

++ de Statistiques

Ou comment survivre dans la jungle des stats

Tome 2 : Suite des aventures dans la jungle des probas

Colin 'Alaska' Coërchon

IIE++

7 mai 2024

Sommaire

- 1 Résumé des éléments importants
 - Rappels de proba
 - Les bases indispensables en statistiques
 - Des méthodes et définitions à connaître
 - Intervalles de confiance
 - Théorie des tests
 - Les tests paramétriques
 - Les autres tests
- 2 Bilan
 - Les différents types d'estimateurs
 - Les essentiels pour juste valider
 - Comment *bien* valider stat ?
- 3 Correction du partiel de l'an dernier
- 4 Place aux questions !

1 Résumé des éléments importants

- Rappels de proba
- Les bases indispensables en statistiques
- Des méthodes et définitions à connaître
- Intervalles de confiance
- Théorie des tests

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Lecture du premier Tome

J'espère que vous avez aimé "la survie dans la jungle des probas" !
Vous allez sûrement beaucoup aimé ce nouveau tome.

Vous n'avez pas lu le premier ? Malheureux ! Mais il n'est jamais trop tard... (Voici le lien : [ici](#))

Lecture du premier Tome

J'espère que vous avez aimé "la survie dans la jungle des probas" !
Vous allez sûrement beaucoup aimé ce nouveau tome.

Vous n'avez pas lu le premier ? Malheureux ! Mais il n'est jamais trop tard... (Voici le lien : [ici](#))

Les différentes lois à connaître

Après nos aventures dans la jungle des probas, vous devez nécessairement connaître toutes ces lois :

- Bernoulli, Binomiale, Géométrique, Poisson, Exponentielle et Uniforme.
- Et surtout : **la loi normale !**

Une autre loi importante

Un oubli de ma part durant le dernier ++ :

La loi du χ^2

Si $X_1, \dots, X_n$ iid suivent la loi $\mathcal{N}(0, 1)$, alors :

$$X := \sum_{i=1}^n X_i^2 \quad \sim \quad \chi_n^2$$

C'est la loi du χ^2 avec n degré de liberté.

Une application connue

La variance empirique

Pour approcher la variance σ^2 d'un échantillon $(X_1, \dots, X_n)$ iid, on définit la variance empirique comme :

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

On a alors le résultat suivant :

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

Tout ceci nous servira bien plus tard...

Les différents types de convergence

- **Convergence presque sûre** : on dit que la suite de variables aléatoires X_n converge presque sûrement vers X ssi il existe un ensemble négligeable N tel que

$$\forall \omega \in \Omega \setminus N, \quad X_n(\omega) \xrightarrow[n \rightarrow +\infty]{p.s.} X(\omega)$$

- **Convergence en probabilité** :

$$X_n \xrightarrow{\mathbb{P}} X \iff \forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

- **Convergence en loi** : pour toute fonction numérique f continue et bornée,

$$X_n \xrightarrow{\mathcal{L}} X \iff \lim_{n \rightarrow +\infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)]$$

Ordre d'importance :

Convergence p.s. $\implies$ Convergence en proba $\implies$ Convergence en Loi

Théorèmes de convergence

Loi forte des grands nombres

Soit (X_n) une suite de variables aléatoire i.i.d tel que $\mathbb{E}[|X_1|] < +\infty$. Alors

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[X_1]$$

Théorème Central Limite (TCL)

Soit (X_n) une suite de variables aléatoire i.i.d tel que $\mathbb{E}[|X_1|] < +\infty$ et $\text{Var}(|X_1|) < +\infty$. Alors :

$$\begin{aligned} \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1) \\ \Leftrightarrow \bar{X}_n &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \end{aligned}$$

1 Résumé des éléments importants

- Rappels de proba
- **Les bases indispensables en statistiques**
- Des méthodes et définitions à connaître
- Intervalles de confiance
- Théorie des tests

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Modèle statistique

Pour faire des stats, on se définit un modèle statistique $\mathcal{M}$.

