

Statistiques et Apprentissage, Amphi de révisions

Automatants

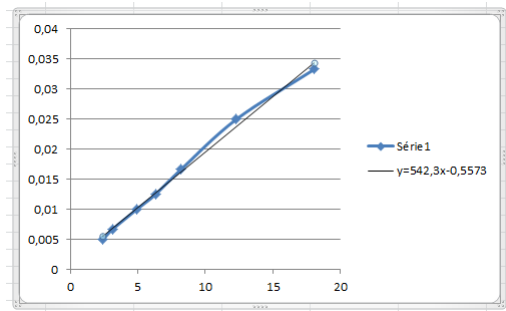
Axel Darmouni Karim Khaldi

6 Mai 2021

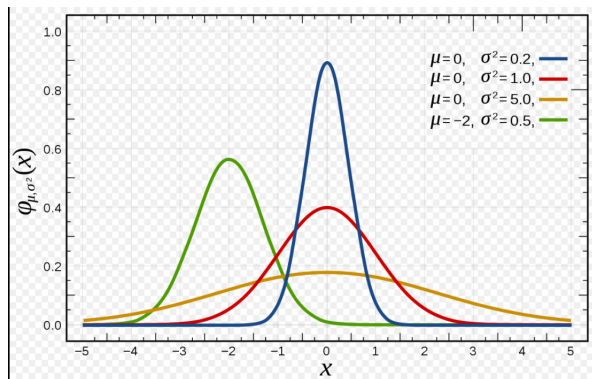


AUTOMATANTS

- Outils de mesures actuels de plus en plus poussés
- But : étudier les phénomènes pour pouvoir les modéliser, et obtenir plus d'informations sur le monde qui nous entoure



Outils d'étude principaux : les modèles paramétriques



Echantillon aléatoire

Soit $X: (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{A})$ variable aléatoire de loi P_X . On appelle **échantillon aléatoire de taille N la variable aléatoire** :

$$E: (\Omega^N, \mathcal{F}^{\otimes N}) \rightarrow (\mathcal{X}^N, \mathcal{A}^{\otimes N})$$
$$(\omega_1, \dots, \omega_N) \mapsto (X(\omega_1), \dots, X(\omega_N))$$

Si $E \sim \mathbb{P}_X^{\otimes N}$, alors il est équivalent de considérer X_1, \dots, X_N , des VAs indépendantes de même loi (abrégé plus tard en iid) \mathbb{P}_X . Le vecteur aléatoire (X_1, \dots, X_N) est alors dit **échantillon d'observation**.

Statistique

En reprenant (X_1, \dots, X_N) échantillon aléatoire dans $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$, et en prenant (Z, \mathcal{C}) un espace mesurable : si T est une application mesurable de $(\mathcal{X}^N, \mathcal{A}^{\otimes N})$ dans (Z, \mathcal{C}) , alors $T_N = T \circ (X_1, \dots, X_N)$ est une VA de $(\Omega^N, \mathcal{F}^{\otimes N})$ dans (Z, \mathcal{C}) appelée **statistique de l'échantillon**.

Modèle Paramétrique

On définit un **Modèle statistique** par une famille de lois de probabilité sur $(\mathcal{X}, \mathcal{A})$. Si la famille est paramétrisable par $\theta \in \Theta \subset \mathbb{R}^d$, c'est un modèle statistique paramétrisable qu'on note $M_\theta = \{\mathbb{P}_\theta, \theta \in \Theta\}$ ($= \{p_\theta, \theta \in \Theta\}$ si la loi admet une densité)

Estimateur

Le problème est le suivant : on considère, pour un échantillon iid (X_1, \dots, X_N) que chaque X_i suit une loi P_{θ^*} , θ^* que l'on cherche à trouver. On définit alors un **estimateur de θ^*** comme une statistique $T(X_1, \dots, X_N)$ visant à estimer θ^* .

Méthode des moments

Ici, on cherche à estimer la loi à partir de sa moyenne, sa variance, ... Pour cela, on cherche pour un échantillon aléatoire iid (X_1, \dots, X_N) de loi appartenant à $\{P_\theta/\theta \in \Theta\}$, θ tel que, à k fixé,

$$\mathbb{E}_{P_\theta}(X_1^k) = \frac{\sum_{i=1}^n X_i^k}{n} \text{ ie tel que } \int_{\mathbb{R}} x^k P_\theta(dx) = \frac{\sum_{i=1}^n X_i^k}{n}$$

Remarques

Ici, on parle de variables aléatoires, mais dans les faits, on traitera de valeurs empiriques. Par exemple, si l'on a les valeurs (x_1, \dots, x_N) , on cherchera si l'on étudie le premier moment, θ tel que,

$$\mathbb{E}_{P_\theta}(X_1) = \frac{\sum_{i=1}^n x_i}{n} = \bar{x}$$

Aussi, par la suite, on notera à la place de \mathbb{E}_{P_θ} : \mathbb{E}_θ .

Exemple

Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre α avec $\alpha \in [0, 1]$.

Estimons α avec le moment d'ordre 1, à l'aide de (x_1, \dots, x_N) un N -échantillon d'observations iid de loi inconnue.

On a :

$$\mathbb{E}_\alpha[X] = \alpha$$

Donc en égalisant $\mathbb{E}_\alpha[X]$ avec l'espérance empirique, on a :

$$\alpha = \frac{\sum_{i=1}^n x_i}{n}$$

Et donc, on pose $\alpha = \bar{x}$ (où \bar{x} est la moyenne empirique), et on a alors comme estimateur de α ,

$$\hat{\alpha} = \frac{\sum_{i=1}^n X_i}{n}$$

Définition

Pour un échantillon d'observations (x_1, \dots, x_N) , on définit :

La vraisemblance du paramètre θ

$$\mathcal{L}(\theta; x_1, \dots, x_N) = \prod_{i=1}^N p(x_i | \theta)$$

Remarque

La vraisemblance peut être vue comme la probabilité de θ pour un jeu de données. Plus celle-ci est proche de 0, moins il paraît possible d'obtenir ces observations avec θ , et plus celle-ci est proche de 1, plus ces observations sont plausibles.

Définitions

Pour un échantillon d'observations (x_1, \dots, x_N) , on définit :

La vraisemblance du paramètre θ

$$\mathcal{L}(\theta; x_1, \dots, x_N) = \prod_{i=1}^N p(x_i | \theta)$$

La Log-vraisemblance

$$\ln(\mathcal{L}(\theta; x))$$

L'estimateur du maximum de vraisemblance

$$\hat{\theta}(x) = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}(\theta; x)$$

Exemple

Reprenons X suivant une loi de Bernoulli de paramètre α avec $\alpha \in [0, 1]$.
Estimons α avec la méthode du maximum de vraisemblance, à l'aide de (x_1, \dots, x_N) un N -échantillon d'observations iid de loi inconnue.

On a :

$$\mathcal{L}(\alpha; x_1, \dots, x_N) = \prod_{i=1}^N p(x_i | \alpha)$$

Soit :

$$\mathcal{L}(\alpha; x_1, \dots, x_N) = \left(\prod_{i=1, x_i=1}^N \alpha \right) \left(\prod_{i=1, x_i=0}^N (1 - \alpha) \right)$$

Exemple

Reprenons X suivant une loi de Bernoulli de paramètre α avec $\alpha \in [0, 1]$.
Estimons α avec la méthode du maximum de vraisemblance, à l'aide de
 (x_1, \dots, x_N) un N -échantillon d'observations iid de loi inconnue.

