

Rappels de statistique : Variables aléatoires, estimations, intervalles de confiance

R. Paegelow - MF Beclin

Statistiques et Analyses de Données - D04

Décembre 2024

Plan du chapitre

Rappels variable aléatoire

Mode de convergence, loi des grands nombres et théorème central limite

Rappel estimation

Étude de l'estimateur "moyenne empirique"

Application directe : étude de l'estimateur d'une proportion

Fluctuations échantillonnales

Estimation par intervalles de confiance (IC)

Estimation par maximum de vraisemblance

Rappels variable aléatoire

Une **variable aléatoire** X est une **application** de Ω ensemble des résultats de l'expérience considérée) dans un ensemble de valeurs E (numériques ou non).

Définition 1.1 : Variable aléatoire

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé et $(E, \mathcal{E})$ un espace mesurable. On appelle variable aléatoire de Ω vers E , toute fonction mesurable X de Ω vers E .

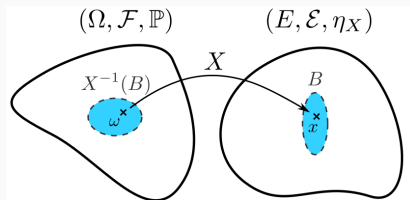


Figure 1: Schéma variable aléatoire

Exemple 1.1 :

Lancer d'une pièce. On peut considérer la variable aléatoire $X = 1$ si pile et $X = 0$ si face.

Exemple 1.2 :

On jette simultanément 2 dés et on note les faces obtenues.

- *Quel ensemble Ω ?*
- On s'intéresse à la plus grande valeur des deux faces. *Quelle variable aléatoire ?*

Différents types de variables aléatoires

- **numériques discrètes** : ensemble des valeurs (numériques) possibles est dénombrable ; par exemple $V = \{v_1, v_2, v_3 \dots v_k\}$ ou encore $V = \mathbb{N}$
- **numériques continues** : ensemble des valeurs (numériques) possibles est continu par exemple $V \subset \mathbb{R}$;
- **qualitatives ordinales** : ensemble des valeurs possibles est qualitatif ordinal ;
- **qualitatives nominales** : ensemble des valeurs possibles est qualitatif non ordinal.

Exemples :

- X : nombre d'enfants dans une famille, $X \in \{0, 1, 2, 3, \dots\}$
- X : taille (en cm) des enfants de 7 à 15 ans, $X \in [100, 170]$
- X : nature de terrains, $X \in \{\text{Peu Argileux, Argileux, Très Argileux}\}$

Pour toute v. a. X , **on sait quel est l'ensemble de ses valeurs possibles** (la valeur sera par exemple dans l'intervalle $[0, 15]$, ou comprise entre $-\infty$ et $+\infty$, ou encore dans l'ensemble $\{\text{Chêne, Hêtre, Bouleau} \dots\}$) mais, pour une expérience donnée, **on ne pourra dire quelle sera sa réalisation effective** ...

On cherche alors à connaître la **loi de probabilité**. Par exemple, dans le cas où E est fini ou dénombrable, on cherche l'application $\mathbb{P}_X$, qui à toutes les valeurs $e \in E$, nous donne la probabilité que $X = e$.

$$\mathbb{P}_X(a) = \mathbb{P}(X^{-1}(a)) = \mathbb{P}(X = a)$$

Dans le cas où $E \subset \mathbb{R}$, on cherche la probabilité de chaque sous ensemble $A \subset E$ (rigoureusement, c'est $A \in \mathcal{B}(E)$)

$$\mathbb{P}_X(A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}(X \in A)$$

Rappel espérance et variance

Définition 1.2 : Espérance

Soit X une variable aléatoire réelle de $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé dans $(E, \mathcal{E})$. On définit l'espérance dans le cas discret

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) \mathbb{P}(\omega) = \sum_{e \in E} e \mathbb{P}_X(e)$$

Dans le cas continue : $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$

$$\mathbb{E}(X) = \int_{\omega \in \Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} e d\mathbb{P}_X(e)$$

Définition 1.3: Variance

Soit X une variable aléatoire réelle de $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé dans $(E, \mathcal{E})$. On définit la variance : $\mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Définition 1.4: Fonction de répartition

Soit X une variable aléatoire réelle de $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé dans $(E, \mathcal{E})$. On définit la fonction de répartition : $F_X(t) = \mathbb{P}(X \leq t)$

Lois discrètes usuelles

- **Loi de Bernoulli $B(p)$**

On s'intéresse à la réalisation d'un évènement A .

Soit la v.a. X : $X = 1$ si A est réalisé, $X = 0$ sinon.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

$$X \sim B(p) ; \quad E(X) = p, \quad E(X^2) = p^2, \quad V(X) = p(1 - p)$$

- **Loi Binomiale $B(n,p)$**

On répète n fois l'épreuve précédente.

