Z325EU07 - Probabilités et Statistiques

Théorèmes Limites et Estimation

Hervé Kerivin

Bureau : B133, Institut d'Informatique - ISIMA

Téléphone : 04 73 40 50 37

E-mail: herve.kerivin@uca.fr

Espérance - Cas Continu

Espérance d'une variable aléatoire continue

Soit X une variable aléatoire continue de densité f. L'espérance de X, noté $E[X]$, est le réel

$$E[X] = \int_{-\infty}^{+\infty} uf(u)du$$

à condition que cette intégrale converge absolument

Loi uniforme (cas continu)

Une variable aléatoire continue X suit une loi uniforme sur $[a, b]$, notée $\mathcal{U}(a, b)$, si sa densité de probabilité est $f(x) = \frac{1}{b-a}$ si $a \leq x \leq b$ et 0 sinon

Loi uniforme (cas continu) - Espérance et variance

1. $E[X] = \frac{a+b}{2}$
2. $\text{var}(X) = \frac{(b-a)^2}{12}$

Loi Normale

Loi normale

Une variable aléatoire X suit une loi normale de paramètres (m, σ), notée $\mathcal{N}(m, \sigma)$ si sa densité de problabilité est

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

- Loi normale aussi appelée loi de Laplace-Gauss
- Une des lois les plus importantes
- Pas de forme analytique pour sa fonction de répartition $\Rightarrow$ $F(x) = P(X \leq x)$ doit être lu dans un table (ou calculé par un logiciel)
- Courbe de la densité = courbe en forme de cloche (i.e., gaussienne)

Loi normale - espérance et variance

$E[X] = m$ et $\text{var}(X) = \sigma^2$

- Courbe de densité : symétrie par rapport à l'axe $x = m$; plus σ est grand, plus elle est ”étalée”

Loi Normale - Somme

Utilisation de la loi normale

- modélisation de phénomènes naturels issus d'évènements aléatoires
- description de la durée de vie d'une pièce en mécanique
- répartition des erreurs de mesures en physique
- poids des personnes en sociologie
- notes à un examen

Somme de variables aléatoires normales

Soient $X_1, \ldots, X_n$ des variables aléaroires mutuellement indépendantes qui suivent des lois normales de paramètres (m_i, σ_i), respectivement. Alors pour tous les réels $\alpha_1, \ldots, \alpha_n$, on a

$$\sum_{i=1}^{n} \alpha_i X_i \sim \mathcal{N}(\sum_{i=1}^{n} \alpha_i m_i, \sqrt{\sum_{i=1}^{n} \alpha_i^2 \sigma_i^2})$$

Variables Centrées Réduites

Variable centrée réduite

Soit X une variable aléatoire admettant une espérance et une variance non nulle. La variable

$$Y = \frac{X - E[X]}{\sigma_X}$$

est la variable centrée réduite associée à X

Variable centrée réduite - Espérance et variance

$E[Y] = 0$ et $\mathrm{var}(Y) = 1$

Loi Normale Centrée Réduite

- centrer une variable : soustraire son espérance à chaque valeur
- réduire une variable : diviser chaque valeur par son écart type
- données indépendantes de l'unité ou de l'échelle choisie
- variables ayant même moyenne et même dispersion.
- objectif : pouvoir mieux comparer les variations
- centrer-réduire : action souvent utilisée dans l'analyse de données

Loi normale centrée réduite

Une variable aléatoire X suit une loi normale centrée réduite, notée $\mathcal{N}(0, 1)$ si sa densité de problabilité est

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

- $F(0) = \frac{1}{2}$ (i.e., symétrique, de centre de symétrie 0)
- $F(-x) = 1 - F(x)$ pour tout $x \in \mathbb{R}$
- $-X$ suit une loi normale centrée réduite
- $X \sim \mathcal{N}(m, \sigma) \Rightarrow Z = \frac{X-m}{\sigma} \sim \mathcal{N}(0, 1)$
- $P(a \leq X \leq b) = P(\frac{a-m}{\sigma} \leq Z \leq \frac{b-m}{\sigma}) = F(\frac{b-m}{\sigma}) - F(\frac{a-m}{\sigma})$

Loi Normale Centrée Réduite - Propriétés

Propriétés de $\mathcal{N}(0,1)$

Soient $X \sim \mathcal{N}(0,1)$ et $x, x_1, x_2 \in \mathbb{R}$

(i) $P(X > x) = 1 - P(X < x) = 1 - F(x)$

(ii) $P(X < -x) = P(X > x) = 1 - F(x)$

(iii) $F(x) \leq \frac{1}{2} \iff x \leq 0; F(x) \geq \frac{1}{2} \implies x \geq 0$

(iv) $P(x_1 \leq X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$

(v) $P(|X| \leq x) = P(-x \leq X \leq x) = P(X \leq x) - P(X \leq -x) = 2F(x) - 1$

Fractiles

Soient X une variable aléatoire et $\alpha \in [0,1]$.

1. Le fractile supérieur d'ordre α de la loi de X est le réel x tel que $P(X \geq x) = \alpha$
2. Le fractile inférieur d'ordre α de la loi de X est le réel x tel que $P(X \leq x) = \alpha$

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0, 1)$

- $P(X \leq 1.79)$?

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0,1)$

- $P(X \leq 1.79)$?
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 1.79) = F(1.79) = 0.9633$
- $P(X \geq 1.25)$?

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0,1)$

- $P(X \leq 1.79)$?
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 1.79) = F(1.79) = 0.9633$
- $P(X \geq 1.25)$?
  - $P(X \geq 1.25) = 1 - P(X \leq 1.25) = 1 - F(1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \geq 1.25) = 1 - 0.8944 = 0.1056$
- $P(X \leq -1.34)$?

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0,1)$

- $P(X \leq 1.79)$?
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 1.79) = F(1.79) = 0.9633$
- $P(X \geq 1.25)$?
  - $P(X \geq 1.25) = 1 - P(X \leq 1.25) = 1 - F(1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \geq 1.25) = 1 - 0.8944 = 0.1056$
- $P(X \leq -1.34)$?
  - $P(X \leq -1.34) = 1 - P(X \leq 1.34) = 1 - F(1.34)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq -1.34) = 1 - 0.9099 = 0.0901$
- $P(-1.25 \leq X \leq 1.79)$?

