

Statistiques et Probabilités L2

26 janvier 2011

Chapitre 1

Vecteurs aléatoires discrets

1.1 Introduction

1.1.1 Variable aléatoire discrète

Un ensemble discret est un ensemble Ω dont le cardinal est plus petit ou égal à celui de \mathbb{N} (qui est aussi celui de \mathbb{Z}). Cela peut être les faces possibles d'une pièce de monnaie : $\Omega = \{P, F\}$, ou bien les résultats du lancer d'un dé : $\Omega = \{1, 2, 3, 4, 5, 6\}$, le nombre de client en attente à un guichet : $\Omega = \mathbb{N}$. Notons $\mathcal{P}(\mathbb{Z})$ l'ensemble des parties de \mathbb{Z} .

Définition Une application mesurable X est une fonction de Ω dans \mathbb{Z} telle que pour toute partie A de \mathbb{Z} , on ait $X^{-1}(A) \in \mathcal{P}(\Omega)$ (l'ensemble des antécédents par X de A appartient au parties de Ω , on dit que X est mesurable).

Définition une mesure de probabilité sur Ω est une fonction mesurable de $\mathcal{P}(\Omega)$ dans $[0, 1]$ en donnant une valeur (une probabilité) à chaque élément (événement) possible de Ω .

par exemple dans le cas de la pièce de monnaie : $P(F) = \frac{1}{2}$.

Définition Une variable aléatoire X est donc une application (mesurable) de $(\Omega, \mathcal{P}(\Omega), P) \longrightarrow (\mathbb{Z}, \mathcal{P}(\mathbb{Z}))$.

Définition On appelle loi de probabilité de X , la mesure image de P par X et on la note P_X .

Ainsi, la probabilité que X appartiennent au sous-ensemble A est $P_X(X \in A) := P_X(A) = P(X^{-1}(A))$. Très souvent, l'ensemble Ω est sous-entendu, invisible et on définit les lois uniquement par les probabilités que la variable aléatoire appartienne à un certain sous-ensemble de \mathbb{Z} .

Formule de Bayes On peut définir une loi conditionnelle de $X : P(X \in A | X \in B) = \frac{P(X \in A \text{ et } X \in B)}{P(X \in B)} = \frac{P(X \in A, X \in B)}{P(X \in B)}$. C'est la probabilité que la variable X appartienne à A sachant que X appartient déjà à B .

1.1.2 Quelques exemples :

- Loi de Bernoulli $b(p) : X$ dans $\{0, 1\}$, $P(X = 1) = p \in [0, 1]$.
- Loi binomiale $B(n, p) : X$ dans $\{0, \dots, N\}$ et pour $0 \leq k \leq N$, $P(X = k) = C_N^k p^k (1 - p)^{n-k}$, nous verrons plus tard que c'est la somme de N Bernoulli indépendantes et de paramètre p .
- La loi géométrique : pour $k \in \mathbb{N}^*$ et $p \in [0, 1]$, $P(X = k) = p(1 - p)^{k-1}$. C'est la loi du nombre d'essais nécessaires pour faire apparaître un évènement de probabilité p .
- Loi hypergéométrique $H(N, R, n) : X$ dans $\{0, \dots, n\}$ et pour $0 \leq k \leq N$, $P(X = k) = \frac{C_R^k \times C_{N-R}^{n-k}}{C_N^n}$, cette loi s'obtient avec l'expérience suivante : Soit une urne avec N boules, R rouges et $N - R$ noires. On tire n boules sans remise. Posons $X_i = 1$ si la i ème boule est rouge et $X_i = 0$ si elle est noire. On cherche la loi de $X = X_1 + \dots + X_n$. On aura pour $\max(0, n - (N - R)) \leq k \leq R$:

$$P(X = k) = \frac{C_R^k \times C_{N-R}^{n-k}}{C_N^n}$$

avec : C_R^k le nombre de choix de k boules rouges parmi R , C_{N-R}^{n-k} le nombre de choix de $n - k$ noires parmi $N - R$, C_N^n le nombre de choix possibles.

- Loi de Poisson $\mathcal{P}(\theta) : X \in \mathbb{N}$ et $P(X = k) = \frac{\theta^k}{k!} e^{-\theta}$

1.2 Moments d'une variable aléatoire

Une loi est connue dès qu'on connaît toutes les probabilités $P(X = k)$, pour $k \in \mathbb{Z}$. Mais si on connaît la famille de la loi, les paramètres de la loi suffisent à la caractériser. Certaines quantités (valeur moyenne ou espérance, écart quadratique moyen ou variance) peuvent donner de précieux renseignements.

1.2.1 Espérance mathématique

Une variable discrète X admettra une espérance si :

$$\sum_{-\infty}^{\infty} |k| P(X = k) < \infty$$

Il faut faire attention à ce que cette quantité soit bien finie, ainsi si on considère la variable aléatoire X de loi sur \mathbb{Z} :

$$P(X = k) = \frac{3}{\pi^2 k^2} \text{ si } k \in \mathbb{Z} \text{ et } P(X = 0) = 0.$$

La loi est symétrique donc on pourrait croire que la valeur moyenne de X est 0, malheureusement $\sum_{-\infty}^{\infty} |k| P(X = k) = \infty$, donc $E(X)$ n'est pas définie.

Pour une variable discrète X , qui admet une espérance, on définira l'espérance $E(X)$ par la formule :

$$E(X) = \sum_{-\infty}^{\infty} kP(X = k)$$

$E(X)$ est la valeur moyenne de différentes valeurs de X , pondérées par leur probabilité. Les propriétés élémentaires de l'espérance sont celles des intégrales et se déduisent par linéarité : Si X et Y sont deux variables aléatoires discrètes :

$$E(aX + bY) = aE(X) + bE(Y)$$

Espérance d'une fonction $\phi(X)$ d'une variable aléatoire discrète Pour une fonction ϕ , si :

$$\sum_{-\infty}^{\infty} |\phi(k)| P(X = k) < \infty$$

Alors, $E(\phi(X)) = \sum_{-\infty}^{\infty} \phi(k) P(X = k)$.

