Fiche de Probabilités

Contents

Part I
Espace probabilisé

1 Définition de la probabilité

Définition d'une tribu Soit Ω un ensemble, $\mathcal{P}(\Omega)$ l'esemble de ses sous-ensembles. $\mathcal{A} \in \mathcal{P}(\Omega)$ est une tribu si:

1. $\Omega \in \mathcal{A}$
2. si $A \in \mathcal{A}$ alors $\bar{A} \in \mathcal{A}$
3. si la suite $A_n \in \mathcal{A}^{\mathbb{N}}$ alors $\bigcup\limits_{n=1}^{+\infty} A_n \in \mathcal{A}$

On appel $\mathcal{B}(\mathbb{R}) = \{]-\infty, a[, a \in \mathbb{R}\}$ la tribu borélienne.

Définition d'une probabilité P est une probabilité sur $(\Omega, \mathcal{A})$ si:

1. $\forall A \in \mathcal{A}, P(A) \in [0,1]$
2. $P(\Omega) = 1$
3. si la suite A_n vérifie $\forall i \neq j, A_i \cap A_j = \emptyset$ Alors $P\left(\bigcup\limits_{n=1}^{+\infty} A_n\right) = \sum\limits_{n=1}^{+\infty} P(A_n)$

Définition d'un espace probabilisé $(\Omega, \mathcal{A}, P)$ est un espace probabilisé

Propriétés élémentaires des probabilités Soient $A, B \in \mathcal{A}$ et $A_n \in \mathcal{A}^{\mathbb{N}}$

1. $P(\bar{A}) = 1 - P(A)$
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
3. $A \subset B \Rightarrow P(A) \leq P(B)$
4. $P\left(\bigcup\limits_{n=1}^{+\infty} A_n\right) \leq \sum\limits_{n=1}^{+\infty} P(A_n)$
5. $P\left(\lim\limits_{n\to+\infty} A_n\right) = \lim\limits_{n\to+\infty} P(A_n)$
6. A et B disjoints $\Rightarrow P(A \cap B) = 0$

2 Indépendance

Définition d'événements indépendants Soit $(A_1, A_2, ..., A_n) \in \mathcal{A}^n$ Ces évènements sont **indépendants** si :

$$\forall k \in \mathbb{N} \forall i_1, ..., i_k \in \{1, ..., n\} \text{ distincts deux à deux, } P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j})$$

Propriété Soit C_n famille de $\mathcal{A}$, formant une partition Ω Alors $\forall A \in \mathcal{A}, P(A) = \sum\limits_{n=1}^{+\infty} P(A \cap C_n)$

3 Probabilité conditionelle

Définition d'une probabilité conditionnelle On définit la **probabilité conditionelle** par rapport à l'évènement B :

$$\forall A \in \mathcal{A}, P(A|B) = \frac{P(A \cap B)}{P(B)} = P_B(A)$$

Propriétés

Formule des probabilités totales Soit C_n famille de $\mathcal{A}$ où aucun des termes n'a une probabilité nulle,

$$\forall A \in \mathcal{A}, P(A) = \sum_{n=1}^{+\infty} P(A|C_n)P(C_n)$$

Formule de Bayes

$$\text{Si } P(A) \neq 0, P(C_i|A) = \frac{P(A|C_i)P(C_i)}{\displaystyle\sum_{n=1}^{+\infty} P(A|C_n)P(C_n)}$$

Part II
Variables

Définition d'une variable aléatoire X **variable aléatoire** est une application mesurable d'un ensemble probabilisé $(\Omega, \mathcal{A}, P)$ dans $(\Omega', \mathcal{A}')$
P_X la **loi de probabilité** de X, est définie pour tout borélien B par :

$$P_X(B) = P(\{\omega | X(\omega) \in B\}) = P(X^{-1}(B))$$

Une variable aléatoire peut être réelle, complexe, vectorielle, multidimensionnelle ou multivariée, ou encore une suite de variables aléatoires

1 Différents types de variables

Variable continue

Définition d'une fonction de répartition Une **fonction de répartition** d'une variable aléatoire X est la fonction :

$$F_X : \left\{ \begin{array}{ccc} \mathbb{R} & \rightarrow & [0,1] \\ x & \mapsto & P(X < x) \end{array} \right.$$

