

Chapitre 2

Quelques domaines d'application

2.1 Chaînes linéaires et gaussiennes

2.1.1 Introduction

La distribution gaussienne joue un rôle essentiel dans la théorie des probabilités. L'importance des variables gaussiennes est en grande partie due au théorème central de la limite. Ce dernier nous informe que toute accumulation de petites fluctuations indépendantes, et de nature quelconque, se traduit asymptotiquement et irrémédiablement par une variable gaussienne. Tout phénomène résultant d'une addition d'effets aléatoires indépendants est donc nécessairement de nature gaussienne. Dans de nombreux problèmes pratiques, il est donc naturel de considérer comme gaussiennes les erreurs de mesures, et les erreurs de modélisation.

2.1.2 Formulation canonique

Les systèmes linéaires et gaussiens $X = (X_n)_{n \geq 0}$ sont définis par la donnée d'une suite de v.a. à valeurs réelles de lois marginales

$$\mathbb{P}^{(X_0, \dots, X_n)}(d(x_0, \dots, x_n)) = \left[\prod_{p=1}^n \frac{e^{-\frac{(x_p - a_p x_{p-1})^2}{2\sigma_p^2}}}{\sqrt{2\pi\sigma_p^2}} dx_p \right] \frac{e^{-\frac{x_0^2}{2\sigma_0^2}}}{\sqrt{2\pi\sigma_0^2}} dx_0 \quad (2.1)$$

Dans le formule ci-dessus $(a_n)_{n \geq 1}$ désigne une suite de nombres réels, et $(\sigma_n)_{n \geq 0}$ une suite de nombres strictement positifs.

En intégrant la formule (2.1) en la coordonnée x_n , on montre facilement que

$$\begin{aligned} & \mathbb{P}^{X_n | (X_0, \dots, X_{n-1})}(dx_n | x_0, \dots, x_{n-1}) \\ &= \mathbb{P}^{X_n | X_{n-1}}(dx_n | x_{n-1}) \\ &= M_n(x_{n-1}, dx_n) =_{\text{déf.}} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp \left\{ -\frac{1}{2\sigma_n^2} (x_n - a_n x_{n-1})^2 \right\} dx_n \end{aligned}$$

pour tout $n \geq 1$, et $x_0, \dots, x_{n-1} \in \mathbb{R}$. De même, on montre que la loi de la condition initiale X_0 est donnée par

$$\mathbb{P}^{X_0}(dx_0) = \eta_0(dx_0) =_{\text{déf.}} \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left\{ -\frac{x_0^2}{2\sigma_0^2} \right\} dx_0$$

La formule multiplicative (2.1) doit donc se lire comme une formule de Bayes séquentielle

$$\begin{aligned} \mathbb{P}^{(X_0, \dots, X_n)}(d(x_0, \dots, x_n)) &= \left[\prod_{p=1}^n \mathbb{P}^{X_p | X_{p-1}}(dx_p | x_{p-1}) \right] \mathbb{P}^{X_0}(dx_0) \\ &= \eta_0(dx_0) M_1(x_0, dx_1) \dots M_n(x_{n-1}, dx_n) \end{aligned}$$

2.1.3 Formulation dynamique

Notons $W = (W_n)_{n \geq 0}$ une suite de v.a. indépendantes et gaussiennes, et telles que

$$\mathbb{P}^{(W_0, \dots, W_n)}(d(w_0, \dots, w_n)) = \prod_{p=0}^n \mathbb{P}^{W_p}(dw_p) = \prod_{p=0}^n \left(\frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{w_p^2}{2} \right\} dw_p \right)$$

On associe à cette séquence de v.a. gaussiennes, le système dynamique

$$\begin{cases} X_n &= a_n X_{n-1} + \sigma_n W_n \\ X_0 &= \sigma_0 W_0 \end{cases} \quad (2.2)$$

Par construction, nous avons pour toute fonction bornée f

$$\begin{aligned}
& \mathbb{E}(f(X_n)|X_{n-1} = x_{n-1}) \\
&= \mathbb{E}(f(x_{n-1} + \sigma_n W_n)|X_{n-1} = x_{n-1}) \\
&= \int_{-\infty}^{+\infty} f(x_{n-1} + \sigma_n w_n) \frac{e^{-\frac{w_n^2}{2}}}{\sqrt{2\pi}} dw_n \\
&= \int_{-\infty}^{+\infty} f(x_n) \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{-\frac{1}{2\sigma_n^2} (x_n - x_{n-1})^2} dx_n = \int M_n(x_{n-1}, dx_n) f(x_n)
\end{aligned}$$

Il est donc équivalent de définir la chaîne X_n soit de façon dynamique, soit directement par la donnée de ses transitions de probabilités.

2.2 Processus de Poisson

Le processus de Poisson est souvent associé à des phénomènes de comptages dans le temps : arrivées de clients dans une file d'attente, nombres de transactions journalières autour d'une action boursière, arrivées et départs d'avions dans des aéroports, nombres d'appels dans un central téléphonique, etc. Nous renvoyons le lecteur à l'ouvrage de F.A. Haight [4] consacré aux différentes applications de ce processus.

D'un point de vue mathématique, le processus de Poisson est défini en terme d'une suite de v.a. $(T_n)_{n \geq 0}$, à valeurs positives, et de lois marginales données par la formule

$$\begin{aligned}
& \mathbb{P}^{(T_0, \dots, T_n)}(d(t_0, \dots, t_n)) \tag{2.3} \\
&= [1_{[0, \infty)}(t_0) \lambda e^{-\lambda t_0} dt_0] [1_{[t_0, \infty)}(t_1) \lambda e^{-\lambda(t_1 - t_0)} dt_1] \\
&\quad \dots \times [1_{[t_{n-1}, \infty)}(t_n) \lambda e^{-\lambda(t_n - t_{n-1})} dt_n]
\end{aligned}$$

Dans la formule ci-dessus, $\lambda > 0$ désigne un paramètre fixé. En interprétant (2.3) comme une formule de Bayes séquentielle, on obtient

$$\mathbb{P}^{(T_0, \dots, T_n)}(d(t_0, \dots, t_n)) = \left[\prod_{p=1}^n \mathbb{P}^{T_p|T_{p-1}}(dt_p|t_{p-1}) \right] \mathbb{P}^{T_0}(dt_0)$$

avec

$$\begin{aligned}\mathbb{P}^{T_0}(dt_0) &= 1_{[0,\infty)}(t_0) \lambda e^{-\lambda t_0} dt_0 \\ \mathbb{P}^{T_n|T_{n-1}}(dt_n|t_{n-1}) &= 1_{[t_{n-1},\infty)}(t_n) \lambda e^{-\lambda(t_n-t_{n-1})} dt_n\end{aligned}$$

Par un simple changement de variable, on montre aussi facilement que

$$U_0 =_{\text{déf.}} T_0 \quad \text{et} \quad U_n =_{\text{déf.}} (T_n - T_{n-1})$$

forment une suite de v.a. exponentielles et indépendantes de paramètre λ . Le processus continu de comptage des sauts

$$t \in \mathbb{R}_+ \mapsto N(t) = \sum_{n \geq 0} 1_{[T_n, \infty)}(t) \in \mathbb{N}$$

est appelé le Processus de Poisson d'intensité λ . Une réalisation de ce processus est donnée dans la figure suivante

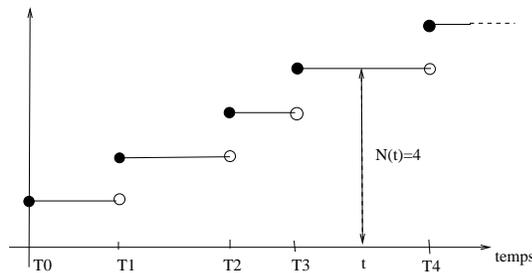


FIG. 2.1 – Processus de Poisson

2.3 Modèles stochastiques en traitement du signal

Le modèle que nous proposons est issu d'un modèle de filtrage de signaux radar. Cette chaîne de Markov est appelé le **modèle de Singer**. L'espace d'état

est ici \mathbb{R}^3 . Le signal correspond à l'évolution d'une cible abstraite ponctuelle formé de trois coordonnées

$$X_n = \begin{pmatrix} X_n^{(1)} \\ X_n^{(2)} \\ X_n^{(3)} \end{pmatrix}$$

La première coordonnée $X_n^{(1)}$ représente les accélérations, la seconde $X_n^{(2)}$ les vitesses, et la dernière $X_n^{(3)}$ les positions.

