

Licence 2 MIO : Intro. à la Statistique Inférentielle

Mouhamad M. Allaya, PhD

Université de Thiès

2 juin 2025



Objectifs

Préambule

Rappels de lois
discrètes et continues

Statistiques usuelles &
Estimations
ponctuelles

Estimation par
intervalle de confiance

Tests de conformité

Références

1. Motiver et présenter quelques outils de Statistiques Décisionnelles ;
2. Donner un aperçu de méthodes d'estimations ponctuelles et par intervalles ;
3. Donner un usage des intervalles de confiance et Tests statistiques pour des paramètres d'intérêts ;
4. Petite sensibilisation à R & exemples computationnels centrés sur l'usage de Statistique Décisionnelle.

Sommaire I

Préambule

Rappels de lois discrètes et continues

Variable aléatoire lois discrètes

Statistiques usuelles & Estimations ponctuelles

Statistiques usuelles Moyenne, Variance et Fréquence empiriques

Estimation ponctuelle Méthode des moments Méthode EMV

Estimation par intervalle de confiance

- I.C pour une moyenne : premier cas
- I.C pour une moyenne : deuxième cas
- I.C pour une variance : Premier cas
- I.C pour une variance : Deuxième cas
- I.C pour une proportion

Tests de conformité

Sommaire II

Test de conformité sur une proportion

Test de conformité sur une moyenne

Test de conformité sur une variance

Références

Statistique : utilité, portée

Construire, Modéliser, Gérer, et Maintenir des données issues d'entrepôts de données pour par exemple

- ▶ Mesurer la portée de campagnes Marketing;
- ▶ Connaitre l'efficacité d'un traitement sur des patients atteints d'une maladie donnée;
- ▶ mesurer les pertes et profits d'une entreprise;

Applications en Marketing, Finance, Assurance, Gestion de portefeuilles, Informatique, etc.

📌 Extraire/résumer de l'information pertinente pour des fins de décision

Exemples préliminaires

- ▶ Observations, mesures :
 - ▶ on lance une pièce 10 fois, on obtient 4 fois Face ;
 - ▶ on rate le bus, le prochain arrive 12 mn plus tard ;
 - ▶ il pleut aujourd'hui lundi alors qu'hier dimanche le ciel était nuageux ;
 - ▶ un certain roulement à billes a fonctionné pendant deux ans et 3 mois
 - ▶ etc.
- ▶ Il s'agit d'observations, de mesures ou d'évènements concrets

Données observées, valeurs issues de l'expérience,
empiriques

Statistique, Modélisation et Probabilité

Définition

Les valeurs observées sont vues comme des observations de phénomènes aléatoires

- ▶ résultat d'un lancer d'une pièce;
- ▶ intervalle de temps entre deux passages du bus;
- ▶ temps qu'il fait le lendemain d'un jour nuageux;
- ▶ durée de fonctionnement d'un certain type de roulement à billes;
- ▶ etc.

Définition

On observe une valeur x . **Modéliser**, c'est considérer qu'il s'agit de la valeur observée d'une variable aléatoire X .

Besoin des probabilités, d'un modèle théorique

Statistique et Probabilité

Préambule

Rappels de lois discrètes et continues

Statistiques usuelles & Estimations ponctuelles

Estimation par intervalle de confiance

Tests de conformité

Références

Ex 1	On observe une proportion de Face $x = 4/10$ La proba. théorique est $1/2$ si la pièce n'est pas truquée
Ex 2	L'intervalle observé est de 12 mn Le bus quitte son terminus toutes les 10 mn
Ex 3	Il pleut lundi, à la suite d'un dimanche nuageux La météo indique une probabilité de pluie de $2/3$
Ex 4	Le roulement à billes a fonctionné 27 mois Le fabricant garantit une durée de vie de 3 ans

Y-a-t-il une contradiction entre les observations et le modèle?

