Intervalle de fluctuation ou intervalle de Confiance : Cas de la proportion

Jeu de Pile ou Face

On joue avec une pièce juste, telle que la proba d'obtenir face soit $p = 1/2$.

Soit X_i la variable indicatrice de succès "obtenir face" au i-ème lancé. X_i suit une loi $\mathcal{B}(p)$ et les X_i sont indépendantes.

Simulation de $N = 100$ échantillons de taille $n = 50$ et calcul des moyennes empiriques de chaque échantillon :

```
n<-50; N<-100; p=0.5 # paramètres de simulation
matrix(rbinom(n*N,1,p),ncol=n)->dataB
# un n-ech. par ligne de dataB
moys<-rowMeans(dataB) # les moyennes de chaque échantillon
length(moys) # N nombre de réalisations de sum(x_i)/n
```

```
## [1] 100
```

Intervalle de fluctuation

$\bar{X}_n$ fluctue autour de son espérance p et avec une probabilité valant approximativement $1-\alpha$ dans l'intervalle :

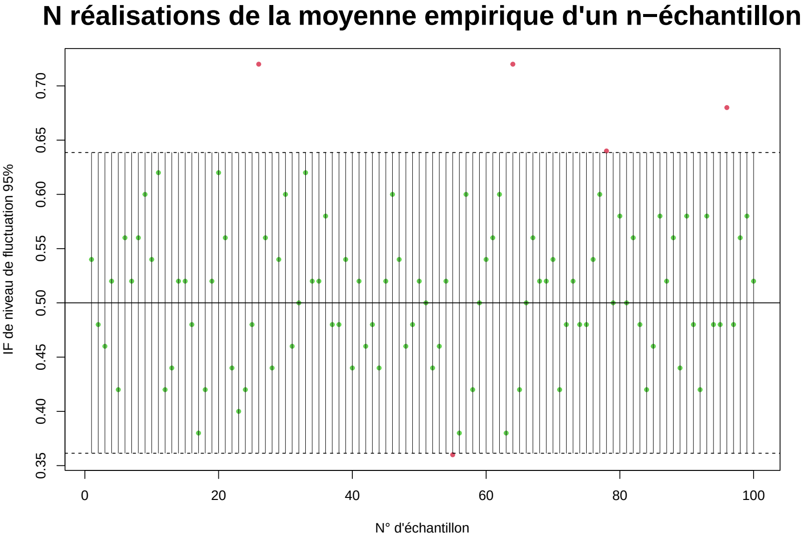
$$IF(p,\alpha) = \left[p - \sqrt{\frac{p(1-p)}{n}} u_{1-\frac{\alpha}{2}}, p + \sqrt{\frac{p(1-p)}{n}} u_{1-\frac{\alpha}{2}} \right]$$

Autrement dit

$$P(\bar{X}_n \in IF(p,\alpha)) \approx 1-\alpha \quad \text{si} \quad np > 10 \text{ et } n(1-p) > 10$$

IF ne dépend pas de l'échantillon. C'est un **intervalle déterministe**.

Illustration avec les N réalisations de $\bar{X}_n$



Environ 5% des échantillons produisent un $\bar{x}_n$ hors de l'IF donc dits non conformes au niveau de fluctuation 95%

Amplitude de l'intervalle de fluctuation

- Pour p et α fixés, l'amplitude $2\sqrt{\frac{p(1-p)}{n}}u_{1-\frac{\alpha}{2}}$ décroit avec n

  On peut trouver une valeur de n minimale n_{min} pour laquelle elle est inférieure à a :

$$2\sqrt{\frac{p(1-p)}{n}}u_{1-\frac{\alpha}{2}} \leq a \Longleftrightarrow n \geq \frac{4p(1-p)u^2_{1-\frac{\alpha}{2}}}{a^2}$$

  En particulier pour le jeu de pile ou face avec une vraie pièce où $p = 0.5$ avec $\alpha = 0.05$ et $a = 0.02$, $n_{min} \geq a^{-2}u^2_{1-\frac{\alpha}{2}} = (1.96)^2(0.25) \cdot 10^4$ soit $n_{min} = 9604$. Pour une amplitude de 0.1 on obtient $n_{min} = 77$ et pour une amplitude de 0.3, $n_{min} = 26$.

