\frac{k}{n} - u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}} \;,\; \frac{k}{n} + u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}}\right]\right\}.$$
- Le couple (n,p) est dit **heureux pour un IC au seuil** $1-\alpha$ si $\mathcal{C}(n,p)$ est très proche ou supérieure de $1-\alpha$. Dans le cas contraire, il est dit **malheureux**.

Probabilité de recouvrement de l'IC standard

- Posons $\mathcal{C}_S(n,p) =$
$$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right).$$
- Notons que $\mathcal{C}_S(n,p) = \sum_{k \in \mathcal{N}_S} \binom{n}{k} p^k (1-p)^{n-k}$ où
$$\mathcal{N}_S = \left\{ k \in \mathbb{N} \;/\; p \in \left[\frac{k}{n} - u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}} \;,\; \frac{k}{n} + u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}} \right] \right\}.$$
- Le couple (n,p) est dit **heureux pour un IC au seuil** $1-\alpha$ si $\mathcal{C}(n,p)$ est très proche ou supérieure de $1-\alpha$. Dans le cas contraire, il est dit **malheureux**.

Probabilité de recouvrement de l'IC standard

- Posons $\mathcal{C}_S(n,p) =$
$$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right).$$
- Notons que $\mathcal{C}_S(n,p) = \sum_{k \in \mathcal{N}_S} \binom{n}{k} p^k (1-p)^{n-k}$ où
$$\mathcal{N}_S = \left\{ k \in \mathbb{N} \;/\; p \in \left[\frac{k}{n} - u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}} \,,\, \frac{k}{n} + u_\alpha \frac{\sqrt{\frac{k}{n}(1-\frac{k}{n})}}{\sqrt{n}} \right] \right\}.$$
- Le couple (n,p) est dit **heureux pour un IC au seuil** $1-\alpha$ si $\mathcal{C}(n,p)$ est très proche ou supérieure de $1-\alpha$. Dans le cas contraire, il est dit **malheureux**.

Représentations graphiques de $\mathcal{C}_S(n,p)$

- Variations de $\mathcal{C}_S(n,p)$ en fonction de n et p : Graphiques .
- Graphiques de $\mathcal{C}_S(n,p)$ en fonction de n et p : Voir les graphiques .

Représentations graphiques de $\mathcal{C}_S(n, p)$

- Variations de $\mathcal{C}_S(n, p)$ en fonction de n et p : Graphiques .
- Graphiques de $\mathcal{C}_S(n, p)$ en fonction de n et p : Voir les graphiques .

Approche directe de l'IC standard

- 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

- Considérons, pour tout entier $n \geq 1$, la variable aléatoire $$T_n = \frac{F_n - p}{\sqrt{\dfrac{F_n(1-F_n)}{n}}}.$$
- **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

- Considérons, pour tout entier $n \geq 1$, la variable aléatoire
$$T_n = \frac{F_n - p}{\sqrt{\dfrac{F_n(1 - F_n)}{n}}}.$$
- **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

- Considérons, pour tout entier $n \geq 1$, la variable aléatoire
$$T_n = \frac{F_n - p}{\sqrt{\dfrac{F_n(1-F_n)}{n}}}.$$
- **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
$$\mathbb{P}\left(F_n - u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}} \leq p \leq F_n + u_\alpha \frac{\sqrt{F_n(1-F_n)}}{\sqrt{n}}\right) \simeq 1-\alpha\,.$$

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

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

Commentaire sur la construction des IC et IF

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

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

Commentaire sur la construction des IC et IF

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

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

Indicateurs de la qualité statistique d'un IC

Pour étudier la qualité statistique de l'IC $[P_1(n, F_n)\,,\ P_2(n, F_n)]$, on s'intéressera naturellement à trois types d'indicateurs :

- sa **probabilité de recouvrement**, i.e. la probabilité $\mathcal{C}(n,p) = \mathbb{P}\left(p \in [P_1(n, F_n)\,,\ P_2(n, F_n)]\right)$;
- sa **longueur espérée**, i.e. l'espérance $\mathcal{L}(n,p) = \mathbb{E}[P_2(n, F_n) - P_1(n, F_n)]$, et sa **longueur espérée moyenne** $\mathcal{M}(n) = \int_0^1 \mathcal{L}(n,p)dp$.
- ses **conditions de validité**, i.e. les conditions de l'observation rendant l'application de l'IC pertinente en statistique.

Indicateurs de la qualité statistique d'un IC

Pour étudier la qualité statistique de l'IC $[P_1(n, F_n)\, ,\, P_2(n, F_n)]$, on s'intéressera naturellement à trois types d'indicateurs :

- sa **probabilité de recouvrement**, i.e. la probabilité $\mathcal{C}(n, p) = \mathbb{P}\left(p \in [P_1(n, F_n)\, ,\, P_2(n, F_n)]\right)$;
- sa **longueur espérée**, i.e. l'espérance $\mathcal{L}(n, p) = \mathbb{E}[P_2(n, F_n) - P_1(n, F_n)]$, et sa **longueur espérée moyenne** $\mathcal{M}(n) = \int_0^1 \mathcal{L}(n, p)dp$.
- ses **conditions de validité**, i.e. les conditions de l'observation rendant l'application de l'IC pertinente en statistique.