Propriétés élémentaires :

1. F_X est croissante au sens large
2. F_X est continue à gauche
3. $\lim\limits_{x \to -\infty} F_X(x) = 0$
4. $\lim\limits_{x \to +\infty} F_X(x) = 1$
5. $P(a \leq X < b) = F_X(b) - F_X(a)$

Ou on trouve aussi : $F_X : x \mapsto P(X \leq x)$ ce qui modifie la dernière propriété en $P(a < X \leq b) = F_X(b) - F_X(a)$

Définition de la densité de probabilité Si F_X est continue et $\mathcal{C}^1$ par morceau sur $\mathbb{R}$, la fonction $f_X = F'_X$ est appelée **densité de probabilité** de X.

Propriétés

1. f_X est positive ou nulle
2. $\lim_{x\to+\infty} F(x) = \int_{\mathbb{R}} f_X(t)dt = 1$, plus généralement $\int_\Omega f_X(t)dt = 1$
3. $\forall A \in \mathcal{A}', P_X(A) = \int_A f_X(t)dt$

Discrète

Loi $P_X : P(X = x_i) = p_i$

Fonction de répartition $F_X(x) = \sum p_i \delta(x - x_i)$

mixte

On décompose en deux parties, une continue et une discrète

2 Variable aléatoire vectorielle

Définitions :

Loi jointe: loi de la variable $X = (X_1, ..., X_n)$
Loi marginale: loi des composantes X_i
La connaissance de la jointe de X suffit pour caractériser X et connaître les lois marginales mais la réciproque est fausse !

Cas discret

Contexte :
Soient X^1 et X^2 v.a discrètes prenant respectivement les valeurs $\{x_1^1, ..., x_n^1\}$ et $\{x_1^2, ..., x_m^2\}$ avec des probabilités p_i^1 et p_j^2
La v.a $X = (X^1, X^2)$ prend ses valeurs dans $\{(x_i^1, x_j^2), 1 \leq i \leq n, 1 \leq j \leq m\}$ avec des probabilités $p_{i,j}$:

$$\sum_{1\leq k\leq m} p_{i,k} = p_i^1 \text{ et } \sum_{1\leq k\leq n} p_{k,j} = p_j^2$$

Fonctions de répartition :
La v.a $X = (X_1, ..., X_n)$ est entrièrement déterminée par sa f.d.r

$$F_X(x) = F_{X_1,...,X_n}(x_1, ..., x_n) = P(]-\infty, x1[\times ... \times]-\infty, x_n[) = P(\{X_1 < x_1\} \cap ... \cap \{X_n < x_n\})$$

Toute v.a marginale X_1 est entièrement caractérisée par la f.d.r de X.

$$\forall k \neq i, \lim_{x_k\to+\infty} F_X(x_1, ..., x_n) = F_{X_i}(x_i)$$

Mêmes propriétés que le cas scalaire.

Cas continu

Contexte :
La v.a $X = (X_1, X_2)$ est caractérisée par la densité de probabilité f_X

Fonction de répartition
La fonction de répartition F_X de X qui est alors continue et partiellement dérivable presque partout vérifie : $f_X(X1, X2) = \frac{\partial^2 F_X(x_1,x_2)}{\partial x_1 \partial x_2}$
On a donc : $F_X(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f_X(t,u)dtdu$
Les f.d.r marginales respectives de X_1 et X_2 vérifient :

$$F_{X_2}(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{+\infty} f_X(t,x_2)dx_2dt, \text{ et de même pour } F_{X_2}$$

Densité marginale

$$f_{x_1} = \int_{-\infty}^{+\infty} f_X(x_1,x_2)dx_2$$

Indépendance des variables
Les v.a X et Y définies de $(\Omega, \mathcal{A}, P)$ dans $(\Omega', \mathcal{A}')$ sont indépendantes si

$$\forall A, B \in \mathcal{A}'^2, P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A)P(Y \in B)$$

Changement de variable
Soit $(X,Y) \to (U,V)$
$f_{U,V}(u,v) = f_{X,Y}(x(u,v), y(u,v))J(u,v)$
avec J jacobien valeur absolue du déterminant : $\begin{vmatrix} \frac{\partial x(u,v)}{\partial u} & \frac{\partial x(u,v)}{\partial v} \\ \frac{\partial y(u,v)}{\partial u} & \frac{\partial y(u,v)}{\partial v} \end{vmatrix}$
Conseil pour les changements de variable non bijectifs : faire un dessin !

