Chapitre 4

Modélisation des observations

4.1 Introduction

Nous avons vu qu'une inférence bayésienne est fondée sur la détermination rigoureuse de trois facteurs :

1. la loi des observations $p(\boldsymbol{x}|\boldsymbol{\theta})$
2. la distribution *a priori* des paramètres $p(\boldsymbol{\theta})$
3. le coût associé aux décisions $C(\boldsymbol{\theta}, \boldsymbol{\delta})$

La détermination de ces trois facteurs se pose de difficultés comparables et ne peuvent se faire que par des considération partiellement subjectives.

Notons que les critiques fréquents de l'approche bayésienne se font souvent sur le point 2, alors que, conceptuellement les points 1 et 3 se trouvent sur le même rang. Le point 1 qui est un point commun entre l'approche classique et l'approche bayésienne est encore plus subtile et fait l'objet de ce chapitre.

La question que l'on se pose est comment déterminer la loi des observations $p(\boldsymbol{x}|\boldsymbol{\theta})$. Il s'agit donc d'une modélisation probabiliste des observations. Différentes approches sont possibles.

La première cosiste à choisir une loi qui nous semble "du bon sense" parmi les lois usuelles. En effet dans une application particulière les observations $\{x_1, \cdots, x_n\}$ représentent les échantillons d'une quantitée physique, par exemple les échantillons dans le temps d'un signal (température, tension, courant, etc.) ou les échantillons dans l'espace d'une image (variation spatiale de la conductivité ou de la vitesse de propagation d'une onde à l'intérieure d'un objet). Nous avons alors en général des informations nécessaires pour faire des hypothèses qui nous conduit à un choix pour $p(x_i|\boldsymbol{\theta})$ parmi les lois usuelles, et en déduire $p(\boldsymbol{x}|\boldsymbol{\theta})$. Par exemple, si les x_i représentent les valeurs mesurées de la température et que nous savons que les x_i ne peuvent dépasser les deux limites [a,b], et, qu'il n'y a aucune raison de dire que la valeur x_i est plus probable que la valeur x_j, alors on peut choisir une loi uniforme

$$p(x_i|a, b) = \frac{1}{b-a}.$$

Par contre, si on ne connait pas les limites mais on connait (ou on suppose connaître) sa valeur moyenne μ et sa variance σ^2 autour de cette valeur moyenne, alors on peut choisir la loi normale

$$p(x_i|a, b) = \mathbf{N}(x_i|\mu, \sigma^2).$$

D'autres exemples peuvent être cités.

- Si x_i représentent les valeurs observées de la durée de vie d'un composant que l'on estime connaître sa moyenne λ, alors on peut choisir la loi exponentielle

$$p(x_i|\lambda) = \mathbf{Ex}(x_i|\lambda).$$

- Si x_i représentent les nombres de photons reçues par un détecteur et on suppose connaître sa moyenne λ, alors on peut choisir la loi de Poisson

$$p(x_i|\lambda) = \mathbf{Pn}(x_i|\lambda).$$

Une foi $p(x_i|\boldsymbol{\theta})$ choisie ou déterminée, il faut encore faire d'autres hypothèses sur les liens entre les x_i et x_j pour pouvoir déterminer $p(\boldsymbol{x}|\boldsymbol{\theta})$. L'hypothèse la plus simple est que les différentes observations x_i sont indépendantes, mais ceci est une hypothèse trés forte qui n'est pas toujours vérifiée. D'autres notions comme interchangeabilité (complète ou partielle) ou la symétrie ou encore l'invariance sont classiquement utilisées. Dans ce qui suit, il y a un bref apérçu d'un certain nombre de ces notions.

4.2 Notions d'indépendance

La loi des observations $p(x_1, \cdots, x_n|\boldsymbol{\theta})$ contient toute information que l'on peut tirer de ses seules observations. Une fois cette loi définie, on peut définir aussi les lois marginales

$$p(x_i|\boldsymbol{\theta}) = \int p(x_1, \cdots, x_n|\boldsymbol{\theta})\, \mathrm{d}x_1 \cdots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \cdots \mathrm{d}x_n, \quad i = 1, \cdots, n,$$

mais aussi de tout autres sous-ensemble des x_i

$$p(x_1, \cdots, x_m|\boldsymbol{\theta}) = \int p(x_1, \cdots, x_n|\boldsymbol{\theta})\, \mathrm{d}x_{m+1} \cdots \mathrm{d}x_n, \quad m \leq n.$$

