

Mathématiques

14 février 2007

Table des matières

I	Probabilités	4
1	Terminologie	5
2	Définitions	5
3	Probabilités élémentaires	7
4	Espaces probabilisés discrets	8
5	Probabilités continues	9
6	Indépendance, probabilités produits	9
7	Probabilités conditionnelles	11
8	Fonction d'une variable ou plusieurs variables aléatoires	16
9	V.a.r.d. particulières	21
9.1	Variable de Bernoulli	21
9.2	Variable Binomiale	21
9.3	Loi discrète uniforme	21
9.4	Variable géométrique	21
9.5	Variable binomiale négative	22
9.6	Variable de Poisson	22
9.7	Variable Gaussienne	22
9.8	Quelques relations limites	22
II	Information, compression	24
10	Entropie d'une loi finie	24
11	Interprétation	24
12	Compression	26
12.1	Sources discrètes	26
12.2	Cas particuliers	26
12.3	Théorème de McMillan (cas d'une source simple)	26
12.4	Entropie conditionnelle	28
III	Chaines de Markov	29

Première partie
Probabilités

1 Terminologie

Évènement proposition dont on sait si elle est vraie ou fausse à l'issue d'une expérience aléatoire.

Si A, B sont des évènements

- $\neg A$ est un évènement
- $A \vee B$ est un évènement
- $A \wedge B$ est un évènement

L'ensemble des évènements a une structure d'algèbre de Boole pour la négation, le "ET" et le "OU".

Évènement certain évènement qui prend la valeur 1 quelque soit l'issue de l'expérience. ex : $H \vee F$

Évènement impossible évènement qui prend la valeur 0 quelque soit l'issue de l'expérience. ex : $H \wedge F$

Ω ensemble.

\mathcal{A} un ensemble de parties de Ω

2 Définitions

Def On dit qu'un sous-ensemble non-vide \mathcal{A} de l'ensemble $\mathcal{P}(\Omega)$ ($\mathcal{P}(\Omega)$: ensemble des parties de Ω) est une algèbre de Boole - on dit aussi tribu finie

- si $\forall A \in \mathcal{A}, \bar{A} \in \mathcal{A}$
- $\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$
- $\emptyset \in \mathcal{A}$

nb : ceci implique que $\Omega \in \mathcal{A}$.

Def On dit qu'un sous-ensemble nonvide \mathcal{A} de $\mathcal{P}(\Omega)$ est une σ -algèbre de Boole (on dit aussi une tribu) si c'est une algèbre de Boole, et si pour toute suite d'éléments de \mathcal{A} ,

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$$

Def On dit qu'un sous-ensemble non-vide \mathcal{A} de $\mathcal{P}(\Omega)$ est une σ -algèbre de Boole si c'est une algèbre de Boole, et si pour toute famille dénombrable $(A_i)_{i \in I}$ d'éléments de \mathcal{A} ,

$$\bigcup_{i=1} A_i \in \mathcal{A}$$

Def On appelle *espace probabilisable* un couple (Ω, \mathcal{A}) où \mathcal{A} est un sous-ensemble non-vide de $\mathcal{P}(\Omega)$ qui est une σ -algèbre de Boole.

Def On appelle Ω l'ensemble fondamental et ses éléments sont appelés *tirages* (ou *épreuves*).

Def On appelle *évènements* tous les éléments de \mathcal{A} .

Def On appelle *probabilités sur un espace probabilisable* (Ω, \mathcal{A}) une fonction $\mathcal{P} : \mathcal{A} \rightarrow [0, 1] \subset \mathbb{R}$ telle que :

1. $\mathcal{P}(\Omega) = 1$
2. Pour toute suite A_1, A_2, \dots d'évènements deux à deux disjoints

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Évènements deux à deux disjoints : $i \neq j \Rightarrow A_i \cap A_j = \emptyset$

Deux évènements disjoints sont dits incompatibles.

2 \Rightarrow 2 bis : $\forall A, B \in \mathcal{A}$ tels que $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$

Def le triplet (Ω, \mathcal{A}, P) ainsi défini, est appelé *espace probabilisé*.

Conséquence Soit (Ω, \mathcal{A}, P) un espace probabilisé :

1. $P(\emptyset) = 0$
 - 2 bis** avec $A = \Omega, B = \emptyset$
 $A \cap B = \emptyset, \Omega \cap \emptyset = \emptyset$: oui.
 - 2 bis** $\Rightarrow P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) = 1 + 0$
2. $\forall A \in \mathcal{A} \quad P(\bar{A}) = 1 - P(A)$
 $A = A, B = \bar{A} : A \cap \bar{A} = \emptyset$
 $P(A \cup \bar{A}) = P(A) + P(\bar{A}) = 1$
3. $\forall A, B \in \mathcal{A} \quad A \subset B \Rightarrow P(A) \leq P(B)$
 $A \cap (B \setminus A) = \emptyset$
 $P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$
 $P(B) = P(A) + P(B \setminus A)$
4. $\forall A, B \in \mathcal{A} \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 - (a) $P(A \cup B) = P(A) + P(B \setminus (A \cup B))$
 - (b) $P(B) = P(B \setminus (A \cap B)) + P(A \cap B)$

Def On dit que $A, B \in \mathcal{A}$ sont incompatibles si $A \cap B = \emptyset$.

Def On appelle *évènement certain* Ω .

Def On appelle *évènement impossible* \emptyset .

Def On appelle *presque sûr* un évènement dont la probabilité vaut 1.
L'évènement certain est toujours presque sûr.

Def On appelle *négligeable* l'évènement dont la probabilité est 0.
L'évènement impossible est toujours négligeable.

3 Probabilités élémentaires

1. Ω : Ensemble fini.
2. $\mathcal{A} : \mathcal{P}(\Omega)$
3. $\forall \omega \in \Omega \quad P(\{\omega\}) = \frac{A}{\text{Card}(\Omega)}$

Tout évènement $E \in \Omega$ admet une partition :

$$E = \cup_{\omega \in E} \{\omega\} = \{\omega_1\} \cup \{\omega_2\} \cup \dots \cup \{\omega_n\}$$

On a :

$$P(E) = \sum_{\omega \in E} P(\{\omega\}) = \frac{n}{\text{Card}(\Omega)} = \frac{\text{Card}(E)}{\text{Card}(\Omega)}$$

4 Espaces probabilisés discrets

1. Ω dénombrable (donc possiblement fini).
2. $\mathcal{A} = \mathcal{P}(\Omega)$

Tout élément de \mathcal{A} est une réunion dénombrable de singletons $\{\omega_i\}$ dont les probabilités sont connues, et qui sont disjoints :

$$\forall E \in \mathcal{A} \quad P(E) = \sum_{\omega \in E} P(\{\omega\})$$

Pour définir une probabilité P dans ce cadre, il suffira de définir $P(\{\omega\})$ pour tout ω avec les conditions :

1. $\forall \omega \in \Omega \quad 0 \leq P(\{\omega\}) \leq 1$
2. $\sum_{\omega \in \Omega} P(\{\omega\}) = 1$

5 Probabilités continues

Paradoxe : une surface est formée de points.

$P(\text{être dans la surface}) = 0$

$P(\text{être sur un point précis}) = 0$

6 Indépendance, probabilités produits

Def Soit (Ω, \mathcal{A}, P) un espace probabilisé.

On dit que $A, B \in \mathcal{A}$ sont indépendants si $P(A \cap B) = P(A) \cdot P(B)$.

Exemple 1 carte tirée dans un jeu de 56 cartes.

R : "on tire un roi"

C : "on tire un carreau" R et C indépendants ?

$$P(R) = \frac{4}{52}$$

$$P(C) = \frac{13}{52}$$

$$P(R \cap C) = \frac{1}{52}, \text{ or } \frac{1}{52} = \frac{4}{52} \cdot \frac{13}{52}$$

Les deux évènements sont donc indépendants.

Exemple Avec un dé à 6 faces

A : "nombre pair"

B : "nombre ≤ 3 "

C : "nombre ≤ 2 "

$$P(A) = \frac{1}{2} \quad P(A \cap B) = \frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{2} : A \text{ et } B \text{ non indépendants}$$

$$P(B) = \frac{1}{2} \quad P(A \cap C) = \frac{1}{6} = \frac{1}{2} \cdot \frac{1}{3} : A \text{ et } C \text{ indépendants}$$

$$P(C) = \frac{1}{3} \quad P(B \cap C) = \frac{1}{3} = \frac{1}{2} \cdot \frac{1}{3} : B \text{ et } C \text{ non indépendants.}$$

Proposition Si $A, B \in \mathcal{A}$ sont indépendants, alors A et B sont indépendants, de même que \bar{A} et \bar{B} .

Proposition Deux évènements incompatibles non nuls ne sont pas indépendants.

Proposition Un évènement négligeable est indépendant de tout autre.