L'évolution du signal X est donné par les équations suivantes

$$\begin{cases} X_n^{(1)} = X_{n-1}^{(1)} + \epsilon_n W_n \\ X_n^{(2)} = (1 - \alpha \Delta) X_{n-1}^{(2)} + \beta \Delta X_n^{(1)} \\ X_n^{(3)} = X_{n-1}^{(3)} + \Delta X_n^{(2)} \end{cases}$$

avec des coefficients de viscosités $\alpha = 1$ et $\beta = 18$, et des conditions initiales gaussiennes $X_0^{(1)} \sim N(30, \sigma_0)$, $X_0^{(2)}$ et $X_0^{(3)} \sim N(500, \sigma_0)$ avec $\sigma_0 \in \{10, 100\}$. Le paramètre $\Delta = 10^{-2}$ correspond à la période d'échantillonnage du radar. La suite $(\epsilon_n)_{n \geq 1}$ est formée de v.a. indépendantes de Bernoulli de paramètre Δ . Ces dernières représentent les instants aléatoires où la cible accélère. Enfin la suite $(W_n)_{n \geq 1}$ représente les amplitudes des accélérations. On peut supposer par exemple que cette suite est formée de v.a. indépendantes et uniformes sur l'intervalle $[0, 60]$.

Le lecteur avisé aura très probablement noté que ce modèle discret correspond à la discrétisation du système dynamique suivant

$$\begin{cases} dX_t^{(1)} = W_t dN_t \\ \frac{dX_t^{(2)}}{dt} = -\alpha X_t^{(2)} + \beta X_t^{(1)} \\ \frac{dX_t^{(3)}}{dt} = X_t^{(2)} \end{cases}$$

où W_t, N_t désignent respectivement une trajectoire gaussienne, et poissonnienne.

Le système dynamique discret peut se représenter sous la forme vectorielle suivante

$$X_n = A X_{n-1} + U_n$$

avec la matrice A , et les vecteur aléatoires U_n , donnés par

$$A = \begin{pmatrix} 1 & 0 & 0 \\ \beta \Delta & (1 - \alpha \Delta) & 0 \\ 0 & \Delta & 1 \end{pmatrix} \quad \text{et} \quad U_n = \begin{pmatrix} \epsilon_n W_n \\ 0 \\ 0 \end{pmatrix}$$

Le processus d'observation est défini par les équations

$$\forall n \geq 0, \quad Y_n = X_n^{(3)} + \Delta V_n$$

Les perturbations de mesures $V = (V_n)_{n \geq 1}$ sont une suite de variables indépendantes des v.a. précédentes, et indépendantes de même loi gaussienne centrée et d'écart type $\sigma_v^2 = 100$. Le problème du filtrage consiste à calculer, ou plutôt à estimer, les lois conditionnelles du vecteur X_n sachant les observations $Y_0 = y_0, \dots, Y_n = y_n$, fournies par le radar.

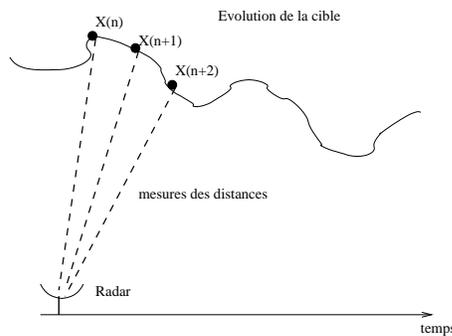


FIG. 2.2 – Mesures Radar

2.4 Modèles stochastiques en finance

2.4.1 Évolution de prix d'actions

L'évolution des prix d'actifs risqués dans les marchés financiers sont souvent modélisés par des processus aléatoires réels de la forme suivante :

$$Y_n = (1 + a_n(Y_{n-1}) \Delta + b_n(Y_{n-1}) \Delta W_n) \times Y_{n-1}$$

Les fonctions (a_n, b_n) représentent respectivement la tendance et la volatilité du marché. Le paramètre $\Delta > 0$ correspond à un pas de temps. Les fluctuations aléatoires du marché sont représentées par les accroissement

$$\Delta W_n = W_n - W_{n-1}$$

d'un processus aléatoire W_n , tel que $W_0 = 0$. Elles sont souvent choisies indépendantes, et de même loi gaussienne centrée et de variance Δ . On conviendra que ces accroissements aléatoires sont tels que

$$(\Delta W_n)^2 \simeq \Delta$$

A titre d'exemple, W_n peut correspondre à une marche aléatoire vers le haut ou vers le bas, et définie par la formule

$$\Delta W_n = \epsilon_n \sqrt{\Delta}$$

avec une suite de variables ϵ_n indépendantes, et distribuées selon la même loi

$$\mathbb{P}(\epsilon_n = 1) = 1 - \mathbb{P}(\epsilon_n = -1) = 1/2$$

En termes financiers, l'accroissement relatif du prix d'un actif sur une petite durée est donné par la formule

$$\frac{\Delta Y_n}{Y_{n-1}} = \frac{Y_n - Y_{n-1}}{Y_{n-1}} = a_n(Y_{n-1}) \Delta + b_n(Y_{n-1}) \Delta W_n$$

Le modèle introduit par Black et Scholes en 1973 correspond à la situation où les fonctions de tendance et volatilité sont constantes et homogènes en temps

$$a_n(x) = a \quad \text{et} \quad b_n(x) = b$$

Dans cette situation, l'équation précédente admet pour solution explicite :

$$\begin{aligned} Y_n &= \left[\prod_{p=1}^n (1 + a \Delta + b \Delta W_p) \right] Y_0 \\ &= \exp \left\{ \sum_{p=1}^n \log(1 + a \Delta + b \Delta W_p) \right\} Y_0 \end{aligned}$$

En utilisant le développement du logarithme autour de l'origine

$$\log(1 + u) \simeq u - \frac{u^2}{2}$$

on montre que

$$\begin{aligned} Y_n &\simeq \exp \left\{ \sum_{p=1}^n (a \Delta + b \Delta W_p) - \frac{1}{2} b^2 n \Delta \right\} Y_0 \\ &\simeq \exp \left\{ \left(a - \frac{b^2}{2} \right) n \Delta + b W_n \right\} Y_0 \end{aligned}$$

En pratique, les paramètres de tendance et volatilité sont de l'ordre de 10^{-p} log d'euros par minute, avec $p = 1, 2, 3, \dots$. On peut donc en première approximation, remplacer cette formulation exponentielle par

$$Y_n \simeq \exp \{a n\Delta + b W_n\} Y_0$$

2.4.2 Estimation des volatilités du marché

En pratique on observe les cours des actifs Y_n mais la tendance et la volatilité du marché sont des paramètres inconnus. L'estimation de ces quantités est l'un des problèmes les plus courants en mathématiques financières.

Pour résoudre numériquement ce problème, le Professeur René Carmona du Bendheim Center for Finance de Princeton, proposa en 2001 un modèle de filtrage de signaux analogue à celui du filtrage de signaux radar décrit dans la section précédente. L'idée est la suivante. Supposons tout d'abord que les tendances du marché sont des paramètres constants homogènes et connus $a_n(x) = a$, disons $a = 6 \cdot 10^{-3}$. On suppose de plus que la volatilité du marché est un paramètre aléatoire b_n oscillant autour de sa condition initiale b_0 selon l'équation suivante

$$b_n = -c (b_{n-1} - b_0) \Delta + d \Delta \widetilde{W}_n$$

où (c, d) désigne un couple de paramètres fixés, et $\Delta \widetilde{W}_n$ une suite d'accroissement aléatoires indépendants, du même type que ΔW_n . Pour fixer les idées on peut convenir que $b_0 = 10^{-1}$, $c = 2$, $d = 1$, et un pas de temps $\Delta = 10^{-3}$. La volatilité du marché financier est observée indirectement par la donnée des prix des actifs

$$Y_n/Y_{n-1} = (1 + a_n \Delta) + b_n \Delta W_n$$

Le problème de filtrage associé consiste donc à calculer les lois conditionnelles

$$\text{Loi}((b_0, \dots, b_n) \mid Y_0, \dots, Y_n)$$

du processus de volatilité connaissant l'évolution du prix des actifs.

2.5 Milieux absorbants

2.5.1 Introduction

En termes physiques, la chaîne de Markov suivante représente l'évolution d'une particule physique dans un puit de potentiel absorbant. Ce modèle peut

aussi s'interpréter comme une désintégration, ou encore comme une absorption de la radioactivité dans des containers de stockages de déchets nucléaires.

Très brièvement, une particule évolue aléatoirement dans un environnement absorbant avec un taux de survie $G(x) \in [0, 1]$, en chaque site x .

Les obstacles sont d'autant plus absorbants que les valeurs de G sont proches de 0. Inversement, la particule évolue librement dans des régions où G vaut 1. Les régions où le potentiel est strictement inférieur à 1 représentent des obstacles dans lesquels la particule peut être piégée.