On a alors

$$\mathcal{L}(\alpha; x_1, \dots, x_N) = \alpha^{\sum_{i=1}^N x_i} \cdot (1 - \alpha)^{N - \sum_{i=1}^N x_i}$$

Sauf que, maximiser un produit, c'est compliqué ; mais maximiser une somme, ça l'est moins ! Donc on regarde $l(\alpha) = \ln(\mathcal{L}(\alpha))$ (pour simplifier la notation, on omet les x_i). On va donc maintenant étudier les cas d'annulation de $\frac{dl}{d\alpha}$.

Exemple

Reprenons X suivant une loi de Bernoulli de paramètre α avec $\alpha \in [0, 1]$.
Estimons α avec la méthode du maximum de vraisemblance, à l'aide de (x_1, \dots, x_N) un N -échantillon d'observations iid de loi inconnue.
On avait déjà $l(\alpha) = N\bar{x}\ln(\alpha) + N(1 - \bar{x})\ln(1 - \alpha)$ (où \bar{x} est la moyenne empirique). Cela donne donc :

$$\frac{N\bar{x}}{\alpha} - \frac{N(1 - \bar{x})}{1 - \alpha} = 0$$

On obtient alors, en jouant avec les termes, que cela vaut 0 si et seulement si $\alpha = \bar{x}$. (On retrouve le même estimateur que pour le moment d'ordre 1, mais c'est du pur hasard)

Comment différencier les estimateurs ?



Comment différencier les estimateurs ?

- On a pu voir qu'on avait différentes méthodes permettant d'obtenir des estimateurs.
- Toutefois, notre but est de trouver des estimateurs optimaux, permettant le plus d'approcher nos données.
- Comment trouver alors de nos estimateurs, ceux réalisant la meilleure modélisation ?

Risque quadratique

On définit le **Risque quadratique** pour un estimateur T du paramètre θ de la manière suivante :

$$R(T, \theta) = \mathbb{E}_{\theta}(\|T - \theta\|^2)$$

Première comparaison

On dit que T est un **meilleur estimateur** que T' , ssi,

$$\begin{cases} R(T, \theta) \leq R(T', \theta), \forall \theta \\ \exists \theta \mid R(T, \theta) < R(T', \theta) \end{cases}$$

Score

On définit le score de la manière suivante :

$$S_{\theta}(X) = \nabla_{\theta} \ln(p_{\theta}(X)) = \left(\frac{\partial \ln p_{\theta}(X)}{\partial \theta_i} \right)_{1 \leq i \leq N}^T$$

C'est un vecteur aléatoire centré, ie $\mathbb{E}_{\theta}(S_{\theta}(X)) = 0$.

Information de Fisher au point θ

On définit l'information de Fisher au point θ par la matrice de covariance du vecteur score. On la note $I(\theta)$.

$$\begin{aligned} I(\theta) &= \text{Var}_{\theta}(S_{\theta}(X)) = \mathbb{E}_{\theta}[S_{\theta}(X)S_{\theta}(X)^T] \text{ car } \mathbb{E}_{\theta}(S_{\theta}(X)) = 0. \\ &= \left(-\mathbb{E}_{\theta} \left(\frac{\partial^2 \ln(p_{\theta}(x))}{\partial \theta_i \partial \theta_j} \right) \right)_{i,j} \text{ sous les conditions de régularité.} \end{aligned}$$

Exemple

On reprend $X \sim B(\alpha)$.

On a $p_\alpha(x) = \alpha^x(1-\alpha)^{1-x}$.

On a alors $\ln(p_\alpha(x)) = x\ln(\alpha) + (1-x)\ln(1-\alpha)$. On a donc :

$$\frac{d\ln(p_\alpha)}{d\alpha}(x) = \frac{x}{\alpha} - \frac{1-x}{1-\alpha} = \frac{x(1-\alpha) - \alpha(1-x)}{\alpha(1-\alpha)} = \frac{x-\alpha}{\alpha(1-\alpha)}$$

On en déduit donc le score $S_\alpha(X) = \left(\frac{X-\alpha}{\alpha(1-\alpha)}\right)$.

On calcule l'information de Fisher. On a

$$I_\alpha = \mathbb{E}_\alpha \left(\frac{(X-\alpha)^2}{(\alpha(1-\alpha))^2} \right)$$

Or, on a, comme $X^2 \sim B(\alpha)$,

$$\mathbb{E}_\alpha[(X-\alpha)^2] = \mathbb{E}_\alpha(X^2) - 2\alpha\mathbb{E}_\alpha(X) + \alpha^2 = \alpha(1-\alpha)$$

On a donc

$$I_\alpha = \frac{1}{\alpha(1-\alpha)}$$

Biais

On définit le **Biais** de l'estimateur T de θ par $b_{\theta}(T) = \mathbb{E}_{\theta}(T) - \theta$ si T intégrable, ie $\mathbb{E}_{\theta}(T) < \infty$.

On dit que T est un **estimateur sans biais**, ssi, $\forall \theta, b_{\theta}(T) = 0$.

Remarque

Le biais, ici, mesure donc à quel point un estimateur est proche en moyenne du paramètre qu'il approche.

Efficacité

Soient T, T' deux estimateurs sans biais de θ pour $\theta \in \Theta$. On dit que T est dit plus efficace que T' si :

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T'), \forall \theta \in \Theta \text{ et } \exists \theta \in \Theta, \mathbb{V}_{\theta_0}(T) < \mathbb{V}_{\theta_0}(T').$$

(On précise que lorsqu'il s'agit d'inégalité de matrices, il s'agit de voir si par exemple $\mathbb{V}_\theta(T') - \mathbb{V}_\theta(T)$ est une matrice symétrique positive, ie à spectre dans \mathbb{R}_+)

On dit que l'estimateur sans biais T est de variance minimale (ou meilleur estimateur sans biais) si

$$\mathbb{V}_\theta(T) \leq \mathbb{V}_\theta(T')$$

pour tout estimateur sans biais T' et pour tout $\theta \in \Theta$.

Borne de Cramer-Rao

$\mathbb{V}_\theta(T) \geq I^{-1}(\theta)$ avec T un estimateur sans biais de $\theta \in \Theta \subset \mathbb{R}^d$ de carré sommable.

Décomposition biais-variance

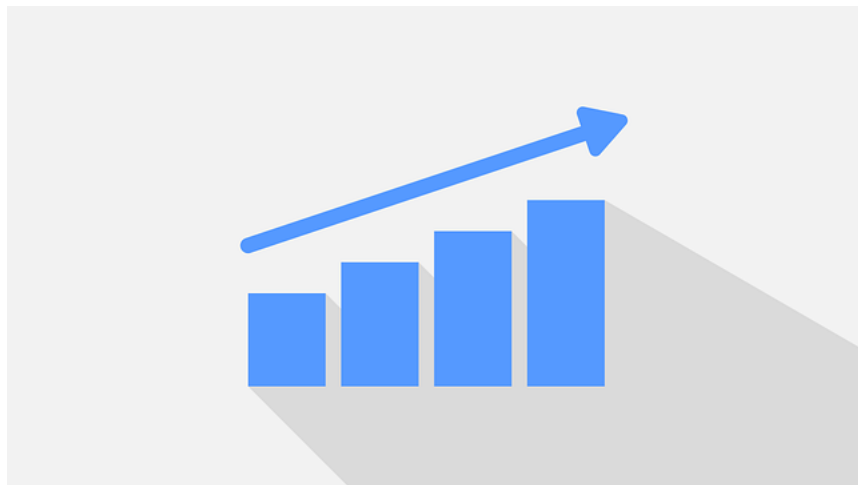
On a d'ailleurs que

$$R(T, \theta) = \|b_\theta(T)\|^2 + \text{Tr}(V_\theta(T))$$

Remarque

On peut donc voir que pour un estimateur sans biais de θ , si sa variance vaut la borne de Cramer-Rao, alors c'est un estimateur qui minimise le risque.