Soit S_n : le nombre de réalisations de A au cours des n épreuves.

$$S_n \in \{0, 1, 2, \dots, n\}$$

$$P(S_n = k) = C_n^k p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k}.$$

$$S_n \sim B(n, p) ; \quad E(S_n) = np, \quad V(S_n) = np(1 - p)$$

Lois continues usuelles

La loi Normale $\mathcal{N}(\mu, \sigma^2)$. Fonction de répartition de la loi Normale :

$$F_{\mu, \sigma^2}(t) = \int_{-\infty}^t \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} dx.$$

Pour des lois continues, il peut exister une fonction de densité qui est la dérivée de la fonction de répartition.

Densité loi normale centrée réduite : $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$

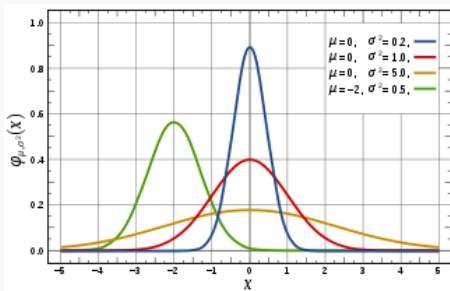


Figure 2: Densité de lois normales.

Quelques propriétés des lois normales :

- Toute combinaison linéaire de lois normales est une loi normale
- Si $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$;
- si $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, indépendantes

$$aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

- *calcul des probabilités ??*
 $\rightsquigarrow$ table de la loi $\mathcal{N}(0, 1)$...ou ordinateurs !
- **Si** $X \sim \mathcal{N}(\mu, \sigma^2)$, $P(\mu - 1.96\sigma < X \leq \mu + 1.96\sigma) = 0.95$

Mode de convergence, loi des grands nombres et théorème central limite

Définition 2.1 : Échantillon

On appelle échantillon (d'une v.a. X) de taille n une suite de n v.a. $(X_1, \dots, X_n)$ indépendantes et identiquement distribuées (i.i.d.)

- Cela correspond aux conditions suivantes :
 - Tous les individus sont sélectionnés dans la même population et sont donc identiques à quelques variations près
 - les individus sont sélectionnés de manière indépendante
- Parfois on introduit une variable X de même loi que les X_i que l'on nomme variable parente de l'échantillon.
- Après une expérience, on recueille un jeu de données constitué des observations $(x_1, \dots, x_n)$. C'est une réalisation de l'échantillon aléatoire $(X_1, \dots, X_n)$: x_1 est la réalisation de la variable X_1 .

Voici les notions de convergence stochastique indispensables pour présenter ensuite des résultats centraux de la théorie des probabilités et les fondements de la statistique.

- **Convergence en Probabilité**

$$Z_n \xrightarrow{\mathbb{P}} Z \quad \text{ssi} \quad \forall \varepsilon > 0, \mathbb{P}(|Z_n - Z| > \varepsilon) \xrightarrow{n \rightarrow +\infty} 0.$$

- **Convergence en moyenne quadratique**

$$Z_n \xrightarrow{L^2} Z \quad \text{ssi} \quad \mathbb{E}[(Z_n - Z)^2] \xrightarrow{n \rightarrow +\infty} 0.$$

- **Convergence en loi**

$$Z_n \xrightarrow{\mathcal{L}} Z \quad \text{ssi} \quad \forall t, F_{Z_n}(t) \xrightarrow{n \rightarrow +\infty} F_Z(t).$$

Théorème 2.1 : Loi des grands nombres

Soit $(X_1, \dots, X_n)$ une suite de variables aléatoires non corrélées de même loi (i.i.d.) que la variable parente X dont le moment d'ordre 2 est fini. Alors on a :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{L^2} \mathbb{E}(X)$$
$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}(X)$$

Théorème 2.2 : Théorème Central Limite

Soit $(X_1, \dots, X_n)$ un échantillon de variable parente X dont le moment d'ordre 2 est fini. Posons $\mu = \mathbb{E}(X)$. Alors, on a

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sqrt{\text{Var}(X)}} \xrightarrow{\mathcal{L}} Z \text{ avec } Z \sim \mathcal{N}(0, 1).$$

Ce résultat est valide quelle que soit la loi de X ! Y compris pour des variables discrètes.

Rappel estimation

Cadre général : On cherche à se faire une idée, aussi précise que possible, sur la valeur inconnue d'un paramètre d'une population à partir d'observations sur **un échantillon** de taille n . Dans la suite :

- Le terme **paramètre** désignera **toujours** une valeur **déterministe** (i.e. non-aléatoire) ; cette valeur sera donc fixe (*mais inconnue !*)
- Lorsqu'un paramètre est inconnu, on en cherchera *une estimation* à partir de réalisations sur un échantillon, **cette estimation dépendra des observations faites ! !**
- *Attention* : une **estimation** ne permet **jamais** de connaître la **vraie valeur** du paramètre !