Part III
Espérance mathématique

1 Définition

Valeur numérique mesurant le degré d'équité d'un jeu de hasard, c'est-à-dire moyenne.
Attention : elle n'existe pas toujours.

2 Calcul

v.a discrète $E[X] = \sum_{\omega \in \Omega} X(\omega)P(\omega) = \sum_{k \in (1,\ldots n)} a_n P(X = a_n)$
v.a continue $E[X] = \int_{\omega \in \Omega} X(\omega)dP(\omega) = \int_{\mathbb{R}} x dP_X(x)$ et si X a une densité de probabilité, $E[X] = \int_{-\infty}^{+\infty} x f_X(x)dx$
Attention : vérifier que X est intégrable.

3 Propriétés

- $E[X+Y] = E[X] + E[Y]$
- Linéarité : $E[aX+b] = aE[X] + b$, $a, b \in \mathbb{R}$

4 Théorème

DE TRANSFERT
Pour toute fonction réelle g, positive ou P-mesurable : $E[g(X)] = \int_{\omega \in \Omega} g(X(\omega))dP(\omega) = \int_{\mathbb{R}} g(x)dP_X(x)$

5 Maximum, Médiane, Mode, Minimum

mediane Valeur pour laquelle $P(X < med) = P(X \geq med) = \frac{1}{2}$)
De même, on peut définir : quartiles, déciles, centiles,...

mode v.a. discrète : valeur de X dont la probabilité est la plus grande
v.a. continue : valeur pour laquelle la densité de probabilité est maximum

Part IV
Variance

1 Définition

C'est la moyenne des écarts à la moyenne. C'est un caractère de dispersion.

2 Calcul

$$Var(X) \begin{cases} = E[(X-E[X])^2] & \textit{(Définition)} \\ = \int_{\mathbb{R}} (x-E[X])^2 f_X(x)dx & \textit{(Théorème de transfert)} \\ = E[X^2] - E[X]^2 & \textit{(Pour les calculs. ATTENTION à vérifier que } X^2 \textit{ est intégrable)} \end{cases}$$

3 Covariance

Définition : $Cov(X,Y) = E[(X-E[X])(Y-E[Y])] = E[XY] - E[X]E[Y]$

4 Propriétés

- $Var(aX+b) = a^2 Var(X)$
- $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$
- Si X et Y sont indépendants alors $E[XY] = E[X]E[Y]$ et $Cov(X,Y) = 0$ donc $Var(X+Y) = Var(X) + Var(Y)$

5 Ecart type

$\sigma_X = \sqrt{Var(X)}$

6 Coefficient de variation mesure

Il indique la dispersion relative. $c_v(X) = \sigma/\mu$ sans unité.

7 Théorème

DE BIENAYMÉ-TCHEBYCHEV

Soit X v.a. de moyenne μ et de variance finie σ^2 $\forall a \in R^+$, c'est-à-dire,

$$\begin{aligned} &P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2} \\ &P(|X-\mu| \geq a\sigma) \leq \frac{1}{a^2} \\ &P(|X-\mu| < a\sigma) \geq 1 - \frac{1}{a^2} \\ &P(-a\sigma < X-\mu < a\sigma) \geq 1 - \frac{1}{a^2} \end{aligned}$$

Part V
Généralisation

1 Dans $\mathbb{R}$

Soit X de densité $f_X(x)$. Grâce au théorème de transfert, on calcule $E[X_i] = \int_{\mathbb{R}^n} \int \dots \int x_i f_X(x) dx_1 \dots dx_n$ et la valeur moyenne $M = E[X] = E[X_1] \dots E[X_n]$. De même, on a $Var[X_i] = \int_{\mathbb{R}^n} \int \dots \int (x_i - E[X_i])^2 f_X(x) dx_1 \dots dx_n$ et $Cov(X_i, X_j) = \int_{\mathbb{R}^n} \int \dots \int (x_i - E[X_i])(x_j - E[X_j]) f_X(x) dx_1 \dots dx_n$