Def Deux σ -algèbres \mathcal{A} et \mathcal{A}' sur Ω sont indépendantes si

$$\forall A \in \mathcal{A}, \forall A' \in \mathcal{A}' \quad A \text{ et } A' \text{ sont indépendants}$$

Def Indépendance mutuelle d'une famille d'évènements :

Les évènements d'une suite finie A_1, A_2, \dots, A_n sont mutuellement indépendants (indépendants dans leur ensemble) si

$$\forall I \subset [1, n] \subset \mathbb{N}, P(\cap_{i \in I} A_i) = \prod_{i \in I} P(A_i)$$

Exemple On tire deux fois à pile ou face.

$$\Omega = \{(P, P), (P, F), (F, P), (F, F)\}$$

$$\text{Probabilités élémentaires : } P((P, P)) = P((P, F)) = P((F, P)) = P((F, F)) = \frac{1}{4}$$

A : "1ère pièce : Pile"

B : "2ème pièce : Face"

C : "même résultat pour les deux pièces"

$$P(A) = P(B) = P(C) = \frac{1}{2}$$

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4}$$

On a donc (A, B) , (A, C) , (B, C) indépendants.

$$P(A \cap B \cap C) = P(\emptyset) \neq \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2}$$

Pas indépendants mutuellement.

Problème Soit l'expérience 1 modélisée par $(\Omega_1, \mathcal{A}_1, P_1)$.

Soit l'expérience 2 modélisée par $(\Omega_2, \mathcal{A}_2, P_2)$.

Sous l'hypothèse d'indépendance des deux expériences, modéliser le couple d'expériences.

Les événements relatifs à Exp1 et Exp2 sont dans deux σ -algèbres $(\mathcal{A}_1$ et $\mathcal{A}_2)$ différentes : comment formaliser leur indépendance ?.

1. $\forall E_1 \in \mathcal{A}_1$, on associe à E_1 l'ensemble $E_1 \times \Omega_2$. On obtient ainsi une σ -algèbre de sous-ensembles de $\Omega_1 \times \Omega_2$.

On construit l'espace probabilisé $(\Omega_1 \times \Omega_2, \mathcal{A}'_1, P'_1)$ défini par :

$$\forall E_1 \in \mathcal{A}_1 \quad P(E_1 \times \Omega_2) = P(E_1)$$

Ce modèle est équivalent à $(\Omega_1 \times \Omega_2, \mathcal{A}'_1, P'_1)$.

2. Même chose pour Exp2 : $(\Omega_1 \times \Omega_2, \mathcal{A}'_2, P'_2)$.
3. On complète les éléments de $\mathcal{A}'_1 \cup \mathcal{A}'_2$ en une σ -algèbre sur $\Omega_1 \times \Omega_2$, soit \mathcal{A} .

L'indépendance d'un événement $A'_1 \in \mathcal{A}'_1$ relatif à Exp1 et de $A'_2 \in \mathcal{A}'_2$ relatif à Exp2 s'écrit :

$$\begin{aligned} P(A'_1 \cap A'_2) &= P'_1(A'_1) \cdot P'_2(A'_2) \\ &= P'_1(A'_1 \times \Omega_2) \cdot P'_2(\Omega_1 \times A'_2) \\ &= P_1(A_1) \cdot P_2(A_2) \end{aligned}$$

Exemple Soit l'expérience 1 modélisée par $(\Omega_1, \mathcal{A}_1, P_1)$.

Soit l'expérience 2 modélisée par $(\Omega_2, \mathcal{A}_2, P_2)$.

On prend 2 pièces indépendantes (distinctes).

Exp1 $\Omega_1 = \{P, F\}$
 $\mathcal{A}_1 = \mathcal{P}(\Omega) = \{\emptyset, \{P, F\}, \{P\}, \{F\}\}$
 $P_1 = \{0, 1, \frac{1}{2}, \frac{1}{2}\}$

Exp2 idem

1.

$$\begin{aligned}\Omega &= \Omega_1 \times \Omega_2 = \{(P, P), (P, F), (F, P), (F, F)\} \\ \mathcal{A}'_1 &= \{\emptyset, \{P, F\} \times \{P, F\}, \{P\} \times \{P, F\}, \{F\} \times \{P, F\}\} \\ P' &= 0, 1, \frac{1}{2}, \frac{1}{2}\end{aligned}$$

2. idem

3. on trouve $\mathcal{A} = \mathcal{P}(\Omega_1, \Omega_2)$
on trouve P à partir des probabilités des produits des singletons.
Par exemple :

$$\begin{aligned}P(\{(P, P)\}) &= P(\{(P, P), (P, F)\} \cap \{(P, P), (P, P)\}) \\ &= P'_1(\{P\} \times \Omega_2) \cdot P'_2(\{P\} \times \Omega_1) \\ &= P_1(\{P\}) \cdot P_2(\{P\}) \\ &= \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\end{aligned}$$

7 Probabilités conditionnelles

Def Soit (Ω, \mathcal{A}, P) un espace probabilisé.

$$B \in \mathcal{A} \quad P(B) \neq 0$$

$\forall A \in \mathcal{A}$, on appelle probabilité de A *conditionnellement* à B, ou "probabilité de A si B" l'expression $P(A|B) = \frac{P(A \cap B)}{P(B)}$.

Théorème sous les mêmes hypothèses, A et B sont indépendants si et seulement si $P(A|B) = P(A)$.

Démonstration :

$$\begin{aligned}P(B) &\neq 0 \\ P(A \cap B) &= P(A|B) \cdot P(B) \\ \text{donc} \\ P(A \cap B) &= P(A) \cdot P(B) \Leftrightarrow P(A|B) = P(A)\end{aligned}$$

Exercice On tire deux boules sans remise, d'une boîte contenant 3 boules rouges et une boule bleue.

Quelle est la probabilité de tirer deux boules rouges ? $\frac{1}{2}$

R_1 : La première boule tirée est rouge. $\frac{3}{4}$

R_2 : La deuxième boule tirée est rouge. $\frac{2}{3}$

$$P(R_1 \cap R_2) = P(R_2|R_1) \cdot P(R_1) = \frac{1}{2}$$

Def Système complet d'évènements :

Une famille dénombrable $(A_i)_{i \in I}$ d'éléments de \mathcal{A} tels que :

1. $\forall i \in I \quad P(A_i) \neq 0$
2. $\forall i, j \in I \quad A_i \cap A_j = \emptyset$
3. $\bigcup_{i \in I} A_i = \Omega$

Proposition (Ω, \mathcal{A}, P)

Soit $(A_i)_{i \in I}$ un système complet d'évènements. Alors :

1. $\forall B \in \mathcal{A} \quad P(B) = \sum_{i \in I} P(B \cap A_i)$
2. $\forall B \in \mathcal{A} \quad P(B) = \sum_{i \in I} P(B|A_i) \cdot P(A_i)$

Démonstration :

1. $\forall i, j \in I \quad i \neq j$
 $(B \cap A_i) \cap (B \cap A_j) = B \cap (A_i \cap A_j) = B \cap \emptyset = \emptyset$
 Les $B \cap A_i$ sont disjoints.
 2. $B \subset \Omega$
 $B = B \cap \Omega = B \cap (\bigcup_{i \in I} A_i) = \bigcup_{i \in I} (B \cap A_i)$
1. et 2. $\Rightarrow P(B) = \sum_{i \in I} P(B \cap A_i)$

Exercice (suite) :

- R_1, \bar{R}_1 : système complet d'évènements
 $R_1 \cap \bar{R}_1 = \emptyset$
 $R_1 \cup \bar{R}_1 = \Omega$
 $P(R_2) = P(R_2 \cap R_1) + P(R_2 \cap \bar{R}_1)$
 $P(R_2) = P(R_2|R_1) \cdot P(R_1) + P(R_2|\bar{R}_1) \cdot P(\bar{R}_1)$
 $\cdot = \frac{2}{3} \cdot \frac{3}{4} + 1 \cdot \frac{1}{4} = \frac{3}{4}$

Théorème *Formule de Bayes* (formule concernant la probabilité des causes)

- (Ω, \mathcal{A}, P)
 $(A_i)_{i \in I}$ système complet d'évènements.
 $\forall B \in \mathcal{A}$ tant que $P(B) \neq 0$, on a :

$$P(A_1|B) = \frac{P(B|A_1) \cdot P(A_1)}{\sum_{i \in I} P(B|A_i) \cdot P(A_i)}$$

Démonstration :

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} \\ &= \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i \in I} P(B|A_i) \cdot P(A_i)} \end{aligned}$$

Exercice d'application On a deux urnes : [RRRRB] et [RBB] :

- On choisit au hasard une des deux urnes.
- Dans l'urne ainsi tirée, on tire une boule.
- Cette boule est bleue.

Quelle est la probabilité que l'urne choisie soit la première ?

U_1 : "On a choisi la première urne."

