Probabilités - Statistiques

David Loiseaux et Fanny Simões

1er Mars 2022

Contenu

Contenu

Modèle mathématique – Tribu

Soit Ω un ensemble. Un sous ensemble $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ des parties de Ω, est une *tribu* (ou σ-algèbre) si

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3. $\forall n \in \mathbb{N}, A_n \in F \implies \bigcup_n A_n \in \mathcal{F}$.

Propriétés

Modèle mathématique – Mesure de probabilité

Une mesure de probabilité sur $(\Omega, \mathcal{F})$ est une application $\mathbb{P} \colon \mathcal{F} \longrightarrow [0, 1]$ telle que

1. $\mathbb{P}(\Omega) = 1$, et
2. (σ-additivité) pour toute famille dénombrable $A_n \in \mathcal{F}$ d'éléments 2 à 2 disjoints

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n) \tag{1}$$

On dit que $(\Omega, \mathcal{F}, \mathbb{P})$ est un *espace probabilisé*.

Propriétés

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé,

1. $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,

Propriétés

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé,

1. $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(\varnothing) = 0$,

Propriétés

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé,

1. $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(\varnothing) = 0$,
3. $A \subseteq B \in \mathcal{F} \implies \mathbb{P}(A) \leq \mathbb{P}(B)$

Propriétés

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé,

1. $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(\emptyset) = 0$,
3. $A \subseteq B \in \mathcal{F} \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
4. (σ-sous-additivité) $\forall n \in \mathbb{N}, A_n \in \mathcal{F}, \mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$,

Propriétés

Si $(\Omega, \mathcal{F}, \mathbb{P})$ est un espace probabilisé,

1. $\forall A \in \mathcal{F}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$,
2. $\mathbb{P}(\varnothing) = 0$,
3. $A \subseteq B \in \mathcal{F} \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
4. (σ-sous-additivité) $\forall n \in \mathbb{N}, A_n \in \mathcal{F}, \mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$,
5. $\forall n \in \mathbb{N}, A_n \subseteq A_{n+1} \in \mathcal{F} \implies P(\cup_n A_n) = \lim_n \mathbb{P}(A_n)$,
6. $\forall n \in \mathbb{N}, B_n \supseteq B_{n+1} \in \mathcal{F} \implies P(\cap_n B_n) = \lim_n \mathbb{P}(B_n)$

Modèle mathématique – Variable aléatoire

Version « intuitive » :
Une variable aléatoire est un « objet aléatoire », qui prend des valeurs quelque part (souvent $\mathbb{R}$) et qui doit pouvoir « s'exprimer avec $\mathcal{F}$ ».

Modèle mathématique – Variable aléatoire

Version « intuitive » :
Une variable aléatoire est un « objet aléatoire », qui prend des valeurs quelque part (souvent $\mathbb{R}$) et qui doit pouvoir « s'exprimer avec $\mathcal{F}$ ».
Version matheuse :
Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Une variable aléatoire X de cet espace probabilisé est une fonction $\mathcal{F}$-mesurable $X : \Omega \to \mathcal{E}$, où $(E, \mathcal{E})$ est un espace mesuré.

Cas fini – Mesure uniforme

On suppose ici $|\Omega| = n < +\infty$, et $\mathcal{F} = \mathcal{P}(\Omega)$. La probabilité uniforme d'un évènement A est défini par

$$\mathbb{P} \colon A \in \mathcal{F} \longmapsto \frac{|A|}{|\Omega|}.$$

Exemple : On lance 2 dés,
$A := \{\omega \in \Omega : \text{ la somme des points est } 4\}$

$$\mathcal{P}(A) = \frac{|\{(1,3),(2,2),(3,1)\}|}{|\Omega|} = \frac{3}{36} = \frac{1}{12}.$$

Arrangements, combinaisons

Soit E un ensemble fini ; et $p \in \mathbb{N}^*$.
Un p-arrangement de E est un p-uplet $(x_1, \ldots, x_p)$ tel que $i \neq j \implies x_i \neq x_j$.
$\longrightarrow$ Tirage ordonné sans remise.

Arrangements, combinaisons

Soit E un ensemble fini ; et $p \in \mathbb{N}^*$.
Un p-arrangement de E est un p-uplet $(x_1, \ldots, x_p)$ tel que
$i \neq j \implies x_i \neq x_j$.
$\longrightarrow$ Tirage ordonné sans remise.
Le nombre de p-arrangements de E est :

