

Introduction à la statistique

Cours d'agrégation, option A

M. Simon

1 Exemple et problématique

De manière générale, le rôle du statisticien est d'extraire des informations de données mesurées, sans connaître *a priori* le modèle sous-jacent.

Exemple 1.1 *On modélise l'intention de vote des électeurs à un référendum par 0 pour non et 1 pour oui. On cherche à estimer la probabilité $\theta \in [0, 1]$ qu'un électeur vote oui. Pour ce faire, on interroge n personnes choisies au hasard, dont les réponses x_1, \dots, x_n sont considérées comme des réalisations de v. a. X_1, \dots, X_n i.i.d de même loi de Bernoulli $(1 - \theta)\delta_0 + \theta\delta_1$. **Objectif** : estimer le paramètre θ (autrement dit, on estime une proportion).*

Exemple 1.2 *Une entreprise de téléphonie suppose que la distribution du nombre journalier des appels passés par ses clients est une loi de Poisson de paramètre θ inconnu. Cette hypothèse, appelée H_0 , lui permet de fixer ses tarifs, et elle souhaite donc la tester. Pour cela, elle mesure sur une journée le nombre d'appels téléphoniques émis, et dispose de l'observation de n v.a. X_1, \dots, X_n i.i.d de loi inconnue. **Objectif** : accepter ou rejeter l'hypothèse H_0 , et, en cas d'acceptation, estimer le paramètre θ de la loi.*

La statistique paramétrique consiste à considérer qu'un phénomène aléatoire obéit à une loi de probabilité, qui appartient à une famille de loi indexée par un paramètre. Le choix de la famille constitue la *modélisation*. La véritable valeur du paramètre est inconnue, et l'on cherche à la déterminer à l'aide d'un *échantillon*. On n'obtiendra jamais une valeur exacte pour ce paramètre inconnu, mais un *intervalle de confiance*, c'est-à-dire un sous-ensemble de l'ensemble des paramètres qui contient la véritable valeur avec au moins une certaine probabilité p , appelée *niveau de confiance*. On parle de *risque* pour désigner la probabilité $1 - p$, et on cherchera à le minimiser. Pour estimer le paramètre, on construit une fonction de l'échantillon, appelée *estimateur*, qui converge, en un certain sens, vers la valeur recherchée du paramètre.

Remarque – Un modèle statistique est toujours faux. Ce n'est qu'une approximation simple de la réalité qu'on espère pas trop mauvaise. Le choix d'un modèle doit toujours être justifié et critiqué. Il repose sur la connaissance du phénomène étudié (par exemple, des représentations graphiques).

2 Rudiments de statistique paramétrique

Définition 2.1

1. *Un modèle statistique est la donnée d'un espace mesurable et d'un tribu (Ω, \mathcal{A}) ainsi que d'une famille $(\mathbb{P}_\theta)_{\theta \in \Theta}$ de lois de probabilités sur cet espace.*
2. *Lorsque $\Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, le modèle est dit paramétrique. Sinon, il est dit non-paramétrique.*

3. Une observation X est une variable aléatoire à valeurs dans Ω dont la loi inconnue μ appartient à $\{\mathbb{P}_\theta ; \theta \in \Theta\}$. On dispose de données x_1, \dots, x_n , qui constituent un échantillon, et que l'on suppose être des réalisations $X_1(\omega), \dots, X_n(\omega)$ de v.a. i.i.d de même loi μ .

Remarque – On supposera toujours que $\theta \rightarrow \mathbb{P}_\theta$ soit injective (pour que l'on puisse dire des choses sur θ). On dit alors que le modèle est *identifiable*.

Exemple 1.1 : $\Omega = \{0, 1\}$, \mathcal{A} est l'ensemble des parties de Ω , $\mathbb{P}_\theta = (1 - \theta)\delta_0 + \theta\delta_1$, et $\Theta =]0, 1[$.

Exemple 1.2 : $\Omega = \mathbb{N}$, \mathcal{A} est l'ensemble des parties de Ω ,
 $\Theta = \{\text{fonctions de répartition nulle sur } \mathbb{R}_- \text{ et constante sur tous les intervalles } [k, k + 1]\}$,
 $\mathbb{P}_\theta = \text{loi associée à la fonction de répartition } \theta$.

Dans toute la suite, on suppose donné un modèle statistique identifiable que l'on note $(\Omega, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$.

3 Méthodes d'estimation

L'un des objectifs de la statistique paramétrique est de déterminer au mieux le paramètre θ de la loi inconnue en utilisant les données.

Définition 3.1 Soit (X_1, \dots, X_n) un échantillon de loi \mathbb{P}_θ et $g : \Theta \rightarrow \mathbb{R}^d$ une fonction mesurable.

Un estimateur \widehat{G}_n de $g(\theta)$ est une fonction mesurable de X_1, \dots, X_n à valeurs dans \mathbb{R}^d , indépendante de θ .

Un estimateur peut avoir différentes propriétés qui illustrent son bon comportement dans l'estimation de $g(\theta)$. On note \mathbb{E}_θ et Var_θ l'espérance et la variance sous la loi \mathbb{P}_θ .

Définition 3.2 On dit que l'estimateur \widehat{G}_n de $g(\theta)$ est

1. sans biais si

$$\forall \theta \in \Theta, \mathbb{E}_\theta[\widehat{G}_n] = g(\theta)$$

La quantité $g(\theta) - \mathbb{E}_\theta[\widehat{G}_n]$ est appelée le biais.

