

III – CHAINES DE MARKOV

Dans ce chapitre, on étudie un premier exemple de *processus stochastique discret*, ie une suite de variables aléatoires. On considère donc une suite de variables $(X_n)_{n \in E}$ à valeurs dans un ensemble E qu'on supposera dénombrable.

On dira que est une **chaîne de Markov** si et seulement si elle vérifie la propriété suivante:

$$\forall i_0, \dots, i_{n+1} \in E ; \mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n)$$

On dira aussi plus conceptuellement que *la donnée de la valeur future ne dépend que de la valeur présente*.

On note alors plus généralement $\mathbb{P}(X_{n+1} = y | X_n = x) = Q_n(x, y)$. Aussi si la probabilité de passer de l'état x à l'état y ne dépend pas de n ¹, alors on dira que la chaîne de markov est **homogène**. On a alors l'application suivante:

$$Q : E^2 \longrightarrow [0; 1] \\ (x, y) \longmapsto \mathbb{P}(X_1 = y | X_0 = x)$$

On appelle cette application **matrice de transition** de la chaîne de Markov, en effet, dans le cas fini c'est exactement une matrice de taille $|E|$. Alors, on en déduit directement la caractérisation principale:

Une chaîne de Markov est caractérisée par sa matrice et sa distribution initiale $\lambda = \mathbb{P}_{X_0}$. On dira alors que $X \sim \text{Markov}(Q, \lambda)$

On peut alors de manière heuristique se représenter une chaîne de Markov, ie une matrice de transition, sous une forme plus géométrique, en effet on peut toujours représenter une chaîne de Markov sous forme d'un **graphe orienté pondéré**, où chaque arête porte la probabilité de transition entre deux sommets comme on le verra plus loin.

Exemples:

Voici quelques exemples de différents type pour illustrer la définition:

- **La chaîne à deux états:** On prends $E = \{1, 2\}$, alors la donnée d'une chaîne de Markov sur E est équivalente à la donnée de deux réels $\alpha, \beta \in]0; 1[$ qui donne la matrice de transition:

$$Q = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix}$$

Et le graphe suivant: TODO

- **La marche aléatoire homogène:** On prends \mathbb{Z}^k , alors la donnée d'une chaîne de Markov sur E est définie par la matrice de transition:

$$Q(x, y) = \begin{cases} \frac{1}{k} ; & \text{si } y \text{ est voisin de } x \\ 0 ; & \text{sinon} \end{cases}$$

De manière générale, sur n'importe quel graphe (localement fini), on peut définir une marche aléatoire homogène via:

$$Q(x, y) = \begin{cases} \frac{1}{\deg(x)} ; & \text{si } y \text{ est voisin de } x \\ 0 ; & \text{sinon} \end{cases}$$

Et, par exemple sur \mathbb{Z}^2 , on a le graphe suivant: TODO

¹Ce qui peut toujours être supposé sans perte de généralité.

Lois marginales:

Par une simple application de la définition d'une chaîne $X \sim \text{Markov}(Q, \lambda)$, on peut montrer que:

$$\forall x_0, \dots, x_{n+1} \in E; \mathbb{P}(X_{n+1} = i_{n+1}, \dots, X_0 = x_0) = \lambda(x_0)Q(x_0, x_1)Q(x_1, x_2) \dots Q(x_n, x_{n+1})$$

Puis alors on a facilement que pour tout y dans E , on a:

$$\mathbb{P}(X_n = y \mid X_0 = x) = \sum_{x_1, \dots, x_{n-1} \in E} \mathbb{P}(X_n = x, \dots, X_0 = x_0) = \sum_{x_1, \dots, x_{n-1} \in E} Q(x, x_1)Q(x_1, x_2) \dots Q(x_{n-1}, y)$$

Ceci s'identifie alors à un produit matriciel ! En effet on a:

$$\sum_{x_1, \dots, x_{n-1} \in E} Q(x, x_1)Q(x_1, x_2) \dots Q(x_{n-1}, y) = (Q^n)_{x,y}$$

Et donc la loi de X_n conditionnellement à $X_0 = x_0$ est donnée par $(Q^n)_{x,y}$, en outre on a donc que:

$$\forall y \in E; \mathbb{P}(X_n = y) = \sum_{x \in E} \lambda(x)Q^n_{x,y}$$

On appelle cette loi, **loi marginale** de la chaîne (X_n) . Elle sera particulièrement importante par la suite car un thème d'étude récurrent sur les chaînes de Markov consiste à étudier le comportement limite de cette loi.