Arrangements, combinaisons

Soit E un ensemble fini ; et $p \in \mathbb{N}^*$.
Un p-arrangement de E est un p-uplet $(x_1, \ldots, x_p)$ tel que $i \neq j \implies x_i \neq x_j$.
$\longrightarrow$ Tirage ordonné sans remise.
Le nombre de p-arrangements de E est :

$$A_n^p = n \times (n-1) \times \cdots \times (n-p+1) = \frac{n!}{(n-p)!}$$

Arrangements, combinaisons

Soit E un ensemble fini ; et $p \in \mathbb{N}^*$.
Un p-arrangement de E est un p-uplet $(x_1, \ldots, x_p)$ tel que $i \neq j \implies x_i \neq x_j$.
$\longrightarrow$ Tirage ordonné sans remise.
Le nombre de p-arrangements de E est :

$$A_n^p = n \times (n-1) \times \cdots \times (n-p+1) = \frac{n!}{(n-p)!}$$

Une p-combinaison de E est un p-ensemble $\{x_1, \ldots, x_p\}$.
$\longrightarrow$ Tirage non-ordonné sans remise.
Le nombre de p-combinaisons de E est

Arrangements, combinaisons

Soit E un ensemble fini ; et $p \in \mathbb{N}^*$.
Un p-arrangement de E est un p-uplet $(x_1, \ldots, x_p)$ tel que $i \neq j \implies x_i \neq x_j$.
$\longrightarrow$ Tirage ordonné sans remise.
Le nombre de p-arrangements de E est :

$$A_n^p = n \times (n-1) \times \cdots \times (n-p+1) = \frac{n!}{(n-p)!}$$

Une p-combinaison de E est un p-ensemble $\{x_1, \ldots, x_p\}$.
$\longrightarrow$ Tirage non-ordonné sans remise.
Le nombre de p-combinaisons de E est

$$\binom{n}{p} = \frac{n!}{p!(n-p)!}$$

Double urne

Probabilités conditionnelles

Soient $A, B \in \mathcal{F}$ deux évènements, avec $\mathbb{P}(B) > 0$. La probabilité $\mathbb{P}(A|B)$ de A sachant B est défini par

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Probabilités conditionnelles

Soient $A, B \in \mathcal{F}$ deux évènements, avec $\mathbb{P}(B) > 0$. La probabilité $\mathbb{P}(A|B)$ de A sachant B est défini par

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

$\longrightarrow$ $\mathbb{P}(\cdot|B)$ est une mesure de probabilité.

Probabilités conditionnelles

Soient $A, B \in \mathcal{F}$ deux évènements, avec $\mathbb{P}(B) > 0$. La probabilité $\mathbb{P}(A|B)$ de A sachant B est défini par

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

$\longrightarrow$ $\mathbb{P}(\cdot|B)$ est une mesure de probabilité.
Remarque : On peut aussi définir une mesure de probabilité sachant une variable aléatoire, ou une tribu, mais c'est plus compliqué.

Probabilités totales

Si $\Omega = \bigcup_n B_n$, on a

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \cap B_n)$$

Probabilités totales

Si $\Omega = \bigcup_n B_n$, on a

$$\mathbb{P}(A) = \sum_n \mathbb{P}(A \cap B_n) = \sum_n \mathbb{P}(B_n)\mathbb{P}(A|B_n).$$

Indépendance

- A et B sont indépendant $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Indépendance

- A et B sont indépendant $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- Deux variables aléatoires X et Y sont indépendantes si pour tout $A, B \in \mathcal{F}$,

$$\mathbb{P}(X \in A \text{ et } Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Exemples

Espérances

Si X est une variable aléatoire pouvant prendre un nombre fini (ou dénombrable) de valeurs, par exemple $X \in \{x_1, \ldots, x_n\}$, on définit l'espérance de X **lorsque c'est bien défini** par

$$\mathbb{E}(X) := \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k).$$

Espérances

Si X est une variable aléatoire pouvant prendre un nombre fini (ou dénombrable) de valeurs, par exemple $X \in \{x_1, \dots, x_n\}$, on définit l'espérance de X **lorsque c'est bien défini** par

$$\mathbb{E}(X) := \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k).$$

$\mathbb{E}(X) = 0 \Leftrightarrow X$ est centrée.

Espérances

Si X est une variable aléatoire pouvant prendre un nombre fini (ou dénombrable) de valeurs, par exemple $X \in \{x_1, \ldots, x_n\}$, on définit l'espérance de X **lorsque c'est bien défini** par

$$\mathbb{E}(X) := \sum_{k=1}^{n} x_k \mathbb{P}(X = x_k).$$

$\mathbb{E}(X) = 0 \Leftrightarrow X$ est centrée.
Plus généralement, lorsque X est intégrable par rapport à $\mathbb{P}$, on définit

$$\mathbb{E}(X) := \int_\Omega X(\omega) \mathrm{d}\mathbb{P}(\omega) = \int_E x \mathrm{d}\mathbb{P}_X(x) = \int_E x f(x) \mathrm{d}\mathbb{P}(x)$$

si f est la densité de $\mathbb{P}_X$ par rapport à $\mathbb{P}$.

