

# Echantillonnage et estimation

## Terminale ES

X. OUVRARD

Enseignant au Lycée International de Ferney-Voltaire

# Plan du cours

1 Echantillonnage

2 Estimation

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.  
On effectue  $n$  tirages avec remise.  
Deux situations bien différentes :

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.  
On effectue  $n$  tirages avec remise.  
Deux situations bien différentes :

Dans l'urne  $U_1$  :

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.  
On effectue  $n$  tirages avec remise.  
Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

★ Proportion  $p$  de boules  
rouges connue.

★ Observation de la  
fréquence d'apparition de la  
boule rouge,  $F_n$ .

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

★ Proportion  $p$  de boules  
rouges connue.

★ Observation de la  
fréquence d'apparition de la  
boule rouge,  $F_n$ .

★  $F_n$  appartient en général à  
un intervalle de fluctuation de  
centre  $p$ , dont la précision  
augmente avec  $n$ .

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

★ Proportion  $p$  de boules  
rouges connue.

★ Observation de la  
fréquence d'apparition de la  
boule rouge,  $F_n$ .

★  $F_n$  appartient en général à  
un intervalle de fluctuation de  
centre  $p$ , dont la précision  
augmente avec  $n$ .

★ Domaine de  
l'échantillonnage et de  
l'intervalle de fluctuation



Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

★ Proportion  $p$  de boules  
rouges connue.

★ Observation de la  
fréquence d'apparition de la  
boule rouge,  $F_n$ .

★  $F_n$  appartient en général à  
un intervalle de fluctuation de  
centre  $p$ , dont la précision  
augmente avec  $n$ .

★ Domaine de  
l'échantillonnage et de  
l'intervalle de fluctuation  
=> paragraphe 1

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.
- ★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .
- ★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .
- ★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

★ Proportion  $p$  de boules rouges connue.

★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .

★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .

★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

◆ Proportion de boules rouges inconnue a priori.

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.
- ★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .
- ★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .
- ★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

- ◆ Proportion de boules rouges inconnue a priori.
- ◆ Tentative d'estimation à partir de la fréquence observée de la proportion de boules rouges.

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.

On effectue  $n$  tirages avec remise.

Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.
- ★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .
- ★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .
- ★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

- ◆ Proportion de boules rouges inconnue a priori.
- ◆ Tentative d'estimation à partir de la fréquence observée de la proportion de boules rouge.
- ◆ Estimation faite à partir d'un intervalle de confiance.

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.  
On effectue  $n$  tirages avec remise.  
Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.
- ★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .
- ★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .
- ★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

- ◆ Proportion de boules rouges inconnue a priori.
- ◆ Tentative d'estimation à partir de la fréquence observée de la proportion de boules rouge.
- ◆ Estimation faite à partir d'un intervalle de confiance.
- ◆ Domaine de l'estimation et de l'intervalle de confiance

Soit deux urnes  $U_1$  et  $U_2$ , contenant des boules rouges et noires.  
On effectue  $n$  tirages avec remise.  
Deux situations bien différentes :

Dans l'urne  $U_1$  :

- ★ Proportion  $p$  de boules rouges connue.
- ★ Observation de la fréquence d'apparition de la boule rouge,  $F_n$ .
- ★  $F_n$  appartient en général à un intervalle de fluctuation de centre  $p$ , dont la précision augmente avec  $n$ .
- ★ Domaine de l'échantillonnage et de l'intervalle de fluctuation  
=> paragraphe 1

Dans l'urne  $U_2$  :

- ◆ Proportion de boules rouges inconnue a priori.
- ◆ Tentative d'estimation à partir de la fréquence observée de la proportion de boules rouge.
- ◆ Estimation faite à partir d'un intervalle de confiance.
- ◆ Domaine de l'estimation et de l'intervalle de confiance  
=> paragraphe 2

# Plan du cours

1 Echantillonnage

2 Estimation



Soit  $X_n$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$ .

Soit  $X_n$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$ .  
Pour  $n$  grand, on a vu que l'on peut admettre que  $X_n$  soit une loi normale  $\mathcal{N}(\mu, \sigma^2)$  d'espérance  $\mu = np$  et d'écart type  $\sigma = \sqrt{np(1-p)}$ .

Soit  $X_n$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$ .  
Pour  $n$  grand, on a vu que l'on peut admettre que  $X_n$  soit une loi normale  $\mathcal{N}(\mu, \sigma^2)$  d'espérance  $\mu = np$  et d'écart type  $\sigma = \sqrt{np(1-p)}$ .

On a vu que pour  $Z_n = \frac{X_n - \mu}{\sigma}$  suivant la loi normale centrée réduite :  $P(-1,96 \leq Z_n \leq 1,96) \approx 0,95$ .