Inégalité de Jensen Si ϕ est une fonction convexe, on peut montrer, si les espérances existent, que :

$$E(\phi(X)) \geq \phi(E(X))$$

en particulier :

$$\begin{aligned} E(|X|) &\geq |E(X)| \\ E(X^2) &\geq (E(X))^2 \\ E(\exp(X)) &\geq \exp(E(X)) \end{aligned}$$

1.2.2 La variance

Lorsqu'elle existe, on appelle variance de X noté $V(X)$ la quantité :

$$V(X) = E[(X - E(X))^2] = E(X^2) - (E(X))^2$$

On appelle écart-type $\sigma = \sqrt{V(X)}$.

1.3 Couple de variable aléatoires discrètes

Soit (X_1, X_2) un couple de variable aléatoire discrète à valeurs dans E^2 . En notant $\|(X_1, X_2)\|^2$ le carré de la norme euclidienne, on supposera dans toute la suite que

$$E(\|(X_1, X_2)\|^2) < \infty$$

La loi de (X_1, X_2) est caractérisée par la connaissance de $P((X_1, X_2) = (k, l))$ pour $(k, l) \in \mathbb{Z}^2$. C'est ce qu'on appelle la loi jointe.

Pour un couple de variable aléatoire, on aura les définitions/propriétés suivantes :

– Loi marginale :

$$P(X_1 = k) = \sum_{l \in \mathbb{Z}} P(X_1 = k, X_2 = l), \quad P(X_2 = l) = \sum_{k \in \mathbb{Z}} P(X_1 = k, X_2 = l)$$

– Espérance :

$$E(X_1, X_2) = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} (k, l) P((X_1, X_2) = (k, l)) = (E(X_1), E(X_2))$$

– Espérance d'une fonction : pour toute fonction ϕ définie sur \mathbb{Z}^2 , on aura

$$E(\phi(X_1, X_2)) = \sum_{k \in \mathbb{Z}} \sum_{l \in \mathbb{Z}} \phi(k, l) P(X_1 = k, X_2 = l)$$

– Dépendance et indépendance : On dit que X_1 est indépendante de X_2 si pour tout $(k, l) \in E^2$

$$P(X_1 = k, X_2 = l) = P(X_1 = k) P(X_2 = l)$$

Si X_1 et X_2 ne sont pas indépendantes on dit qu'elles sont dépendantes.

– Loi conditionnelle : On rappelle que si A et B sont deux ensembles, tels que $P(B) > 0$ alors

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

En appliquant la formule de Bayes on peut définir la loi conditionnelle ainsi :

$$P(X_1 = k | X_2 = l) = \frac{P(X_1 = k, X_2 = l)}{P(X_2 = l)}$$

si $P(X_2 = l) > 0$.

– Espérance conditionnelle : On aura l'espérance conditionnelle $E(X_1 | X_2 = l)$ qui est la valeur moyenne de X_1 lorsque $X_2 = l$ qui vaudra :

$$E(X_1 | X_2 = l) = \sum_{k \in E} k P(X_1 = k | X_2 = l)$$

On pourra aussi "ne pas fixer" la valeur de X_2 , l'espérance conditionnelle $E(X_1 | X_2)$ sera alors une variable aléatoire (une fonction de X_2).

– Variance conditionnelle : On appelle variance de X_1 sachant que $X_2 = l$: $V(X_1 | X_2 = l)$ la quantité :

$$V(X_1 | X_2 = l) = E \left[(X_1 - E(X_1 | X_2 = l))^2 | X_2 = l \right]$$

Si on ne fixe pas la valeur de X_2 , on peut définir la variable aléatoire, variance conditionnelle :

$$V(X_1 | X_2) = E \left[(X_1 - E(X_1 | X_2))^2 | X_2 \right]$$

Et on aura le théorème de la variance totale :

$$V(X_1) = E[V(X_1 | X_2)] + V[E(X_1 | X_2)]$$

– Covariance :

On aura

$$Cov(X_1, X_2) = E((X_1 - E(X_1))(X_2 - E(X_2))) = E(X_1 X_2) - E(X_1)E(X_2)$$

– Variance d'une somme de variables aléatoires

$$\text{On aura } V(X_1 + X_2) = V(X_1) + V(X_2) + 2Cov(X_1, X_2)$$

– Matrice de covariance :

$$E \begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \end{pmatrix} (X_1 - E(X_1), X_2 - E(X_2)) = \begin{pmatrix} V(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & V(X_2) \end{pmatrix}$$

Remarques :

- Si X_1 est indépendant de X_2 la matrice de covariance est diagonale.
- Toutes ces propriétés se généralisent immédiatement pour $n > 2$.

1.4 Généralisation aux vecteurs aléatoires discrets

Si (X_1, \dots, X_n) est un n -uplet et si $(k_1, \dots, k_n) \in \mathbb{Z}^n$, la loi $P(X_1, \dots, X_n) = (k_1, \dots, k_n)$ caractérisera totalement (X_1, \dots, X_n) , on aura entre autre :

- Les variables (X_1, \dots, X_n) seront indépendantes si

$$P(X_1 = k_1, \dots, X_n = k_n) = \prod_{i=1}^n P(X_i = k_i)$$

- Si l'espérance existe : $E(X_1, \dots, X_n) = (E(X_1), \dots, E(X_n))$

- Si les variances de (X_1, \dots, X_n) existent. $V(X_1, \dots, X_n) = \begin{pmatrix} V(X_1) & \dots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_1, X_n) & \dots & V(X_n) \end{pmatrix}$

- Si les espérances de (X_1, \dots, X_n) existent, l'espérance d'une combinaison linéaire $E(\sum_{i=1}^n a_i X_i) = \sum_{i=1}^n a_i E(X_i)$

- Si les variances de (X_1, \dots, X_n) existent. La variance d'une somme :

$$V\left(\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

- On pourra facilement définir des probabilités conditionnelles grâce à la règle de Bayes.

Exemple de loi Un vecteur aléatoire (X_1, \dots, X_r) suit une loi multinomiale de paramètres n, p_1, \dots, p_r avec $0 < p_i < 1$ et $p_1 + \dots + p_r = 1$ si :

1. $X_1(\Omega) = \dots = X_r(\Omega) = \{0, \dots, n\}$
2. La loi de (X_1, \dots, X_r) est donnée par :

$$P(X_1 = k_1, \dots, X_r = k_r) = \frac{n!}{k_1! \dots k_r!} p^{k_1} \dots p^{k_r}$$

si $k_1 + \dots + k_r = n$, sinon la probabilité du r -uplet est nulle.