Définition 3.1 :

Soit θ un paramètre d'une population P et $(X_1, \dots, X_n)$ un n -échantillon issu de P . On appelle estimateur de θ toute fonction $f(X_1, \dots, X_n)$; pour des observations $(x_1, \dots, x_n)$, une estimation de θ est donnée par $f(x_1, \dots, x_n)$.

Exemple 3.1 :

La moyenne empirique $\overline{X}_n$ est un estimateur de la moyenne théorique.

Remarques :

1. Ne pas confondre *estimateur* et *estimation* ;
2. De part sa définition, un estimateur est une variable aléatoire !
3. Quelles propriétés sont souhaitables pour un *bon* estimateur ?
4. Estimation $\rightarrow$ valeur *d'estimation ponctuelle* du paramètre ;
estimation + fluctuations échantillonnales $\rightarrow$ intervalle ayant une (forte) probabilité de contenir la vraie valeur du paramètre.

Bon estimateur : quelles propriétés ?

Estimateur sans biais ; Estimateur asymptotiquement sans biais

Définition 3.2 : Biais d'un estimateur

Soit $\hat{\theta}$ un estimateur de θ . Le biais de l'estimateur θ est :

$$b(\hat{\theta}) = \mathbb{E}(\hat{\theta}) - \theta$$

Définition 3.3 : Biais asymptotique

Soit $\hat{\theta}_n$ un estimateur de θ obtenue avec un échantillon de taille n .
Le biais asymptotique de l'estimateur $\hat{\theta}_n$ est

$$\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}_n) - \theta$$

Il est souhaitable pour un estimateur d'être sans biais (de biais nul). En effet, l'estimateur est une variable aléatoire est l'estimation est une réalisation de celle-ci. Si l'estimateur est sans biais, cela signifie qu'en moyenne, il tombe juste.

Dans certaines situations, on ne peut pas avoir un estimateur sans biais, on se contentera d'un estimateur asymptotiquement sans biais. La taille du biais diminue alors avec la taille de l'échantillon. Le biais tend vers 0 lorsque la taille de l'échantillon tend vers ∞ . Si on est dans une situation où on dispose de très grand échantillon, cela ne pose pas de problème.

Estimateur convergent (ou correct)

Un estimateur convergent est un estimateur $\hat{\theta}_n$ dont la variance $\hat{\theta}_n$ tend vers 0 quand la taille n de l'échantillon tend vers l'infini. **La précision de l'estimateur augmente lorsque n grandit.**

Estimateur absolument convergent

C'est un estimateur sans biais convergent :

$$\lim_{n \rightarrow +\infty} \mathbb{E}(\hat{\theta}_n) = \theta;$$

$$\lim_{n \rightarrow +\infty} \text{Var}(\hat{\theta}_n) = 0.$$

Il est possible d'avoir deux estimateurs absolument convergents $\hat{\theta}_n$ et $\hat{\theta}'_n$ d'une même quantité. On privilégiera celui de variance minimale. On dira qu'il est plus **efficace** que l'autre.

Loi de l'estimateur

Il est aussi intéressant de regarder les éventuelles convergences stochastiques d'un estimateur. En effet, en appliquant par exemple le théorème central limite ou d'autre résultat, on peut connaître la loi asymptotique d'un estimateur.

Convergence de l'estimateur

Il est aussi intéressant de regarder les éventuelles convergences stochastiques d'un estimateur. En effet, en appliquant par exemple le théorème central limite ou d'autre résultat, on peut connaître la loi asymptotique d'un estimateur. Cela peut être très intéressant pour exhiber des intervalles de confiance et faire des tests statistiques (cf. cours n°2).

Étude de l'estimateur "moyenne empirique"

Soit $(X_1, \dots, X_n)$ un échantillon de variable parente X de variance et d'espérance finie. On veut estimer le paramètre $\theta = \mathbb{E}(X)$.

Un estimateur de l'espérance $\theta = \mathbb{E}(X)$ est la moyenne empirique

$$\hat{\theta}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Cet estimateur est sans biais.

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X) = \mathbb{E}(X) = \theta$$

Cet estimateur est convergent en norme quadratique et donc aussi en probabilité.

$$\bar{X}_n \xrightarrow{P} \theta.$$

Il s'agit de la loi des grands nombres (LGN) que nous avons vus dans la section précédente (Théorème 1).