Y-a-t-il une contradiction entre la pratique et la théorie?

Rappels de probabilité

- ▶ Une variable aléatoire (VA) est une fonction des évènements qui n'est pas calculable (comme une fonction usuelle) mais seulement observable
 - ▶ A chaque expérience (mesure, évènement élémentaire) est associé un nombre, qu'on peut seulement observer, et non prévoir
 - ▶ Elle est définie par
 - ▶ l'ensemble des valeurs qu'elle peut prendre
 - ▶ sa distribution (ou répartition) sur ces valeurs
- 1 Discrète : sur $\{0, 1\}$, sur $\{0, 1, \dots, n\}$, sur \mathbb{N}
 - 2 Continue : sur $[0, 1]$, sur \mathbb{R}

Variable aléatoire discrète I

On dit que $X \sim \mathcal{B}(\theta)$ si

$$\forall x \in \{0, 1\} \quad \mathbb{P}(X = x) = \theta^x (1 - \theta)^{1-x}$$

avec $\theta \in]0, 1[$.

Variable aléatoire discrète II

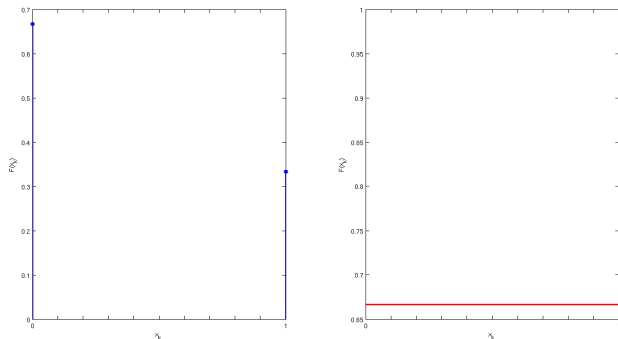


Figure – v.a de Bernoulli de paramètre $\theta = 1/3$

Variable aléatoire discrète III

Domaines d'utilisation

1. En médecine, l'état d'un patient;
2. En assurance, la survenue de sinistre;
3. etc.

Variable aléatoire discrète I

On dit que que $X \sim \mathcal{B}in(n, \theta)$ si

$$\forall k \in \{0, 1, \dots, n\} \quad \mathbb{P}(X = k) = C_n^k \theta^k (1 - \theta)^{n-k}$$

avec $\theta \in]0, 1[$.

Variable aléatoire discrète II

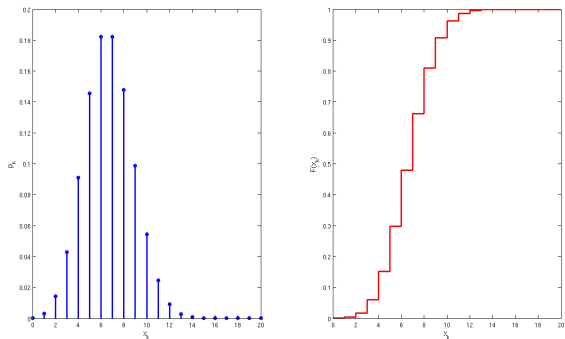


Figure – v.a binomiale de paramètres $n = 20, \theta = 1/3$

Variable aléatoire discrète III

Domaines d'utilisation

1. Détection de produits/articles défectueux dans une production ;
2. Détection de l'efficacité du médicament dans les essais cliniques ;
3. Finance (*Pricing* de produits dérivés)
4. etc.