De même on peut calculer les lois conditionnelles

$$p(x_i|x_1, \cdots, x_m\boldsymbol{\theta}) = \frac{p(x_1, \cdots, x_n|\boldsymbol{\theta})}{p(x_1, \cdots, x_{i-1}, x_{i+1}, x_n|\boldsymbol{\theta})}, \quad i = 1, \cdots, n,$$

et

$$p(x_{m+1}, \cdots, x_n|x_1, \cdots, x_m; \boldsymbol{\theta}) = \frac{p(x_1, \cdots, x_n|\boldsymbol{\theta})}{p(x_1, \cdots, x_m|\boldsymbol{\theta})}.$$

Dans le cas où les x_i sont indépendants on a

$$p(x_1, \cdots, x_n|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i|\boldsymbol{\theta}),$$

et, évidament, on a

$$p(x_1, \cdots, x_m|\boldsymbol{\theta}) = \prod_{i=1}^{m} p(x_i|\boldsymbol{\theta}),$$

et

$$p(x_{m+1}, \cdots, x_n|x_1, \cdots, x_m; \boldsymbol{\theta}) = p(x_{m+1}, \cdots, x_n|\boldsymbol{\theta}),$$

Ce qui signifie que les observations du "passé" $\{x_1, \cdots, x_m\}$ ne donne aucune information sur les observations du "future" $\{x_{m+1}, \cdots, x_n\}$; on dit qu'il n'y a pas eu d'apprentissage.

4.3 Notions d'interchangeabilité

Indépendance est une notion trés forte. D'autres notions un peu moins strictes sont celles de l'interchangeabilité totale ou partielle, finie ou infinie, que l'on va les définir par la suite.

Définition 7 [Interchangeabilité] Un ensemble fini de variables $\{x_1, \cdots, x_n\}$ est dit *interchangeable* vis-à-vis de la mesure de probabilité P si

$$P(x_1, \cdots, x_n) = P(x_{\pi(1)}, \cdots, x_{\pi(n)})$$

pour toutes permutations π définie sur l'ensemble $\{1, \cdots, n\}$.

Exemple 7 Nous avons lancé n dés simultanément et observé leurs valeurs. Si on suppose que les dés sont tous identiques, alors les variables $\{x_1, \cdots, x_n\}$ sont interchangeables.

Définition 8 Un ensemble de variables $\{x_1, x_2, \cdots\}$ est dit *infiniment interchangeable* vis-à-vis de la mesure de probabilité P si tous ses sous-ensembles sont interchangeables.

Exemple 8 Supposons que $\{x_1, \cdots, x_n\}$ représentent les n valeurs mesurées d'une quantité physique ou chimique d'un objet. Si on suppose que ces mesures sont obtenues dans les mêmes conditions et suivant exactement le même protocole, alors on peut les considérer comme interchangeables.

Exemple 9 Supposons maintenant que $\{x_1, \cdots, x_m\}$ représentent les m valeurs mesurées d'une quantité par le laboratoire A et $\{x_{m+1}, \cdots, x_n\}$ les $n - m$ mesures de la même quantité par le laboratoire B. Ici, on n'a plus qu'une interchangeabilité partielle dans l'ensemble $\{x_1, \cdots, x_n\}$.

Dans ces exemples, il y a au moins un point commun entre les valeurs $\{x_1, \cdots, x_m\}$. En effet, tous représentent les mesures sur une quantité spécifique d'un objet. Si nous notons cette quantité par θ, que pouvons nous dire sur sa distribution $Q(\theta)$ et sur la loi $p(x_1, \cdots, x_n|\theta)$? Bien entendu, il nous faut d'autres hypothèses sur les x_i pour pouvoir répondre à ce type de questions. Les sections qui suivent donnent des réponses partielles à ces questions.

4.4 Modélisation des variables binaires interchangeables

4.4.1 Modéle binomial

Supposons que les $\{x_1, \cdots, x_n\}$ sont des variables binaires, *i.e.*; $x_i = \{0, 1\}$ et interchangeables, ce qui signifie que toutes permutations de ces variables ne modifie en rien leur loi de probabilité. Alors la question que l'on se pose est quelle est la forme générale de cette loi.