Caractérisation récursive:

Si X_0 est variable aléatoire sur E de loi λ et que $(\xi_n)_{n \in \mathbb{N}^*}$ est une suite de variables aléatoires iid à valeurs dans un ensemble F et indépendante de X_0 , alors pour tout fonction mesurable $G : E \times F \rightarrow E$, la suite définie par récurrence:

$$X_{n+1} = G(X_n, \xi_{n+1})$$

Est une chaîne de markov sur E de loi initiale λ . En outre, toute chaîne de markov sur E se réalise de la sorte, c'est un autre point de vue sur les chaînes de Markov. On a alors que :

$$Q(x, y) = \mathbb{P}(G(x, \xi_1) = y)$$

Par exemple la marche aléatoire homogène sur \mathbb{Z} se réalise via une suite iid (ξ_n) de Rademacher de paramètre p avec:

$$X_{n+1} = X_n + \xi_n$$

Fonction d'une chaîne de Markov:

Comme souvent lorsque l'on introduit un nouvel objet, on cherche à comprendre quels sont les fonctions qui préservent ces objets. On se donne une chaîne de markov (X_n) sur E et une fonction $f : E \rightarrow F$. On cherche à savoir si la suite de variables aléatoires suivantes et une chaîne de Markov sur l'image de F :

$$Y_n = f(X_n)$$

On supposera sans perte de généralité f surjective. Alors peut montrer que c'est facilement le cas si f est injective. Mais on peut aller plus loin, en effet si on considère la partition de E en classe pour la relation 'avoir la même image par f ', alors on peut définir une chaîne de markov sur F si la probabilité de transition entre les classes ne dépend pas du représentant choisi, ie si la définition suivante à un sens:

$$Q(\tilde{x}, \tilde{y}) := Q(x, \tilde{y})$$

Alors on définit:

$$Q(\tilde{x}, \tilde{y}) := Q(x, \tilde{y}) := \sum_{y \in \tilde{y}} Q(x, y)$$

C'est une chaîne de markov sur le quotient E/\sim . On peut alors identifier les éléments du quotient avec ceux de l'image par le théorème d'isomorphisme et alors définir \tilde{Q} sur F^2 par:

$$\tilde{Q}(f(x), f(y)) = Q(\tilde{x}, \tilde{y})$$

Par exemple pour la marche aléatoire homogène sur \mathbb{Z} de probabilité $\frac{1}{2}$ et la fonction $f(z) = |z|$, on trouve que cette fonction convient et que:

$$\forall x, y \in \mathbb{N}; \tilde{Q}(x, y) = \mathbb{1}_{\{(x,y)=(0,1)\}} + \frac{1}{2}\mathbb{1}_{\{|y-x|=1\}}$$

Classes de communication et chaînes irréductibles:

On peut alors définir sur E une relation d'équivalence, dite relation de communication et on dira que x **mène à** y , et on note $x \rightarrow y$ si et seulement si:

$$\forall x, y \in E ; \mathbb{P}(\exists n ; X_n = y \mid X_0 = x) > 0$$

Et on dira alors que x **communiqu**e avec y si et seulement si:

$$x \rightarrow y \text{ et } y \rightarrow x$$

Qu'on note alors $x \longleftrightarrow y$. Cette relation définit une relation d'équivalence sur E et ses classes sont appelés **classes de communication** de E . Si une chaîne de Markov n'a qu'une classe de communication, on dira qu'elle est **irréductible**.

Ce concept est fondamental est similaire à celui de composantes connexes en topologie, en effet, on verra qu'on se réduira surtout à l'étude de chaînes de Markov irréductibles sans perte de généralité.

III – TEMPS D'ARRÊT ET RÉCURRENCE

Dans ce chapitre on s'intéressera plus particulièrement aux propriétés de *récurrence* d'une chaîne de Markov $(X_n) \sim \text{Markov}(Q, \lambda)$, qui consiste à répondre à la question suivante:

Etant donnée une chaîne de Markov partant d'un état donné, y reviendra elle ? Et si oui avec quelle probabilité ?

Pour ceci, on définit la notion de **temps d'arrêt**, de **récurrence** et de **récurrence positive**.