Cette loi correspond au tirage avec remise dans une urne de n boules. Ces n boules sont de r couleurs différentes c_i . p_i est la proportion de boules de couleur c_i dans l'urne.

Chapitre 2

Introduction aux chaînes de Markov à espace d'états fini

2.1 Introduction

On considère un ensemble E fini, par exemple $E = \{1, \dots, N\}$, où $N \in \mathbb{N}$. Une chaîne de Markov (homogène) à espace d'états fini est une suite de variables aléatoires $(X_1, \dots, X_n, \dots) \in E^{\mathbb{N}}$, telle que : $P(X_{n+1} = i | X_n, \dots, X_0) = P(X_{n+1} = i | X_n)$, i.e. la loi de X_{n+1} conditionnellement à tout le passé du processus ne dépend que du passé immédiat X_n .

Le processus $(X_n)_{n \in \mathbb{N}}$, est donc défini par toutes les probabilités conditionnelles définies sur E^2 : $P(X_{n+1} = i | X_n = j) := p_{ij}$.

Si on représente tous les états possibles sur le simplexe de \mathbb{R}^N , c'est-à-dire l'état i sera représenté par $(0, \dots, 0, 1, 0, \dots, 0)$ avec le 1 en i ème position, on pourra noter $(P(X_{n+1} = e_i | X_n))_{1 \leq i \leq N} = E(X_{n+1} | X_n) = MX_n$, avec $M =$

$$\begin{pmatrix} p_{11} & \cdots & p_{1N} \\ \vdots & \ddots & \vdots \\ p_{N1} & \cdots & p_{NN} \end{pmatrix} \text{ et } E(X_{n+1} | X_n) \text{ est l'espérance conditionnelle de } X_{n+1}$$

sachant X_n . On appellera M la matrice de transition de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$. On remarquera que cette matrice est "stochastique" c'est-à-dire que la somme sur des lignes sur une colonne vaudra toujours 1 ($\sum_{i=1}^N p_{ij} = 1$), ainsi $P(X_{n+1} | X_n = j)$ est bien une probabilité.

2.2 Exemple : hamster paresseux

2.2.1 La chaîne de Markov

Un hamster paresseux ne connaît que 3 endroits dans sa cage :

1. Les copeaux où il dort (état $(1, 0, 0)$)

2. La mangeoire où il mange (état $(0, 1, 0)$)
3. La roue où il fait de l'exercice (état $(0, 0, 1)$).

Ses journées sont semblables les unes aux autres et on représente son activité par une chaîne de Markov. Toutes les minutes, il peut soit changer d'activité, soit continuer celle qu'il était en train de faire :

- Quand il dort, il a 9 chances sur 10 de ne pas se réveiller la minute suivante
- Quand il se réveille, il y a 1 chance sur 2 qu'il aille manger et 1 chance sur 2 qu'il parte faire de l'exercice.
- Le repas ne dure qu'une minute, après il fait autre chose.
- Après avoir mangé, il y a 3 chances sur 10 qu'il parte courir dans sa roue, mais surtout 7 chances sur 10 qu'il retourne dormir.
- Courir est fatigant, il a donc 8 chances sur 10 de retourner dormir au bout d'une minute, sinon (avec 2 chances sur 10) il continue en oubliant qu'il est déjà un peu fatigué.

2.2.2 La matrice de transition

Si on note $(X_n)_{n \in \mathbb{N}}$, la suite des états du hamster, la matrice de transition sera $M = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix}$. Cette matrice de transition nous permet de faire facilement des prévisions sur les états futurs. Prenons l'hypothèse que le hamster dort lors de la première minute de l'étude. $X_1 = (1, 0, 0)^T$, au bout d'une minute on peut prédire :

$$\hat{X}_1 = MX_0 = \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.9 \\ 0.05 \\ 0.05 \end{pmatrix}$$

Ainsi, après une minute, on a 90% de chance que le hamster dorme encore, 5% qu'il mange et 5% qu'il court dans sa roue. On admettra de plus le lemme suivant :

lemme Soit (X_1, \dots, X_n, \dots) une chaîne de Markov homogène de matrice de transition M , on aura

$$\forall n, k \in \mathbb{N}^*, E(X_{n+k} | X_n) = M^k X_n$$

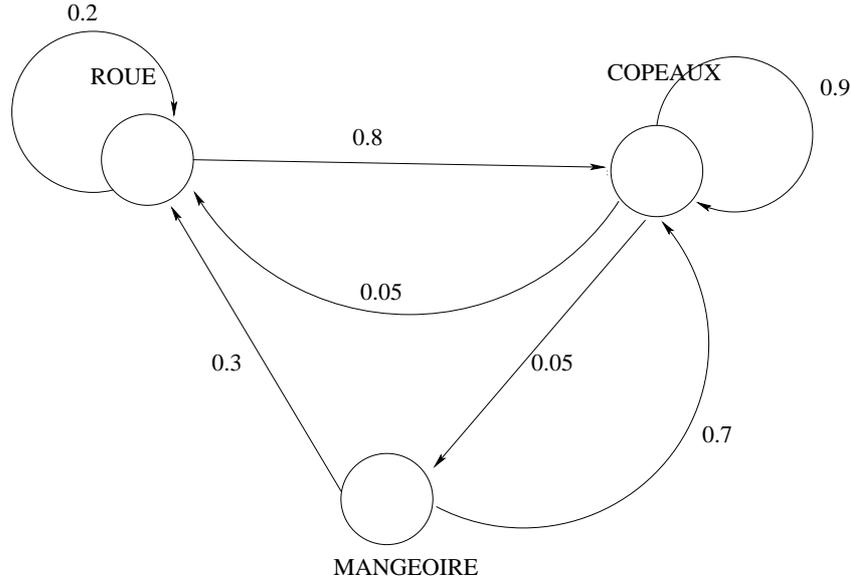
En appliquant ce lemme on aura alors $E(X_2 | X_0) = E(E(X_2 | X_1) | X_0)$ d'où

$$\begin{aligned} \hat{X}_2 = M\hat{X}_1 = MMX_0 &= \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \\ &= \begin{pmatrix} 0.9 & 0.7 & 0.8 \\ 0.05 & 0 & 0 \\ 0.05 & 0.3 & 0.2 \end{pmatrix} \begin{pmatrix} 0.9 \\ 0.05 \\ 0.05 \end{pmatrix} = \begin{pmatrix} 0.885 \\ 0.045 \\ 0.045 \end{pmatrix} \end{aligned}$$