Propriétés

Lorsque bien défini (ou positif),

1. (Linéarité) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,

Propriétés

Lorsque bien défini (ou positif),

1. (Linéarité) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
2. (Linéarité) $\alpha \in \mathbb{R} \implies \mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$,

Propriétés

Lorsque bien défini (ou positif),

1. (Linéarité) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
2. (Linéarité) $\alpha \in \mathbb{R} \implies \mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$,
3. (Inégalité de Jensen) Si φ est une fonction convexe sur $\mathbb{R}$ et X est réelle,
$$\varphi\left(\mathbb{E}(X)\right) \leq \mathbb{E}\left(\varphi(X)\right).$$
Typiquement,
$$|\mathbb{E}(X)| \leq \mathbb{E}|X|.$$

Propriétés

Lorsque bien défini (ou positif),

1. (Linéarité) $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$,
2. (Linéarité) $\alpha \in \mathbb{R} \implies \mathbb{E}(\alpha X) = \alpha \mathbb{E}(X)$,
3. (Inégalité de Jensen) Si φ est une fonction convexe sur $\mathbb{R}$ et X est réelle,
$$\varphi\left(\mathbb{E}(X)\right) \leq \mathbb{E}\left(\varphi(X)\right).$$
Typiquement,
$$|\mathbb{E}(X)| \leq \mathbb{E}|X|.$$
4. (Inégalité de Hölder) Lorsque bien défini, si $\frac{1}{p} + \frac{1}{q} = 1$, on a
$$\mathbb{E}(|XY|) \leq \mathbb{E}\left(|X|^p\right)^{1/p} \mathbb{E}\left(|Y|^q\right)^{1/q}.$$

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.
Propriétés

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.

Propriétés

1. $\mathrm{var}(aX) = a^2\mathrm{var}(X)$

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.

Propriétés

1. $\mathrm{var}(aX) = a^2\mathrm{var}(X)$
2. $\mathrm{var}(X + a) = \mathrm{var}(X)$

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.

Propriétés

1. $\mathrm{var}(aX) = a^2\mathrm{var}(X)$
2. $\mathrm{var}(X + a) = \mathrm{var}(X)$
3. si X et Y sont indépendants $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

Variance, écart type

$\longrightarrow$ Pareil que l'espérance mais au carré et recentré.

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2.$$

X est dite *réduite* si $\mathrm{var}(X) = 1$.

Propriétés

1. $\mathrm{var}(aX) = a^2\mathrm{var}(X)$
2. $\mathrm{var}(X + a) = \mathrm{var}(X)$
3. si X et Y sont indépendants $\mathrm{var}(X + Y) = \mathrm{var}(X) + \mathrm{var}(Y)$

L'écart type de X est la racine de la variance

$$\sqrt{\mathrm{var}(X)}$$

Exemple

Loi (forte) des grands nombres

Soit X une variable aléatoire réelle, et $(X_i)_{i\in\mathbb{N}}$ *iid* de même loi que X. Alors

$$X \text{ intégrable} \Leftrightarrow \overline{X}_n \xrightarrow[n\to\infty]{} \mathbb{E}(X) \text{ presque sûrement;}$$

où $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ est l'estimateur de la moyenne empirique de X.

Application : Monté-Carlo

TODO

Théorème limite central

Si $(X_n)_n$ est une suite de loi iid admettant une variance, alors

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \mathbb{E}(X) = \sqrt{n}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{var}(X)).$$

Théorème limite central

Si $(X_n)_n$ est une suite de loi iid admettant une variance, alors

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i - \mathbb{E}(X) = \sqrt{n}(\overline{X}_n - \mathbb{E}(X)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{var}(X)).$$

$\sqrt{n}$ est moralement la bonne vitesse car

$$\mathrm{var}\left(\sqrt{n}\overline{X}_n\right) = \mathrm{var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{var}(X_i) = \mathrm{var}(X).$$

Contenu

Notebook

Contenu

Modèle statistique paramétrique

But : Étant donné un échantillon $(X_i)_{1\leq i\leq n}$ iid de même loi que X,
$\longrightarrow$ Qui est X sachant $(X_i)_i$?

Modèle statistique paramétrique

But : Étant donné un échantillon $(X_i)_{1\leq i\leq n}$ iid de même loi que X,
$\longrightarrow$ Qui est X sachant $(X_i)_i$?
Modèle : On va supposer que X suit une loi parmi une famille de loi indexée par un paramètre $\theta \in \mathbb{R}^n$ fixée à l'avance. On note $\mathcal{P} = (\mathcal{P}_\theta)_\theta$ cette famille de loi.

Modèle statistique paramétrique

But : Étant donné un échantillon $(X_i)_{1\leq i\leq n}$ iid de même loi que X,
$\longrightarrow$ Qui est X sachant $(X_i)_i$?
Modèle : On va supposer que X suit une loi parmi une famille de loi indexée par un paramètre $\theta \in \mathbb{R}^n$ fixée à l'avance. On note $\mathcal{P} = (\mathcal{P}_\theta)_\theta$ cette famille de loi.
Exemple :

1. $X \sim \mathcal{N}(-12, 78^2)$ ou uniforme sur $[-32, 14]$
   $\longrightarrow \mathcal{P}_0 = \mathcal{N}(-12, 78^2)$ et $\mathcal{P}_1 = \mathcal{U}_{[-32,14]}$.