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0,1)$

- $P(X \leq 1.79)$?
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 1.79) = F(1.79) = 0.9633$
- $P(X \geq 1.25)$?
  - $P(X \geq 1.25) = 1 - P(X \leq 1.25) = 1 - F(1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \geq 1.25) = 1 - 0.8944 = 0.1056$
- $P(X \leq -1.34)$?
  - $P(X \leq -1.34) = 1 - P(X \leq 1.34) = 1 - F(1.34)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq -1.34) = 1 - 0.9099 = 0.0901$
- $P(-1.25 \leq X \leq 1.79)$?
  - $P(-1.25 \leq X \leq 1.79) = P(X \leq 1.79) - P(X \leq -1.25)$
  - $P(-1.25 \leq X \leq 1.79) = P(X \leq 1.79) - 1 + P(X \leq 1.25) = F(1.79) + F(1.25) - 1$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(-1.25 \leq X \leq 1.79) = 0.9633 + 0.8944 - 1 = 0.8577$
- $P(-1.25 \leq X \leq 1.25)$?

Loi Normale Centrée Réduite - Lecture dans une Table

Soit $X \sim \mathcal{N}(0,1)$

- $P(X \leq 1.79)$?
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 1.79) = F(1.79) = 0.9633$
- $P(X \geq 1.25)$?
  - $P(X \geq 1.25) = 1 - P(X \leq 1.25) = 1 - F(1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \geq 1.25) = 1 - 0.8944 = 0.1056$
- $P(X \leq -1.34)$?
  - $P(X \leq -1.34) = 1 - P(X \leq 1.34) = 1 - F(1.34)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq -1.34) = 1 - 0.9099 = 0.0901$
- $P(-1.25 \leq X \leq 1.79)$?
  - $P(-1.25 \leq X \leq 1.79) = P(X \leq 1.79) - P(X \leq -1.25)$
  - $P(-1.25 \leq X \leq 1.79) = P(X \leq 1.79) - 1 + P(X \leq 1.25) = F(1.79) + F(1.25) - 1$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(-1.25 \leq X \leq 1.79) = 0.9633 + 0.8944 - 1 = 0.8577$
- $P(-1.25 \leq X \leq 1.25)$?
  - $P(-1.25 \leq X \leq 1.25) = 2P(X \leq 1.25) - 1 = 2F(1.25) - 1$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(-1.25 \leq X \leq 1.79) = 2 * 0.8944 - 1 = 0.7888$

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.8204$
- x tel que $P(X < x) = 0.23$

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.8204$
- x tel que $P(X < x) = 0.23$
  - $0.23 < 0.5 \Longrightarrow x < 0$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = -0.7388$
- x tel que $P(X \geq x) = 0.025$?

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.8204$
- x tel que $P(X < x) = 0.23$
  - $0.23 < 0.5 \implies x < 0$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = -0.7388$
- x tel que $P(X \geq x) = 0.025$?
  - x tel que $P(X < x) = 0.975$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 1.9600$
- x tel que $P(-1.25 < X \leq x) = 0.5$?

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.8204$
- x tel que $P(X < x) = 0.23$
  - $0.23 < 0.5 \Longrightarrow x < 0$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = -0.7388$
- x tel que $P(X \geq x) = 0.025$?
  - x tel que $P(X < x) = 0.975$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 1.9600$
- x tel que $P(-1.25 < X \leq x) = 0.5$?
  - $P(-1.25 < X \leq x) = P(X \leq x) - P(X < -1.25) = P(X \leq x) - 1 + P(X < 1.25)$
  - $P(X \leq x) = 0.5 + 1 - P(X < 1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq x) = 1.5 - 0.8944 = 0.6056$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.2689$
- x tel que $P(|X| \leq x) = 0.970$?

Loi Normale Centrée Réduite - Fractile et Table

Soit $X \sim \mathcal{N}(0,1)$

- x tel que $P(X \leq x) = 0.794$?
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.8204$
- x tel que $P(X < x) = 0.23$
  - $0.23 < 0.5 \Longrightarrow x < 0$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = -0.7388$
- x tel que $P(X \geq x) = 0.025$?
  - x tel que $P(X < x) = 0.975$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 1.9600$
- x tel que $P(-1.25 < X \leq x) = 0.5$?
  - $P(-1.25 < X \leq x) = P(X \leq x) - P(X < -1.25) = P(X \leq x) - 1 + P(X < 1.25)$
  - $P(X \leq x) = 0.5 + 1 - P(X < 1.25)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq x) = 1.5 - 0.8944 = 0.6056$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 0.2689$
- x tel que $P(|X| \leq x) = 0.970$?
  - $P(|X| \leq x) = P(-x \leq X \leq x) = 2P(X \leq x) - 1$
  - $P(X \leq x) = \frac{0.97+1}{2} = 0.985$
  - Lecture dans la table des fractiles de $\mathcal{N}(0,1)$: $x = 2.1701$

Loi Normale - Exemples

Soit X une variable aléatoire

- $X \sim \mathcal{N}(2, 3)$, $P(X \leq 4.79)$?

Loi Normale - Exemples

Soit X une variable aléatoire

- $X \sim \mathcal{N}(2,3)$, $P(X \leq 4.79)$?
  - $Y = \frac{X-2}{3} \sim \mathcal{N}(0,1)$
  - $P(X \leq 4.79) = P(Y \leq \frac{4.79-2}{3}) = P(Y \leq 0.76)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 4.79) = 0.7764$
- $X \sim \mathcal{N}(-5,7)$, $P(X < 2.35)$?

Loi Normale - Exemples

Soit X une variable aléatoire

- $X \sim \mathcal{N}(2,3)$, $P(X \leq 4.79)$?
  - $Y = \frac{X-2}{3} \sim \mathcal{N}(0,1)$
  - $P(X \leq 4.79) = P(Y \leq \frac{4.79-2}{3}) = P(Y \leq 0.76)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 4.79) = 0.7764$
- $X \sim \mathcal{N}(-5,7)$, $P(X < 2.35)$?
  - $Y = \frac{X+5}{7} \sim \mathcal{N}(0,1)$
  - $P(X < 2.35) = P(Y < \frac{2.35+5}{7}) = P(Y < 1.05)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X < 2.35) = 0.8531$
- $X \sim \mathcal{N}(15,3)$, $P(9 \leq X < 24)$?