Qu'est-ce qu'il se passe si je veux augmenter N ?



Qu'est-ce qu'il se passe si je veux augmenter N ?

- Généralement, nos mesures peuvent ne pas être suffisantes pour réaliser une bonne estimation à l'aide de notre statistique.
- Toutefois, on aimerait savoir si en augmentant nos mesures, notre statistique pourrait être meilleure pour la description des phénomènes que l'on observe !

Convergence L^2

On dit que X_N converge par la norme L^2 vers X si

$$\mathbb{E}(\|X_N - X\|^2) \xrightarrow{N \rightarrow \infty} 0$$

On note cela $X_N \xrightarrow{L^2} X$

Convergence presque sûre

On dit que X_N converge presque sûrement vers X si

$$\mathbb{P}(\lim_{N \rightarrow \infty} X_N = X) = 1$$

On note cela $X_N \xrightarrow{p.s.} X$

Convergence en Probabilité

On dit que X_N converge en probabilité vers X si

$$\mathbb{P}(\|X_N - X\| > \epsilon) \xrightarrow{N \rightarrow \infty} 0$$

On note cela $X_N \xrightarrow{\mathbb{P}} X$

Convergence en loi

On dit que X_N converge en loi vers X si

$$\mathbb{P}(X_N \leq x) \rightarrow \mathbb{P}(X \leq x)$$

Pour tout point x en lequel la fonction de répartition de X est continue. On note cela $X_N \xrightarrow{\mathcal{L}} X$. On notera toutefois que X_N converge seulement vers une variable ayant la même loi que X , sans y être nécessairement égale !

Loi forte des grands nombres

$$\bar{X} \xrightarrow{p.s} \mathbb{E}(X_1) \text{ où } \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \text{ si les } X_i \text{ iid et } \mathbb{E}(|X_1|) < \infty$$

Théorème Central Limite

$$\sqrt{N} \left(\frac{\bar{X} - \mathbb{E}(X_1)}{\sqrt{\text{Var}(X_1)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1) \text{ si les } X_i \text{ iid et } 0 < \text{Var}(X_1) < \infty$$

Théorèmes importants

Théorème de Slutsky

$Y_n + Z_n \xrightarrow{\mathcal{L}} Y + c$, $Y_n Z_n \xrightarrow{\mathcal{L}} cY$ et plus généralement $f(Y_n, Z_n) \xrightarrow{\mathcal{L}} f(Y, c)$ si $f \in \mathcal{C}^0$, avec $(Y_n)_n$ et $(Z_n)_n$ des suites de VAs \mathbb{R} tel que $Y_n \xrightarrow{\mathcal{L}} Y$, une VA \mathbb{R} et $Z_n \xrightarrow{\mathcal{L}} c \in \mathbb{R}$.

Théorème de continuité

Soit $(Y_n)_n$ des VAs \mathbb{R} et $h \in \mathcal{C}^0$, si $Y_n \rightarrow Y$ (en loi, en probabilité ou presque sûrement), alors $h(Y_n) \rightarrow h(Y)$ (en loi, en probabilité ou presque sûrement).

Remarque

On utilise régulièrement de façon conjointe le Théorème Central Limite, le Théorème de Slutsky et la Loi forte des grands nombres afin de faire disparaître des constantes lors des calculs de convergences.

Caractéristiques asymptotiques

Rappel : $T_N = T(X_1, \dots, X_N)$ où les X_i sont un échantillon iid suivant une loi P_θ .

Petites def

- T_N est **asymptotiquement sans biais**, ssi, $\lim_{N \rightarrow \infty} b_\theta(T_N) = 0, \forall \theta$.
- T_N est **consistant / convergent**, ssi, $\forall \theta, T_N \xrightarrow{\mathbb{P}} \theta$.
- T_N est **fortement consistant / convergent**, ssi, $\forall \theta, T_N \xrightarrow{P.S} \theta$.

Vitesse de convergence

Soit T_N un estimateur de θ consistant. S'il existe une suite réelle (a_N) d'éléments strictement positifs et si Z est une variable aléatoire non constante presque sûrement, telle que $\lim_{N \rightarrow +\infty} a_N = +\infty$ et $a_N(T_N - \theta) \xrightarrow{\mathcal{L}} Z$, on appelle alors (a_N) la **vitesse de convergence** de l'estimateur (T_N) .

D'autres définitions

- Un estimateur T_N de θ est dit **asymptotiquement normal** s'il existe $\Sigma(\theta)$ telle que :

$$\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma(\theta))$$

La matrice de covariance Σ ne dépend pas de N . On l'appelle **variance asymptotique associée à l'estimateur normal T_N** .

- Si T_N et $T_{N'}$ sont deux estimateurs de θ asymptotiquement normaux, de variances asymptotiques associées Σ et Σ' , alors T_N est **asymptotiquement plus efficace** que $T_{N'}$ si on a que pour tout θ , $\Sigma'(\theta) - \Sigma(\theta)$ est une matrice symétrique positive, et qu'il existe θ' tel que $\Sigma'(\theta') - \Sigma(\theta')$ est symétrique définie positive.
- Un estimateur est dit **asymptotiquement efficace** s'il est asymptotiquement normal et que sa matrice de variance asymptotiquement associée, $\Sigma(\theta) = I^{-1}(\theta)$.

Pourquoi le maximum de vraisemblance c'est génial

Soit (X_1, \dots, X_N) iid pour p_θ . Sous des hypothèses de régularité vérifiées dans le cadre du cours, l'estimateur du maximum de vraisemblance $\hat{\theta}^{MV}$ est :

- Convergent
- Asymptotiquement Normal
- Asymptotiquement Efficace

On a donc, finalement, que :

$$\sqrt{N}(\hat{\theta}^{MV} - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, I^{-1}(\theta))$$

Mini exo d'illustration

On reprend X suivant une loi de Bernoulli de paramètre α . On avait $\hat{\alpha}^{MV} = \bar{X}_N$.

On a $E(|X|) < \infty$. Par Loi Forte des Grands Nombres, $\bar{X}_N \xrightarrow{ps} \mathbb{E}[X] = \alpha$.

On a donc que $\hat{\alpha}^{MV}$ est fortement convergent, donc convergent (la convergence presque sûre implique la convergence en probabilité).

On a $\mathbb{V}(X) < \infty$. Donc, par théorème central limite,

$$\sqrt{N} \left(\frac{\bar{X}_N - \mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

On a alors, comme $\mathbb{V}(X) = \alpha(1 - \alpha)$, que

$$\sqrt{N} \left(\frac{\bar{X}_N - \alpha}{\sqrt{\alpha(1 - \alpha)}} \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

On a donc que $\sqrt{N}(\bar{X}_N - \alpha) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \alpha(1 - \alpha))$. En reprenant ce qu'on avait trouvé précédemment, on a que notre estimateur est asymptotiquement efficace.

