Probabilités discrètes

Généralités

Espérance mathématique : $E(X) = \sum_{w \in \Omega} w.P(X = w)$

On notera parmi les propriétés de E :

- E est linéaire
- $E(E(X)) = E(X)$
- la formule de transfert, $E(f(X)) = \sum_{w \in \Omega} f(w).P(X = w)$
- $E(X) = 0$ pour une variable aléatoire centrée.

Variance : $Var(X) = E(X^2) - E(X)^2 = E\left((X - E(X))^2\right)$

On donne les propriétés suivantes :

- $Var(\lambda X) = \lambda^2 Var(X)$
- X est constante si et seulement si $Var(X) = 0$.

Ecart-type : $\sigma(X) = \sqrt{Var(X)}$

On a : $\sigma(\lambda X) = |\lambda| \sigma(X)$.

Couple (X,Y)

Loi conjointe et Loi marginale :

La loi conjointe de (X,Y) est $P((X,Y) \in A \times B) = P(X \in A, Y \in B)$.

Si $P(X \in A) = P(X \in A, Y \in B)$, alors on dit que X est une loi marginale.

Variables aléatoires indépendantes :

X et Y sont indépendantes si $\forall A, B \;\; P(X \in A, Y \in B) = P(X \in A) \times P(Y \in B)$.

Covariance : $C_{XY} = E(XY) - E(X).E(Y)$.

En particulier, on trouve $C_{XX} = Var(X)$.

Si X et Y sont indépendantes alors $E(XY) = E(X).E(Y)$, d'où $C_{XY} = 0$.

La matrice des covariances s'écrit : $C = \begin{pmatrix} C_{XX} & C_{XY} \\ C_{YX} & C_{YY} \end{pmatrix}$. Elle est symétrique et positive.

Coefficient de corrélation : $\rho_{XY} = \dfrac{C_{XY}}{\sigma_X . \sigma_Y}$

On a les propriétés suivantes :

- $\rho_{\alpha X, \beta Y} = \rho_{X,Y}$
- $|\rho_{X,Y}| \leq 1$

Loi de S=X+Y :

$S = w$ peut s'écrire $\bigcup_{\alpha} (X = \alpha \text{ et } Y = w - \alpha)$. Ainsi $P(S = w) = \sum_{\alpha} P(X = \alpha, Y = w - \alpha)$.

De plus, si X et Y sont indépendantes alors $P(S = w) = \sum_{\alpha} P(X = \alpha) \times P\,(Y = w - \alpha)$, le produit de convolution des lois X et Y.

<u>Variance d'une somme de variables aléatoires</u> :
$Var(X + Y) = Var(X) + Var(Y) + 2.C_{XY}$.
Si X et Y sont indépendantes, alors on a la formule : $Var(X + Y) = Var(X) + Var(Y)$.

Fonction génératrice

Considérons une variable aléatoire X entière.
Sa fonction génératrice est définit par : $G_X(z) = \sum_n p(X = n).z^n$ de rayon $R \geq 1$.

Cette fonction caractérise la loi X, et $P(X = n) = \frac{1}{n!} G_X{}^{(n)}(0)$.

On calcule :

- $E(X) = G'_X(1)$
- $Var(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

Si X et Y sont indépendantes, alors $G_{X+Y} = G_X . G_Y$.

Loi binomiale et Loi de Poisson

<u>Loi binomiale de paramètre d et d'ordre n</u> :
Considérons une expérience pendant laquelle un événement A peut arriver avec la probabilité d. On répète cette expérience n fois de façon indépendante.
X : nombre de fois où A s'est produit
Y : indicatrice de l'événement A (Y=1 si A s'est produit, 0 sinon).
Y_1, Y_2, …, Y_n : échantillon de modèle Y.
On a $X = \sum Y_i$.

On montre facilement que $P(X = k) = \begin{cases} 0, \text{ si } k > n \\ C_n^k d^k (1-d)^{n-k}, \text{ sinon} \end{cases}$.

On établit ensuite que :

- $Var(X) = nd(1-d)$
- $E(X) = nd$
- en utilisant $G_X(z) = (1 + dz - d)^n$.