Ces trappes peuvent représenter des **niveaux de sécurité** dans tout type d'environnements, tels des aéroports, des milieux carcéraux, des chaînes de production, des réseaux de télécommunications, des containers de stockage de radioactivité, des tissus cellulaires, etc.

En pharmacologie, ces modèles d'absorption sont aussi utilisés pour modéliser l'évolution des taux de leucocytes dans des traitements du cancer. Dans ce contexte, les trappes reflètent des chutes de niveaux de leucocytes dans lesquels un individu risque de décéder.

En biologie et en recherche médicale, ces modèles d'absorption sont aussi utilisés pour modéliser l'évolution de photons émis par un laser sur un tissu cellulaire. Les trappes représentent des régions cellulaires, telles des tumeurs plus sombres absorbant les photons. L'analyse de ces modèles permet de localiser et d'analyser ces régions de gonflement pathologiques des tissus.

Enfin, ces modèles peuvent aussi s'interpréter comme des **processus économiques** tels les évolutions de portefeuilles, ou tout autre indicateur économique, dans des milieux financiers ou géopolitiques-politiques.

Dans tous ces domaines d'applications, l'un des problèmes majeurs est de calculer les probabilités de défaillances, autrement dit les probabilités pour que l'absorption de la particule ne soit pas effective à des instants donnés.

2.5.2 Description des modèles

Pour fixer les idées, nous conviendrons par la suite que l'espace des états est donné par le réseau \mathbb{Z}^d . Plus formellement, on adjoint à l'espace des sites \mathbb{Z}^d , un point cimetière " c ". On considère alors une particule évoluant sur l'espace augmenté $E = \mathbb{Z}^d \cup \{c\}$, selon une transition $Y_n \rightsquigarrow Y_{n+1}$ se décomposant en deux étapes

$$Y_n \xrightarrow{\text{absorption}} \widehat{Y}_n \xrightarrow{\text{évolution}} Y_{n+1}$$

Pour construire le mécanisme d'absorption, on considère à chaque instant n , une collection de v.a. indépendantes $(\epsilon_n(x))_{x \in \mathbb{Z}^d}$, de Bernoulli à valeurs dans $\{0, 1\}$, de paramètre $G(x)$. Sur l'évènement $\epsilon_n(x) = 0$, le site x devient une pure trappe hautement absorbante. Inversement, si $\epsilon_n(x) = 1$, le site x reste viable, et la particule peut le traverser sans encombres.

Le mécanisme d'évolution libre est analogue à celui de la marche aléatoire sur \mathbb{Z}^d présenté dans l'exemple 1.3.4. Pour le décrire, on se donne une mesure de probabilité p , sur l'ensemble des $2d$ vecteurs unitaires directionnels

$$\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$$

ainsi qu'une suite de vecteurs indépendants $(U_n)_{n \geq 1}$ de même loi p sur \mathcal{U} . On suppose que les suites de v.a. d'exploration U_n , et de v.a. d'absorption $\epsilon_n(x)$, sont indépendantes. On convient enfin que $\epsilon_n(c) = 0 = g(c)$.

Les transitions élémentaires de la chaîne sont définies récursivement de la façon suivante. Supposons que la particule se trouve à l'instant n sur un site $Y_n = x \in \mathbb{Z}^d$ (si $Y_n = c$, on pose $Y_{n+1} = c$).

1. Avec une probabilité $(1 - G(x))$, la particule est tuée, puis placée dans l'état cimetière. Dans ce cas, on pose $\widehat{Y}_n = c$. Dans le cas contraire, la particule reste active et l'on pose $\widehat{Y}_n = x$. Plus formellement, on a

$$\widehat{Y}_n = \epsilon_n(Y_n) Y_n + (1 - \epsilon_n(Y_n)) c$$

2. Comme il n'y a semble-t-il pas de vie après la mort, lorsque la particule a été tuée, elle reste inactive à l'instant suivant. Dans ce cas, on pose $Y_{n+1} = c$. Dans le cas contraire, la particule est encore active, et effectue un mouvement exploratoire analogue à celui d'une marche aléatoire sur \mathbb{Z}^d . Plus formellement, on pose dans ce dernier cas

$$Y_{n+1} = \widehat{Y}_n + U_{n+1}$$

Par construction, la suite

$$Y_0 \rightarrow \widehat{Y}_0 \rightarrow \dots \rightarrow Y_n \rightarrow \widehat{Y}_n \rightarrow Y_{n+1} \rightarrow \widehat{Y}_{n+1} \rightarrow \dots$$

forme une chaîne de Markov de transitions

$$\begin{aligned} \mathbb{P}(\widehat{Y}_n = y \mid Y_n = x) &= G(x) 1_x(y) + (1 - G(x)) 1_c(y) \\ \mathbb{P}(Y_{n+1} = z \mid \widehat{Y}_n = y) &= 1_c(y) 1_c(z) + 1_{\mathbb{Z}^d}(y) K(y, z) \end{aligned}$$

avec la probabilité de transition

$$K(y, z) = \sum_{u \in \mathcal{U}} p(u) 1_{y+u}(z)$$

La figure suivante montre deux réalisations de trajectoires absorbées dans des "poches" d'obstacles associés à des régions où le potentiel $G(x) < 1$. Lorsque $G(x) = 1$, la particule ne subit pas le mécanisme d'absorption, et évolue librement comme une marche aléatoire sur \mathbb{Z}^2 . Les traits pointillés témoignent du fait qu'une particule visitant une poche d'obstacles s'essouffle. Sa durée de vie diminue à chaque instant, en "subissant" les v.a. trappes de Bernoulli.

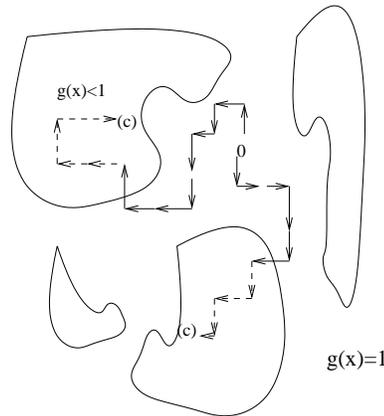


FIG. 2.3 – Particule absorbée

Exercice 2.5.1 On note T l'instant d'absorption de la particule

$$T = \inf \{n \geq 0 : \widehat{Y}_n = c\} = \inf \{n \geq 0 : \epsilon_n(Y_n)_n = c\}$$

1. Montrer que

$$\mathbb{P}(T > n) = \mathbb{E} \left(\prod_{p=0}^n G_p(X_p) \right)$$

où X_n désigne une chaîne de Markov sur \mathbb{Z}^d de transitions de probabilités K , et de même loi que Y_0 sur \mathbb{Z}^d . Lorsque le potentiel est uniformément majoré $G(x) \leq e^{-\lambda}$, avec $\lambda > 0$, montrer que

$$\mathbb{P}(T > n) \leq e^{-\lambda(n+1)}$$

Interpréter ce résultat.

2. Vérifier que la loi d'une particule non absorbée est donnée pour toute fonction bornée f sur \mathbb{Z}^d par la formule renormalisée de Feynman-Kac

$$\mathbb{E}(f(Y_n) \mid T > n) = \frac{\mathbb{E} \left(f(X_n) \prod_{p=0}^n G_p(X_p) \right)}{\mathbb{E} \left(\prod_{p=0}^n G_p(X_p) \right)}$$

2.6 Dynamiques de population avec branchements

Dans la section 1.3.6 nous avons étudié des processus de branchement élémentaires dans lesquels chaque individu donne naissance à un certain nombre d'enfants. Ces modèles simplifiés sont assez éloignés de la réalité. Tout d'abord ils ne tiennent pas compte des explorations des individus dans l'espace. De plus, les taux de branchement sont totalement indépendants des régions plus ou moins accueillantes dans lesquelles se trouvent les populations.

