

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



N° Réf :.....

Centre Universitaire de Mila

Institut des Sciences et de la Technologie

Département de Mathématiques et Informatique

Mémoire préparé En vue de l'obtention du diplôme de licence

En:- Filière Mathématiques Fondamentales

Thème

L'approche Bayésienne
L'approche Bayésienne

Préparé par :

- *Bahloul Ramla*
- *Bénidir Bassma*
- *Bouzidi Wafa*
- *Benhedj Hadda*

Encadré par : *Zerari Amel* Grade :M.A.B

Année universitaire : 2013/2014

REMERCIEMENTS

Nous remercierons dieu tout puissant pour nous avoir offert la force et la patience durant toutes ces années.

Nous tenons à exprimer Un remerciement particulier à notre encadreur « Zerari Amel » pour avoir dirigé ce travail, pour sa Présence, son aide et surtout pour ses précieux conseils.

Mille merci nos trop chères familles, que dieu vous garde pour nous.

Nous tenons à exprimer, nos sincères remerciements à tout le personnel de l'institut des sciences et de la technologie surtout les enseignants qui nous ont formé durant les années d'étude, et tous ceux qui nous ont apporté une aide au pour la réalisation de ce projet.

Sans oublier bien-sûre tous les amis et collègues d'études pour leur enjouement et soutient moral.

Table des matières

Introduction Générale	2
1 Introduction Probabilité	4
1.1 Notions générales de la probabilité:	4
1.1.1 Probabilité combinatoires:	5
1.2 Propriétés des probabilités:	6
1.3 Probabilités conditionnelles:	7
1.4 Probabilités marginales:	8
1.4.1 Distribution marginale:	9
1.5 Couple de variables aléatoires discrètes:	9
1.5.1 Généralité	9
1.5.2 Fonction de répartition d'un couple marginale:	10
1.5.3 Loi de probabilité conditionnelle ,Espérance conditionnelle:	10
1.6 Couple de variables aléatoires continues:	11
1.6.1 Fonction densité marginales du couple (X, Y) :	11
1.6.2 Variables conditionnelles:	13
1.7 Lois de probabilité discrètes:	16
1.7.1 Définition d'une variable discrète	16
1.7.2 Loi de Dirac	17

1.7.3	Loi binomiale ou loi des tirages avec remise	18
1.8	Loi de probabilité continues	21
1.8.1	Généralités	21
1.8.2	Loi uniforme	22
1.8.3	Loi exponentielle	23
1.8.4	Loi gamma	25
1.8.5	Lois bêta	26
1.8.6	Loi de Laplace-Gauss ou loi normale	26
1.8.7	Ttéorème de Bayes:	
2	L'approche bayésienne:	32
	Introduction	32
2.1	L'approche bayésienne:	33
2.2	.La loi a priori et a posteriori	34
2.2.1	Loi a priori:	34
2.2.2	La loi a posteriori	34
2.3	Le problème de choix de l'a priori	34
2.3.1	Lois subjectives	35
2.3.2	Les Lois a priori conjuguées:	35
2.3.3	Principe des règles d'arrête:	37
2.4	Distributions a priori et a postériori:	37
2.5	La vraisemblance:	39
2.5.1	Principe de vraisemblance:	39
2.6	Lois conjuguées des familles exponentielles	40
2.7	Loi a priori non informatives:	41

2.7.1	La loi a priori de Jeffreys:	41
2.7.2	Les bases de la théorie de la decision:	42
2.7.3	La fonction de coût:	42
2.7.4	Fonctions de coût usuelles:	43
2.7.5	Facteur de bayes:	44
	Conclusion	44
	BIBLIOGRAPHE	45

Introduction

Les premières personnes à s'être intéressées aux problèmes des probabilités furent des mathématiciens français, Blaise Pascal et Pierre de Fermat qui répondaient aux questions soulevées par un adepte des jeux de hasard, le chevalier de Méré.

A cette époque, la théorie des probabilités se développa uniquement en relation avec les jeux de hasard. Mais avec Pierre Simon Laplace et Karl Friedrich Gauss, les bases de la théorie furent étendues à d'autres applications et phénomènes.

Le calcul des probabilités fournit une modélisation efficace des situations non déterministes c'est à dire des phénomènes aléatoires ou stochastiques. En ce qui concerne les premières, le résultat d'une expérience suit une loi rigoureuse connue (taux de croissance d'une population bactérienne).

On peut donc ainsi prévoir le résultat pour un événement donné. En revanche dans le cas des phénomènes aléatoires, le résultat de l'expérience n'est pas connu avec certitude mais fluctue autour d'un résultat moyen qui est régi par une loi.

Il existe deux manières d'introduire la notion de probabilité:

*La probabilité a priori <<Subjective>> d'un événement est un nombre qui caractérise la croyance que l'on a que cet événement est réalisé avec plus ou moins de certitude

avant l'exécution de l'expérience: l'événement est réalisé (Probabilité 1) et l'événement n'est pas réalisé (Probabilité 0).

*La probabilité empirique assimilée à une fréquence est définie à partir d'expériences indéfiniment renouvelables. La probabilité d'un événement est alors la fréquence d'apparition de cet événement.

L'étude des tests bayésiens a fait l'objet de plusieurs livres et articles récents [Jean Leray 2007- Jérôme DUPUIS Septembre 2007- Lionel Riou França Mai 2009- Christian P-Robert Paris 2006- Jean Michel Marin- Judith Rousseau 2009-2010] L'approche bayésienne offre plus de souplesse à l'analyse des données expérimentales.

On a un phénomène aléatoire que l'on veut comprendre.

Ce phénomène aléatoire est a priori complexe et on ne peut calculer directement les probabilités de ses éventualités.

Ce mémoire est divisé en deux chapitres, dans **le premier chapitre**, nous présentons quelques définitions et propriétés des éléments des probabilités.

Dans **le deuxième chapitre** a pour objet d'introduire les notions indispensables à l'inférence bayésienne, et l'analyse bayésienne.

Chapitre 1

Introduction Probabilité

Le passage d'une description de type ensembliste des phénomènes aléatoires à l'élaboration d'un véritable modèle mathématique se fait en introduisant les mesures de probabilité.

1.1 Notions générales de la probabilité:

Définition

On appelle probabilité P toute application de l'ensemble des événements Ω dans l'intervalle $[0, 1]$, tel que :

$$P : \varepsilon(\Omega) \rightarrow [0, 1]$$

$$A \rightarrow P(A)$$

Satisfaisant les propriétés (ou axiomes) suivantes:

- 1) $\forall A \in \varepsilon(\Omega) \quad P(A) \geq 0$
- 2) $P(\Omega) = 1$
- 3) $\forall A, B \in \varepsilon(\Omega)$

$$4) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Si

$$A \cap B = \phi$$

Alors

$$P(A \cup B) = P(A) + P(B)$$

1.1.1 Probabilité combinatoires:

Soit Ω un espace fondamental fini constitué de N événements élémentaires sur lequel on fait l'hypothèse d'équiprobabilité de réalisation des N événements élémentaires on suppose ainsi que tous les événements élémentaires ont «la même chance» de se réaliser. dans ce cas la probabilité P_i d'un événement élémentaire quelconque W_i telle que:

$$P_i = \frac{1}{N}$$

avec

$$P_i = P(W_i)$$

satisfaisant: (P_i) avec

$$1) \quad \forall P_i \geq 0$$

$$2) \quad \sum_i P_i = 1$$

Soit A un événement quelconque constitué K événement élémentaires de Ω , on en déduit:

$$P(A) = \frac{K}{N}$$

avec

$$P(A) = \sum_{w \in A} P_i, W_i \in A$$

Cette formule s'énonce souvent comme:

$$P(A) = \frac{\text{card } A}{\text{card } \Omega} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$$

Cette formule permet de ramener les calculs de probabilités à des décomptes d'événements élémentaire effectués par des techniques d'analyse combinatoire qui ne sont pas des probabilités.

1.2 Propriétés des probabilités:

Des axiomes précédents découlent les propriétés additives des probabilité, d'usage permanent

* Cas de deux événements quelconques

Si A et B sont deux événements quelconques, alors:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

voici pour quoi:

A et B étant deux événements quelconques, alors:

$$A = A' \cup (A \cap B)$$

avec

$$A' \cap (A \cap B) = \phi$$

$$(A \cap B) \neq \phi$$

, ces événement peuvent se alors;

$$P(A) = P(A') + P(A \cap B)$$

décomposer comme la réunion de deux et

$$P(A') = P(A) - P(A \cap B)$$

événements incompatibles:

$$B = B' \cup (A \cap B)$$

avec

$$B' \cap (A \cap B) = \phi$$

d'où

$$P(B') = P(B) - P(A \cap B)$$

$$P(A \cup B) = P(A') + P(A \cap B) + P(B')$$

d'où

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Dans l'exemple du lancer d'un dé à 6 faces, non pipé, on considère l'événement A («le résultat est pair»)

ona alors:

$$A = \{2, 4, 6\} \quad \text{et} \quad B = \{3, 6\} \quad \text{donc} \quad A \cup B = \{2, 3, 4, 6\} \quad \text{et} \quad A \cap B = \{6\} \quad \text{avec}$$

$$P(A) = 3/6 \quad P(B) = 2/6$$

$$P(A \cup B) = 4/6 \quad \text{et} \quad P(A \cap B) = 1/6$$

on vérifie alors que:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/6 + 2/6 = 4/6$$

1.3 Probabilités conditionnelles:

Il arrive souvent de devoir résoudre un problème du type suivant:

une boîte contient 100 ampoules électriques dont 5 sont défectueuses on retire une ampoule et elle est défectueuse.

