

Université Paris 7- Denis Diderot

Notes de cours pour  
le Module

Methodes Statistiques de bases

Dominique Picard<sup>1</sup>

---

1. Copyright © 2010 Université Paris-Diderot Dominique Picard



# Table des matières

<b>1</b>	<b>1. METHODOLOGIE STATISTIQUE</b>	<b>5</b>
1.1	Introduction. Modélisation statistique. . . . .	5
1.2	Modèle Linéaire gaussien . . . . .	7
1.2.1	Exemples de modèles linéaires . . . . .	7
1.3	Identifiabilité, Domination . . . . .	8
1.4	Structure des modèles dominés. . . . .	11
1.4.1	Dominante Privilégiée . . . . .	11
1.4.2	Distance en variation et Modèles dominés . . . . .	13
1.5	Introduction à l'exhaustivité. . . . .	14
1.5.1	Sous-Expérience, perte d'information. . . . .	14



# Chapitre 1

## 1. METHODOLOGIE STATISTIQUE

### 1.1 Introduction. Modélisation statistique.

À la base, le statisticien dispose d'une **observation**  $x$  à valeurs dans un espace  $\mathcal{X}$ . La modélisation consiste à faire l'hypothèse que cette observation est la réalisation  $X(\omega)$  d'une variable aléatoire  $X$ , à valeurs dans  $(\mathcal{X}, \mathcal{A})$ . ( $\mathcal{A}$  est alors une tribu sur  $\mathcal{X}$  et  $\omega$  appartient à un ensemble  $\Omega$ .)

Formellement, on a un triplet  $(\Omega, \mathcal{F}, P)$  c'est à dire que  $\mathcal{F}$  est une tribu sur  $\Omega$ , et  $P$  une mesure de probabilité sur  $\mathcal{F}$ .

$X$  est une application mesurable de  $(\Omega, \mathcal{F})$  dans  $(\mathcal{X}, \mathcal{A})$ , et la loi de  $X$ ,  $P_X$ , est la mesure image de  $P$  par  $X$  définie pour tout ensemble  $A$  de  $\mathcal{A}$  par la formule

$$P^X(A) = P(X^{-1}(A)).$$

En statistique, nous verrons que  $\mathcal{X}$  est souvent  $\mathbb{R}^n$  ou un sous ensemble de  $\mathbb{R}^n$ ,  $\mathcal{A}$  est généralement sa tribu borelienne, et l'espace  $(\Omega, \mathcal{F})$  joue un rôle très auxiliaire. La plupart du temps, on peut identifier  $(\Omega, \mathcal{F})$  et  $(\mathcal{X}, \mathcal{A})$ , de sorte que  $X$  devient l'application identité. On dira alors que le modèle est en position canonique.

**Définition 1.** On appelle **Modèle statistique** ou **Expérience** la donnée de la famille :

$$(\Omega, \mathcal{F}, X, \mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta) \quad \text{où}$$

- $\mathcal{X}, \mathcal{A}$  est l'espace des réalisations de la variable aléatoire  $X$  définie sur  $(\Omega, \mathcal{F})$ .
- $\Theta$  est l'ensemble des **paramètres**.
- $P_\theta$  est une loi de probabilité sur  $(\Omega, \mathcal{F})$ .

Le statisticien fait donc l'hypothèse que son observation  $x$  est la réalisation d'une variable aléatoire (i.e. il existe  $\omega, x = X(\omega)$ ) et qu'il existe  $\theta$  tel que  $\omega$  est tiré selon la loi  $P_\theta$  (la loi de  $X$  est alors  $P_\theta^X$ ).

**Définition 2.** Dans une expérience  $\mathcal{E} = (\Omega, \mathcal{F}, X, \mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$ , on appelle **statistique** toute variable aléatoire de la forme  $T \circ X$  où  $T$  est mesurable de  $(\mathcal{X}, \mathcal{A})$  dans un espace arbitraire muni d'une tribu.

Le statisticien va donc disposer de toutes les "statistiques" comme outil pour "deviner"  $\theta$ .

#### Premiers exemples.

1. Considérons l'exemple du **sondage** : Soit  $N$  et  $n$  fixés. On considère une population de  $N$  éléments qui comprend une proportion  $\theta$  de défectueux. On extrait au hasard  $n$  éléments et on compte le nombre de défectueux parmi cette population extraite. Ce nombre  $x$  est la réalisation d'une variable aléatoire  $X$ . La loi de  $X$  sous  $P_\theta$  est une hypergéométrique que l'on peut écrire sous la forme :

$$P_\theta^X = \sum_{k=0}^n \frac{C_{\theta N}^k C_{(1-\theta)N}^{n-k}}{C_N^n} \delta_k$$

on a bien sur  $\Theta = [0, 1]$ ,  $\mathcal{X} = \{0, \dots, n\}$ ,  $\mathcal{A}$  est la tribu des parties de  $\mathcal{X}$ . Par commodité, on pourra prendre  $\Omega = \mathcal{X}$  et  $X$  est alors l'identité.

2. Supposons que l'on observe  $n$  données  $x_1, \dots, x_n$  qui chacune représente une mesure d'une quantité physique  $\mu$ , inconnue que l'on cherche à estimer. Chacune de ces données  $x_i$  est entachée d'une erreur due à la mesure. Faire des statistiques consiste à "modéliser" cette erreur, c'est à dire à considérer par exemple que  $x_i$  peut s'écrire  $\mu + e_i$  où  $e_i$  (l'erreur, qui est tout aussi inconnue que  $\mu$ ) est la **réalisation** d'une variable  $\varepsilon_i$ . De sorte que  $x_i$  est aussi la **réalisation** d'une variable  $X_i = \mu + \varepsilon_i$ .

Il est très important de faire la différence entre les variables que nous considérerons, d'un point de vue théorique pour construire ou valider des procédures et les réalisations de ces variables, qui sont les données numériques que l'on traite par le calcul ou en utilisant des logiciels.

Nous modéliserons ici les erreurs  $\varepsilon_i$  par des variables aléatoires indépendantes, identiquement distribuées de loi  $N(0, \sigma^2)$ , de sorte que  $X_1, \dots, X_n$  sont i.i.d.  $N(\mu, \sigma^2)$ .

On a vu que nous avons en fait souvent pris  $\Omega = \mathcal{X}$ ,  $\mathcal{F} = \mathcal{A}$  en considérant que  $X$  était l'identité. Dans ce cas nous résumerons la donnée du modèle statistique à  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$

### Echantillonnage

**Définition 3.** On appelle **modèle d'échantillonnage** associé au modèle  $(\Omega, \mathcal{F}, X, \mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  le modèle

$$(\Omega^n, \mathcal{F}_n, X_n, \mathcal{X}^n, \mathcal{A}_n, P_\theta^{\otimes n}, \theta \in \Theta)$$

où  $\mathcal{F}_n$  et  $\mathcal{A}_n$  sont les tribus produit respectivement sur  $\Omega^n$  et  $\mathcal{X}^n$ , et pour tout  $\theta \in \Theta$ ,  $P_\theta^{\otimes n}$  est la probabilité produit de  $n$  copies indépendantes de la loi  $P_\theta$ , notée aussi  $P_\theta^{\otimes n}$ . De plus, si  $\omega^n = (\omega_1, \dots, \omega_n)$  est un élément générique de  $\Omega^n$ ,  $X_n(\omega^n) = (X(\omega_1), \dots, X(\omega_n))$ .