Maintenant, où est-ce qu'on range θ ?



Maintenant, où est-ce qu'on range θ ?

- On a réussi à trouver un estimateur sympa de θ .
- Cependant, on n'a pas vraiment d'idée de la valeur de θ . On sait que notre estimateur converge vers θ en faisant tendre N vers l'infini, mais en pratique, l'infini, on n'y a pas accès.
- Du coup, comment estimer θ à un certain niveau de précision ?

Région et intervalle de confiance

Soit un niveau de risque $\alpha \in]0, 1[$ fixé.

Une **région de confiance** $I(\mathbf{X})$ de niveau $1 - \alpha$ est une statistique sur l'échantillon X à valeurs dans $P(\Theta)$ (les parties de Θ) tel que

$$\mathbb{P}(\theta^* \in I(\mathbf{X})) \geq 1 - \alpha.$$

Si $\mathbb{P}(\theta^* \in I(\mathbf{X})) = 1 - \alpha$, on dit que la région est de **taille** $1 - \alpha$.

Si $\Theta \subset \mathbb{R}$, alors $I(\mathbf{X}) \in \mathbb{R}$ et on parle donc d'**intervalles de confiance**.

On définit le **taux de couverture** par $\tau_C = \mathbb{P}(\theta^* \in I(\mathbf{X}))$.

Fonction Pivotal

Une **fonction pivotale** est une statistique $S(X; \theta)$ dépendant de θ , mais dont la loi de probabilité ne dépend pas de θ . On parle de **loi de probabilité libre**.

Lois utiles

- Si $(Z_i)_i \sim \mathcal{N}(0, 1)$ indé., alors $\sum_{i=1}^N Z_i^2 \sim \chi^2(N)$ (Loi du Chi 2).
- Si $U \sim \mathcal{N}(0, 1)$ et $Z \sim \chi^2(\nu)$ indé., alors $\frac{U}{\sqrt{\frac{Z}{\nu}}} \sim T(\nu)$ (Loi de Student).

Fonction Pivotale

Une **fonction pivotale** est une statistique $S(X; \theta)$ dépendant de θ , mais dont la loi de probabilité ne dépend pas de θ . On parle de **loi de probabilité libre**.

Exemples de fonction pivotale

- Si $X \sim \mathcal{N}(\mu, 1)$, alors $S(X; \mu) = X - \mu \sim \mathcal{N}(0, 1)$ est une fonction pivotale pour μ .
- Si X_1, \dots, X_N iid suivant une loi $\mathcal{N}(\mu, \sigma^2)$, en supposant μ, σ^2 inconnus et en notant $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, $S = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$ et en admettant que $\sqrt{N-1} \frac{\bar{X} - \mu}{S} \sim Student(N-1)$ (théorème de Cochran), on a que $F(X_1, \dots, X_N; \mu) = \sqrt{N-1} \frac{\bar{X} - \mu}{S}$ est une fonction pivotale pour μ .

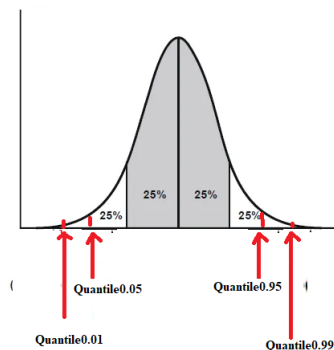
Quantile

Pour F une fonction de répartition, on pose $\forall r \in]0, 1[$,

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

Si $F \in \mathcal{C}^0$, alors $F(q_r) = r$.

Si F est strictement croissante, alors $q_r = F^{-1}(r)$.



Quantile

Pour F une fonction de répartition, on pose $\forall r \in]0, 1[$,

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

Si $F \in \mathcal{C}^0$, alors $F(q_r) = r$.

Si F est strictement croissante, alors $q_r = F^{-1}(r)$.

Intervalle de confiance d'ordre r

Soient $\alpha \in]0, 1[$ fixé et $S(X; \theta)$ une fonction pivotale pour $\theta \in \Theta$.

Si sa fonction de répartition F est continue, alors $\forall \gamma \in [0, \alpha]$,

$$I_\alpha^\gamma(X) = S^{-1}(X; [q_\gamma, q_{\gamma+1-\alpha}]) = \{\theta \in \Theta \mid q_\gamma \leq S(X; \theta) \leq q_{\gamma+1-\alpha}\}$$

est un intervalle de confiance de taille $1 - \alpha$.

Remarque importante

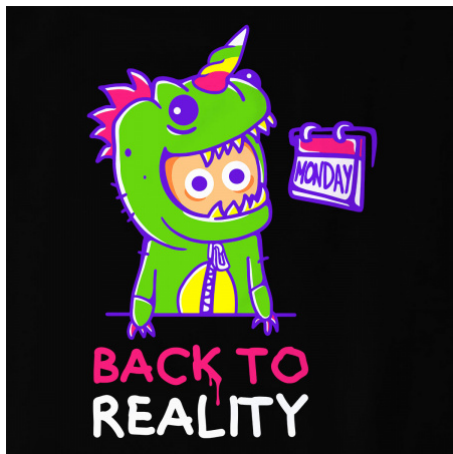
En pratique, pour $\alpha < \frac{1}{2}$, et pour F^{-1} convexe sur $[0, \frac{1}{2}]$, alors l'intervalle le plus petit est obtenu pour $\gamma = \frac{\alpha}{2}$. C'est le cas par exemple pour les lois Gaussiennes ou de Student.

Cas Loi Symétrique Centrée (par exemple $\mathcal{N}(0, 1)$)

Pour le cas où on étudie une variable suivant une loi normale centrée $\mathcal{N}(0, \sigma^2)$ ou une loi de Student, on a en prenant $\gamma = \frac{\alpha}{2}$, par symétrie de la loi, que $q_{\frac{\alpha}{2}} = -q_{1-\frac{\alpha}{2}}$.

On a alors pour α dans $]0, 1[$ fixé, que

$$I_{\alpha}^{\frac{\alpha}{2}}(X) = \{\theta \in \Theta \mid -q_{1-\frac{\alpha}{2}} \leq S(X; \theta) \leq q_{1-\frac{\alpha}{2}}\}$$



- En fait, on a un souci : souvent, nos statistiques ne nous permettent pas directement d'obtenir un bel intervalle de confiance.
- Mais, on a des théorèmes sympas sur nos statistiques en faisant tendre notre N (ie le nombre de mesures) vers l'infini !
- Donc on va étudier des **intervalles de confiance asymptotiques** !

Région et intervalle de confiance asymptotique

Une **région de confiance asymptotique** $I(\mathbf{X})$ de niveau $1 - \alpha$ est une statistique sur l'échantillon X à valeurs dans $P(\Theta)$ (les parties de Θ) tel que $\lim_{N \rightarrow \infty} \mathbb{P}(\theta^* \in I(\mathbf{X})) \geq 1 - \alpha$.

Si $\lim_{N \rightarrow \infty} \mathbb{P}(\theta^* \in I(\mathbf{X})) = 1 - \alpha$, la région de confiance asymptotique est dite de taille $1 - \alpha$.

Si $\Theta \subset \mathbb{R}$, on choisit la région de confiance dans l'ensemble des intervalles sur \mathbb{R} . On parle alors **d'intervalle de confiance asymptotique**.