2. asymptotiquement sans biais si

$$\forall \theta \in \Theta, \lim_{n \rightarrow \infty} \mathbb{E}_\theta[\widehat{G}_n] = g(\theta)$$

3. consistant (ou convergent) si

$$\forall \theta \in \Theta, \widehat{G}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\theta} g(\theta) \quad (\text{convergence en probabilité})$$

4. fortement consistant (ou fortement convergent) si

$$\forall \theta \in \Theta, \widehat{G}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}_\theta\text{-p.s.}} g(\theta) \quad (\text{convergence presque sûre})$$

Exemple 1.1 : On estime θ par

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

\bar{X}_n est sans biais, et fortement convergent d'après la loi des grands nombres. Si on voulait estimer la variance $\sigma = \sigma(\theta) := \theta(1 - \theta)$, alors, deux choix sont possibles, qui donnent tous les deux un estimateur fortement convergent :

- d'une part, $S_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ est un estimateur biaisé, car $\mathbb{E}_\theta[S_n^2] = \frac{n-1}{n}\sigma$;
- d'autre part, l'estimateur $\frac{n}{n-1}S_n^2$ est toujours un estimateur fortement convergent de σ , cette fois sans biais.

L'inégalité de Bienaymé-Tchebichev nous permet de montrer le lemme suivant :

Lemme 3.1 *Si \hat{G}_n est un estimateur asymptotiquement sans biais, et si $\text{Var}_\theta[\hat{G}_n]$ tend vers 0 lorsque $n \rightarrow +\infty$, alors \hat{G}_n est convergent.*

3.1 Méthode des moments

On suppose donné un échantillon X_1, \dots, X_n de loi \mathbb{P}_θ . On veut estimer (en supposant que la quantité soit bien définie)

$$g(\theta) = \mathbb{E}_\theta[\phi(X_1)].$$

On utilise pour cela la convergence donnée par la loi des grands nombres :

$$\frac{\phi(X_1) + \dots + \phi(X_n)}{n} \xrightarrow[n \rightarrow \infty]{} g(\theta).$$

Par exemple, on estime $\mathbb{E}_\theta[X_1]$ par \bar{X}_n , et $\text{Var}_\theta[X_1]$ par

$$\frac{X_1^2 + \dots + X_n^2}{n} - \left(\frac{X_1 + \dots + X_n}{n} \right)^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 = S_n^2.$$

Avantage de cette méthode : il est souvent très facile d'exhiber des estimateurs simples, sans de grosses difficultés de calculs. **Inconvénient de cette méthode** : les estimateurs obtenus peuvent avoir de mauvaises performances.

3.2 Maximum de vraisemblance

On suppose donné un échantillon X_1, \dots, X_n de loi \mathbb{P}_θ . On supposera ici que \mathbb{P}_θ admet une densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d , notée f_θ , ou par rapport à la mesure de comptage sur un ensemble dénombrable.

Définition 3.3 *La vraisemblance de l'échantillon (X_1, \dots, X_n) est la fonction définie sur Θ par*

$$V_{X_1, \dots, X_n}(\theta) = \prod_{i=1}^n f_\theta(X_i)$$

On note $\hat{\theta}(X_1, \dots, X_n)$ le point qui maximise la fonction $V_{X_1, \dots, X_n}(\theta)$. Cette fonction, à valeurs dans Θ est appelée estimateur du maximum de vraisemblance de θ .

Remarque – Cet estimateur est naturel puisqu'il conduit à privilégier la valeur de θ la "plus probable" au vu de l'observation. Plutôt que de maximiser la vraisemblance, on peut de manière équivalente maximiser le logarithme de la vraisemblance (la log-vraisemblance).

Avantages de cet estimateur :

1. Il peut être construit dans de nombreux cas.
2. On sait montrer les bonnes propriétés de cet estimateur dans des cas très généraux (familles exponentielles par exemple).

Exemple 1.1 : La vraisemblance pour ce modèle est donnée par :

$$V_{X_1, \dots, X_n}(\theta) = \theta^{X_1 + \dots + X_n} (1 - \theta)^{n - (X_1 + \dots + X_n)}$$

Le maximum est atteint en un unique point, \bar{X}_n .

Exemple 1.2 : L'application de cette méthode n'est pas possible si on ne connaît pas la distribution du nombre journalier d'appels. Si la distribution est une loi de Poisson, la vraisemblance s'écrit :

$$V_{X_1, \dots, X_n}(\theta) = e^{-n\theta} \theta^{X_1 + \dots + X_n} \prod_{i=1}^n \frac{1}{X_i!}$$

Le maximum est atteint en un unique point, \bar{X}_n .

3.3 Autres propriétés des estimateurs

Supposons pour simplifier $d = 1$.

Définition 3.4 Le risque quadratique de l'estimateur \hat{G}_n sous \mathbb{P}_θ est

$$\mathcal{R}(\theta, G_n) = \mathbb{E}_\theta [|G_n - g(\theta)|^2] = \text{Var}_\theta[\hat{G}_n] + \underbrace{(\mathbb{E}_\theta[\hat{G}_n] - g(\theta))^2}_{\text{biais}}$$

Pour minimiser le risque quadratique, on préfère donc des estimateurs sans biais, et de variance la plus petite possible.

Définition 3.5 Soit G'_n un autre estimateur de $g(\theta)$. On dit que G_n est préférable à G'_n si, pour tout $\theta \in \Theta$,

$$\mathcal{R}(\theta, G_n) \leq \mathcal{R}(\theta, G'_n)$$

Exemple 1.1 L'estimateur $G_n = \bar{X}_n$ de θ est préférable à l'estimateur $G'_n = X_1$.

Définition 3.6 On dit que l'estimateur G_n est asymptotiquement normal si il existe une suite normalisante $(v_n)_{n \in \mathbb{N}^*}$ qui croît vers l'infini telle que, pour tout $\theta \in \Theta$,

$$v_n(G_n - g(\theta)) \xrightarrow[n \rightarrow \infty]{\text{loi sous } \mathbb{P}_\theta} \mathcal{N}(0, 1)$$

Exemple 1.1 D'après le théorème central limite, l'estimateur \bar{X}_n est asymptotiquement normal, avec comme suite normalisante $v_n = \frac{\sqrt{n}}{\theta(1-\theta)}$.

4 Régions de confiance

4.1 Intervalles de confiance

On voudrait construire, à l'aide d'un échantillon, un intervalle dans lequel on espère que se trouve le paramètre recherché θ , ou plus précisément : on contrôlera la probabilité que θ appartienne à cet intervalle. Pour quantifier la qualité d'un intervalle de confiance I , on choisit un *niveau de confiance* $1 - \alpha$ proche de 1, tel que $\theta \in I$ avec probabilité $1 - \alpha$.