B : "La boule tirée est bleue"

U_1, \bar{U}_1 : système complet d'évènements.

$$\begin{aligned} P(U_1|B) &= \frac{P(B|U_1).P(U_1)}{P(B|U_1).P(U_1) + P(B|\bar{U}_1).P(\bar{U}_1)} \\ &= \frac{\frac{1}{5} \times \frac{1}{2}}{\frac{1}{5} \times \frac{1}{2} + \frac{2}{3} \times \frac{1}{2}} = \frac{3}{13} \end{aligned}$$

Exercice Test biologique de dépistage

S^+ : "positif"

S^- : "négatif"

M^- : "sujet porteur"

M^+ : "sujet non porteur" : \bar{M}^-

$S_e = P(S^+|M^+)$: sensibilité

$S_p = P(S^-|M^-)$: spécificité

$VPP = P(M^+|S^+)$: valeur prédictive positive

$VPN = P(M^-|S^-)$: valeur prédictive négative

Def Soit (Ω, \mathcal{A}, P) un espace probabilisé.

On appelle variable aléatoire réelle (v.a.r.) une application.

$X : \Omega \rightarrow \mathbb{R}$ tel que

$\forall t \in \mathbb{R} \quad \{\omega \in \Omega : X(\omega) \leq t\} \in \mathcal{A}$

$B \subset \mathbb{R}$

$P(X \in B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\})$

Exemple $B = \{x\}$

$P(X = x) = P(X^{-1}(x)) = P(\{\omega \in \Omega : X(\omega) = x\})$

Def On dit que 2 v.a.r. X et Y sur (Ω, \mathcal{A}, P) ont la même loi de probabilité si :

$\forall t \in \mathbb{R} \quad P(X \leq t) = P(Y \leq t)$

Def Soit (Ω, \mathcal{A}, P) et $X : \Omega \rightarrow \mathbb{R}$ une v.a.r..

On appelle fonction de répartition de X la fonction :

$F_x : \mathbb{R} \rightarrow [0; 1] \subset \mathbb{R}$

$\forall t \in \mathbb{R} \quad F_x(t) = P(X \leq t)$

Propriétés de $F_x - F_x$ est positive (≥ 0)

- $\lim_{t \rightarrow -\infty} F_x(t) = 0$

- $\lim_{t \rightarrow +\infty} F_x(t) = 1$

- F_x est continue à droite et admet des limites à gauche (cadlag)

Théorème Soit (Ω, \mathcal{A}, P) un espace probabilisé.

P a la propriété de continuité monotone séquentielle.

$\forall A_1 \subset A_2 \subset A_3 \subset \dots \subset A_i \subset \dots$

Une suite décroissante d'éléments de \mathcal{A} .

$\lim_{i \rightarrow \infty} P(A_i) = P(\cup_{i=1}^{\infty} A_i)$

$$A_1 = \{\omega \in \Omega : X(\omega) \leq -1\} \in \mathcal{A}(xv.a.r.)$$

$$A_2 = \{\omega \in \Omega : X(\omega) \leq -2\} \in \mathcal{A}$$

...

$$A_i = \{\omega \in \Omega : X(\omega) \leq -i\} \in \mathcal{A}$$

$$\lim P(A_i) = P(\bigcap_{i=1}^{\infty} A_i)$$

$$\lim_{t \rightarrow -\infty} F(t) = P(\emptyset) = 0$$

Def (Ω, \mathcal{A}, P)

$$X : \Omega \rightarrow \mathbb{R} \quad \text{v.a.r.}$$

X est une v.a.r. discrète si $X(\Omega)$ est dénombrable (fini au sens dénombrable).

Def x est une variable aléatoire densité s'il existe une fonction $f : \mathbb{R} \mapsto$ intégrable telle que :

$$\forall t \in \mathbb{R} \quad P(X \leq t) = \int_{-\infty}^t f(x) dx$$

$$\text{nb : il est nécessaire que } \int_{-\infty}^{+\infty} f(x) dx = 1$$

Exemple :

$$\text{si } x \leq 0 \quad f(x) = 0$$

$$\text{si } x \geq 0 \quad f(x) = ke^{-kx}$$

$$P(X \leq t) = \int_{-\infty}^t f(x) dx = 0 \text{ si } t \leq 0$$

$$P(X \leq t) = \int_0^t ke^{-kx} dx = 1 - e^{-kt} \text{ sinon}$$

Exemple Variable gaussienne centrée réduite :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$\forall t \in \mathbb{R} \quad P(X = t) = F(t) - \lim_{h \rightarrow 0} (F(t+h) - F(t-h))$$

$$\text{Si } F \text{ est continue } \forall t \in \mathbb{R} \quad P(X = t) = 0$$

Def Espérance d'une v.a.r. discrète

Soit $X : \Omega \rightarrow \mathbb{R}$ v.a.r. discrète (Ω, \mathcal{A}, P) .

$$E(X) = \sum_{x \in X(\Omega)} x.P(X = x) \text{ si absolument convergente.}$$

Exemple $P(X = 0) = \frac{1}{2} \quad P(X = 1) = \frac{1}{2}$

$X(\Omega) = \{0, 1\}$ (ensemble des valeurs possibles). X est une v.a.r. discrète.

$$F_X(t) = P(X \leq t)$$

$$\begin{aligned}
E(X) &= \sum_{x \in \{0,1\}} x.P(X=x) \\
&= 0.P(X=0) + 1.P(X=1) \\
&= 0.\frac{1}{2} + 1.\frac{1}{2} \\
&= \frac{1}{2}
\end{aligned}$$

Proposition $E(X) = \sum_{\omega \in \Omega} X(\omega).P(\{\omega\})$

Def Variance d'une v.a.r. discrète.

Soit $X : \Omega \rightarrow \mathbb{R}$ une v.a.r. discrète telle que $E(X)$ existe.

$$\begin{aligned}
V(X) &= E[(X - E(X))^2] \\
&= \sum_{x \in X(\Omega)} [x - E(x)]^2.P(X=x)
\end{aligned}$$

Proposition $V(X) = E(X^2) - E(X)^2$.

Def Ecart type : $\sigma_x = \sqrt{V(X)}$.

Proposition Soient X, Y deux v.a.d. sur le même (Ω, \mathcal{A}, P) .

Si $X \leq Y$, c'est à dire $\forall \omega \quad X(\omega) \leq Y(\omega)$,

Alors $E(X) \leq E(Y)$.

Théorème Inégalité de Markov

Soit ω une v.a.d. supérieure ou égale à 0, ($\forall \omega \in \Omega \quad X(\omega) \geq 0$)

Alors $\forall \alpha > 0 \quad P(X \geq \alpha) \leq \frac{E(X)}{\alpha}$

Def fonction indicative :

E : ensemble, $A \subset E$

$I_A : E \rightarrow \mathbb{R}$

$\forall x \in E$

$I_A(x) = 1$ si $x \in A$

$I_A(x) = 0$ si $x \notin A$

Application indicatrice d'un évènement.

I_A v.a.d.

$A \in \mathcal{A}$

$I_A : \Omega \rightarrow \mathbb{R}$

$I_A(x) = 1$ si $x \in A$

$I_A(x) = 0$ si $x \notin A$

Démonstration Inégalité de Markov :

X v.a.r. $X \geq 0$

$\forall \alpha > 0 \quad X \geq \alpha, I_{\{X \geq \alpha\}}$ en effet :

ou bien ω tel quel $X(\omega) \geq \alpha \Rightarrow I_{\{X \geq \alpha\}} = 1 \Rightarrow X \geq \alpha$

ou bien ω tel quel $X(\omega) < \alpha \Rightarrow \alpha I_{\{X \geq \alpha\}} = 0$ et $X \geq 0$

$E(X) \geq E(\alpha I_{\{X \geq \alpha\}})$

Proposition $\forall X$ v.a.r. $\forall \lambda \in \mathbb{R}$

$$E(\lambda X) = \lambda E(X)$$

$$E(\lambda I_{\{X \geq \alpha\}}) = \alpha E(I_{\{X \geq \alpha\}})$$

avec

$$E(\lambda I_{\{X \geq \alpha\}}) = \sum_{j \in \{0,1\}} j \cdot P(I_{\{X \geq \alpha\}})$$

$$= 0 \cdot P(I_{\{X \geq \alpha\}}) + 1 \cdot P(I_{\{X \geq \alpha\}})$$

$$= P(X \leq \alpha) + P(X \geq \alpha)$$

$$= P(X \geq \alpha)$$

$$E(\alpha) \geq \alpha P(X \geq \alpha)$$

Théorème Inégalité de Bien-aymé Tchébitchev

Soit X une v.a.r. discrète telle que

$E(X)$ et $V(X)$ existant, alors :

$$\forall \lambda > 0 \quad P(|X| \geq \lambda) \leq \frac{E(X^2)}{\lambda^2}$$

Démonstration $P(|X| \geq \lambda) = P(X^2 \geq \lambda^2) \leq \frac{E(X^2)}{\lambda^2}$

Théorème conséquent avec les mêmes hypothèses :

$$\forall \lambda \geq 0 \quad P(|X - E(X)| \geq \lambda) \leq \frac{V(X)}{\lambda^2}$$

Def Variable aléatoire dégénérée : variable X pour laquelle $V(X) = 0$.