Fonction asymptotiquement pivotale

Une **fonction asymptotiquement pivotale** est une statistique $S_N(\theta) = S(X_1, \dots, X_N; \theta)$ dépendant de θ mais dont la loi limite ne dépend pas de θ : ie, il existe Y une VA de loi connue et libre telle que $S_N(\theta) \xrightarrow{\mathcal{L}} Y$.

Proposition importante intervalle asymptotique

Construction d'intervalles de confiance asymptotiques

Soit $\alpha \in]0, 1[$ fixé. Soit $S_N(\theta)$ une fonction asymptotiquement pivotale pour $\theta \in \Theta$, $S_N \xrightarrow{\mathcal{L}} Y$.

Si la fonction de répartition F de Y est continue, alors $\forall \gamma, 0 \leq \gamma \leq \alpha$:

$$I_\alpha^\gamma(X_1, \dots, X_N) = S_N^{-1}([q_\gamma, q_{\gamma+1-\alpha}]) = \{\theta \in \Theta \mid q_\gamma \leq S_N(\theta) \leq q_{\gamma+1-\alpha}\}$$

est un intervalle de confiance asymptotique de taille $(1-\alpha)$ pour θ . (PS : les q_x sont des quantiles par rapport à la loi de Y !)

Remarque

Comme pour ce que l'on a dit précédemment, si Y suit une loi normale ou une loi de student, on prendra $\gamma = \frac{\alpha}{2}$. De plus, si Y suit une loi normale centrée ou de Student, on pourra utiliser $q_{\frac{\alpha}{2}} = -q_{1-\frac{\alpha}{2}}$, ce qui facilitera la construction de l'intervalle.

Théorème Important

Soit $T_N = T(X_1, \dots, X_N)$ un estimateur de θ , asymptotiquement normal, soit vérifiant $\sqrt{N}(T_N - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(\theta))$, avec $V(\theta) > 0$ et V continue en θ . On a que :

$$\frac{\sqrt{N}(T_N - \theta)}{\sqrt{V(\theta)}} \text{ et } \frac{\sqrt{N}(T_N - \theta)}{\sqrt{V(T_N)}}$$

Sont des fonctions asymptotiquement pivotales pour θ , tendant vers une variable aléatoire Z de loi $\mathcal{N}(0, 1)$.

Méthode Plug-In

Si $T_N \xrightarrow{\mathbb{P}} \theta$ et $S_N(\theta)$, une fonction asymptotiquement stable pour θ telle que $S_N(\theta) = U_N(\theta)\phi(\theta)$ avec $U_N(\theta)$ une statistique (plus simple) de l'échantillon et ϕ une fonction continue non nulle sur Θ , alors $U_N(\theta)\phi(T_N)$ est aussi asymptotiquement pivotale.

En pratique, si l'on regarde $S(X_1, \dots, X_N; \theta)$ une statistique asymptotiquement pivotale, dans les exercices que l'on fait, la taille de l'échantillon est fixe. L'intervalle de confiance asymptotique doit être donc vu comme une **approximation** de notre estimation, en supposant que les N à partir desquels on regarde sont assez grands.

Exemple

On prend X_1, \dots, X_N suivant une loi de Bernoulli de paramètre θ . On a alors $\mathbb{E}[X_i] = \theta$ et $\mathbb{V}[X_i] = \theta(1 - \theta)$. On a alors par Théorème Central Limite que

$$\sqrt{N} \frac{\bar{X}_N - \theta}{\sqrt{\theta(1 - \theta)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

On a alors que $S_N(X; \theta) = \sqrt{N} \frac{\bar{X}_N - \theta}{\sqrt{\theta(1 - \theta)}}$ est asymptotiquement pivotale.

On a donc que

$$\lim_{N \rightarrow \infty} \mathbb{P}(q_{\frac{\alpha}{2}} \leq S_N(X; \theta) \leq q_{1 - \frac{\alpha}{2}}) = 1 - \alpha$$

En utilisant $q_{\frac{\alpha}{2}} = -q_{1 - \frac{\alpha}{2}} = 1 - \alpha$, on arrive à simplifier l'intervalle à étudier, mais pour entourer θ avec S_N , ça va pas être rigolo.

Suite de l'Exemple

Rappel : On prend X_1, \dots, X_N des variables i.i.d. suivant une loi de Bernoulli de paramètre θ . On a que $S_N(X; \theta) = \sqrt{N} \frac{\bar{X}_N - \theta}{\sqrt{\theta(1-\theta)}}$ est asymptotiquement pivotale. On cherche une variable aléatoire nous permettant d'encadrer θ sans trop se prendre la tête.

Or, on a que $\bar{X}_N \xrightarrow{\mathbb{P}} \theta$ par Loi Forte des Grands Nombres. On utilise alors la méthode Plug-In ! On a donc que

$$\sqrt{N} \frac{\bar{X}_N - \theta}{\sqrt{\bar{X}_N(1 - \bar{X}_N)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

Cela nous donne donc un intervalle **beaucoup** plus simple à étudier, puis qu'il n'y a qu'un composant dépendant de θ . En remuant bien, on trouve que l'on a comme intervalle de confiance asymptotique

$$\left[\bar{X}_N - q_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{X}_N(1 - \bar{X}_N)}}{\sqrt{N}}, \bar{X}_N + q_{1-\frac{\alpha}{2}} \frac{\sqrt{\bar{X}_N(1 - \bar{X}_N)}}{\sqrt{N}} \right]$$

- Changement de paradigme : dans l'estimation paramétrique, on étudiait θ comme un paramètre. On étudiait alors des résultats en fréquence, et on tentait d'estimer θ à partir de nos données.
- Dans cette partie-là de l'amphi, on étudiera θ comme une variable aléatoire dans un ensemble $\Theta \subset \mathbb{R}^p$. On note cette variable aléatoire ϑ et θ sa réalisation. On étudiera alors la plausibilité que notre paramètre soit θ en fonction des données.
- Toutefois, on se donnera une loi de probabilité traduisant la connaissance que l'on a a priori sur la variable aléatoire, que l'on appelle le **prior**.

Modèle stat bayésien

Un modèle statistique bayésien est un ensemble de familles de lois de probabilité M_Θ muni d'une loi de probabilité π_ϑ sur $(\Theta, \mathcal{B}(\Theta))$, appelée **distribution a priori** ou **prior**.

But

Etant donné un échantillon iid $X = (X_1, \dots, X_N)$, le principe de l'estimation bayésienne sera la détermination de la loi conditionnelle dite *a posteriori* $\mathbb{P}(\vartheta|X)$.

Détermination de la loi a posteriori

Pour ce faire, on utilise la formule de Bayes, fonctionnant pour tout $x \in \mathcal{X}$ tel que $\rho_X(x) \neq 0$. En effet, on a que dans ce cas, pour tout θ dans Θ

$$\rho(\theta | x) = \frac{\rho_{(\vartheta, X)}(\theta, x)}{\rho_X(x)} = \frac{\rho(x | \theta)\pi(\theta)}{\rho_X(x)} \text{ est définie.}$$

Remarques

- On reconnaît $\rho(x | \theta)$: c'est la vraisemblance que nous avons étudiée précédemment !
- $\rho_X(x) = \int_{\Theta} \rho(x | \theta)\pi(\theta)d\theta$ est souvent difficile à calculer, mais elle ne joue que le rôle d'une constante de normalisation. On note donc souvent $\rho(\theta | x) \propto \rho(x | \theta)\pi(\theta)$. Il n'est souvent pas nécessaire de la calculer avec les méthodes qui vont suivre.