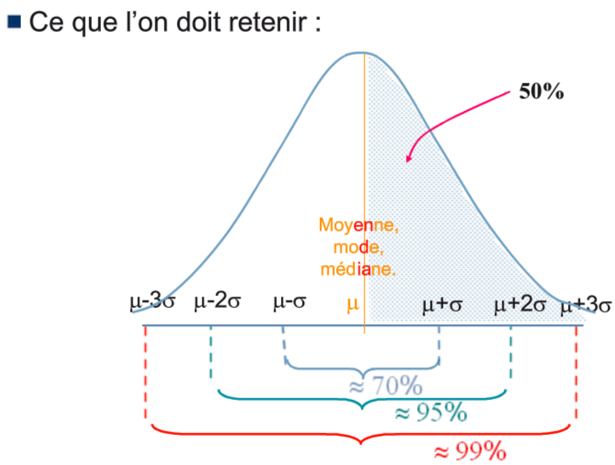
2 Dans $\mathbb{R}^n$

On définit $M = (\mu_1, ..., \mu_n)$
On note $Var(X) = E[(X - E[X])^2]$ si on considère $(X - E[X])^2 = (X - E[X])(X - E[X])^T$.On note aussi la matrice de variance-covariance $\Sigma = E[(X - E[X])^2] = [\sigma_{ij}]_{1 \leq i,j \leq n}$ est symétrique, semi-définie positive donc inversible.

Part VI
Loi normale

1 Caractéristiques

Graph



Densité $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[\frac{-(x-\mu)^2}{2\sigma^2}$

Espérance $E[X] = \mu (= méd = mode)$

Variance $Var(X) = \sigma^2$

Variables gaussiennes Si X suit une loi $\mathcal{N}(\mu, \sigma)$ et $Y = aX + b$

- Y normale aussi
- $E[Y] = a\mu + b$
- $Var(Y) = a^2\sigma^2$

On peut toujours se ramener à Z de loi centrée ($E[Z] = 0$) et réduite ($Var(Z) = 1$) par le changement de variable $X = \frac{Z - \sigma_x}{\mu_x}$ dont on connait les valeurs recueillies dans les tables, de densité $f_Z(z) = \dfrac{1}{\sqrt{2\pi}} exp(-\dfrac{z^2}{2})$

2 Généralisation

Densité normale multivariée $\mathcal{N}(M, \Sigma)$ $f_X(x_1, ..., x_n) = \dfrac{1}{(2\pi)^{\frac{n}{2}} (\det(\Sigma))^{\frac{1}{2}}} exp(-\dfrac{1}{2}(x - M)^T \Sigma^{-1} (x - M))$

Vecteur gaussien $X \Leftrightarrow$toute combinaison de ses X_i est une v.a. normale. Si X gaussien, A matrice à n colonnes et b vecteur de $\mathbb{R}^n$, alors $AX + b$ est un vecteur gaussien de loi $\mathcal{N}(AM + b, A\Sigma \cdot A^T)$.
Si X gaussien

- Toutes les v.a. marginales X_i sont gaussiennes.
- Si X_i et X_j non corréleés, alors elles sont indépendantes (équivalence)
- et si les X_i et X_j son indépendantes 2 à 2 alors les $X_1..X_n$ sont globalment indépendants.

Si tous les X_i sont des v.a. gaussiennes indépendantes, alors X est un vecteur gaussien.
ATTENTION, par exemple $f_{X,Y}(0, y)$ n'est pas forcément gaussien.