### Exemples d'échantillonnage

1. En médecine ou en fiabilité on s'intéresse souvent au temps de 'survie' d'un individu ou d'une machine. Prenons le cas des machines : Supposons que nous disposions des temps de panne de  $n$  machines à laver de même marque. On peut faire l'hypothèse que ces machines n'étant pas reliées, leurs pannes sont indépendantes. Il s'agit ensuite de modéliser la loi d'un temps de panne. Plusieurs éventualités sont possibles. Nous allons en envisager deux très différentes.

i) Supposons d'abord que l'on fasse l'hypothèse que la machine ne s'use pas : nous avons alors pour tout  $a \leq b$ ,  $t \in \mathbb{R}^+$ ,

$$P(X \in [a + t, b + t] | X \geq t) = P(X \in [a, b] | X \geq 0)$$

On peut montrer que nécessairement, cette loi admet une densité de la forme :

$$f(x) = \lambda \exp -\lambda x.$$

On pourra considérer  $\Phi(x) = P(X > x)$  et montrer que si  $\Phi$  est continue alors le résultat est facile à obtenir.

C'est ce qu'on appelle une loi exponentielle de paramètre  $\lambda > 0$ . Notre modèle est alors un  $n$  échantillon du modèle  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), X, \mathbb{R}^+, \mathcal{B}(\mathbb{R}^+), P_\lambda, \lambda \in \mathbb{R}_*^+)$ ,  $X$  est l'identité et  $P_\lambda$  la loi exponentielle de paramètre  $\lambda$ .  $(\mathcal{B}(\mathbb{R}^+))$  est la tribu borelienne de  $\mathbb{R}^+$

ii) Supposons maintenant que notre machine ne puisse pas tomber en panne avant un temps connu ( $t = 1$ , par exemple). On prend en compte de cette façon le temps où la machine est sous garantie. On pourra alors considérer un modèle comme ci-dessus mais où  $P_\lambda$  est maintenant la loi de *Pareto* de paramètre  $\lambda > 0$  dont la fonction de répartition est donnée par :

$$G_\lambda(x) = P_\lambda(X \leq x) = 1 - x^{-\lambda}, \text{ si } x \geq 1, 0, \text{ sinon.}$$

*Exercice : Etudier le comportement de cette loi face au vieillissement.*

2. Un autre exemple très classique est le suivant : On observe  $(X_1, \dots, X_n)$   $n$ -variables aléatoires réelles identiquement distribuées de loi  $P$  sur  $\mathbb{R}$ , muni de  $\mathcal{B}(\mathbb{R})$ , sa tribu borelienne et on se propose d'estimer  $P$ , sans autres hypothèses sur  $P$ . Le modèle est alors un modèle d'échantillonnage où l'ensemble des paramètres  $\Theta$  est égal à l'ensemble de toutes les lois de probabilités sur  $\mathbb{R}$ .

**Modèles paramétriques, non-paramétriques** Comme on l'a vu précédemment  $\Theta$  est souvent un sous-ensemble d'un espace  $\mathbb{R}^d$ . Nous dirons quand c'est le cas que le modèle est **paramétrique**.

Le cas ci-dessus où  $\mathcal{X} = \mathbb{R}$ ,  $\mathcal{A}$  est sa tribu borelienne et  $\Theta$  est l'ensemble de toutes les mesures de probabilités sur  $(\mathcal{X}, \mathcal{A})$  est un exemple de modèle **non-paramétrique**.

## 1.2 Modèle Linéaire gaussien

**Définition 4.** Etant donné une matrice  $M$  de dimension  $n \times p$ , On appelle **modèle linéaire gaussien multidimensionnel** associé à la matrice "exogène"  $M$ , une observation  $Y$  dont la loi est  $N_n(M\beta, \sigma^2 I_n)$ .  $\beta$  est un paramètre inconnu de  $\mathbb{R}^p$ .

*Remarque :* On observe donc, à la fois le vecteur  $Y$  (aléatoire) et la matrice  $M$  supposée déterministe (non aléatoire). On cherche à utiliser cette observation pour tirer des informations sur le paramètre  $\beta$  inconnu. Le modèle précédent peut aussi s'écrire :

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, M = \begin{pmatrix} M_{11} & \dots & M_{1p} \\ \vdots & & \vdots \\ M_{n1} & \dots & M_{np} \end{pmatrix}, Y = M\beta + \varepsilon, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

où les  $\varepsilon_i$  sont i.i.d.  $N(0, \sigma^2)$ .  $\Delta$

Le modèle linéaire gaussien est un modèle statistique, dans lequel, on a

- $\mathcal{X} = \mathbb{R}^n$
- $\mathcal{A}$  est la tribu borelienne de  $\mathbb{R}^n$ .
- $\Theta = \{\theta = (\beta, \sigma^2) \in \mathbb{R}^p \otimes \mathbb{R}_+\}$
- $P_{(\beta, \sigma^2)} = N(M\beta, \sigma^2 I_n)$ .

### 1.2.1 Exemples de modèles linéaires

1. Le modèle précédent de mesure d'une quantités physique est un modèle linéaire.
2. Comparaison de 2 populations de même variance : Supposons que l'on dispose de 2 échantillons  $X_1, \dots, X_n$  i.i.d.  $N(\mu_1, \sigma^2)$  et  $X'_1, \dots, X'_m$  i.i.d.  $N(\mu_2, \sigma^2)$  indépendants. On se demande si ces échantillons sont comparables, autrement dit est-ce que  $\mu_1 = \mu_2$  ?  
On concatène les 2 échantillons pour former le vecteur

$$Y = (X_1, \dots, X_n, X'_1, \dots, X'_m)^* = (Y_1, \dots, Y_{m+n})^*$$

Si on considère la matrice  $M$  de taille  $n \times 2$ , telle que

$$M_{11} = \dots = M_{n1} = 1, M_{n+1,1} = \dots = M_{n+m,1} = 0$$

$$M_{12} = \dots = M_{n2} = 0, M_{n+1,2} = \dots = M_{n+m,2} = 1$$

et le vecteur  $\beta = (\mu_1, \mu_2)^*$ , il est facile de mettre notre modèle sous la forme (1.2).