Soit  $X_n$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$ .  
Pour  $n$  grand, on a vu que l'on peut admettre que  $X_n$  soit une loi normale  $\mathcal{N}(\mu, \sigma^2)$  d'espérance  $\mu = np$  et d'écart type  $\sigma = \sqrt{np(1-p)}$ .

On a vu que pour  $Z_n = \frac{X_n - \mu}{\sigma}$  suivant la loi normale centrée réduite :  $P(-1,96 \leq Z_n \leq 1,96) \approx 0,95$ .

Ce qui revient à  $P(-1,96 \leq \frac{X_n - \mu}{\sigma} \leq 1,96) \approx 0,95$ ,

Soit  $X_n$  une variable aléatoire suivant une loi binomiale  $\mathcal{B}(n, p)$ .  
Pour  $n$  grand, on a vu que l'on peut admettre que  $X_n$  soit une loi normale  $\mathcal{N}(\mu, \sigma^2)$  d'espérance  $\mu = np$  et d'écart type  $\sigma = \sqrt{np(1-p)}$ .

On a vu que pour  $Z_n = \frac{X_n - \mu}{\sigma}$  suivant la loi normale centrée réduite :  $P(-1,96 \leq Z_n \leq 1,96) \approx 0,95$ .

Ce qui revient à  $P(-1,96 \leq \frac{X_n - \mu}{\sigma} \leq 1,96) \approx 0,95$ ,  
ou encore à :  $P(\mu - 1,96\sigma \leq X_n \leq \mu + 1,96\sigma) \approx 0,95$

On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

$$\text{Alors } P\left(\frac{\mu - 1,96\sigma}{n} \leq F_n \leq \frac{\mu + 1,96\sigma}{n}\right) \approx 0,95.$$

On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

$$\text{Alors } P\left(\frac{\mu - 1,96\sigma}{n} \leq F_n \leq \frac{\mu + 1,96\sigma}{n}\right) \approx 0,95.$$

$$\text{Or } \frac{\mu}{n} = p \text{ et } \frac{\sigma}{n} = \sqrt{\frac{p(1-p)}{n}}$$



On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

Alors  $P\left(\frac{\mu - 1,96\sigma}{n} \leq F_n \leq \frac{\mu + 1,96\sigma}{n}\right) \approx 0,95$ .

Or  $\frac{\mu}{n} = p$  et  $\frac{\sigma}{n} = \sqrt{\frac{p(1-p)}{n}}$

On note  $I_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}} ; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right]$ .

On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

$$\text{Alors } P\left(\frac{\mu - 1,96\sigma}{n} \leq F_n \leq \frac{\mu + 1,96\sigma}{n}\right) \approx 0,95.$$

$$\text{Or } \frac{\mu}{n} = p \text{ et } \frac{\sigma}{n} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{On note } I_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}} ; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right].$$

Alors, pour  $n$  grand  $P(F_n \in I_n)$  est voisine de 0,95.

On pose  $F_n = \frac{X_n}{n}$  la variable aléatoire correspondant à la fréquence de succès.

$$\text{Alors } P\left(\frac{\mu - 1,96\sigma}{n} \leq F_n \leq \frac{\mu + 1,96\sigma}{n}\right) \approx 0,95.$$

$$\text{Or } \frac{\mu}{n} = p \text{ et } \frac{\sigma}{n} = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{On note } I_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}} ; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right].$$

Alors, pour  $n$  grand  $P(F_n \in I_n)$  est voisine de 0,95.

### Définition

$I_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}} ; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right]$  est appelé un intervalle de fluctuation asymptotique de  $F_n$  avec la probabilité 0,95 (ou au seuil de 95%)

## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Soit  $F_n$  la fréquence dans un échantillon de taille  $n$ .*

*Si  $I_n$  est l'intervalle de fluctuation asymptotique au seuil de 95 %, alors la règle de décision est la suivante :*

- si  $F_n \in I_n$  : on accepte l'hypothèse au seuil de risque de 5%.*
- si  $F_n \notin I_n$  : on rejette l'hypothèse selon laquelle la proportion vaut  $p$  au seuil de risque de 5 %.*

## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Prise de décision à partir d'un échantillon :*

*On considère une population dans laquelle on émet l'hypothèse qu'un caractère a la proportion  $p$ .*

## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Prise de décision à partir d'un échantillon :*

*On considère une population dans laquelle on émet l'hypothèse qu'un caractère a la proportion  $p$ .*

*Soit  $F_n$  la fréquence dans un échantillon de taille  $n$ .*

## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Prise de décision à partir d'un échantillon :*