Indicateurs de la qualité statistique d'un IC

Pour étudier la qualité statistique de l'IC $[P_1(n, F_n)\,,\ P_2(n, F_n)]$, on s'intéressera naturellement à trois types d'indicateurs :

- sa **probabilité de recouvrement**, i.e. la probabilité $\mathcal{C}(n, p) = \mathbb{P}\left(p \in [P_1(n, F_n)\,,\ P_2(n, F_n)]\right)$;
- sa **longueur espérée**, i.e. l'espérance $\mathcal{L}(n, p) = \mathbb{E}[P_2(n, F_n) - P_1(n, F_n)]$, et sa **longueur espérée moyenne** $\mathcal{M}(n) = \int_0^1 \mathcal{L}(n, p)dp$.
- ses **conditions de validité**, i.e. les conditions de l'observation rendant l'application de l'IC pertinente en statistique.

Probabilité de recouvrement

- On a $\mathcal{C}(n,p) = \sum_{k \in \mathcal{N}} \binom{n}{k} p^k (1-p)^{n-k}$ où $\mathcal{N} = \left\{k \in \mathbb{N} \ / \ p \in \left[P_1(n, \frac{k}{n}) \, , \, P_2(n, \frac{k}{n})\right]\right\}$.
- Les expressions de ces probabilités en fonction de n et p sont très compliquées.
- Graphiques de $\mathcal{C}(n,p)$ pour différents IC : Voir les graphiques .

Probabilité de recouvrement

- On a $\mathcal{C}(n,p) = \sum_{k \in \mathcal{N}} \binom{n}{k} p^k (1-p)^{n-k}$ où $\mathcal{N} = \left\{k \in \mathbb{N} \ / \ p \in \left[P_1(n, \frac{k}{n}) \, , \, P_2(n, \frac{k}{n})\right]\right\}$.
- Les expressions de ces probabilités en fonction de n et p sont très compliquées.
- Graphiques de $\mathcal{C}(n,p)$ pour différents IC : Voir les graphiques .

Probabilité de recouvrement

- On a $\mathcal{C}(n,p) = \sum_{k \in \mathcal{N}} \binom{n}{k} p^k (1-p)^{n-k}$ où $\mathcal{N} = \left\{k \in \mathbb{N} \;/\; p \in \left[P_1(n, \frac{k}{n}),\, P_2(n, \frac{k}{n})\right]\right\}$.
- Les expressions de ces probabilités en fonction de n et p sont très compliquées.
- Graphiques de $\mathcal{C}(n,p)$ pour différents IC : Voir les graphiques .

Conditions de validité pour IC_S

- **Théorème** : *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 $M_{\gamma,\alpha} = \mathbb{P}\left(a_{\gamma,\alpha} < V \leq b_{\gamma,\alpha}\right)$ *avec :*
** V v.a. de Poisson de paramètre γ,*
** $a_{\gamma,\alpha}$ la partie entière de $\frac{1}{2}(u_\alpha^2 + 2\gamma - u_\alpha\sqrt{u_\alpha^2 + 4\gamma})$,*
** $b_{\gamma,\alpha}$ la partie entière de $\frac{1}{2}(u_\alpha^2 + 2\gamma + u_\alpha\sqrt{u_\alpha^2 + 4\gamma})$.*

- 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 .

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** (1998) 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.

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.
- **Proposition** : *Pour tous n et p, on a $IC_W \subseteq IC_{AC}$*, d'où $\mathcal{L}_W(n,p) \leq \mathcal{L}_{AC}(n,p)$;
- 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 : $IC_S = [0,479015 \ ; \ 0,540985]$.

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. (Brown-Cai-Dasgupta, 2001)

Conclusion

Aspects mathématiques des IF et IC

- Techniques mathématiques mises en jeu (limites de suites, approximations, calculs algébriques, représentations graphiques, calculs de probabilités, ...)
- Nombreuses définitions et des modalités d'approche très diverses pour construire un IC.
- Relation d'inversion entre IF et IC.

Aspects statistiques des IF et IC

- Il est nécessaire d'avoir des IC pour des valeurs de p extrêmement petites.
- La valeur de n est souvent imposée par le contexte réel.
- Au-delà des aspects mathématiques de l'IC, c'est son intérêt du point de vue des applications statistiques qui guide la construction des IC.
- Les IC font l'objet d'ajustements pratiques, ou de corrections de continuité, pour les améliorer.
- L'appréciation de la qualité d'un IC est complexe : nombreux indicateurs.
- Les conditions des grandes binomiales sont indicatives.

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)

- **Téléphones** : (YD) +33(0)3 81 66 62 32 / (BS) +33(0)3 81 66 63 00
  Adresses électroniques :

  yves.ducel@univ-fcomte.fr
  bruno.saussereau@univ-fcomte.fr

- **Adresse postale** : IREM - Département de mathématiques
  UFR Sciences et techniques de l'Université de Franche-Comté
  16, route de Gray, F-25030 Besançon cedex

Sites web :

- **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