Temps d'arrêt:

On dira qu'une variable aléatoire T à valeurs dans $\mathbb{N} \cup \{\infty\}$ est un **temps d'arrêt** si et seulement si, on a:

$$\forall n \in \mathbb{N}; \{T \leq n\} \in \sigma(X_0, \dots, X_n)$$

En d'autres termes, $\{T \leq n\}$ n'est déterminée que par les n dernières valeurs. On peut aussi de manière équivalente le définir par la même propriété pour $\{T > n\}$ ou $\{T = n\}$. Quelques exemples classiques:

- Une variable aléatoire constante $T \equiv n$ est un temps d'arrêt. En effet $\{T \equiv n\} \in \{\emptyset, \Omega\}$
- Une variable aléatoire codant le **temps d'atteinte** d'une partie A donnée ci-dessous, est un temps d'arrêt:

$$H_A := \inf \{n \geq 0; X_n \in A\}$$

- Une variable aléatoire codant le **temps de premier passage** d'une partie A donnée ci-dessous, est un temps d'arrêt:

$$T_A := \inf \{n \geq 1; X_n \in A\}$$

Les deux derniers exemples sont fondamentaux, pour prouver que ce sont des temps d'arrêt, on remarque que par exemple $\{H_A > n\} = \{X_0 \notin A, \dots, X_n \notin A\}$.

Propriétés des temps d'arrêt:

On considère alors deux temps d'arrêts T, M , on a alors les propriétés suivantes:

- La variable $\min\{T, M\}$ est un temps d'arrêt.
- La variable $\max\{T, M\}$ est un temps d'arrêt.
- La variable $T + M$ est un temps d'arrêt.

Propriété de Markov faible:

On considère une chaîne de Markov $X \sim \text{Markov}(Q, \lambda)$, un état $x \in E$, un entier n . Alors on a la propriété suivante dite **propriété de Markov faible**, qui donne que conditionnellement à $X_n = x$, la chaîne décalée $(X_{k+n})_{k \in \mathbb{N}}$ est indépendante de tout événement $A \in \sigma(X_0, \dots, X_n)$, ie:

$$\forall B \subset E^{\mathbb{N}}; \mathbb{P}(X_{k+n} \in B \mid \{X_n = x\} \cap A) = \mathbb{P}(X_{k+n} \in B \mid X_n = x)$$

Moralement, c'est une propriété *d'oubli du passé*. L'information du passé peut être oubliée.

Propriété de Markov forte:

Grâce à la notion de temps d'arrêt, on peut former la **propriété de Markov forte** qui donne alors pour tout temps d'arrêt fini T que conditionnellement à $X_T = x$, la chaîne décalée $(X_{k+T})_{k \in \mathbb{N}}$ est indépendante de tout événement $A \in \sigma(X_0, \dots, X_T)$, ie:

$$\forall B \subset E^{\mathbb{N}}; \mathbb{P}(X_{k+T} \in B \mid \{T < \infty, X_T = x\} \cap A) = \mathbb{P}(X_{k+T} \in B \mid X_T = x)$$

Probabilités d'atteinte, analyse à un pas:

On peut donc chercher pour $x \in E$ et $A \subseteq E$ la probabilité d'atteinte de A en partant de x , définie par:

$$h_A(x) := \mathbb{P}(H_A < \infty \mid X_0 = x)$$

Ce sera une idée courante par la suite pour étudier la récurrence. En outre, il existe une méthode usuelle pour étudier ces problèmes appelées **analyse à un pas**. Elle consiste à **conditionner** cette probabilité une étape plus loin puis utiliser la propriété de Markov, en effet on a après disjonction de cas sur x vérifie:

$$h_A(x) = \mathbb{1}_{x \in A} + \mathbb{1}_{x \notin A} \sum_{y \in E \setminus A} \mathbb{P}(X_1 = y) \mathbb{P}(H_A < \infty \mid X_1 = y, X_0 = x)$$

Par la propriété de Markov, ceci est égal à:

$$h_A(x) = \mathbb{1}_{x \in A} + \mathbb{1}_{x \notin A} \sum_{y \in E \setminus A} \mathbb{P}(X_1 = y) \mathbb{P}(H_A < \infty \mid X_1 = y)$$

Qui donne un **système linéaire** ! En effet, on obtient:

$$h_A(x) = \mathbb{1}_{x \in A} + \mathbb{1}_{x \notin A} \sum_{y \in E \setminus A} \mathbb{P}(X_1 = y) h_A(y)$$

Il est souvent facilement résoluble, notamment dans le cas fini. Cette méthode se généralise aussi au calcul de probabilité du type $\mathbb{P}(H_A = n)$ ou d'espérance du type $\mathbb{E}[H_A]$.