3. Droite de régression. Supposons que l'on sache par des arguments théoriques ( agronomiques, biologiques, économiques, physiques,...) que 2 quantités  $x$  (par exemple le temps) et  $y$  (par exemple la taille d'un animal) sont liées par une équation affine de la forme  $y = ax + b$ , dont on veut identifier les coefficients  $a$  et  $b$ . Une façon de procéder est de mesurer  $y_i$  pour différentes valeurs de  $x_i$  (appelée variable contrôlée) et de modéliser les erreurs par des  $N(0, \sigma^2)$  indépendantes. On a alors la représentation (1.2), avec

$$M_{11} = x_1, \dots, M_{n1} = x_n,$$

$$M_{12} = 1, \dots, M_{n2} = 1,$$

$$\beta = (a, b)^*$$

Cet exemple peut se généraliser en remplaçant la relation affine par une relation de la forme :

$$y = \sum_{j=0}^p \beta_j f_j(x)$$

Une régression polynomiale s'obtient par exemple en prenant

$$f_0 = 1, f_1(x) = x, \dots, f_p(x) = x^p$$

4. On appelle **Analyse de la variance** (Anova) le cas où la matrice  $M$  est uniquement constituée de 1 et de 0. Donnons un exemple : Dans des conditions de culture de référence (0), une variété de blé a un rendement moyen de  $\mu$ . On la soumet, dans des parcelles expérimentales à un traitement à 2 facteurs :
- 1er facteur (par exemple, un engrais) auquel, outre le niveau 0 de référence, on donne 2 niveaux, notés 1 et 2 (par exemple, 2 doses différentes d'engrais).
  - 2eme facteur (par exemple, un niveau d'ensoleillement) auquel on donne soit le niveau de référence 0 soit le niveau 1.
- Le modèle de base choisi est le suivant :

$$y = \mu + \alpha_i + \beta_j \tag{1.1}$$

Il est dit additif : Les effets des facteurs s'ajoutent simplement sans interférences.  $\alpha_i$  représente l'effet du 1er facteur au niveau  $i = 0, 1, 2$ ,  $\beta_j$  représente l'effet du 2eme facteur au niveau  $j = 0, 1$ .  $\alpha_0 = \beta_0 = 0$ . Il est clair qu'on aurait pu aussi rajouter "une interaction" de la forme  $\gamma_{ij}$ , mais par souci de simplicité, nous ne l'avons pas fait ici.

Le but est d'obtenir des informations (estimation ou test) sur les  $\alpha_i$  et les  $\beta_j$ . Pour cela, on réalise une expérimentation : On divise un champs en parcelles numérotées (6, dans l'exemple qui suit). Sur chaque parcelle, on applique les facteurs à un niveau prescrit. La description des niveaux affectés aux parcelles s'appelle le plan de l'expérience. Ici, il est donné par le tableau suivant.

Parcelle	1	2	3	4	5	6
Facteur 1	0	1	2	0	1	0
Facteur 2	0	0	0	0	0	1

Si l'on suppose que l'on modélise le rendement sur chaque parcelle par un effet de type (1.1) auquel s'ajoute une erreur  $N(0, \sigma^2)$ , et si l'on suppose les erreurs indépendantes, on obtient une équation du type  $Y = M\beta + \varepsilon$ , où  $Y$  est le vecteur des rendements,  $\varepsilon$  est le vecteur des erreurs,  $\beta = (\mu, \alpha_1, \alpha_2, \beta_1)^*$  et  $M$  est la matrice suivante

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

*Exercice* : Sous quelle condition un modèle linéaire est-il un modèle d'échantillonnage ?  $\triangle$

### 1.3 Identifiabilité, Domination

**Définition 5.** On dit qu'un modèle  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  en position canonique est **identifiable** si l'application de  $\Theta$  dans l'espace des probabilités sur  $(\mathcal{X}, \mathcal{A})$  qui à  $\theta$  associe  $P_\theta$  est injective.

La définition précédente est assez naturelle. On conçoit en effet avoir fort peu de chance d'identifier  $\theta$ , au vu d'une observation  $X$  dont la loi est  $P_\theta^X$  s'il n'y a pas injectivité. Notons qu'un modèle donné n'est pas automatiquement identifiable. Nous en verrons des exemples plus loin, mais il suffit de reprendre l'exemple ci-dessus de l'ANOVA en modèle additif à 2 facteurs. Nous avons fait l'hypothèse  $\alpha_0 = \beta_0 = 0$ . Cette hypothèse avait en fait pour but de rendre le modèle identifiable. Montrer en effet (en exercice) que si les paramètres  $\alpha_0$  et  $\beta_0$  sont inconnus, alors, le modèle précédent n'est pas identifiable.



**Définition 6.** On dit qu'un modèle en position canonique  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  est **dominé** s'il existe une mesure  $\mu$  sur  $(\mathcal{X}, \mathcal{A})$ , positive,  $\sigma$ -finie telle que pour tout  $\theta \in \Theta$ ,  $P_\theta$  est dominée par  $\mu$ .  $\mu$  est appelée mesure dominante du modèle.

Si la définition précédente était assez naturelle, celle-ci l'est un peu moins. Disons seulement ici que sans cette propriété, beaucoup de résultats seront faux. Nous verrons ( et heureusement ) que beaucoup de modèles 'standards' sont dominés. Ce n'est toutefois pas toujours le cas. Le premier exemple ci-dessous montre que l'ensemble de toutes les lois sur  $\mathbb{R}, \mathcal{B}(\mathbb{R})$  n'est pas dominé.

*Remarques :*

1. Rappelons les notions utilisées dans la définition précédente : On dit qu'une mesure positive  $\mu$  sur  $(\mathcal{X}, \mathcal{A})$  est  $\sigma$ -finie si il existe une suite  $A_n$  d'ensembles de  $\mathcal{A}$  telle que  $\mathcal{X} = \cup_n A_n$  et  $\mu(A_n) < \infty, \forall n$ .
2. On dit que la mesure  $\mu$  domine  $P_\theta$  (noté  $P_\theta \ll \mu$ ), si pour tout  $A$  de  $\mathcal{A}$ ,  $\mu(A) = 0$  entraîne  $P_\theta(A) = 0$ .
3. La domination est une relation transitive. On en déduit donc que si un modèle est dominé, il l'est par une infinité de mesures positives. (En particulier toutes les mesures qui dominent  $\mu$ .)
4.  $P_\theta \ll \mu$  équivaut, d'après le théorème de Radon-Nykodym à l'existence d'une densité de  $P_\theta$  par rapport à  $\mu$ ,  $\frac{dP_\theta}{d\mu}$ , définie  $\mu$  presque sûrement, telle que pour toute fonction  $F$ ,  $P_\theta$ -intégrable on a :

$$\int_{\mathcal{X}} F(x) dP_\theta = \int_{\mathcal{X}} F(x) \frac{dP_\theta}{d\mu} d\mu$$

5. Dans un modèle dominé, on peut définir des fonctions, appelées fonctions de vraisemblance, de  $\mathcal{X} \otimes \Theta$  dans  $\mathbb{R}_+$  qui à  $(x, \theta)$  associent

$$p(x, \theta) = \frac{dP_\theta}{d\mu}(x)$$

△

### Identifiabilité, domination : Exemples

1. Si  $\delta_x$  désigne la masse de Dirac au point  $x$  et  $\mathcal{B}(E)$  la tribu borelienne de l'espace topologique  $E$ , on peut remarquer que

$$(\mathbb{R}, \mathcal{B}(\mathbb{R}), \delta_\theta, \Theta)$$

est dominé si et seulement si  $\Theta$  est un ensemble dénombrable de  $\mathbb{R}$ .