Loi Normale - Exemples

Soit X une variable aléatoire

- $X \sim \mathcal{N}(2,3)$, $P(X \leq 4.79)$?
  - $Y = \frac{X-2}{3} \sim \mathcal{N}(0,1)$
  - $P(X \leq 4.79) = P(Y \leq \frac{4.79-2}{3}) = P(Y \leq 0.76)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X \leq 4.79) = 0.7764$
- $X \sim \mathcal{N}(-5,7)$, $P(X < 2.35)$?
  - $Y = \frac{X+5}{7} \sim \mathcal{N}(0,1)$
  - $P(X < 2.35) = P(Y < \frac{2.35+5}{7}) = P(Y < 1.05)$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(X < 2.35) = 0.8531$
- $X \sim \mathcal{N}(15,3)$, $P(9 \leq X < 24)$?
  - $Y = \frac{X-15}{3} \sim \mathcal{N}(0,1)$
  - $P(9 \leq X < 20) = P(\frac{9-15}{3} \leq Y < \frac{24-15}{3}) = P(-2 \leq Y < 3)$
  - $P(9 \leq X < 20) = P(Y < 3) + P(Y < 2) - 1$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0,1)$
  - $P(9 \leq X < 20) = 0.99865 + 0.9772 - 1 = 0.97585$

Loi Normale - Exemples (suite)

Soit X une variable aléatoire

- $X \sim \mathcal{N}(-3, 4)$, x tel que $P(-4.8 \leq X < x) = 0.529$?

Loi Normale - Exemples (suite)

Soit X une variable aléatoire

- $X \sim \mathcal{N}(-3, 4)$, x tel que $P(-4.8 \leq X < x) = 0.529$?
  - $Y = \frac{X+3}{4} \sim \mathcal{N}(0, 1)$
  - $P(-4.8 \leq X < x) = P(\frac{-4.8+3}{4} \leq Y < \frac{x+3}{4}) = P(-0.45 \leq Y < \frac{x+3}{4})$
  - $P(-0.45 \leq Y < \frac{x+3}{4}) = P(Y < \frac{x+3}{4}) + P(Y \leq 0.45) - 1 = 0.529$
  - Lecture dans la table de la fonction de répartition de $\mathcal{N}(0, 1)$
  - $P(Y < \frac{x+3}{4}) = 0.529 + 1 - 0.6736 = 0.8554$
  - Lecture dans la table des fractiles de $\mathcal{N}(0, 1)$
  - $\frac{x+3}{4} = 1.0581$
  - $x = 1.2324$

Central Limit Theorem: Poly_Tunis (F18) pp. 37-41 for some exercises

Inégalité de Markov

- X : variable aléatoire non-negative
- $t > 0$
- $E[X] = \sum_{x \geq t} xP(X = x) + \sum_{x < t} xP(X = x) \geq \sum_{x \geq t} xP(X = x)$
- $E[X] \geq t \sum_{x \geq t} P(X = x) = tP(X \geq t)$

Inégalité de Markov

Soient X une variable aléatoire non-négative admettant une espérance (finie) et $t > 0$. Alors

$$P(X \geq t) \leq \frac{E[X]}{t}$$

- Probabilité que X soit au moins k fois son espérance est au plus $\frac{1}{k}$
- Pas d'information nécessaire sur la variance ou la loi de X

Inégalité de Markov - Exemples

- Une pièce a une probabilité de 20% de tomber sur Face. La pièce est lancée 20 fois. Trouver une borne supérieure pour la probabilité d'avoir au moins 16 Face

Inégalité de Markov - Exemples

- Une pièce a une probabilité de 20% de tomber sur Face. La pièce est lancée 20 fois. Trouver une borne supérieure pour la probabilité d'avoir au moins 16 Face
  - $\Omega = \{Pile, Face\}^{20}$
  - X: nombre de Face obtenus, $X(\Omega) = \{0, 1, \dots, 20\}$
  - $X \sim \mathcal{B}(20, \frac{1}{5})$
  - $E[X] = 20.\frac{1}{5} = 4$
  - Inégalité de Markov $P(X \geq 16) \leq \frac{E[X]}{16} = \frac{1}{4}$
  - $P(X \geq 16) = \sum_{k=16}^{20} \binom{20}{16} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{20-k} \approx 1.38.10^{-8}$
- Soit X une variable aléatoire telle que $P(X = 0) = \frac{24}{25}$ and $P(X = 5) = \frac{1}{25}$. Trouver une borne supérieure pour $P(X \geq 5)$

Inégalité de Markov - Exemples

- Une pièce a une probabilité de 20% de tomber sur Face. La pièce est lancée 20 fois. Trouver une borne supérieure pour la probabilité d'avoir au moins 16 Face
  - $\Omega = \{Pile, Face\}^{20}$
  - X: nombre de Face obtenus, $X(\Omega) = \{0, 1, \ldots, 20\}$
  - $X \sim \mathcal{B}(20, \frac{1}{5})$
  - $E[X] = 20.\frac{1}{5} = 4$
  - Inégalité de Markov $P(X \geq 16) \leq \frac{E[X]}{16} = \frac{1}{4}$
  - $P(X \geq 16) = \sum_{k=16}^{20} \binom{20}{16} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{20-k} \approx 1.38.10^{-8}$
- Soit X une variable aléatoire telle que $P(X = 0) = \frac{24}{25}$ and $P(X = 5) = \frac{1}{25}$. Trouver une borne supérieure pour $P(X \geq 5)$
  - $X(\Omega) = \{0, 5\}$
  - $E[X] = 0.\frac{24}{25} + 5.\frac{1}{25} = \frac{1}{5}$
  - Inégalité de Markov $P(X \geq 5) \leq \frac{E[X]}{5} = \frac{1}{25} = P(X = 5)$

Inégalité de Bienaymé-Tchebychev

- Soient X variable aléatoire et $Y = (X - E[X])^2$
- Y est non-negative et $E[Y] = \text{var}(X)$
- $t > 0$
- Inégalité de Markov $P(Y \geq t^2) \leq \frac{E[Y]}{t^2} = \frac{\text{var}(X)}{t^2}$
- $(Y \geq t^2) = (|X - E[X]| \geq t)$
- $P(|X - E[X]| \geq t) \leq \frac{\text{var}(X)}{t^2}$

Théorème - Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire admettant une espérance et une variance. Alors pour tout $\epsilon > 0$

$$P(|X - E[X]| \geq \epsilon) \leq \frac{\text{var}(X)}{\epsilon^2}$$

- Probabilité pour que X se trouve à l'extérieur de l'intervalle centré en $E[X]$ et de rayon ϵ est majorée par $\frac{\text{var}(X)}{\epsilon^2}$
- Si $\epsilon^2 \leq \text{var}(X)$ alors on trouve 1 comme majorant

$\Longrightarrow$ efficacité de l'inégalité vient de ϵ^2 grand devant $\text{var}(X)$

Inégalité de Bienaymé-Tchebychev - Exemples

- Une pièce non-faussée est lancée 100 fois. Trouver une borne supérieure pour la probabilité que le nombre de Face obtenus soit au moins 60 ou au plus 40.