Définition 4.1 Soit $\alpha \in [0, 1]$. On cherche à estimer $g(\theta)$ pour $\theta \in \Theta$ à l'aide d'une observation X_1, \dots, X_n .

Une région de confiance de $g(\theta)$ de niveau de confiance $1 - \alpha$ est un ensemble (dépendant de l'observation) $I_n(X_1, \dots, X_n) \subset g(\Theta)$, mesurable, tel que

$$\forall \theta \in \Theta, \mathbb{P}_\theta[g(\theta) \in I_n] = 1 - \alpha$$

On parle d'intervalle de confiance par excès lorsque le signe d'égalité est remplacé par \geq .

Remarque – Les valeurs usuelles de α sont 0.01, 0.05, 0.1. A niveau de confiance fixé, un intervalle de confiance est d'autant meilleur qu'il est de "taille petite".

Remarque – Un intervalle de confiance n'est pas unique. Si les lois \mathbb{P}_θ sont symétriques, un choix raisonnable est de choisir un intervalle symétrique.

On donne maintenant quelques méthodes pour construire des intervalles de confiance.

4.1.1 Application de l'inégalité de Bienaymé-Tchebichev

Rappelons que si X est une variable aléatoire de carré intégrable, alors

$$\forall \varepsilon > 0, \mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\text{Var}[X]}{\varepsilon^2}$$

Appliquons cette inégalité à l'exemple 1.1, où on estime θ à l'aide de \bar{X}_n :

$$\forall \varepsilon > 0, \mathbb{P}_\theta [|\bar{X}_n - \theta| \geq \varepsilon] \leq \frac{\theta(1-\theta)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

On obtient ainsi un intervalle de confiance par excès de niveau $1 - \alpha$ en posant $\alpha = 1/(4n\varepsilon^2)$ et on obtient

$$I_n = \left[\bar{X}_n - \frac{1}{2\sqrt{n\alpha}}, \bar{X}_n + \frac{1}{2\sqrt{n\alpha}} \right]$$

Remarque – La majoration obtenue par l'application de l'inégalité de Bienaymé-Tchebichev n'est pas très précise, en particulier si la vraie valeur du paramètre est loin de 1/2.

4.1.2 Méthodes des quantiles (peut être sauté si manque de temps)

Définition 4.2 Soit F la fonction de répartition d'une variable aléatoire. Le quantile q_r d'ordre $r \in]0, 1[$ de la loi F est défini par

$$q_r = \inf\{x \in \mathbb{R} ; F(x) \geq r\}$$

Si F est continue, alors $F(q_r) = r$. Si de plus F est strictement croissante, alors q_r est l'unique solution de l'équation $F(x) = r$.

Reprenons l'**Exemple 1.1**. Sous la loi \mathbb{P}_θ , on sait que $n\bar{X}_n$ suit la loi binomiale de paramètre (n, θ) , qu'on note $\mathcal{B}(n, \theta)$. Notons $q_{n,\theta}^-$ et $q_{n,\theta}^+$ les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de la loi $\mathcal{B}(n, \theta)$. Alors, pour tout $\theta \in]0, 1[$,

$$\mathbb{P}_\theta \left[n\bar{X}_n \in [q_{n,\theta}^-, q_{n,\theta}^+] \right] \geq 1 - \alpha$$

(Le fait qu'on n'ait seulement une inégalité provient du fait que les variables aléatoires sont discrètes)

Il faut garder à l'esprit que θ est inconnu, il faut donc inverser les quantiles par rapport au paramètre θ . Les fonctions $\theta \rightarrow q_{n,\theta}^-$ et $\theta \rightarrow q_{n,\theta}^+$ étant croissantes (comparer les fonctions de répartition de $\mathcal{B}(n, \theta_1)$ et de $\mathcal{B}(n, \theta_2)$), on peut définir les "inverses" R_n et L_n de $\frac{q_{n,\theta}^-}{n}$ et $\frac{q_{n,\theta}^+}{n}$ par

$$R_n(x) = \inf \left\{ \theta \in]0, 1[; \frac{q_{n,\theta}^-}{n} \geq x \right\}$$

$$L_n(x) = \inf \left\{ \theta \in]0, 1[; \frac{q_{n,\theta}^+}{n} \geq x \right\}$$

Une écriture équivalente est

$$x \geq \frac{q_{n,\theta}^-}{n} \Leftrightarrow \theta \leq R_n(x)$$

$$x \leq \frac{q_{n,\theta}^+}{n} \Leftrightarrow \theta \geq L_n(x)$$

Dès lors, on a pour tout $\theta \in]0, 1[$,

$$\mathbb{P}_\theta [\theta \in [L_n(\bar{X}_n) ; R_n(\bar{X}_n)]] \geq 1 - \alpha$$

On obtient donc un intervalle de confiance par excès pour θ de niveau $1 - \alpha$ en considérant

$$I_n = [L_n(\bar{X}_n) ; R_n(\bar{X}_n)]$$

Les bornes de l'intervalle s'obtiennent numériquement via le calcul des quantiles de la loi $\mathcal{B}(n, \theta)$.

4.1.3 Application des grandes déviations et de l'inégalité de Hoeffding

Lemme 4.1 (Inégalité de Hoeffding) Soit (Y_1, \dots, Y_n) une suite de variables indépendantes centrées et, pour tout i , $a_i \leq Y_i \leq b_i$ p.s., alors

$$\forall \lambda > 0, \mathbb{P} \left[\sum_{i=1}^n Y_i \geq \lambda \right] \leq \exp \left(\frac{-2\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2} \right)$$

Pour une preuve, aller voir dans les développements d'agreg.