En effet si  $\Theta = \{\theta_i, i \in \mathbb{N}^*\}$  est dénombrable, alors  $\mu = \sum_{\Theta} \delta_\theta$  domine le modèle. (On montre que cette mesure est  $\sigma$ -finie en décomposant  $\mathbb{R}$  en l'union des  $A_i = \{\theta_i\}, i \in \mathbb{N}^*$ , de  $\mu$ - mesure positive, et  $A_0 = \mathbb{R} \setminus (\cup_{i \in \mathbb{N}^*} \{\theta_i\})$  de  $\mu$ - mesure nulle.)

Si  $\Theta$  n'est pas dénombrable, et si  $\mu$  est une mesure dominante du modèle, alors si on considère  $\frac{d\delta_\theta}{d\mu}$ , la densité de  $\delta_\theta$  par rapport à  $\mu$ , on a

$$1 = \delta_\theta(\theta) = \frac{d\delta_\theta}{d\mu}(\theta) \mu(\{\theta\})$$

Ceci entraîne que  $\forall \theta \in \Theta, \mu(\{\theta\}) \neq 0$ .

Nous avons le lemme suivant :

**Lemme 1.** Si  $B$  est tel que  $\mu(B) < +\infty$  et  $\mu\{x\} > 0, \forall x \in B$ ,  $B$  est nécessairement un ensemble dénombrable.

En effet, soit  $A = \{x, \mu\{x\} > 0\}$ ,  $A = \cup_{k>0} E_k, E_k = \{x, \mu\{x\} \geq \frac{1}{k}\}$ . On a  $\mu(E_k) \geq \frac{1}{k} \text{card}\{E_k\}$ . Et donc,  $\mu(E_k) < \infty$  implique que  $E_k$  est un ensemble fini. Donc si  $B \subset A$  est tel que  $\mu(B) < \infty$ , nécessairement  $B$  est union dénombrable d'ensembles finis donc dénombrable.

On en déduit que  $\mu$  n'est pas donc pas  $\sigma$ -finie. En effet, si on considère un recouvrement de  $\mathbb{R}$  par des ensembles mesurables  $\mathcal{A}_n$  tels que  $\mu(\mathcal{A}_n) < \infty$ , on en déduit un recouvrement de  $\Theta$  par une suite d'ensembles de mesure finie et chargeant chaque points, donc dénombrables en utilisant le lemme précédent. Ceci conduit au fait que  $\Theta \subset \mathbb{R}$  est dénombrable. Une conséquence, plus intéressante en pratique est que le modèle non paramétrique cité plus haut de toutes les probabilités sur  $\mathbb{R}$  n'est pas dominé.

2. Le modèle linéaire gaussien  $Y = M\beta + \varepsilon$  est un modèle dominé (on peut prendre  $\mu$  la mesure de Lebesgue sur  $\mathbb{R}^n$ ). Il est identifiable si et seulement si  $M$  est injective (exercice). Une vraisemblance est :

$$\frac{dP_\theta}{d\mu}(y) = \frac{\exp\left\{-\frac{\|y - M\beta\|^2}{2\sigma^2}\right\}}{(\sqrt{2\pi}\sigma)^n}$$

3. Considérons l'exemple du sondage : Soit  $N$  et  $n$  fixés. On considère une population de  $N$  éléments qui comprend une proportion  $\theta$  de defectueux. On extrait au hasard  $n$  éléments et on compte le nombre de defectueux parmi cette population extraite. La loi  $P_\theta$  de ce nombre de defectueux est une hypergéométrique que l'on peut écrire sous la forme :

$$P_\theta = \sum_{k=0}^n \frac{C_{\theta N}^k C_{(1-\theta)N}^{n-k}}{C_N^n} \delta_k$$

Ce modèle est dominé par la mesure  $\mu = \sum_{k=0}^n \delta_k$  et une vraisemblance est

$$\frac{dP_\theta}{d\mu}(k) = \frac{C_{\theta N}^k C_{(1-\theta)N}^{n-k}}{C_N^n}$$

Montrer en exercice que ce modèle est identifiable et que ça n'est pas un modèle d'échantillonnage.

4. Montrer en exercice qu'un modèle est identifiable (resp. dominé) si et seulement si le modèle échantillonné l'est.
5. On observe  $n$  sites de pontes. Pour chaque site, le nombre d'oeufs suit une loi de Poisson de paramètre  $\lambda$ . Chaque oeuf a une probabilité  $p$  de donner un insecte. On considère que les oeufs sont indépendants entre eux, indépendants du nombre  $X$  et que les sites de pontes sont aussi indépendants entre eux. On peut faire 2 types d'observations
- Soit on observe  $X$  le nombre d'oeufs et  $Y$  le nombre d'insectes (observation complète), sur chaque site de ponte.
  - Soit on observe seulement  $Y$  le nombre d'insectes (observation partielle), également sur chaque site de ponte.
- On a  $\Theta = \{(\lambda, p) \in \mathbb{R}_+^* \otimes [0, 1]\}$
- Dans le premier cas, l'espace  $\mathcal{X} = (\Lambda_1, \dots, \Lambda_n)$  avec  $\Lambda_i = \mathbb{N} \otimes \mathbb{N}$ , muni de la tribu des parties.  $P_\theta^{\otimes n}$  est la loi de  $n$  copies indépendantes de la loi  $P_\theta$  du couple  $(X, Y)$ .
- Calculons cette loi : On a

$$P_\theta(Y = y | X = x) = C_x^y p^y (1-p)^{x-y} \mathbf{1}_{0 \leq y \leq x}, \quad P_\theta(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

$$P_\theta(Y = y, X = x) = C_x^y p^y (1-p)^{x-y} \mathbf{1}_{0 \leq y \leq x} e^{-\lambda} \frac{\lambda^x}{x!}$$

On obtient donc,

$$P_\theta^{\otimes n}((x_1, y_1), \dots, (x_n, y_n)) = \prod_{i=1}^n C_{x_i}^{y_i} p^{y_i} (1-p)^{x_i - y_i} \mathbf{1}_{0 \leq y_i \leq x_i} e^{-\lambda} \frac{\lambda^{x_i}}{x_i!}$$

L'exercice ci-dessus montre qu'il nous suffit d'étudier le cas  $n = 1$ . Le modèle est dominé, nous laissons la démonstration au lecteur. Pour vérifier que le modèle est identifiable, prenons  $\lambda, p, \lambda', p'$  et supposons que

$$C_x^y p^y (1-p)^{x-y} e^{-\lambda} \frac{\lambda^x}{x!} = C_x^y p'^y (1-p')^{x-y} e^{-\lambda'} \frac{\lambda'^x}{x!} \quad \forall x \geq 0, 0 \leq y \leq x \quad (1.2)$$

Supposons d'abord que  $p$  et  $p'$  sont dans  $]0, 1[$ . Alors l'équation (1.2) devient :

$$\left(\frac{p(1-p')}{p'(1-p)}\right)^y \left(\frac{\lambda(1-p)}{\lambda'(1-p')}\right)^x = \exp\{(\lambda - \lambda')\}$$

que l'on peut encore écrire  $A^y B^x = \exp\{(\lambda - \lambda')\} \quad \forall x \geq 0, 0 \leq y \leq x$ . En passant au logarithme, on vérifie aisément que ceci n'est possible que si  $A = B = 1, \lambda' = \lambda$ .