Quelle est la probabilité qu'une deuxième ampoule retirée soit elle aussi défectueuse.

Il s'agit donc d'une réévaluation de la probabilité compte tenu de l'information supplémentaire disponible. ces problèmes sont résolus en étudiant les probabilités conditionnelles.

Soit deux événements A et B , la probabilité conditionnelle de l'événement B , étant donné l'événement A , est la probabilité que B arrive, étant donné que A s'est produit. cette probabilité conditionnelle sera désignée $P(B/A)$.

Comme A est arrivé, il faut considérer uniquement les résultats qui correspondent à cet événement.

L'espace d'échantillonnage est alors réduit aux événements dans A .

La probabilité que B arrive, étant donné que A s'est produit, est alors:

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

$$P(A) \neq 0$$

1.4 Probabilités marginales:

Généralisons sous forme symbolique.

Si X désigne la variable aléatoire en colonne et x_j la modalité de rang j de cette variable, la probabilité marginale de X se note:

$$P_r(X = x_j) = P_j = \frac{n_{+j}}{n_{++}} = \frac{n_{.j}}{n_{..}}$$

Si Y désigne la variable aléatoire en ligne et y_i la modalité de rang i de cette variable, la probabilité marginale de Y se note:

$$P_r(Y = y_i) = P_i = \frac{n_{i+}}{n_{++}} = \frac{n_{i.}}{n_{..}}$$

Par commodité, les modalités de rang i , respectivement j seront notées par la suite modalité i , respectivement modalité j .

La suite des probabilités marginales en lignes définissent la distribution marginale en lignes.

1.4.1 Distribution marginale:

La notion de distribution marginale est l'opposée de celle de distribution conjointe.

Il s'agit de la distribution individuelle d'une variable aléatoire sonsténir compte de la covariation d'autres variables aléatoires.

Apartir de la distribution conjointe d'un ensemble de variable aléatoires est définie la distribution marginale de toutes les variables qui le composent.

Les lois marginales déduite d'une distribution multinormale sont normales de mêmes espérances et variances. C'est une très forte priété de cohérence de cette distribution.

si on donne la loi du couple (X, Y) , la loi de probabilité de X et celle de Y sont appelées lois de probabilité marginale.

1.5 Couple de variables aléatoires discrètes:

1.5.1 Généralité

On se consaie à l'étude simultanée de 2 variables aléatoires X, Y discrètes. (X, Y) est appelé couple aléatoires ou variable aléatoire à 2 dimensions ou vecteur aléatoire de dimension 2.

Loi de probabilité marginales du couple (X, Y) :

La loi du couple (X, Y) est définie par

- 1) Les valeurs de (X, Y) $\{(x_i, y_j) \mid i \in \{1, \dots, n\} \mid j \in \{1, \dots, p\}\}$
- 2) Les probabilités correspondantes

$$\begin{aligned}
 P_{ij} &= P((X, Y) = (x_i, y_j)) \\
 &= P(X = x_i, Y = y_j) \\
 &= P((X = x_i) \cap (Y = y_j))
 \end{aligned}$$

Définition:

Si on donne la loi du couple (X, Y) , la loi de probabilité de X et celle de Y sont appelées; lois de probabilité marginales.

propriété: pour $X : P(X = x_i) = \sum_{j=1}^p p_{ij}$ (somme d'une ligne du tableau)

pour $Y : P(Y = y_j) = \sum_{i=1}^n p_{ij}$ (somme d'une colonne)

$$\sum_{i=1}^n \sum_{j=1}^p p_{ij} = 1$$

1.5.2 Fonction de répartition d'un couple marginale:

La fonction de répartition du couple (X, Y) est une fonction de \mathbb{R}^2 dans $[0, 1]$ définie par

$$F_{X,Y}(x, y) = P(X \leq x \cap Y \leq y)$$

1.5.3 Loi de probabilité conditionnelle ,Espérance conditionnelle:

On appelle variable aléatoire conditionnelle X sachant $Y = y_j$, notée

$$X | Y = y_j$$

la variable aléatoire discrète de valeurs $\{x_i, i = 1, \dots, n\}$ et dont les probabilité sont

$$P(X = x_i | Y = y_j) = \frac{p_{ij}}{p \cdot j}$$

on appelle espérance conditionnelle de X sachant $Y = y_j$ la quantité:

$$E(X | Y = y_j) = \sum_{i=1}^n x_i P(X = x_i | Y = y_j).$$

Théorème de l'espérance conditionnelle:

$$E(X) = \sum_{j=1}^p E(X | Y = y_j) P(Y = y_j)$$

Preuve:

$$\begin{aligned} \sum_{j=1}^p E(X | Y = y_j) P(Y = y_j) &= \sum_{j=1}^p \sum_{i=1}^n x_i \frac{p_{ij}}{p_{\cdot j}} \times p_{\cdot j} = \sum_{i=1}^n x_i \sum_{j=1}^p p_{ij} \\ &= \sum_{i=1}^n p_{i\cdot} \cdot x_i = E(X) \end{aligned}$$

1.6 Couple de variables aléatoires continues:

Soient (X, Y) deux variables aléatoires continues.

1.6.1 Fonction densité marginales du couple (X, Y) :

Définition:

Fonction densité du couple (X, Y) est définie par:

$$f_{X,Y}(x, y) = \frac{d^2 F_{X,Y}(x, y)}{dxdy}.$$

On peut également donner la fonction de répartition conjointe en fonction de répartition conjointe en fonction de la fonction densité:

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) dudv$$

$$\forall (x, y) \in \mathbb{R}^2$$

ona alors

$$P(a < X \leq b \cap c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy$$

que l'on peut également écrire

$$P((X, Y) \in]a, b[\times]c, d]) = \int_a^b \int_c^d f_{X,Y}(x, y) dx dy .$$

Exemple:

$$f_{(X,Y)} = \begin{cases} kx^2 + y^2 - xy & \text{si } x \in [-1, 1] \text{ et } y \in [-1, 0] \\ 0 & \text{sinon} \end{cases}$$

- a) déterminer k pour que f soit effectivement une fonction densité d'un couple (X, Y) .
- b) calculer $P \left[\{0 \leq X \leq 1\} \cap \left\{ \frac{1}{2} \leq Y \leq 1 \right\} \right]$.
- c) fonctions densité marginales .
- d) fonction de répartition conjointe.

Solution:

a) $\int_{\mathbb{R}^2} f(x, y) dx dy = 1$

$$\Leftrightarrow \int \int_D kx^2 + y^2 - xy \, dx dy = 1$$

$$\Leftrightarrow \int_{-1}^1 \left(\int_{-1}^0 kx^2 + y^2 - xy dy \right) dx = 1$$

$$\Leftrightarrow \int_{-1}^1 \left(kx^2 + \frac{1}{3} + \frac{x}{2} \right) dx = 1$$

$$\Leftrightarrow k \left[\frac{x^3}{3} \right]_{-1}^1 + \frac{1}{3} [x]_{-1}^1 + \left[\frac{x^2}{4} \right]_{-1}^1 = 1$$

$$\Leftrightarrow \frac{2}{3}k + \frac{2}{3} = 1 \Leftrightarrow k = \frac{1}{2}$$

b) $\int_0^1 \int_{-\frac{1}{2}}^0 \left(\frac{x^2}{2} + y^2 - xy \right) dy dx = \int_0^1 \left(\frac{x^2}{4} + \frac{1}{24} + \frac{x}{8} \right) dx$

$$= \frac{1}{4} \left[\frac{x^3}{3} \right]_0^1 + \frac{1}{4} + \frac{1}{8} \left[\frac{x^2}{2} \right]_0^1$$

$$= \frac{1}{12} + \frac{1}{24} + \frac{1}{16} = \frac{3}{24} + \frac{1}{24} = \frac{1}{8} + \frac{1}{16} = \frac{3}{16} .$$

c) On a $f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy$.

si $x \notin [-1, 1]$, $f_X(x) = 0$.

si $x \in [-1, 1]$, $f_X(x) = \int_{-1}^0 \left(\frac{x^2}{2} + y^2 - xy \right) dy$

$$= \frac{x^2}{2} + \frac{1}{3} + \frac{x}{2} .$$

ona $f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx$

si $y \notin [-1, 0]$, $f_Y(y) = 0$.