Inégalité de Bienaymé-Tchebychev - Exemples

- Une pièce non-faussée est lancée 100 fois. Trouver une borne supérieure pour la probabilité que le nombre de Face obtenus soit au moins 60 ou au plus 40.
  - $\Omega = \{Pile, Face\}^{100}$
  - X : nombre de Face obtenus, $X(\Omega) = \{0, 1, \dots, 100\}$
  - $X \sim \mathcal{B}(100, \frac{1}{2})$, $E[X] = 100.\frac{1}{2} = 50$, $\mathrm{var}(X) = 100.\frac{1}{2}.\frac{1}{2} = 25$
  - $P((X \leq 40) \cup (X \geq 60)) = P((X - 50 \leq -10) \cup (X - 50 \geq 10)) = P(|X - 50| \geq 10)$
  - Inégalité de Bienaymé-Tchebychev $P(|X - 50| \geq 10) \leq \frac{\mathrm{var}(X)}{10^2} = \frac{25}{100} = \frac{1}{4}$
- Exemple précédent de la pièce lancée 20 fois
  - $P(X \geq 16) = P((X \geq 16) \cup (X \leq -8)) = P((X - 4 \geq 12) \cup (X - 4 \leq -12)) = P(|X - 4| \geq 12)$
  - Inégalité de Bienaymé-Tchebychev
    $P(|X - 4| \geq 12) \leq \frac{\mathrm{var}(X)}{12^2} = \frac{20.\frac{1}{5}\cdot\frac{4}{5}}{144} = \frac{1}{45}$

Moyenne Empirique

- Fréquence relative d'un évènement = manière intuitive de voir une probabilité
- Probabilité (modèle mathématique) = valeur d'une fonction de répartition d'une variable aléatoire représentant l'évènement

Moyenne empirique

Soient $X_1, \ldots, X_n$ des variables aléatoires (deux à deux) indépendantes, de même loi, d'espérance m et de variance σ^2. La moyenne empirique de $X_1, \ldots, X_n$ est la variable aléatoire

$$\overline{X_n} = \frac{X_1 + \ldots + X_n}{n}$$

- $E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n]$
- $\text{var}(X_1 + \ldots + X_n) = \text{var}(X_1) + \ldots + \text{var}(X_n)$

Moyenne empirique - espérance et variance

$E[\overline{X_n}] = m$ et $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$

Loi Faible des Grands Nombres

Loi faible des grands nombres

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes, de même loi, d'espérance m et de variance σ^2. Quand n est grand, alors $\overline{X_n}$ est proche de m, c'est-à-dire

$$\lim_{n\to\infty} P(|\overline{X_n} - m| > \epsilon) = 0 \quad \forall\, \epsilon > 0$$

- Inégalité de Bienaymé-Tchebychev : $P(|\overline{X_n} - m| > \epsilon) \leq \frac{\mathrm{var}(\overline{x_n})}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2 n}$
- $\overline{X_n}$ convergne en probabilité vers m
- Interpération : avec un échantillon assez large, il y a une probabilité très élevée que la moyenne des observations soit proche de l'espérance

Loi Faible des Grands Nombres - Loi de Bernoulli

- Loi de Bernoulli de paramètre p
- X_i variable aléatoire telle que $X_i = 1$ si la $i^{\text{ème}}$ observation est un succès, 0 sinon
- $E[X_i] = p$
- Loi faible des grands nombres : pour tout $\epsilon > 0$, $P(|\overline{X_n} - p| < \epsilon) \to 1$ quand $n \to \infty$
- Interprétation : Pour un grand nombre d' expérience de Bernoulli, on peut espérer que la proportion de réalisation d'un évènement soit proche de p

$\Rightarrow$ modèle mathématique des probabilité est conforme à l'interprétation de la fréquence relative

Loi Faible des Grands Nombres - Exemple

Lancer d'un dé *n* fois

Loi Faible des Grands Nombres - Exemple

Lancer d'un dé n fois

- X_i variable aléatoire telle que $X_i = 1$ si Face est obtenu, 0 sinon
- $X_i \sim \mathcal{B}(1, \frac{1}{2})$
- $S_n = \sum_{i=1}^{n} X_i$:i variable aléatoire indiquant le nombre de Face obtenus
- $\overline{X_n} = \frac{S_n}{n} \in [0, 1]$: moyenne empirique (i.e., proportion de Face obtenus)
- Loi faible des grands nombres : pour un grand nombre n de lancers, la valeur de $\overline{X_n}$ sera très proche de $\frac{1}{2}$, i.e.,

$$P(|\overline{X_n} - \frac{1}{2}| < \epsilon) \to 1 \quad \text{quand } n \to \infty$$
$$P(|\overline{X_n} - \frac{1}{2}| \geq \epsilon) \to 0 \quad \text{quand } n \to \infty$$

- Valeurs empiriques de $P(|\overline{X_n} - \frac{1}{2}| \leq \frac{5}{100}) = P(\frac{45}{100} \leq \overline{X_n} < \frac{55}{100})$?

Loi Faible des Grands Nombres - Exemple

Lancer d'un dé n fois

- X_i variable aléatoire telle que $X_i = 1$ si Face est obtenu, 0 sinon
- $X_i \sim \mathcal{B}(1, \frac{1}{2})$
- $S_n = \sum_{i=1}^{n} X_i$:i variable aléatoire indiquant le nombre de Face obtenus
- $\overline{X_n} = \frac{S_n}{n} \in [0, 1]$: moyenne empirique (i.e., proportion de Face obtenus)
- Loi faible des grands nombres : pour un grand nombre n de lancers, la valeur de $\overline{X_n}$ sera très proche de $\frac{1}{2}$, i.e.,

$$P(|\overline{X_n} - \frac{1}{2}| < \epsilon) \to 1 \quad \text{quand } n \to \infty$$
$$P(|\overline{X_n} - \frac{1}{2}| \geq \epsilon) \to 0 \quad \text{quand } n \to \infty$$

- Valeurs empiriques de $P(|\overline{X_n} - \frac{1}{2}| \leq \frac{5}{100}) = P(\frac{45}{100} \leq \overline{X_n} < \frac{55}{100})$?