Dans ce qui suit, nous allons essayer de raffiner ces modèles stochastiques pour rendre compte de ces deux paramètres, et essayer de coller au mieux à la réalité scientifique. Nous conviendrons que les individus évoluent à chaque instant n dans un espace d'état E_n . Les populations d'individus seront représentées par des vecteurs de E_n^p . Le paramètre entier $p \geq 0$ correspond à la taille des populations. Lorsque $p = 0$, on conviendra que l'espace d'état se réduit à un état cerceuil, ou cimetière $E_n^0 = \{c\}$. L'espace d'état du système est donc donné par l'ensemble

$$S_n = \cup_{p \geq 0} E_n^p$$

La dynamique d'exploration de chaque individu est associée à des transitions markoviennes M_n de E_{n-1} vers E_n . Les mécanismes de branchements dépendent

de fonctions potentiel $G_n : E_N \rightarrow [0, \infty)$ représentant les différents degrés de "fertilité" du milieu. Plus précisément, un individu sur un site $x_n \in E_n$ donnera naissance à un nombre aléatoire d'enfants $g_n(x_n)$ avec

$$\mathbb{E}(g_n(x_n)) = G_n(x_n)$$

Pour illustrer ce modèle, on peut supposer que les populations évoluent dans le réseau du plan $E_n = \mathbb{Z}^2$ selon une marche aléatoire simple, et les variables aléatoires de branchement sont de Bernoulli

$$\begin{aligned} \mathbb{P}(g_n(x_n) = [G_n(x)] + 1) &= G_n(x_n) - [G_n(x_n)] \\ &= 1 - \mathbb{P}(g_n(x_n) = [G_n(x)]) \end{aligned}$$

où $[a]$ désigne la partie entière d'un nombre $a \in [0, \infty)$. On peut aussi choisir des branchements poissonniens

$$\forall m \in \mathbb{N} \quad \mathbb{P}(g_n(x_n) = m) = \frac{G_n(x_n)^m}{m!} \exp(-G_n(x_n))$$

Pour définir plus formellement notre dynamique de population, il convient d'introduire une suite de variable $(g_n^i(x_n))_{i \geq 1}$ indépendantes, et de même loi que $g_n(x_n)$. On supposera de plus que les variables sur des sites distincts sont indépendantes.

Nous sommes enfin en mesure de construire récursivement le modèle. Initialement, la population \mathcal{X}_0 est formé d'un seul individu \mathcal{X}_0^1 dans l'état $x_0 \in E_0$

$$\mathcal{X}_0 = \mathcal{X}_0^1 = x_0 \in E_0^{p_0} = E_0^1$$

Cet individu donne naissance à

$$\widehat{p}_0 = g_0^1(\mathcal{X}_0^1)$$

enfants que l'on note

$$\widehat{\mathcal{X}}_0 = (\widehat{\mathcal{X}}_0^1, \dots, \widehat{\mathcal{X}}_0^{\widehat{p}_0}) \in E_0^{\widehat{p}_0}$$

Chacun d'entre eux explore aléatoirement l'espace $\widehat{\mathcal{X}}_0^i \rightsquigarrow \mathcal{X}_1^i$ selon la transition M_1 . Cette transition revient à simuler \widehat{p}_0 variables aléatoires \mathcal{X}_1^i de loi $M_1(\widehat{\mathcal{X}}_0^i, dx_1)$. Lorsque cette étape d'exploration est terminée, nous avons une population formée de

$$p_1 = \widehat{p}_0$$

individus. Cette transition peut s'exprimer de façon synthétique par la formule suivante :

$$\widehat{\mathcal{X}}_0 = (\widehat{\mathcal{X}}_0^1, \dots, \widehat{\mathcal{X}}_0^{\widehat{p}_0}) \rightsquigarrow \mathcal{X}_1 = (\mathcal{X}_1^1, \dots, \mathcal{X}_1^{p_1})$$

Durant l'étape de branchement suivante chaque individu \mathcal{X}_1^i donne naissance à $g_1^1(\mathcal{X}_1^i)$ enfants. À la fin de ce processus nous avons une population formée de \widehat{p}_1 individus

$$\widehat{\mathcal{X}}_1 = (\widehat{\mathcal{X}}_1^1, \dots, \widehat{\mathcal{X}}_1^{\widehat{p}_1}) \in E_1^{\widehat{p}_1} \quad \text{avec} \quad \widehat{p}_1 = \sum_{i=1}^{p_1} g_1^1(\mathcal{X}_1^i)$$

Chacun de ces individus explore aléatoirement l'espace E_2 selon la transition M_2 , etc.

$$\mathcal{X}_n = (\mathcal{X}_n^i)_{1 \leq i \leq p_n} \xrightarrow{\text{branchement}} \widehat{\mathcal{X}}_n = (\widehat{\mathcal{X}}_n^i)_{1 \leq i \leq \widehat{p}_n} \xrightarrow{\text{exploration}} \mathcal{X}_{n+1}$$

Si le système meurt au bout d'un certain temps n , nous avons $\widehat{p}_n = 0$. Dans ce cas, on pose

$$\widehat{\mathcal{X}}_p = \mathcal{X}_{p+1} = 0$$

pour tout les instants suivants $p \geq n$.

Exercice 2.6.1 1. Pour toutes fonctions bornées f_n sur E_n , exprimer les variables aléatoires $\sum_{i=1}^{\widehat{p}_n} f_n(\widehat{\mathcal{X}}_n^i)$ en fonction de p_n , \mathcal{X}_n , et des variables g_n^i (on utilisera la convention $\sum_{\emptyset} = 0$, lorsque la population est éteinte).

2. On considère les mesures empiriques aléatoires

$$s(x_n) =_{def} \sum_{i=1}^p \delta_{x_n^i} \quad \text{pour chaque} \quad x_n = (x_n^i)_{1 \leq i \leq p} \in E_n^p$$

Pour toute fonction bornée f_n sur E_n , calculer les moyennes conditionnelles

$$\mathbb{E}(s(\widehat{\mathcal{X}}_n)(f_n) \mid \mathcal{X}_n) \quad \text{et} \quad \mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1}) \mid \widehat{\mathcal{X}}_n)$$

En déduire que

$$\mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1}) \mid \mathcal{X}_n) = s(\mathcal{X}_n)(G_n M_{n+1}(f_{n+1}))$$

3. Vérifier que les premiers moments des dynamiques de population sont donnés par la formule suivante

$$\mathbb{E}(s(\mathcal{X}_{n+1})(f_{n+1})) = \mathbb{E}_{x_0}(f_{n+1}(X_{n+1})) \prod_{k=0}^n G_k(X_k)$$

2.7 Algorithme génétique

En termes biologiques, les algorithmes génétiques représentent les dynamiques de population à nombre constant d'individus. Ces individus explorent des régions plus ou moins accueillantes, selon des mécanismes de mutation et de sélection. Les individus meurent ou donnent naissance à des enfants suivant la qualité de leur milieu, mesurée en terme d'une fonction potentiel.

Dans la section 3.5.3, nous montrerons que ces modèles génétiques sont une simple expression "microscopique" des modèles d'évolution-absorption étudiés dans la section 2.5.

Leurs interprétations sont à l'image des différents domaines d'applications décrits à la page 50.

Lorsque la fonction potentiel représente des **niveaux de sécurité**, tels des risques de collisions dans des aéroports, l'étape de sélection permet par exemple de choisir les configurations les moins sécurisées. L'étape de mutation consiste alors à explorer plus en profondeur les risques de collisions suivantes. Dans ce contexte, l'algorithme génétique peut s'interpréter comme des séquences d'évolutions probables conduisant à des collisions d'avion.

En pharmacologie, la fonction potentiel peut représenter les différents niveaux de leukocytes dans un organisme vivant. Dans ce contexte, l'étape de mutation représente les différentes possibilités d'évolution de ces niveaux. La sélection permet d'évaluer les chutes possibles, et les entrées dans des niveaux mortels. L'algorithme génétique correspondant permet de décrire les histoires possibles conduisant aux décès d'un organisme.

Enfin, dans des **modèles économiques**, les potentiels peuvent représenter des valeurs de portefeuilles. L'étape de mutation consiste à décrire les évolutions envisageables de ces quantités partir de niveaux donnés. L'étape de sélection permet de choisir ces propositions suivant leurs tendances à la hausse ou à la baisse. Le modèles génétiques correspondent alors à des évolutions de portefeuilles vers la ruine ou la fortune dans des milieux financiers.

Enfin, dans les **modèles d'apprentissage** développés en intelligence artificielle, les potentiels représentent la qualité d'une proposition émise. L'étape de mutation consiste tout d'abord à élaborer des séquences de propositions envisageables, plus ou moins bonnes. L'étape de sélection permet ensuite d'affiner le raisonnement, en choisissant les propositions les mieux adaptées pour compléter une action donnée, ou pour reconstruire une information partiellement observée.

En termes mathématiques, un algorithme génétique est une chaîne de Markov $X_n = (X_n^i)_{1 \leq i \leq N}$ sur un espace produit E^N . Pour fixer les idées on pourra supposer que l'espace d'état est donné par $E = \mathbb{Z}^d$ ou $E = \mathbb{R}^d$, avec $d \in \mathbb{N} - \{0\}$. Chacune des composantes X_n^i représente la position de l'individu de label i , avec $i = 1, \dots, N$.