si $y \in [-1, 0]$, $f_Y(y) = \int_{-1}^1 \left(\frac{1}{2}x^2 + y^2 - xy \right) dx$.

$$= \frac{1}{2} \left[\frac{x^3}{3} \right]_{-1}^1 + y^2 [x]_{-1}^1 - y \left[\frac{x^2}{2} \right]_{-1}^1$$

$$= \frac{1}{3} + 2y^2 .$$

$$d) F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du .$$

$$\text{- si } x < -1 \text{ ou } y < -1, F_{X,Y}(x, y) = 0$$

$$\text{- si } (x, y) \in [-1, 1] \times [-1, 0],$$

$$\begin{aligned} F(x, y) &= \int_{-1}^x \int_{-1}^y (ku^2 + v^2 - uv) dv du \\ &= \int_{-1}^x ku^2 [v]_{-1}^y + \left[\frac{v^3}{3} \right]_{-1}^y - u \left[\frac{v^2}{2} \right]_{-1}^y du \\ &= \int_{-1}^x ((y+1)ku^2 + \frac{y^3+1}{3} - u \frac{y^2-1}{2}) du \\ &= \frac{y+1}{2} \left[\frac{u^3}{3} \right]_{-1}^x + \frac{y^3+1}{3}(x+1) - \frac{y^2-1}{2} \left[\frac{u^2}{2} \right]_{-1}^x \\ F_{X,Y}(x, y) &= \frac{y+1}{2} \frac{x^3+1}{3} + (x+1) \frac{y^3+1}{3} - \frac{y^2-1}{2} \frac{x^2-1}{2}. \end{aligned}$$

$$\text{- si } y > 0 \text{ et } x \in [-1, 1],$$

$$F_{X,Y}(x, y) = \frac{x^3+1}{6} + \frac{x+1}{3} + \frac{1}{4}(x^2-1).$$

$$\text{- si } y \in [-1, 0] \text{ et } x > 1,$$

$$F(x, y) = \frac{y+1}{3} + \frac{2}{3}(y^3+1).$$

$$\text{- si } y > 0 \text{ et } x > 1,$$

$$F(x, y) = 1.$$

1.6.2 Variables conditionnelles:

On veut définir la loi de probabilité de la variable conditionnelle $Y | X = x$ avec X, Y deux variables aléatoires de fonction densité conjointe $f_{x,y}$

-Fonction de répartition conditionnelle de $Y | X = x$

$P(Y \leq y | X = x)$ a priori n'est pas défini car $P(X = x) = 0$, X étant une variable continue .

On note $F_X = x$ cette fonction de répartition conditionnelle, que l'on définit de la manière suivante:

$$\begin{aligned}
F_{X=x}(y) &= \lim_{\substack{\epsilon \rightarrow 0 \\ \eta \rightarrow 0}} P(Y \leq x - \epsilon \leq X \leq x + \eta) = \lim_{\substack{\epsilon \rightarrow 0 \\ \eta \rightarrow 0}} \frac{P(Y \leq y \cap x - \epsilon \leq X \leq x + \eta)}{F_X(x + \eta) - F_X(x - \epsilon)} \\
&= \lim_{\substack{\epsilon \rightarrow 0 \\ \eta \rightarrow 0}} \frac{P(Y \leq y \cap X \in [x - \epsilon, x + \eta])}{F_X(x + \eta) - F_X(x - \epsilon)} \\
&= \frac{\frac{dF}{dx}(x, y)}{f_X(x)} = \frac{\frac{dF_{X,Y}}{dx}(x, y)}{f_x(x)}
\end{aligned}$$

Définition:

La fonction de répartition de la variable conditionnelle $Y | X = x$, si elle existe, est égale à

$$F_{X,Y}(y) = \frac{\frac{dF_{X,Y}}{dx}(x, y)}{f_x(x)}$$

Définition:

La fonction densité de la variable conditionnelle $Y | X = x$ est

$$f_{X=x}(y) = \frac{f(x, y)}{f_X(x)} \Rightarrow f_{Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$$

Exemple:

$$\text{reprendre } \begin{cases} f(x, y) = kx^2 + y^2 - xy & \text{si } x \in [-1, 1] \text{ et } y \in [-1, 0] \\ f(x, y) = 0 & \text{sinon} \end{cases}$$

a) Donner la fonction densité conditionnelle de $X | Y = 0$.

b) Calculer $P(0 < X < 1 | Y = 0)$.

c) Calculer l'espérance de $X | Y = 0$.

Solution:

$$\text{a) } f_{Y=0}(x) = \frac{f(x, 0)}{f_Y(0)} = \frac{\frac{x^2}{3}}{\frac{1}{3}} = \frac{3x^2}{2} \text{ si } x \in [-1, 1] . \text{ Ainsi}$$

$$f_{Y=0}(x) = \begin{cases} \frac{3x^2}{2} & \text{si } x \in [-1, 1] \\ 0 & \text{sinon} \end{cases}$$

$$\text{b) } P(0 < X < 1 | Y = 0) = \int_0^1 \frac{3x^2}{2} dx = \frac{3}{2} \left[\frac{x^3}{3} \right]_0^1 = \frac{1}{2} .$$

$$\text{c) } E(X | Y = 0) = \int_{-\infty}^{+\infty} x f_{Y=0}(x) dx = \int_{-1}^1 \frac{3x^3}{2} dx = \frac{3}{2} \left[\frac{x^4}{4} \right]_{-1}^1 = 0 .$$

Indépendance en probabilité covariance-coefficient de corrélation:

Comme dans le cas des variables discrètes, intuitivement on peut penser que X et Y sont indépendantes si et seulement si tout évènement lié à X est indépendant de tout évènement lié à Y .

i.e. $\{X \leq x\}$ et $\{Y \leq y\}$ sont indépendants $\forall x, y$

i.e.

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x)P(Y \leq y)$$

Définition:

deux variables aléatoires X, Y continues sont indépendantes en probabilité si et seulement si l'une des propriétés suivantes équivalentes sont vérifiées (si elles ont un sens)

$$(1) f_{X,Y}(x, y) = f_X(x) \times f_Y(y) \quad \forall x, y$$

$$(2) F_{X,Y}(x, y) = F_X(x) \times F_Y(y) \quad \forall x, y$$

$$(3) f_{X=x}(y) = f_Y(y).$$

$$(4) f_{Y=y}(x) = f_X(x) .$$

Preuve:

$$(2) \implies (1) \frac{d^2}{dx dy}$$

$$(1) \implies (2) \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv = \int_{-\infty}^x f_x(u) du \int_{-\infty}^y f_Y(v) dv.$$

$$(3) \implies (1) \text{ on a } f(x, y) = f_X(x) f_Y(y) \forall x, y \text{ tq } f_X(x) \neq 0$$

$$(4) \iff (1) \text{ pareil.}$$

Théorème:

Si X et Y sont indépendants alors

$$(1) Cov(X, Y) = 0$$

$$(2) \rho(X, Y) = 0$$

$$(3) E(XY) = E(X)E(Y)$$

$$(4) V(X + Y) = V(X) + V(Y)$$

$$(5) \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

Ici, cela se traduit par

$$\int_{\mathbb{R}^2} xy f_{X,Y}(x, y) dx dy - E(X)E(Y)$$

on peut également exprimer la covariance à l'aide de l'espérance conditionnelle:

$$\text{Cov}(X, Y) = \int_{\mathbb{R}} x E(Y | X = x) f_X(x) dx - E(X)E(Y)$$

1.7 Lois de probabilité discrètes:

Pour trouver un modèle décrivant un ensemble de données, il est nécessaire de connaître parfaitement les lois statistiques les plus utilisées. Le choix d'une loi est lié :

- à la nature du phénomène étudié afin de choisir entre loi discrète et loi continue,
- à la forme de la distribution (histogramme),
- à la connaissance et à l'interprétation des principales caractéristiques de l'ensemble de données : espérance, médiane, variance, écart-type, coefficients d'asymétrie et de dissymétrie, etc.,
- au nombre de paramètres des lois, une loi dépendant de plusieurs paramètres peut s'adapter plus facilement à une distribution.

1.7.1 Définition d'une variable discrète

Une variable aléatoire discrète prend ses valeurs sur un ensemble fini ou dénombrable de points. La loi de probabilité d'une telle variable est appelée loi discrète.