Variable aléatoire continue I

On dit que que $X \sim \mathcal{N}(0, 1)$ si

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Variable aléatoire continue II

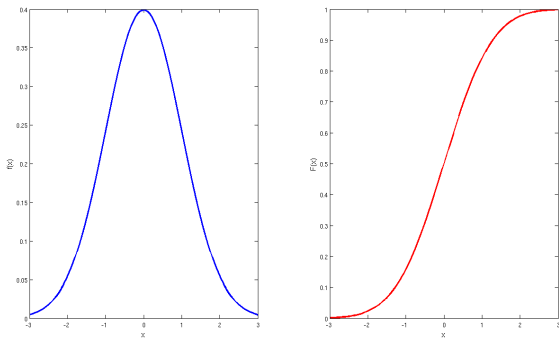


Figure – v.a normales de paramètres $(0, 1)$

Variable aléatoire continue III

Domaines d'utilisation

1. distribution de revenus ;
2. répartition de la taille de la population ;
3. etc.

Variable aléatoire continue I

On dit que $X \sim \chi_k^2$, $k \in \mathbb{N}^*$ si :

$$\forall x \in \mathbb{R}_+, f_X(x) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}$$

Variable aléatoire continue II

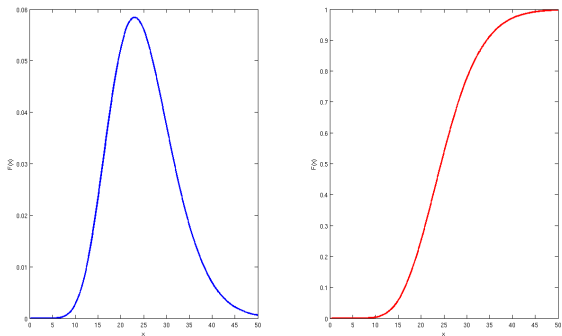


Figure – v.a de Chi2 de paramètre $k = 25$

Variable aléatoire continue III

Domaines d'utilisation

1. Tests d'indépendance, d'adéquation
2. Intervalles de confiance

Variable aléatoire continue I

On dit que $X \sim \mathcal{F}(k)$, $k \in \mathbb{N}^*$ si :

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma((k+1)/2)}{\Gamma(k/2)} \left(1 + \frac{x^2}{k}\right)^{-\left(\frac{k+1}{2}\right)}$$

Variable aléatoire continue II

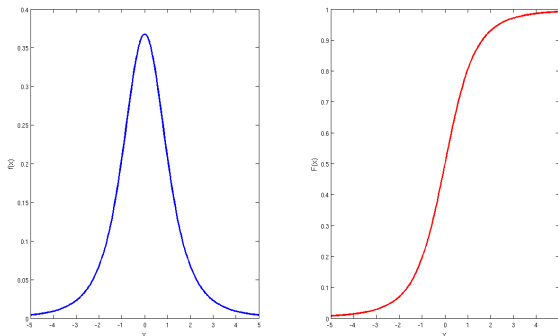


Figure – v.a de Student de paramètre $k = 3$

Variable aléatoire continue III

Domaines d'utilisation

1. Caractérisation de tumeurs par IRM
2. Intervalles de confiance
3. Tests de comparaison
4. Régression

Variable aléatoire continue I

On dit que $X \sim \mathcal{F}(k_1, k_2)$, $k_1, k_2 \in \mathbb{N}^*$ si :

$$\forall x \in \mathbb{R}_+^*, f_X(x) = \frac{\left(\frac{k_1 x}{k_1 x + k_2}\right)^{k_1/2} \left(1 - \frac{k_1 x}{k_1 x + k_2}\right)^{k_2/2}}{xB(k_1, k_2)}$$

Variable aléatoire continue II

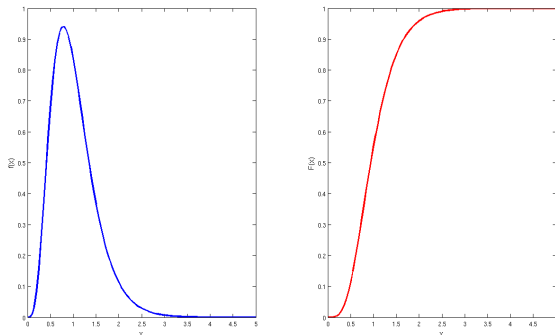


Figure – v.a de Fisher de paramètres $k_1 = 10, k_2 = 100$

Variable aléatoire continue III

Domaines d'utilisation

1. Images de radars
2. En econometrie
3. Tests d'hypothèses

Autres lois importantes I

Géométrique

$X \sim \text{Geom}(p)$ (où $p > 0$) : le nombre de lancés d'une pièce de monnaie jusqu'au premier face, pour une pièce qui tombe sur face avec une probabilité p .