$n = 10$: 0.2460937
$n = 25$: 0.309962
$n = 50$: 0.5201123
$n = 75$: 0.6443008

$n = 100$: 0.728747
$n = 250$: 0.9099566
$n = 500$: 0.9846362
$n = 750$: 0.9963632

$n = 1000$: 0.9992216
$n = 2500$: 0.9999997
$n = 5000$: 1

Loi Forte des Grands Nombres

Loi forte des grands nombres

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes, de même loi, d'espérance m et de variance σ^2. Quand n tend vers l'infini, alors la probabilité que la moyenne des observations soit égale à m vaut 1, c'est-à-dire

$$P(\lim_{n \to \infty} \overline{X_n} = m) = 1$$

- Interprétation : quand le nombre n d'essais tend vers l'infini, la moyenne des observations converge vers l' espérance
- Pour une variable aléatoire X dont on ne connait pas la loi, la loi des grands nombres permet d'avoir une idée plus ou moins précise de la loi de X

Théorème Central Limite

Deux questions

- Qu'est-ce qu'un grand nombre ?
- Que veut dire proche de m ?

Théorème central limite

Soient $X_1, \dots, X_n$ des variables aléatoires mutuellement indépendantes, de même loi, d'espérance m et de variance σ^2. La loi de $\overline{X_n}$ tend vers la loi normale $\mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$ quand n tend vers l'infini, ou encore la loi de la variable aléatoire centrée réduite $\frac{\overline{X_n}-m}{\sigma/\sqrt{n}}$ converge vers $\mathcal{N}(0,1)$ quand n tend vers l'infini, i.e., pour tout $z \in \mathbb{R}$

$$\lim_{n\to\infty} P\left(\frac{\overline{X_n}-m}{\sigma/\sqrt{n}} \leq z\right) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} dx$$

Théorème le plus important de la statistique

- permet d'approximer la sum ou la moyenne de variables aléatoires indépendantes et identiquement distribuées par un variable aléatoire de loi mormale
- extrêmement utile car il est généralement facile de faire des calculs avec la loi normale

Théorème Central Limite

- $\overline{X_n} \approx \mathcal{N}(m, \frac{\sigma}{\sqrt{n}})$ pour des grandes valeurs de n
  - même espérance
  - plus petite variance pour $\overline{X_n}$ (que pour X_i)
- $S_n = \sum_{i=1}^{n} X_i \approx \mathcal{N}(nm, \sigma\sqrt{n})$ pour des grandes valeurs de n
- $Z_n = \frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} = \frac{S_n - nm}{\sigma\sqrt{n}} \approx \mathcal{N}(0, 1)$ pour des grandes valeurs de n

Théorème Central Limite - Exemples

Considérons une pièce équilibrée qui est lancée n fois

- X_i : résultat du $i^{\text{ème}}$ lancer
- $X_i = 1$ si Face est obtenu, 0 sinon, $X_i \sim \mathcal{B}(1, \frac{1}{2})$
- $E[X_i] = \frac{1}{2}$, $\text{var}(X_i) = \frac{1}{4}$
- Estimer la probabilité d'obtenir au moins 55 Face pour $n = 100$

Théorème Central Limite - Exemples

Considérons une pièce équilibrée qui est lancée n fois

- X_i : résultat du $i^{\text{ème}}$ lancer
- $X_i = 1$ si Face est obtenu, 0 sinon, $X_i \sim \mathcal{B}(1, \frac{1}{2})$
- $E[X_i] = \frac{1}{2}$, $\text{var}(X_i) = \frac{1}{4}$
- Estimer la probabilité d'obtenir au moins 55 Face pour $n = 100$
  - $E[S_{100}] = 100E[X_1] = 100\frac{1}{2} = 50$
  - $\text{var}(S_{100}) = 100\text{var}(X_i) = 100.\frac{1}{4} = 25$, $\sigma_{S_{100}} = 5$
  - Théorème Central Limite : $\frac{S_{100} - E[S_{100}]}{\sigma_{S_{100}}} \approx \mathcal{N}(0, 1)$
  - $P(S_{100} > 55) = P(\frac{S_{100} - 50}{5} < \frac{55-50}{5}) = P(Z_{100} > 1)$
  - $P(S_{100} > 55) = 1 - P(Z_{100} \leq 1) = 1 - 0.8413 = 0.1587$
- Estimer la probabilité d'obtenir au moins 220 Face pour $n = 400$

Théorème Central Limite - Exemples

Considérons une pièce équilibrée qui est lancée n fois

- X_i : résultat du $i^{\text{ème}}$ lancer
- $X_i = 1$ si Face est obtenu, 0 sinon, $X_i \sim \mathcal{B}(1, \frac{1}{2})$
- $E[X_i] = \frac{1}{2}$, $\text{var}(X_i) = \frac{1}{4}$
- Estimer la probabilité d'obtenir au moins 55 Face pour $n = 100$
  - $E[S_{100}] = 100E[X_1] = 100\frac{1}{2} = 50$
  - $\text{var}(S_{100}) = 100\text{var}(X_i) = 100.\frac{1}{4} = 25$, $\sigma_{S_{100}} = 5$
  - Théorème Central Limite : $\frac{S_{100} - E[S_{100}]}{\sigma_{S_{100}}} \approx \mathcal{N}(0, 1)$
  - $P(S_{100} > 55) = P(\frac{S_{100}-50}{5} < \frac{55-50}{5}) = P(Z_{100} > 1)$
  - $P(S_{100} > 55) = 1 - P(Z_{100} \leq 1) = 1 - 0.8413 = 0.1587$
- Estimer la probabilité d'obtenir au moins 220 Face pour $n = 400$
  - $E[S_{400}] = 200$, $\text{var}(S_{400}) = 100$, $\sigma_{S_{400}} = 10$
  - Théorème Central Limite : $\frac{S_{400} - E[S_{400}]}{\sigma_{S_{400}}} \approx \mathcal{N}(0, 1)$
  - $P(S_{400} > 220) = P(\frac{S_{400}-200}{10} < \frac{220-200}{10}) = P(Z_{400} > 2)$
  - $P(S_{400} > 220) = 1 - P(Z_{400} \leq 2) = 1 - 0.9772 = 0.0228$
- $\frac{55}{100} = \frac{220}{400}$ mais $P(S_{100} > 55) > P(S_{400} > 220)$ (à cause de la loi faible des grands nombres)

Théorème Central Limite - Exemples (suite)

La somme des résultats de 10000 lancers d'un même dé est 35487. Ce dé est-il truqué ?

Théorème Central Limite - Exemples (suite)

La somme des résultats de 10000 lancers d'un même dé est 35487. Ce dé est-il truqué ?