Une loi de probabilité discrète est caractérisée par l'énumération des valeurs x_i , appartenant à \mathbb{R} ou à un intervalle de \mathbb{R} , prises par la variable aléatoire X et par les probabilités associées, c'est-à-dire les nombres réels positifs p_i tels que

$$Pr(X = x_i) = p_i$$

$$0 \leq p_i \leq 1 \quad \sum_i p_i = 1$$

Moments

$$E(X) = \sum_i x_i p_i \quad Var(X) = \sum_i x_i^2 p_i - [E(X)]^2$$

Domaine d'utilisation

Les lois discrètes sont utilisées pour modéliser les résultats des jeux de hasard, les sondages d'opinion, les phénomènes biologiques, les processus aléatoires (files d'attente, évolution de l'état de matériels)

Les plus utilisées sont la loi uniforme, la loi binomiale et les lois dérivées, la loi hypergéométrique, la loi de Poisson.

Exemple

Soit X la variable aléatoire prenant trois valeurs 0, 1, 2 avec

les probabilités :

$$Pr(X = 0) = 1/2 \quad Pr(X = 1) = 1/3 \quad Pr(X = 2) = 1/6$$

$$(1/2 + 1/3 + 1/6 = 1)$$

Espérance mathématique :

$$E(X) = 0 \times 1/2 + 1 \times 1/3 + 2 \times 1/6 = 2/3$$

Variance :

$$Var(X) = (\sigma_x)^2 = E(X^2) - [E(X)]^2 = 1 \times 1/3 + 4 \times 1/6 - (2/3)^2 = 5/9$$

1.7.2 Loi de Dirac

La loi de Dirac au point a de \mathbb{R} est la loi de probabilité δ_a , définie sur \mathbb{R} par :

$$\delta_a(x) = \begin{cases} 1 & \text{si } x = a \\ 0 & \text{si } x \neq a \end{cases}$$

Cette loi est la loi la plus simple, associée à un phénomène déterministe X dont le résultat de toute expérience est égale à a .

Moments

$$E(X) = a \qquad \text{Var}(X) = 0$$

1.7.3 Loi binomiale ou loi des tirages avec remise

Définition et propriétés d'une variable de Bernoulli

On considère une expérience dont le résultat ne peut prendre que deux valeurs appelées, par convention, succès ou échec : un candidat est reçu ou non à un examen, une pièce usinée est bonne ou défectueuse, une porte est ouverte ou fermée...

À une expérience de ce type, est associée une variable aléatoire X prenant la valeur 1 pour le succès et la valeur 0 pour l'échec, avec les probabilités respectives p et $(1 - p) = q$.

Cette variable est appelée variable de Bernoulli.

La loi de probabilité d'une variable de Bernoulli est définie par :

$$Pr(X = 1) = p$$

$$Pr(X = 0) = 1 - p = q$$

Ses moments sont :

$$E(X) = p$$

$$Var(X) = p(1 - p) = pq$$

Domaine d'utilisation : elle est utilisée pour modéliser des matériels qui seront soit survivants (valeur 1), soit défectueux (valeur 0) à un instant donné.

Elle s'applique aux jeux de hasard de type binaire comme pile ou face...

Définition d'une variable binomiale

On réalise n épreuves indépendantes de la même expérience telles que :

- chaque épreuve ne peut avoir que deux résultats, s'excluant mutuellement, soit le succès, soit l'échec,
- la probabilité p de succès est constante à chaque épreuve, la probabilité d'échec est également constante et égale à $1 - p = q$.

Probabilité d'obtenir k succès au cours de ces n épreuves

Soit X la variable aléatoire qui « compte » le nombre de succès au cours de n épreuves.

- Si au cours de n épreuves, on obtient k succès, on obtient également $(n - k)$ échecs.

La probabilité de réalisation d'un tel événement est égale à $p^k(1 - p)^{n-k}$ (les épreuves sont indépendantes).

- Il y a différentes façons d'obtenir k succès au cours de n épreuves, chaque épreuve pouvant être, indépendamment les unes des autres, un succès ou un échec. Le nombre de réalisations possibles de l'événement « obtenir k succès au cours de n épreuves » est le nombre de combinaisons sans répétitions de n objets pris k à k , soit :

$$C_n^k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$\text{D'où} \quad \Pr(X = k) = C_n^k p^k (1 - p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1 - p)^{n-k}$$

Cette expression étant un terme du développement du binôme $[p + (1 - p)]^n$, la variable X est appelée variable binomiale. On vérifie facilement que :

$$\sum_{k=0}^n \Pr(X = k) = 1$$

La loi de la variable X est appelée loi binomiale de paramètres (n, p) , notée $B(n, p)$.

Une variable binomiale est égale à la somme de n variables aléatoires indépendantes de Bernoulli, la loi binomiale est donc la loi des épreuves répétées.

Moments

Une variable binomiale, suivant la loi $B(n, p)$, peut être considérée comme la somme de n variables indépendantes de Bernoulli. D'où les résultats :

$$E(X) = np$$

$$Var(X) = np(1 - p) = npq$$

Domaine d'utilisation

– La loi binomiale décrit des phénomènes ne pouvant prendre que deux états s'excluant mutuellement, succès ou échec dans un jeu, bonne pièce ou pièce défectueuse dans une fabrication, lot acceptable ou lot refusé, défaillance ou fonctionnement d'un matériel.

Exemple

On veut réaliser une étude clinique sur des malades se présentant à une consultation hospitalière. Pour cette étude, seuls les malades répondant à un ensemble de critères C sont retenus. Des statistiques antérieures ont montré que 20% des consultants présentent les critères C .

10 malades viennent consulter le premier jour.

Soit X la variable aléatoire « nombre de malades retenus » c'est-à-dire répondant à l'ensemble des critères C . La variable X suit la loi binomiale $B(10; 0, 20)$.

La probabilité qu'aucun malade ne soit recruté ce jour est égale à :

$$Pr(X = 0) = C_{10}^0 (0, 20)^0 (0, 80)^{10} = 0, 107$$

La probabilité pour qu'il y ait au moins un malade recruté est égale à :

$$Pr(X \geq 1) = 1 - Pr(X = 0) = 1 - 0, 107 = 0, 893$$

1.8 Loi de probabilité continues

1.8.1 Généralités

Une variable aléatoire continue prend ses valeurs sur un ensemble infini non dénombrable de points, elle décrit par exemple la durée de vie d'une batterie de voiture, l'heure d'arrivée des voitures à un péage donné d'autoroute...

Il existe une fonction f non négative, définie pour toute valeur x appartenant à \mathbb{R} et vérifiant, pour toute partie A de \mathbb{R} , la propriété :

$$Pr(X \in A) = \int_A f(x) dx$$

et telle que :

$$\int_{\mathbb{R}} f(x) dx = 1$$

La fonction f est la densité de probabilité de la variable aléatoire X .

La fonction de répartition de la variable aléatoire X est définie par :

$$F(a) = Pr(X < a) = \int_{-\infty}^a f(x) dx$$

Pour toutes les valeurs a et b appartenant à \mathbb{R} , on a donc la relation :

$$Pr(a \leq X < b) = F(b) - F(a)$$

On en déduit :

$$Pr(X = x) = 0 \quad Pr(x \leq X < x + dx) = f(x) dx$$

Espérance mathématique :

$$E(X) = \int_{\mathbb{R}} x f(x) dx$$

Variance :

$$Var(X) = E(X^2) - [E(X)]^2 = \int_{\mathbb{R}} x^2 f(x) dx - [E(X)]^2$$

L'espérance et la variance existent si les intégrales sont définies.

1.8.2 Loi uniforme

Définition

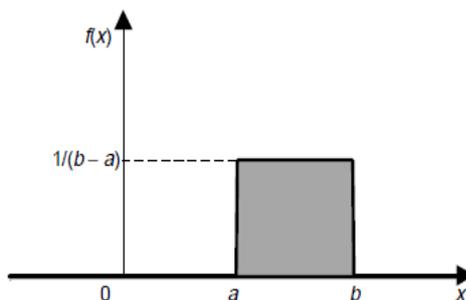
Une variable aléatoire réelle X , suit une loi uniforme sur l'intervalle $[a, b]$, si sa loi de probabilité admet une densité f égale à :

$$f(x) = \frac{1}{b-a} 1_{[a,b]}$$

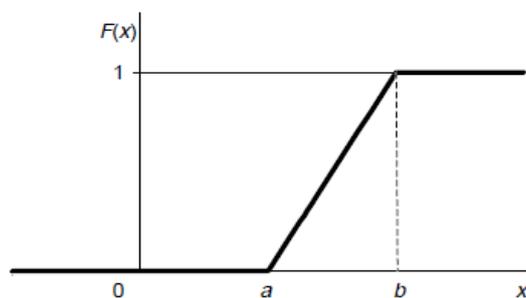
$1_{[a,b]}$ est la fonction caractéristique du segment $[a, b]$.

Fonction de répartition :

$$F(x) = \begin{cases} 0 & \text{si } x \leq a \\ (x-a)/(b-a) & \text{si } a < x < b \\ 1 & \text{si } x \geq b \end{cases}$$



Loi uniforme sur $[a, b]$. Densité



Loi uniforme sur $[a, b]$. Fonction de répartition.