Montrer que $\mathbb{E} \left[(\hat{\theta}_n - \theta)^2 \right] = \text{Biais } (\hat{\theta}_n)^2 + \text{Var}(\hat{\theta}_n).$

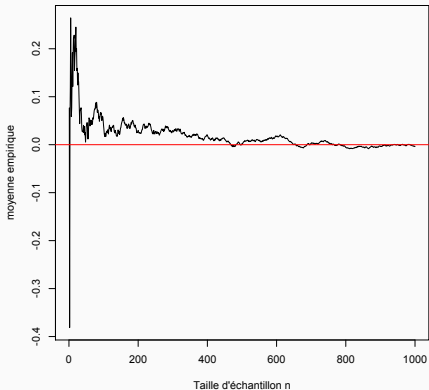
$$\begin{aligned}\mathbb{E} \left[(\hat{\theta}_n - \theta)^2 \right] &= \mathbb{E} \left[(\hat{\theta}_n)^2 \right] + \theta^2 - 2\theta \mathbb{E} \left[\hat{\theta}_n \right] \\ &= \text{Var}(\hat{\theta}_n) + \mathbb{E} \left[\hat{\theta}_n \right]^2 + \theta^2 - 2\theta \mathbb{E} \left[\hat{\theta}_n \right] \\ &= \text{Var}(\hat{\theta}_n) + \left[\mathbb{E} \left(\hat{\theta}_n \right) - \theta \right]^2\end{aligned}$$

$\hat{\theta}_n$ est sans biais. Donc

$$\begin{aligned}\mathbb{E} \left[(\hat{\theta}_n - \theta)^2 \right] &= \text{Var}(\hat{\theta}_n) = \text{Var} \left(\frac{1}{n} \sum_{i=1}^n X_i \right) \\ &= \frac{1}{n^2} \left[\sum_{i=1}^n \text{Var}(X_i) + 2 \sum_i \sum_{j>i} \text{Cov}(X_i, X_j) \right] \\ &= \frac{1}{n} \text{Var}(X) \rightarrow 0\end{aligned}$$

La Loi des Grands Nombres (LGN)

```
> x<-runif(1000,-1,1)
> plot(cumsum(x)/(1:1000),type="l",xlab="Taille d'échantillon n",
ylab="moyenne empirique")
> abline(h=0,col=2)
```



Quelle que soit la loi de X , en pratique pour n assez grand on peut approximer la loi de $\bar{X}_n$ par une loi normale de moyenne $\theta = \mathbb{E}(X)$ et de variance $\frac{Var(X)}{n}$. Ce résultat vient du théorème central limite (Théorème 2).

Application directe : étude de l'estimateur d'une proportion

Soit $(X_1, \dots, X_n)$ un échantillon de variable parente X binaire, qui prend la valeur 1 si l'individu a une caractéristique donnée et 0 sinon. **Donc**
 $X \sim \mathcal{B}(p)$.

On veut estimer le paramètre $p = \mathbb{E}(X)$.

Un estimateur de p est $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$, la proportion empirique d'individus ayant cette caractéristique

- Cet estimateur est sans biais. $\mathbb{E}(\hat{p}_n) = p$ et on a de plus $Var(\hat{p}_n) = \frac{Var(X)}{n} = \frac{p(1-p)}{n} \rightarrow 0$.
- Cet estimateur est convergent en norme quadratique et donc aussi en probabilité (LGN).
 $\hat{p}_n \xrightarrow{L^2} p$ et $\hat{p}_n \xrightarrow{P} p$.
- On connaît la loi exacte de $\sum_i X_i \sim \mathcal{BIN}(n, p)$. En pratique, quand $n \geq 30, np \geq 5$ et $\mathcal{N}(1-p) \geq 5$ il est plus simple d'utiliser une approximation grâce au TLC on peut dire que $\sqrt{n} \frac{\hat{p}_n - p}{\sqrt{p(1-p)}}$ suit approximativement une loi normale centrée réduite.

Attention, on ne connaît pas p au dénominateur ! Que faire en pratique ?

Fluctuations échantillonnales

Echantillons → information *partielle* sur la population . . .

Conséquences : On ne connaît pas les fluctuations de toutes quantités évaluées sur des échantillons

On cherche alors à quantifier de façon rigoureuse les fluctuations inhérentes aux approches statistiques.

⇒ caractériser la loi de ces fluctuations !

Une population comporte 25% d'individus atteints d'une maladie M . On tire un n -échantillon . . .

- Pas de raisons que la proportion de malades de l'échantillon soit exactement égale à 25% !
- D'après la section précédente, pour un échantillon de taille n , l'estimateur $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n X_i$ est sans biais et convergent.
- Il est vraisemblable que ce pourcentage ne soit pas trop éloigné de 25% (et ceci d'autant plus que n sera grand ?) tout en sachant qu'il va fluctuer. **Mais comment ?** Il nous faut la **loi de probabilité pour caractériser cette "incertitude" !**

- Population : 100 sportifs dont on mesure la taille.