Modèle statistique

Définition : C'est une famille de probabilités sur un espace d'observation $\mathcal{X}$ définie comme telle :

$$\mathcal{M} = \{\mathbb{P}_\theta \mid \theta \in \Theta\}$$

Avec Θ l'espace des paramètres.

Bref, en stats, on a donc des modèles de probabilités, mais surtout, des **paramètres à estimer** !

Les estimateurs

Définition

En stat, on a donc $(X_1, \dots, X_n) \in \mathcal{M}$. Et on souhaite alors estimer le **vrai paramètre** θ^* . On construit alors un **estimateur** $\hat{\theta}$ tel que :

$$\hat{\theta} = T(X_1, \dots, X_n) \text{ approche correctement le paramètre } \theta^*$$

Les estimateurs

Définition

En stat, on a donc $(X_1, \dots, X_n) \in \mathcal{M}$. Et on souhaite alors estimer le **vrai paramètre** θ^* . On construit alors un **estimateur** $\hat{\theta}$ tel que :

$$\hat{\theta} = T(X_1, \dots, X_n) \text{ approche correctement le paramètre } \theta^*$$

Dans l'idéal, on cherche un estimateur $\hat{\theta}$ ayant au moins 2 propriétés :

- 1 **Estimateur non biaisé** : $\mathbb{E}_{\theta^*}[\hat{\theta}] = \theta^*$
- 2 **Estimateur consistant** : $\hat{\theta} \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta^*$

Ce n'est pas toujours le cas...

Statistique

Définition

Une statistique $T(X_1, \dots, X_n)$ est simplement une fonction de l'échantillon (avec donc $T : \mathcal{X} \rightarrow \mathbb{R}^m$).

Estimateur $\neq$ Statistique

Chaque estimateur est une statistique, mais **toutes les statistiques ne sont pas des estimateurs !**

En d'autres termes, y'a des statistiques qui ne servent vraiment à rien.

1 Résumé des éléments importants

- Rappels de proba
- Les bases indispensables en statistiques
- Des méthodes et définitions à connaître
- Intervalles de confiance
- Théorie des tests

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Le risque quadratique

Définition

On définit le risque quadratique de l'estimateur T_n par :

$$\begin{aligned} R(T_n, \theta) &= (\mathbb{E}_\theta[T_n] - \theta)^2 + \text{Var}_\theta(T_n) \\ &= (b(T_n, \theta))^2 + \text{Var}_\theta(T_n) \end{aligned}$$

Avec $b(T_n, \theta)$ le **biais** de l'estimateur

Note : Un bon estimateur T_n est un estimateur qui **minimise** le risque quadratique.

Comment répondre à la question :

Trouver un estimateur à l'aide de la méthode des moments.

On va voir ça tout de suite !

Méthode des moments

La méthode des moments générale est à l'ordre K . Il faut **impérativement** savoir l'appliquer à l'ordre 1 !

Méthode des moments à l'ordre 1

Soit X une V.A., on suppose qu'on connaît $\mathbb{E}[X] = m_1(\theta)$ et qu'on souhaite estimer le paramètre θ . La méthode des moments à l'ordre 1 nous assure qu'il suffit de résoudre l'équation :

$$\hat{m}_1 = \overline{X}_n = m_1(\theta)$$

La solution de cette équation nous donne alors un estimateur $\hat{\theta}$.