<u>Loi de Poisson de paramètre a</u> : $P(X = k) = \frac{a^k}{k!} e^{-a}$

Si l'on pose $a = nd$, alors $C_n^k d^k (1-d)^k \underset{n \to \infty}{\longrightarrow} \frac{a^k}{k!} e^{-a}$.

En pratique, on fait l'approximation de la loi binomiale d'ordre n et de paramètre d par la loi de Poisson de paramètre $a = nd$, pour $n \geq 30$.

On établit que :

- $Var(X) = a$
- $E(X) = a$
- en utilisant $G_X(z) = e^{a(z-1)}$.

Théorème : stabilité des lois de Poisson pour la somme
Supposons que X suit la loi de Poisson de paramètre a et que Y suit la loi de Poisson de paramètre b, alors X+Y suit la loi de Poisson de paramètre a+b.

Probabilités générales

Note : Seul le cas diffus est développé dans ce chapitre.

Généralités

On définit dans le cas diffus la probabilité pour la loi X : $P(X \in A) = \int_A f(x)dx$ où f représente la densité de la loi X. La densité vérifie la propriété fondamentale des probabilités : $P(X \in \Omega) = \int_\Omega f(x)dx = 1$.

L'espérance se définit, dans le cas diffus, comme suit : $E(X) = \int x\,dp(x) = \int x.f(x)\,dx$.
Les propriétés restent les mêmes que pour les probabilités discrètes.
La formule de transfert s'écrit: $E(h(X)) = \int h(x)f(x)dx$.

La variance, la covariance et l'écart-type, … restent inchangés.

Loi de S = X+ Y :
Soit h la densité de la loi S, f la densité de la loi X et g la densité de la loi Y.
Pour X et Y indépendants, la densité de S est le produit de convolution des densité de X et Y, soit :

$$h(s) = (f * g)(s) = \int_{-\infty}^{s} f(s-y)g(y)dy.$$

De façon générale, la densité d'un somme de n variables indépendantes est le produit de convolution des densités.

Loi de -X : $F_{-X}(x) = 1 - F_X(-x)$ et $f_{-X}(x) = f_X(-x)$

Loi conjointe d'un couple (X,Y)

Considérons la densité $h(x,y)$, on a $P((X,Y) \in A \times B) = \iint_{A\times B} h(x,y)dxdy$. On pourra noter que la densité de la loi en X est $\int_B h(x,y)dy$, de même pour la densité de la loi en Y.
Pour des variables aléatoires indépendantes, on a $P((X,Y) \in A \times B) = P(X \in A) \times P(Y \in B)$.

Théorème : X et Y sont indépendants si $h(x,y) = f(x)\,g(y)$.

Fonction de répartition

On définit la fonction répartition comme suit : $F(x) = \mathrm{Prob}(]-\infty, \mathrm{x}] \cap \Omega) = \mathrm{Prob}(X \leq x)$.
On a :

- $\lim_{x\to\infty} F(x) = 1$
- $\lim_{x\to-\infty} F(x) = 0$

Dans le cas diffus, on a $F(x) = \int_a^x f$ et $F'(x) = f(x)$, où f représente la densité et $a = Inf(\Omega)$.
Ainsi on peut calculer la probabilité, $P(x \in A) = \int_A f$.

Fonction caractéristique

Soit X une variable aléatoire de densité f et u un réel quelconque.
La fonction caractéristique de X est définit par $\Phi_X(u) = E\left(e^{iuX}\right) = \int e^{iux} f(x)dx$.

On a les propriétés suivantes :

- Transformée de Fourier : $\Phi_X(u) = \hat{f}\left(\frac{-u}{2\pi}\right)$
- $\Phi'_X(0) = i.E(X)$
- $Var(X) = -\Phi''_X(0) + (\Phi'_X(0))^2$
- si X et Y sont indépendants, $\Phi_{X+Y} = \Phi_X \times \Phi_Y$.