Appliquons cette inégalité à l'**Exemple 1.1**. On a :

$$\forall \varepsilon > 0, \mathbb{P}_\theta [|\bar{X}_n - \theta| \geq \varepsilon] \leq 2 \exp(-2n\varepsilon^2)$$

On obtient un intervalle de confiance de niveau $1 - \alpha$ en considérant

$$\left[\bar{X}_n - \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)} ; \bar{X}_n + \sqrt{\frac{1}{2n} \log \left(\frac{2}{\alpha} \right)} \right]$$

4.2 Intervalles de confiance dans le cas gaussien

On suppose ici que (X_1, \dots, X_n) est un échantillon i.i.d. de loi gaussienne $\mathcal{N}(\mu, \sigma^2)$. On a donc deux paramètres potentiels, μ et σ . On va construire des intervalles de confiance (exacts) dans plusieurs cas. On note Φ la fonction de répartition de la loi normale centrée réduite, et $x_\alpha := \Phi^{-1}(\alpha)$ pour tout $\alpha \in [0, 1]$. Autrement dit,

$$\int_{-\infty}^{x_\alpha} e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} = \alpha.$$

4.2.1 Estimation de la moyenne μ

1. **Si l'écart-type σ est connu.** On sait que

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

donc l'intervalle

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

est un intervalle de confiance de niveau $1 - \alpha$ pour le paramètre μ . On remarque qu'on a construit une statistique dont la loi ne dépend pas du paramètre (μ) à estimer. La statistique fait intervenir les données (σ), les observations (\bar{X}_n), et le paramètre à estimer, que l'on peut exhiber ensuite de manière simple.

2. **Si l'écart-type σ est inconnu.** Dans ce cas, la statistique précédente ne peut plus être utilisée car on n'a pas la valeur de σ . Par contre, on sait, d'après le théorème de Cochran, que

$$T_n := \frac{\bar{X}_n - \mu}{S_n} \sqrt{n-1} \sim \text{Student}(n-1).$$

De la même façon que pour la loi normale centrée réduite, on peut trouver la valeur de $t_{\alpha/2}$ tel que $\mathbb{P}(-t_{\alpha/2} \leq T_n \leq t_{\alpha/2}) = 1 - \alpha$. Et l'intervalle de confiance pour μ de niveau $1 - \alpha$ s'écrit alors

$$\left[\bar{X}_n - t_{\alpha/2} \frac{S_n}{\sqrt{n-1}} ; \bar{X}_n + t_{\alpha/2} \frac{S_n}{\sqrt{n-1}} \right].$$

4.2.2 Estimation de l'écart-type σ

1. **Si la moyenne μ est connue.** On utilise le fait que

$$\Sigma_n^2 := \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

est un estimateur sans biais (de variance minimale!) de σ^2 et de plus, $n\Sigma_n^2/\sigma^2$ suit une loi χ_n^2 (à n degrés de liberté). On note $u_{\alpha/2}$ et $u_{1-\alpha/2}$ les quantiles de la loi χ_{n-1}^2 (cette fois, ils ne sont pas opposés car la densité de cette loi n'est pas paire, contrairement à la loi normale par exemple). L'intervalle de confiance pour σ^2 de niveau $1 - \alpha$ est donné par

$$\left[\frac{n\Sigma_n^2}{u_{1-\alpha/2}} ; \frac{n\Sigma_n^2}{u_{\alpha/2}} \right]$$

2. **Si la moyenne μ est inconnue.** On sait cette fois que nS_n^2/σ^2 suit la loi χ_{n-1}^2 . On note $v_{\alpha/2}$ et $v_{1-\alpha/2}$ les quantiles de la loi χ_{n-1}^2 (attention aux degrés de liberté!). On obtient alors l'intervalle de confiance pour σ^2 , de niveau $1 - \alpha$, par :

$$\left[\frac{nS_n^2}{v_{1-\alpha/2}} ; \frac{nS_n^2}{v_{\alpha/2}} \right]$$

4.3 Exercices

1. Une usine fabrique des balances, pouvant commettre des erreurs de mesure dont l'écart-type vaut 0,1 kg. Un inspecteur décide de faire des tests sur 100 balances. Il pèse 100 fois un même objet. Résultat moyen des pesées : 55,4 kg. Donner un intervalle de confiance pour le poids de l'objet de niveau de confiance 0,95. Expliquer comment les constructeurs de balance ont pu fournir une valeur de σ , et donner aussi un intervalle de confiance.

2. On a mesuré le poids de raisins sur une souche, en prenant 10 souches au hasard dans une vigne. On a obtenu les résultats suivants (en kg) :

2,4 3,2 3,6 4,1 4,3 4,7 5,4 5,9 6,5 6,9

On modélise le poids de chaque souche comme étant distribué selon une loi $\mathcal{N}(m, \sigma^2)$. Le vigneron cherche à estimer m . Donner une estimation et un intervalle de confiance pour m , de niveau de confiance 0,95.

4.4 Intervalles de confiance asymptotiques

Lorsque nous disposons de suffisamment de données X_1, \dots, X_n , et pour les modèles plus classiques, il est souvent plus simple de remplacer la loi de la statistique par son asymptotique lorsque n est grand. Le théorème central limite s'avère alors un excellent outil, et on parle d'*intervalle de confiance asymptotique*.

Définition 4.3 On dit que la suite $(I_n)_{n \geq 1}$ est un intervalle de confiance asymptotique de niveau $1 - \alpha$ pour $g(\theta)$ si, pour tout $\theta \in \Theta$,

$$\lim_{n \rightarrow \infty} \mathbb{P}_\theta [g(\theta) \in I_n(X_1, \dots, X_n)] = 1 - \alpha$$

Notons $v = \theta(1 - \theta)$ la variance de X_1 . Le théorème central limite nous dit que

$$\mathcal{L} \left(\sqrt{\frac{n}{v}} (\bar{X}_n - \theta) \right) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1) \quad \text{sous } \mathbb{P}_\theta$$

Ainsi, si $q_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi $\mathcal{N}(0, 1)$,

$$\mathbb{P}_\theta \left[\theta \in \left[\bar{X}_n - q_{1-\alpha/2} \sqrt{\frac{v}{n}}; \bar{X}_n + q_{1-\alpha/2} \sqrt{\frac{v}{n}} \right] \right] \xrightarrow[n \rightarrow \infty]{} 1 - \alpha$$

Le point important est que l'on ne sait pas à quelle vitesse a lieu cette convergence. Il existe des résultats théoriques pour contrôler cette erreur mais ils sont en général trop pessimistes et font intervenir des paramètres de la loi de l'échantillon que l'on ne dispose pas. En pratique, les livres conseillent d'utiliser cette approximation pour $n \geq 30$.