- Suite $(X_i)_{i\geq 1}$ de variables aléatoires indépendantes de même loi uniforme sur $\{1,\dots,6\}$ (i.e., $X_i \sim \mathcal{U}(6)$)
- $E[X_i] = \frac{6+1}{2} = \frac{7}{2}$, $\text{var}(X_i) = \frac{6^2-1}{12} = \frac{35}{12}$
- $E[S_n] = nE[X_1] = \frac{7n}{2}$ et $\text{var}(S_n) = n\text{var}(X_1) = \frac{35n}{12}$
- Théorème central limite : $\frac{S_n - E[S_n]}{\sqrt{\text{var}(S_n)}} \sim \mathcal{N}(0,1)$ quand $n \to \infty$
- Approximation pour $n = 10000$: $\frac{S_{10000} - \frac{7*10000}{2}}{\sqrt{\frac{35*10000}{12}}}$ égale à $\mathcal{N}(0,1)$
- Pour tout $z > 0$, on a $P(-z \leq \frac{S_{10000}-35000}{50\sqrt{\frac{35}{3}}} \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-z}^{z} e^{-\frac{x^2}{2}} dx$
- $z = 2.93$ (car $2.93 * 50 * \sqrt{\frac{35}{3}} = 500.39$)
- $P(35000 - 500.39 \leq S_{10000} \leq 35000 + 500.39) = 2F(2.93) - 1 \approx 0.9964$
- Somme obtenue est compatible avec un dé pas truqué

Théorème Central Limite - Converge

Loi binomiale par loi de Poisson (rappel)

On peut approcher la loi binomiale $\mathcal{B}(n, p)$ par la loi de poisson$\mathcal{P}(np)$

Approximation acceptable si

- $np \leq 10$
- n est grand (e.g., $n \geq 50$)

Approximation de la loi binomiale

On peut approcher la loi binomial $\mathcal{B}(n, p)$ par la loi normale $\mathcal{N}(np, \sqrt{np(1-p)})$

Approximation acceptable si

- $n > 30$
- $np > 5$
- $np(1-p) > 5$

Approximation de la loi de Poisson

On peut approcher la loi $\mathcal{P}(\lambda)$ par la loi normale $\mathcal{N}(n\lambda, \sqrt{n\lambda})$

Approximation acceptable si $n\lambda > 15$

Théorème Central Limit (version 2)

Théorème central limite

Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes, de même loi, d'espérance m_i et de variance σ_i^2, respectivement. La loi de la variable aléatoire $\frac{\sum_{i=1}^{n}(X_i - m_i)}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}$ converge vers $\mathcal{N}(0, 1)$ quand n tend vers l'infini

- Si les lois des X_i sont proches d'une loi normale, alors pour tout $n \geq 4$, le théorème central limite donne une bonne approximation
- Si les lois des X_i sont relativement proches d'une loi normale (e.g., loi uniforme), alors pour tout $n \geq 12$, le théorème central limite donne une bonne approximation
- Si les lois des X_i ne sont pas proches d'une loi normale, alors pour tout $n \geq 100$, le théorème central limite donne une bonne approximation

Estimateur

Dans la pratique, on ne connait pas forcément les valeurs de tous les paramètres du système que l'on étudie. On a parfois besoin d'estimer ces valeurs

Estimateur

Soit $(X_1, \ldots, X_n)$ un échantillons de n observations d'une varaible aléatoire X. La fonction de l'échantillon qui estimera un paramètre θ est appelé estimateur ; son écart-type est appelé erreur standard, noté se

Estimateur = fonction de $X_1, \ldots, X_n$

Estimateur sans biais

Un estimateur T d'un paramètre θ est dit sans biais si $E[T] = \theta$

La variance d'un estimateur sans biais T mesure l'écart entre les valeurs de T et θ ($\mathrm{var}(T) = E[(T - E[T])^2] = E[(T - \theta)^2]$)

Estimateur efficace

Un estimateur T d'un paramètre θ est dit efficace si c'est un estimateur sans biais de variance minimum, i.e., $\sigma^2_T \leq \sigma^2_{T'}$ pour tout estimateur sans biais T' de θ

Estimateur - Moyenne Empirique

Soit $(X_1, \ldots, X_n)$ un échantillons de n observations (indépendantes) d'une variable aléatoire X dont l'espérance est m

Moyenne empirique

La moyenne empirique

$$\overline{X_n} = \frac{\sum_{i=1}^{n} X_i}{n}$$

est un estimateur sans biais pour l'espérance m

- $\overline{X_n} = \frac{S_n}{n}$
- $E[\overline{X_n}] = E[\frac{S_n}{n}] = \frac{1}{n}E[\sum_{i=1}^{n} X_i]$
- $E[\overline{X_n}] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{1}{n}nm = m$

Erreur standard de $\overline{X_n}$?

- $\text{var}(\overline{X_n}) = \text{var}(\frac{S_n}{n}) = \frac{1}{n^2}\text{var}(S_n) = \frac{1}{n^2}n\text{var}(X) = \frac{\text{var}(X)}{n}$
- $\text{se}(\overline{X_n}) = \sqrt{\frac{\text{var}(X)}{n}}$

Estimateur - Variance Empirique

Soit $(X_1, \ldots, X_n)$ un échantillons de n observations (indépendantes) d'une variable aléatoire X dont l'espérance est m et l' écart type σ_X

Variance empirique

La variance empirique

$$s^2_{n-1} = \frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{n-1}$$

est un estimateur sans biais pour la variance σ^2_X

- $E[s^2_{n-1}] = E[\frac{\sum_{i=1}^{n}(X_i - \overline{X_n})^2}{n-1}] = \frac{1}{n-1}E[\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X_n}^2 + \overline{X_n}^2)]$
- $E[s^2_{n-1}] = \frac{1}{n-1}E[\sum_{i=1}^{n} X_i^2 - 2\overline{X_n}\sum_{i=1}^{n} X_i + \sum_{i=1}^{n}\overline{X_n}^2] = \frac{1}{n-1}E[\sum_{i=1}^{n} X_i^2 - n\overline{X_n}^2]$
- $E[s^2_{n-1}] = \frac{1}{n-1}(n[E[X_1^2] - nE[\overline{X_n}^2])$
- $E[X_1^2] = E[X_1^2 - 2X_1\overline{X_n} + \overline{X_n}^2 + 2X_1\overline{X_n} - \overline{X_n}^2] = \sigma^2_X + m^2$
- $E[s^2_{n-1}] = \frac{1}{n-1}(n(\sigma^2_X + m^2) - n(\frac{\sigma^2_X}{n} + m^2)) = \sigma^2_X$

Intervalle de Confiance

Estimation par intervalle

L'estimation par intervalle permet de déterminer un intervalle contenant la vraie valeur du paramètre avec une certaine probabilité d'erreur