Comment obtenir $\mathbb{P}(\theta | x)$?

Obtention de la loi

Il y a 4 étapes principales à connaître, pour obtenir cette loi de probabilité:

- 1 On commence par calculer la vraisemblance $\mathcal{L}(\theta; x) = \rho(x | \theta)$.
- 2 Une fois celle-ci calculée, on applique la formule de Bayes.
- 3 Ensuite, on reconnaît la densité d'une loi connue, ou l'on réalise une simulation numérique d'un échantillon de la densité (second cas peu probable dans le cadre du cours).
- 4 Enfin, on déduit la loi a posteriori à l'aide de ladite densité.

Lois utiles à connaître

Il est donc utile de connaître les densités (ou fonctions de masses si l'on travaille avec des lois discrètes) de lois usuelles. On assume ici (et PHC assumera !) que les densités/fdm des lois de Bernoulli, Binomiale, de Poisson, Uniforme, Exponentielle et Gaussienne sont connues ! Parmi les lois peu courantes, on retrouvera ici :

- La loi $\beta(a, b)$ de densité

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1} \mathbf{1}_{[0,1]}(x)$$

où Γ désigne la fonction Γ d'Euler.

- La loi $\gamma(\alpha, \beta)$ de densité

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{\mathbb{R}_+}(x)$$

Exemple

Soit (X_1, \dots, X_N) échantillon iid pour Bernoulli(θ), avec $\theta \in [0, 1]$ inconnu. On prend comme distribution a priori la distribution uniforme sur $[0, 1]$. On a alors $\pi(\theta) = 1_{[0,1]}(\theta)$. Cela nous donne un modèle Bayésien. Soit $x = (x_1, \dots, x_N)$ un échantillon d'observation. On cherche donc à déterminer $p(\theta | x)$, en utilisant la formule $p(\theta | x) \propto p(x | \theta)\pi(\theta)$.

Etape 1 : Calcul de la vraisemblance

On a $p(x_i | \theta) = \theta^{x_i}(1 - \theta)^{(1-x_i)}$, et donc

$$p(x | \theta) = \prod_{i=1}^N p(x_i | \theta) = \theta^{\sum_{i=1}^N x_i} (1 - \theta)^{N - \sum_{i=1}^N x_i}$$

On note S la statistique $S(X_1, \dots, X_N) = \sum_{i=1}^N X_i$ et $S(x)$ sa réalisation en x . On a alors

$$p(x | \theta) = \theta^{S(x)}(1 - \theta)^{N - S(x)}$$

Exemple : Suite

On a alors $p(x | \theta) = \theta^{S(x)}(1 - \theta)^{N - S(x)}$. De plus, comme $\theta \in [0, 1]$, on a $\pi(\theta) = 1$.

Etape 2 : Formule de Bayes

On a donc $p(\theta | x) \propto p(x | \theta)\pi(\theta) = \theta^{S(x)}(1 - \theta)^{N - S(x)}$

Etape 3 : Reconnaissance d'une densité

On vient de voir que la densité $f(\theta)$ d'une loi $\beta(a, b)$ est donnée proportionnelle à $\theta^{a-1}(1 - \theta)^{b-1}$. On a donc que $p(\theta | x)$ est proportionnelle à la densité d'une loi $\beta(S(x) + 1, N - S(x) + 1)$. Or, deux densités proportionnelles sont égales par la normalisation...

Etape 4 : Dédution du posterior

On a donc que la variable aléatoire $[\vartheta | X = x]$ suit une loi $\beta(S(x) + 1, N - S(x) + 1)$.

Comment choisir le prior ?

Il y a différentes manières, plus détaillées dans le cours, pour faire le choix du $\pi(\theta)$, qui conditionne pas mal le résultat sur $p(\theta | x)$.

- **Cas idéal** Connaissance réelle du phénomène étudié, on peut déterminer une distribution a priori.
- **Les priors conjugués** Des priors tels que les lois a posteriori et les priors soient dans la même famille de lois de probabilité.
- **Priors non-informatifs** Loi uniforme, prior de Jeffrey, prior impropre, ...
Ils traduisent notre manque de connaissance du processus.

Région de crédibilité

Pour un échantillon d'observations donné x , une **région de crédibilité** $C_\alpha \subset \Theta$ au niveau $1 - \alpha$, $\alpha \in]0, 1[$, vérifie

$$\mathbb{P}(\theta \in C_\alpha | X = x) = \int_{C_\alpha} \rho(\theta | x) d\theta \geq 1 - \alpha$$

En cas d'égalité, on dit de taille $1 - \alpha$.

Intervalle de crédibilité

Si la fonction de répartition de $(\vartheta | X = x)$ est continue, alors $\forall \gamma \in [0, \alpha]$, $C_\alpha^\gamma = [q_\gamma, q_{\gamma+1-\alpha}]$ est un intervalle de crédibilité de taille $1 - \alpha$.

Vu que $\rho(\theta | x)$ est connue, la détermination des régions de crédibilité est directe.

Contrairement au région de confiance, ici x est fixé et on étudie la distribution $(\vartheta | X = x)$.

Introduction

Maintenant que l'on connaît la distribution ($\vartheta | X = x$), on aimerait pouvoir calculer une estimation ponctuelle $\hat{\theta}$ pour prendre des décisions. Soit θ le vrai paramètre, qu'on ne connaît pas, et $\forall \delta \in \Theta$, $L(\theta, \delta)$ une **fonction de perte**, qui représente le coût de se tromper.

Alors on choisit :

$$\hat{\theta} = \underset{\delta \in \Theta}{\operatorname{argmin}} [\mathbb{E}_{(\rho(\theta|x))}(L(\theta, \delta))] = \int_{\Theta} L(\theta, \delta) \rho(\theta | x) d\theta$$

Processus Test and Learn



Introduction

- On sort du cadre actuel des estimations pour s'intéresser à la validation ou invalidation d'hypothèses sur nos paramètres.
- On s'intéresse à des tests binaires, ici : on a une hypothèse H_0 que l'on confronte à une hypothèse H_1 , et l'on aimerait savoir laquelle des deux est vraie (les deux ne pouvant pas l'être, et on suppose que si H_0 est fausse, alors H_1 est vraie et vice-versa).

Comment vérifier la validité d'une hypothèse ?

- Pour savoir laquelle des hypothèses est vraie, on va d'abord supposer H_0 .
- On va ensuite chercher ce que l'on appelle une zone de rejet. Il s'agit d'un ensemble \mathcal{R} où la probabilité d'avoir notre échantillon dedans est très faible, dans la mesure où H_0 est vraie.
- Si l'on trouve que notre échantillon est dans \mathcal{R} , on rejette H_0 et l'on accepte H_1 .

Définitions

Soit un modèle statistique M sur $(\mathcal{X}, \mathcal{A})$ et $P^* \in M$, une loi de probabilité inconnue.

- On définit une **Hypothèse statistique sur P^*** par une affirmation du type $P^* \in M_0 \subset M$.

M_0 peut être un modèle paramétrique, un singleton, etc.

- Soit $M_0 \subset M$ et $M_1 \subset M$ tel que $M_0 \cap M_1 = \emptyset$.
On définit un **Test d'hypothèses statistiques** par une procédure confrontant deux hypothèses statistiques :

- $H_0 : P^* \in M_0$ (Hypothèse nulle / conservative)
- $H_1 : P^* \in M_1$ (Hypothèse alternative)

et définissant une **règle de décision** permettant d'accepter ou rejeter H_0 à partir de l'observation d'un N-échantillon i.i.d pour P^* .