[1] 150.48 188.25 153.08 180.39 163.05 161.37 180.97 182.00

[9] 174.53 185.65 159.69 173.29 177.98 192.27 175.10 160.59

⋮

[89] 179.86 182.14 171.54 174.37 171.40 161.30 193.65 156.27

[97] 164.03 168.00 165.21 176.56

Taille moyenne de la population : 172.6 cm.

↪ *Quelles fluctuations pour la moyenne d'échantillons de taille 20 pris dans cette population ?*

Une idée ? : Prendre plusieurs échantillons de taille $n = 20$ dans cette population et s'intéresser aux fluctuations des moyennes échantillonnales correspondantes.

Echantillon 1 :

[1] 148.22 163.27 163.05 182.18 156.85 162.43 193.31 171.75 186.02 170.57
[11] 177.51 183.51 174.53 193.65 168.28 176.69 180.06 179.04 164.64 192.27

Moyenne : 174.39

Echantillon 2 :

[1] 179.86 192.27 168.95 173.29 150.48 183.51 172.41 161.37 174.37 177.82
[11] 170.50 179.11 175.23 174.53 171.63 180.03 163.27 180.65 153.08 148.22

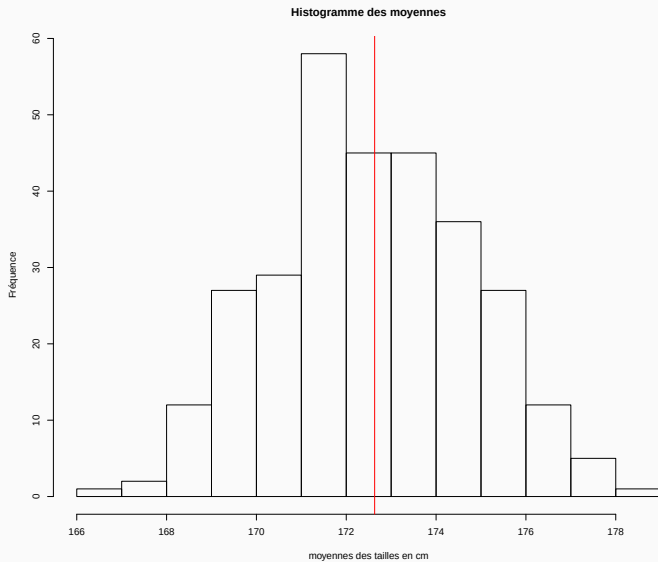
Moyenne : 171.53

On calcule ainsi la moyenne associée à 300 échantillons de taille 20 :

[1] 173.06 172.31 175.22 168.40 174.62 176.96 170.07 173.15 175.46 174.76
[11] 175.61 173.74 173.78 170.81 168.65 169.77 174.53 169.57 172.60 171.12 ...
[281] 172.63 170.80 174.25 172.82 173.47 170.74 171.02 174.30 174.33 171.33
[291] 170.45 173.60 174.15 169.23 168.67 172.31 172.85 171.41 171.99 174.94

↪

300 valeurs de moyennes échantillonnales




```
fmean=functio\mathcal{N}(x)
{
  retur\mathcal{N}(mea\mathcal{N}(sample(Pop,20)))
}
```

```
Vect=rep(0,300)
Vect=sapply(Vect,fmean)
```

```
hist(Vect,main="Histogramme des moyennes",
      xlab="moyennes des tailles en cm",
      ylab="Fréquence")
abline(v=mea\mathcal{N}(Pop),col="red")
```

Lorsque l'on considère un très grand nombre d'échantillons de (même) taille n , on peut donc avoir une idée de la loi de fluctuations de la variable aléatoire que l'on considère . . .

⇒ caractériser la loi des fluctuations d'une variable aléatoire sur les n -échantillons ?

En pratique, on ne dispose que d'un seul échantillon.

Mais les distributions de probabilité vont permettre de faire cette caractérisation !

Estimation par intervalles de confiance (IC)

L'estimation ponctuelle à partir d'un échantillon ne renseigne pas sur la précision de l'approximation de θ . On voudrait donc obtenir un **intervalle aléatoire**, pas trop grand, à partir de l'échantillon prélevé, tel que **la probabilité qu'il contienne θ soit acceptable**.

Le **niveau de confiance** est la probabilité qu'un intervalle (de confiance) contienne θ , on la désigne par **$1 - \alpha$** .

Le nombre **α** est le risque que l'on prend de se tromper en affirmant que θ est bien dans l'intervalle proposé.