Moments

$$E(X) = \frac{(b+a)}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

1.8.3 Loi exponentielle

Une variable aléatoire réelle positive X suit une loi exponentielle, de paramètre λ positif, si sa densité de probabilité est donnée par :

$$\begin{cases} f(x) = \lambda e^{-\lambda x} & \text{si } x \geq 0 \\ f(x) = 0 & \text{sinon} \end{cases}$$

X est appelée variable exponentielle.

Fonction de répartition :

$$F(a) = Pr(X < a) = \int_0^a \lambda e^{-\lambda x} dx = 1 - e^{-\lambda a}$$

Moments

Espérance et variance :

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

L'espérance d'une variable exponentielle est égale à son écart-type.

Domaine d'utilisation

– La distribution exponentielle est associée aux processus de Poisson. Un tel processus génère des événements dont les temps d'occurrence sont indépendants et distribués suivant une loi exponentielle.

– La loi exponentielle est utilisée en fiabilité, le paramètre λ représente le taux de défaillance alors que son inverse $\theta = 1/\lambda$ est le temps moyen de bon fonctionnement MTBF (Mean Time Between Failure). Avec le paramètre θ , la densité de probabilité s'écrit :

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

et les moments sont égaux à :

$$E(X) = \theta$$

$$Var(X) = \theta^2$$

Exemple

On suppose que le temps, en heures, nécessaire pour réparer une machine est une variable aléatoire suivant une loi exponentielle de paramètre $\lambda = 0,5$.

La densité de probabilité est $f(t) = 0,5e^{-0,5t}$ et la fonction de répartition

$$F(t) = 1 - e^{-0,5t}$$

La probabilité pour que le temps de réparation dépasse 2 heures est :

$$Pr(T > 2) = 1 - Pr(T < 2) = 1 - F(2) = e^{-1} = 0,368$$

Sachant que la réparation a déjà dépassé 9 heures, quelle est la probabilité qu'elle prenne au moins 10 heures ?

La loi exponentielle étant une loi sans « mémoire », on obtient :

$$Pr(T > 10 | T > 9) = Pr(T > 10 - 9 = 1) = e^{-0,5} = 0,606$$

1.8.4 Loi gamma

Définition

La loi exponentielle est un cas particulier de la famille des lois gamma.

Une variable aléatoire réelle positive X suit une loi gamma $\gamma(t; \lambda)$ ou $\Gamma(t; \lambda)$, de paramètres positifs t et λ , si sa densité de probabilité est donnée par :

$$f(x) = \begin{cases} \frac{\lambda e^{-\lambda x} (\lambda x)^{t-1}}{\Gamma(t)} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Γ est la fonction eulérienne définie par l'intégrale pour $t > 0$:

$$\Gamma(t) = \int_0^{\infty} e^{-y} y^{t-1} dy$$

Le paramètre t est un paramètre de forme tandis que $1/\lambda$ est un paramètre d'échelle.

Pour les représentations graphiques, on peut prendre $\lambda = 1$.

Selon les valeurs du paramètre t , la densité de la loi gamma a différentes formes.

En particulier, si $t = 1$, on retrouve la loi exponentielle.

Si le paramètre λ est différent de 1, la variable aléatoire $Y = \lambda X$ suit une loi $\gamma(t, 1)$ ou $\gamma(t)$ de densité :

$$f(y) = \begin{cases} \frac{e^{-y} y^{t-1}}{\Gamma(t)} & \text{si } y \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Moments

Par intégrations par parties et en utilisant les propriétés de la fonction Γ , on obtient :

$$E(X) = \frac{t}{\lambda}$$

$$Var(X) = \frac{t}{\lambda^2}$$

1.8.5 Lois bêta

Une variable aléatoire réelle X , prenant ses valeurs dans l'intervalle $[0, 1]$, suit une loi bêta, notée $\beta(n; p)$, de paramètres positifs n et p , si sa densité de probabilité est donnée par :

$$\begin{cases} f(x) = \frac{1}{B(n,p)} x^{n-1} (1-x)^{p-1} & 0 \leq x \leq 1 \\ f(x) = 0 & \text{sinon} \end{cases}$$

où $B(n; p)$ est la fonction eulérienne définie par :

$$B(n, p) = \int_0^1 x^{n-1} (1-x)^{p-1} dx = B(p, n)$$

$$B(n; p) = \frac{\Gamma(n)\Gamma(p)}{\Gamma(n+p)}$$

La forme de la densité de X varie selon la valeur des paramètres n et p .

Moments

$$E(X) = \frac{n}{n+p} \quad Var = \frac{np}{(n+p+1)(n+p)^2}$$

Propriétés et domaines d'utilisation

– Les lois bêta, dépendant de deux paramètres, s'adaptent bien à la description de nombreux phénomènes aléatoires positifs (temps d'attente, durées de vie...) ; elles sont liées aux lois de Fisher-Snedecor utilisées en statistique.

1.8.6 Loi de Laplace-Gauss ou loi normale

Définition:

Une variable aléatoire réelle X , prenant ses valeurs dans \mathbb{R} , suit une loi de Laplace-Gauss ou loi normale, de paramètres m et σ , si sa densité de probabilité est donnée par :

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

La fonction f définit une densité. En effet :

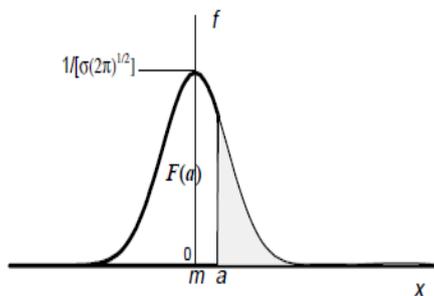
$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

Cette loi est notée, en général $N(m; \sigma)$. On dit indifféremment qu'une variable suivant une telle loi est une variable normale ou gaussienne.

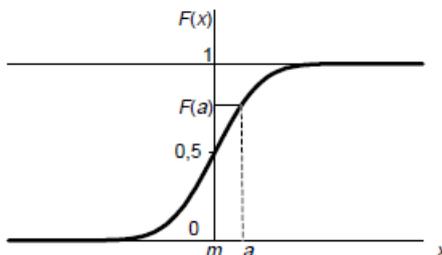
Fonction de répartition :

$$Pr(X < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Cette courbe a un axe de symétrie vertical pour $x = m$ et du fait de sa forme, elle est souvent appelée « courbe en cloche ».



Densité de la loi normale



Fonction de répartition de la loi normale

Moments

Espérance et variance :

$$E(X) = m \quad \text{Var}(X) = \sigma^2$$

Variable aléatoire centrée réduite

La variable centrée réduite associée à la variable aléatoire X est la variable :

$$U = \frac{X - m}{\sigma}$$

Ses moments d'ordre impair sont nuls, en particulier $E(U) = 0$ et les moments d'ordre pair sont égaux à :

$$\mu_{2k} = 1 \times 3 \times \dots \times (2k - 1) = \frac{(2k)!}{2^k k!}$$

en particulier $\text{Var}(U) = 1$.

Densité de probabilité de la variable U :

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

La variable U suit la loi normale $N(0; 1)$ dont les paramètres sont $m = 0$ et $\sigma = 1$.

Fonction de répartition :

$$Pr(X < a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{u^2}{2}}$$

ona:

$$U = \frac{X-m}{\sigma} \quad \text{et} \quad X = \sigma U + m$$

permettant de passer d'une variable à l'autre.

Domaine d'utilisation

– La loi normale est une des lois de probabilité la plus utilisée. Elle dépend de deux paramètres, la moyenne m , paramètre de position, et l'écart-type σ , paramètre mesurant la dispersion de la variable aléatoire autour de sa moyenne.

– Elle s'applique à de nombreux phénomènes, en physique, en économie (erreurs de mesure). De plus, elle est la forme limite de nombreuses distributions discrètes.

– Elle représente la loi de distribution d'une variable aléatoire X dépendant d'un grand nombre de facteurs agissant sous forme additive, chacun ayant une variance faible par rapport à la variance résultante.

Application

Soit $(X_1 \dots X_n)$ un échantillon de n observations indépendantes, issu d'une population suivant la loi $N(m; \sigma)$. La variable aléatoire X , moyenne de l'échantillon :

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

est une combinaison linéaire de n variables aléatoires indépendantes suivant la même loi $N(m; \sigma)$, elle suit donc une loi normale dont les paramètres sont :

$$E(X) = m \quad \text{Var}(X) = \frac{\sigma^2}{n}$$

La variable aléatoire X suit la loi $N(m; \sigma/\sqrt{n})$

1.8.7 Théorème de Bayes:

Un corollaire au théorème des probabilités est connu sous le nom de formule de Bayes si $\{A_1, A_2, \dots, A_i, \dots, A_n\}$ est un système complet d'événements, et quel que soit

l'événement B tel que $P(B) \neq 0$, alors:

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{\sum_{i=1}^n P(B/A_i)P(A_i)}$$

Voici pour quoi:

D'après la formule des probabilités composées

$$P(A_i \cap B) = P(A_i/B)P(B) = P(B/A_i)P(A_i)$$

D'après la formule des probabilités totales

$$P(B) = \sum_{i=1}^n P(B/A_i)P(A_i)$$

D'après la formule des probabilités conditionnelles

$$P(A_i/B) = \frac{P(A_i \cap B)}{P(B)}$$

d'où

$$P(A_i/B) = \frac{P(B/A_i) P(A_i)}{\sum_{i=1}^n P(B/A_i) P(A_i)}$$

Exemple:

une première boîte contient 2 boules blanches et 1 boule rouge une deuxième boîte contient 1 boule blanche et 3 boules rouges on choisit au hasard une boîte et on tire une boule. si la boule est blanche, quelle est la probabilité qu'elle vient de la première boîte?