Exemple 1

Si $X \sim \mathcal{E}(\theta)$, alors $\mathbb{E}[X] = \frac{1}{\theta}$ et donc :

$$\hat{m}_1 = m_1(\theta) \implies \bar{X}_n = \frac{1}{\theta} \implies \theta = \frac{1}{\bar{X}_n}$$

La méthode des moments à l'ordre 1 donne donc $\hat{\theta} = \frac{1}{\bar{X}_n}$

Exemple 2

Si X est une V.A. avec $\mathbb{E}[X] = 2\sqrt{\frac{2}{\pi\theta}}$, alors on résout :

$$\bar{X}_n = 2\sqrt{\frac{2}{\pi\theta}} \implies \frac{\bar{X}_n^2}{4} = \frac{2}{\pi\theta} \implies \theta = \frac{8}{\pi\bar{X}_n^2}$$

La méthode des moments à l'ordre 1 donne donc $\hat{\theta} = \frac{8}{\pi\bar{X}_n^2}$

Méthode des moments

Méthode des moments à l'ordre K

Soit X une V.A., on suppose qu'on connaît $\forall k \in \llbracket 1, K \rrbracket$, $m_k(\theta) = \mathbb{E}[X^k]$ et qu'on souhaite estimer le paramètre θ .

De plus, on note $\hat{m}_k$ les moments à l'ordre k : $\hat{m}_k = \frac{1}{n} \sum_{i=1}^n X_i^k$.

La méthode des moments à l'ordre K nous assure qu'il suffit de résoudre l'équation ci-contre pour trouver $\hat{\theta}$.

$$\begin{cases} \hat{m}_1 &= m_1(\theta) \\ \vdots &= \vdots \\ \hat{m}_K &= m_K(\theta) \end{cases}$$

On a alors :

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \|\hat{m} - m(\theta)\|^2$$

On notera $\overline{X^2}_n = \frac{1}{n} \sum_{i=1}^n X_i^2$ le moment empirique d'ordre 2.

Exemple 1

Si $X \sim \mathcal{E}(\theta)$, alors $\mathbb{E}[X] = \frac{1}{\theta}$ et $\mathbb{E}[X^2] = \frac{2}{\theta^2}$. On résoud donc :

$$\begin{cases} \hat{m}_1 = m_1(\theta) \\ \hat{m}_2 = m_2(\theta) \end{cases} \implies \begin{cases} \overline{X}_n = \frac{1}{\theta} & (1) \\ \overline{X^2}_n = \frac{2}{\theta^2} & (2) \end{cases}$$

On a donc :

- ❶ L'équation (1) donne un estimateur : $\hat{\theta}_1 = \frac{1}{\overline{X}_n}$.
- ❷ L'équation (2) donne un autre estimateur : $\hat{\theta}_2 = \sqrt{\frac{2}{\overline{X^2}_n}}$.
- ❸ On peut aussi trouver $\hat{\theta}_3$ en cherchant :

$$\arg \min_{\theta > 0} \left[\left(\overline{X}_n - \frac{1}{\theta} \right)^2 + \left(\overline{X^2}_n - \frac{2}{\theta^2} \right)^2 \right]$$

Vraisemblance

Définition de la vraisemblance

Pour un échantillon $(x_1, \dots, x_n)$ i.i.d, on définit la vraisemblance L par :

$$L(x_1, \dots, x_n; \theta) = f_{X_1 \dots X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_i}(x_i; \theta)$$

Définition de la log-vraisemblance

Pour un échantillon $(x_1, \dots, x_n)$ i.i.d, on définit la vraisemblance $\mathcal{L}$ par :

$$\mathcal{L}(x_1, \dots, x_n; \theta) = \ln(L(x_1, \dots, x_n; \theta)) = \sum_{i=1}^n \ln(f_{X_i}(x_i; \theta))$$

Estimateur du Maximum de Vraisemblance (EMV)

Il y a énormément d'estimateurs imaginables, mais si on doit en choisir **un seul**, c'est bien l'EMV !

Définition

Pour un échantillon $(x_1, \dots, x_n)$ i.i.d, l'EMV est un estimateur qui maximise la vraisemblance (*et donc aussi la log-vraisemblance*). On a donc :

$$\hat{\theta}^{\text{MV}} \in \arg \max_{\theta \in \Theta} \mathcal{L}(x_1, \dots, x_n; \theta)$$

Comment répondre à la question :

Trouver l'estimateur du maximum de vraisemblance.