8 Fonction d'une variable ou plusieurs variables aléatoires

Exemple X : nombre de dé tiré.

$$X(\Omega) = \{1, \dots, 6\} \quad \forall i, 1 \leq i \leq 6 \quad P(X = i) = \frac{1}{6}$$

$$Y = X^2 = f(x)$$

$$X : \Omega \rightarrow \mathbb{R} \quad Y : \Omega \rightarrow^X \mathbb{R} \xrightarrow{f} \mathbb{R}$$

En fait il faudrait plutôt écrire foX que $f(x)$ car X est une v.a. c'est à dire une application (petite subtilité).

$$Y(\Omega) = \{1, 4, 9, 16, 25, 36\}$$

$$P(Y = 1) = P(Y = 2) = \dots = \frac{1}{6}$$

Exemple $Z = |X - 3| \quad X : \Omega \rightarrow \mathbb{R}$.

$$\forall \omega \quad Z(\omega) = |X(\omega) - 3| \quad g(x) = |x - 3|$$

$$Z = goX$$

$$Z(\omega) = \{0, 1, 2, 3\}$$

$$Z(\omega) = 0 \Leftrightarrow X(\omega) = 3$$

$$\begin{aligned}
\Rightarrow \{\omega|Z(\omega) = 0\} &= \{\omega|X(\omega) = 3\} \\
P(\{\omega|Z(\omega) = 0\}) &= P(\{\omega|X(\omega) = 3\}) \\
P(Z = 0) &= P(X = 3) \\
P(Z = 1) &=? \\
\{\omega|Z(\omega) = 1\} &= \{\omega|X(\omega) = 2\} \cup \{\omega|X(\omega) = 4\} \\
\text{de plus, } \{\omega|X(\omega) = 2\} \cap \{\omega|X(\omega) = 4\} &= \emptyset \\
\text{donc } P(Z(\omega) = 1) &= P(X(\omega) = 2) + P(X(\omega) = 4) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.
\end{aligned}$$

Exemple X v.a. $X(\Omega) = \{0, 1\}$ $P(X = 0) = P(X = 1) = \frac{1}{2}$
 Y v.a. $Y(\Omega) = \{0, 1\}$ $P(Y = 0) = P(Y = 1) = \frac{1}{2}$
Si X, Y indépendants
 $Z = X + Y \quad \forall \omega \in \Omega \quad Z(\omega) = X(\omega) + Y(\omega)$
 $Z(\Omega) = \{0, 1, 2\}$

$$\begin{aligned}
P(Z(\omega) = 0) &= P(X(\omega) = 0 \cap Y(\omega) = 0) = P(X(\omega) = 0) \cdot P(Y(\omega) = 0) = \frac{1}{4} \\
\{\omega|Z(\omega) = 1\} &= (\{\omega|X(\omega) = 0\} \cap \{\omega|Y(\omega) = 1\}) \\
&\quad + (\{\omega|X(\omega) = 1\} \cap \{\omega|Y(\omega) = 0\}) \\
P(Z(\omega) = 1) &= P(X(\omega) = 0) \cdot P(Y(\omega) = 1) + P(X(\omega) = 1) \cdot P(Y(\omega) = 0) \\
&= \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}
\end{aligned}$$

Proposition Soit X_1, X_2, \dots, X_n des v.a.r.d. définies sur son même espace probabilisé (Ω, \mathcal{A}, P) et $f : \mathbb{R}^n \rightarrow \mathbb{R}$.

La fonction $Y : \Omega \rightarrow \mathbb{R}$ définie par $\forall \omega \in \Omega \quad Y(\omega) = f(X_1(\omega) \dots X_n(\omega))$.

Ceci est une v.a.d. sur (Ω, \mathcal{A}, P) .

On dit que Y est fonction des n v.a. X_1, \dots, X_n .

Remarque $\forall y \in Y(\Omega)$

$$Y^{-1}(y) = \bigcup_{(x_1, \dots, x_n) \in f^{-1}(y)} \{\omega|X_1(\omega) = x_1\} \cap \dots \cap \{\omega|X_n(\omega) = x_n\}$$

L'union se fait sur une famille dénombrable puisque

$f^{-1}(y) \subset X_1(\Omega) \times \dots \times X_n(\Omega)$, qui est dénombrable car produit fini d'ensemble dénombrable.

Proposition 1. Soit X une v.a.r.d.

$$\forall \lambda \in \mathbb{R} \quad E(\lambda X) = \lambda E(X)$$

2. Soit X et Y deux v.a.r.d.

$$E(X + Y) = E(X) + E(Y)$$

Démonstration On suppose Ω dénombrable.

$$\begin{aligned}
 E(\lambda X) &= \sum_{\omega \in \Omega} \lambda X(\omega) P(\omega) \\
 &= \lambda \sum_{\omega \in \Omega} X(\omega) P(\omega) \\
 &= \lambda(E(X)) \\
 E(X + Y) &= \sum_{\omega \in \Omega} (X + Y)(\omega) P(\omega) \\
 &= \sum_{\omega \in \Omega} X(\omega) P(\omega) + \sum_{\omega \in \Omega} Y(\omega) P(\omega) \\
 &= E(X) + E(Y)
 \end{aligned}$$

Def v.a.d. indépendantes :

Les v.a.r.d. X_1, \dots, X_n sont indépendantes si :

$\forall t_1, \dots, t_n \in \mathbb{R}$ les évènements $(X_1 = t_1), \dots, (X_n = t_n)$ sont mutuellement indépendantes, c'est à dire tels que :

$$P(X_1 = t_1 \cap \dots \cap X_n = t_n) = \prod_{i=1}^n P(X_i = t_i)$$

Théorème L'espérance du produit de 2 v.a.d. indépendantes

X, Y v.a.d. indépendantes

Alors $E(XY) = E(X)E(Y)$.

Conséquence du théorème Variable de la somme de 2 v.a.d. indépendantes :

X, Y v.a.d. indépendantes.

$V(X + Y) = V(X) + V(Y)$

Démonstration

$$\begin{aligned}
 V(X + Y) &= E[(X + Y)^2] - (E(X + Y))^2 \\
 &= E(X^2 + 2XY + Y^2) - (E(X) + E(Y))^2 \\
 &= E(X^2) + 2E(XY) + E(Y^2) - E^2(X) + 2E(X)E(Y) - E^2(Y) \\
 &= E(X^2) + E(Y^2) - E^2(X) - E^2(Y) \\
 &= E(X^2) - E^2(X) + E(Y^2) - E^2(Y) \\
 &= V(X) + V(Y)
 \end{aligned}$$

Proposition Soient X_1, \dots, X_n indépendants.

$$V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n)$$

$$V\left(\sum X_i\right) = \sum V(X_i)$$

Conséquence Espérance et variance de la moyenne de v.a.d. i.i.d. (indépendantes identiquement distribuées) X_1, \dots, X_n des v.a.d. i.i.d. c'est à dire de même loi que X .

On suppose l'existence de $E(X)$ et de $V(X)$.

On définit $M = \frac{1}{n} \sum_{i=1}^n X_i$ alors $E(M) = E(X)$ et $V(M) = \frac{V(X)}{n}$.

Démonstration

$$\begin{aligned} E(M) &= E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} E\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n} \sum_{i=1}^n nE(X_i) \\ &= \frac{nE(X)}{n} \\ &= E(X) \end{aligned}$$

$$\begin{aligned} V(M) &= V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} V\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n V(X_i) \\ &= \frac{nV(X)}{n^2} \\ &= \frac{V(X)}{n} \end{aligned}$$

Théorème Loi faible des grands nombres.

Soit X_1, X_2, \dots une suite infinie de v.a. indépendantes et de même loi, admettant une espérance et une variance.

Soit M_1, M_2, \dots la suite des v.a. égales, pour chaque n , à la moyenne des n premiers X_i :

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

Alors la suite des M_n converge en probabilité vers la v.a. dégénérée égales à $E(X)$ c'est à dire :

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|M_n - E(X)| \geq \varepsilon) = 0$$

Démonstration :

$$\forall n \quad M_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$E(M_n) = E(X)$$

Comme les X_i sont i.i.d., $V(M_n) = \frac{V(X)}{n}$

$$P(|M_n - E(M_n)| \varepsilon) \leq \frac{V(M_n)}{\varepsilon^2} \text{ I.B.T.}$$

$$P(|M_n - E(M_n)| \varepsilon) \leq \frac{V(M_n)}{n\varepsilon^2} \longrightarrow \lim_{n \rightarrow \infty} \frac{V(X)}{n\varepsilon^2} = 0$$

Cas particuliers Soit A un évènement tel que $P(A) = p$

X : v.a. indicatrice de A , $X(\omega) = 1$ si $\omega \in A$, $X(\omega) = 0$ si $\omega \notin A$.

X_1, X_2, \dots i.i.d. comme X .

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

$$E(M_n) = E(X) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) = 0 + p = p$$

$$\text{LFGN : } \forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|M_n - p| \varepsilon) = 0$$

9 V.a.r.d. particulières

9.1 Variable de Bernoulli

Expérience type : "Pile ou face". p , $1-p$.