La région de rejet

La règle de décision permettant l'acceptation ou le rejet de H_0 est ainsi définie par une **région de rejet ou région critique** $\mathcal{R} \subset \chi^N$. Soit l'échantillon $(x_1, \dots, x_N) \in \chi^N$. Si $(x_1, \dots, x_N) \in \mathcal{R}^c$, alors on accepte H_0 . Si $(x_1, \dots, x_N) \in \mathcal{R}$, alors on rejette H_0 .

Concrètement, un test sera défini par un triplet (M_0, M_1, R) où (M_0, M_1) sont deux familles de loi de probabilité disjointes, et R une région de rejet associée.

Construction de la région de rejet

On construit la région de rejet à l'aide d'une **statistique de test** T : $\chi^N \rightarrow \mathbb{R}$ dont on connaît la loi.

La région de rejet sera alors de la forme :

$$\mathcal{R} = \{(x_1, \dots, x_N) \in \chi^N \mid T(x_1, \dots, x_N) \in W\}, \text{ avec } W \subset \mathbb{R}.$$

Définitions

Pour mesurer la pertinence du test (ie si \mathcal{R} est une région de rejet pertinente), on introduit plusieurs notions :

- Le **risque de première espèce** ou probabilité de rejeter H_0 alors qu'elle est vraie.
- Le **risque de seconde espèce** ou probabilité d'accepter H_0 alors que H_1 est vraie.
- On définit alors la **puissance du test**, soit "1 - risque de deuxième espèce", qui correspond à la probabilité de rejeter H_0 alors que H_1 est vraie.

En termes mathématiques

- Le **risque de première espèce** ou probabilité de rejeter H_0 alors qu'elle est vraie correspond à calculer $\mathbb{P}((X_1, \dots, X_N) \in \mathcal{R})$ dans le cas où la loi P^* que l'on cherche est bien dans M_0 .
- Le **risque de seconde espèce** ou probabilité d'accepter H_0 alors que H_1 est vraie correspond à calculer $\mathbb{P}((X_1, \dots, X_N) \notin \mathcal{R})$ dans le cas où la loi P^* que l'on cherche est bien dans M_1 .
- La puissance du test, qui correspond à la probabilité de rejeter H_0 alors que H_1 est vraie, correspond juste alors à calculer $1 - \mathbb{P}((X_1, \dots, X_N) \notin \mathcal{R})$ en supposant que H_1 est vraie.

Autre Définition

On définit la **taille du test**, pour un test (M_0, M_1, \mathcal{R}) par $\alpha = \sup_{P \in M_0} P(\mathcal{R})$. Tout majorant de la taille du test sera dit un **niveau** de test.

On va alors chercher la plus petite taille d'un test qui conduirait à rejeter H_0 . C'est ce qu'on appelle la **p-valeur**.

La p-valeur

Soit $\{(M_0, M_1, \mathcal{R}_t), t \in \mathcal{T}\}$, une collection de tests indicés par $t \in \mathcal{T} \subset \mathbb{R}$, et α_t la taille du test de région de rejet \mathcal{R}_t . La p-valeur α^* est la statistique définie sur χ^N (ie l'ensemble des mesures) par

$$\alpha^*(X_1, \dots, X_N) = \inf \{ \alpha_t, t \in \mathcal{T} \mid (X_1, \dots, X_N) \in \mathcal{R}_t \}$$

En pratique, on en tire les conclusions suivantes :

- $p \leq 0.01$: Très forte présomption contre l'hypothèse H_0 .
- $0.01 \leq p \leq 0.05$: Forte présomption contre l'hypothèse H_0 .
- $0.05 \leq p \leq 0.1$: Faible présomption contre l'hypothèse nulle.
- $0.1 < p$: Pas de présomption contre l'hypothèse H_0 .

Test paramétrique

Maintenant qu'on a introduit tous ces concepts, on revient au coeur du problème, l'étude de modèles paramétriques.

Définition

On prend un modèle statistique paramétrique $M_\Theta = \{P_\theta \mid \theta \in \Theta\}$. Soit $P_\theta \in M_\Theta$ la loi inconnue. On prend Θ_0 et Θ_1 des sous-ensembles disjoints de Θ . Un **test paramétrique** est un test de la forme :

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$$

Sous-définitions

Le **test paramétrique d'hypothèses simples** prend Θ_0 et Θ_1 comme singletons : ie il confronte $H_0 : \theta = \theta_0$ à $H_1 : \theta = \theta_1$. Tout autre test paramétrique est dit **composite**. On notera toutefois deux tests particuliers : $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$ dit test bilatère et $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$ (ou $<$) dit test unilatère.

Construction d'une zone de rejet

Si l'on a un test d'hypothèses $H_0 : \theta \in \Theta_0$ vs $H_1 : \theta \in \Theta_1$, et α un niveau de risque, comment construire une zone de rejet de taille inférieure à α ?

Méthode

Avec notre échantillon iid (X_1, \dots, X_N) pour P_θ , on trouve une statistique T telle que la loi de $T(X_1, \dots, X_N)$ soit connue.

On peut alors déterminer un ensemble W_α tel que

$\sup_{\theta \in \Theta_0} P_\theta(T(X_1, \dots, X_N) \in W_\alpha) \leq \alpha$. La région de rejet est alors donnée par

$$R_\alpha = \{(x_1, \dots, x_N) \in \mathcal{X}^N \mid T(x_1, \dots, x_N) \in W_\alpha\}$$

Construction d'une zone de rejet

Méthode

Avec notre échantillon iid (X_1, \dots, X_N) pour P_θ , on trouve une statistique T telle que la loi de $T(X_1, \dots, X_N)$ soit connue.

On peut alors déterminer un ensemble W_α tel que $\sup_{\theta \in \Theta_0} P_\theta(T(X_1, \dots, X_N) \in W_\alpha) \leq \alpha$. La région de rejet est alors donnée par $R_\alpha = \{(x_1, \dots, x_N) \in \mathcal{X}^N \mid T(x_1, \dots, x_N) \in W_\alpha\}$

Remarques

C'est là que notre travail précédent sur les intervalles de confiance va être utile ! En effet, on pourra, si notre statistique suit une loi connue, trouver un intervalle de confiance de taille $1 - \alpha$ à l'aide des méthodes vues précédemment, ce qui nous permettra en prenant le complémentaire de définir une région de rejet.

De plus, on utilisera aussi plus souvent des intervalles de confiances asymptotiques de taille $1 - \alpha$, puisque c'est asymptotiquement qu'on trouve plus souvent nos propriétés.

Exemple

Soit X suivant une loi $\mathcal{N}(\mu, \sigma^2)$. On suppose σ inconnu. Le test d'hypothèse est le suivant : $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$. On a réalisé N échantillons.

On note $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, $S = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$.