L'évolution de cette chaîne se décompose en deux mécanismes bien distincts. Le premier correspond à une sélection des individus, selon un certain critère de qualité. Le second est une exploration pure de l'espace des états. En biologie, cette étape est souvent appelé mutation, par référence au fait que les codes génétiques des individus changent au cours du temps.

$$X_n = (X_n^i)_{1 \leq i \leq N} \xrightarrow{\text{sélection}} \widehat{X}_n = (\widehat{X}_n^i)_{1 \leq i \leq N} \xrightarrow{\text{mutation}} X_{n+1} = (X_{n+1}^i)_{1 \leq i \leq N}$$

2.7.1 Sélection/Adaptation

Pour décrire l'étape de sélection, on se fixe une fonction potentiel, strictement positive, G sur l'espace E . Une fois connue la configuration de la chaîne $X_n = (X_n^i)_{1 \leq i \leq N}$ au temps n , l'étape de sélection consiste à simuler N v.a. $(\widehat{X}_n^i)_{1 \leq i \leq N}$ indépendantes de même loi

$$\sum_{i=1}^N \frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)} \delta_{X_n^i}$$

Autrement dit chaque v.a. \widehat{X}_n^k choisit l'une des valeurs X_n^i , avec la probabilité $\frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)}$. On remarquera que ce procédé de sélection peut aussi s'interpréter comme un **mécanisme de naissances et morts**. Dans cette interprétation, les individus X_n^i disparaissent, ou donnent naissance à un certain nombre de copies.

Il existe divers variantes pour sélectionner les individus les mieux adaptés au potentiel G . Dans le cas où le potentiel G est à valeurs dans $[0, 1]$, il est bien plus naturel "d'accepter" chaque individu X_n^i avec une probabilité $G(X_n^i)$, et de le remplacer (avec une probabilité $[1 - G(X_n^i)]$) par un individu choisi avec la loi discrète

$$\sum_{i=1}^N \frac{G(X_n^i)}{\sum_{j=1}^N G(X_n^j)} \delta_{X_n^i}$$

Plus formellement, ce mécanisme de sélection est équivalent à poser pour chaque $i = 1, \dots, N$

$$\widehat{X}_n^i = \begin{cases} X_n^i & \text{avec probabilité } G(X_n^i) \\ \widetilde{X}_n^i & \text{avec probabilité } 1 - G(X_n^i) \end{cases}$$

où \widetilde{X}_n^i désigne une v.a. de loi $\sum_{j=1}^N \frac{G(X_n^j)}{\sum_{k=1}^N G(X_n^k)} \delta_{X_n^j}$.

Pour des fonctions potentiel pouvant s'annuler sur certaines régions de l'espace, il est possible que tous les individus aient des potentiels nul. Dans cette situation, l'algorithme est stoppé.

2.7.2 Mutation/Exploration

Durant la phase de mutation, les individus sélectionnés explorent l'espace indépendamment les uns des autres, selon des transitions de probabilités élémentaires $M(x, y)$. Autrement dit, nous avons

$$\widehat{X}_n^i \rightsquigarrow X_{n+1}^i$$

où X_{n+1}^i désigne une v.a. de loi $M(\widehat{X}_{n-1}^i, \cdot)$.

Plus formellement, nous avons

$$\mathbb{P}(X_{n+1}^1 \in dx^1, \dots, X_{n+1}^N \in dx^N) = M(\widehat{X}_{n-1}^1, dx^1) \dots M(\widehat{X}_{n-1}^N, dx^N)$$

À titre d'exemple, si $M(x, y)$ désigne la matrice de transition d'une marche aléatoire sur \mathbb{Z}^d , on peut réaliser dynamiquement ces explorations locales en posant

$$X_{n+1}^i = \widehat{X}_{n-1}^i + U_{n+1}^i$$

où U_{n+1}^i désigne une suite de v.a. indépendantes de même loi p , sur l'ensemble des vecteurs unitaires directionnels $\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$.

Un exemple schématique d'évolution de $N = 4$ individus est représenté dans la figure suivante. Les nombres entiers entre parenthèse correspondent au nombre d'individus sur le site en question, après l'étape de sélection.

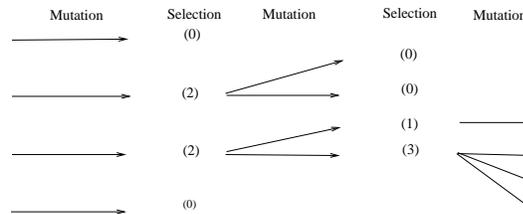


FIG. 2.4 – Algorithme génétique ($N = 4$)

Exercice 2.7.1 Décrire mathématiquement, et schématiquement, l'algorithme génétique sur \mathbb{Z} associé à la fonction de potentiel indicatrice $G(x) = 1_{[-L,L]}$, avec $L \geq 1$. On conviendra que les mutations sont données par les transitions d'une marche aléatoire sur \mathbb{Z} , et l'on initialisera les individus en l'origine.

Exercice 2.7.2 Décrire l'algorithme génétique sur \mathbb{R} associé à des mutations gaussiennes

$$M(x, dy) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y-x)^2 \right\} dy$$

et un potentiel quadratique centré autour d'un point $a \in \mathbb{R}$

$$G(x) = \exp \left\{ -\frac{1}{2}(x-a)^2 \right\}$$

2.8 Modèles d'arbres généalogiques

2.8.1 Modèles non homogènes

L'algorithme génétique décrit dans la section précédente peut être étendu de façon naturelle à des espaces d'états E_n dépendants du paramètre temporel $n \in \mathbb{N}$. Dans ce contexte, les individus X_n^i vivent à chaque instant n dans l'espace E_n . Les sélections s'effectuent dans ces mêmes espaces, tandis que les mutations s'expriment comme des passages aléatoires d'un état de E_n vers un nouvel état dans E_{n+1} .

Plus formellement, les populations d'individus sont données par des N -uplets

$$X_n = (X_n^i)_{1 \leq i \leq N} \in E_n^N \quad \text{et} \quad \hat{X}_n = (\hat{X}_n^i)_{1 \leq i \leq N} \in E_n^N$$

où E_n^N désigne l'espace produit $(E_n \times \dots \times E_n)$, avec N termes.

Dans ce contexte, les transitions de sélection/mutations s'expriment entre des états non homogènes :

$$X_n \in E_n^N \xrightarrow{\text{sélection}} \hat{X}_n = E_n^N \xrightarrow{\text{mutation}} X_{n+1} = E_{n+1}^N$$

Supposons que les sélections à chaque instant n soient aussi dictées par des potentiels non homogènes

$$G_n : x_n \in E_n \mapsto G_n(x_n) \in [0, 1]$$

Dans ce cas, le mécanisme de sélection dans E_n , s'exprime sous la forme suivante :

$$\widehat{X}_n^i = \begin{cases} X_n^i & \text{avec proba } G_n(X_n^i) \\ \widetilde{X}_n^i & \text{avec proba } 1 - G_n(X_n^i) \end{cases}$$

où \widetilde{X}_n^i désigne une v.a. de loi

$$\sum_{j=1}^N \frac{G_n(X_n^j)}{\sum_{k=1}^N G_n(X_n^k)} \delta_{X_n^j}$$

Les individus ainsi sélectionnés \widehat{X}_n^i vivent dans l'espace E_n .

Durant la mutation, ces individus passent de l'état E_n vers un nouvel état E_{n+1} , selon des transitions de probabilités $M_{n+1}(x_n, dx_{n+1})$ de E_n vers E_{n+1} . Autrement dit, nous avons

$$\widehat{X}_n^i (\in E_n) \rightsquigarrow X_{n+1}^i (\in E_{n+1})$$

où X_{n+1}^i désigne une v.a. de loi $M_{n+1}(\widehat{X}_{n-1}^i, \cdot)$. Ces nouveaux individus sont alors sélectionnés en fonction d'un potentiel G_{n+1} sur E_{n+1} , puis ils mutent de E_{n+1} vers E_{n+2} selon une transition markovienne de E_{n+1} vers E_{n+2} , etc.

2.8.2 Modèles trajectoriels

Dans ce qui précède, nous n'avons pas précisé la nature des espaces E_n , ni a fortiori celle des transitions $M_n(x_{n-1}, dx_n)$. Tout ceci semble donc bien abstrait ! Revenons donc sur terre en supposant que les espaces E_n sont donnés par des espaces produits

$$E_n = \underbrace{E' \times \dots \times E'}_{(n+1)\text{-fois}}$$

où E' désigne un ensemble quelconque, suffisamment réguliers. Pour fixer les idées, on pourra supposer que $E = \mathbb{Z}^d$, ou $E = \mathbb{R}^d$. On conviendra que les points de E_n sont donnés par des $(n+1)$ -uplets représentant des trajectoires de longueur n dans l'espace E'

$$x_n = (x'_0, \dots, x'_n) \in E_n = (E')^{n+1}$$

L'algorithme génétique précédent est, à chaque étape $n \in \mathbb{N}$, formé de N variables aléatoires trajectorielles, à valeurs dans E_n , que l'on notera

$$X_n^i = (X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i) \in E_n = (E')^{n+1}, \quad 1 \leq i \leq N.$$

et

$$\widehat{X}_n^i = (\widehat{X}_{0,n}^i, \widehat{X}_{1,n}^i, \dots, \widehat{X}_{n,n}^i) \in E_n = (E')^{n+1}, \quad 1 \leq i \leq N.$$

Chacun des individus X_n^i , et \widehat{X}_n^i représente un chemin dans E' de l'origine jusqu'au temps n .