Soit C_1 l'événement qu'on choisit la première boîte, C_2 celui qu'on choisit la deuxième boîte et B l'événement que la boule tirée est blanche. on a alors:

$$P(B/C_1) = 2/3$$

$$P(B/C_2) = 1/4$$

$$P(C_1) = P(C_2) = 1/2$$

$$P(C_1/B) = \frac{(2/3)(1/2)}{(2/3) \cdot (1/2) + (1/4) \cdot (1/2)} = 8/11$$

La formule de Bayes sera utilisée pour calculer les probabilités a posteriori.

Chapitre 2

L'approche bayésienne:

Introduction

Dans de nombreuses situations d'expériences aléatoires, il semble raisonnable d'imaginer que le praticien a une certaine idée du phénomène aléatoire qu'il est en train d'observer. Or, la démarche statistique classique repose essentiellement sur un principe de vraisemblance qui consiste à considérer que ce qui a été observé rend compte de manière exhaustive du phénomène. Mais l'observation ne fournit qu'une image et celle-ci peut être mauvaise. Certes cet inconvénient est en général gommé par les considérations asymptotiques et un certain nombre de théorèmes permettent d'évaluer la bonne qualité des estimateurs si le nombre d'observations est suffisant.

L'analyse bayésienne des problèmes statistiques propose d'introduire dans la démarche d'inférence, l'information dont dispose a priori le praticien. Dans le cadre de la statistique paramétrique, ceci se traduira par le choix d'une loi sur le paramètre d'intérêt.

Dans l'approche classique, le modèle statistique est le triplet $(\mathfrak{N}, \mathbf{A}, P_\theta, \theta \in \Theta)$. Ayant un a priori sur le paramètre, modélisé par une densité de probabilité que nous noterons $\pi(\theta)$, on "ré-actualise" cet a priori au vu de l'observation en calculant la densité a posteriori $\pi(\theta | x)$, et c'est à partir de cette loi que l'on mène l'inférence.

On peut alors, par exemple, de manière intuitive pour le moment, retenir l'espérance

mathématique ou encore le mode de cette densité a posteriori comme estimateur de θ . Le paramètre devient donc en quelque sorte une variable aléatoire, à laquelle on associe une loi de probabilité dite loi a priori.

On sent bien d'emblée que les estimateurs bayésiens sont très dépendants du choix de l'a priori.

Différentes méthodes existent pour déterminer ces lois a priori. On peut se référer à des techniques bayésiennes empiriques, où l'on construit la loi a priori sur la base d'une expérience passée, usant de méthodes fréquentistes, pour obtenir forme et valeurs de paramètres pour cette loi. Nous verrons que l'on peut aussi modéliser l'absence d'information sur le paramètre au moyen des lois dites lois non informatives.

L'approche bayésienne se différencie donc de l'approche classique dans le sens où le paramètre n'est plus considéré comme étant totalement inconnu ;

2.1 L'approche bayésienne:

Incertitude sur le paramètre θ est représentée par une probabilité π sur Θ .

Le paramètre inconnu devient une variable aléatoire comme les observations.

π est la loi a priori sur θ .

On interprète la loi des observations f_θ comme la loi conditionnelle des observations sachant θ

$$f(x | \theta) = f_\theta(x)$$

Traiter chaque inconnu (paramètre θ , prédicat z, \dots) comme une variable aléatoire, donner lui une distribution (en utilisant souvent l'indépendance), et les calculs sa distribution a posteriori sachant les données, utilisant le théorème de Bayes.

Définition 2.1

Quand $x \sim f(x | \theta)$, une fonction T de x (aussi appelé statistique) est exhaustive si la distribution de x conditionnellement à $T(x)$ ne dépend pas de θ .

2.2 .La loi a priori et a posteriori

2.2.1 Loi a priori:

Le choix des lois a priori est une étape fondamentale dans l'analyse bayésienne. ce choix peut avoir différentes motivations les stratégies sont diverses. Elle peuvent se baser sur des

expériences du passé ou sur un intuition, une idée que le praticien a du phénomène aléatoire qu'il en train de suivre.

Elle peuvent être également motivées par aspects calculabilité.

Enfin, ces stratégies peuvent également tenir compte du fait qu'on ne sait rien par le truchement des lois non informatives.

2.2.2 La loi a posteriori

C'est la loi conditionnelle de θ sachant x , sa densité notée $\pi(\theta | x)$, en vertu de la formule de Bayes de (1, 1), et on a aussi :

a/ La loi du couple (θ, x) : sa densité est notée $\varphi(\theta, x)$ on a donc

$$\varphi(\theta, x) = f(x | \theta) \pi(\theta)$$

b/ La loi marginale de x : sa densité est notée $m(x)$, on a donc

$$m(x) = \int_{\Theta} f(x | \theta) \pi(\theta) d\theta$$

2.3 Le problème de choix de l'a priori

L'aspect de l'analyse bayésienne le plus critiqué et le plus délicat est certainement le choix de la loi a priori des paramètres ; en particulier, le recours à des lois usuelles comme la loi normale, gamma, bêta, ne peut pas être défendu comme approche systématique plus aisée.

La détermination de la loi a priori donc est basée sur l'information a priori.

2.3.1 Lois subjectives

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique. L'idée est d'utiliser les données antérieures. Par exemple, dans un cadre paramétrique, cela revient à choisir une valeur particulière du paramètre.

Dans un cas concret, il peut être judicieux de baser son raisonnement sur les dires d'experts, notamment à l'aide de questionnaires. Il est alors nécessaire de veiller à ce que les questions soient compréhensibles¹, par exemple en prenant comme base les quantiles plutôt que les moments. Pour plusieurs experts, il peut être utile de pondérer leurs réponses et d'utiliser des modèles hiérarchiques.

Ainsi, la difficulté ici n'est pas mathématique mais plus psychométrique pour réduire les biais sur les réponses fournies. Nous allons nous concentrer sur le second aspect de la détermination.

2.3.2 Les Lois a priori conjuguées:

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. ce calcul est facilité lorsque loi a priori et loi a posteriori ont la même forme.

Dans ce cas, on parle de loi a priori conjuguée.

La loi a priori conjuguée:

Une famille \mathcal{F} de lois sur Θ est dite conjuguée si, pour tout π appartenant à cette famille, la loi $\pi(\theta | x)$ appartient également à celle-ci.

Dans ce cas, le praticien induit directement la forme de son estimateur dès qu'il a choisi sa loi a priori.

<i>loi des observation</i> $P(x \theta)$	loi a priori $P(\theta T)$	<i>loi a posteriori</i> $P(\theta x, T) \propto P(\theta T)P(x \theta)$
	variables discrètes	
Binomiale $Bin(x n, \theta)$	Bêta $Bet(\theta \alpha, \beta)$	Bêta $Bet(\theta \alpha + x, \beta + n - x)$
Binomiale $Bin(x n, \theta)$	$P(\theta) = \theta(1 - \theta)^{-1}$	Bêta $Bet(\theta x, n - x), x \neq 0, x \neq n$
<i>Binomiale négative</i> $NegBin(x n, \theta)$	Bêta $Bet(\theta \alpha, \beta)$	Bêta $Bet(\theta \alpha + n, \beta + x)$
<i>Multinomiale</i> $M_k(x \theta_1, \dots, \theta_k)$	Dirichlet $Di_k(\theta \alpha_1, \dots, \alpha_k)$	<i>Dirichlet</i> $Di_k(\theta \alpha_1 + x_1, \dots, \alpha_k + x_k)$
Poisson $P_n(x \theta)$	Gamma $Gam(\theta \alpha, \beta)$	Gamma $Gam(\theta \alpha + x, \beta + 1)$
Gamma $Gam(x v, \theta)$	Gamma $Gam(\theta \alpha, \beta)$	Gamma $Gam(\theta \alpha + v, \beta + x)$
Bêta $Bet(x \alpha, \theta)$	Exponentielle $Ex(\theta \lambda)$	Exponentielle $Ex(\theta \lambda - \log(1 - x))$
Normale $N(x \theta, \sigma^2)$	Normale $N(\theta \mu, T^2)$	Normale $N(\mu \frac{\mu\sigma^2 + T^2 X}{\sigma^2 + T^2}, \frac{\sigma^2 T^2}{\sigma^2 + T^2})$
	variables continues	
Normale $N(x \mu, 1/\theta)$	Gamma $Gam(\theta \alpha, \beta)$	<i>Gamma</i> $Gam(\theta \alpha + \frac{1}{2}, \beta + \frac{1}{2}(\mu - x)^2)$
Normale $N(x \theta, \theta^2)$	<i>Normale inverse généralisée</i> $INg(\theta \alpha, \mu, \sigma) \propto \theta ^{-\alpha}$ $\exp[-\frac{1}{2\sigma^2}(\frac{1}{\theta} - \mu)^2]$	<i>Normale inverse généralisée</i> $INg(\theta \alpha_n, \mu_n, \sigma_n)$

Famille exponentielle

Un type particulier de lois de probabilité permet une détermination directe des familles de lois conjuguées. De telles lois sont dites familles exponentielles.