Exemple 10 Les n valeurs d'un mot de n bits dans la mémoire d'un ordinateur.

Une réponse à cette question se trouve dans la proposition suivante :

Proposition 7 [Modélisation des variables binaires] Si l'ensemble $\{x_1, x_2, \cdots\}$ de variables binaires ($x_j \in \{0, 1\}$) est infiniement interchangeable, alors il existe une

distribution Q telle que

$$p(x_1, \cdots, x_n) = \int_0^1 \prod_{j=1}^{n} \theta^{x_j} (1-\theta)^{1-x_j} \, \mathrm{d}Q(\theta),$$

où

$$Q(\theta) = \lim_{n\to\infty} P\left[\bar{x}_n \leq \theta\right] \quad \text{avec} \quad \theta = \lim_{n\to\infty} \bar{x}_n, \quad \text{et où} \quad \bar{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j.$$

De plus on a

$$p(y_n = \sum_{j=1}^{n} x_j) = \int_0^1 \binom{n}{y_n} \theta^{y_n} (1-\theta)^{n-y_n} \, \mathrm{d}Q(\theta),$$

$$p(x_1, \cdots, x_n|\theta) = \prod_{j=1}^{n} p(x_j|\theta) = \prod_{j=1}^{n} \theta^{x_j} (1-\theta)^{1-x_j}$$

et

$$p(x_{m+1}, \cdots, x_n|x_1, \cdots, x_m) = \int_0^1 \prod_{j=m+1}^{n} \theta^{x_j} (1-\theta)^{1-x_j} \, \mathrm{d}Q(\theta|x_1, \cdots, x_m)$$

avec

$$\mathrm{d}Q(\theta|x_1, \cdots, x_m) = \frac{\prod_{j=1}^{m} \theta^{x_j} (1-\theta)^{1-x_j} \, \mathrm{d}Q(\theta)}{\int_0^1 \prod_{j=1}^{m} \theta^{x_j} (1-\theta)^{1-x_j} \, \mathrm{d}Q(\theta)}.$$

D'un point de vue bayésien l'interprétation de cette proposition a une grande importance, car en effet

1. les x_i sont jugées *indépendantes conditionnellement à la connaissance de* θ,
2. la loi de $x_i|\theta$ est la loi de Bernouilli, *i.e.*; $p(x_i|\theta) = \mathbf{Ber}(x_i|\theta)$,
3. θ aussi est considéré d'admettre une loi *a priori* , *i.e.*; $p(\theta) = \mathrm{d}Q(\theta)$
4. une proposition est fournie pour déterminer la valeur de θ si on pouvait accéder à un trés grand nombre d'observations.

Notons cependant, que cette dernière phrase est trés hypothétique, car en pratique, on ne pourra jamais accéder à ce limite, et d'ailleurs souvent la question est d'estimer θ à partir d'un nombre fini d'échantillons.

4.4.2 Modèle Multinomial

Supposons que les $\{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\}$ sont des vecteurs binaires, *i.e.*; chaque vecteur $\boldsymbol{x}_j$ a K composantes $x_{jk}, k = 1, \cdots, K$ et que ces composantes sont binaires $x_{jk} = \{0, 1\}$. Supposons aussi que chaque vecteur $\boldsymbol{x}_j$ contient au moins un élément de valeur 1. Supposons aussi que ces vecteurs sont interchangeables. La question est : que peut-on dire sur la loi $p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n)$?

Exemple 11 les n valeurs représentées en binaires de n mots de K bits dans la mémoire d'un ordinateur en supposant qu'aucun de ces mots est égale à zéro.

Proposition 8 [Modélisation des vecteurs aléatoires binaires] Si l'ensemble $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots\}$ de vecteurs aléatoires binaires est infiniement interchangeable, alors il existe une distribution Q telle que

$$p(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = \int_\Theta \prod_{j=1}^{n} \theta_1^{x_{j1}} \cdots \theta_K^{x_{jK}} \left(1 - \sum_{k=1}^{K} \theta_k\right)^{1-\sum_{k=1}^{K} x_{jk}} \mathrm{d}Q(\boldsymbol{\theta}),$$

où

$$\Theta = \left\{ \boldsymbol{\theta} = (\theta_1, \cdots, \theta_K); \quad 0 \leq \theta_k \leq 1, \quad \sum_{k=1}^{K} \theta_k \leq 1 \right\}$$

et

$$Q(\boldsymbol{\theta}) = \lim_{n\to\infty} P\left[\bar{x}_{1n} \leq \theta_1 \cup \cdots \cup \bar{x}_{Kn} \leq \theta_K\right]$$

avec

$$\theta_k = \lim_{n\to\infty} \bar{x}_{kn} \text{ et où } \bar{x}_{kn} = \frac{1}{n} \sum_{j=1}^{n} x_{kj}.$$