D'autre part, le défaut de ce raisonnement est que l'on ne connaît pas v puisqu'il dépend du paramètre. Pour corriger ceci, on utilise des outils plus élaborés. Le lemme de Slutsky permet de surmonter certaines difficultés. La méthode delta permet d'obtenir des régions de confiance plus intéressantes. Elle permet surtout d'obtenir la normalité asymptotique de nombreux estimateurs pour l'estimation de $g(\theta)$.

Lemme 4.2 (Lemme de Slutsky) Soient $(X_n)_{n \in \mathbb{N}}$ et $(Y_n)_{n \in \mathbb{N}}$ deux suites de vecteurs de \mathbb{R}^d et \mathbb{R}^p , tels que

- $X_n \xrightarrow[n \rightarrow \infty]{} X$ en loi, où X est un vecteur aléatoire,
- $Y_n \xrightarrow[n \rightarrow \infty]{} c$ en probabilité, où c est un vecteur déterministe.

Alors, $(X_n, Y_n) \xrightarrow[n \rightarrow \infty]{} (X, c)$ en loi.

Pour une preuve, cf le poly de Vincent Rivoirard.

En reprenant l'exemple 1.1, et puisque $\bar{X}_n(1 - \bar{X}_n) \xrightarrow[n \rightarrow \infty]{} \theta(1 - \theta)$, on obtient :

$$\mathcal{L} \left(\sqrt{\frac{n}{\bar{X}_n(1 - \bar{X}_n)}} (\bar{X}_n - \theta) \right) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}(0, 1) \quad \text{sous } \mathbb{P}_\theta$$

d'où :

$$\mathbb{P}_\theta \left[\theta \in \left[\bar{X}_n - q_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}}; \bar{X}_n + q_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1 - \bar{X}_n)}{n}} \right] \right] \xrightarrow[n \rightarrow \infty]{} 1 - \alpha$$

Pour $n\theta(1 - \theta)$ assez grand, on obtient un intervalle satisfaisant. Un problème survient lorsque θ est proche de 0 ou 1 en raison de la mauvaise qualité de l'estimation de $\theta(1 - \theta)$.

4.5 Application à la méthode de Monte-Carlo

La méthode de Monte-Carlo permet le calcul de valeurs approchées d'intégrales multiples en utilisant des réalisations i.i.d de loi uniforme. Si $(X_n)_{n \geq 1}$ est une suite de v.a. i.i.d de loi uniforme sur $[0, 1]^m$, et si $f : [0, 1]^m \rightarrow \mathbb{R}$ est une fonction mesurable, alors la loi des grands nombres entraîne la convergence presque sûre suivante :

$$\frac{1}{n} (f(X_1) + \dots + f(X_n)) \xrightarrow[n \rightarrow \infty]{} \mathbb{E}[f(X_1)] = \int_{[0,1]^m} f(x) dx$$

Si la variance de $f(X_1)$ est majorable, on peut utiliser les intervalles de confiance pour la moyenne. On voit alors que la convergence a lieu à la vitesse \sqrt{n} . Cette vitesse est assez lente, comparée aux méthodes déterministes, mais on remarquera qu'aucune régularité sur f n'est requise ici. La méthode de Monte-Carlo reste valable lorsque f est à support quelconque, en prenant par exemple des X_i de loi gaussienne.

5 Tests d'hypothèse

5.1 Généralités

L'une des fonctions des statistiques est de prendre des décisions, à partir d'observations d'un phénomène aléatoire (ou modélisé comme tel). Par exemple, peut-on considérer qu'un médicament donné est plus efficace qu'un autre? Les gènes codant la couleur des yeux et celle des cheveux sont-ils sur les mêmes chromosomes? Étant donné le nombre de composants défectueux présents à la sortie de l'usine, l'industriel doit-il en arrêter la production? Il y a deux points communs à toutes ces questions : on y répond par oui ou non, et le phénomène sous-jacent est aléatoire. Les tests statistiques vont permettre d'apporter une réponse à ces problèmes de décision.

Dans le cadre de notre modèle statistique $(\Omega, \mathcal{A}, (\mathbb{P}_\theta)_{\theta \in \Theta})$, on se donne Θ_0 et Θ_1 disjoints tels que $\Theta_0 \cup \Theta_1 = \Theta$. Au vu d'une observation X , on veut déterminer si $\theta \in \Theta_0$ ou non. On note :

- l'hypothèse nulle $H_0 : \theta \in \Theta_0$,
- l'hypothèse alternative $H_1 : \theta \in \Theta_1$.

L'hypothèse H_0 est dite *simple* si Θ_0 est réduit à un singleton, et *composite* sinon. Nous verrons que H_0 et H_1 ne jouent pas des rôles symétriques.

Définition 5.1 *On se donne un échantillon (X_1, \dots, X_n) de loi \mathbb{P}_θ . On dit qu'on fait un test d'hypothèse H_0 contre l'hypothèse H_1 lorsque, après avoir déterminé si une certaine observation $G(X_1, \dots, X_n)$ appartient ou non à une zone $R \subset \mathbb{R}$, on prend la décision suivante :*

- si $X \notin R$, alors on accepte H_0 ,
- si $X \in R$, alors on rejète H_0 .

L'ensemble R est appelé zone de rejet de H_0 .