- Pour p et n fixés, lorsque α augmente l'amplitude diminue. Autrement dit la précision augmente lorsque le niveau de fluctuation baisse. Et α qui garantit une amplitude $\leq a$ satisfait $\alpha \geq 2(1 - \Phi(a\sqrt{n/(4p(1-p))}))$. Pour $n = 50$ et $a = 0.02$ (précision $\pm 1\%$) on devrait prendre $\alpha \geq 88,8\%$ soit $1 - \alpha \leq 11,2\%$.

Intervalle de confiance

Si n assez grand ($np > 10$ et $n(1-p) > 10$) l'intervalle de confiance pour p de niveau de confiance approximatif $1-\alpha$ est donné par

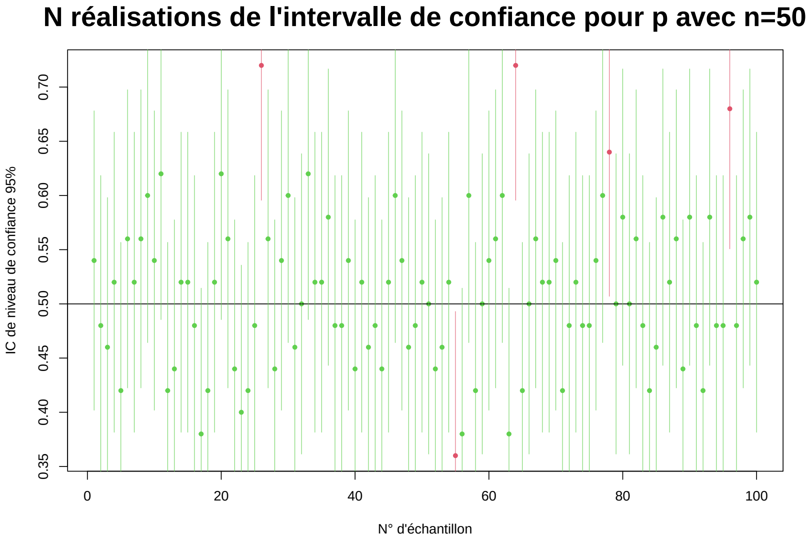
$$IC(F_n, \alpha) = \left[F_n - \sqrt{\frac{F_n(1-F_n)}{n}} u_{1-\frac{\alpha}{2}}, F_n + \sqrt{\frac{F_n(1-F_n)}{n}} u_{1-\frac{\alpha}{2}} \right]$$

Autrement dit

$$P(p \in IC(F_n, \alpha)) \approx 1-\alpha \quad \text{si} \quad np > 10 \text{ et } n(1-p) > 10$$

L'intervalle de confiance est un **intervalle aléatoire** qui dépend de l'échantillon et du niveau de confiance $1-\alpha$.

Calculs des IC pour les N réalisations de $\bar{X}_n$



On observe qu'environ 5% des intervalles de confiance calculés **contiennent** la valeur du paramètre $p = 1/2$.

Exercice

- Refaire tourner le script en modifiant les valeurs de n, N et α dans le premier tronçon d'instruction `simulations` et observer ce qui change (les objets déterministes) et ce qui ne change pas (les objets aléatoires)
- Faire la même expérience numérique dans le modèle normal avec $\mu = 2$ et $\sigma = 2$:
  - dans le tronçon `simulations` utiliser le générateur aléaoire de la loi normale `rnorm()`
  - dans le tronçon `fluctuation` adapter le calcul de l'intervalle de fluctuation au cas du modèle normal
  - dans le tronçon `confiance` adapter le calcul des intervalles de confiance pour μ dans le cas σ inconnu (indication : pour calculer la variance empirique corrigée de chacun des N échantillons de taille n on peut utiliser `apply(dataG,MARGIN=1,var)` qui calcule les N variances corrigées)