Part VII
Lois conditionnelles et espérance conditionnelle

1 Définition

Cas discret Soient $X : \Omega \to E$ et $Y : \Omega \to F$ v.a. discrètes, alors la loi conditionnelle de Y sachant $X = x$ est : $\forall y \in F, P_{Y/X=x}(y) = \frac{P_{Y,X}(y,x)}{P_X(x)} = P(Y = y/X = x)$.
X et Y indépendants $\Leftrightarrow \forall x, y, P(Y = y/X_x) = P(Y = y)$
L'espérence conditionnelle de Y sachant $X = x$ est : $E[Y/X = x] = \sum_{y \in F} y P_{Y/X=x}(y)$ (c'est l'espérance par rapport à la loi conditionnelle).

Cas continu Soit (X, Y) un couple de densité $f_{X,Y}(x, y)$ sur $\mathbb{R}^2$ alors la densité conditionnelle de Y sachant $X = x$ est : $\forall y \in \mathbb{R}, f_{Y/X=x}(y) = \frac{f_{Y,X}(y,x)}{f_X(x)}$.
X et Y indépendants $\Leftrightarrow \forall x, y, f_{Y/X=x}(y) = f_Y(y)$
L'espérence conditionnelle de Y sachant $X = x$ est : $E[Y/X = x] = \int_{\mathbb{R}} y f_{Y/X=x}(y) dy$ (c'est l'espérance par rapport à la densité conditionnelle).

2 Propriété

- L'espérance conditionnelle est linéaire et possède les propriétés classiques de l'espérance.
- Si X et Y indépendants, $E[Y/X = x] = E[Y]$
- Si h fonction mesurable et $h(X)$ intégrable,$E[Yh(X)/X] = h(X)E[Y/X]$, et aussi $E[h(X)/X] = h(X)$

Remarque : $E[E[Y/X]] = E[Y]$ car $E[Y/X = x)$ est fonction de x

Part VIII
Les différentes convergences

1 Convergence en moyenne d'ordre alpha

Définition $X_n \xrightarrow{L^\alpha} X$
X si $\lim_{n\to+\infty} E[|X_n - X|^\alpha] = 0$

Loi faible des grands nombres pour cette convergence Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de variance finie, alors $\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X]$ et $\lim_{n\to+\infty} \frac{1}{n} Var(X) = 0$

2 Convergence en probabilité

Définition $X_n \xrightarrow{P} X$
$(X_n)_{n\in\mathbb{N}}$ converge en probabilité si $\lim_{n\to+\infty} P(|X_n - X| \geq \varepsilon) = 0$

Loi faible des grands nombres pour cette convergence Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de variance finie, alors $\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X]$ et $\lim_{n\to+\infty} \frac{1}{n\varepsilon^2} Var(X) = 0$

3 Convergence presque sure

Définition $X_n \xrightarrow{p.s.} X$
$(X_n)_{n\in\mathbb{N}}$ converge presque sûrement vers X si $P(\{\omega \in \Omega, \lim_{n\to+\infty} X_n(\omega) = X(\omega)\}) = 1$

Loi Forte des grands nombres pour cette convergence Soit $(X_n)_{n\in\mathbb{N}}X$ une suite de v.a.i.i.d. d'espérance finie, alors $\frac{1}{n}\sum_{i=1}^{n} X_i \to E[X]$ et $\lim_{n\to+\infty} \frac{1}{n\varepsilon^2} Var(X) = 0$

4 Convergence en loi

Définition $X_n \xrightarrow{\mathcal{L}} X$
$(X_n)_{n\in\mathbb{N}}$ converge en loi vers X si en tout point de continuité de F_X, $F_{X_n}(x)$ converge vers $F_X(x)$

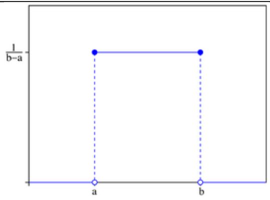
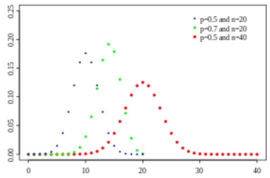
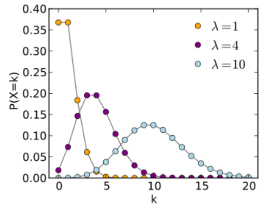
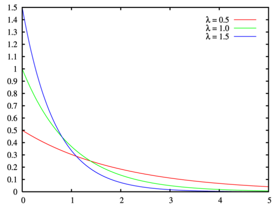
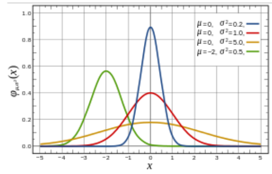
5 Relations entre les convergences