Il est important de souligner que la population initiale

$$X_0^i = X_0'^i \in E_0 = E' \quad \text{avec} \quad 1 \leq i \leq N$$

est tout simplement formée de N variables aléatoires à valeurs dans E' .

Sélection trajectorielle :

Dans notre cadre trajectoriel, le mécanisme de sélection, peut s'interpréter comme une sélection de trajectoires, en fonction des différents potentiels

$$G_n(X_n^i) = G_n(X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i)$$

L'expression de cette transition reste inchangée. Nous avons à nouveau

$$\widehat{X}_n^i = \begin{cases} X_n^i & \text{avec proba } G_n(X_n^i) \\ \tilde{X}_n^i & \text{avec proba } 1 - G_n(X_n^i) \end{cases}$$

où $\tilde{X}_n^i = (\tilde{X}_{0,n}^i, \tilde{X}_{1,n}^i, \dots, \tilde{X}_{n,n}^i)$ désigne une v.a. de loi

$$\sum_{j=1}^N \frac{G_n(X_n^j)}{\sum_{k=1}^N G_n(X_n^k)} \delta_{X_n^j} = \sum_{j=1}^N \frac{G_n(X_{0,n}^j, \dots, X_{n,n}^j)}{\sum_{k=1}^N G_n(X_{0,n}^k, \dots, X_{n,n}^k)} \delta_{(X_{0,n}^j, \dots, X_{n,n}^j)}$$

Les trajectoires ainsi sélectionnés \widehat{X}_n^i vivent désormais dans l'espace de chemins $E_n = (E')^{n+1}$.

Mutation trajectorielle :

L'étape de mutation dépend uniquement de la nature des transitions $M_{n+1}(x_n, dx_{n+1})$ de E_n dans E_{n+1} . Supposons que ces dernières correspondent aux transitions de probabilités d'un processus historique associé à une évolution markovienne sur E' .

Dans cette situation, on rappelle que ces transitions $M_{n+1}(x_n, dy_{n+1})$ s'expriment sous la forme suivante :

$$M_{n+1}((x'_0, \dots, x'_n), d(y'_0, \dots, y'_{n+1})) = \delta_{(x'_0, \dots, x'_n)}(d(y'_0, \dots, y'_n))M'_{n+1}(y'_n, dy'_{n+1})$$

M'_{n+1} est une transition de Markov de E' vers lui même. Autrement dit, simuler une variable aléatoire trajectorielle de loi

$$M_{n+1}((x'_0, \dots, x'_n), d(y'_0, \dots, y'_{n+1}))$$

revient à conserver tout d'abord le segment de trajectoire

$$x_n = (x'_0, \dots, x'_n)$$

On lui adjoint ensuite une extension élémentaire aléatoire de loi $M'_{n+1}(x'_n, dy'_{n+1})$

$$x'_n \rightsquigarrow y'_{n+1}$$

Durant l'étape de mutation correspondante

$$\begin{aligned} \widehat{X}_n^i &= \underbrace{(\widehat{X}_{0,n}^{i}, \dots, \widehat{X}_{n,n}^{i})}_{\downarrow} \in E_n^N \\ X_{n+1}^i &= \left(\underbrace{(X_{0,n+1}^{i}, \dots, X_{n,n+1}^{i})}_{\downarrow}, X_{n+1,n+1}^{i} \right) \in E_{n+1}^N \end{aligned}$$

chaque chemin s'étend selon un déplacement élémentaire de loi M'_{n+1} , c'est à dire

$$\begin{aligned}
X_{n+1}^i &= \left(\underbrace{[X'_{0,n+1}, \dots, X'_{n,n+1}]}, X'_{n+1,n+1} \right) \in E_{n+1} = (E')^{n+2} \\
&\quad \parallel \\
&= \left(\underbrace{[\widehat{X}'_{0,n}, \dots, \widehat{X}'_{n,n}]}, X'_{n+1,n+1} \right) = \left(\widehat{X}_n^i, X'_{n+1,n+1} \right)
\end{aligned}$$

avec une variable aléatoire $X'_{n+1,n+1}$ de loi $M'_{n+1}(\widehat{X}'_{n,n}, \cdot)$.

À titre d'exemple, dans le cas où M'_{n+1} désigne la matrice de transition d'une marche aléatoire sur $E' = \mathbb{Z}^d$, on peut réaliser dynamiquement ces explorations locales en posant

$$X'_{n+1,n+1} = \widehat{X}'_{n,n} + U_{n+1}^i$$

où U_{n+1}^i désigne une suite de v.a. indépendantes de même loi p , sur l'ensemble des vecteurs unitaires directionnels

$$\mathcal{U} = \{u \in \mathbb{Z}^d : |u| = 1\}$$

Lignes ancestrales :

L'algorithme génétique trajectorien décrit ci-dessus correspond bien à une évolution d'arbres généalogiques.

Dans cette interprétation, chaque trajectoire

$$X_n^i = (X_{0,n}^i, X_{1,n}^i, \dots, X_{n,n}^i) \in E_n = (E')^{n+1}$$

représente la ligne ancestrale de l'individu courant $X_{n,n}^i$ à la date n . Les coordonnées $X_{p,n}^i$, avec $0 \leq p \leq n$, correspondent aux différents ancêtres de cet individu, à chaque niveau temporel $0 \leq p \leq n$.

Dans ce contexte, $X_{0,n}^i$ représente l'ancêtre initial, $X_{1,n}^i$ sa première descendance, $X_{2,n}^i$ sa seconde, etc. Dans le modèle général que nous avons développé, les disparitions et les sélections de lignes ancestrales dépendent de la qualité de la

trajectoire de descendance complète. Les différents degrés d'adaptation d'une lignée complète

$$x_n = (x'_0, \dots, x'_n) \in E_n$$

sont mesurés par une fonction potentiel $G_n(x'_0, \dots, x'_n)$. Lorsque ces fonctions ne dépendent que des composantes terminales, c'est à dire lorsque l'on a

$$G_n(x'_0, \dots, x'_n) = G'_n(x'_n) ,$$

la sélection des lignes ancestrale ne dépend que de la qualité d'adaptation du **dernier descendant**. Dans cette situation, nous laissons le soin au lecteur de se convaincre que ce modèle trajectorien correspond à l'arbre généalogique d'une population d'individus explorant l'espace E' selon M'_n , et s'adaptant en fonction des potentiels de selection G'_n .

2.9 Chaînes renforcées

Ces modèles de renforcement sont associés à des évolutions aléatoires où chaque transition dépend de la mesure d'occupation des sites visités dans le passé. Ces processus permettent de représenter des stratégies humaines d'exploration de sites comme des rues, des magasins, des restaurants, ou tout autre endroits d'une ville. Dans ce contexte, un site donné est d'autant plus attractif s'il répond à un certain critère de qualité, ou lorsqu'il a été déjà visité de nombreuses fois par le passé !

Le critère de qualité est représenté par une fonction potentiel $G : E \rightarrow [0, 1]$, sur un espace d'états E . Les valeurs de G en un site x sont d'autant plus grandes que le site est attrayant. Ainsi, notre explorateur se trouvant en $X_n = x$, au temps n , choisit d'y rester avec une probabilité $G(x)$, soit préfère retourner vers l'un des sites précédemment visité X_0, \dots, X_{n-1} . Dans cette situation, il choisit un nouveau site avec une probabilité

$$\sum_{p=0}^{n-1} \frac{G(X_p)}{\sum_{q=0}^{n-1} G(X_q)} \delta_{X_p}$$

Cette sélection aléatoire peut être vue comme une transition élémentaire

$$X_n \rightsquigarrow \widehat{X}_n$$

de probabilités de transitions

$$G(X_n) \delta_{X_n} + (1 - G(X_n)) \sum_{p=0}^{n-1} \frac{G(X_p)}{\sum_{q=0}^{n-1} G(X_q)} \delta_{X_p}$$

Partant du site sélectionné \widehat{X}_n , l'individu effectue une nouvelle exploration de la région $\widehat{X}_n \rightsquigarrow X_{n+1}$, selon une transition de probabilité M_{n+1} . Autrement dit, nous avons

$$\mathbb{P}(X_{n+1} \in dy \mid \widehat{X}_n = x) = M_{n+1}(x, dy)$$

Cette transition M_n peut, par exemple, représenter une exploration uniforme des sites voisins à \widehat{X}_n .