*On considère une population dans laquelle on émet l'hypothèse qu'un caractère a la proportion  $p$ .*

*Soit  $F_n$  la fréquence dans un échantillon de taille  $n$ .*

*Si  $I_n$  est l'intervalle de fluctuation asymptotique au seuil de 95 %, alors la règle de décision est la suivante :*

## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Prise de décision à partir d'un échantillon :*

*On considère une population dans laquelle on émet l'hypothèse qu'un caractère a la proportion  $p$ .*

*Soit  $F_n$  la fréquence dans un échantillon de taille  $n$ .*

*Si  $I_n$  est l'intervalle de fluctuation asymptotique au seuil de 95 %, alors la règle de décision est la suivante :*

*- si  $F_n \in I_n$  : on accepte l'hypothèse au seuil de risque de 5%.*



## Remarque

*Cette approximation est valable dès que  $n \geq 30$ ,  $np \geq 5$  et  $n(1 - p) \geq 5$ .*

*Prise de décision à partir d'un échantillon :*

*On considère une population dans laquelle on émet l'hypothèse qu'un caractère a la proportion  $p$ .*

*Soit  $F_n$  la fréquence dans un échantillon de taille  $n$ .*

*Si  $I_n$  est l'intervalle de fluctuation asymptotique au seuil de 95 %, alors la règle de décision est la suivante :*

- si  $F_n \in I_n$  : on accepte l'hypothèse au seuil de risque de 5%.*
- si  $F_n \notin I_n$  : on rejette l'hypothèse selon laquelle la proportion vaut  $p$  au seuil de risque de 5 %.*

# Plan du cours

① Echantillonnage

② Estimation

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)  
On désire estimer au mieux la probabilité  $p$  à partir de  $n$  expériences indépendantes.

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)

On désire estimer au mieux la probabilité  $p$  à partir de  $n$  expériences indépendantes.

Soit  $X_n$  la variable aléatoire comptant le nombre de succès au bout des  $n$  expériences.

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)

On désire estimer au mieux la probabilité  $p$  à partir de  $n$  expériences indépendantes.

Soit  $X_n$  la variable aléatoire comptant le nombre de succès au bout des  $n$  expériences.

Soit  $F_n$  la variable aléatoire donnant la fréquence de succès.

$$F_n = \frac{X_n}{n}.$$

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)

On désire estimer au mieux la probabilité  $p$  à partir de  $n$  expériences indépendantes.

Soit  $X_n$  la variable aléatoire comptant le nombre de succès au bout des  $n$  expériences.

Soit  $F_n$  la variable aléatoire donnant la fréquence de succès.

$$F_n = \frac{X_n}{n}.$$

$F_n$  peut-être considéré comme un estimateur de  $p$ .

On considère une expérience de Bernoulli dont on ne connaît pas la probabilité de succès  $p$  (ni celle de l'échec !)

On désire estimer au mieux la probabilité  $p$  à partir de  $n$  expériences indépendantes.

Soit  $X_n$  la variable aléatoire comptant le nombre de succès au bout des  $n$  expériences.

Soit  $F_n$  la variable aléatoire donnant la fréquence de succès.

$$F_n = \frac{X_n}{n}.$$

$F_n$  peut-être considéré comme un estimateur de  $p$ .

Jusqu'à quel point est-ce fiable ? Quelle est la marge d'erreur ?



## Définition

*On appelle intervalle de confiance au niveau 0,95 pour l'estimation de  $p$  l'intervalle noté  $I_C$  tel que :  $P(p \in I_C) \geq 0,95$*

## Définition

On appelle *intervalle de confiance* au niveau 0,95 pour l'estimation de  $p$  l'intervalle noté  $I_C$  tel que :  $P(p \in I_C) \geq 0,95$

## Propriété

Soit  $X_n$  une variable aléatoire suivant la loi binomiale  $\mathcal{B}(n, p)$ , alors pour tout  $p \in ]0; 1[$  il existe un entier  $n_0$  tel que pour tout  $n \geq n_0$ ,  
$$P\left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}}\right) > 0,95.$$

## Preuve

Montrons que

$$K_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}}; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right] \subset$$
$$\left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] = J_n.$$

## Preuve

Montrons que

$$K_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}}; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right] \subset \left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] = J_n.$$

Cela revient à montrer que  $1,96\sqrt{p(1-p)} \leq 1$  pour  $p \in [0; 1]$ .

Soit  $g : p \mapsto p(1-p)$   $g$  est dérivable sur  $[0; 1]$  et  $g'(p) = 1 - 2p$ .