2.3 Classe de récurrence

2.3.1 Graphe associé à une matrice de transition

On peut dessiner sur un graphe toutes les “liaisons” entre les états, c’est-à-dire les coordonnées strictement positive de la matrice de transition. Ainsi dans l’exemple précédent on aura :



On dira que l’état e_i est accessible par e_j si il existe une suite finie de points de E : $x_0 = e_j, \dots, x_n = e_i$ et telle que pour tout $k \in \{0, \dots, n-1\}$, $p_{x_{k+1}x_k} > 0$, c’est-à-dire que la probabilité de passer de l’état x_{k+1} en venant de l’état x_k est strictement positive. On notera alors $e_j \rightarrow e_i$.

Définition : Nous dirons qu’un point e_j est récurrent si $\forall e_i \in E, e_j \rightarrow e_i \Rightarrow e_i \rightarrow e_j$. Un point e_i non récurrent sera dit transitoire.

Remarques :

- Pour tout $e_j \in E$, il y a au moins un point e_i tel que $e_j \rightarrow e_i$, car $\sum_i p_{ij} = 1$.
- $e_j \rightarrow e_i$ si et seulement si il existe un entier $n \geq 1$ tel que la coordonnée ij de $M^n := M^n(i, j) > 0$. En effet,

$$M^n(i, j) = \sum_{(i_1, \dots, i_{n-1})} p_{i_1 j} p_{i_2 i_1} \dots p_{i i_{n-1}}.$$

Dans la somme précédente, tous les termes sont positifs ou nuls ; la somme n’est donc non nulle que si un des termes au moins ne l’est pas : c’est la définition de $e_j \rightarrow e_i$.

- Lorsque $M^n(i, j) > 0$, nous dirons qu'il existe un chemin de longueur n qui va de e_j à e_i .
- La relation $e_j \rightarrow e_i$ est transitive. En effet, si $M^n(i, j) > 0$ et $M^q(h, i) > 0$, alors $M^{n+q}(h, j) > 0$ puisque

$$M^{n+q}(h, j) = \sum_i M^n(i, j) M^q(h, i) \geq M^n(i, j) M^q(h, i) > 0$$

- On peut avoir ou non $e_j \rightarrow e_j$, mais si le point e_j est récurrent, alors il existe au moins un point e_i tel que $e_j \rightarrow e_i$ alors $e_i \rightarrow e_j$ et par transitivité $e_j \rightarrow e_j$. Par contre un point transitoire peut ou non satisfaire $x \rightarrow x$: considérons les deux matrices de transition suivantes sur l'ensemble $\{e_1, e_2\}$:

$$M_1 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \text{ et } M_2 = \begin{pmatrix} 0.5 & 0 \\ 0.5 & 1 \end{pmatrix}$$

Dans les deux cas, le point e_1 est transitoire, mais la relation $e_1 \rightarrow e_1$ est fautive pour M_1 et vraie pour M_2 .

- Si $e_j \rightarrow e_i$ et si e_j est récurrent, alors e_i est récurrent : il suffit d'appliquer la propriété de transitivité de la relation \rightarrow .

L'ensemble des points transitoires peut être vide, mais pas l'ensemble des points récurrents si l'ensemble E est non vide. On en déduit la proposition suivante :

Proposition : Sur l'ensemble \mathcal{R} des points récurrents, la relation $e_j \rightarrow e_i$ est une relation d'équivalence.

Définition : On appelle classe de récurrence les classes d'équivalence de \mathcal{R} pour cette relation d'équivalence.

2.4 Mesure invariante

Une mesure de probabilité μ sur E pourra toujours être représentée par un vecteur $\mu := (\mu_1, \dots, \mu_N)^T$ tel que les μ_i soient tous positifs ou nuls et $\sum_{i=1}^N \mu_i = 1$ i.e. $\mu_i = P(X = i)$. La mesure μ sera dite invariante pour M si et seulement si $\mu = M\mu$. Pour une matrice de transition, il existe toujours une mesure invariante. En effet, pour une mesure μ_0 quelconque considérons la suite $\mu_n := \frac{1}{n+1} \sum_{i=0}^n M^i \mu_0$, c'est une suite de probabilité, donc de points dans le compact $[0, 1]^N$, elle admet donc une sous-suite convergente $\mu_{n_k} \xrightarrow{k \rightarrow \infty} \mu$. Mais alors

$$M\mu_{n_k} = \frac{1}{n_k + 1} \sum_{i=0}^{n_k} M^{i+1} \mu_0 = \mu_{n_k} + \frac{M^{n_k+1} \mu_0 - \mu_0}{n_k + 1}$$

et lorsque $k \rightarrow \infty$, $\frac{(M^{n_k+1} \mu_0 - \mu_0)}{n_k + 1} \rightarrow 0$, puisqu'il s'agit de la différence de deux mesures de probabilité divisée par $n_k + 1$. Alors, nous voyons en passant à la limite que $M\mu = \mu$.

CHAPITRE 2. INTRODUCTION AUX CHAÎNES DE MARKOV À ESPACE D'ÉTATS FINI 11

De plus, bien souvent (si la chaîne est irréductible récurrente et apériodique), on aura M^n qui convergera vers une matrice dont chaque ligne est l'unique distribution invariante μ : $\lim_{n \rightarrow \infty} M^n \mu_0 = \mu$.

Chapitre 3

Vecteurs aléatoires réels

3.1 Introduction

Un vecteur aléatoire X est une application de (Ω, \mathcal{A}, P) dans un espace vectoriel réel : \mathbb{R}^n muni de sa tribu borélienne (l'ensemble des parties de \mathbb{R}^n stable par réunion et intersection dénombrable). On identifiera X au n-uple de variables aléatoires formé par ses composantes : $X = (X_1, \dots, X_n)$.