Le lecteur aura certainement noté que X_n n'est pas une chaîne de Markov, mais le processus historique possède toujours la propriété markovienne.

Exercice 2.9.1 Examiner la situation où le potentiel est constant $G(x) = \epsilon$, avec $\epsilon \in [0, 1]$. Établir une analogie avec la loi des grands nombres.

2.10 Algorithme de Robbins-Monro

2.10.1 Introduction

Cet algorithme aléatoire permet d'étudier les lignes de niveaux

$$U_a = \{x \in E : U(x) = a\}, \quad a \in \mathbb{R}^d$$

d'une fonction $U : E \rightarrow \mathbb{R}^d$ sur un espace d'états donné E .

Potentiel gradient convexe

Cet exemple correspond au potentiel gradient $U(x) = \nabla V(x)$ d'une fonction $V : E = \mathbb{R}^2 \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et strictement convexe. Dans ce cas, l'ensemble

$$U_0 = \{x \in \mathbb{R}^2 : \nabla V(x) = 0\}$$

se résume à l'unique minimum x_0 de V .

Fonction de répartition

Un autre exemple plus probabiliste consiste à prendre pour U la fonction de répartition d'une variable aléatoire réelle Y

$$U : x \in E = \mathbb{R} \mapsto U(x) = \mathbb{P}(Y \leq x) = \mathbb{E}(1_{Y \leq x}).$$

Dans ce contexte, l'ensemble $U_{1/2} = \{x \in \mathbb{R}^d : \mathbb{P}(Y \leq x) = 1/2\} = \{x_{1/2}\}$ est à nouveau réduit à un point $x_{1/2}$, appelé le médiane de la distribution de Y .

2.10.2 Description du modèle

Supposons désormais qu'il n'existe qu'un seul point $x_a \in \mathbb{R}^d$ prenant la valeur $U(x_a) = a$, et pour chaque $x \neq x_a$, on a

$$\langle x - x_a, U(x) - U(x_a) \rangle > 0. \quad (2.4)$$

Lorsque $d = 1$, l'inégalité (2.4) indique que les nombres $(x - x_a)$ et $(U(x) - U(x_a))$ ont les mêmes signes. Dans cette situation, la relation (2.4) se traduit par l'une des équivalences suivantes

$$\begin{aligned} x < x_a &\iff U(x) - U(x_a) < 0 \\ x > x_a &\iff U(x) - U(x_a) > 0. \end{aligned}$$

Un algorithme déterministe naturel de recherche du point $x_a \in \mathbb{R}^d$ est donc défini en posant

$$X_{n+1} - X_n = \gamma_n (U(x_a) - U(X_n)) = \gamma_n (a - U(X_n))$$

On choisit $X_0 \in \mathbb{R}^d$, et on utilise une suite de pas positifs $\gamma_n \downarrow 0$ afin de stopper l'évolution de la suite X_n . Les schémas de décroissances de γ_n pour lesquels $\lim_{n \rightarrow \infty} X_n = x_a$ doivent satisfaire les deux conditions suivantes

$$\sum_n \gamma_n = \infty \quad \text{et} \quad \sum_n \gamma_n^2 < \infty$$

L'introduction de l'aléatoire dans les algorithmes déterministes précédents ne se justifie que dans les deux situations suivantes. Dans les deux cas, le choix de schémas de décroissances de γ_n vérifiant les conditions précédentes entraîne que

$$\lim_{n \rightarrow \infty} X_n = x_a$$

2.10.3 Applications

Évaluation indirecte bruitée

L'évaluation de la fonction $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$ ne peut être faite directement sinon par des mesures approchées. Plus précisément, chaque tentative d'évaluation de $U(x)$ en un point $x \in \mathbb{R}^d$, donne une valeur approximative

$$x \longrightarrow \boxed{\text{capteur/mesure/évaluation}} \longrightarrow U(x) + \epsilon$$

où ϵ représente une perturbation aléatoire à valeurs dans \mathbb{R}^d . La nature statistique de ces perturbations de mesures dépendent de l'instrument de mesure utilisé.

Pour construire l'algorithme de Robbins-Monro, on remplace simplement à chaque étape n la valeur inconnue $U(X_n)$ par son observation bruitée

$$Y_{n+1} =_{\text{déf.}} U(X_n) + \epsilon_{n+1}.$$

La perturbation ϵ_n désigne ici une variable aléatoire à valeurs dans \mathbb{R}^d de loi donnée. Cela conduit à l'algorithme stochastique

$$\begin{aligned} X_{n+1} - X_n &= \gamma_n (U(x_a) - U(X_n)) - \gamma_n \epsilon_{n+1} \\ &= \gamma_n (U(x_a) - Y_{n+1}) = \gamma_n (a - Y_{n+1}) \quad (2.5) \end{aligned}$$

Formulation intégrale

La fonction $U : \mathbb{R}^d \rightarrow \mathbb{R}^d$ est de forme intégrale

$$U(x) = \int_E \mathcal{U}(x, y) \mu(dy)$$

où μ désigne une distribution de probabilité donnée sur un espace auxiliaire E , et

$$\mathcal{U} : \mathbb{R}^d \times E \longrightarrow \mathbb{R}^d$$

est une fonction bornée. De telles intégrales étant bien souvent difficilement calculables analytiquement, il est parfois plus aisé de simuler une variable aléatoire Z de loi μ sur E . On remarquera que U peut alors s'écrire

$$U(x) = \mathbb{E}(\mathcal{U}(x, Z))$$

Pour construire l'algorithme de Robbins-Monro correspondant à cette situation, on remplace à chaque étape n la valeur difficilement accessible $U(X_n)$ par sa "valeur simulée" $\mathcal{U}(X_n, Z_{n+1})$, où $(Z_n)_{n \geq 1}$ désigne une suite de variables aléatoires indépendantes de même loi que Z .

Ceci nous conduit à la chaîne de Markov suivante

$$X_{n+1} - X_n = \gamma_n (U(x_a) - \mathcal{U}(X_n, Z_{n+1})) = \gamma_n (a - \mathcal{U}(X_n, Z_{n+1}))$$

où $(Z_n)_{n \geq 1}$ désigne une suite de variables aléatoires indépendantes de même loi que Z .

2.11 Filtre de Kalman-Bucy

2.11.1 Introduction

Le filtre de Kalman-Bucy est sans nul doute l'algorithme stochastique le plus couramment utilisé en ingénierie, et plus particulièrement en traitement du signal. Le problème du filtrage consiste à estimer les lois conditionnelles des états d'un signal par rapport à une séquence d'observations partielles et bruitées. Plus précisément, les valeurs exactes du X_n ne sont pas directement observées. Néanmoins, Un capteur de mesure nous transmet à chaque étape n certaines informations Y_n sur la valeur de X_n :

$$\begin{array}{ccccccc} X_0 & \rightarrow & X_1 & \rightarrow & X_2 & \rightarrow & X_3 & \rightarrow & \dots & \text{signal} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow & & \dots & \\ Y_0 & & Y_1 & & Y_2 & & Y_3 & & \dots & \text{observation} \end{array}$$

Le but du filtrage est alors de calculer les lois conditionnelles

$$\text{Loi}(X_n | Y_0, \dots, Y_{n-1}) \quad \text{et} \quad \text{Loi}(X_n | Y_0, \dots, Y_n)$$

ou plus généralement la distribution conditionnelle de toute la trajectoire

$$\text{Loi}((X_0, \dots, X_n) | Y_0, \dots, Y_n)$$

La première distribution est appelée le **prédicteur optimal**, la seconde porte le nom de **filtre optimal**.

2.11.2 Description du modèle

Dans le cadre du filtrage linéaire gaussien, le couple signal/observation (X_n, Y_n) est représenté par une chaîne de Markov à valeurs dans \mathbb{R}^{p+q} , et définie récursivement par les équations suivantes :

$$\begin{cases} X_n = A_n X_{n-1} + a_n + B_n W_n, & n \geq 1 \\ Y_n = C_n X_n + c_n + D_n V_n, & n \geq 0 \end{cases} \quad (2.6)$$

Les séquences W_n , et V_n représentent des suites de variables aléatoires indépendantes à valeurs dans \mathbb{R}^{d_w} , et dans \mathbb{R}^{d_v} . On convient que ces séquences sont indépendantes de la variable initiale X_0 . Les lettres (A_n, B_n, C_n, D_n) , et (a_n, c_n) représentent respectivement des matrices, et des vecteurs déterministes de dimensions appropriées.