Définition 7.1 : Intervalle de confiance

Un intervalle de confiance est un intervalle aléatoire déterminé par deux bornes inférieure et supérieure **aléatoires** t_1 et t_2 telles que :

$$P(t_1 \leq \theta \leq t_2) = 1 - \alpha.$$

On veut construire un intervalle de confiance associé à un certain niveau de confiance.

- À taille d'échantillon fixée n , lorsqu'on augmente le niveau de confiance $1 - \alpha$, la largeur de l'IC augmente
- À niveau de confiance fixé $1 - \alpha$, lorsqu'on augmente la taille de l'échantillon n , la largeur de l'IC diminue.

Soit $X_1, \dots, X_n$ un échantillon de variable i.i.d parente à partir d'un échantillon $X_1, \dots, X_n$ de variable parente $X \sim \mathcal{N}(\mu, \sigma^2)$, la variance étant connue

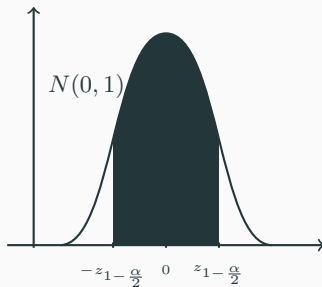
- **Estimateur** : $\bar{X} = \frac{1}{n} \sum_i X_i$ (estimation ponctuelle $\hat{\mu} = \bar{X}$)
- **loi de l'estimateur** : $Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
- **Intervalle de confiance** de niveau $(1 - \alpha)$. On note, $z_\beta = \phi^{-1}(\beta)$ le quantile d'ordre β de la loi normale centrée réduite. En utilisant une table de quantiles on a la valeur $z_{1-\frac{\alpha}{2}} > 0$ telle que :

$$\mathbb{P} \left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$
$$\Leftrightarrow \mathbb{P} \left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$\mathbb{P} \left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \leq z_{1-\frac{\alpha}{2}} \right) = 1 - \alpha$$

$$\Leftrightarrow \mathbb{P} \left(\bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) = 1 - \alpha$$

$$IC = \left[\bar{X} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{X} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$



Si $X_i \sim N(\mu, \sigma^2)$ alors $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ quelle que soit la valeur de n . Si la loi de X_i est quelconque mais que sa variance est connue, alors le théorème central limite nous dit que pour n assez grand on peut approximer la loi de $\bar{X}$ par cette même loi $N(\mu, \frac{\sigma^2}{n})$.

Pour trouver les intervalles de confiance, on “centre et normalise” $\bar{X}$, c'est à dire que l'on considère la variable aléatoire $Y = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ qui suit (approximativement dans le cas TLC) une loi normale centrée réduite. On peut alors utiliser les quantiles de $N(0, 1)$ et les tables pour trouver l'intervalle de confiance.

On s'intéresse à la proportion de réalisation p_A de l'événement A .

- **Estimateur** : $\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{X_i=A}$ (estimation ponctuelle)
- **Loi de l'estimateur** : $\hat{p}_n \sim \mathcal{B}(n, p_A)$. Pour n assez grand, grâce au TCL on va approcher la version centrée et réduite de cette loi par une loi normale centrée réduite :

$$Y = \frac{\hat{p}_n - p_A}{\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}} \sim N(0, 1)$$

- **Intervalle de confiance** :

On procède alors exactement comme dans le cas de l'estimation d'une moyenne à variance connue, de sorte que

$$IC = \left[\hat{p}_n - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} ; \hat{p}_n + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right]$$

Un intervalle de confiance peut être :

- Unilatéral : limite d'un seul côté :
 - inférieur, par t_1 : $P(t_1 \leq \theta) = 1 - \alpha$ (unilatéral à droite),
 - supérieur, par t_2 : $P(\theta \leq t_2) = 1 - \alpha$ (unilatéral à gauche).
- Bilatéral : limites des deux côtés
 - $P(t_1 \leq \theta) = \alpha_1$,
 - $P(\theta \leq t_2) = 1 - \alpha_2$, avec $\alpha_1 + \alpha_2 = \alpha$.

En pratique, on choisit souvent $\alpha_1 = \alpha_2 = \alpha/2$ (intervalle bilatéral symétrique).

Estimation d'une espérance avec variance connue :

- **Conditions d'utilisation** : $X \sim \mathcal{N}(\mu, \sigma^2)$ et n quelconque **ou** X de loi quelconque et n grand (*TCL*).
- **Estimateur** : $\bar{X}_n = \frac{1}{n} \sum_i X_i$; on a alors que $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$, d'où :

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$$

$\rightsquigarrow$

$$P\left(-z_{1-\alpha/2} < \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_{1-\alpha/2}\right) = 1 - \alpha$$

avec z_α quantile d'ordre α de la loi $\mathcal{N}(0, 1)$.