3.2 Fonction de répartition et densité

3.2.1 Fonction de répartition

F est une application de \mathbb{R}^n dans $[0; 1]$ définie par :

$$F(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n)$$

3.2.2 Densité

f , si elle existe est définie par :

$$f(x_1, \dots, x_n) = \frac{\partial^n F}{\partial x_1 \dots \partial x_n}(x_1, \dots, x_n)$$

Pour illustrer ces notions nous allons étudier en détail le cas $n = 2$, la généralisation au cas n quelconque est immédiate.

3.2.3 Couple de variable aléatoires continues

Soit (X, Y) un couple de variables aléatoires continues à valeurs dans \mathbb{R}^2 . On se placera dans le cadre où la loi de (X, Y) est caractérisée par la connaissance de $f(x, y)$ pour $(x, y) \in \mathbb{R}^2$. On aura les propriétés suivantes :

- Loi marginale de X et Y :

$$h(x) = \int_{\mathbb{R}} f(x, y) dy \text{ et } g(y) = \int_{\mathbb{R}} f(x, y) dx$$

- Espérance

$$E(X, Y) = (E(X), E(Y))$$

- Dépendance et indépendance : On dit que X est indépendante de Y si pour tout $(x, y) \in \mathbb{R}^2$

$$f(x, y) = h(x)g(y)$$

- Densité conditionnelle : en supposant que les dénominateurs soient bien non nuls, les densité conditionnelles à y puis à x seront :

$$h^y(x) := \frac{f(x, y)}{g(y)} \text{ et } g^x(y) := \frac{f(x, y)}{h(x)}$$

- L'espérance conditionnelle : On aura $E(X | Y = y)$ qui sera la valeur moyenne de la variable X lorsque la variable Y vaut y :

$$E(X | Y = y) = \int_{\mathbb{R}} x h^y(x) dx$$

On pourra aussi "ne pas fixer" la valeur de Y , l'espérance conditionnelle $E(X | Y)$ sera alors une variable aléatoire (fonction de Y).

- Matrice de covariance :

$$E \begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \end{pmatrix} (X_1 - E(X_1), X_2 - E(X_2)) = \begin{pmatrix} V(X_1) & Cov(X_1, X_2) \\ Cov(X_1, X_2) & V(X_2) \end{pmatrix}$$

Remarques :

- Si X_1 est indépendant de X_2 la matrice de covariance est diagonale.
- Toutes ces propriétés se généralisent immédiatement pour $n > 2$.

3.2.4 Application à la somme de variables aléatoires discrètes

Si (X_1, \dots, X_n) est un n -uplet on aura ainsi :

- $E(X_1, \dots, X_n) = (E(X_1), \dots, E(X_n))$
- $V(X_1, \dots, X_n) = \begin{pmatrix} V(X_1) & \dots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_1, X_n) & \dots & V(X_n) \end{pmatrix}$
- $E(\sum_{i=1}^n X_i) = \sum_{i=1}^n E(X_i)$ et

$$V \left(\sum_{i=1}^n X_i \right) = \sum_{i,j=1}^n Cov(X_i, X_j) = \sum_{i=1}^n V(X_i) + 2 \sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

3.2.5 Exemple des vecteurs gaussiens

On dira qu'une variable aléatoire X suit la loi gaussienne de paramètres m et σ^2 : $X \sim \mathcal{N}(m, \sigma^2)$, si la densité de la loi de X est $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$ pour $x \in \mathbb{R}$. On considère le couple (X_1, X_2) , avec X_1 et X_2 gaussiens, si $X_1 \sim \mathcal{N}(m_1, \sigma_1^2)$ de densité f_1 et $X_2 \sim \mathcal{N}(m_2, \sigma_2^2)$ de densité f_2 . Si, en plus, X_1 est indépendant de X_2 alors la loi du couple (X_1, X_2) aura pour densité $f(x_1, x_2) = f_1(x_1) \times f_2(x_2)$. On dira alors que (X_1, X_2) est un couple gaussien de loi $\mathcal{N}\left(\begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right)$.

Maintenant, on admettra (cela sera démontré l'année prochaine) que l'image par une application linéaire de ce couple gaussien reste un vecteur gaussien.

Ainsi si on considère l'application représentée par la matrice $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

Le vecteur $\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = A \times \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ sera un couple gaussien. Par linéarité sa moyenne sera $E\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = E\left[A \times \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right] = AE\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = A \times \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$.

Sa matrice de covariance sera $\Sigma = E\left(\begin{pmatrix} Y_1 - E(Y_1) \\ Y_2 - E(Y_2) \end{pmatrix} \times (Y_1 - E(Y_1), Y_2 - E(Y_2))\right) = E\left(A \begin{pmatrix} X_1 - E(X_1) \\ X_2 - E(X_2) \end{pmatrix} \times (X_1 - E(X_1), X_2 - E(X_2)) A^T\right) = A \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} A^T$.

On dira alors que (Y_1, Y_2) est le couple gaussien de loi $\mathcal{N}\left(A \times \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}, A \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} A^T\right)$, si on note $\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = A \times \begin{pmatrix} m_1 \\ m_2 \end{pmatrix}$ et $\Sigma = A \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} A^T$, la densité du couple (Y_1, Y_2) sera

$$\forall (y_1, y_2) \in \mathbb{R}^2 : g(y_1, y_2) = \frac{1}{2\pi} \frac{1}{\sqrt{\det \Sigma}} e^{-\frac{1}{2}(y_1 - \mu_1, y_2 - \mu_2)\Sigma^{-1} \begin{pmatrix} y_1 - \mu_1 \\ y_2 - \mu_2 \end{pmatrix}}$$

Plus généralement, si $X = (X_1, \dots, X_n)^T$ est un vecteur gaussien de dimension n , de vecteur moyen m et de matrice de covariance Σ inversible, alors la densité de X sera pour tout $x \in \mathbb{R}^n$:

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

On dira aussi que la loi de X est : $X \sim \mathcal{N}(m, \Sigma)$. Les vecteurs gaussiens sont caractérisés par la propriété suivante :

Proposition $X = (X_1, \dots, X_n)^T$ est un vecteur gaussien si et seulement si toute combinaison linéaire $Z = a_1 X_1 + \dots + a_n X_n$ est une variable aléatoire gaussienne.