On va voir ça tout de suite !

Méthode de calcul de l'EMV

Dans la plupart des modèles statistiques, il y a **une seule méthode** pour déterminer un EMV.

Méthode de l'EMV

- Conditions du 1^{er} ordre.

Objectif : Trouver $\hat{\theta}^{MV}$ en résolvant ceci :

$$\nabla_{\theta} \mathcal{L}(x_1, \dots, x_n; \theta) = 0$$

- Conditions du 2^{ème} ordre.

Objectif : Vérifier que $\hat{\theta}^{MV}$ maximise la vraisemblance :

$$\nabla_{\theta}^2 \mathcal{L}(x_1, \dots, x_n; \hat{\theta}) \prec 0$$

Attention, " $\prec 0$ " signifie "symétrique définie négative".

Méthode de l'EMV en dimension 1

- Conditions du 1^{er} ordre.

Objectif : Trouver $\hat{\theta}^{MV}$ en résolvant ceci :

$$\frac{\partial \mathcal{L}}{\partial \theta}(x_1, \dots, x_n; \theta) = 0$$

- Conditions du 2^{ème} ordre.

Objectif : Vérifier que $\hat{\theta}^{MV}$ maximise la vraisemblance :

$$\frac{\partial^2 \mathcal{L}}{\partial \theta^2}(x_1, \dots, x_n; \hat{\theta}) < 0$$

Un exemple ?

Cf. question 5, exercice 2, partiel de statistiques 2023.

Comment répondre à la question :

Est-ce que la statistique $T(X_1, \dots, X_n)$ est une statistique exhaustive ?

On va voir ça tout de suite !

Statistique exhaustive

Pour répondre à cette question, on utilise **TOUJOURS** ce lemme :

Lemme de factorisation

Soit $(x_1, \dots, x_n)$ un échantillon i.i.d.

- **Si** : il existe 3 fonctions T , g et h telles que l'on a :

$$f(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n), \theta) \times h(x_1, \dots, x_n)$$

- **Alors** : $T(X_1, \dots, X_n)$ est une statistique exhaustive pour notre modèle.

Exemple 1 : La loi géométrique

Soit $X \sim \mathcal{G}(p)$. Alors $\forall k \in \mathbb{N}^*$, $\mathbb{P}(X = k) = p(1 - p)^{k-1}$, et donc :
 $f(x, p) = p(1 - p)^{x-1} \mathbb{1}_{\mathbb{N}^*}(x)$. On a alors :

$$\begin{aligned}
 f(x_1, \dots, x_n; p) &= p^n (1 - p)^{\sum_{i=1}^n (x_i - 1)} \prod_{i=1}^n \mathbb{1}_{\mathbb{N}^*}(x_i) \\
 &= \exp \left(n \ln(p) + \ln(1 - p) \sum_{i=1}^n (x_i - 1) \right) \times \exp \left(\sum_{i=1}^n \ln(\mathbb{1}_{\mathbb{N}^*}(x_i)) \right) \\
 &= \underbrace{\exp \left(n(\ln(p) - \ln(1 - p)) + \ln(1 - p) \sum_{i=1}^n x_i \right)}_{g(T(x_1, \dots, x_n), p)} \times \underbrace{\exp \left(\sum_{i=1}^n \ln(\mathbb{1}_{\mathbb{N}^*}(x_i)) \right)}_{h(x_1, \dots, x_n)}
 \end{aligned}$$

En posant $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$, d'après le lemme de factorisation, T est une statistique exhaustive

D'autres exemples

Exemple 2 : La loi de Poisson

Exemple plus simple.

Cf. **exercice 1, question 2**, contrôle continu de statistiques 2023.

Exemple 3 : Une loi normale particulière

Exemple plus difficile.

Cf. **exercice 2, question 3**, partiel de statistiques 2023.

Comment répondre à la question :

Est-ce que notre modèle est une famille exponentielle ?

On va voir ça tout de suite !