Remarquons maintenant que si  $p = 0$ , si on se place en  $y = 0$ , (1.2) s'écrit :

$$e^{\lambda - \lambda'} \left[ \frac{(1 - p')\lambda'}{\lambda} \right]^x = 1, \forall x \geq 0$$

En passant au logarithme, on vérifie aisément que ceci implique  $p' = 0$ ,  $\lambda = \lambda'$ .

On fait un raisonnement analogue pour le cas  $p = 1$ . On en conclut que le modèle est aussi identifiable pour l'ensemble des paramètres  $\Theta$ .

6. Si nous considérons maintenant le cas où on observe seulement  $Y$  l'espace est maintenant  $\mathcal{X} = (\Lambda_1, \dots, \Lambda_n)$  avec  $\Lambda_i = \mathbb{N}$ , muni de la tribu des parties.  $Q_\theta^{\otimes n}$  est la loi de  $n$  copies indépendantes de la loi  $Q_\theta$  de la variable  $Y$ . On a

$$\begin{aligned} Q_\theta(Y = y) &= \sum_{y \leq x} e^{-\lambda} \frac{\lambda^x}{x!} C_x^y p^y (1-p)^{x-y} \mathbf{1}_{0 \leq y} \\ &= \frac{e^{-\lambda} \lambda^y p^y}{y!} \sum_{y \leq x} \frac{((1-p)\lambda)^{(x-y)}}{(x-y)!} \\ &= \frac{e^{-\lambda p} (p\lambda)^y}{y!} \end{aligned}$$

On en déduit que  $Q_\theta$  est une loi de Poisson de paramètre  $\lambda p$ . Le modèle est clairement non identifiable.

*Remarque* : On peut retrouver la loi de  $Y$  en remarquant que

$$Y = \sum_{i=1}^x Z_i$$

où les  $Z_i$  sont i.i.d. de loi de Bernoulli de paramètre  $p$  (i.e.  $P(Z_i = 1) = p = 1 - P(Z_i = 0)$ ) et indépendantes de  $X$ . Calculer la fonction génératrice  $\mathbb{E}s^Y$  pour  $s$  réel de module strictement inférieur à 1.

△

## 1.4 Structure des modèles dominés.

Le but de cette section est d'abord de montrer que parmi les mesures dominantes, certaines peuvent avoir des propriétés particulières. Nous allons ensuite nous attacher à donner une formulation mathématique à l'observation que l'on a pu faire au travers des exemples précédents : une famille dominée est une famille 'pas trop grande' en un sens à déterminer.

### 1.4.1 Dominante Privilégiée

Nous avons vu que les mesures dominantes ne sont pas uniques. Toutefois certaines sont plus intéressantes que d'autres.

**Proposition 1.** *Si un modèle  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  est dominé par une mesure  $\mu$ , il est dominé par une probabilité  $Q$ , vérifiant de plus  $Q(A) = 0 \iff P_\theta(A) = 0, \forall \theta \in \Theta$ . Une telle probabilité est appelée **dominante privilégiée** du modèle.*

*Remarque* : Pas plus qu'une mesure dominante, une probabilité dominante privilégiée n'est unique. N'importe qu'elle probabilité équivalente à la probabilité de départ convient. △

*Preuve de la Proposition :*

La démonstration de la proposition repose sur le lemme suivant :

**Lemme 1.** *Toute mesure  $\sigma$ -finie est dominée par une probabilité.*

*Démonstration du lemme :*

Nous savons qu'il existe des ensembles mesurables,  $A_n$ ,  $n \in \mathbb{N}$  tels que  $\mathcal{X} = \cup_n A_n$  et  $\mu(A_n) < \infty$ ,  $\forall n \in \mathbb{N}$ . Soit  $\lambda_n$  une suite de réels tels que  $\lambda_n \in [0, 1]$ ,  $\sum \lambda_n = 1$  et  $\lambda_n > 0 \iff \mu(A_n) > 0$ . Soit  $P$  la mesure de densité par rapport à  $\mu$  :

$$\frac{dP}{d\mu}(x) = \sum_{n \in \mathbb{N}} \frac{\lambda_n}{\mu(A_n)} 1_{A_n}(x)$$

Il est clair que  $P(A) = \sum_{n \in \mathbb{N}} \frac{\lambda_n \mu(A \cap A_n)}{\mu(A_n)}$  donc  $P$  est une probabilité. Par ailleurs  $P$  est équivalente à  $\mu$  : En effet

$$P(A) = 0 \iff \mu(A \cap A_n) = 0, \forall n \iff \mu(A) = 0.$$

□

Passons maintenant à la démonstration de la proposition : En utilisant le lemme, considérons  $P$  probabilité, équivalente à  $\mu$ .  $P$  domine toutes les  $P_\theta$  par transitivité, mais il va nous falloir travailler plus pour fabriquer une dominante privilégiée  $Q$ . Posons  $F_\theta = \frac{dP_\theta}{dP}$  et  $A_\theta = \{F_\theta > 0\} \in \mathcal{A}$ . Considérons  $\mathcal{C} \subset \mathcal{A}$ , l'ensemble des réunions dénombrables d'éléments  $A_\theta$  et

$$M = \sup_{C \in \mathcal{C}} P(C).$$

On vérifie que  $M \leq 1$ , et que  $M$  est atteinte : Il existe une suite  $C_n$ , telle que  $P(C_n) \geq M - n^{-1}$ . Posons  $C^* = \cup_n C_n$ ,  $C^*$  appartient à  $\mathcal{C}$  puisque cet ensemble est stable par réunion dénombrable. On a  $M \leq P(C^*) \leq M$ . Par ailleurs, toujours par définition de  $\mathcal{C}$ , il existe une suite  $\{\theta_j, j \in \mathbb{N}\}$  telle que  $C^* = \cup_j A_{\theta_j}$ . Soit  $\lambda_n$  une suite de réels,  $\lambda_n \in ]0, 1]$ ,  $\sum \lambda_n = 1$ , Définissons la mesure  $Q$  dont la densité par rapport à  $P$  est

$$\frac{dQ}{dP}(x) = \sum_{j \in \mathbb{N}} \lambda_j F_{\theta_j}(x).$$

Il est évident  $Q(A) = \sum_{j \in \mathbb{N}} \lambda_j P_{\theta_j}(A)$  donc  $Q$  est une probabilité et si  $A \in \mathcal{A}$  est tel que  $P_\theta(A) = 0, \forall \theta \in \Theta$  alors  $Q(A) = 0$ .