Définition 5.2

1. On dit qu'on commet une erreur de première espèce lorsque H_0 est vraie, mais est rejetée par le test. L'erreur de première espèce est donc la fonction définie sur Θ_0 par :

$$\theta \in \Theta_0 \rightarrow \mathbb{P}_\theta(X \in R)$$

2. On dit qu'on commet une erreur de seconde espèce lorsque H_1 est vraie, mais que H_0 est acceptée. L'erreur de seconde espèce est donc la fonction définie sur Θ_1 par :

$$\theta \in \Theta_1 \rightarrow \mathbb{P}_\theta(X \notin R)$$

3. On dit que le test est de niveau α si α majore l'erreur de première espèce :

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(X \in R) \leq \alpha$$

Plus α est petit, plus la zone de rejet sera petite. La valeur $1 - \alpha$ est aussi appelée *confiance* du test. Autrement dit, une proportion α des situations où H_0 serait vraie donneront en fait un résultat en faveur de H_1 .

On considère que rejeter H_0 à tort est plus préjudiciable que qu'accepter H_0 à tort. La stratégie consistera à construire le test (et donc la zone de rejet) de façon à minimiser en priorité l'erreur de première espèce. En particulier, l'hypothèse nulle représentera souvent ce que l'on pense être la réalité.

Exemple 5.1 On vient d'acheter un dé à six faces sensé fournir des résultats distribués selon la loi uniforme sur $\{1, \dots, 6\}$. On souhaite valider (ou invalider) cette information du fabricant à partir de l'observation de n lancers de dé. On est donc amené à poser les hypothèses :

$$H_0 : \text{le dé est équilibré,} \quad \text{et} \quad H_1 : \text{le dé est pipé}$$

Exemple 5.2 Un industriel souhaite évaluer la qualité de ses produits. Après l'observation de n pièces, il voudrait estimer la proportion de composants défectueux : si cette proportion est trop grande, il sera contraint d'arrêter temporairement sa production le temps de rectifier les machines d'usinage. Une telle éventualité ayant un certain coût, il ne voudra faire ce choix qu'en cas de grande nécessité.

Notre observation est donc une suite de v.a i.i.d de loi de Bernoulli de paramètre p : la valeur 1 représente un composant défectueux. Après avoir déterminer un seuil p_s (au delà duquel la production devra être arrêtée), la question que l'industriel se pose est : p est-il plus grand ou plus petit que p_s ?

Du point de vue de l'industriel, réparer ses machines entraînerait des frais non négligeables. Il ne souhaite pas en arriver là si cela n'est pas nécessaire, souhaitant minimiser le risque de décider $p > p_s$, alors qu'en réalité $p \leq p_s$. Ainsi, on choisit :

$$H_0 : p \leq p_s \quad \text{et} \quad H_1 : p > p_s$$

Si au contraire, nous nous plaçons du point de vue d'un organisme de qualité, l'intérêt serait d'être le plus sûr possible que le seuil n'est pas atteint. On inverserait alors les deux hypothèses.

On choisit souvent le niveau $\alpha = 0.05$. Une fois le niveau fixé, on choisit en second lieu de contrôler la probabilité de rejeter H_0 lorsque cela est nécessaire.

Définition 5.3 On appelle puissance du test la fonction définie sur Θ_1 par :

$$\theta \in \Theta_1 \rightarrow \mathbb{P}_\theta(X \in R)$$

Autrement dit, la puissance est la probabilité d'accepter H_1 alors que H_1 est vraie. L'objectif est donc de construire, parmi les tests de niveau α , ceux dont la puissance est maximale. Les deux motivations sont en conflit :

- plus le niveau est faible, plus la zone de rejet est petite, et alors la puissance diminue,
- plus Θ_1 est grand, plus la zone de rejet est grande, ce qui augmente la puissance, mais également le niveau.

Exemple 5.3 *Le test d'hypothèses $H_0 : p = 0.2$ et $H_1 : p \neq 0.2$ est plus puissant que $H_0 : p = 0.2$ et $H_1 : p = 0.4$.*

Toutes ces considérations peuvent nous aider à déterminer H_0 et H_1 . Par exemple, on prendra pour H_0 :

- une hypothèse qu'on pense être vraie,
- une hypothèse de prudence (critère de coût, de sécurité, ...)
- la seule facile à formuler

La démarche est alors la suivante :

1. Choix de H_0 et H_1 .
2. Détermination de la statistique de test $\phi(X)$: on doit en connaître la loi sous H_0 .
3. Étude de l'allure de la zone de rejet en fonction de la forme de H_1 .
4. Calcul de la zone de rejet, à niveau α préalablement fixé.
5. Utilisation des données, calcul, et conclusion.

Comme précédemment, il peut être plus intéressant d'utiliser les théorèmes limites et de considérer des échantillons de plus en plus grands.

Définition 5.4 *Soit (X_1, \dots, X_n) une observation, et R_n une zone de rejet de niveau α associée aux hypothèses H_0 et H_1 . On dit que le test est de niveau asymptotique α si*

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \Theta} \mathbb{P}_\theta(R_n) \leq \alpha$$

5.2 Test du χ^2

Le test du χ^2 est un exemple relativement simple de test statistique qui permet de tester :

1. l'adéquation à une loi de probabilité sur un ensemble fini (par exemple, est-il raisonnable de penser que les résultats que j'observe sont des réalisations i.i.d d'une loi (p_1, \dots, p_k) sur $\{1, \dots, k\}$?),
2. l'indépendance de deux caractères mesurés sur un même individu,
3. l'homogénéité de plusieurs échantillons (par exemple, deux médicaments ont-ils le même effet sur la population ?).

Considérons un phénomène aléatoire de loi inconnue $P = p_1\delta_1 + \dots + p_k\delta_k$. Des considérations extérieures nous font penser que la loi du phénomène est $Q = q_1\delta_1 + \dots + q_k\delta_k$. Partant de n observations, on désire tester l'hypothèse $H_0 : P = Q$ contre l'hypothèse $H_1 : P \neq Q$.

On note les effectifs des n observations indépendantes par N_1^n, \dots, N_k^n (le nombre N_i^n représente le nombre de fois où on a obtenu i dans l'échantillon de taille n). On désigne par p_i^n la distribution empirique associée définie par :

$$p_i^n = \frac{N_i^n}{n}$$

et on note $P_n := (p_1^n, \dots, p_k^n)$.