$$\mathbb{P}(X = x) = p(1 - p)^{x-1}, \quad x = 1, 2, \dots$$

Poisson

$X \sim \text{Poisson}(\lambda)$ (où $\lambda > 0$) : une distribution de probabilité sur les entiers utilisée pour modéliser la fréquence des événements rares.

$$\mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \quad x = 0, 1, 2, \dots$$

Autres lois importantes II

Uniforme

$X \sim \text{Unif}(a, b)$ (où $a < b$) : densité de probabilité égale à chaque valeur entre a et b sur la droite réelle.

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } a \leq x \leq b \\ 0, & \text{sinon} \end{cases}$$

Exponentielle

$X \sim \text{Exp}(\lambda)$ (où $\lambda > 0$) : densité de probabilité décroissante sur les réels non négatifs.

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & \text{sinon} \end{cases}$$

Espérance, Variance

Deux paramètres très importants :

- ▶ Espérance (localisation)

$$\mathbb{E}(X) = \begin{cases} \sum_i x_i p_i \\ \int_{-\infty}^{+\infty} x f(x) dx \end{cases}$$

- ▶ Variance (dispersion)

$$\mathbb{V}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \mathbb{E}(X - \mathbb{E}(X))^2$$

Exemples :

$$X \sim B(p), \quad \mathbb{E}(X) = p, \quad \mathbb{V}(X) = p(1-p)$$

$$X \sim \text{Bin}(n, p), \quad \mathbb{E}(X) = np, \quad \mathbb{V}(X) = np(1-p)$$

$$X \sim N(2, 3), \quad \mathbb{E}(X) = 2, \quad \mathbb{V}(X) = 3$$

Echantillon

Partant d'un échantillon de la population, on veut déduire des informations sur cette population.

P.b : Comment choisir une partie de la population qui reproduit le plus fidèlement possible ses caractéristiques?

Réponse : **Échantillonnage** (ne sera pas abordé dans ce cours)

Définition

Le n -uplet (X_1, \dots, X_n) formé de n variables aléatoires indépendantes et de même loi est appelé **n -échantillon** ou **échantillon** de taille n . La réalisation unique (x_1, \dots, x_n) de l'échantillon (X_1, \dots, X_n) est l'ensemble des **valeurs observées**.

Statistiques usuelles

Définition

Une statistique Y sur un échantillon (X_1, \dots, X_n) est une variable aléatoire, fonction mesurable des X_k ;

$$Y = f(X_1, \dots, X_n).$$

Les statistiques sont utilisées pour estimer les caractéristiques de la population totale. Les statistiques les plus utilisées sont la **moyenne empirique**, la **variance empirique**, la **fréquence empirique**.

Moyenne, Variance et Fréquence empiriques I

Posons $E(X) = \mu$, $V(X) = \sigma^2$ (inconnues)

Définition

On appelle moyenne empirique de l'échantillon (X_1, \dots, X_n) de X , la statistique

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Définition

On appelle variance empirique de l'échantillon (X_1, \dots, X_n) de X , la statistique

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2$$

Moyenne, Variance et Fréquence empiriques II

Soit une population comportant deux modalités A et B. Soit π la proportion d'individus de la population possédant la modalité A. $1 - \pi$ est donc la proportion des individus de la population possédant la modalité B.

On extrait de la population un échantillon de taille n . Soit K_n la v.a. qui représente le nombre d'individus dans l'échantillon ayant la modalité A.