Def On appelle variable de Bernoulli une v.d. X telle que :

$$V(\Omega) = \{0, 1\} \quad P(X = 1) = p \quad P(X = 0) = 1 - p = q$$

Paramètres : p

$$E(X) = p \quad V(X) = p(1 - p) = pq, \text{ car } E(X^2) - p^2 = 0^2P(X = 0) + 1^2P(X = 1) - p^2 = 0 + p - p^2$$

nb : $V(X)$ maximale pour $p = \frac{1}{2}$.

9.2 Variable Binomiale

Expérience type : n tirages à pile ou face (indépendant) et on compte le nombre de pile.

Def une variable binomiale de loi $\mathcal{B}(n, p)$ - n : nombre de tirages, p : probabilité - est une v.a. somme de n variables de Bernoulli indépendantes et de même paramètre p .

Proposition loi d'une variable binomiale

$$X \text{ suit une loi binomiale } \mathcal{B}(n, p) \quad X \rightsquigarrow \mathcal{B}(n, p)$$

$$\text{Alors } X(\Omega) = \{0, \dots, n\}$$

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}$$

Démonstration Une suite particulière de tirage où $X = k$ a pour probabilité $p^k(1 - p)^{n-k}$.

$$E(X) = np \quad V(X) = np(1 - p)$$

9.3 Loi discrète uniforme

Expérience type : dé.

Def On dit qu'on a une loi discrète uniforme sur $[a, b] \subset \mathbb{N}$ si $\forall k \in [a, b] \subset$

$$\mathbb{N} \quad P(X = k) = \frac{1}{b-a+1} = \frac{1}{n}$$

$$E(X) = \frac{a+b}{2} \quad V(X) = \frac{n^2-1}{12}$$

9.4 Variable géométrique

Expérience type : on effectue des tirages de Bernoulli indépendantes de même paramètres jusqu'à succès.

X : nombre de tirages.

$$P(X = k) = (1 - p)^{k-1} p$$

$$E(X) = \frac{1}{p} \quad V(X) = \frac{q}{p^2}$$

9.5 Variable binomiale négative

Expérience type : on fait des tirages de Bernoulli indépendants de mêmes paramètres jusqu'à r succès.

$$\begin{aligned} P(X = k) &= C_{k-1}^{r-1} p^r (1-p)^{k-r} \\ E(X) &= \frac{rq}{p} \end{aligned}$$

9.6 Variable de Poisson

Expérience type : comptage d'évènements rares ou de fréquence proportionnelle à une mesure de temps et d'espace.

limite de $\mathcal{B}(n, p)$ quand n tend vers l'infini, np tend vers une limite finie.

Def variable de Poisson de paramètre λ une variable discrète X :

$$\begin{aligned} X &\rightsquigarrow \mathcal{P}(\lambda) & X(\Omega) &= \mathbb{N} \\ \forall k \in \mathbb{N} & & P(X = k) &= e^{-\lambda} \frac{\lambda^k}{k!} \end{aligned}$$

$$E(X) = V(X) = \lambda$$

9.7 Variable Gaussienne

Loi de densité : $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

$$X \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{+\infty} x f(x) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} x e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \mu \\ V(X) &= E(X^2) - E^2(X) \\ &= E(X^2) - \mu^2 \\ &= \sigma^2 \end{aligned}$$

Importance de la loi de Gauss : Théorème central limite.

Def v.a. X centrée : $E(X) = 0$, réduite : $V(X) = 1$.

v.a.r. gaussienne centrée réduite : $X \rightsquigarrow \mathcal{N}(0, 1)$ de densité : $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

9.8 Quelques relations limites

Théorème de Moivre-Laplace convergence d'une loi Binomiale vers loi normale.

Soit X_1, X_2, \dots, X_n une suite de v.a.d. de Bernoulli indépendantes et de

même espérance p .

On pose :

$$\forall n \in \mathbb{N}^* \quad S_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

$$S_n^* = \frac{S_n + E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - np}{\sqrt{npq}}$$

$$\text{Alors } \forall t \in \mathbb{R} \quad \lim P(S_n^* < t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{x^2}{2}} dx$$

Règle pratique : En statistique on approche $\mathcal{B}(n, p)$ pour $\mathcal{N}(np, npq)$ quand
 $n \geq 100 \quad np \geq 10 \quad nq \geq 10$

Théorème de Poisson Convergence d'une loi Binomiale vers une loi de Poisson.

Soit p_1, p_2, \dots, p_n une suite de réels de $[0, 1]$.

Si $\exists \lambda \quad \lim_{n \rightarrow \infty} np = \lambda$ alors

$$\forall k \in \mathbb{N} \lim_{n \rightarrow \infty} C_n^k p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

Règle pratique : On approche $\mathcal{B}(n, p)$ par $\mathcal{P}(np)$; quand $n \geq 100, np \leq 10$.

Def Soient X, Y deux v.a.r.

Leur covariance est :

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))]$$

Leur corrélation est :

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

$$(X(\omega) = x) = \{\omega(X(\omega)) = x\} = \{\omega(X(\omega))^2 = x^2\} = \{\omega(Y(\omega)) = x^2\}$$

$$\forall x \in \mathbb{R} \quad X(\omega) = x \Leftrightarrow X^2(\omega) = x^2 \text{ car } X \geq 0$$

$$P(Y(\omega) = x^2) = P(X(\omega) = x).$$

Proposition Si X et Y sont indépendantes

$$Cov(X, Y) = 0$$

$$Corr(X, Y) = 0$$

Démonstration

$$\begin{aligned} E[(X - E(X))(Y - E(Y))] &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(Y)E(X) - E(X)E(Y) + E(X)E(Y) \\ &= E(XY) - E(X)E(Y) \\ &= 0 \end{aligned}$$

nb : la réciproque est fausse.

Deuxième partie

Information, compression

10 Entropie d'une loi finie

Def X v.a telle que $X(\Omega)$ fini.

On appelle *entropie de X* l'expression :

$$\begin{aligned} H(X) &= - \sum_{i \in X(\Omega)} P_i \log(P_i) \\ &= \sum_{i \in X(\Omega)} P_i \log\left(\frac{1}{P_i}\right) \end{aligned}$$

où $P_i = P(X = i)$,

avec la convention $P_i = 0 \Rightarrow P_i \log(P_i) = 0$.

Unité de mesure : si $a = 2$: 1 bit.

11 Interprétation

incertitude d'un évènement $A \in \mathcal{A}$ (Ω, \mathcal{A}, P)

$$I(A) = \log\left(\frac{1}{P_A}\right)$$

- Si $P(A) = 1$ $I(A) = 0$
- Si $P(A) = \epsilon$ petit $I(A)$ grande
- A, B indépendant :

$$\begin{aligned} I(A \cap B) &= \log \frac{A}{P(A \cap B)} \\ &= \log\left(\frac{1}{P(A)} \times \frac{A}{P(B)}\right) \\ &= I(A) + I(B) \end{aligned}$$

- X v.a.

$$\begin{aligned} I(X(\omega)) &= \log \frac{1}{P[X(\omega)]} \\ H(X) &= E(I(X)) \end{aligned}$$

Exemple Variable de Bernoulli de paramètre p :

$$X : X(\Omega) = \{0, 1\}$$

$$P(X = 1) = p$$

$$P(X = 0) = 1 - p$$

$$H(X) = -(1-p) \log_2(1-p) - p \log_2 p$$

$$\begin{aligned} p &= \frac{1}{2} \\ H &= -\frac{1}{2} \log_2\left(\frac{1}{2}\right) - \frac{1}{2} \log_2\left(\frac{1}{2}\right) \\ &= -\log_2\left(\frac{1}{2}\right) \\ &= \log_2(2) = 1 \end{aligned}$$

Proposition X v.a. telle que $X(\Omega) = S$ fini

$H(X)$ est nominale quand la loi de X est uniforme sur S :

$$\forall x \in S \quad P(X = x) = \frac{1}{\text{Card}(S)}$$

et dans ce cas $H(X) = \log(\text{Card}(S))$.

Démonstration Multiplicateurs de Lagrange (maximum sans contrainte).

Def entropie relative

$$\text{Card}(X(\Omega)) = N$$

$$h(X) = \frac{H(X)}{\log(N)} \leq 1$$

12 Compression

12.1 Sources discrètes

Def Source :

1. Un ensemble $S = \{m_1, m_2, \dots, m_n\}$ de messages.
2. Un mécanisme de génération aléatoire de suites de messages.
Les messages sont émis aux instants successifs $0, 1, 2, \dots$ repérés par les éléments de \mathbb{N} .
Donc on considère une suite X_0, X_1, \dots, X_n de v.a. à valeurs dans S .

12.2 Cas particuliers

Source simple les X_k sont indépendants et de même loi.