Chapitre 4

Estimation paramétrique

4.1 Loi dépendant d'un paramètre

Un grand nombre de lois dépendent d'un paramètre. Ainsi on peut donner comme exemple :

- La loi de Bernoulli de paramètre $p \in [0, 1]$: $X \in \{0, 1\}$ et $P(X = 1) = p$, $E(X) = p$ et $V(X) = p(1 - p)$
- La loi Binomiale de paramètres $n \in \mathbb{N}^*$ et $p \in [0, 1]$: $X \in \{0, \dots, n\}$ et pour $k \in \{0, \dots, n\}$ $P(X = k) = C_n^k p^k (1 - p)^{n-k}$, $E(X) = np$, $V(X) = np(1 - p)$.
- La loi de Poisson de paramètre $\lambda > 0$: $X \in \mathbb{N}$ et pour $k \in \mathbb{N}$, $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $E(X) = \lambda$, $V(X) = \lambda$.
- La loi exponentielle de paramètre $\theta > 0$, de densité $f(x) = \theta e^{-\theta x} \mathbf{1}_{\mathbb{R}^+}(x)$, $E(X) = \frac{1}{\theta}$, $V(X) = \frac{1}{\theta^2}$.
- Loi du χ^2 à n degrés de libertés :

$$f(x) = \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbf{1}_{\mathbb{R}^+}(x)$$

- Plus généralement la loi Gamma de paramètres $p > 0$ et $\theta > 0$ de densité

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

avec pour $p \in \mathbb{R}^{+*}$, $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$. On aura alors $\mathcal{E}(\theta) = \Gamma(1, \theta)$ et $\chi^2(n) = \Gamma(\frac{n}{2}, \frac{1}{2})$.

- La loi gaussienne de paramètres m et σ^2 , de densité $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-m)^2}$

Bien souvent, on disposera d'observations (x_1, \dots, x_n) supposées être les réalisations des variables (X_1, \dots, X_n) indépendantes et identiquement distribuées. On supposera que les observation appartiennent à une certaine famille de lois (par exemple des lois de Bernoulli si leurs valeurs sont dans $\{0, 1\}$ ou des gaussiennes si leurs valeurs sont réelles) et on cherchera une fonction (appelée statistique)

des observations (x_1, \dots, x_n) qui permette de d'estimer le ou les paramètres de la loi. C'est la démarche des statistiques paramétriques.

4.2 Loi des statistiques

Le but des statistiques paramétriques est d'estimer un paramètre à l'aide d'une fonction des observations $(x_1, \dots, x_n) : T(x_1, \dots, x_n) := \hat{\theta}_n$. Pour choisir la meilleure fonction possible il faut étudier le comportement théorique de la variable aléatoire $T(X_1, \dots, X_n) = \hat{\theta}_n$. Notons θ le paramètre (éventuellement vectoriel), on considérera pour cela :

– Le biais : $b := E(\hat{\theta}_n - \theta)$

– La variance : $V(\hat{\theta}_n)$

– Plus généralement le risque quadratique : $Q = E(\hat{\theta}_n - \theta)^2 = V(\hat{\theta}_n) + b^2$.

Le premier but est d'obtenir un estimateur consistant du paramètre, à savoir, tel le biais et la variance ou bien le risque quadratique de $\hat{\theta}_n$ converge vers 0 quand $n \rightarrow \infty$. Enfin on essaiera aussi de préciser "la vitesse" de cette convergence pour avoir éventuellement des intervalles de confiance. On utilisera 3 résultats de probabilité :

La loi faible des grands nombres : Si $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. et $E(|X_i|^2) < \infty$, alors $V(\bar{X}) = \frac{V(X_i)}{n} \rightarrow 0$ et $\bar{X} \xrightarrow{P} m$, car $P(|\bar{X} - E(\bar{X})| > \varepsilon) \leq \frac{V(\bar{X})}{\varepsilon^2}$, par l'inégalité de B.T.

La loi forte des grands nombres : Si $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. et $E(|X_i|) < \infty$, alors $\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} E(X_i)$

Le théorème de la limite centrale : Si $(X_n)_{n \in \mathbb{N}}$ est une suite i.i.d. et $E(|X_i|^2) < \infty$, alors

$$\sqrt{n}(\bar{X} - E(X_i)) \xrightarrow{\mathcal{L}} \mathcal{N}(0, V(X_i))$$

Le lemme de Slutsky : Si $X_n \xrightarrow{\mathcal{L}} X$ et $Y_n \xrightarrow{P} C$, où C est une constante, alors :

$$X_n Y_n \xrightarrow{\mathcal{L}} CX$$

Si $C \neq 0$, on aura aussi

$$\frac{X_n}{Y_n} \xrightarrow{\mathcal{L}} \frac{X_n}{C}$$

4.3 Estimation de la moyenne et la variance d'une loi normale

L'hypothèse de normalité des observations est une hypothèse très courante et souvent justifiée. On va donc donner les estimateurs des paramètres de cette loi pour l'observation d'un n échantillon (X_1, \dots, X_n) .

4.3.1 Estimateur de m :

On a $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ qui est un estimateur sans biais et convergent de m . En effet : $\bar{X} \rightarrow m$ par la loi forte des grands nombres. On remarquera de plus que $E(\bar{X}) = m$ cet estimateur est donc sans biais pour tout $n \in \mathbb{N}^*$.

4.3.2 Estimateur de σ^2 :

- Si m est connu (par exemple les données sont centrées) on aura pour estimateur de σ^2 :

$$\hat{\sigma}^2 := \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

Comme $E(\hat{\sigma}^2) = \sigma^2$ c'est un estimateur sans biais, pour tout n . De plus, par la loi forte des grands nombres $\hat{\sigma}^2 \xrightarrow{p.s.} \sigma^2$, il est donc convergent.

- Si m est inconnu, on aura pour estimateur de σ^2 :

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

cet estimateur est sans biais car, on peut montrer :

$$\sum_{i=1}^n (X_i - \bar{X})^2 \sim \sigma^2 \chi^2(n-1)$$

On aura de plus $\sum_{i=1}^n (X_i - \bar{X})^2 = \sum_{i=1}^n X_i^2 - n\bar{X}^2$, donc par la loi forte des grands nombres $S^2 \xrightarrow{p.s.} \sigma^2$.