Loi exponentielle de paramètre a et de début t_0

<u>Problème de durée de vie</u> :
Par exemple un appareil est mis en marche à l'instant t_0. La variable aléatoire X désigne l'instant où l'appareil tombe en panne. On fait l'hypothèse de non-vieillissement de la machine, c'est à dire $a(t) = a$.
Si F est la fonction de répartition de X, on démontre que : $1 - F(t) = e^{-a(t-t_0)}$, $t > t_0$.
Ensuite, on établit les résultats suivants :

- $E(X) = \frac{1}{a} + t_0$
- $Var(X) = \frac{1}{a^2}$
- densité : $f(t) = a.e^{-a(t-t_0)}$

Loi normale (Gauss)

<u>Loi normale réduite</u> $N(0,1)$:

La loi est de densité : $f_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$. On rappelle que $\int e^{-\left(\frac{1}{2}\right).x^2} dx = \sqrt{2\pi}$.

Supposons que X suit $N(0,1)$.
Alors il vient :

- $E(X) = 0$
- $Var(X)$=1 donc $\sigma_X = 1$
- la fonction caractéristique $\Phi_X(u) = \exp\left(-\frac{u^2}{2}\right)$

<u>Loi normale générale</u> $N(m,\sigma)$:
On suppose que X suit $N(0,1)$ et on pose $Y = \sigma X + m$, alors Y suit $N(m,\sigma)$.
De même il vient :

- $E(Y) = m$
- $Var(Y) = \sigma^2$ donc $\sigma_X = \sigma$
- la densité s'écrit $f_{m,\sigma}(y) = f_{0,1}\left(\frac{y-m}{\sigma}\right) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2}\left(\frac{y-m}{\sigma}\right)^2\right)$

Théorème : Stabilité des lois normales

Si Y_1 suit $N(m_1, \sigma_1)$, si Y_2 suit $N(m_2, \sigma_2)$ et si Y_1, Y_2 sont indépendantes, alors $(Y_1 + Y_2)$ suit $N(m, \sigma)$ avec $\begin{cases} m = m_1 + m_2 \\ \sigma^2 = {\sigma_1}^2 + {\sigma_2}^2 \end{cases}$.

Moyenne de lois normales

Y suit $N(m, \sigma)$, et $Y_1, Y_2, \ldots, Y_n$ un échantillon.

La moyenne $\overline{Y} = \frac{1}{n}\sum_i Y_i$ suit $N\left(m, \frac{\sigma}{\sqrt{n}}\right)$, car $Var(\overline{Y}) = \frac{1}{n}Var(Y)$.

Couple Gaussien : (X_1, X_2)

Hypothèses : X_1 suit $N(m_1, \sigma_1)$, X_2 suit $N(m_2, \sigma_2)$, $|r| < 1$.

La densité de la loi du couple s'exprime :

$$h(x_1, x_2) = \left(\frac{1}{2\pi.\sigma_1\sigma_2.\sqrt{1-r^2}}\right) \times \exp\left\{\frac{-1}{2.\sqrt{1-r^2}}\left[\left(\frac{x_1 - m_1}{\sigma_1}\right)^2 + \left(\frac{x_2 - m_2}{\sigma_2}\right)^2 - 2r\left(\frac{x_1 - m_1}{\sigma_1}\right)\left(\frac{x_2 - m_2}{\sigma_2}\right)\right]\right\}$$

avec r le coefficient de corrélation, $r = \frac{C_{X_1X_2}}{\sigma_1\sigma_2}$.

Dans le cas d'un couple Gaussien, $C_{X_1X_2} = 0$ si et seulement si X_1 et X_2 sont indépendantes. En général, la réciproque est fausse.

Loi conditionnelle :

La loi conditionnelle à une valeur y_0 : $P(X \in A / Y = y_0) = \frac{\int_A h(x, y_0)dx}{\int_R h(x, y_0)dx}$.

C'est une loi normale $N(m, \sigma)$, avec $m = E(X) + (y_0 - E(Y))\frac{\sigma_X}{\sigma_Y}r$ et $\sigma = \sqrt{1 - r^2}.\sigma_X$.

Loi des grands nombres – Théorème central limite

Convergence en probabilité (CVP)

Soient $(Y_n)_n$ et un réel a.

$$\left(Y_n \underset{\text{CVP}}{\longrightarrow} a\right) \Leftrightarrow \left(\forall \varepsilon > 0,\ \lim_{n\to\infty} \mathrm{Prob}(|\mathrm{Y_n} - a| > \varepsilon) = 0\right).$$

Inégalité de Markoff

Hypothèses : $a > 0$, X à valeurs réelles positives tel que $m = E(X) < \infty$.