$g'(p) \geq 0 \Leftrightarrow \frac{1}{2} \geq p$ . Et donc  $g$  est croissante sur  $\left[0; \frac{1}{2}\right]$  et

décroissante sur  $\left[\frac{1}{2}; 1\right]$ . Donc sur  $[0; 1]$ ,  $g$  atteint son maximum

en  $\frac{1}{2}$ .  $g\left(\frac{1}{2}\right) = \frac{1}{4}$ . Et ainsi  $1,96\sqrt{p(1-p)} \leq 1,96\sqrt{\frac{1}{4}} = 0,98$ .

## Preuve

Montrons que

$$K_n = \left[ p - 1,96\sqrt{\frac{p(1-p)}{n}}; p + 1,96\sqrt{\frac{p(1-p)}{n}} \right] \subset \left[ p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right] = J_n.$$

Cela revient à montrer que  $1,96\sqrt{p(1-p)} \leq 1$  pour  $p \in [0; 1]$ .

Soit  $g : p \mapsto p(1-p)$   $g$  est dérivable sur  $[0; 1]$  et  $g'(p) = 1 - 2p$ .

$g'(p) \geq 0 \Leftrightarrow \frac{1}{2} \geq p$ . Et donc  $g$  est croissante sur  $\left[0; \frac{1}{2}\right]$  et

décroissante sur  $\left[\frac{1}{2}; 1\right]$ . Donc sur  $[0; 1]$ ,  $g$  atteint son maximum

en  $\frac{1}{2}$ .  $g\left(\frac{1}{2}\right) = \frac{1}{4}$ . Et ainsi  $1,96\sqrt{p(1-p)} \leq 1,96\sqrt{\frac{1}{4}} = 0,98$ .

Ainsi,  $P(F_n \in K_n) \leq P(F_n \in J_n)$  et donc :  $0,95 \leq P(F_n \in J_n)$

## Propriété

Lorsque  $n$  est assez grand (c.à.d. pour  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ ), l'intervalle  $I_C = \left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$  est un intervalle de confiance au niveau 95% pour l'estimation de  $p$ .  $I_C$  est appelé intervalle de confiance pour  $p$  au niveau asymptotique 95%.

$$\text{Or } p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq F_n - p \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq$$

$$p - F_n \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} + F_n \leq p \leq \frac{1}{\sqrt{n}} + F_n \Leftrightarrow p \in I_C.$$

$$P\left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}}\right) = P(p \in I_C) > 0,95. \quad \square$$

## Propriété

Lorsque  $n$  est assez grand (c.à.d. pour  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ ), l'intervalle  $I_C = \left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$  est un intervalle de confiance au niveau 95% pour l'estimation de  $p$ .  $I_C$  est appelé intervalle de confiance pour  $p$  au niveau asymptotique 95%.

## Preuve

D'après la propriété précédente, il existe  $n_0$  tel que pour  $n \geq n_0$  :

$$P \left( p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \right) > 0,95.$$

## Propriété

Lorsque  $n$  est assez grand (c.à.d. pour  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ ), l'intervalle  $I_C = \left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$  est un intervalle de confiance au niveau 95% pour l'estimation de  $p$ .  $I_C$  est appelé intervalle de confiance pour  $p$  au niveau asymptotique 95%.

## Preuve

D'après la propriété précédente, il existe  $n_0$  tel que pour  $n \geq n_0$  :

$$P \left( p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \right) > 0,95.$$

$$\text{Or } p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq F_n - p \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq$$

$$p - F_n \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} + F_n \leq p \leq \frac{1}{\sqrt{n}} + F_n \Leftrightarrow p \in I_C.$$



## Propriété

Lorsque  $n$  est assez grand (c.à.d. pour  $n \geq 30$ ,  $np \geq 5$  et  $n(1-p) \geq 5$ ), l'intervalle  $I_C = \left[ F_n - \frac{1}{\sqrt{n}}; F_n + \frac{1}{\sqrt{n}} \right]$  est un intervalle de confiance au niveau 95% pour l'estimation de  $p$ .  $I_C$  est appelé intervalle de confiance pour  $p$  au niveau asymptotique 95%.

## Preuve

D'après la propriété précédente, il existe  $n_0$  tel que pour  $n \geq n_0$  :

$$P\left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}}\right) > 0,95.$$

$$\text{Or } p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq F_n - p \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} \leq$$

$$p - F_n \leq \frac{1}{\sqrt{n}} \Leftrightarrow -\frac{1}{\sqrt{n}} + F_n \leq p \leq \frac{1}{\sqrt{n}} + F_n \Leftrightarrow p \in I_C.$$

$$P\left(p - \frac{1}{\sqrt{n}} \leq F_n \leq p + \frac{1}{\sqrt{n}}\right) = P(p \in I_C) > 0,95. \quad \square$$