4.3.3 Intervalle de confiance pour m et σ^2

En connaissant la loi des estimateurs de m et σ^2 , on peut donner un intervalle contenant les vrais paramètres avec une grande probabilité (par exemple 95%, mais on peut le remplacer par n'importe quoi). Mais en se rapportant aux exemples de lois on trouve :

$$T = \frac{\sqrt{n}(\bar{X} - m)}{\sqrt{S^2}} \sim \frac{\mathcal{N}(0,1)}{\frac{1}{\sqrt{n-1}}\sqrt{\chi^2(n-1)}} = \mathcal{ST}(n-1)$$

On aura ainsi

$$P(-C < T < C) = P(-C < \mathcal{ST}(n-1) < C)$$

Par exemple, si on choisit C tel que $P(-C < \mathcal{ST}(n-1) < C) = 0.95$, on aura alors

$$P\left(-C < \frac{\sqrt{n}(\bar{X} - m)}{\sqrt{S^2}} < C\right) \Leftrightarrow P\left(\bar{X} - \frac{C}{\sqrt{n}}\sqrt{S^2} < m < \bar{X} + \frac{C}{\sqrt{n}}\sqrt{S^2}\right) = 0.95$$

Ce qui veut donc dire : Avec une probabilité de 95%, m appartient à l'intervalle

$$\left[\bar{X} - \frac{C}{\sqrt{n}}\sqrt{S^2}, \bar{X} + \frac{C}{\sqrt{n}}\sqrt{S^2}\right]$$

De même, on peut choisir deux constantes telles que

$$P(B < \chi^2(n-1) < C) = 0.95$$

on aura alors

$$P\left(B(n-1) < \frac{(n-1)S^2}{\sigma^2} < C(n-1)\right) = 0.95$$

d'où

$$P\left(\frac{S^2}{C} < \sigma^2 < \frac{S^2}{B}\right) = 0.95$$

4.4 Estimation par maximum de vraisemblance

On décrit ici une méthode permettant d'obtenir dans de très nombreux cas des estimateurs possédant de bonnes qualités.

4.4.1 Vraisemblance d'un échantillon i. i. d.

Soit X une variable aléatoire dont la densité $f(x)$, dépend d'un paramètre (éventuellement vectoriel) θ . Soit (x_1, \dots, x_n) une réalisation de l'échantillon i.i.d. (X_1, \dots, X_n) , la vraisemblance (qui est en faite la densité de (X_1, \dots, X_n)) de (x_1, \dots, x_n) est alors :

$$L_\theta(x_1, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

On aura par exemple :

Bernoulli Ici, il n'y a pas a proprement parler de densité, mais une probabilité qui joue son rôle et qui peut se résumer de la façon suivante : $P(X = x) := p^x (1-p)^{1-x}$. On aura alors pour un échantillon (x_1, \dots, x_n) , avec $0 < p < 1$:

$$L_p(x_1, \dots, x_n) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = (1-p)^n \left(\frac{p}{1-p}\right)^{\sum_{i=1}^n x_i}$$

Loi normale Ici on a deux paramètres m et σ^2 , on aura donc

$$L_{(m,\sigma^2)}(x_1, \dots, x_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2}(x_i-m)^2} = \left(\frac{1}{\sqrt{2\pi\sigma}} \right)^n e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i-m)^2}$$

4.4.2 Estimateur du maximum de vraisemblance

Pour (x_1, \dots, x_n) une réalisation de l'échantillon i.i.d. (X_1, \dots, X_n) , on appelle estimateur du maximum de vraisemblance le paramètre $\hat{\theta}_n$ qui maximise $\theta \mapsto L_\theta(x_1, \dots, x_n)$. Lorsque le support de la loi (i.e. l'ensemble où la densité de X_i est strictement positive) ne dépend pas du paramètre θ , on peut souvent obtenir l'estimateur du maximum de vraisemblance en cherchant les points où la dérivée première s'annule et où la dérivée seconde est négative :

$$\frac{\partial}{\partial \theta} L_{\hat{\theta}_n}(x_1, \dots, x_n) = 0 \text{ et } \frac{\partial^2}{\partial \theta^2} L_{\hat{\theta}_n}(x_1, \dots, x_n) < 0$$

Remarque

- L'estimateur du maximum de vraisemblance $\hat{\theta}_n$ n'existe pas toujours et n'est pas toujours unique. Il faudra examiner chaque cas particulier.
- $\hat{\theta}_n$ est une fonction (on dit statistique) des observations x_1, \dots, x_n . Pour étudier son comportement, on l'étudiera comme fonction des variables aléatoires X_1, \dots, X_n , $\hat{\theta}_n$ sera alors une *variable aléatoire*.
- Dans la majorité des cas pratiques, θ est vectoriel. Ainsi $\frac{\partial}{\partial \theta} L_{\hat{\theta}_n}(x_1, \dots, x_n) = 0$ signifiera que ce le vecteur gradient qui doit s'annuler.
De même $\frac{\partial^2}{\partial \theta^2} L_{\hat{\theta}_n}(x_1, \dots, x_n) < 0$, signifiera que c'est la matrice Hessienne qui doit être définie négative.

Théorème (admis) Si l'estimateur du maximum de vraisemblance existe et est unique alors il est consistant.

On traitera quelques exemples pour illustrer le cours.

Chapitre 5

Introduction au tests statistiques

5.1 Principes des tests

Un test statistique permet de décider entre deux hypothèses H_0 et H_1 . Cette décision se fera à partir d'une réalisation d'un échantillon. L'hypothèse H_0 est une hypothèse privilégiée appelée hypothèse de base. H_1 est l'hypothèse alternative. On associe à un test un niveau α , ou risque de première espèce (généralement entre 10% et 1%) . Une fois que le niveau α est fixé on peut en déduire la puissance η ou de manière équivalente le risque de deuxième espèce $\beta = 1 - \eta$.

Définition Soit (X_1, \dots, X_n) un n-échantillon et θ un paramètre de la loi de X_i . Soit les hypothèse H_0 et H_1 :

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

On définit une statistique $T(X_1, \dots, X_n)$ et la règle de décision du test sera de rejeter l'hypothèse de base H_0 si $T(X_1, \dots, X_n)$ appartient à la région de rejet W .