Définition

La v.a. $F = \frac{K_n}{n}$ s'appelle **fréquence empirique**.

Sa réalisation f est la proportion d'individus dans l'échantillon ayant la modalité A.

Estimateurs I

Définition

On appelle estimateur de θ toute fonction

$T := f(X_1, X_2, \dots, X_N)$ servant à estimer θ .

Qualités d'un estimateur

On appelle biais de l'estimateur la fluctuation B définit par :

$$B(T) := \mathbb{E}(T) - \theta.$$

T est :

- ▶ sans biais si $B(T) = 0$;
- ▶ asymptotiquement sans biais si $\lim_{N \rightarrow +\infty} \mathbb{E}(T) = \theta$
- ▶ convergent si s.b. (ou a.s.b.) et $\lim_{N \rightarrow +\infty} \mathbb{V}(T) = 0$
- ▶ efficace si s.b. et de variance minimale.

Estimateurs II

Erreur quadratique moyenne

La précision d'un estimateur $\hat{\theta}$ est mesurée par l'erreur quadratique moyenne définie par :

$$\mathbb{E}[\hat{\theta} - \theta]^2.$$

Par conséquent, pour rendre un estimateur meilleur il suffit de minimiser son erreur quadratique. Un petit développement donne l'égalité suivante :

$$\mathbb{E}[\hat{\theta} - \theta]^2 = V(\hat{\theta}) + (\mathbb{E}[\hat{\theta}] - \theta)^2 = V(\hat{\theta}) + b(\theta)^2.$$

Ainsi, rendre cette erreur minimale revient à rendre minimal $V(\hat{\theta})$ et $b(\theta)$ (i.e estimateur sans biais). Un tel estimateur est appelé **estimateur efficace**.

Soient T et T' 2 estimateurs s.b.de θ . T plus efficace que T' si $\mathbb{V}(T') \geq \mathbb{V}(T)$.

Échantillon gaussien

on se donne un échantillon de lois normale de paramètres μ et σ^2 , i.e

- ▶ (X_1, X_2, \dots, X_N) i.i.d

Soient T et T' 2 estimateurs s.b.de θ . T plus efficace que T' si $\mathbb{V}(T') \geq \mathbb{V}(T)$.

Échantillon gaussien

on se donne un échantillon de lois normale de paramètres μ et σ^2 , i.e

- ▶ (X_1, X_2, \dots, X_N) i.i.d
- ▶ $X_i \sim \mathcal{N}(\mu, \sigma^2)$, pour $i = 1, 2, \dots, N$

Soient T et T' 2 estimateurs s.b.de θ . T plus efficace que T' si $\mathbb{V}(T') \geq \mathbb{V}(T)$.

Échantillon gaussien

on se donne un échantillon de lois normale de paramètres μ et σ^2 , i.e

- ▶ (X_1, X_2, \dots, X_N) i.i.d
- ▶ $X_i \sim \mathcal{N}(\mu, \sigma^2)$, pour $i = 1, 2, \dots, N$
- ▶ $\theta := (\mu, \sigma^2)$

Soient T et T' 2 estimateurs s.b.de θ . T plus efficace que T' si $\mathbb{V}(T') \geq \mathbb{V}(T)$.

Échantillon gaussien

on se donne un échantillon de lois normale de paramètres μ et σ^2 , i.e

- ▶ (X_1, X_2, \dots, X_N) i.i.d
- ▶ $X_i \sim \mathcal{N}(\mu, \sigma^2)$, pour $i = 1, 2, \dots, N$
- ▶ $\theta := (\mu, \sigma^2)$

Objectif : estimer μ et σ^2

Estimer une moyenne I

Théorème

La moyenne empirique définie par

$$\bar{X} := \frac{1}{N} \sum_{i=1}^N X_i \quad (1)$$

est un estimateur de la moyenne μ de l'échantillon.