4.5 Modèlisation des variables reélles

Une extension des idées précédentes au cas des variables reélles n'est pas, mathématiquement parlant, trés facile. Cependant nous citerons, ici aussi sans une prétention de rigueur mathématique, les propositions suivantes :

Proposition 9 [Modélisation des variables reélles] Si l'ensemble $\{x_1, x_2, \cdots\}$ de variables reélles ($x_j \in \mathbb{R}$) est infiniement interchangeable, alors il existe une distribution Q telle que

$$p(x_1, \cdots, x_n) = \int_\Theta \prod_{j=1}^{n} p(x_j|\boldsymbol{\theta})\, \mathrm{d}Q(\boldsymbol{\theta}),$$

$$p(x_1, \cdots, x_n|\boldsymbol{\theta}) = \prod_{j=1}^{n} p(x_j|\boldsymbol{\theta})$$

et

$$p(x_{m+1}, \cdots, x_n|x_1, \cdots, x_m) = \int_\Theta \prod_{j=m+1}^{n} p(x_j|\boldsymbol{\theta})\, \mathrm{d}Q(\boldsymbol{\theta}|x_1, \cdots, x_m)$$

avec

$$\mathrm{d}Q(\boldsymbol{\theta}|x_1, \cdots, x_m) = \frac{\prod_{j=1}^{m} p(x_j|\boldsymbol{\theta})\, \mathrm{d}Q(\boldsymbol{\theta})}{\int_\Theta \prod_{j=1}^{m} p(x_j|\boldsymbol{\theta})\, \mathrm{d}Q(\boldsymbol{\theta})}.$$

Une différence de taille cependant est qu'ici la forme de la loi $p(x_i|\boldsymbol{\theta})$ n'est pas définie. Il faut alors faire d'autres hypothèses supplémentaires pour pouvoir la déterminer. La symétrie et l'invariance par group de transformations sont deux notions que l'on peut utiliser pour ce fin.

4.6 Modèlisation des variables reélles via la symétrie

Définition 9 [Propriéte de symétrie sphérique] Un ensemble de variables $\{x_1, \cdots, x_n\}$ a une propriéte de symétrie sphérique (S.S.) vis-à-vis de la mesure de probabilité P si

$$P(x_1, \cdots, x_n) = P(y_1, \cdots, y_n)$$

pour toutes transformation $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ avec $\boldsymbol{A}$ une matrice orthonormale $\boldsymbol{A}^{-1} = \boldsymbol{A}^t$ et $|\boldsymbol{A}| = 1$.

Définition 10 [Propriéte de symétrie sphérique centrée] Un ensemble de variables $\{x_1, \cdots, x_n\}$ a une propriéte de symétrie sphérique centrée (S.S.C.) vis-à-vis de la mesure de probabilité P si l'ensemble $\{x_1 - \bar{x}, \cdots, x_n - \bar{x}\}$ avec $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ a la propriéte de symétrie sphérique vis-à-vis de la mesure de probabilité P.

Proposition 10 [Modélisation des variables reélles de S.S.] Si l'ensemble $\{x_1, x_2, \cdots\}$ de variables reélles ($x_j \in \mathbf{R}$) avec la mesure de probabilité P est infiniement interchangeable, et si pour tout n, l'ensemble $\{x_1, \cdots, x_n\}$ a une propriété de symétrie sphérique, alors il existe une distribution Q sur $\mathbf{R}^+$ telle que

$$p(x_1, \cdots, x_n) = \int_{\mathbf{R}^+} \prod_{j=1}^{n} p(x_j|\lambda)\, \mathrm{d}Q(\lambda),$$

où

$$p(x_j|\lambda) = \mathbf{N}(x_j|0, \lambda)$$

est la distribution gaussienne centrée et

$$Q(\lambda) = \lim_{n\to\infty} P\left[s_n^{-2} \leq \lambda\right] \quad \text{avec} \quad \lambda^{-1} = \lim_{n\to\infty} s_n^2 \quad \text{et où} \quad s_n^2 = \frac{1}{n}\sum_{j=1}^{n} x_j^2.$$