Intervalle de confiance

Étant donné $0 < \alpha < 1$, un intervalle $[a, b]$ est dit intervalle de confiance à $(1 - \alpha)100\%$ d'un paramètre θ si $P(a \leq \theta \leq b) = 1 - \alpha$

Coefficient de confiance

$1 - \alpha$ est appelé le coefficient (niveau) de confiance

- Un intervalle de confiance est un intervalle qui a une probabilité connue de contenir un paramètre inconnu
- Un intervalle de confiance a la forme estimateur $\pm$ marge d'erreur
  - augmentation de la taille de l'échantillon (i.e., n) $\Longrightarrow$ marge d'erreur diminue
  - augmentation du coefficient de confiance $\Longrightarrow$ augmentation de la marge d'erreur
- On dit qu'on est sûr à $(1 - \alpha)100\%$ que le paramètre est dans l'intervalle

Estimation d'une Espérance

- Soit X une variable aléatoire de moyenne m et d'écart type σ_X
- On cherche à estimer m à partir d'un échantillon $(X_1, \ldots, X_n)$ de n observations
- Deux cas : variance σ_X^2 est connue ou pas

Fractiles (rappel)

Soit Z une variable aléatoire. Le fractile supérieur d'ordre α de la loi de Z est le réel z qui vérifie $P(Z \geq z) = \alpha$. Le fractile inférieur d'ordre α de la loi de Z est le réel z qui vérifie $P(Z \leq z) = \alpha$.

Estimation d'une Espérance - variance connue

Si X suit une loi normale de variance σ_X^2 connue, un intervalle de confiance pour l' espérance, de niveau de confiance $1 - \alpha$, est

$$\left[\overline{X_n} - z_{\frac{\alpha}{2}} \frac{\sigma_X}{\sqrt{n}}, \overline{X_n} + z_{\frac{\alpha}{2}} \frac{\sigma_X}{\sqrt{n}}\right]$$

où $z_{\frac{\alpha}{2}}$ est le fractile supérieur d'ordre $\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0,1)$

Théorème central limit: si X ne suit pas une loi normale, alors $\frac{\overline{X_n} - m}{\frac{\sigma_X}{\sqrt{n}}}$ est approximé par la loi normale $\mathcal{N}(0,1)$ quand n est suffisamment grand (≥ 30)

Estimation d'une Espérance - Exemple

Un échantillon aléatoire de 225 cours de probabilités et statistiques a été sélectionné sur les 5 dernières années et le nombre d'étudiants absents pour chacuns de ces cours a été enregistré. Nous avons obtenu une moyenne empirique de 11.6 et un écart-type de 4.1. Estimer le nombre moyen d'absences par cours sur les 5 dernières années avec un niveau de confiance de 90%

Estimation d'une Espérance - Exemple

Un échantillon aléatoire de 225 cours de probabilités et statistiques a été sélectionné sur les 5 dernières années et le nombre d'étudiants absents pour chacuns de ces cours a été enregistré. Nous avons obtenu une moyenne empirique de 11.6 et un écart-type de 4.1. Estimer le nombre moyen d'absences par cours sur les 5 dernières années avec un niveau de confiance de 90%

- $n = 225$ suffisamment grand : approximation par la loi normale centrée réduite
- niveau de confiance = 0.9 $\Longrightarrow \alpha = 0.1$
- fractile supérieur d'ordre 0.05 : $z_{0.05} = 1.6449$
- intervalle de confiance à 90% : $\left[11.6 - 1.6449\frac{4.1}{\sqrt{225}}, 11.6 + 1.6449\frac{4.1}{\sqrt{225}}\right] = [11.15, 12.05]$
- Incorrect de dire qu'il y a une probabilité de 0.9 que m soit entre 11.15 et 12.05 (cette probabilité est soit 1 ou 0)
- Incorrect de dire que tous les cours ont entre 11.15 et 12.05 étudiants absents
- Correct de dire que le nombre moyen d'absences est entre 11.15 et 12.05 avec un niveau de confiance de 90%

Estimation d'une Espérance - Variance Inconnue

Loi de Student

La variable $\frac{\overline{X_n}-m}{\frac{S_{n-1}}{\sqrt{n}}}$ suit une loi de Student à $(n-1)$ degrés de liberté, notée $\mathcal{T}_{n-1}$

- densité de la loi de Student est symmétrique (comme $\mathcal{N}(0,1)$)
- tables pour obtenir les fractiles de la loi de Student
- Loi de Student à ν degrés de liberté est approximativement la loi normale centrée réduite quand ν est grand (> 30)

Estimation d'une Espérance - variance inconnue

Si X suit une loi normale de variance inconnue, un intervalle de confiance pour l' espérance, de niveau de confiance $1-\alpha$, est

$$\left[\overline{X_n} - t_{n-1,\frac{\alpha}{2}}\frac{s_{n-1}}{\sqrt{n}}, \overline{X_n} + t_{n-1,\frac{\alpha}{2}}\frac{s_{n-1}}{\sqrt{n}}\right]$$

où $t_{n-1,\frac{\alpha}{2}}$ est le fractile supérieur d'ordre $\frac{\alpha}{2}$ de la loi de Student $\mathcal{T}_{n-1}$

Estimation d'une Espérance - Variance Inconnue - Exemple

Un échantillon de 20 sacs de 5kg de pommes-de-terre, provenant d'une exploitation agricole, a une moyenne empirique égale à 5.12 kg et un écart-type empirique (i.e., erreur standard) égale à 0.14 kg. Donner un intervalle de confiance à 95% du poids moyen de tous les sacs produits par cette exploitation.

Estimation d'une Espérance - Variance Inconnue - Exemple

Un échantillon de 20 sacs de 5kg de pommes-de-terre, provenant d'une exploitation agricole, a une moyenne empirique égale à 5.12 kg et un écart-type empirique (i.e., erreur standard) égale à 0.14 kg. Donner un intervalle de confiance à 95% du poids moyen de tous les sacs produits par cette exploitation.