Source Markovienne $\forall n, k \in \mathbb{N} \quad \forall x_n, x_{n-1}, \dots, x_{n-k} \in S$
 $P(X_n | (X_{n-1}) \cap \dots \cap (X_{n-k} = x_{n-k})) = P(X_n = x_n | X_{n-1} = x_{n-1})$
(markovienne d'ordre 1).

Source markovienne stationnaire $\forall n \geq 0 \quad \forall m, m' \in S$
 $P(X_n = m | X_{n-1} = m') = P(X_1 = m | X_0 = m')$

12.3 Théorème de McMillan (cas d'une source simple)

Problème de la compression Ensemble S de symboles, $Card(S) = N$
 X_k sont indépendants de même loi.
 $\forall k \quad X_k(\Omega) = S$

- Si chaque X_k est équidistribué, alors l'entropie de la suite X_1, \dots, X_n est la somme $H(X_1) + H(X_2) + \dots + H(X_n) = nH(X)$ maximum.
- Si la loi de X n'est pas uniforme

Théorème de Mc Millan Soit X_1, \dots, X_n, \dots une suite de v.a. à valeur dans un ensemble F fini à N éléments, les X_i étant indépendants et de même loi qu'une v.a. X d'entropie $H(X)$. Alors :

$$\forall \varepsilon > 0 \quad \sigma > 0 \quad \exists n_0 \in \mathbb{N} \text{ tel que } \forall n > n_0 \quad \exists C \in \mathcal{P}(F^n)$$

tel que

1. $P(X) \geq 1 - \varepsilon$
2. $\forall (x_1, \dots, x_n) \in C,$
 $2^{-n(H+\sigma)} \leq P(x_1, \dots, x_n) \leq 2^{-n(H-\sigma)}$
3. $(1 - \varepsilon) \cdot 2^{n(H-\sigma)} \leq Card C \leq 2^{n(H+\sigma)}$

1. La plupart des suites émises par la source sont dans C .
2. Les suites qui sont dans C sont approximativement équidistribuées (car $2^{n(H-\sigma)}$ est proche de $2^{n(H+\sigma)}$)

3. nombre d'éléments de C

Démonstration base : 2

$\forall i, j$, l'incertitude lancée par l'observation que X_j prenne la valeur x_i est :

$$I(x_i) = \log\left(\frac{1}{P(X_j = x_i)}\right)$$

donc $P(X_j = i) = 2^{-I(x_i)}$

On définit la v.a. $I(X) : \omega \mapsto I(X(\omega)) = \log \frac{1}{P(X=\omega)}$

1. Soit la suite des v.a. $I(X_k)$ et S_n la suite des v.a. $S_n = \sum_{k=1}^n I(X_k)$
 $E(I(X)) = H(X)$ donc (L.F.G.N.) $\frac{S_n}{n}$ converge en probabilité vers

$$H(X) : \forall \varepsilon > 0 \quad \forall \sigma > 0 \quad \exists n_0 : \forall n > n_0$$

$$P\left(\left|\frac{S_n}{n} - H\right| \leq \sigma\right) \geq 1 - \varepsilon$$

On prend $n = n_0$ et on définit C comme l'ensemble des suites (x_1, \dots, x_n) qui vérifient

$$n(H - \sigma) \leq \sum_{i=1}^n I(x_i) \leq n(H + \sigma)$$

Donc $P(X) \geq 1 - \varepsilon$.

2. $\forall (x_1, \dots, x_n) \in F^n$

$$P(X_1, \dots, X_n) = (x_1, \dots, x_n) = \prod_{i=1}^n P(X = x_i)$$

$$= \prod_{i=1}^n 2^{-I(x_i)}$$

$$= 2^{-\sum_{i=1}^n I(x_i)}$$

Donc 2. $\Rightarrow \forall (x_1, \dots, x_n) \in C$,
 $P(x_1, \dots, x_n) \leq 2^{-n(H-\sigma)}$

3. Soit $\nu = \text{Card}C$
 $\nu \cdot 2^{-n(H+\sigma)} \leq P(C) \leq \nu \cdot 2^{-n(H-\sigma)}$

3.
 $\text{Rightarrow} \quad 1 - \varepsilon \leq P(C) \leq 1$
donc $\nu \cdot 2^{-n(H+\sigma)} \leq 1$ et $1 - \varepsilon \leq \nu \cdot 2^{-n(H-\sigma)}$
d'où $(1 - \varepsilon) \cdot 2^{n(H-\sigma)} \leq \nu \leq 2^{n(H+\sigma)}$

Conséquence : Si n_1 est le plus petit entier $\geq n(H + \sigma)$, alors $\nu \leq 2^{n_1}$.
On peut coder C avec n_1 bits.

Algorithme de compression déduit du théorème de MacMillan Soit n_2 le plus petit entier tel que $\nu + 1 < 2^{n_2}$.
 Les 2^{n_2} suites de n_2 bits permettent de coder C en générant une suite pour coder un caractère d'échappement.

Soit une suite de n_0 messages :

- Si la suite est dans C, on la code par n_2 bits qui constituent le numéro de cette suite au sein de C.
- Sinon on écrit les n_2 bits de la suite d'échappement suivis de la transcription binaire des n_0 messages originaux.

L : longueur du message une fois codé.

$$E(L) : (1 - \bar{\epsilon})n_2 + \bar{\epsilon} \cdot (n_2 + n_0 \log_2 N) = n_2 + \bar{\epsilon} \cdot n_0 \log_2 N$$

le taux de compression peut être rendu aussi proche que l'on veut de $\frac{H}{\log_2 N}$

$$\frac{n_2}{n \cdot \log_2 N} + \bar{\epsilon} \sim \frac{n_0 H}{n_0 \log_2 N}$$

nb : Si les X_i sont équidistribués, $H(X) = \log_2 N$.

12.4 Entropie conditionnelle

X v.a. sur (Ω, \mathcal{A}, P) à valeurs dans F fini, $A \subset \infty$

Def de la v.a. $(X|A)$ de loi

$$\forall x \in F \quad P((X|A) = x) = P(X = x|A) = \frac{P((X = x) \cap A)}{P(A)}$$

L'entropie de cette v.a. peut être calculée.

Soit Y une autre v.a. sur (Ω, \mathcal{A}, P) à valeurs dans F .

On pose $\forall j \in F \quad B_j = Y^{-1}(\{j\})$

Donc $B_j = (Y = j)$.

$\forall j$, l'entropie de X conditionnelle à $Y = j$ est :

$$- \sum_{i \in X(\Omega)} P(X = i|Y = j) \log P(X = i|Y = j)$$

Pour apprécier globalement l'influence de Y sur l'info apportée par X on prend l'espérance.

Def Entropie conditionnelle de X conditionnelle à Y .

$$H_Y(X) = - \sum_{j \in F} P(Y = j) \sum_{i \in F} P(X = i | Y = j) \log P(X = i | Y = j)$$

Proposition Soient X, Y v.a. sur (Ω, \mathcal{A}, P) à valeurs dans F fini. On appelle (X, Y) la v.a. (appelée conjointe) à valeurs dans $F \times F$ telle que :

$$P((X, Y) = (i, j)) = P(X = i \cap Y = j) = p = j$$

Donc $H(X, Y) = - \sum_i - \sum_j P_{ij} \log P_{ij}$.

1. (a) $H(X, Y) = H(Y) + H_Y(X)$
- (b) $H(X) \leq H(X, Y)$
- (c) $H_Y(X) \leq H(X)$
- (d) $H(X, Y) \leq H(X) + H(Y)$
2. Si X et Y indépendants :
 - (a) $H_Y = H(X)$
 - (b) $H(X, Y) = H(X) + H(Y)$

Troisième partie

Chaines de Markov

nb : On utilisera ici les notations transposées.

Def Chaîne de Markov :

Suite de v.a. $X_0, X_1, \dots, X_k, \dots$ à valeurs dans un ensemble S fini tels que :

$$\forall n, k \in \mathbb{N}$$

$$\forall x_n, x_{n-1}, \dots, x_{n-k} \in S$$

$$P(X_n = x_n | X_{n-1} = x_{n-1} \dots \cap X_{n-k} = x_{n-k}) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

et $\forall n \geq 0$,

$$\text{forall } m, m' \in S \quad P(X = m | X_{n-1} = m') = P(X_1 = m | X_0 = m')$$

$F = \{1, \dots, q\}$: espace des états.

On note :

$$\forall i, j \in F \quad P_{ij} = P(X_n = i | X_{n-1} = j)$$

Def On appelle matrice de transition M de la chaîne de Markov la matrice des P_{ij} :

$$M \in M_q(\mathbb{R}) \quad M = (m_j^i) \\ \forall i, j \quad m_j = P_{ij} = P(X_n = i | X_{n-1} = j)$$

Proposition Soit V_{n-1} le vecteur colonne de coefficients $(v_{n-1})^i$ et V_n le vecteur colonne de coefficient $(v_n)^i$ tel que :

$$\forall i \quad (v_{n-1})^i = P(X_{n-1} = i) \quad (v_n)^i = P(X_n = i)$$

Alors la loi de X_n telle que $V_n = MV_{n-1} = MMV_{n-2}$.