De plus on a

$$p(x_1, \cdots, x_n|\lambda) = \prod_{j=1}^{n} \mathbf{N}(x_j|\lambda) = \mathbf{N}_n(\boldsymbol{x}|\boldsymbol{0}, \lambda\boldsymbol{I})$$

et

$$p(x_{m+1}, \cdots, x_n|x_1, \cdots, x_m) = \int_{\mathbf{R}^+} \prod_{j=m+1}^{n} p(x_j|\lambda)\, \mathrm{d}Q(\lambda|x_1, \cdots, x_m)$$

avec

$$\mathrm{d}Q(\lambda|x_1, \cdots, x_m) = \frac{\prod_{j=1}^{m} p(x_j|\lambda)\, \mathrm{d}Q(\lambda)}{\int_{\mathbf{R}^+} \prod_{j=1}^{m} p(x_j|\lambda)\, \mathrm{d}Q(\lambda)}.$$

Proposition 11 [Modélisation des variables reélles de S.S.C.] Si l'ensemble $\{x_1, x_2, \cdots\}$ de variables reélles ($x_j \in \mathbf{R}$) avec la mesure de probabilité P est infiniement interchangeable, et si pour tout n, l'ensemble $\{x_1, \cdots, x_n\}$ a une propriété de symétrie sphérique centrée, alors il existe une distribution Q sur $\mathbf{R} \times \mathbf{R}^+$ telle que

$$p(x_1, \cdots, x_n) = \int_{\mathbf{R}\times\mathbf{R}^+} \prod_{j=1}^{n} p(x_j|\mu, \lambda)\, \mathrm{d}Q(\mu, \lambda),$$

où

$$p(x_j|\mu,\lambda) = \mathbf{N}(x_j|\mu,\lambda)$$

est la distribution normale et

$$Q(\mu,\lambda) = \lim_{n\to\infty} P\left[(\bar{x}_n \leq \mu) \cap (s_n^{-2} \leq \lambda)\right]$$

avec

$$\mu = \lim_{n\to\infty} \bar{x}_n, \quad \lambda^{-1} = \lim_{n\to\infty} s_n^2, \quad \text{et où} \quad \bar{x}_n = \frac{1}{n}\sum_{j=1}^{n} x_j \quad \text{et} \quad s_n^2 = \frac{1}{n}\sum_{j=1}^{n}(x_j-\mu)^2.$$

De plus on a

$$p(x_1,\cdots,x_n|\mu,\lambda) = \prod_{j=1}^{n} \mathbf{N}(x_j|\mu,\lambda) = \mathbf{N}_n(\boldsymbol{x}|\mu\mathbf{1},\lambda\boldsymbol{I})$$

et

$$p(x_{m+1},\cdots,x_n|x_1,\cdots,x_m) = \int_{\mathbf{R}^+} \prod_{j=m+1}^{n} p(x_j|\mu,\lambda)\,\mathrm{d}Q(\mu,\lambda|x_1,\cdots,x_m)$$

avec

$$\mathrm{d}Q(\mu,\lambda|x_1,\cdots,x_m) = \frac{\displaystyle\prod_{j=1}^{m} p(x_j|\mu,\lambda)\,\mathrm{d}Q(\mu,\lambda)}{\displaystyle\int_{\mathbf{R}\times\mathbf{R}^+} \prod_{j=1}^{m} p(x_j|\mu,\lambda)\,\mathrm{d}Q(\mu,\lambda)}.$$