On définit le niveau α du test (ou erreur de 1ère espèce) et la puissance η du test (ou l'erreur de seconde espèce $\beta = 1 - \eta$) par :

- P_{H_0} (Choisir H_1) = $\alpha = \sup_{\theta \in \Theta_0} P_{\theta}(T \in W)$
- P_{H_1} (Choisir H_1) = $\eta = \inf_{\theta \in \Theta_1} P_{\theta}(T \in W)$
- P_{H_1} (Choisir H_0) = $\beta = \sup_{\theta \in \Theta_1} P_{\theta}(T \notin W)$

Ici P_{H_0} signifie la probabilité sous H_0 , c'est-à-dire en supposant que l'hypothèse H_0 est vraie, de même P_{H_1} signifie la probabilité sous H_1 , c'est-à-dire en supposant que l'hypothèse H_1 est vraie.

Construction d'un test

- On pose souvent α à priori, un choix standard est $\alpha = 5\%$.
- Choix des hypothèses H_0 et H_1 . Il ne faut pas oublier que H_0 est avantagée, donc H_0 est souvent l'hypothèse pessimiste, ou bien neutre. H_1 l'hypothèse alternative est celle qui, si elle est acceptée, montre l'efficacité ou bien l'innocuité d'une décision.
- Choix de la statistique de test $T(X_1, \dots, X_n)$. Celle-ci peut être choisie "intuitivement", par exemple la moyenne empirique si on veut que le test porte sur une moyenne théorique. Celle-ci est aussi souvent déterminée par la "rapport de vraisemblance", qui donne un test optimale (sous de bonnes conditions).
- Détermination explicite de la région de rejet W . La région de rejet W est toujours associée à la statistique $T(X_1, \dots, X_n)$. Par exemple si vous voulez tester $H_0 : E(X) = 0$ contre $H_1 : E(X) = 1$, on aura $T(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i := \bar{X}$ et $W = \bar{X} > C$. Cela veut dire qu'on rejettera la nullité de la moyenne si on observe une moyenne théorique "trop grande". Les constantes intervenant dans la définition de la région de rejet seront calculées de façon que $P_{H_0}(W) = \alpha$.
- Calcul de la puissance du test $P_{H_1}(W) = \eta$. Plus un test sera puissant mieux cela sera.
- Conclusion pour la réalisation (x_1, \dots, x_n) de l'échantillon.

Remarques

- Le choix de H_0 influe grandement sur la décision du test. Les hypothèses H_0 et H_1 ne sont pas interchangeables.
- Une hypothèse H_0 ou H_1 est dite simple si elle est associée à un singleton (Θ_0 ou Θ_1), sinon elle sera dite multiple, ou bien composite. Si H_1 est multiple de la forme $\theta > \theta_0$ ou $\theta < \theta_0$, on parlera de test unilatéral. Si H_1 est multiple de la forme $\theta \neq \theta_0$ on parlera de test bilatéral.

5.2 Test du rapport de vraisemblance

5.2.1 Deux hypothèse simples : Lemme de Neyman et Pearson

Dans le modèle de Neyman-Pearson, les deux ensembles qui définissent les hypothèses Θ_0 et Θ_1 sont constitués d'un seul élément :

$$\mathbf{H0} : \theta = \theta_0$$

$$\mathbf{H1} : \theta = \theta_1$$

Le niveau du test α étant fixé (par exemple 5%) on cherche la forme de la région de rejet W qui, en vérifiant $P_{H_0}(W) = \alpha$, maximisera la puissance du test. On parlera alors du test le plus puissant. On aura le résultat suivant :

Lemme de Neyman et Pearson Soit L_0 la vraisemblance des observations sous l'hypothèse H_0 et L_1 la vraisemblance des observations sous l'hypothèse

H_1 . La région de rejet du test le plus puissant de H_0 contre H_1 est de la forme

$$\frac{L_0}{L_1} < K$$

5.2.2 Généralisation à H_1 multiple

Supposons que pour le test l'hypothèse alternative H_1 soit multiple mais unilatérale, c'est-à-dire que :

$$\mathbf{H0} : \theta = \theta_0$$

$$\mathbf{H1} : \theta > \theta_0$$

ou bien de la forme

$$\mathbf{H0} : \theta = \theta_0$$

$$\mathbf{H1} : \theta < \theta_0$$

Si la forme du test le plus puissant ne dépend pas du choix de $\theta_1 \in \Theta_1$, par exemple

$$\forall \theta_1 \in \Theta_1, \frac{L_0}{L_1} < K \Leftrightarrow \bar{X} < C$$

alors le test sera dit uniformément plus puissant (U.P.P.).

5.3 Test gaussiens

On dispose d'un échantillon gaussien (X_1, \dots, X_n) i.i.d. On veut tester $H_0 : m = m_0$ contre $H_1 : m \in \Theta_1$ avec $m_0 \notin \Theta_1$ au niveau α . Notons

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \text{ et } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

On dispose alors de la statistique $T = \sqrt{n} \frac{\bar{X} - m_0}{\sqrt{S^2}}$, qui suit sous H_0 la loi d'une Student à $n-1$ degrés de libertés. La région de rejet du test sera alors à adapter suivant l'alternative Θ_1 . Par exemple si $\Theta = m > m_0$, $W = \{T > C\}$ avec C tel que $P_{H_0}(T > C) = \alpha$.

5.4 Test d'adéquation du χ^2

C'est un test pour une variable X prenant valeur dans un ensemble fini E de cardinal N . Soit $P^0 = (p_1^0, \dots, p_N^0)$ la loi sous l'hypothèse de base H_0 , i.e. $P^0(X = i) = p_i^0$. On veut tester, au niveau α :

$$\mathbf{H0} : P = P_0$$

$$\mathbf{H1} : P \neq P_0$$

Notons $n_j = \sum_{i=1}^n \mathbf{1}_{\{j\}}(X_i)$ le nombre d'observation valant j dans l'échantillon et $\hat{p}_j = \frac{n_j}{n}$ la fréquence de l'observation j dans l'échantillon. La statistique de test :

$$T = n \sum_{j=1}^N \frac{(\hat{p}_j - p_j^0)^2}{p_j^0} = \sum_{j=1}^N \frac{(n_j - np_j^0)^2}{np_j^0}$$

aura pour loi asymptotique, sous $H_0 : \chi^2(N-1)$. On pourra donc construire la région de rejet $W = T > C$, avec C vérifiant $P(\chi^2(N-1) > C) = \alpha$.