On définit la pseudo-distance du χ^2 entre deux distributions A et B discrètes sur $\{1, \dots, k\}$ par :

$$\chi^2(A, B) = \sum_{i=1}^k \frac{(A_i - B_i)^2}{B_i}$$

On montre alors que :

Proposition 5.1

1. La variable aléatoire $\sqrt{n} \left(\frac{p_1^n - p_1}{\sqrt{p_1}}, \dots, \frac{p_k^n - p_k}{\sqrt{p_k}} \right)$ converge en loi vers la loi normale centrée sur \mathbb{R}^k de matrice de covariance $I_k - \sqrt{P}(\sqrt{P})^*$.
2. Sous H_0 , $n \cdot \chi^2(P_n, Q)$ converge en loi vers la loi $\chi^2(k - 1)$.
3. Sous H_1 , $n \cdot \chi^2(P_n, Q)$ converge en probabilité vers $+\infty$.

Pour une preuve, cf le poly de F. Malrieux, prépa agreg de Rennes + Vincent Rivoirard.

Le fait que la loi limite ne dépende du paramètre P qu'à travers sa taille k est très important ici. On est ainsi capable de construire une statistique de test : $n \cdot \chi^2(P_n, Q)$, et de choisir une région de \mathbb{R} , telle que la statistique appartienne à cette région avec une probabilité (asymptotique) donnée.

Plus précisément, notons χ_α^2 le quantile d'ordre $1 - \alpha$ de la loi $\chi^2(k - 1)$. Un test du χ^2 de niveau asymptotique α consiste en ce qui suit, pour n "assez grand" :

- Si $n \cdot \chi^2(P_n, Q) > \chi_\alpha^2$, alors on choisit H_1 ,
- Si $n \cdot \chi^2(P_n, Q) \leq \chi_\alpha^2$, alors on choisit H_0 .

Avant de donner plusieurs exemples, ajoutons un autre résultat, qui n'est qu'une conséquence directe de la loi des grands nombres mais qui joue un rôle essentiel dans l'élaboration des tests du χ^2 :

Proposition 5.2 Sous H_1 , $\chi^2(P_n, Q)$ converge presque sûrement vers $\sum_{i=1}^k \frac{(p_i - q_i)^2}{q_i}$.

Remarque – Les tests du χ^2 ne donnent vraiment d'information que si l'hypothèse H_0 est rejetée. En effet, si H_0 n'est pas rejetée, il se peut bien que ce soit parce que la loi p de l'échantillon est dans H_1 , mais tout près de H_0 . Ceci est encore renforcé quand on est obligé de créer des classes pour des lois continues par exemple (voir le paragraphe suivant) : des tas de lois fourniront les même vecteurs de probabilité sur l'ensemble fini. *Moralité* : on se sert du test du χ^2 (ou de tout autre test non-paramétrique) pour invalider un modèle. Si H_0 est rejetée, alors il faut changer de modèle. Si non, c'est que le modèle (bien que simpliste, approximatif... et vraisemblablement faux) est satisfaisant.

5.2.1 Test d'adéquation

Soit (X_1, \dots, X_n) un n -échantillon réel de loi inconnue μ . On suppose que les X_i prennent les valeurs dans k classes I_1, \dots, I_k constituées de k intervalles ou de k nombres selon que la loi μ est discrète ou continue. On note \hat{f}_i la fréquence empirique associée à la classe i :

$$\hat{f}_i = \frac{1}{n} \sum_{m=1}^n \mathbb{1}_{X_m \in I_i}$$

Soit ν une loi de probabilité connue, de même nature que μ . On doit tester l'hypothèse H_0 : "la loi des X_i est ν " contre H_1 : "la loi des X_i n'est pas ν ".

Les $f_i := \nu(I_i)$ sont calculables a priori, et on a alors sous H_0 :

$$n \cdot \sum_{i=1}^k \frac{(f_i - \hat{f}_i)^2}{f_i} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(k - 1)$$

Remarque – Le test du χ^2 s'appuie sur deux résultats asymptotiques (une convergence en loi et une convergence en probabilité). Or, on ne dispose jamais que d'un nombre fini d'observations. Toute la question est de savoir si l'on a le droit de faire comme si la limite en loi était une égalité. En pratique, on dit que pour que le test soit valide, il faut que, pour tout $i = 1, \dots, k$, $n \cdot f_i$ soit supérieur ou égal à 5. Si ce n'est pas le cas, il faut regrouper des classes à faible effectif pour atteindre le seuil exigé.

Exemple 5.4 (*Historique*) Pour tester sa théorie génétique, Mendel croisa des pois tous jaunes et lisses et obtint à la première génération des pois jaunes ou verts et lisses ou ridés. Plus précisément, il obtint 315 pois jaunes et lisses, 108 pois verts et lisses, 101 pois jaunes et ridés, et 32 pois verts et ridés. Est-ce que ces observations confirment ou infirment la théorie mendélienne ? Sous cette approche, la proportion p_0 de chacune des classes précédentes est $p_0 = \left(\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \right)$. On teste donc $H_0 : p = p_0$ contre $H_1 : p \neq p_0$. Quel est le résultat obtenu ?

5.2.2 Test d'homogénéité

Soit (X_1, \dots, X_n) un n -échantillon réel de loi inconnue μ_1 et (Y_1, \dots, Y_m) un m -échantillon réel de loi inconnue μ_2 , les deux échantillons étant indépendants l'un de l'autre. On suppose que les X_i et Y_i prennent leurs valeurs dans k classes (I_1, \dots, I_k) . On note \hat{f}_i^X et \hat{f}_i^Y les fréquences empiriques associées à la classe $i \in \{1, \dots, k\}$. On désire tester l'hypothèse H_0 : " μ_1 et μ_2 sont identiques" contre H_1 : " μ_1 et μ_2 sont différentes".