Alors, $\mathrm{Prob}(\mathrm{X} \geq a.m) \leq \frac{1}{a}$.

Inégalité de Tchebycheff

Hypothèses : $b > 0$, Y tel que $E(Y^2) < \infty$.

Alors, $\mathrm{Prob}(|\mathrm{Y - E(Y)}| > b\sigma) \leq \frac{1}{b^2}$ ou encore $\mathrm{Prob}(|\mathrm{Y - E(Y)}| > b) \leq \frac{\sigma^2}{b^2}$.

Loi des grands nombres

Hypothèses : X tel que $E(X^2)$ existe, et $X_1, X_2, \ldots, X_n$ un échantillon.

Alors $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ converge en probabilité vers l'espérance $E(X)$.

Convergence en loi

$(X_n)_n$ converge en loi vers X si et seulement si la fonction de répartition F_{X_n} converge simplement vers F_X, c'est à dire $P(X_n \le x) \xrightarrow[n\to\infty]{} P(X \le x)$.

Dans le cas d'une variable aléatoire entière, cela signifie aussi $P(X_n = l) \xrightarrow[n\to\infty]{} P(X = l)$.

Théorème Paul Levy

$(X_n)_n$ converge en loi vers X si et seulement si la fonction caractéristique Φ_{X_n} converge vers Φ_X.

Théorème central limite

Soit X tel que $E(X^2)$ existe, $\int |x|^3 f(x)dx < \infty$ et $X_1, X_2, \ldots, X_n$ un échantillon.

On suppose $E(X) = 0$, $\sigma(X) = \sigma$.

Soit $Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$.

Alors $(Y_n)_n$ converge vers la loi normale $N(0,\sigma)$.

Utilisation pratique du théorème central limite

Soit une variable aléatoire X et $X_1, X_2, \ldots, X_n$ un échantillon.

- Pour $n \ge 30$, $Y = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ suit (approximativement) une loi normale $N(\sqrt{n}.E(X), \sigma_X)$.
- Pour $n \ge 30$, $Y = \frac{1}{n}\sum_{i=1}^{n} X_i$ suit (approximativement) une loi normale $N\left(E(X), \frac{\sigma_X}{\sqrt{n}.}\right)$.

En conclusion, il est normal de trouver une loi normale !

Loi du chi-deux : $\aleph^2$

Soit $X_1, X_2, \ldots, X_n$ un échantillon de la loi $N(0,1)$.

$U_n = \sum_{i=1}^{n} X_i^{\ 2}$.

On dit que U_n suit la loi du $\aleph^2$ à n degrés de liberté et pour $n \ge 30$, on considère que $\frac{1}{\sqrt{n}}(U_n - n)$ suit (approximativement) la loi $N(0, \sqrt{2})$.

Processus Markoviens discrets en temps et en valeurs

Définitions

Processus : $X(t)$ variable aléatoire qui dépend du temps.

Un processus discret en valeurs est tel que $\forall t \ X(t)$ est une variable aléatoire discrète.

Discrets en temps revient à considérer une suite d'instants $t_0, t_1, \ldots, t_n$ au lieu de prendre le temps en continue. Les $X(t_i)$ sont notés $X(i)$.

Définitions Markovien

Un processus discret en temps et en valeurs est dit Markovien si et seulement si : $\forall m, \ \forall n \geq m+1, \ \forall e_i$,

$$P\left(X_{m+1}=e_{m+1},\cdots,X_n=e_n \middle| X_m=e_m X_{m+1}=e_{m+1},\cdots,X_n=e_n\right)=P\left(X_{m+1}=e_{m+1},\cdots,X_n=e_n \middle| X_m=e_m, X_{m-1}=e_{m-1},\cdots, X_0=e_0\right)$$

En d'autres termes, l'état futur ne dépend (directement) que du présent et pas du passé.