Modèle statistique paramétrique

But : Étant donné un échantillon $(X_i)_{1\leq i\leq n}$ iid de même loi que X,
$\longrightarrow$ Qui est X sachant $(X_i)_i$?
Modèle : On va supposer que X suit une loi parmi une famille de loi indexée par un paramètre $\theta \in \mathbb{R}^n$ fixée à l'avance. On note $\mathcal{P} = (\mathcal{P}_\theta)_\theta$ cette famille de loi.
Exemple :

1. $X \sim \mathcal{N}(-12, 78^2)$ ou uniforme sur $[-32, 14]$
   $\longrightarrow \mathcal{P}_0 = \mathcal{N}(-12, 78^2)$ et $\mathcal{P}_1 = \mathcal{U}_{[-32,14]}$.
2. X suit une loi exponentielle dont on ne connaît pas la moyenne :
   $\longrightarrow \mathcal{P}_{\mathrm{m}} = \mathcal{E}(m)$.

Modèle statistique paramétrique

But : Étant donné un échantillon $(X_i)_{1\leq i\leq n}$ iid de même loi que X,
$\longrightarrow$ Qui est X sachant $(X_i)_i$?
Modèle : On va supposer que X suit une loi parmi une famille de loi indexée par un paramètre $\theta \in \mathbb{R}^n$ fixée à l'avance. On note $\mathcal{P} = (\mathcal{P}_\theta)_\theta$ cette famille de loi.
Exemple :

1. $X \sim \mathcal{N}(-12, 78^2)$ ou uniforme sur $[-32, 14]$
   $\longrightarrow \mathcal{P}_0 = \mathcal{N}(-12, 78^2)$ et $\mathcal{P}_1 = \mathcal{U}_{[-32,14]}$.
2. X suit une loi exponentielle dont on ne connaît pas la moyenne :
   $\longrightarrow \mathcal{P}_{\mathrm{m}} = \mathcal{E}(m)$.
3. X suit une loi normale dont on ne connaît pas la moyenne, ni la variance :
   $\longrightarrow \mathcal{P}_{\mu,\sigma^2} = \mathcal{N}(\mu, \sigma^2)$.

Hypothèses (cas paramétrique)

Trouver $X \Leftrightarrow$ trouver le paramètre θ *tq* $X \sim \mathcal{P}_\theta$.

Hypothèses (cas paramétrique)

Trouver $X \Leftrightarrow$ trouver le paramètre θ *tq* $X \sim \mathcal{P}_\theta$.
Souvent, pour simplifier on teste des hypothèses : H_0 VS H_1

$$H_0 : \ll \ \theta \in A \ \gg \quad \textbf{VS} \quad H_1 : \ll \ \theta \in B \ \gg$$

Hypothèses (cas paramétrique)

Trouver $X \Leftrightarrow$ trouver le paramètre θ *tq* $X \sim \mathcal{P}_\theta$.
Souvent, pour simplifier on teste des hypothèses : H_0 VS H_1

$$H_0 : \ll \ \theta \in A \ \gg \quad \textbf{VS} \quad H_1 : \ll \ \theta \in B \ \gg$$

et on peut regarder

		Réalité	
		H_0	H_1
Décision	H_0	OK	Erreur de 2ème espèce
	H_1	Erreur de 1ère espèce	OK

Exemple

p-valeur

Si on choisit H_0 ou H_1 en fonction d'un test $T(X_1, \ldots, X_n) \in [0, 1]$

p-valeur

Si on choisit H_0 ou H_1 en fonction d'un test $T(X_1, \ldots, X_n) \in [0, 1]$ avec

- $T \approx 1$ on choisit H_1,
- $T \approx 0$ on choisit H_0,
- $0 << T << 1$ on choisit un seuil à l'avance.

Version matheuse :

$$p(x) = \sup_{\theta \text{ tq } H_0 \text{ est vrai}} \mathbb{P}\left(T(\tilde{X}_1, \ldots, \tilde{X}_n) \geq T(x) \text{ avec } \tilde{X}_i \sim \mathcal{P}_\theta \text{ iid}\right)$$

p-valeur

Si on choisit H_0 ou H_1 en fonction d'un test $T(X_1, \dots, X_n) \in [0, 1]$ avec

- $T \approx 1$ on choisit H_1,
- $T \approx 0$ on choisit H_0,
- $0 << T << 1$ on choisit un seuil à l'avance.