Définition

Soient μ , mesure σ finie sur X , et Θ l'espace des paramètres. On définit C et H , respectivement fonction de X et Θ dans \mathbb{R}_+ , et R et T , fonction de Θ et X dans \mathbb{R}^K . La famille de distributions de densité (par rapport à μ)

$$f(x | \theta) = C(\theta)H(x) \exp\{R(\theta)T(x)\}$$

est dite famille exponentielle de dimension K . Dans le cas particulier où $\Theta \in \mathbb{R}^k$, $X \subset \mathbb{R}^K$

et :

$$f(x | \theta) = C(\theta)H(x) \exp \{\theta x\}$$

La famille est dite naturelle.

2.3.3 Principe des règles d'arrêt:

Si une suite d'expérience, ξ_1, ξ_2, \dots , admet une règle d'arrêt, τ , qui indique quand doivent s'arrêter les expériences, l'inférence sur θ ne doit dépendre de τ qu'à travers l'échantillon résultant.

En analyse séquentielle. Une règle d'arrêt τ peut être définie comme suit : si les expériences ξ_i produisent des observations $x_i \in \mathfrak{N}_i$, avec $x_i \sim f(x_i | \theta)$, considérons la suite correspondante $A_i \subset \mathfrak{N}_1 \times \dots \times \mathfrak{N}_i$ telle que le critère τ prend la valeur n si $(x_1, \dots, x_n) \in A_n$, i.e., l'expérience s'arrête après la n -ième observation seulement si les n premières observation sont en A_n . La vraisemblance de (x_1, \dots, x_n) est alors

$l(\theta | x_1, \dots, x_n) = f(x_1 | \theta)f(x_2 | x_1, \theta) \dots f(x_n | x_1, \dots, x_{n-1}, \theta)1_{A_n}(x_1, \dots, x_n)$, et donc dépend seulement de τ via l'échantillon x_1, \dots, x_n .

Consiste à observer de $x_i \sim N(\theta, 1)$ et à prendre τ comme le premier entier n tel que

$$|\bar{x}_n| = \left| \sum_{i=1}^n \frac{x_i}{n} \right| > 1,96 / \sqrt{n}$$

Dans ce cas la règle d'arrêt est évidemment incompatible avec la modélisation fréquentiste, parce que avec un tel-échantillon on rejettera toujours l'hypothèse nulle $H_0: \theta = 0$ au seuil de 5%. En revanche, une approche bayésienne évite cette difficulté.

2.4 Distributions a priori et a postériori:

supposons distribution a priori sur θ . $\pi(\theta)$, soit disponible, c'est-à-dire que nous disposons d'un modèle complètement bayésien une fois données ces deux distributions nous pouvons en construire plusieurs autres, a savoir:

a) La distribution jointe de (θ, x) :

$$\varphi(\theta, x) = f(x/\theta)\pi(\theta)$$

b) La distribution marginal de x :

$$m(x) = \int \varphi(\theta, x) d\theta = \int f(x/\theta)\pi(\theta) d\theta$$

c) La distribution a posteriori de θ obtenue par la formule de bayes:

$$\pi(\theta/x) = \frac{f(x/\theta)\pi(\theta)}{\int f(x/\theta)\pi(\theta) d\theta} = \frac{f(x/\theta)\pi(\theta)}{m(x)}$$

d) La distribution prédictive de y ou $y \sim g(y/\theta, x)$ obtenue par

$$g(y/x) = \int g(y/\theta, x)\pi(\theta/x) d\theta$$

Exemple:

Une boule de billard W roule sur une ligne de longueur un, supposons qu'elle s'arrête en P . une deuxième boule O roule alors n fois dans les mêmes conditions, on note X le nombre de fois que la boule O s'arrête à gauche de W .

Quelle inférence pouvons-nous mener sur P ?

Dans la terminologie moderne, le problème est de déterminer la distribution a posteriori de P conditionnellement à X , quand la distribution a priori de P est uniforme sur $[0, 1]$ et $X \sim B(n, p)$, variable aléatoire binomiale comme:

$$P(X = \theta/P) = \binom{n}{\theta} P^\theta (1-P)^{n-\theta}$$

$$P(a < P < b \text{ et } X = \theta) = \int_a^b \binom{n}{\theta} P^\theta (1-P)^{n-\theta} dp \quad \text{et}$$

$$P(X = \theta) = \int_0^1 \binom{n}{\theta} P^\theta (1-P)^{n-\theta} dp$$

nous trouvons que:

$$P(a < p < b \mid X = \theta) = \frac{\int_a^b \binom{n}{\theta} P^\theta (1-P)^{n-\theta} dp}{\int_0^1 \binom{n}{\theta} P^\theta (1-p)^{n-\theta} dp} = \frac{\int_a^b p^\theta (1-p)^{n-\theta} dp}{B(\theta+1, n-\theta+1)}$$

donc la distribution de P conditionnellement à $X = \theta$ est une distribution bêta,

$$B(\theta + 1, n - \theta + 1).$$

2.5 La vraisemblance:

L'inférence statistique cherche à apprendre sur la valeur dans la population d'un paramètre θ à partir d'un échantillon de données y .

Inférence traditionnelle : une fois y connu, $P(y \mid \theta)$ (fonction de vraisemblance) quantifie à quel point certaines valeur de θ sont compatibles avec les valeur observées y .

$P(y \mid \theta)$ résume toute l'information que les données fournissent sur le paramètre θ .

2.5.1 Principe de vraisemblance:

les vraisemblances de y sous les hypothèses contiennent toute l'information qui peut être extraite des données.

La statistique Bayésienne obéit au principe de vraisemblance.

La statistique fréquentiste n'obéit pas au principe de vraisemblance.

– L'information apportée par une observation de x sur θ est entièrement contenue dans la fonction de vraisemblance $l(\theta \mid x)$. De plus, si x_1 et x_2 sont deux observations qui dépendent du même paramètre θ , et telles qu'il existe une constante c satisfaisant

$$l_1(\theta \mid x_1) = c l_2(\theta \mid x_2)$$

Pour tout θ , elles apportent la même information sur θ et doivent conduire à la même inférence.

Notont que le principe de vraisemblance n'est valide que lorsque

- i) L'inférence concerne le même paramètre θ .
- ii) θ prend en compte tous les facteurs inconnus du modèle.

2.6 Lois conjuguées des familles exponentielles

Soit $f(x | \theta) = h(x) \exp\{\theta x - \psi(x)\}$

appartenant à une famille exponentielle. Alors cette loi admet une famille conjuguée comme :

$$\pi(\theta | \mu, \lambda) = K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(x)\}$$

Où $K(\mu, \lambda)$ est la constante de normalisation de la densité. Donc la densité a posteriori est :

$$\begin{aligned} \pi(\theta | \mu + x, \lambda + 1) &= \frac{h(x) \exp\{\theta x - \psi(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(\theta)\}}{\int h(x) \exp\{\theta x - \psi(\theta)\} K(\mu, \lambda) \exp\{\theta\mu - \lambda\psi(\theta)\} d\theta} \\ &= \frac{\exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\}}{\int \exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\} d\theta} \\ &= K(\theta | \mu + x, \lambda + 1) \exp\{\theta x - \psi(\theta)\} \exp\{\theta\mu - \lambda\psi(\theta)\} \\ &= K(\theta | \mu + x, \lambda + 1) \exp\{\theta(x + \mu) - \psi(\theta)(1 + \lambda)\} \end{aligned}$$

On peut mettre comme exemple de lois conjuguées des familles exponentielles le calcul de la loi a posteriori lorsque σ^2 est inconnu.

Dans la partie suivante, quand l'information a priori n'est pas disponible sur le modèle, les lois a priori doivent être construites à partir de la distribution d'échantillonnage. Elles sont appelées les lois non informatives.