Les variables W_n , et V_n sont supposées gaussiennes, centrées, et de matrices de covariance respectives

$$R_n^w = \mathbb{E}(W_n W_n') \quad \text{et} \quad R_n^v = \mathbb{E}(V_n V_n')$$

On suppose enfin que la variable initiale X_0 est une variable gaussienne de moyenne et de matrice de covariance données :

$$\hat{X}_0^- = \mathbb{E}(X_0) \quad \text{and} \quad \hat{P}_0^- = \mathbb{E}((X_0 - \mathbb{E}(X_0)) (X_0 - \mathbb{E}(X_0))')$$

Ces hypothèses sur la linéarité des modèles, et la nature gaussienne de toutes les variables en jeu n'est pas anodine. Les mondes linéaires et gaussiens sont très stables : toute opération linéaire et tout type de conditionnement entre variables ont à nouveau une nature gaussienne. Il est donc impossible d'échapper à ces trous noirs gaussiens par des opérations linéaires ou par des conditionnements !

On notera par la suite $\mathcal{N}(m, R)$ la mesure gaussienne sur \mathbb{R}^d de moyenne $m \in \mathbb{R}^d$, et de matrice de covariance $R \in \mathbb{R}^{d \times d}$

$$\mathcal{N}(m, R)(dx) = \frac{1}{(2\pi)^{d/2} \sqrt{|R|}} \exp[-2^{-1}(x - m)R^{-1}(x - m)'] dx$$

D'après la discussion précédente le prédicteur et le filtre optimal sont des mesures gaussiennes

$$\text{Loi}(X_n | Y_0, \dots, Y_{n-1}) = \mathcal{N}(\widehat{X}_n^-, P_n^-) \quad \text{et} \quad \text{Loi}(X_n | Y_0, \dots, Y_n) = \mathcal{N}(\widehat{X}_n, P_n)$$

Les moyennes $(\widehat{X}_n^-, \widehat{X}_n)$ correspondent aux espérances conditionnelles

$$\widehat{X}_n^- = \mathbb{E}(X_n | Y_0, \dots, Y_{n-1}) \quad \text{et} \quad \widehat{X}_n = \mathbb{E}(X_n | Y_0, \dots, Y_n)$$

Les matrices (P_n^-, P_n) représentent les matrices de covariance d'erreurs entre les estimateurs conditionnels $(\widehat{X}_n^-, \widehat{X}_n)$ et le signal X_n observé, c'est à dire

$$P_n^- = \mathbb{E}([X_n - \widehat{X}_n^-][X_n - \widehat{X}_n^-]') \quad \text{et} \quad P_n = \mathbb{E}([X_n - \widehat{X}_n][X_n - \widehat{X}_n]')$$

2.11.3 Les équations du filtrage

Examinons brièvement le processus d'apprentissage du filtre.

À chaque instant n , une observation \mathbf{Y}_n est délivrée par un capteur de mesure. Le prédicteur optimal est alors mis à jour. Cette transition est parfois appelée étape de correction. Elle consiste à rajouter une observation dans la formule de conditionnement

$$\text{Loi}(X_n | Y_0, \dots, Y_{n-1}) \xrightarrow{\text{mise à jour}} \text{Loi}(X_n | Y_0, \dots, Y_{n-1}, \mathbf{Y}_n)$$

Une fois la mise à jour effectuée, le filtre optimal entre dans une étape de prédiction. Cette étape revient à prédire le mieux possible les valeurs de X_{n+1} , avant de recevoir et traiter la nouvelle information Y_{n+1}

$$\text{Loi}(X_n | Y_0, \dots, Y_n) \xrightarrow{\text{prédiction}} \text{Loi}(X_{n+1} | Y_0, \dots, Y_n)$$

En résumé, les lois conditionnelles gaussiennes évoluent selon deux étapes d'apprentissage naturelles

$$\mathcal{N}(\widehat{X}_n^-, P_n^-) \xrightarrow{\text{mise à jour}} \mathcal{N}(\widehat{X}_n, P_n) \xrightarrow{\text{prédiction}} \mathcal{N}(\widehat{X}_{n+1}^-, P_{n+1}^-)$$

L'idée géniale de Kalman et de Bucy a été de poursuivre d'un cran ce raisonnement naturel, et de montrer que les paramètres recherchés pouvaient à nouveau se calculer récursivement

$$(\hat{X}_n^-, P_n^-) \xrightarrow{\text{mise à jour}} (\hat{X}_n, P_n) \xrightarrow{\text{prédiction}} (\hat{X}_{n+1}^-, P_{n+1}^-)$$

L'étape de prédiction est relativement simple. En effet, il suffit de remarquer que l'on a

$$\begin{aligned} \hat{X}_{n+1}^- &= \mathbb{E}(A_{n+1}X_n + a_{n+1} + B_{n+1}W_{n+1} \mid (Y_0, \dots, Y_n)) \\ &= A_{n+1}\hat{X}_n + a_{n+1} \end{aligned}$$

et

$$\begin{aligned} P_{n+1}^- &= \mathbb{E}((A_{n+1}(X_n - \hat{X}_n) + B_{n+1}W_{n+1})(A_{n+1}(X_n - \hat{X}_n) + B_{n+1}W_{n+1})') \\ &= A_{n+1} P_n A_{n+1}' + B_{n+1} R_{n+1}^w B_{n+1}' \end{aligned}$$

En résumé, l'étape de prédiction est définie par les formules

$$\begin{aligned} \hat{X}_{n+1}^- &= A_{n+1}\hat{X}_n + a_{n+1} \\ P_{n+1}^- &= A_{n+1} P_n A_{n+1}' + B_{n+1} R_{n+1}^w B_{n+1}' \end{aligned}$$

L'étape de mise à jour est essentiellement basée sur le fait suivant, que nous admettons. La différence $(\hat{X}_n - \hat{X}_n^-)$ peut s'écrire sous la forme

$$\hat{X}_n - \hat{X}_n^- = \mathbf{G}_n (Y_n - \hat{Y}_n^-)$$

avec une matrice de gain \mathbf{G}_n convenable, et

$$\hat{Y}_n^- = \mathbb{E}(Y_n \mid (Y_0, \dots, Y_{n-1})) = C_n \hat{X}_n^- + c_n$$

Pour calculer la matrice \mathbf{G}_n , on commence par observer que

$$\mathbb{E}((X_n - \hat{X}_n)(Y_n - \hat{Y}_n^-)') = 0$$

et

$$(Y_n - \widehat{Y}_n^-) = C_n(X_n - \widehat{X}_n^-) + D_n V_n$$

Ceci permet de montrer que l'on a

$$\mathbb{E}((X_n - \widehat{X}_n^-)(Y_n - \widehat{Y}_n^-)') = \mathbf{G}_n \mathbb{E}((Y_n - \widehat{Y}_n^-)(Y_n - \widehat{Y}_n^-)')$$

et

$$\begin{aligned} \mathbb{E}((X_n - \widehat{X}_n^-)(Y_n - \widehat{Y}_n^-)') &= \mathbb{E}((X_n - \widehat{X}_n^-)(C_n[X_n - \widehat{X}_n^-] + V_n)') \\ &= P_n^- C_n' \end{aligned}$$

On en déduit alors aisément la formule

$$\mathbf{G}_n = P_n^- C_n' (C_n P_n^- + D_n R_n^v D_n')^{-1}$$

Enfin, en utilisant la décomposition

$$X_n - \widehat{X}_n = (X_n - \widehat{X}_n^-) + (\widehat{X}_n^- - \widehat{X}_n)$$

et par des arguments de symétrie, on obtient

$$\begin{aligned} P_n &= P_n^- - \mathbb{E}((\widehat{X}_n^- - \widehat{X}_n)(\widehat{X}_n^- - \widehat{X}_n)') \\ &= P_n^- - \mathbf{G}_n \mathbb{E}((Y_n - \widehat{Y}_n^-)(Y_n - \widehat{Y}_n^-)') \mathbf{G}_n' \\ &= P_n^- - \mathbf{G}_n C_n P_n^- \end{aligned}$$

En résumé, l'étape de correction est définie par les formules

$$\begin{aligned} \widehat{X}_n &= \widehat{X}_n^- + \mathbf{G}_n (Y_n - [C_n \widehat{X}_n^- + c_n]) \\ P_n &= P_n^- - \mathbf{G}_n C_n P_n^- \end{aligned}$$

avec la matrice de gain \mathbf{G}_n déterminée par l'équation suivante :

$$\mathbf{G}_n = P_n^- C_n' (C_n P_n^- + D_n R_n^v D_n')^{-1}$$