Version matheuse :

$$p(x) = \sup_{\theta \text{ tq } H_0 \text{ est vrai}} \mathbb{P}\left(T(\tilde{X}_1, \dots, \tilde{X}_n) \geq T(x) \text{ avec } \tilde{X}_i \sim \mathcal{P}_\theta \text{ iid}\right)$$

Version intuitive : La p-valeur d'un échantillon $x = (x_1, \dots, x_n)$ = la proba de faire un choix plus extrême (selon T) que x

p-valeur

Si on choisit H_0 ou H_1 en fonction d'un test $T(X_1, \ldots, X_n) \in [0, 1]$ avec

- $T \approx 1$ on choisit H_1,
- $T \approx 0$ on choisit H_0,
- $0 << T << 1$ on choisit un seuil à l'avance.

Version matheuse :

$$p(x) = \sup_{\theta \text{ tq } H_0 \text{ est vrai}} \mathbb{P}\left(T(\tilde{X}_1, \ldots, \tilde{X}_n) \geq T(x) \text{ avec } \tilde{X}_i \sim \mathcal{P}_\theta \text{ iid}\right)$$

Version intuitive : La p-valeur d'un échantillon $x = (x_1, \ldots, x_n)$ = la proba de faire un choix plus extrême (selon T) que x
Attention aux mauvaises interpretations !
Ex : Moralement, si H_0 est vrai, on a $p(X_1, \ldots, X_n) \sim \mathcal{U}_{[0,1]}$!

Lois utiles :

- Loi normale.

Lois utiles :

- Loi normale.
- Loi du χ^2 (Chi-deux) :

$$\chi^2(n) \sim \left(\mathcal{N}(0,1)^2\right)^{\otimes n} \sim X_1^2 + \cdots + X_n^2,$$

où $(X_i)_i$ est iid $\sim \mathcal{N}(0,1)$.

Lois utiles :

- Loi normale.
- Loi du χ^2 (Chi-deux) :

$$\chi^2(n) \sim \left(\mathcal{N}(0,1)^2\right)^{\otimes n} \sim X_1^2 + \cdots + X_n^2,$$

  où $(X_i)_i$ est iid $\sim \mathcal{N}(0,1)$.
- Loi de Student $\mathcal{T}$ à d degrés de liberté.

$$\mathcal{T}(d) \sim \ll \frac{\mathcal{N}(0,1)}{\sqrt{\chi^2(d)/d}} \gg \sim \frac{X}{\sqrt{\left(Y_1^2 + \cdots + Y_d^2\right)/d}}$$

  où X et les Y_i sont des lois normales indépendantes.

Tests d'espérance, variance connue

On considère un échantillon $X_1, \dots, X_n$ ($\in L^2$).

$$H_0 : m = m_0 \quad \text{VS} \quad H_1 : m \neq m_0$$

Tests d'espérance, variance connue

On considère un échantillon $X_1, \dots, X_n$ ($\in L^2$).

$$H_0 : m = m_0 \quad \text{VS} \quad H_1 : m \neq m_0$$

Test : $T = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m_0\right)$

Tests d'espérance, variance connue

On considère un échantillon $X_1, \dots, X_n$ ($\in L^2$).

$$H_0 : m = m_0 \quad \text{VS} \quad H_1 : m \neq m_0$$

Test : $T = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m_0\right)$

Sous H_0 : $T \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.

Sous H_1 : $|T| \xrightarrow[n\to\infty]{\text{p.s.}} +\infty$.

Tests d'espérance, variance connue

On considère un échantillon $X_1, \dots, X_n$ ($\in L^2$).

$$H_0 : m = m_0 \quad \text{VS} \quad H_1 : m \neq m_0$$

Test : $T = \frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m_0\right)$
Sous H_0 : $T \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}(0,1)$.
Sous H_1 : $|T| \xrightarrow[n\to\infty]{\text{p.s.}} +\infty$.

$\longrightarrow$ On approxime $T \approx \mathcal{N}(0,1)$ et on regarde si la p-valeur est en dessous du seuil qu'on s'est fixé avant (souvent 5%, ou 1%).

Tests d'espérance, variance inconnue

Modèle $\mathcal{P} = \left\{\mathcal{N}(m, \sigma^2) : m \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\right\}$
$\longrightarrow$ On doit approximer σ.

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{1 \leq i \leq n} (X_i - \overline{X}_n)^2 \sim \frac{1}{n-1} \sum_{1 \leq i \leq n} \sigma^2 \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim {}^{1}\sigma^2 \chi^2(n-1)$$

On considère alors le test

$$T = \frac{\sqrt{n}}{\hat{\sigma}_n} \left(\overline{X}_n - m_0\right) = \frac{\sigma}{\hat{\sigma}_n} \cdot \frac{\sqrt{n}}{\sigma} \left(\overline{X}_n - m_0\right)$$

1. Théorème de Cochran

Tests d'espérance, variance inconnue

Modèle $\mathcal{P} = \left\{\mathcal{N}(m, \sigma^2) : m \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\right\}$
$\longrightarrow$ On doit approximer σ.