Le point difficile est de démontrer que  $Q$  est dominante. Remarquons d'abord que  $C^*$  a la propriété fondamentale suivante (support maximal) :

$$\forall \theta \in \Theta, P(A_\theta) = P(A_\theta \cap C^*) \quad (1.3)$$

En effet, sinon,  $P(A_\theta \cap (C^*)^c)$  serait strictement positif et  $P(A_\theta \cup C^*) = P(C^*) + P(A_\theta \cap (C^*)^c)$  serait strictement plus grand que  $M$ .

Considérons  $A$  tel que  $Q(A) = 0$ , on a, bien entendu  $P_{\theta_j}(A) = 0 \forall j \in \mathbb{N}^*$ , mais montrons qu'on a bien  $P_\theta(A) = 0 \forall \theta \in \Theta$  :

$$P_\theta(A) = \int_A F_\theta dP = \int_{A \cap A_\theta} F_\theta dP = \int_{A \cap A_\theta \cap C^*} F_\theta dP$$

En effet, à cause de (1.3),  $P(A_\theta \cap (C^*)^c) = 0$  et donc  $P_\theta(A_\theta \cap (C^*)^c) = 0$  par domination. Par conséquent,  $P_\theta(A \cap A_\theta \cap (C^*)^c) = 0$ . On en déduit,

$$\begin{aligned} P_\theta(A) &= \int_{A \cap A_\theta \cap (\cup_j A_{\theta_j})} F_\theta dP \\ &\leq \sum_j \int_{A \cap A_\theta \cap A_{\theta_j}} F_\theta dP \\ &= \sum_j \int_{A \cap A_\theta \cap A_{\theta_j}} \frac{F_\theta}{F_{\theta_j}} F_{\theta_j} dP \\ &= \sum_j \int_{A \cap A_\theta \cap A_{\theta_j}} \frac{F_\theta}{F_{\theta_j}} dP_{\theta_j} \end{aligned}$$

Or, par définition  $Q(A) = 0 \iff P_{\theta_j}(A) = 0 \forall j \in \mathbb{N}$ . En conséquence la somme précédente est nulle. □

### 1.4.2 Distance en variation et Modèles dominés

**Définition 7.** Soit  $(\mathcal{X}, \mathcal{A})$  un ensemble et sa tribu et désignons par  $\mathcal{P}$  l'ensemble de toutes les probabilités sur  $(\mathcal{X}, \mathcal{A})$ . On peut munir  $\mathcal{P}$  de la **distance en variation** définie comme suit :

$$d(P, Q) = \sup_{A \in \mathcal{A}} |P(A) - Q(A)|$$

*Remarque :* Il est clair que  $d$  est une distance :

- $d(P, Q) = d(Q, P)$ .
- $d(P, Q) = 0 \iff P = Q$ .
- $d(P, Q) \leq d(P, R) + d(R, Q)$ .

△

**Lemme 2.** Si  $S$  domine  $P$  et  $Q$ , soit  $p = \frac{dP}{dS}$ ,  $q = \frac{dQ}{dS}$ , alors

$$d(P, Q) = \frac{1}{2} \int |p - q| dS = \int_{p > q} (p - q) dS = \int_{q > p} (q - p) dS$$

*Démonstration du lemme*

- Remarquons tout d'abord qu'une telle mesure existe toujours : Il suffit de prendre  $S = P + Q$ .
- Nous avons  $0 = \int (p - q) dS = \int_{p > q} (p - q) dS - \int_{q > p} (q - p) dS$ .  
De plus,  $\int |p - q| dS = \int_{p > q} (p - q) dS + \int_{q > p} (q - p) dS$ .
- Maintenant,  $P(A) - Q(A) = Q(A^c) - P(A^c)$ , ce qui implique :  
 $|P(A) - Q(A)| = \frac{1}{2} (|P(A) - Q(A)| + |Q(A^c) - P(A^c)|)$   
 $= \frac{1}{2} (|\int_A dP - dQ| + |\int_{A^c} dP - dQ|)$ .  
On en déduit :  $|P(A) - Q(A)| \leq \frac{1}{2} (\int_A |p - q| dS + \int_{A^c} |p - q| dS)$ , soit,  
 $d(P, Q) \leq \frac{1}{2} \int |p - q| dS$ .
- Soit maintenant  $A = \{p > q\}$ , on a  $d(P, Q) \geq |P(A) - Q(A)| = \int_{p > q} (p - q) dS$ .

□

**Proposition 2.** On a :

- $0 \leq d(P, Q) \leq 1$ .
- $d(P, Q) = 1 \iff P$  et  $Q$  sont étrangères. (i.e. il existe  $A \in \mathcal{A}$ , tel que  $P(A) = 1, Q(A) = 0$ .)

*Démonstration de la Proposition*

- D'abord on a  $0 \leq d(P, Q) = \frac{1}{2} \int |p - q| dS \leq \frac{1}{2} \int (p + q) dS$ .
- Supposons maintenant :  
 $1 = d(P, Q) = \int_{p \geq q} (p - q) dS = \int_{p \geq q} p dS - \int_{p \geq q} q dS$ .  
Les 2 quantités de cette différence étant comprises entre 0 et 1, ceci n'est possible que si  $\int_{p \geq q} p dS = 1$  et  $\int_{p \geq q} q dS = 0$ . Il suffit alors de prendre  $A = \{p \geq q\}$ , et de vérifier qu'on a bien  $P(A) = 1, Q(A) = 0$ .

**Théorème 1.** Soit  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  un modèle statistique.

- Si la famille  $\{P_\theta, \theta \in \Theta\}$  est séparable pour la distance en variation, alors le modèle est dominé.
- Si la tribu  $\mathcal{A}$  est dénombrablement engendrée, la réciproque est vraie.

*Remarques :*

1. Un ensemble est dit séparable pour une métrique, s'il contient un ensemble dénombrable dense pour cette métrique.
2. Une tribu est dite dénombrablement engendrée si elle s'écrit  $\sigma(A_n, n \in \mathbb{N})$ . △

**Corollaire 1.** Si  $\Theta$  peut être muni d'une métrique pour laquelle il est séparable, et si l'application  $\phi$  de  $\Theta$  dans  $\mathcal{P}$ , muni de la distance en variation

$$\theta \in \Theta \xrightarrow{\phi} P_\theta \in \mathcal{P}$$

est continue, alors le modèle est dominé.