- **Intervalle de confiance bilatéral symétrique**

$$\bar{X}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Estimation d'une espérance avec variance inconnue)

- **Conditions d'utilisation** : n quelconque et $X \sim \mathcal{N}(\mu, \sigma^2)$ ou n grand et X de loi quelconque (*TCL*).
- **Estimateur** : $\bar{X}_n = \frac{1}{n} \sum_i X_i$. $\bar{X}_n \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$ mais σ^2 inconnue
...on l'estime par $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

$$\frac{\bar{X}_n - \mu}{S/\sqrt{n}} \sim T_{(n-1)}$$

$$P\left(-t_{1-\alpha/2;(n-1)} < \frac{\bar{X}_n - \mu}{S/\sqrt{n}} \leq t_{1-\alpha/2;(n-1)}\right) = 1 - \alpha$$

avec $t_{\alpha;(n-1)}$ quantile d'ordre α de la loi de Student $T_{(n-1)}$; on a $t_{\alpha;(n-1)} = -t_{1-\alpha;(n-1)}$.

- **Intervalle de confiance bilatéral symétrique**

$$\bar{X} - t_{1-\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{1-\alpha/2} \frac{S}{\sqrt{n}}$$

Application : estimation d'une proportion p

- **Conditions d'utilisation** : n grand ; $np > 5$; $\mathcal{N}(1 - p) > 5$ (TCL).
- **Estimateur** : $P_n = \frac{\text{nbre succès sur échantillon}}{n}$.

$$\frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim \mathcal{N}(0, 1)$$

$$\rightsquigarrow P \left(z_{\alpha/2} < \frac{P_n - p}{\sqrt{\frac{p(1-p)}{n}}} \leq z_{1-\alpha/2} \right) = 1 - \alpha$$

mais p inconnue !

- le remplacer par P_n

Application : intervalle de confiance de niveau $(1 - \alpha)$ pour p

- Intervalle de confiance bilatéral symétrique**

$$\hat{p}_n - z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \leq p \leq \hat{p}_n + z_{1-\alpha/2} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}}$$

Attention : vérification a posteriori de l'utilisation de TCL en utilisant les bornes de l'IC pour valider empiriquement $np > 5$ et $N(1 - p) > 5$

La longueur d'un IC pour une confiance donnée est d'autant plus petite que n est grand ...

⇒ **précision d'une estimation augmente avec n**

Si Δ représente une précision fixée à l'avance, il suffit de résoudre :
demie-longueur de l'IC $= \Delta$ pour déterminer n assurant la précision voulue.

Cas d'une espérance : $z_{1-\alpha/2} \frac{s}{\sqrt{n}} = \Delta$

Notations standards utilisées en estimation pour désigner un **estimateur** d'un **paramètre** θ on utilise souvent $\hat{\theta}$.

Exemples : $\hat{\mu}$; $\hat{\sigma}$; $\hat{\sigma}^2$; $\hat{\pi}$...

- Avantage : facilité d'utilisation ! (*plus besoin de donner des “noms” particuliers aux estimateurs*);
- Inconvénient : même notation pour l'estimateur et l'estimation d'où les possibles confusions ... **attention de ne pas confondre une v.a. (estimateur) et l'une de ses réalisations possibles (estimation)**

Estimation par maximum de vraisemblance

Maintenant que nous avons défini l'échantillon et l'estimateur, identifié les qualités souhaitables d'un estimateur et appris à déterminer des intervalles de confiance pour des estimateurs bien connus, nous pouvons nous interroger sur la manière de trouver un estimateur pour des situations autres que l'estimation d'une espérance ou d'une proportion.

Il y a plusieurs façons de résoudre cette tâche.

Les statisticiens utilisent une grandeur appelée **la vraisemblance**.

La **vraisemblance** est une fonction qui mesure, pour chaque valeur du paramètre θ recherché, combien cette valeur est vraisemblable au vu des données.

On cherche alors à maximiser la vraisemblance. L'estimateur déterminé par cette méthode est appelé, **estimateur par maximum de vraisemblance** (MLE : Maximum Likelihood estimation en anglais). En plus de nous donner des estimateurs dans beaucoup d'autre circonstance que celle évoquée précédemment, sous certaines conditions (que nous ne détaillerons pas dans ce cours).