Famille exponentielle

Définition (avec θ de dimension 1)

- **Si** : on peut écrire

$$f(x, \theta) = \exp(T(x)\eta(\theta) - A(\eta)) \times h(x)$$

- **Alors** : la densité de la loi est **une famille exponentielle**.

Propriétés

- De nombreuses lois connues ont pour vraisemblance une famille exponentielle. **Ex** : Gaussienne, Gamma, Beta, Bernoulli, Poisson, Géométrique, etc...
- Une famille exponentielle implique **l'existence d'un EMV**.

Exemple 1 : La loi de Poisson

$$f(x) = \frac{\lambda^x}{x!} e^{-\lambda} = \exp(x \ln(\lambda) - \lambda) \times \frac{1}{x!}$$

C'est donc une famille exponentielle avec :

$$T(x) = x, \quad h(x) = \frac{1}{x!}, \quad \eta = \ln(\lambda) \quad \text{et} \quad A(\eta) = \exp(\lambda)$$

Exemple 2 : La loi exponentielle

$$f(x) = \lambda e^{-\lambda x} = \exp(-\lambda x + \ln(\lambda)) \times 1$$

C'est donc une famille exponentielle avec :

$$T(x) = x, \quad h(x) = 1, \quad \eta = -\lambda \quad \text{et} \quad A(\eta) = -\ln(-\eta)$$

Plein d'autres exemples sur Wikipedia : [ici](#)

1 Résumé des éléments importants

- Rappels de proba
- Les bases indispensables en statistiques
- Des méthodes et définitions à connaître
- **Intervalles de confiance**
- Théorie des tests

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Score (θ de dimension 1)

Le Score

Le score $S(X, \theta)$ est défini par :

$$S(X, \theta) = \frac{d}{d\theta} \mathcal{L}(X; \theta)$$

De plus, on a toujours : $\mathbb{E}_\theta[S(X, \theta)] = 0$

Information de Fisher (θ de dimension 1)

Information de Fisher

L'information de Fisher $I(\theta)$ est définie par :

$$I(\theta) = \mathbb{E} \left[\left(\frac{d}{d\theta} \mathcal{L}(X; \theta) \right)^2 \right] \quad \text{OU} \quad I(\theta) = -\mathbb{E}_{\theta} \left[\frac{d^2}{d\theta^2} \mathcal{L}(X; \theta) \right]$$

Et pour un échantillon de taille n $((X_1, \dots, X_n))$, on a :

$$\begin{aligned} I_n(\theta) &= -\mathbb{E}_{\theta} \left[\frac{d^2}{d\theta^2} \mathcal{L}(X_1, \dots, X_n; \theta) \right] \\ &= n I_1(\theta) = n I(\theta) \end{aligned}$$

Borne de Cramér-Rao

Borne de Cramér-Rao

Soit $(X_1, \dots, X_n)$ un échantillon de taille n . Pour un estimateur $\hat{\theta}$ d'un paramètre θ , si l'estimateur est **non biaisé**, alors :

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n I(\theta)} \quad \left(= \frac{1}{I_n(\theta)} \right)$$

Transformation par continuité

Transformation par continuité

Si on a :

- $(X_n)_{n \in \mathbb{N}}$ suite de v.a. telle que : $X_n \xrightarrow{\mathbb{P}} a \in \mathbb{R}$
- g fonction continue en a .

Alors : $g(X_n) \xrightarrow{\mathbb{P}} g(a)$

Utilité de ce théorème ?

La transformation continue est très utile pour montrer la **consistance** d'un estimateur. Par exemple, si $\overline{X}_n \xrightarrow{\mathbb{P}} \frac{1}{\theta}$, comme la fonction $x \mapsto \frac{1}{x}$ est continue sur $\mathbb{R}_+^*$, alors on a directement :

$$\frac{1}{\overline{X}_n} \xrightarrow{\mathbb{P}} \theta$$

Et donc, l'estimateur $\hat{\theta} = \frac{1}{\overline{X}_n}$ est **consistant**.