La démonstration du corollaire est triviale si on remarque que l'image d'un espace séparable par une application continue est séparable.

**Démonstration du Théorème**

- Supposons donc  $(P_\theta, \theta \in \Theta, d)$  séparable. Cela signifie qu'il existe une famille  $(\theta_n)_{n \in \mathcal{N}^*}$  dense pour la métrique  $d$ . Considérons la mesure de probabilité

$$S = \sum_{n \in \mathcal{N}^*} \frac{1}{2^n} P_{\theta_n}$$

et vérifions que  $S$  est une dominante privilégiée.

En effet, soit  $A \in \mathcal{A} / S(A) = 0$ , on a  $P_{\theta_n}(A) = 0 \forall n \in \mathcal{N}^*$  par construction de  $S$ .

Montrons que ceci implique que  $P_\theta(A) = 0 \forall \theta \in \Theta$  :

En effet, sinon soit  $\theta$  tel que  $P_\theta(A) > \alpha > 0$ . Par densité, il existe,  $\theta_n$  tel que

$d(P_\theta, P_{\theta_n}) \leq \alpha/2$ , mais ceci implique que  $|P_\theta(A) - P_{\theta_n}(A)| \leq \alpha/2$ . Nous aboutissons à une contradiction.

- Pour démontrer la réciproque, nous laissons en exercice, le résultat suivant (la construction est usuelle et repose sur la considération des fonctions  $\{\sum_{i=1}^N \alpha_i 1_{A_i}, \alpha_i \in \mathcal{Q}, N \in \mathcal{N}\}$  :

*Si  $\mathcal{A}$  est dénombrablement engendrée et  $\mu$   $\sigma$ -finie alors  $\mathcal{L}_1(\mu) = \{f / \int |f| d\mu < \infty\}$ , est un ensemble séparable.*

Supposons que  $S$  domine la famille  $\{P_\theta, \theta \in \Theta\}$ , alors  $\{P_\theta, \theta \in \Theta\}$  s'injecte naturellement dans  $\mathcal{L}_1(S)$ . En effet si on note  $p_\theta = \frac{dP_\theta}{dS}$ , on a  $2d(P_\theta, P_{\theta'}) = \int |p_\theta - p_{\theta'}| dS$ . Or cette dernière quantité est exactement la norme dans  $\mathcal{L}_1(S)$  de  $p_\theta - p_{\theta'}$ . La séparabilité de  $\{P_\theta, \theta \in \Theta\}$  se déduit alors de celle de  $\mathcal{L}_1(S)$ .

**1.5 Introduction à l'exhaustivité.**

Ce problème concerne les expériences où l'on dispose de beaucoup de données, et que (pour des problèmes de stockage par exemple), on cherche à réduire la taille de ces données. C'est souvent le cas par exemple en traitement de l'image : Une image peut être la donnée de  $256 \times 256$  niveaux de gris par "pixels" (i.e. picture elements : éléments d'image). Cette donnée définit une qualité d'image. Toutefois, si l'on désire transmettre cette image rapidement, il peut être utile de résumer cette donnée. Bien entendu, on souhaite généralement que cette réduction se fasse avec le moins de perte possible au niveau de la qualité de l'image. La notion de perte de qualité est relativement difficile à définir de façon quantitative. Nous allons étudier maintenant la notion d'exhaustivité, qui correspond à une réduction de la taille des données, sans perte d'information statistique.

**1.5.1 Sous-Expérience, perte d'information.**

Sans aller plus loin dans la formalisation à ce niveau, disons que nous allons considérer que 'l'information' que nous amène une expérience est l'ensemble des 'statistiques' (cf définition) qu'elle permet de construire.

**Définition 8.**  $\mathcal{E} = (\Omega, \mathcal{F}, X, \mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  étant une expérience, et  $T$  une application mesurable de  $(\mathcal{X}, \mathcal{A})$  dans  $(\Pi, \mathcal{T})$ , on appelle **sous expérience** associée à  $T$ , l'expérience :

$$\mathcal{E}^T = (\Omega, \mathcal{F}, T \circ X, \Pi, \mathcal{T}, P_\theta, \theta \in \Theta)$$

*Remarques :*

1. Un exemple consiste dans le cas du modèle d'échantillonnage où l'on observe  $(X_1, \dots, X_n)$  à prendre  $T$  telle que  $T(x_1, \dots, x_n) = x_1$ . Les statistiques de l'expérience  $\mathcal{E}$  sont calculées à partir de  $(X_1, \dots, X_n)$ . En revanche, celles de la sous expérience  $\mathcal{E}^T$  ne sont calculées qu'à partir de  $X_1$ , de sorte que la moyenne  $\bar{X}$  par exemple n'est plus accessible à partir de  $\mathcal{E}^T$ . On voit qu'il peut en résulter une perte d'information et des pertes de précisions statistiques importantes.
2. Dans le paragraphe précédent nous avons rencontré un problème de sous expérimentation qui se pose naturellement dans l'exemple des sites de ponte.
3. L'exemple suivant est aussi très important en statistique. Il s'agit de la **réduction de données en classes**. L'expérience originale est un échantillonnage de taille  $n$ , chaque variable  $X_i$  étant à valeurs dans un ensemble  $X, \mathcal{F}$ . Soit  $A_1, \dots, A_k$  une partition mesurable de  $X, \mathcal{F}$ . Supposons maintenant qu'au lieu d'observer chaque  $X_i$  on sous expérimente en n'observant que la classe  $A_l$  dans laquelle il est tombé. i.e. on observe  $T(X_i) = \sum_{l=1}^k I\{X_i \in A_l\}$ . On peut encore réduire l'information en n'observant que :  $(N_1, \dots, N_k)$ , où  $N_l = \sum_{i=1}^n I\{X_i \in A_l\}$  compte le nombre de  $X_i$  qui sont tombés dans chaque classe.

4. La question fondamentale de l'exhaustivité est la suivante : on dira que  $T$  est exhaustive si l'expérience  $\mathcal{E}^T$  est aussi informative que  $\mathcal{E}$ . On peut alors naturellement se demander à quelles conditions sur  $T$  une telle propriété est réalisée.

Il y a un cas où le problème est clair : S'il existe  $U$ , mesurable de  $(\Pi, \mathcal{T})$  dans  $(\mathcal{X}, \mathcal{A})$  tel que  $U(T(x)) = x, \forall x \in \mathcal{X}$ . Dans ce cas,  $\mathcal{E}^T$  est une sous expérience de  $\mathcal{E}$ , mais  $\mathcal{E} = (\mathcal{E}^T)^U$  est aussi une sous expérience de  $\mathcal{E}^T$ . Les 2 expériences sont donc clairement équivalentes au sens où elles permettent de calculer les mêmes statistiques.  $\triangle$

Néanmoins, il existe des cas intéressants où on n'a pas l'existence de  $U$ , mesurable de  $(\Pi, \mathcal{T})$  dans  $(\mathcal{X}, \mathcal{A})$  tel que  $U(T(x)) = x, \forall x \in \mathcal{X}$ , et où on n'a pas perte d'information statistique.