Démonstration

$$\forall i \in F \quad v_n^i = P(X_n = i) \\ = \sum_{j=1}^q P(X_n = i \cap X_{n-1} = j) \\ = \sum_{j=1}^q P(X_n = i \cap X_{n-1} = j) \cdot P(X_{n-1} = j) \\ = \sum_{j=1}^q m_j^i v_{n-1}^j$$

Def On appelle matrice stockastique une matrice carrée (m_j^i) telle que :

1. $\forall i, j \in \{1, \dots, q\} \quad 0 \leq m_j^i \leq 1$
2. $\forall j \in \{1, \dots, q\} \quad \sum_{i=1}^q m_j^i = 1$

On appelle vecteur de probabilité un vecteur colonne (v^i) tel que :

1. $\forall i \quad 0 \leq v^i \leq 1$
2. $\sum_{i=1}^q v^i = 1$

Proposition Pour qu'un vecteur colonne de $M_1^q(\mathbb{R})$ représente une loi de probabilité sur F , il faut et il suffit qu'il soit stockastique.

Une matrice $M \in M_q(\mathbb{R})$ représente des $P(X = i | Y = j)$ pour deux v.a. X, Y à valeurs dans F si et seulement si elle est stockastique.

Def On appelle chaîne de Markov homogène d'espace d'états $F = \{1, \dots, q\}$ de matrice de transition $M \in M_q(\mathbb{R})$ et de distribution initiale $V_0 = (v_0^i)_{i \in F} = (P(X_0 = i))_{i \in F}$

Une suite de v.a. $X_0, X_1, \dots, X_n, \dots$ à valeurs dans F telle que $\forall n$ la loi de X_n est représentée par $\forall n = M^n V_0$.

Représentation graphique Un graphe :

- Un sommet par élément de F

– Une flèche de j vers i si $m_j^i \neq 0$

Def On dit qu'une chaîne de Markov est irréductible si un graphe est fortement connexe, c'est à dire si $\forall i, j \in S$, il existe un chemin orienté allant de i à j , et un chemin orienté allant de j à i .

Ceci équivaut à : $\forall i, j \in S \quad \exists n \in \mathbb{N}^*$ tel que, en posant $(n_j^i) = N = M^n$:

$$n_j^i \neq 0$$

Def Quelque soit un état i de S , on appelle période $d(i)$ de cet état

$$d(i) = PGCD(\{n \in \mathbb{N}^* : (M^n)_i^i \neq 0\})$$

Si $d(i) = 1$, on dit que i est apériodique.

Def Une chaîne de Markov est apériodique si ses états sont tous apériodiques.

Théorème 1. Soit une CM homogène apériodique d'espace d'état $S = \{s_1, \dots, s_k\}$ et de matrice de transition M .

$$\exists N \in \mathbb{N} \text{ tel que } \forall i \in [1, k], \forall n \geq N \quad (M^n)_j^i > 0$$

2. Soit une CMH apériodique et irréductible

$$\exists N \in \mathbb{N} \text{ tel que } n > N \Rightarrow \forall i, j (M^n)_j^i > 0$$

Def On dit qu'une CMH de matrice de transition M est régulière (donc réversible) si $\forall N \in \mathbb{N}$ tel que $n \geq N \Rightarrow \forall i, j \quad (M^n)_j^i > 0$

Théorème ergodique Soit une CMH régulière, de matrice de transition M , alors :

1. Il existe un vecteur de probabilité V telle que $MV = V$

Ce vecteur est unique et pour tout vecteur de probabilité w , $\lim_{n \rightarrow \infty} (M^n)_j^i w = v^i$

2. $\forall j$, on a $\lim_{n \rightarrow \infty} (M^n)_j^i = v^i$, c'est à dire la suite $(M^n)_{n \in \mathbb{N}}$ converge vers une limite \tilde{M} dont les colonnes sont égales entre elles.

Notation

$$\sum_i a_i = \sum_i^+ a_i + \sum_i^0 a_i + \sum_i^- a_i$$

Def Soit μ_1, μ_2 deux vecteurs de probabilité,

$$\begin{aligned} d(\mu_1, \mu_2) &= \frac{1}{2} \sum_i |\mu_1^i - \mu_2^i| \\ &= \frac{1}{2} \sum_i^+ (\mu_1^i - \mu_2^i) + \frac{1}{2} \sum_i^- (\mu_2^i - \mu_1^i) \\ &= \sum_i^+ (\mu_1^i - \mu_2^i) \end{aligned}$$

Lemme

$$\begin{aligned}
0 &= \sum_i \mu_1^i - \sum_i \mu_2^i \\
&= \sum_i (\mu_1^i - \mu_2^i) \\
&= \sum_i^+ (\mu_1^i - \mu_2^i) + \sum_i^0 + \sum_i^- \\
&= \sum_i^+ (\mu_1^i - \mu_2^i) - \sum_i^+ (\mu_2^i - \mu_1^i) \\
\text{Donc } \sum_i^+ (\mu_1^i - \mu_2^i) &= \sum_i^+ (\mu_2^i - \mu_1^i)
\end{aligned}$$

nb : $d(\mu_1, \mu_2) = \frac{1}{2} \sum |\mu_1^i - \mu_2^i| \leq \frac{1}{2} (\sum |\mu_1^i| + \sum |\mu_2^i|) = 1$

Lemme Q matrice stockastique, μ_1, μ_2 vecteurs de probabilité.

1. $d(Q\mu_1, Q\mu_2) \leq d(\mu_1, \mu_2)$
2. ($\exists \alpha \in \mathbb{R}$ tq $\forall i, j \quad q_{i,j} \geq \alpha$)
 $\Rightarrow d(Q\mu_1, Q\mu_2) \leq (1 - \alpha)d(\mu_1, \mu_2)$

Demonstration du lemme

$$\begin{aligned}
\forall i \quad (Q\mu_1)^i &= \sum_j q_j^i \mu_1^j \\
(Q\mu_2)^i &= \sum_j q_j^i \mu_2^j \\
d(Q\mu_1, Q\mu_2) &= \sum_i^+ ((Q\mu_1)^i - (Q\mu_2)^i) \\
&= \sum_i^+ \sum_j q_j^i (\mu_1^j - \mu_2^j) \\
&\leq \sum_i^+ \sum_j^+ q_j^i (\mu_1^j - \mu_2^j) = \sum_i^+ (\mu_1^i - \mu_2^i) \sum_i^+ q_j^i \\
&\leq d(\mu_1, \mu_2)
\end{aligned}$$

$$\sum_i \sum_j q_j^i \mu_1^j = \sum_i \sum_j q_j^i \mu_2^j \Rightarrow \sum_i \sum_j q_j^i (\mu_1^j - \mu_2^j) = 0$$

donc si $\sum_i^+ \sum_j q_j^i (\mu_1^j - \mu_2^j)$ n'est pas nulle, la somme \sum^+ se fait sur strictement moins de termes que dans $\sum_i \sum_j$.

idem sur $\sum^+ q_j^i$. Si $\exists \alpha > 0$ tel que $\forall i \quad \alpha < q_j^i$, alors $\sum^+ q_j^i < 1 - \alpha$ puisque $\sum_i q_j^i = 1$.

Démonstration du théorème 1. Soit μ_0 un vecteur de probabilité.

On pose $\mu_n = M^n \mu_0$

(a) la suite $(\mu_n)_{n \in \mathbb{N}}$ est de Cauchy.

$\forall \varepsilon > 0 \quad N \in \mathbb{N}$ tels que $\forall p, n \in \mathbb{N}$

$n > N \Rightarrow d(\mu_n, \mu_{n+p}) < \varepsilon.$

$$\begin{aligned} \forall v, p \in \mathbb{N} \quad d(\mu_n, \mu_{n+p}) &= d(M_{\mu_0}^m, M_{\mu_0}^{m+p}) \\ &\leq (1 - \alpha) d(M^{n-s_0} \mu_0, M^{n+p-s_0} \mu_0) \\ &\leq (1 - \alpha)^k d(M^{n-ks_0} \mu_0, M^{n+p-ks_0} \mu_0) \\ &\leq \dots \\ &\leq (1 - \alpha)^k d(\dots) \end{aligned}$$

nb : $d(\dots) \leq 1$

Il suffit de prendre k tel que $(1 - \alpha)^n < \varepsilon$, puis n tel que $n \geq ks_0$ pour obtenir $d(\mu_n, \mu_{n+p}) < \varepsilon.$

(b) (\mathbb{R}^2, d) espace métrique complet, Cauchy \Rightarrow converge.

$\forall \mu_i$ la suite des $\mu_n = M^n \mu_0$ converge vers une limite $\tilde{\mu}_0.$

Cette limite $\tilde{\mu}_0$ est la même que celle de la suite des $M \mu_0$, qui est égale à $M \tilde{\mu}_0$ par continuité du produit matriciel.