Théorème de Slutsky

Théorème de Slutsky

Si on a :

- $(A_n)_{n \in \mathbb{N}^*}$ suite de v.a. telle que : $A_n \xrightarrow{\mathbb{P}} A \in \mathbb{R}$
- $(B_n)_{n \in \mathbb{N}^*}$ suite de v.a. telle que : $B_n \xrightarrow{\mathbb{P}} B \in \mathbb{R}$
- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.r. telle que : $X_n \xrightarrow{\mathcal{L}} X$ v.a.r.

Alors :

$$A_n X_n + B_n \xrightarrow{\mathcal{L}} AX + B$$

Méthode Delta

Méthode Delta

Si on a :

- $(X_n)_{n \in \mathbb{N}^*}$ suite de v.a.r. telle que : $X_n \xrightarrow{\mathbb{P}} a$ (v.a.r.)
- $g \mapsto g(x)$ fonction continue de $\mathbb{R}$ dans $\mathbb{R}$ dérivable en a , avec $g'(a) \neq 0$.
- $\exists V > 0, \exists (r_n)_{n \in \mathbb{N}^*}$ suite croissante tels que :

$$r_n (X_n - a) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V)$$

Alors :

$$r_n (g(X_n) - g(a)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (g'(a))^2 V)$$

Souvent, on a : $r_n = \sqrt{n}$ et la convergence en loi est obtenue par le **TCL** !

Un exemple

Méthode Delta avec la loi exponentielle

Soit $(X_1, \dots, X_n)$ un échantillon i.i.d de la loi exponentielle $\mathcal{E}(\lambda)$. On a donc : $\mathbb{E}[X_1] = \frac{1}{\lambda}$ et $\text{Var}(X_1) = \frac{1}{\lambda^2}$. On cherche à estimer le paramètre λ .

D'après la LGN (faible), on a : $\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = \frac{1}{\lambda}$.

D'après le TCL, on a alors :

$$\sqrt{n} \left(\bar{X}_n - \frac{1}{\lambda} \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{\lambda^2} \right)$$

On va appliquer la méthode avec la fonction $g : x \mapsto \frac{1}{x}$. g est de classe $\mathcal{C}^1$ sur $\mathbb{R}_+^*$

D'après la **méthode Delta**, on a donc :

$$\sqrt{n} \left(\frac{1}{\bar{X}_n} - \lambda \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{\lambda^2} \times \left(g' \left(\frac{1}{\lambda} \right) \right)^2 \right) = \mathcal{N}(0, \lambda^2)$$

Un autre exemple

Exemple concret

Cf. **exercice 2, question 9** (la dernière), partiel de statistiques 2023.

Comment répondre à la question :

Donner un intervalle de confiance de niveau asymptotique $1 - \alpha$.

On va voir ça tout de suite !

Intervalles de confiance

On considère le cas $X \sim \mathcal{N}(\mu, \sigma^2)$:

- **Cas 1** : Estimation de l'espérance μ lorsque la variance σ^2 est connue.

$$IC_{1-\alpha}(\mu) = \left[\bar{x}_n - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x}_n + u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

- **Cas 2** : Estimation de l'espérance μ lorsque la variance σ^2 est inconnue.

$$IC_{1-\alpha}(\mu) = \left[\bar{x}_n - t_{1-\alpha/2} \frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{1-\alpha/2} \frac{s_n}{\sqrt{n}} \right]$$

- **Cas 3** : Estimation de la variance σ^2 lorsque l'espérance μ est inconnue.

$$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)s_n^2}{v_{1-\alpha/2}}, \frac{(n-1)s_n^2}{v_{\alpha/2}} \right]$$

Ça manque d'information ?

Vous voulez devenir un crackito des intervalles de confiance et avoir 100% de réussite à des questions sur ça ?

Ça manque d'information ?

Vous voulez devenir un crackito des intervalles de confiance et avoir 100% de réussite à des questions sur ça ?