Proposition 12 [Modélisation des vecteurs aléatoires reélles de S.S.C.] Si l'ensemble $\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots\}$ de variables reélles ($\boldsymbol{x}_j \in \mathbb{R}^k$) avec la mesure de probabilité P est infiniement interchangeable, et si pour tout n et pour tout vecteur $\boldsymbol{c} \in \mathbf{R}^k$, l'ensemble $\{\boldsymbol{c}^t\boldsymbol{x}_1,\cdots,\boldsymbol{c}^t\boldsymbol{x}_n\}$ a une propriété de symétrie sphérique centrée, alors

$$p(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_n|\boldsymbol{\mu},\boldsymbol{\lambda}) = |\boldsymbol{\lambda}|^{\frac{1}{2}}\,(2\pi)^{-\frac{k}{2}}\,\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\boldsymbol{\lambda}(\boldsymbol{x}-\boldsymbol{\mu})\right],$$

où

$$\boldsymbol{\mu} = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{x}_j \quad \text{et} \quad \boldsymbol{\lambda}^{-1} = \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}(\boldsymbol{x}_j-\boldsymbol{\mu})(\boldsymbol{x}_j-\boldsymbol{\mu})^t$$

4.7 Modèlisation via l'invariance d'origine et la positivité

Proposition 13 [Modélisation des variables reélles positives] Si l'ensemble $\{x_1,x_2,\cdots\}$ de variables reélles positives ($x_j \in \mathbf{R}_+$) avec la mesure de probabilité P est infiniement interchangeable, et si pour tout n et pour tout sous-ensemble $A \subseteq \mathbb{R}_+^n$

$$p(\boldsymbol{x} \in A) = p(\boldsymbol{x} \in A + \boldsymbol{a})$$

pour tout $\boldsymbol{a} \in \mathbf{R}^n$ tel que $\boldsymbol{a}^t\mathbf{1} = 0$ et tel que $A + \boldsymbol{a} \subseteq \mathbf{R}_+^n$, alors

$$p(x_1,\cdots,x_n) = \int_0^\infty \prod_{j=1}^{n} p(x_j|\theta)\,\mathrm{d}Q(\theta),$$

où

$$p(x_j|\theta) = \mathbf{Ex}(x_j|\theta) = \theta \exp\left[-\theta x_j\right]$$

et où

$$Q(\theta) = \lim_{n\mapsto\infty} \Pr\left\{\bar{x}_n^{-1} \leq \theta\right\}.$$

Proposition 14 [Modélisation des variables entiers positives] Si l'ensemble $\{x_1, x_2, \cdots\}$ de variables entiers positives ($x_j \in \mathbf{Z}_+$) avec la mesure de probabilité P est infiniement interchangeable, et si pour tout n et pour tout sous-ensemble $A \subseteq \mathbf{Z}_+^n$

$$p(\boldsymbol{x} \in A) = p(\boldsymbol{x} \in A + \boldsymbol{a})$$

pour tout $\boldsymbol{a} \in \mathbb{R}^n$ tel que $\boldsymbol{a}^t\mathbf{1} = 0$ et tel que $A + \boldsymbol{a} \subseteq \mathbf{Z}_+^n$, alors

$$p(x_1, \cdots, x_n) = \int_0^\infty \prod_{j=1}^n p(x_j|\theta)\, \mathrm{d}Q(\theta),$$

où

$$p(x_j|\theta) = \mathbf{Geo}(x_j|\theta) = \theta(1-\theta)^{x_j - 1}$$

et où

$$Q(\theta) = \lim_{n\mapsto\infty} \Pr\left\{\bar{x}_n^{-1} \leq \theta\right\}.$$

4.8 Modèlisation via la Statistiques suffisantes

Définition 11 Pour un ensemble de variables $\{x_1, \cdots, x_n\}$, un vecteur $\boldsymbol{t} = \{t_1, \cdots, t_k\}$, où $t_k(\boldsymbol{x})$ sont des fonctions réelles, est appelé une *statistique* de dimension k.

Définition 12 Considérons l'ensemble de variables $\boldsymbol{x} = \{x_1, x_2, \cdots\}$ avec une mesure de probabilité P, et $\boldsymbol{x}_1$ et $\boldsymbol{x}_2$ deux sous-ensembles différents de $\boldsymbol{x}$. Alors, une Statistique $\boldsymbol{t}(\boldsymbol{x}_1) = \{t_1(\boldsymbol{x}_1), \cdots, t_k(\boldsymbol{x}_1)\}$ de dimension k, est appelé une *Statistique suffisante prédictive* pour $\boldsymbol{x}_1$ si pour tout $\boldsymbol{x}_1$ et $\boldsymbol{x}_2$ possibles

$$p(\boldsymbol{x}_1|\boldsymbol{x}) = p(\boldsymbol{x}_1|\boldsymbol{t})$$

Proposition 15 A COMPLETER

A COMPLETER