Etat récurrent:

On définit une variable aléatoire appelée **nombre de visites** en x par:

$$V_x = \sum_{i=1}^n \mathbb{1}_{X_i = x}$$

Alors on définit les notions suivantes pour un état $x \in E$:

- Il est dit **récurrent** si et seulement si $\mathbb{P}(V_x = \infty \mid X_0 = x) = 1$.
- Il est dit **transcient** si et seulement si $\mathbb{P}(V_x < \infty \mid X_0 = x) = 1$.

On peut alors montrer que ces notions sont équivalentes à celles-ci:

- Il est dit **récurrent** si et seulement si $\mathbb{P}(T_x < \infty \mid X_0 = x) = 1$.
- Il est dit **transcient** si et seulement si $\mathbb{P}(T_x < \infty \mid X_0 = x) < 1$.

Et que l'on a une caractérisation¹ en termes de **série** de la récurrence, en effet on a:

$$x \in E \text{ est récurrent} \iff \sum_{n \in \mathbb{N}} (Q^n)_{x,x} = \infty$$

Par exemple on peut étudier la marche aléatoire uniforme sur \mathbb{Z} . On peut montrer que:

$$\mathbb{P}(X_n = m) = \mathbb{1}_{\{n+m \text{ pair}\}} \binom{n}{\frac{n+m}{2}} p^{\frac{n+m}{2}} (1-p)^{\frac{n-m}{2}}$$

Donc:

$$\sum_{n \in \mathbb{N}} (Q^n)_{0,0} = \sum_{n \in \mathbb{N}} \mathbb{1}_{\{n \text{ pair}\}} \binom{n}{\frac{n}{2}} p^{\frac{n}{2}} (1-p)^{\frac{n}{2}}$$

Cette dernière est égale à:

$$\sum_{k \in \mathbb{N}} \binom{2k}{k} (p(1-p))^k = \sum_{k \in \mathbb{N}} u_k$$

¹On utilise pour montrer ceci que $\mathbb{P}(V_x > n) = \mathbb{P}(T_x < \infty)^n$ et est donc distribué géométriquement. On calcule ensuite son espérance et on obtient alors le critère.

Puis en utilisant la formule de Stirling, on trouve l'équivalent:

$$u_k \sim \frac{(4p(1-p))^k}{\sqrt{\pi k}}$$

Et donc finalement on a:

- La marche aléatoire est **récurrente** si $p = \frac{1}{2}$.
- La marche aléatoire est **transiente** si $p \neq \frac{1}{2}$.

Donc la marche aléatoire symétrique sur \mathbb{Z} est récurrente, cette étude se généralise en le théorème de Polya:

La marche aléatoire symétrique sur \mathbb{Z}^d est récurrente si et seulement si $d \leq 2$.

On peut montrer assez facilement que la récurrence est une **propriété des classes de communication**, et donc en particulier, si (X_n) est irréductible, on dira que la chaîne est récurrente si n'importe quel état est récurrent. En outre dans ce cas, on a que si la chaîne est récurrente, l'état initial n'importe pas et:

$$\forall x \in E ; \mathbb{P}(T_x < \infty) = 1$$

Etat récurrent positif:

Il existe alors une propriété encore plus forte de récurrence qui sera utile par la suite, pour $x \in E$, on dira que x est **récurrent positif** si et seulement si:

$$\mathbb{E}[T_x \mid X_0 = x] < \infty$$

C'est bien une propriété plus forte, en effet si cette espérance est finie, la probabilité de la variable correspondante est finie mais en revanche si T_x est finie presque partout, a priori rien n'indique que son espérance soit finie. Cette propriété sera **très** utile par la suite, et on peut la caractériser par les séries, en effet on a que:

$$x \text{ est récurrent positif} \iff \frac{1}{n} \sum_{k \leq n-1} (Q^k)_{x,x} > 0$$

La récurrence positive est aussi une propriété de classes de communication, on dira donc qu'une chaîne irréductible est récurrente positive si et seulement si n'importe quel état l'est.

III – ETUDE ASYMPTOTIQUE

Dans cette dernière partie, on s'intéresse plus particulièrement aux propriétés **asymptotiques** de la chaîne de Markov. On énoncera notamment un analogue de la loi des grands nombres appelé **théorème ergodique**, et on cherchera des conditions pour que la chaîne de Markov admette une loi marginale limite, qu'on appellera **probabilité invariante**.