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{1 \leq i \leq n} (X_i - \overline{X}_n)^2 \sim \frac{1}{n-1} \sum_{1 \leq i \leq n} \sigma^2 \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim {}^{1}\sigma^2 \chi^2(n-1)$$

On considère alors le test

$$T = \frac{\sqrt{n}}{\hat{\sigma}_n} \left(\overline{X}_n - m_0\right) = \frac{\sigma}{\hat{\sigma}_n} \cdot \frac{\sqrt{n}}{\sigma} \left(\overline{X}_n - m_0\right)$$

Sous $H_0 : T \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{T}(n-1)$ (Thm de Slutsky + Thm de Cochran).
Sous $H_1 : |T| \xrightarrow[n \to \infty]{\text{p.s.}} +\infty$.

1. Théorème de Cochran

Tests d'espérance, variance inconnue

Modèle $\mathcal{P} = \left\{\mathcal{N}(m,\sigma^2) : m \in \mathbb{R}, \sigma^2 \in \mathbb{R}_+\right\}$
$\longrightarrow$ On doit approximer σ.

$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{1\le i\le n}(X_i - \overline{X}_n)^2 \sim \frac{1}{n-1}\sum_{1\le i\le n}\sigma^2\frac{(X_i-\overline{X}_n)^2}{\sigma^2} \sim {}^1\sigma^2\chi^2(n-1)$$

On considère alors le test

$$T = \frac{\sqrt{n}}{\hat{\sigma}_n}\left(\overline{X}_n - m_0\right) = \frac{\sigma}{\hat{\sigma}_n}\cdot\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m_0\right)$$

Sous $H_0 : T \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{T}(n-1)$ (Thm de Slutsky + Thm de Cochran).
Sous $H_1 : |T| \xrightarrow[n\to\infty]{\text{p.s.}} +\infty$.
$\longrightarrow$ On approxime $T \approx \mathcal{T}(n-1)$ et on regarde si la p-valeur est en dessous du seuil qu'on s'est fixé avant (souvent 5%, ou 1%).

1. Théorème de Cochran

Tests de variance, espérance connue

Modèle $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2) : \sigma^2 \in \mathbb{R}_+\right\}$

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 < \sigma_0^2$$

Tests de variance, espérance connue

Modèle $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2) : \sigma^2 \in \mathbb{R}_+\right\}$

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 < \sigma_0^2$$

Test :

$$T = \frac{1}{\sigma_0^2}\overline{(X - \mathbb{E}(X))^2}_n = \frac{1}{\sigma_0^2}\left[(X_1 - \mathbb{E}(X))^2 + \cdots + (X_n - \mathbb{E}(X))^2\right]$$

Tests de variance, espérance connue

Modèle $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2) : \sigma^2 \in \mathbb{R}_+\right\}$

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 < \sigma_0^2$$

Test :

$$T = \frac{1}{\sigma_0^2}\overline{(X - \mathbb{E}(X))^2}_n = \frac{1}{\sigma_0^2}\left[(X_1 - \mathbb{E}(X))^2 + \cdots + (X_n - \mathbb{E}(X))^2\right]$$

Sous H_0 : $T \sim \chi^2(n)$ (Loi du Chi-deux)
Sous H_1 : $T \sim \chi^2(n, \mathbb{E}(X) - \sigma_0^2)$ (Loi du Chi-deux décentrée)

Tests de variance, espérance connue

Modèle $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2) : \sigma^2 \in \mathbb{R}_+\right\}$

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{VS} \quad H_1 : \sigma^2 < \sigma_0^2$$

Test :

$$T = \frac{1}{\sigma_0^2}\overline{(X - \mathbb{E}(X))^2}_n = \frac{1}{\sigma_0^2}\left[(X_1 - \mathbb{E}(X))^2 + \cdots + (X_n - \mathbb{E}(X))^2\right]$$

Sous $H_0 : T \sim \chi^2(n)$ (Loi du Chi-deux)
Sous $H_1 : T \sim \chi^2(n, \mathbb{E}(X) - \sigma_0^2)$ (Loi du Chi-deux décentrée)

On regarde si la p-valeur est en dessous du seuil qu'on s'est fixé avant (souvent 5%, ou 1%)

Test de Fisher

Maximum de vraisemblance

Toujours en modèle paramétrique : H_0 VS H_1

$$H_0 : \ll \ \theta \in A \ \gg \quad \textbf{VS} \quad H_1 : \ll \ \theta \in B \ \gg,$$

Mais on suppose : les $\mathcal{P}_\theta$ ont une *densité* par rapport à une autre mesure μ

Maximum de vraisemblance

Toujours en modèle paramétrique : H_0 VS H_1

$$H_0 : \ll \ \theta \in A \ \gg \quad \textbf{VS} \quad H_1 : \ll \ \theta \in B \ \gg,$$

Mais on suppose : les $\mathcal{P}_\theta$ ont une *densité* par rapport à une autre mesure μ (typiquement Lebesgue ou comptage).
Exemple : les lois de $\mathcal{P} = \left\{\mathcal{N}(\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+\right\}$ ont des densités :

$$L(\cdot; \mu, \sigma^2) : x \longmapsto L(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma}\right)$$

Par rapport à la mesure de Lebesgue sur $\mathbb{R}$.