Étudions l'exemple important suivant :

Rappelons que l'on appelle **modèle multinomial** le modèle où

- l'observation  $X$  est à valeurs dans  $\{(n_1, \dots, n_k), n_l \in \mathcal{N}, \sum_{l=1}^k n_l = n\}$  muni de la tribu de ses parties.
- $\theta = (\theta_1, \dots, \theta_k) \in \Theta = \{(\theta_1, \dots, \theta_k), \theta_l \in [0, 1], \sum_{l=1}^k \theta_l = 1\}$
- $P_\theta\{(n_1, \dots, n_k)\} = \frac{n!}{n_1! \dots n_k!} \theta_1^{n_1} \dots \theta_k^{n_k}$ .

Ce modèle provient généralement comme on l'a indiqué dans l'exemple plus haut d'une sous expérimentation qui n'observe pas quels individus sont tombés dans telle classe, mais ne fait que les compter. Nous allons montrer qu'en fait ce genre de sous expérimentation ne réduit pas l'information statistique. Plus précisément, prenons pour simplifier le cas où le nombre de classes  $k = 2$ .

Soit  $(X_1, \dots, X_n)$  un n-échantillon de la loi de Bernoulli de paramètre  $\theta \in [0, 1]$ . Nous allons considérer que 2 types d'observation sont possibles :

1. Soit on observe effectivement  $(X_1, \dots, X_n)$ , et on a

$$\mathcal{E} = (\{0, 1\}^n, \mathcal{A}_n, B(\theta)^{\otimes n}, \theta \in [0, 1]).$$

2. Soit on observe seulement un compteur  $T = \sum_{i=1}^n X_i$  et on a la sous-expérience

$$\mathcal{E}^T = (\{0, \dots, n\}, \mathcal{B}_n, \text{Bin}(n, \theta), \theta \in [0, 1]),$$

où  $\text{Bin}(n, \theta)$  est la loi binomiale (i.e.  $P_\theta(k) = C_n^k \theta^k (1 - \theta)^{n-k} \mathbf{1}_{0 \leq k \leq n}$ ).

$\mathcal{E}^T$  est clairement une sous expérience de  $\mathcal{E}$  mais l'application  $T$  n'a clairement pas d'inverse. Nous allons maintenant montrer que s'il ne nous est pas possible de reconstruire  $(X_1, \dots, X_n)$  à partir de  $T$ , en revanche, il est possible de reconstruire à partir de  $T$ .  $(Y_1, \dots, Y_n)$  de même **loi que**  $(X_1, \dots, X_n)$ . C'est à dire qu'il nous est possible de reconstruire l'expérience  $\mathcal{E}$ .

Considérons la procédure suivante :

- A partir  $T$ , nous construisons  $(Y_1, \dots, Y_n) \in \{0, 1\}^n$  tel que :
- Conditionnellement à  $T = k$ , on tire au hasard (uniformément) un ensemble de  $k$  éléments parmi  $n$ . On obtient ainsi l'ensemble aléatoire  $A_k$ .
- On définit  $Y_i = 1$  si  $i \in A_k$ ,  $Y_i = 0$  sinon.

Quelle est la loi de  $(Y_1, \dots, Y_n)$  ? Soit  $(\varepsilon_1, \dots, \varepsilon_n) \in \{0, 1\}^n$ , on a

$$\begin{aligned} & P_\theta(Y_1 = \varepsilon_1 \cap \dots \cap Y_n = \varepsilon_n) \\ &= \sum_{k=0}^n \mathbf{1}_{\{\sum_i \varepsilon_i = k\}} P_\theta(Y_1 = \varepsilon_1 \cap \dots \cap Y_n = \varepsilon_n | T = k) P_\theta(T = k) \\ &= \sum_{k=0}^n \mathbf{1}_{\{\sum_i \varepsilon_i = k\}} \frac{1}{C_n^k} C_n^k \theta^k (1 - \theta)^{n-k} \end{aligned}$$

Il est facile de vérifier que

$$P_\theta(Y_1 = \varepsilon_1 \cap \dots \cap Y_n = \varepsilon_n) = P_\theta(X_1 = \varepsilon_1 \cap \dots \cap X_n = \varepsilon_n)$$

Par cette procédure nous avons reconstruit non pas  $(X_1, \dots, X_n)$  à partir de  $T$ , mais  $(Y_1, \dots, Y_n)$ , qui a la même loi. En d'autres termes, à partir de l'expérience  $\mathcal{E}^T$ , on a reconstruit l'expérience  $\mathcal{E}$ .

*Remarque* : Il est fondamental de remarquer ici que cette reconstruction n'a été possible en connaissant  $T$ , mais **sans connaître**  $\theta$  parce que

$$P_{\theta}(Y_1 = \varepsilon_1 \cap \dots \cap Y_n = \varepsilon_n | T = k) = \frac{1}{C_n^k} \quad (1.4)$$

**ne dépend pas de  $\theta$ .**  $\triangle$

La définition classique de l'exhaustivité dans un modèle dominé est la suivante :

**Définition 9.** *Supposons que le modèle  $\mathcal{E} = (\Omega, \mathcal{F}, \mathcal{X}, \mathcal{X}', \mathcal{A}, P_{\theta}, \theta \in \Theta)$  est dominé.  $T$  étant une application mesurable de  $(\mathcal{X}, \mathcal{A})$  dans  $(\Pi, \mathcal{T})$ , on dira que  $T$  est exhaustive si et seulement si pour tout  $A$  dans  $\mathcal{A}$ , pour tout  $B$  dans  $\mathcal{T}$ , pour tous  $\theta$  et  $\theta'$  dans  $\Theta$ ,*

$$P_{\theta}(X \in A | T \in B) = P_{\theta'}(X \in A | T \in B). \quad (1.5)$$

Il est clair que vérifier (1.5) dans le modèle précédent équivaut à vérifier (1.4), qui s'est avérée fondamentale dans notre construction. Le lien dans un cadre général entre cette condition et la reconstruction de l'expérience originelle à partir de la sous-expérience est du à LeCam<sup>1, 2, 3</sup>. C'est un travail assez complexe que nous n'aborderons pas ici.

45

---

1. Le Cam, Lucien and Yang, Grace Lo, "Asymptotics in statistics. Some basic concepts", Springer-Verlag, New York, 1990  
 2. Strasser, Helmut, Mathematical theory of statistics, Walter de Gruyter & Co., Berlin, 1985  
 3. Genon-Catalot, Valentine and Picard, Dominique, Éléments de statistique asymptotique, Springer-Verlag, Paris, 1993,  
 4. Copyright @ 2010 Université Paris-Diderot Dominique Picard  
 5. Copyright @ 2010 Université Paris-Diderot Dominique Picard