Mesures invariantes:

Soit π une mesure sur E (pas nécessairement finie), alors on dira qu'elle est **invariante** pour Q si et seulement si:

$$\pi = \pi Q$$

On remarque que tout multiple d'une mesure invariante est une mesure invariante et toute combinaison linéaire de mesures invariantes est invariante. En outre si la mesure est de probabilité, toute combinaison **convexe** de probabilités invariantes reste une probabilité invariante.

Loi limite:

La première rencontre avec une probabilité invariante que l'on peut avoir vient souvent de la propriété suivante:

Si $X_n \rightarrow X$ avec X de loi π , alors π est une probabilité invariante.

Par exemple, on peut considérer la chaîne à deux états, alors on a par diagonalisation:

$$Q^n = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}^n = \begin{pmatrix} 1 & -\frac{\alpha}{\beta} \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -\alpha-\beta+1 \end{pmatrix}^n \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ -\beta & \beta \end{pmatrix}$$

Puis on calcule:

$$Q^n = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta + \alpha o(1) & \alpha - \alpha o(1) \\ \beta - \beta o(1) & \alpha + \beta o(1) \end{pmatrix} \rightarrow \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix}$$

Donc si $\lambda = (\lambda_1, \lambda_2)$ est la probabilité initiale, alors on a:

$$\lambda Q^n \rightarrow \frac{1}{\alpha+\beta} \begin{pmatrix} \beta(\lambda_1 + \lambda_2) \\ \alpha(\lambda_1 + \lambda_2) \end{pmatrix} = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$$

Donc $\pi = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$ est une probabilité invariante pour la chaîne à deux états.

Aperiodicité:

Une question naturelle est alors:

Existe-il une réciproque à ce résultat, ie si X admet une probabilité invariante, as-t-on que $\mathbb{P}_{X_n} \rightarrow \pi$?

En général c'est faux, c'est néanmoins le cas si la chaîne est **apériodique**, pour définir ce concept, on définit la **période** d'un état $x \in E$ par:

$$\text{per}(x) := \text{pgcd}\{n \in \mathbb{N} ; (Q^n)_{x,x} > 0\}$$

Moralement, on regarde les cycles de longueur n qui partent de x et y reviennent et on prends le pgcd de leur longueurs. Si une chaîne est apériodique, alors la réciproque est vraie, ie si X_n admet une probabilité invariante π , alors pour toute distribution initiale λ , on a que:

$$\mathbb{P}_{X_n} \rightarrow \pi$$

Conditions suffisantes pour obtenir une probabilité invariante:

Il existe aussi des conditions plus simples pour trouver une probabilité invariante. En voici une liste non exhaustive:

- **La méthode classique:** Résoudre le système d'inconnue $\pi = (\pi_1, \dots, \pi_n)$ suivant:

$$\begin{cases} \pi Q = \pi \\ \sum_{i \in \mathbb{N}} \pi_i = 1 \end{cases}$$

- **Les mesures réversibles:** On appelle **mesure réversible** toute mesure $\pi = (\pi_1, \dots, \pi_n)$ telle que:

$$\forall x, y \in E ; \pi_x Q(x, y) = \pi_y Q(y, x)$$

Si une telle mesure existe, elle est invariante, c'est en général plus simple que la méthode précédente.

- **Les matrices bistochastiques:** On appelle **matrice bistochastique** une matrice telle que toute ligne ou colonne somme à 1 et tout ses coefficients sont positifs. Si Q est bistochastique, alors elle admet la mesure uniforme comme probabilité invariante.

Existence de mesures invariantes:

Les concepts précédents définis permettent alors de donner des conditions suffisantes pour l'existence de mesures invariantes. Dans la suite de cette section on supposera les chaînes de Markov **irréductibles**. Alors on a:

- Toute chaîne de Markov **récurrente** admet une **mesure invariante**.
- Toute chaîne de Markov **récurrente positive** admet une **probabilité invariante**.

C'est le critère principal qui nous permettra d'affirmer l'existence de telles mesures.