Donc $M \tilde{\mu}_0 = \tilde{\mu}_0$

Soit π_1, π_2 deux vecteurs de probabilité tels que :

$M \pi_1 = \pi_1$ et $M \pi_2 = \pi_2$

$$\begin{aligned} d(\pi_1, \pi_2) &= d(M^{s_0} \pi_1, M^{s_0} \pi_2) \\ &\leq (1 - \alpha) d(\pi_1, \pi_2) \\ &\Rightarrow d(\pi_1, \pi_2) = 0 \\ &\Rightarrow \pi_1 = \pi_2 \end{aligned}$$

2. $\forall j$, on considère μ tel que $\mu^1 = 1$ et $i \neq j \Rightarrow \mu^i = 0$

$\lim_{n \rightarrow \infty} M^n \mu = \pi$, car indépendance de μ donc de $j.$

$\forall j \quad \lim_{n \rightarrow \infty} P(X^n = i | X_0 = j) = \pi^i$ (indépendance de j).

\rightarrow la matrice $\tilde{M} = \lim_{n \rightarrow \infty} M^n$ dont chaque élément \tilde{m}_j^i est égal à $\lim_{n \rightarrow \infty} P(X^n = i | X_0 = j) = \pi^i$ est constitué de colonnes toutes égales à $\pi.$

$$\tilde{M} = (\pi \pi \dots \pi)$$

Conséquence du théorème ergodique Loi faible des grands nombre pour les chaines de Markov homogènes ergodiques (régulières).

X_k^i les v.a. : $X_k^i = 1$ si $X_k = i$, $X_k^i = 0$ sinon.

$$S_n^i = \sum_{k=1}^n X_k^i$$

X_k^{ij} les v.a. : $X_k^{ij} = 1$ si $X_k = i$ et $X_{k-1} = j$, $X_k^{ij} = 0$ sinon.

$$S_n^{ij} = \sum_{k=1}^n X_k^{ij}$$

$$\forall \varepsilon > 0 \quad \lim_{n \rightarrow \infty} P(|S_n^i - l^i| \geq \varepsilon) = 0$$

$$\lim_{n \rightarrow \infty} P(|S_n^{ij} - m_j^i l^i| \geq \varepsilon) = 0$$

avec $L = (l^i) = \pi$ vecteur limite.

Théorème de McMillan par les CMH ergodiques.

A chaque étape, l'entropie ne croit que de l'entropie conditionnelle $H_{X_{n-1}}(X_n)$

$$H = H(X_0) + \sum_{k=1}^n H_{X_{k-1}}(X_k)$$

Théorème Inégalité de Chernoff

Soient X_1, \dots, X_n v.a. de Bernoulli indépendantes de paramètres p_1, \dots, p_n .

On pose $S = \sum_{i=1}^n X_i$ et $\mu = E(S) = \sum_{i=1}^n p_i$.

$$\forall \sigma > 0 \quad P[S > (1 + \sigma)\mu] < \left[\frac{e^\sigma}{(1 + \sigma)^{1 + \sigma}} \right]^\mu$$

Théorème même hypothèses

$$\forall \sigma \quad 0 < \sigma < 1$$

$$\forall \sigma > 0 \quad P[S < (1 - \sigma)\mu] < \left[\frac{e^{-\sigma}}{(1 + \sigma)^{1 - \sigma}} \right]^\mu$$

Théorème Inégalité de Hoeffding

Soit X_1, \dots, X_n des v.a. indépendantes telles que $\exists a_1, \dots, a_n, b_1, \dots, b_n$ tels que :

$\forall i \quad P(a_i \leq X_i \leq b_i) = 1$ et soit $S_n = \sum_{i=1}^n X_i$, alors :

$$\forall \varepsilon > 0 \quad P(S_n - E(S_n) \geq \varepsilon) \leq e^{\frac{-2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

$$\text{et} \quad P(S_n - E(S_n) \leq -\varepsilon) \leq e^{\frac{-2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}}$$

Quatrième partie

Statistique

On a une boîte de 10 pièces, dont :
 des déséquilibrées : $P(F) = \frac{3}{4}$ modèle 1
 des équilibrées : $P(F) = \frac{1}{2}$ modèle 0

X : nombre de "Face"
 H_0 équilibré : $X \rightsquigarrow B(n = 10, p = \frac{1}{2})$
 H_1 déséquilibré : $X \rightsquigarrow B(n = 10, p = \frac{3}{4})$

1. Statistique Bayésienne : Si on sait que $P(p = \frac{3}{4}) = \frac{1}{3}$ (probabilité a priori).

$$\begin{aligned}
 P(p = \frac{3}{4} | X = 8) &= \frac{P(p = \frac{3}{4} \cap X = 8)}{P(X = 8)} \\
 &= \frac{P(X = 8 | p = \frac{3}{4})P(p = \frac{3}{4})}{P(X = 8 | p = \frac{3}{4}) + P(X = 8 | p = \frac{1}{2})P(p = \frac{1}{2})} \\
 &= \frac{C_{10}^8 (\frac{3}{4})^8 (\frac{1}{4})^2 \cdot \frac{1}{3}}{C_{10}^8 (\frac{3}{4})^8 (\frac{1}{4})^2 \cdot \frac{1}{3} + C_{10}^8 (\frac{1}{2})^{10} \frac{2}{3}} \\
 &= 0.7
 \end{aligned}$$

0.7 : probabilité a posteriori (après l'expérience).

2. Statistique classique : On ne connaît pas $P(p = \frac{3}{4})$.

Théorème des tests d'hypothèse - Neyman-Pearson

- (a) Définir les hypothèses
 H_0 : hypothèse nulle
 H_1 : hypothèse alternative

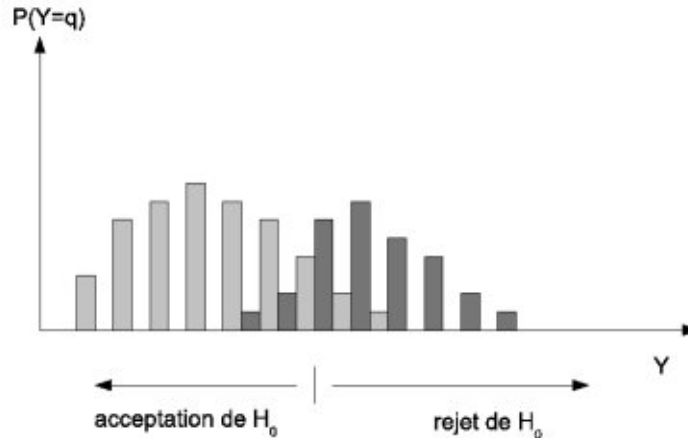
D : Décision, R : Réalité

D \ R	H_0	H_1
H_0		Erreur de type 2
H_1	Erreur de type 1	

$H_0 : p = \frac{1}{2}$: équilibré
 H_1 : : déséquilibré.

3. **Def** l'expression est la variable relative utilisée pour la décision.
 Ici : N titrages, Y : nombre de "Face".
4. Loi de Y sois les 2 hypothèses :
 Ici : $H_0 : Y \rightsquigarrow B(10, \frac{1}{2})$ $H_1 : Y \rightsquigarrow B(10, \frac{3}{4})$.

5. Sens des inégalité et construction du test (règle de décision)
- domaine de rejet (de H_0)
 - domaine d'acceptation (de H_1)



Domaine d'acceptation : $A = [0, s]$ (s : seuil).
 Décision : on accepte H_0 si $H_1 \leq s$, on rejette H_0 sinon.
 nb : rapport de vraisemblance :

$$\frac{P(Y = k|H_0)}{P(Y = k|H_1)}$$

6. Choix de niveau α de test :

Def $\alpha = P_{H_0}(\text{rejeter } H_0) = P_{H_0}(Y > s)$, en général $\alpha = 5\%$.

s	α
9	$P_{H_0}(Y = 10) = C_{10}^{10} \frac{1}{2^{10}} \simeq 10^{-3}$
8	$10^{-3} + P_{H_0}(Y = 9) = 10^{-3} + C_{10}^8 \frac{1}{2^{10}} = 10^{-2}$
7	$10^{-2} + P_{H_0}(Y = 8) = 10^{-2} + C_{10}^8 \frac{1}{2^{10}} = 5.5 \cdot 10^{-2} (> 5\%)$

On prend $s=7$, ce qui donne $\alpha = 5.5\%$.

7. Calcul de la puissance du test (π)

Def On appelle risque de la 2eme espèce $\beta = P_{H_1}(\text{rejeter } H_1)$
 puissance $\pi = 1 - \beta = P_{H_1}(\text{accepter } H_1) = P_{H_1}(\text{rejeter } H_0)$

Ici :

$$\begin{aligned} \pi &= P_{H_1}(Y > 7) \\ &= C_{10}^8 \left(\frac{3}{4}\right)^8 \left(\frac{1}{4}\right)^2 + C_{10}^9 \left(\frac{3}{4}\right)^9 \frac{1}{4} + C_{10}^{10} \left(\frac{3}{4}\right)^{10} \end{aligned}$$

Si besoin : courbe de puissance.