2.7 Loi a priori non informatives:

Partons d'un exemple pour introduire de loi a priori non informative. On considère le modèle statistique bayésien suivant:

le $x_i | \theta$ sont i.i.d (identiquement et indépendamment distribués) et suivent une loi de Bernoulli de paramètre $\theta \in]0, 1[$. Il n'y a pas une unique loi a priori non informative pour le paramètre θ , on peut en fait proposer différentes lois a priori non informatives.

En l'absence d'information sur θ , il est naturel de proposer une loi uniforme sur θ car elle donne une probabilité égale aux intervalles de longueur l donnée à savoir L

on peut également proposer la loi a priori de Haldane $\pi(\theta) = [\theta(1-\theta)]^{-1} \mathbf{1}_{]0,1[}(\theta)$, en arguant que $E[\theta | X]$ est égale à l'estimateur du maximum de vraisemblance.

une alternative a été proposée par Jeffreys en 1960.

Soit un paramètre réel on appelle loi a priori non informative de Jeffreys la loi (éventuellement impropre) de densité $\pi_J(\theta) \propto [I(\theta)]^{\frac{1}{2}} \mathbf{1}_{\Theta}(\theta)$ où $I(\theta)$ désigne l'information de Fisher apportée par x sur θ .

2.7.1 La loi a priori de Jeffreys:

Jeffreys (1946, 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souve compatible lorsque cette structure existe. Les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par $I(\theta) = E_{\theta} \left[\left(\frac{\partial \log f(X|\theta)}{\partial \theta} \right)^2 \right]$ dans le cas unidimensionnel. Sous certaines conditions de régularité,

$$\text{cette information est aussi égale à } I(\theta) = -E_{\theta} \left[\frac{\partial^2 \log f(X|\theta)}{\partial \theta^2} \right].$$

La loi a priori de Jeffreys est $\pi^*(\theta) \propto I^{1/2}(\theta)$, définie à un coefficient de normalisation près quand π^* est propre. Elle vérifie effectivement l'exigence d'invariance par reparamétrisation, puisque, pour une transformation bijective h donnée, nous avons la transformation (jacobienne) $I(\theta) = I(h(\theta))(h'(\theta))^2$

Le choix d'une loi a priori dépendant de l'information de Fisher se justifie par le fait que

$I(\theta)$ est largement accepté comme un indicateur de la quantité d'information apportée par le modèle (ou l'observation) sur θ (Fisher, 1956). Par conséquent, au moins à un niveau qualitatif, il paraît intuitivement justifié que les valeurs de θ pour lesquelles $I(\theta)$ est plus grande doivent être plus probables a priori.

En d'autres termes, $I(\theta)$ mesure la capacité du modèle à discriminer entre θ et $\theta \pm d\theta$ via la pente moyenne de $\log f(x|\theta)$. Favoriser les valeurs de θ pour lesquelles $I(\theta)$ est plus grande équivaut à minimiser l'influence de la loi a priori et est donc aussi non informatif que possible. En fait, la loi de Jeffreys est fréquemment impropre.

Exemple

Si $x \sim B(n, P)$

$$f(x|P) = \binom{n}{x} P^x (1-P)^{n-x}$$

$$\text{donc } I(P) = n \left[\frac{1}{P} + \frac{1}{1-P} \right] = \frac{n}{P(1-P)}$$

donc la loi de Jeffreys pour ce modèle est:

$$\pi^*(P) \propto [P(1-P)]^{-\frac{1}{2}}$$

et est alors propre, car il s'agit de la distribution Beta $(\frac{1}{2}, \frac{1}{2})$

Dans le cas où θ est un paramètre multidimensionnel, on définit la matrice d'information de Fisher, pour $\theta \in \mathbb{R}^K$, et $I(\theta)$ aux éléments suivants:

$$I_{ij}(\theta) = -E_{\theta} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log f(x/\theta) \right] \quad i, j = 1, \dots, K$$

et la loi non informative de Jeffreys est alors définie par:

$$\pi^*(\theta) \propto [\det(I(\theta))]^{\frac{1}{2}}$$

2.7.2 Les bases de la théorie de la décision:

2.7.3 La fonction de coût:

On appelle fonction de coût, toute fonction L de $\Theta \times A$ dans \mathbb{R} .

$L(\theta, a)$ évalue le coût d'une décision a quand le paramètre vaut θ . Elle permet donc,

en quelque sorte, de quantifier la perte encourue par une mauvaise décision, une mauvaise évaluation de θ .

2.7.4 Fonctions de coût usuelles:

Coût quadratique:

La fonction de coût quadratique est la fonction définie par :

$$L(\theta, \delta(x)) = (\theta - \delta(x))^2.$$

Une variante de cette fonction de coût est une fonction de coût quadratique pondérée de la forme : $L(\theta, \delta(x)) = \omega(\theta)(\theta - \delta(x))^2$.

Coût 0-1:

Définition 2.2 On appelle coût 0 – 1, l'application L définie par :

$$L(\theta, \delta(x)) = \begin{cases} 0 & \text{si la décision est bonne} \\ 1 & \text{sinon} \end{cases}$$

on retrouve en utilisant cette fonction de coût, les résultats de la théorie des tests d'hypothèses.

Un problème de test est un problème de choix (de prise de décision) entre $H_0 : \theta \in \Theta_0$ et $H_1 : \theta \in \Theta_1$.

On définit donc la décision de la manière suivante:

$\delta(X) = 1$ on accepte H_0 ;

$\delta(X) = 0$ on rejette H_0 . (nb: ceci ne dépend pas de θ).

On a un espace d'actions de la forme : $A = \{0, 1\}$

Soit ω la région de rejet i.e. le sous ensemble de \mathfrak{X} qui conduit à rejeter H_0 . On peut construire une fonction de coût de la manière suivant:

supposons $\theta \in \Theta_0$,

Si $x \in \omega$, on prend la décision de rejeter i.e. $\delta(x) = 0$, mais la décision n'est pas bonne, on va pénaliser et $L(\theta, \delta(x)) = 1$

Si $x \notin \omega$, on ne rejette pas, on prend la décision $\delta(x) = 1$, la décision est bonne $L(\theta, \delta(x)) = 0$.

Le coût s'écrit donc:

$$L(\theta, \delta(x)) = \begin{cases} 1 - \delta(x) & \text{si } \theta \in \Theta_0 \\ \delta(x) & \text{sinon} \end{cases}$$

Ce qu'on peut écrire $L(\theta, \delta(x)) = 1(x \in \omega)$ et on calcule la fonction de risque:

$$R(\theta, \delta) = E[L(\theta, \delta(X))] = \int_{\mathfrak{X}} L(\theta, \delta(x)) dp_{\theta}(x) = p_{\theta}(x \in \omega), \quad \theta \in \Theta_0.$$

2.7.5 Facteur de bayes:

L'idée à l'origine de cette notion est de limiter l'importance du choix a priori de a_0 et a_1 introduits ci-dessus.

Le facteur de bayes est défini par

$$B_{0/1} = \frac{P^{\pi}(\Theta_0/X)}{p^{\pi}(\Theta_1/x)} \times \frac{\pi(\Theta_1)}{\pi(\Theta_0)}$$

Conclusion

Dans ce travail, nous avons étudié la théorie mathématique des lois de probabilité et les aspects décisionnels de l'inférence statistique.

Nous avons étudié une modélisation bayésienne parce qu'elle offre plus de souplesse et d'objectivité dans le traitement statistique des données.

Bibliography

- [1] A Mallet, A J Valleron, F Carrat, S Tézenas, V Morice. (2014). Université Pierre et Marie Curie Biostatistique, 100.
- [2] Christian P Robert (2006). Le choix bayésien Principe et pratique. 19-17, 49-48, 130, 139-140.
- [3] Christian P Robert, Jean Michel Marin, Gilles CELEUX. SÉLECTION BAYÉSIENNE VARIABLE EN RÉGRESSION LINÉAIRE. 61.
- [4] D Mouchiroud. (2002). Mathématique Outils pour la Biologie Deug SV1-UCBL. 9-10, 13-12, 20-22.
- [5] Jean Michel Marin, Christian P Robert. Paris. Les bases de la statistique bayésienne. 89.
- [6] Jerome Dupuis. (2007). Statistique bayésienne et algorithmes. 10-16
- [7] Judith ROUSSEAU. (2010). STATISTIQUE BAYÉSIENNE. 22, 39.
- [8] Laurence GRAMMONT. (2004). COURS DE PROBABILITÉ 2ième année d'économie et de gestion, semestre 2. 12, 19-23, 29-34, 34-38.
- [9] Lionel RIOUFRAÇA. (2009). Statistique Bayésienne Elément de Culture Générale. 5, 18, 109.

-
- [10] Monique Le Guen. (2006). Tableaux croisés et diagrammes en mosaïque, pour visualiser, Les probabilités marginales et conditionnelles. 7.
- [11] Renée Veysseyre.(2006). Aide mémoire Statistique et probabilités pour l'ingénieur. 67-69, 89-100.