propriétés

- ▶ $\mathbb{E}(\bar{X}) = \mu$ donc s.b.
- ▶ $\mathbb{V}(\bar{X}) = \frac{1}{N} \mathbb{V}(X) \rightarrow 0$ donc convergent;
- ▶ $\forall T$ un autre estimateur de μ , $\mathbb{V}(\bar{X}) \leq \mathbb{V}(T)$ donc \bar{X} efficace.
- ▶ Loi des grands Nombres : $\Rightarrow \frac{\bar{X} - \mu}{\sigma / \sqrt{N}}$ converge en loi vers $\mathcal{N}(0, 1)$

Estimer une moyenne II

- ▶ LFoGN $\Rightarrow \bar{X}$ converge presque sûrement vers μ
- ▶ LFaGN $\Rightarrow \frac{\bar{X} - \mu}{\sigma/\sqrt{N}}$ converge vers $\mathcal{N}(0, 1)$

Sensibilité par rapport aux données aberrantes!

Exemple

Une PME de 9 employés et un directeur à sa tête. On suppose que les 9 employés ont un salaire de 1500 euro/mois chacun et 3500 euro/mois pour le directeur. On voit bien que le $\bar{x} = \frac{1}{10}(9 \times 1500 + 3500) = 1700$

Estimer une variance I

Théorème

La variance empirique définie par

$$S^2 := \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 \quad (2)$$

est un estimateur de la variance σ^2 de l'échantillon.

propriétés

- ▶ $\mathbb{E}(S^2) = \frac{N-1}{N} \sigma^2$ (σ^2 variance théorique de l'échantillon);
- ▶ $\mathbb{V}(S^2) = \frac{N-1}{N^3} ((N-1)\mu_4 - (N-3)\sigma^4)$ avec μ_4 le moment d'ordre de X_i
- ▶ LFoGN $\Rightarrow S^2$ converge presque sûrement vers σ^2
- ▶ LFaGN $\Rightarrow \frac{S^2 - \frac{N-1}{N} \sigma^2}{\sqrt{\mathbb{V}(S^2)}}$ converge vers $\mathcal{N}(0, 1)$

Estimer une variance II

Exemple

Reprenant l'exemple précédent de ce PME on a

$$\bar{x} = 1700 \quad \Rightarrow \quad \bar{x}^2 = (1700)^2 = 2890000$$

On calcule la somme des carrés des observations :

$$\sum_{i=1}^{10} x_i^2 = 32500000$$

On en déduit la variance empirique

$$S^2 = \frac{1}{10} \sum_{i=1}^{10} x_i^2 - \bar{x}^2 = 360000$$

Références I

Courte bibliographie

- [1] Doucouré F. B. *Statistique et Probabilités pour économistes et gestionnaires : Tome 2 Estimation- Tests paramétriques et non paramétriques*, Edition Arima, (2010)
- [2] Graffigne C. *Cours de Statistique Fondamentale*, polycopié (2000).
- [3] Grammont L. *Cours de Statistiques inférentielles*, polycopié (2003).
- [4] Ramachandran K. M., Tsokos C. P. *Mathematical Statistics with Applications*, Academic Press, (2009).
- [5] Veysseyre R. *Aide mémoire : Statistique et Probabilité pour l'ingénieur*, 2^{ème} édition, Dunod, (2001)

Préambule

Rappels de lois
discrètes et continues

Statistiques usuelles &
Estimations
ponctuelles

Estimation par
intervalle de confiance

I.C pour une moyenne : premier
cas

I.C pour une moyenne :
deuxième cas

I.C pour une variance : Premier
cas

I.C pour une variance : Deuxième
cas

I.C pour une proportion

Tests de conformité

Références

Références II

Sites consultables

▶ [Wikipedia](#)

▶ [statistics](#)

▶ [math.umaine](#)

▶ [Introduction à R](#)