La solution magique !

C'est possible, car toutes les informations dont vous avez besoin sont sur le document "**Tuto intervalles de confiance**" !

Enjoy !

1 Résumé des éléments importants

- Rappels de proba
- Les bases indispensables en statistiques
- Des méthodes et définitions à connaître
- Intervalles de confiance
- Théorie des tests
 - Les tests paramétriques
 - Les autres tests

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Les notations usuelles

Définition des hypothèses

$$\begin{cases} H_0 : \theta \in \Theta_0 & \text{(Hypothèse nulle)} \\ H_1 : \theta \in \Theta_1 & \text{(Hypothèse alternative)} \end{cases}$$

Erreurs de type I et II

- Zone de rejet W : Ensemble des valeurs pour lesquelles H_0 est rejetée.
- $\alpha = \mathbb{P}_{H_0}(W)$: Probabilité de rejeter H_0 alors que H_0 est vraie (**Erreur de type I**).
- $\beta = \mathbb{P}_{H_1}(W^c)$: Probabilité de ne pas rejeter H_0 alors que H_1 est vraie (**Erreur de type II**).

On peut résumer la définition de α et β par ceci :

	H_0 vraie	H_1 vraie
Rejeter H_0	α (Erreur de type I)	$1 - \beta$ (Puissance du test)
Ne pas rejeter H_0	$1 - \alpha$ (Spécificité)	β (Erreur de type II)

Table – Tableau des décisions de test et réalité des hypothèses

Erreur dans le poly

Il est noté sur le poly que β correspond à la puissance du test (et $1 - \beta$ l'erreur de type II). Mais, sur n'importe quel autre document sérieux en statistique, c'est l'inverse...

Je prends donc l'inverse :

- β : erreur de type II.
- $1 - \beta$: puissance du test.

Types de tests statistiques

Les principaux types de tests

- Test simple :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

- Test unilatéral :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

- Test bilatéral :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$$

Méthode de Neyman-Pearson

Lemme de Neyman-Pearson

Pour un niveau de signification α donné, on définit la zone de rejet comme :

$$W = \{(x_1, \dots, x_n) \in \mathbb{R}^n \mid \Lambda > K_\alpha\} \quad \text{avec : } \Lambda = \frac{L(x_1, \dots, x_n; \theta_1)}{L(x_1, \dots, x_n; \theta_0)}$$

où K_α est une constante qui dépend uniquement de α telle que : $\mathbb{P}_{H_0}(W) = \alpha$.

Comment répondre aux questions du type :

- Construire le test simple suivant.
- Donner la zone de rejet permettant de garantir un risque de première espèce égal à α .
- Quel est la puissance du test ?
- ...

On va voir ça tout de suite !

Des exemples

Exemple du poly

Les exemples du polycopié sur les tests (le dernier) sont franchement très bons.

Je me suis entraîné avec ceux là personnellement, et puis bon, j'ai vraiment la flemme d'en recopier un en entier en $\text{\LaTeX}$.

Un exemple concret

Cf. **tout l'exercice 1** du partiel de statistiques 2023.

1 Résumé des éléments importants

- Rappels de proba
- Les bases indispensables en statistiques
- Des méthodes et définitions à connaître
- Intervalles de confiance
- **Théorie des tests**
 - Les tests paramétriques
 - **Les autres tests**

2 Bilan

3 Correction du partiel de l'an dernier

4 Place aux questions !

Les autres tests

D'autres tests plus complexes existent :

- Test d'adéquation
- Test d'indépendance
- Test d'ajustement de Kolmogorov

Trop d'honnêteté ?

En toute franchise, je n'avais pas vraiment travaillé ces tests... Ils sont plus complexes, et ne semblent pas trop tomber aux partiels.

En bref, je ne pourrais pas vous être d'une grande aide pour ça, bonne chance à vous !