Maximum de vraisemblance

Toujours en modèle paramétrique : H_0 VS H_1

$$H_0 :\ll\ \theta \in A\ \gg \quad \mathbf{VS} \quad H_1 :\ll\ \theta \in B\ \gg,$$

Mais on suppose : les $\mathcal{P}_\theta$ ont une *densité* par rapport à une autre mesure μ (typiquement Lebesgue ou comptage).
Exemple : les lois de $\mathcal{P} = \left\{\mathcal{U}_{[a,b)} : a \leq b \in \mathbb{R}\right\}$ ont des densités :

$$L(\cdot; a, b) : x \longmapsto L(x; a, b) = \frac{1}{b-a}\mathbb{1}_{[a,b)}(x)$$

Par rapport à la mesure de Lebesgue sur $\mathbb{R}$.

Maximum de vraisemblance

Toujours en modèle paramétrique : H_0 VS H_1

$$H_0 : \ll \ \theta \in A \ \gg \quad \textbf{VS} \quad H_1 : \ll \ \theta \in B \ \gg,$$

Mais on suppose : les $\mathcal{P}_\theta$ ont une *densité* par rapport à une autre mesure μ (typiquement Lebesgue ou comptage).
Exemple : les lois de $\mathcal{P} = \{\mathrm{Poisson}(\lambda) : \lambda \in \mathbb{R}_+\}$ ont des densités :

$$L(\cdot;\lambda) : k \longmapsto L(k;\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Par rapport à la mesure de comptage sur $\mathbb{N}$.

Maximum de vraisemblance

On introduit, pour un échantillon $x = (x_1, \ldots, x_n)$ de loi dans $(\mathcal{P}_\theta^{\otimes n})_\theta$ le rapport

$$\lambda(x) = \frac{\sup_{\theta \in B} L(x, \theta)}{\sup_{\theta \in A} L(x, \theta)}.$$

Maximum de vraisemblance

On introduit, pour un échantillon $x = (x_1, \dots, x_n)$ de loi dans $(\mathcal{P}_\theta^{\otimes n})_\theta$ le rapport

$$\lambda(x) = \frac{\sup_{\theta \in B} L(x, \theta)}{\sup_{\theta \in A} L(x, \theta)}.$$

Et on regarde le test, pour un seuil s à fixer au préalable :

$$T(x) = \mathbb{1}_{\lambda(x) \geq s}$$

Maximum de vraisemblance

On introduit, pour un échantillon $x = (x_1, \ldots, x_n)$ de loi dans $(\mathcal{P}_\theta^{\otimes n})_\theta$ le rapport

$$\lambda(x) = \frac{\sup_{\theta \in B} L(x, \theta)}{\sup_{\theta \in A} L(x, \theta)}.$$

Et on regarde le test, pour un seuil s à fixer au préalable :

$$T(x) = \mathbb{1}_{\lambda(x) \geq s}$$

De plus si $\hat{\theta}$ est une statistique “exhaustive”, alors il existe une fonction $\tilde{\lambda}$ tel que

$$\lambda(x) = \tilde{\lambda}(\hat{\theta}(x))$$

Apprentissage statistique

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \leq i \leq n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.

Apprentissage statistique

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \leq i \leq n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
But : estimer Y à partir d'un nouveau X.

Apprentissage statistique

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \leq i \leq n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
But : estimer Y à partir d'un nouveau X. Par exemple : $f(X)$.

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \le i \le n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
But : estimer Y à partir d'un nouveau X. Par exemple : $f(X)$.
On doit définir une mesure de qualité d'une estimation ; un *coût* ou une *perte* :

$$c : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$

souvent avec $c \geq 0$, $c(Y, Y) = 0$ et vérifiant l'inégalité triangulaire.

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \leq i \leq n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
But : estimer Y à partir d'un nouveau X. Par exemple : $f(X)$.
On doit définir une mesure de qualité d'une estimation ; un *coût* ou une *perte* :

$$c : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$

souvent avec $c \geq 0$, $c(Y, Y) = 0$ et vérifiant l'inégalité triangulaire.
$\longrightarrow$ On veut trouver $f : \mathcal{X} \to \mathcal{Y}$ tq la variable aléatoire $c(f(X), Y)$ soit, en moyenne la plus petite possible.