On définit une sorte de fréquence empirique commune aux deux échantillons en posant :

$$\hat{f}_i := \frac{n \cdot \hat{f}_i^X + m \cdot \hat{f}_i^Y}{n + m}$$

On a alors, sous H_0 :

$$n \cdot \sum_{i=1}^k \frac{(\hat{f}_i - \hat{f}_i^X)^2}{\hat{f}_i} + m \cdot \sum_{i=1}^k \frac{(\hat{f}_i - \hat{f}_i^Y)^2}{\hat{f}_i} \xrightarrow[n, m \rightarrow \infty]{\mathcal{L}} \chi^2(k-1)$$

Exemple 5.5 On cherche à invalider l'affirmation qui consiste à dire que toutes les lessives se valent. On utilise 3 lessives appelées A, B, et C. A la sortie du lavage, on classe les vêtements en 3 catégories : très sale (TS), légèrement sale (LS), et propre (P).

	TS	LS	P
A	30	65	205
B	23	56	121
C	75	125	300

Peut-on dire, au niveau 0.05, que toutes les lessives sont identiques ?

5.2.3 Test d'indépendance

Soit $((X_1, Y_1), \dots, (X_n, Y_n))$ un n -échantillon réel de loi inconnue μ . On note μ_1 et μ_2 les lois marginales de μ et l'on suppose que les (X_i, Y_i) prennent respectivement leurs valeurs sur k classes (I_1, \dots, I_k) et ℓ classes (J_1, \dots, J_ℓ) . On note $\hat{f}_{i,j}$ la fréquence empirique associée à (I_i, J_j) , définie par :

$$\hat{f}_{i,j} := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k \in I_i \text{ et } Y_k \in J_j}$$

On désire tester l'hypothèse H_0 : " $\mu = \mu_1 \otimes \mu_2$ " contre H_1 : " $\mu \neq \mu_1 \otimes \mu_2$ ".

Pour cela, on définit $f_{i,j}$ par :

$$f_{i,j} := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{X_k \in I_i} \cdot \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{Y_k \in I_j}$$

On a alors, sous H_0 :

$$n \sum_{i=1}^k \sum_{j=1}^{\ell} \frac{(f_{i,j} - \widehat{f}_{i,j})^2}{f_{i,j}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2((k-1)(\ell-1))$$

(Pour une preuve, cf le poly de Rennes)

Exemple 5.6 (Yeux et cheveux). Parmi une population de 124 personnes, on note la couleur de leurs yeux et de leurs cheveux :

	blonds	bruns	roux	noirs
bleus	25	9	7	3
gris	13	17	7	10
marrons	7	13	5	8

Les deux critères sont-ils indépendants au niveau 0.05 ?

5.3 Test de Kolmogorov-Smirnov

Soit μ une mesure de probabilité sur \mathbb{R} de fonction de répartition F , et $(X_n)_{n \geq 1}$ une suite de v.a. i.i.d. de loi μ .

Définition 5.5

1. La mesure empirique d'un échantillon (X_1, \dots, X_n) est définie par

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$$

2. La fonction de répartition empirique d'un échantillon (X_1, \dots, X_n) est définie par

$$\forall t \in \mathbb{R}, F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{X_i \leq t}$$

Remarque – La loi μ_n est une mesure aléatoire, de fonction de répartition F_n . La fonction F_n est constante par morceaux.

Une conséquence de la loi des grands nombres est que la mesure empirique μ_n converge étroitement vers la loi μ :

$$\forall f \in \mathcal{C}_b^0(\mathbb{R}), \frac{1}{n} \sum_{i=1}^n f(X_i) = \int_{\mathbb{R}} f d\mu_n \xrightarrow[n \rightarrow \infty]{} \int_{\mathbb{R}} f d\mu$$

Ce résultat de convergence étroite des mesures empiriques peut être affiné en une convergence uniforme des fonctions de répartition, comme le précise le théorème de Glivenko-Cantelli :

Théorème 5.1 (Glivenko-Cantelli) Presque sûrement, la fonction de répartition empirique converge uniformément vers la fonction de répartition de μ :

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \|F_n - F\| \xrightarrow[n \rightarrow \infty]{p.s.} 0$$

Supposons désormais que F est continue, et donc que μ n'a pas de masses ponctuelles. Alors, on peut préciser la vitesse de convergence dans le théorème précédent :

Théorème 5.2 *Si F est continue, alors*

$$\sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_{KS}$$

où D_{KS} est une variable aléatoire dont la loi de probabilité sur \mathbb{R}_+ , appelée distribution de Kolmogorov-Smirnov, est donnée par sa fonction de répartition F_{KS} :

$$\forall t > 0, F_{KS}(t) = 1 + 2 \sum_{k=1}^{+\infty} (-1)^k e^{-2k^2 t^2}$$

5.3.1 Test de Kolmogorov-Smirnov

On voudrait savoir si l'échantillon obtenu X_1, \dots, X_n est bien de loi μ_0 . Il est très important de vérifier que la mesure μ_0 n'a pas d'atomes.

D'après le théorème précédent, on va identifier la loi de $\sqrt{n} \|F_n(x) - F(x)\|_\infty$ à celle de Kolmogorov-Smirnov pour n assez grand. On peut utiliser les quantiles à 0.95 et 0.99 pour obtenir un test de niveau 0.05 et 0.01 sur l'adéquation de la loi de l'échantillon μ .

Plus précisément, on teste H_0 : "la fonction de répartition réelle F est F_0 " contre H_1 : " $F \neq F_0$ ". On fixe α le niveau du test. On note $D_n = \sqrt{n} \|F_n - F_0\|_\infty$, qui se calcule aisément. Alors :

$$D_n \begin{cases} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} D_{KS} & \text{sous } H_0 \\ \xrightarrow[n \rightarrow \infty]{p.s.} +\infty & \text{sous } H_1 \text{ car } \|F_n - F_0\|_\infty \xrightarrow[n \rightarrow \infty]{p.s.} \|F - F_0\|_\infty > 0 \end{cases}$$

On choisit comme zone de rejet $R_\alpha = [q_{1-\alpha}; +\infty[$ où $q_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$ de la loi de Kolmogorov-Smirnov.