Propriété : Loi des n-uplets

Dans le cas d'un processus Markovien, on a :

$$P\left(X_0=e_0,\cdots,X_n=e_n\right)=\prod_{k=1}^{n}\underbrace{P\left(X_k=e_k \middle| X_{k-1}=e_{k-1}\right)}_{transition}\times\underbrace{P\left(X_0=e_0\right)}_{initialisation}.$$

Les chaînes de Markov

Les chaînes de Markov représentent un cas particuliers important des processus Markoviens.

Définition : Chaîne de Markov

Ce sont les processus Markoviens vérifiant la propriété suivante :

$$\forall n,k \ P\left(X_n=e' \middle| X_0=e\right)=P\left(X_{n+k}=e' \middle| X_k=e\right)$$

Il y a invariance des transitions, ou autrement les règles ne changent pas au cours du jeu. Il y a non-vieillissement du processus.

On suppose de plus que les valeurs possibles de X sont en nombre finis : $e_1, \ldots, e_n$ encore notées $1, \ldots, n$.

Notons $G_{i,j}=P\left(X_1=j \middle| X_0=i\right)=\underbrace{P\left(X_{k+1}=j \middle| X_k=i\right)}_{transition}$, pour $i, j=1,\ldots,n$.

On note G la matrice des transitions, $G=\left(G_{i,j}\right)$.

Théorème :

$$P\left(X_k=j \middle| X_0=i\right)=\left(G^k\right)_{i,j}.$$

Ainsi la matrice de transition pour la $k^{ème}$ étape est G^k.

Exemple fondamental : La marche aléatoire

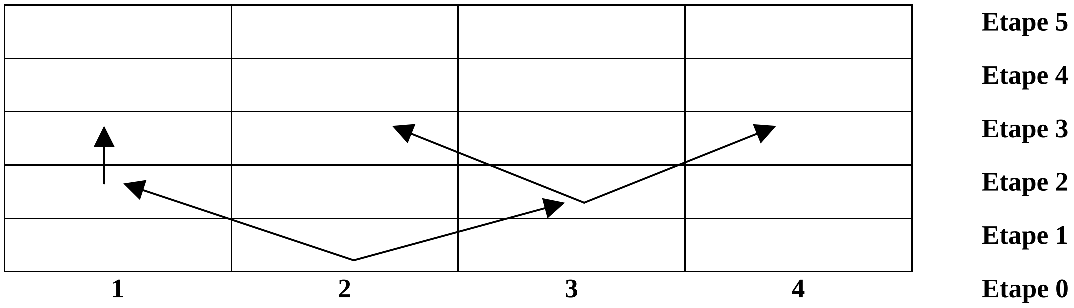
On considère un promeneur se déplaçant étape par étape soit à droite soit à gauche de façon équiprobable. Si il se trouve sur un bord, il y restent (bord absorbant). Les règles n'évoluent pas au cours du temps. Il s'agit d'une chaîne de Markov.

La matrice G de transition est : $\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 0 & \boxed{1/2} & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ $\begin{matrix} \rightarrow \text{bord absorbant à droite} \\ \\ \\ \rightarrow \text{bord absorbant à gauche} \end{matrix}$

On remarque que cette matrice est *stochastique* car la somme des coefficients de chaque ligne fait 1.

Le coefficient encadré s'écrit $P\left(X_1=3 \middle| X_0=2\right)=1/2$ et représente la probabilité d'arriver en position 2, en partant de la position 3.

On peut représenter la marche aléatoire avec le tableau ci-dessous.



On cherche à connaître le comportement asymptotique ou à long terme du promeneur.
Pour cela, on calcule G^k et on prend sa limite en l'infini :

$$G^{\infty} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 1/3 & 0 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \rightarrow \text{proba d'arriver en } j \text{, en partant de } i = 2$$

Ainsi, on peut résumer le contenu de cette matrice en disant que :
- le promeneur parti de 1 reste en 1 (bord absorbant)
- le promeneur parti de 2 a une probabilité 2/3 d'échouer à gauche et 1/3 d'échouer à droite
- le promeneur parti de 3 a une probabilité 1/3 d'échouer à gauche et 2/3 d'échouer à droite
- le promeneur parti de 1 reste en 1 (bord absorbant).