- 1 Résumé des éléments importants
- 2 Bilan
 - Les différents types d'estimateurs
 - Les essentiels pour juste valider
 - Comment *bien* valider stat ?
- 3 Correction du partiel de l'an dernier
- 4 Place aux questions !

Propriétés des estimateurs en statistique

- **Estimateur non biaisé** : Un estimateur $\hat{\theta}$ est dit non biaisé pour un paramètre θ si :

$$\mathbb{E}[\hat{\theta}] = \theta$$

- **Estimateur consistant** : Un estimateur $\hat{\theta}$ est consistant (ou convergent) pour θ si :

$$\hat{\theta} \xrightarrow{\mathbb{P}} \theta$$

- **Estimateur fortement consistant** : Un estimateur $\hat{\theta}$ est fortement consistant pour θ si :

$$\hat{\theta} \xrightarrow[n \rightarrow +\infty]{p.s.} \theta$$

- **Estimateur efficace** : Un estimateur $\hat{\theta}$ est dit efficace s'il est non biaisé ET s'il atteint l'égalité dans l'inégalité de Cramér-Rao :

$$\text{Var}(\hat{\theta}) = \frac{1}{n I(\theta)} \quad \left(= \frac{1}{I_n(\theta)} \right)$$

Propriétés asymptotiques des estimateurs

Deux de ces propriétés ont leur versions "plus faibles" en étant vraies uniquement lorsque $n \rightarrow +\infty$:

- **Estimateur asymptotiquement non biaisé** : Un estimateur $\hat{\theta}$ est dit asymptotiquement non biaisé pour un paramètre θ si :

$$\mathbb{E}[\hat{\theta}] \xrightarrow{n \rightarrow +\infty} \theta$$

- **Estimateur asymptotiquement efficace** : Un estimateur $\hat{\theta}$ est dit asymptotiquement efficace s'il est asymptotiquement non biaisé ET s'il converge vers la borne de Cramér-Rao :

$$\text{Var}(\hat{\theta}) \xrightarrow{n \rightarrow +\infty} \frac{1}{n I(\theta)} \quad \left(= \frac{1}{I_n(\theta)} \right)$$

- 1 Résumé des éléments importants
- 2 Bilan
 - Les différents types d'estimateurs
 - Les essentiels pour juste valider
 - Comment *bien* valider stat ?
- 3 Correction du partiel de l'an dernier
- 4 Place aux questions !

Comment valider stat ?

- Connaître le **TCL**.
- Savoir ce qu'est un **estimateur**.
- Savoir calculer le **risque quadratique** d'un estimateur.
- **Méthode des moments** à l'ordre 1.
- Vraisemblance, log-vraisemblance, et calcul de l'**EMV** (θ de dimension 1).
- Savoir donner une **statistique exhaustive**.
- Savoir calculer l'**information de Fisher**.
- Savoir **construire un test** (au moins le test simple).

- 1 Résumé des éléments importants
- 2 Bilan
 - Les différents types d'estimateurs
 - Les essentiels pour juste valider
 - Comment *bien* valider stat ?
- 3 Correction du partiel de l'an dernier
- 4 Place aux questions !

Il faut rajouter aux précédentes connaissances ces notions :

- Les différents types de convergences de v.a.
- **EMV** avec θ de dimension 2.
- Les **intervalles de confiance**.
- Reconnaître **une famille exponentielle**.
- Transformation par continuité (pour la consistance des estimateurs).
- Savoir construire tous les tests paramétriques.

Et si vous êtes un big tryharder :

- Méthode des moments à l'ordre K .
- Théorème de Slutsky.
- **Méthode Delta** pour les intervalles de confiance.
- Les autres tests (*Tryharder ultime*).

- 1 Résumé des éléments importants
- 2 Bilan
- 3 Correction du partiel de l'an dernier**
- 4 Place aux questions !

- 1 Résumé des éléments importants
- 2 Bilan
- 3 Correction du partiel de l'an dernier
- 4 Place aux questions !**

Des questions ?