- $n = 20$
- on suppose que la distribution des poids des sacs est normale
- $\overline{X_{20}} = 5.14$ et $s_{19} = 0.14$
- niveau de confiance = 0.95 $\Longrightarrow \alpha = 0.05$
- fractile supérieur d'ordre 0.025 : $t_{19,0.025} = 2.093$
- intervalle de confiance à 95% : $\left[5.12 - 2.093\frac{0.14}{\sqrt{20}}, 5.12 + 2.093\frac{0.14}{\sqrt{20}}\right] = [5.054, 5.186]$

Estimation d'une Variance

- Soit X une variable aléatoire de moyenne m et d'écart type σ_X
- On cherche à estimer σ_X^2 à partir d'un échantillon $(X_1, \ldots, X_n)$ de n observations
- Deux cas : espérance m est connue ou pas

Loi du chi-deux

Soit $(Y_1, \ldots, Y_n)$ un échantillon de n observations d'une variable aléatoire Y de loi $\mathcal{N}(0, 1)$. On appelle loi du chi-deux à n degrés de liberté la loi de la variable aléatoire $\sum_{i=1}^{n} Y_i^2$ et on note $\chi^2_{(n)}$

- densité de la loi du chi-deux n'est pas symmétrique (contrairement à $\mathcal{N}(0, 1)$ et $\mathcal{T}_\nu$)
- tables pour obtenir les fractiles de la loi du chi-deux

Espérance inconnue

La variable aléatoire $(n-1)\frac{s^2_{n-1}}{\sigma^2_X}$ suit la loi du chi-deux à $(n-1)$ degrés de liberté

Estimation d'une Variance - Espérance Inconnue

Estimation d'une variance - Espérance Inconnue

Si X suit une loi normale d'espérance inconnue, un intervalle de confiance pour la variance, de niveau de confiance $1 - \alpha$, est

$$\left[\frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)s_{n-1}^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right]$$

où $\chi^2_{n-1,1-\frac{\alpha}{2}}$ et $\chi^2_{n-1,\frac{\alpha}{2}}$ sont les fractiles supérieurs d'ordre $1-\frac{\alpha}{2}$ et $\frac{\alpha}{2}$, respectivement, de la loi $\chi^2_{(n-1)}$

Estimation d'une Variance - Espérance Inconnue - Exemple

Échantillon $\{78, 85, 91, 76\}$ d'une variable aléatoire normale X. Donner un intervalle de confiance à 95% pour l'écart-type de X.

Estimation d'une Variance - Espérance Inconnue - Exemple

Échantillon $\{78, 85, 91, 76\}$ d'une variable aléatoire normale X. Donner un intervalle de confiance à 95% pour l'écart-type de X.

- $n = 4$, $\overline{X_4} = 82.5$ et $s_3^2 = 47$
- $\alpha = 0.05$
- fractiles : $\chi^2_{3,0.975} = 9.348$ et $\chi^2_{3,0.025} = 0.216$
- intervalle de confiance à 95% de la variance : $\left[\frac{3*47}{9.348}, \frac{3*47}{0.216}\right] = [15.083, 652.778]$
- intervalle de confiance à 95% de l'écart-type : $\left[\sqrt{\frac{3*47}{9.348}}, \sqrt{\frac{3*47}{0.216}}\right] = [3.884, 25.549]$
- taille de ce dernier intervalle $\Longrightarrow$ manque de précision de l'écart-type (à cause de la taille de l'échatillon)

Estimation d'une Variance - Espérance Connue

L'espérance m de X est connue

Espérance connue

La variable aléatoire $\dfrac{\sum_{i=1}^{n}(X_i - m)^2}{\sigma_X^2}$ suit une loi du chi-deux à n degrés de liberté

Estimation d'une variance - Espérance connue

Si X suit une loi normale d'espérance m connue, un intervalle de confiance pour la variance, de niveau de confiance $1 - \alpha$, est

$$\left[\frac{\sum_{i=1}^{n}(X_i - m)^2}{\chi^2_{n,1-\frac{\alpha}{2}}}, \frac{\sum_{i=1}^{n}(X_i - m)^2}{\chi^2_{n,\frac{\alpha}{2}}} \right]$$

où $\chi^2_{n,1-\frac{\alpha}{2}}$ et $\chi^2_{n,\frac{\alpha}{2}}$ sont les fractiles supérieurs d'ordre $1 - \frac{\alpha}{2}$ et $\frac{\alpha}{2}$, respectivement, de la loi $\chi^2_{(n)}$

Si l'approximation par la loi normale ne peut pas être faite, alors il n'est pas possible de déterminer un intervalle de confiance

Estimation d'une Proportion

- X suit une loi de Bernoulli de paramètre p
- Estimation de p
- Échantillon $X_1, \ldots, X_n$ de n obervations de X

Estimateur de p

La moyenne empirique $\overline{X_n} = \frac{S_n}{n}$ est une estimateur sans biais de p

$\mathcal{N}(0, 1)$

La variable aléatoire $\frac{\overline{X_n} - p}{\sqrt{\frac{\overline{X_n}(1 - \overline{X_n})}{n}}}$ suit une loi normale centrée réduite.

Estimation de p

Un intervalle de confiance pour le paramètre p, de niveau de confiance $1 - \alpha$, est

$$\left[\overline{X_n} - z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X_n}(1 - \overline{X_n})}{n}}, \overline{X_n} + z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{X_n}(1 - \overline{X_n})}{n}} \right]$$

où $z_{\frac{\alpha}{2}}$ est le fractile supérieur d'ordre $\frac{\alpha}{2}$ de la loi normale $\mathcal{N}(0, 1)$

Estimation d'une Proportion - Exemple

Chaque année, les étudiants en première année à l'UCA peuvent choisir de suivre le cours d'algorithmique. Afin d'estimer la proportion d'étudiants en permière année qui étudie l'algorithmique, un échantillon de 1000 étudiants sur les 10 dernières années a été choisi et 637 de ces étudiants ont suivi le cours d'algorithmique. Donner un intervalle de confiance à 95% de cette proportion

Estimation d'une Proportion - Exemple

Chaque année, les étudiants en première année à l'UCA peuvent choisir de suivre le cours d'algorithmique. Afin d'estimer la proportion d'étudiants en permière année qui étudie l'algorithmique, un échantillon de 1000 étudiants sur les 10 dernières années a été choisi et 637 de ces étudiants ont suivi le cours d'algorithmique. Donner un intervalle de confiance à 95% de cette proportion

- Loi de Bernoulli : succès (étudie algorithmique)
- Loi binomiale de paramètre p et $n = 1000$.
- $\overline{X_{1000}} = \frac{637}{1000}$
- niveau de confiance = 0.95 $\Longrightarrow \alpha = 0.05$
- quantile supérieur d'ordre 0.025 : $z_{0.025} = 1.96$
- intervalle de confiance à 95% :
$$\left[\frac{637}{1000} - 1.96 * \sqrt{\frac{.637*(1-.637)}{1000}}, \frac{637}{1000} + 1.96 * \sqrt{\frac{.637*(1-.637)}{1000}}\right] = [.607, .667]$$