Définition 8.1 : Vraisemblance pour une variable aléatoire discrète

Soit X une variable aléatoire suivant une loi discrète $\mathbb{P}$ dépendant d'un paramètre θ . Soit x une réalisation de X , la vraisemblance $\mathcal{L}$ est une fonction de θ qui s'écrit alors :

$$\mathcal{L}(\theta; x) = \mathbb{P}_\theta(X = x)$$

Exemple 8.1 :

Soit $X \sim \mathcal{B}(p)$, soit x une réalisation de la variable aléatoire. $\mathcal{L}(p; x) = \mathbb{P}_\theta(X = x)$. Si $x = 0$, $\mathcal{L}(p; x) = 1 - p$ et si $x = 1$, $\mathcal{L}(p; x) = p$ D'ou $\mathcal{L}(p; x) = p^x + (1 - p)^{(1-x)}$

Définition 8.2: Vraisemblance pour une variable aléatoire discrète

Soit X une variable aléatoire suivant une loi continue qui admet une densité f dépendant d'un paramètre θ . Soit x une réalisation de X , la vraisemblance $\mathcal{L}$ est une fonction de θ qui s'écrit alors :

$$\mathcal{L}(\theta; x) = f_{\theta}(x)$$

Exemple 8.2 :

Soit $X \sim \mathcal{N}(\mu, \sigma^2)$ avec $\sigma = 1$, soit x une réalisation de la variable aléatoire X . $\mathcal{L}(\mu; x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Définition 8.3 : Vraisemblance d'un échantillon

Soit $X_1, \dots, X_n$ un échantillon i.i.d. de loi parente X suivant une loi qui dépend d'un (ou plusieurs paramètres) θ et $\mathbf{x} = (x_1, \dots, x_n)$ les réalisations de l'échantillon. La vraisemblance s'écrit alors :

$$\mathcal{L}(\theta; \mathbf{x}) = \mathcal{L}(\theta; x_1, \dots, x_n) = \prod_{i=1}^n \mathcal{L}(\theta; x_i)$$

Ainsi, l'estimateur du maximum de vraisemblance s'écrit :

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \mathcal{L}(\theta, \mathbf{x})$$

Exemple 8.3 :

Soit $X_1, \dots, X_n$ i.i.d $\sim \mathcal{B}(p)$ et $\mathbf{x} = (x_1, \dots, x_n)$ les réalisations de ces v.a. On cherche p . On procède par maximum de vraisemblance :

$$\hat{p}_{MLE} = \arg \max_{p \in [0,1]} \mathcal{L}(p, \mathbf{x}) = \arg \max_{p \in [0,1]} \prod_{i=1}^n \mathcal{L}(p, x_i) =$$

$$\arg \max_{p \in [0,1]} \prod_{i=1}^n p^{x_i} (1-p)^{(1-x_i)}$$

Souvent, pour faciliter les calculs, on passe au $\log$:

$$\begin{aligned}\arg \max_{\theta \in \Theta} \mathcal{L}(\theta, \mathbf{x}) &= \arg \max_{\theta \in \Theta} \log \left(\prod_{i=1}^n \mathcal{L}(\theta; x_i) \right) \\ &= \arg \max_{\theta \in \Theta} \sum_{i=1}^n \log(\mathcal{L}(\theta, x_i))\end{aligned}$$

En reprenant l'exemple précédent,

$$\hat{p}_{MLE} = \arg \max_p \sum_{i=1}^n x_i \log(p) + \sum_{i=1}^n (1 - x_i) \log(1 - p)$$

$$\begin{aligned}
\frac{\partial \mathcal{L}(p, x_i)}{\partial p} &= \frac{x_i}{p} - \frac{(1 - x_i)}{1 - p} \\
\frac{\partial \log(\mathcal{L}(\theta, \mathbf{x}))}{\partial p} &= \frac{\sum_{i=1}^n x_i}{p} - \frac{n}{1 - p} + \frac{\sum_{i=1}^n x_i}{1 - p} \\
\frac{\partial \log(\mathcal{L}(\theta, \mathbf{x}))}{\partial p} = 0 &\Rightarrow \sum_{i=1}^n x_i - np = 0
\end{aligned}$$

Ainsi :

$$\hat{p}_{MLE} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Exercice : Pour chacune des trois questions, réaliser les étapes suivantes.

- Écrire la vraisemblance.
- Numériquement, générer un échantillon i.i.d de loi exponentielle en choisissant une valeur de paramètre.
- Tracer numériquement la vraisemblance et estimer numériquement la valeur de λ (ou μ, σ^2) qui la maximise.
- Trouver analytiquement l'estimateur du maximum de vraisemblance
- Vérifier numériquement cette expression pour plusieurs valeurs de lambda.

1) (μ, σ^2) pour un échantillon $(X_1 \dots, X_n) i.i.d \sim \mathcal{N}(\mu, \sigma^2)$.

2) λ pour un échantillon $(X_1 \dots, X_n) i.i.d \sim \mathcal{Exp}(\lambda)$. La densité de la loi exponentielle est $f_{\mathcal{Exp}}(x) = \lambda e^{-\lambda x}$

3) λ pour un échantillon $(X_1 \dots, X_n) i.i.d \sim \mathcal{Exp}(\lambda)$. loi de Poisson de paramètre λ : $\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$