Théorème ergodique:

Dans la suite on suppose que Q est irréductible et récurrente positive. On note π sa probabilité invariante. Alors pour toute fonction $f : E \rightarrow \mathbb{R}$, telle que $f \in L^1(\pi)$ alors on a:

$$\frac{1}{n} \sum_{k \leq n} f(X_k) \xrightarrow{p.s.} \sum_{x \in E} f(x) \pi(x)$$

C'est en fait un analogue à la loi des grands nombres (dont la preuve découle non trivialement), qui moralement signifie que:

La moyenne empirique des valeurs de f sur la chaîne de Markov converge vers l'espérance de f contre une mesure invariante.

Généralisation du théorème ergodique:

On peut généraliser ce théorème à $f : E^2 \rightarrow \mathbb{R}$ telle que $\sum_{x, y \in E} |f(x, y)| \pi(x) Q(x, y) < \infty$ par:

$$\frac{1}{n} \sum_{k \leq n} f(X_k, X_{k+1}) \xrightarrow{p.s.} \sum_{x, y \in E} f(x, y) \pi(x) Q(x, y)$$

Et plus généralement à toute fonction $f : E^k \rightarrow \mathbb{R}$ quitte à rajouter des termes dans l'espérance limite.

Application du théorème ergodique:

Ce théorème nous permet par exemple de calculer:

- La proportion asymptotique d'un état fixé $y \in E$, on choisit la fonction $f(x) = \mathbb{1}_{\{x=y\}}$.
- La proportion asymptotique de changement d'états, on choisit la fonction $f(x, y) = \mathbb{1}_{\{x \neq y\}}$.

III – STATISTIQUES INFÉRENTIELLES

Dans ce chapitre on introduit la démarche inverse à celle du probabiliste, ie celle du statisticien:

Etant donnée des réalisations d'une variable aléatoire dont je peux raisonnablement cerner le type, que puis je dire sur cette distribution ?

Plus formellement, on observe une variable aléatoire X à valeurs dans un espace probabilisable (Ω, \mathcal{A}) appelé **espace des observations** et dont on ne connaît pas la distribution.

On définit un **modèle statistique** (paramétrique) comme le choix d'une famille de mesures de probabilités $(\mathbb{P}_\theta)_{\theta \in \Theta}$ sur $X(\Omega)$ pour un espace des paramètres $\Theta \subseteq \mathbb{R}^d$. On suppose que cette famille contient la "vraie" distribution de X . Notre objectif est alors d'utiliser l'observation pour réussir à obtenir des informations sur cette vraie distribution. Quelques exemples:

- **Le modèle de Bernoulli:** Ici on choisit $\Theta = [0; 1]$ et le modèle:

$$(\Omega, \mathcal{A}, \mathbb{P}_\theta) = (\{0, 1\}, \mathcal{P}(\Omega), \text{Bernoulli}(\theta))$$

Notre observation modélise le résultat d'une expérience de Bernoulli.

- **Le modèle de Bernoulli général:** Ici on choisit $\Theta = [0; 1]$ et le modèle:

$$(\Omega, \mathcal{A}, \mathbb{P}_\theta) = (\{0, 1\}^{\mathbb{N}}, \mathcal{P}(\Omega), \text{Bernoulli}(\theta)^{\otimes \mathbb{N}})$$

Notre observation modélise le résultat d'une infinité de réalisations iid d'une variable de Bernoulli.

- **Le modèle exponentiel:** Ici on choisit $\Theta = \mathbb{R}_+$ et le modèle:

$$(\Omega, \mathcal{A}, \mathbb{P}_\theta) = (\mathbb{R}_+^{\mathbb{N}}, \mathcal{B}(\Omega), \mathcal{E}(\theta)^{\otimes \mathbb{N}})$$

Notre observation modélise le résultat d'une infinité de réalisations iid d'une variable exponentielle.

- **Le modèle Gaussien:** Ici on choisit $\Theta = \mathbb{R} \times \mathbb{R}_+$ et le modèle:

$$(\Omega, \mathcal{A}, \mathbb{P}_\theta) = (\mathbb{R}, \mathcal{B}(\Omega), \mathcal{N}(\theta)^{\otimes \mathbb{N}})$$

Notre observation modélise le résultat d'une infinité de réalisations iid d'une variable gaussienne.

De manière générale, si $(\Omega, \mathcal{A}, \mathbb{P}_\theta)$ est un modèle statistique, l'observation est une réalisation d'une variable à valeurs dans ce modèle, et le modèle statistique $(\Omega^n, \mathcal{A}^{\otimes n}, \mathbb{P}_\theta^{\otimes n})$ correspond à observer $n \in \mathbb{N} \cup \{\infty\}$ observation iid du premier modèle.