On admet que $\sqrt{N-1} \frac{\bar{X} - \mu}{S}$ suit une loi Student($N-1$). On a alors sous H_0 ,

$$T(X_1, \dots, X_N) = \sqrt{N-1} \frac{\bar{X} - \mu_0}{S} \sim \text{Student}(N-1)$$

. On a donc que, pour α un niveau de risque fixé :

$$P_{\mu_0} \left(q_{\frac{\alpha}{2}}^{\text{Student}(N-1)} \leq T(X_1, \dots, X_N) \leq q_{1-\frac{\alpha}{2}}^{\text{Student}(N-1)} \right) = 1 - \alpha$$

En prenant le complémentaire de cette cet intervalle, on a alors

$$R_\alpha = \left\{ (x_1, \dots, x_N) / \sqrt{N-1} \frac{\bar{x} - \mu_0}{s} < q_{\frac{\alpha}{2}}^{\text{Student}(N-1)} \right. \\ \left. \text{ou } \sqrt{N-1} \frac{\bar{x} - \mu_0}{s} > q_{1-\frac{\alpha}{2}}^{\text{Student}(N-1)} \right\}$$

En prenant le complémentaire de cette cet intervalle, on a alors

$$R_\alpha = \left\{ (x_1, \dots, x_N) / \sqrt{N-1} \frac{\bar{x} - \mu_0}{s} < q_{\frac{\alpha}{2}}^{Student(N-1)} \right. \\ \left. \text{ou } \sqrt{N-1} \frac{\bar{x} - \mu_0}{s} > q_{1-\frac{\alpha}{2}}^{Student(N-1)} \right\}$$

Et donc, si l'on dispose de données numériques sur les x_i , N , et μ_0 , on pourra alors réaliser le test d'hypothèses en s'aidant d'une table de quantiles de la loi de Student **qui sera fournie**.

Définitions

Pour un test paramétrique $(\Theta_0, \Theta_1, \mathcal{R})$, on définit la fonction puissance du test, $\forall \theta \in \Theta_0 \cup \Theta_1 : \pi_\theta = P_\theta((X_1, \dots, X_N) \in \mathcal{R})$

- La taille du test est donnée alors par $\alpha = \sup_{\theta \in \Theta_0} \pi(\theta)$.
- Pour un test d'hypothèses simples ie $\Theta_0 = \{\theta_0\}$, et $\Theta_1 = \{\theta_1\}$, on a :
 - $\alpha = \pi(\theta_0)$ la taille du test et risque de première espèce.
 - $\pi(\theta_1)$ la puissance du test, et $\beta = 1 - \pi(\theta_1)$ le risque de deuxième espèce.

Comparaison de tests

Soient $T = (\Theta_0, \Theta_1, \mathcal{R})$ et $T' = (\Theta_0, \Theta_1, \mathcal{R}')$ deux tests paramétriques de fonctions puissances respectives π et π' .

- Si $\Theta_1 = \theta_1$, T est **plus puissant** que T' si $\pi(\theta_1) > \pi'(\theta_1)$.
- Si Θ_1 est quelconque, on dit que T est **uniformément plus puissant** que T' si $\forall \theta \in \Theta_1, \pi(\theta) > \pi'(\theta)$.

Test du rapport de vraisemblance

On traite d'un test d'hypothèses simples sur un N-échantillon i.i.d (X_1, \dots, X_N) pour une loi de densité ρ_θ , soit $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

Test du rapport de vraisemblance

On définit le **Test du rapport de vraisemblance de niveau α** comme un test utilisant la statistique

$$\lambda(X_1, \dots, X_N) = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_N)}{\mathcal{L}(\theta_0; X_1, \dots, X_N)}$$

et la région critique $R_\alpha^* = \{(x_1, \dots, x_N) \in \mathcal{X}^N \mid \lambda(x_1, \dots, x_N) > c_\alpha\}$, où c_α tel que $\mathbb{P}_{\theta_0}((X_1, \dots, X_N) \in R_\alpha^*) \leq \alpha$.

On notera que ce test est le plus puissant parmi les tests de niveau α .

Test du rapport de vraisemblance

On définit le **Test du rapport de vraisemblance de niveau α** comme un test utilisant la statistique

$$\lambda(X_1, \dots, X_N) = \frac{\mathcal{L}(\theta_1; X_1, \dots, X_N)}{\mathcal{L}(\theta_0; X_1, \dots, X_N)}$$

et la région critique $R_\alpha^* = \{(x_1, \dots, x_N) \in \mathcal{X}^N \mid \lambda(x_1, \dots, x_N) > c_\alpha\}$, où c_α tel que $\mathbb{P}_{\theta_0}((X_1, \dots, X_N) \in R_\alpha^*) \leq \alpha$.

On notera que ce test est le plus puissant parmi les tests de niveau α (lemme de Neyman-Pearson).

Méthode pour déterminer la zone de rejet à l'aide de cette statistique

- On calcule le rapport des vraisemblances.
- Ensuite, on étudie l'inégalité $\lambda(x_1, \dots, x_n) > c_\alpha$ jusqu'à simplifier au plus l'expression de la région de rejet.

Test du χ^2 cas multinomial

La loi multinomiale

On appelle loi multinomiale $\mathcal{M}(N, p)$, pour $N \in \mathbb{N}$, $p \in]0, 1[^k$, $\sum_i p_i = 1$, la loi de probabilité sur $[[0, n]]^k$ définie par la fonction de masse (avec $x_1, \dots, x_k \in [[0, n]]^k$ de somme égale à n)

$$\mathbb{P}(x_1, \dots, x_k) = \frac{n!}{\prod_{i=1}^k x_i!} \prod_{i=1}^k p_i^{x_i}$$

Statistique de Pearson

Soit $X \sim \mathcal{M}(N, p)$, $p \in]0, 1[^k$ \ $\sum_i p_i = 1$.

On définit la Statistique de Pearson par

$$T(X) = \sum_{i=1}^k \frac{(X_i - Np_i)^2}{Np_i} \xrightarrow{\mathcal{L}} \chi^2(k-1)$$

Celle-ci est une mesure relative de l'écart entre les effectifs réalisés et ceux théoriques.

Test du χ^2 cas multinomial

Test de Pearson

Le **Test de Pearson du χ^2 pour le modèle multinomial** est le suivant : on prend $X \sim \mathcal{M}(N, p)$, où p inconnu.

Soit $p_0 \in]0, 1[^k \setminus \sum_i p_{0,i} = 1$, et le test $H_0 : p = p_0$ et $H_1 : p \neq p_0$.

Alors sous H_0 , $T(X) = \sum_{i=1}^k \frac{(X_i - Np_i)^2}{Np_i} \xrightarrow{\mathcal{L}} \chi^2(k-1)$, on en déduit donc la

région asymptotique de rejet : $R_\alpha = \{x \setminus T(x) > q_{1-\alpha}^{\chi^2}\}$.

Tests d'adéquation ou d'ajustement

Soit (X_1, \dots, X_N) un échantillon iid de loi P^* inconnue. Le but de ce type de test est de vérifier si P^* appartient à une famille paramétrique particulière $M_\Theta = \{p_\theta \mid \theta \in \Theta\}$.

De manière plus précise, ici, on réalise une phase préliminaire d'estimation. Avec un échantillon donné (x_1, \dots, x_N) , on réalise le test suivant, sur notre famille M_Θ : $H_0 : P^* = P_{\theta_0}$ vs $H_1 : P^* \neq P_{\theta_0}$, où θ_0 est l'estimateur du maximum de vraisemblance évalué sur notre échantillon x_1, \dots, x_N .

Test d'ajustement du χ^2

On partitionne alors le support de P^* , on compte le nombre d'observations dans chaque partition et on applique alors le test de Pearson pour le modèle multinomial.

Il existe d'autres tests comme celui de Kolmogorov-Smirnov ou Cramer-Von Mises qui utilisent la fonction de répartition à la place de la vraisemblance.

Merci à tous et bon courage à vous pour l'exam !



AUTOMATANTS