Apprentissage statistique

$\longrightarrow$ Données : un échantillon $D_n = (X_i, Y_i)_{1 \leq i \leq n}$; $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
But : estimer Y à partir d'un nouveau X. Par exemple : $f(X)$.
On doit définir une mesure de qualité d'une estimation ; un *coût* ou une *perte* :

$$c : \mathcal{Y} \times \mathcal{Y} \longrightarrow \mathbb{R}$$

souvent avec $c \geq 0$, $c(Y, Y) = 0$ et vérifiant l'inégalité triangulaire.
$\longrightarrow$ On veut trouver $f : \mathcal{X} \to \mathcal{Y}$ tq la variable aléatoire $c(f(X), Y)$ soit, en moyenne la plus petite possible.
$\longrightarrow$ Trouver un minimiseur f du *risque* :

$$\mathcal{R}(f) = \mathbb{E}(c(f(X), Y)|D_n)$$

Exemples :

Regression linéaire.

Si le problème est, pour un $\theta \in \mathbb{R}^n$, sous la forme

$$Y = \sum_i \theta_i X_i + \boldsymbol{\varepsilon} = X \cdot \theta + \boldsymbol{\varepsilon}$$

où $X = (X_1, \ldots, X_n)'$ et $Y = (Y_1 \ldots, Y_n)'$; avec $\mathbb{E}\left(\boldsymbol{\varepsilon}|X\right) = 0$, et un coût quadratique $c(a, b) = \|a - b\|^2$.

Exemples :

Regression linéaire.

Si le problème est, pour un $\theta \in \mathbb{R}^n$, sous la forme

$$Y = \sum_i \theta_i X_i + \boldsymbol{\varepsilon} = X \cdot \theta + \boldsymbol{\varepsilon}$$

où $X = (X_1, \ldots, X_n)'$ et $Y = (Y_1 \ldots, Y_n)'$; avec $\mathbb{E}\left(\boldsymbol{\varepsilon}|X\right) = 0$, et un coût quadratique $c(a, b) = \|a - b\|^2$.

Alors (Gauss-Markov) l'estimateur des moindres carrés est donné par

$$\hat{\theta} = \mathrm{argmin}_{\theta \in \mathbb{R}^n} \mathbb{E}(\|Y - \theta X\|^2) = (X'X)^{-1} X' Y.$$

Exemples :

Regression linéaire.

Si le problème est, pour un $\theta \in \mathbb{R}^n$, sous la forme

$$Y = \sum_i \theta_i X_i + \boldsymbol{\varepsilon} = X \cdot \theta + \boldsymbol{\varepsilon}$$

où $X = (X_1, \ldots, X_n)'$ et $Y = (Y_1 \ldots, Y_n)'$; avec $\mathbb{E}\left(\boldsymbol{\varepsilon}|X\right) = 0$, et un coût quadratique $c(a, b) = \|a - b\|^2$.

Alors (Gauss-Markov) l'estimateur des moindres carrés est donné par

$$\hat{\theta} = \text{argmin}_{\theta \in \mathbb{R}^n} \mathbb{E}(\|Y - \theta X\|^2) = (X'X)^{-1}X'Y.$$

Si on suppose $\mathbb{E}(\boldsymbol{\varepsilon}|X) \sim \mathcal{N}(0, \sigma^2)$, alors

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2(X'X)^{-1}).$$

Exemples :

Regression linéaire.
Si le problème est, pour un $\theta \in \mathbb{R}^n$, sous la forme

$$Y = \sum_i \theta_i X_i + \boldsymbol{\varepsilon} = X \cdot \theta + \boldsymbol{\varepsilon}$$

où $X = (X_1, \ldots, X_n)'$ et $Y = (Y_1 \ldots, Y_n)'$; avec $\mathbb{E}\left(\boldsymbol{\varepsilon}|X\right) = 0$, et un coût quadratique $c(a, b) = \|a - b\|^2$.
Alors (Gauss-Markov) l'estimateur des moindres carrés est donné par

$$\hat{\theta} = \mathrm{argmin}_{\theta \in \mathbb{R}^n} \mathbb{E}(\|Y - \theta X\|^2) = (X'X)^{-1} X' Y.$$

Si on suppose $\mathbb{E}(\boldsymbol{\varepsilon}|X) \sim \mathcal{N}(0, \sigma^2)$, alors

$$\hat{\theta} \sim \mathcal{N}(\theta, \sigma^2 (X'X)^{-1}).$$

Attention aux grandes dimensions ! (*cf* Notebook)

Exemples :

En considérant f^* un *oracle* : $f^* \in \operatorname{argmin}_f(\mathcal{R}(f))$.
Modèle paramétrique.
Si on a une bonne intuition, on peut simplifier en regardant un modèle de fonction $M = \{f_\theta : \theta \in \mathbb{R}^n\}$

$$\mathcal{R}(f, f^*) = d(M, f^*) + \inf_\theta \mathcal{R}(f_\theta)$$

$\longrightarrow$ en espérant que la distance $d(S, f^*)$ est petite et que $\inf_\theta \mathcal{R}(f_\theta)$ est facile à calculer.
$\longrightarrow$ c'est de l'optimisation.

Contenu

Références

1. Probabilités - Barbé,Ledoux
2. Introduction aux statistiques en grande dimension - Giraud
3. Test Statistiques - Fromont