$$\begin{array}{ccc} & X_n \xrightarrow{L^\alpha} X & \\ & \Downarrow & \\ X_n \xrightarrow{p.s.} X \Rightarrow & X_n \xrightarrow{P} X & \Rightarrow X_n \xrightarrow{\mathcal{L}} X \end{array}$$

6 Théorème central limite

Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.i.i.d. de variance σ^2 finie.
On note $S_n = \sum_{i=1}^{n} X_i$ et $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
on a alors $\frac{S_n - E[S_n]}{\sqrt{Var(S_n)}} \xrightarrow{\mathcal{L}} Z$ de loi $\mathcal{N}(0,1)$ sans aucune hypothèse sur la loi des X_i

Part IX
Les différentes loi à connaître

Nom	Formule	Espérance	Variance	Graph
Uniforme $[a, b]$	$f_X(x) = \frac{1}{b-a} 1_{[a,b]}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	
Binomiale $\mathcal{B}(n, p)$	$P(X = k) = (\begin{smallmatrix} n \\ k \end{smallmatrix}) p^k (1-p)^{n-k}$	np	$np(1-p)$	
Poisson λ	$P(X = k) = \frac{\lambda^k e_{-\lambda}}{k!}$	λ	λ	
Exponentielle (λ)	$f_X(x) = \lambda e_{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	
Normale $\mathcal{N}(\mu, \sigma)$	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right]$	μ	σ^2	

Part X
Simulation de variables aléatoires

1 Generateur de nombres pseudo-aléatoires

Nombres pseudo-aléatoires : obtenus par différentes opérations (méthode de la congruence linéaire ou multiplication avec retenue) considérées comme irréversibles

2 Générer des réalisations d'un v.a. uniforme sur [0,1]

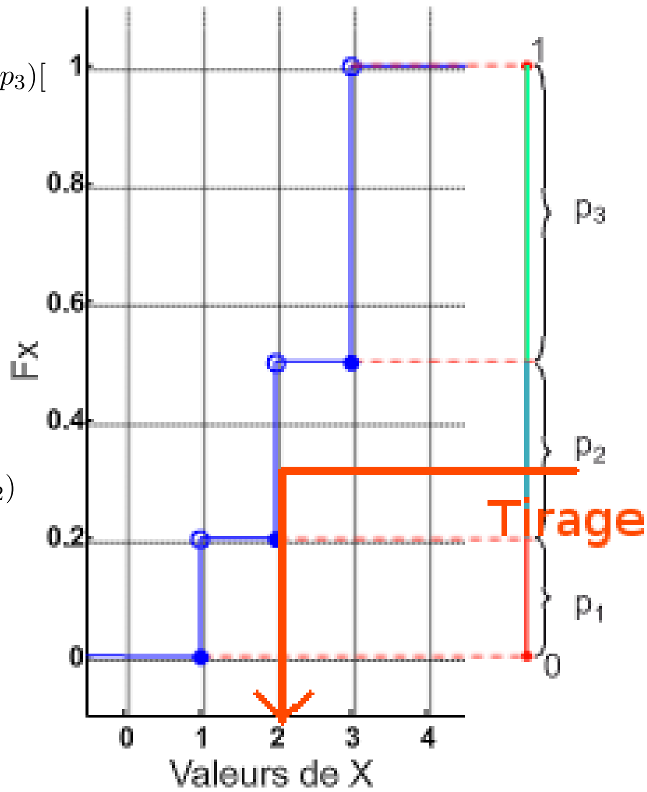
On décompose N en base 2 :$N = \sum_{n>0} a_n 2^n$ On va utiliser : $U = \sum_{n>0} a_n 2^{-n} \in [0, 1]$ On peut donc simuler des réalisations d'une loi uniforme sur $[0, 1]$

3 Méthode de la fonction de répartition inverse

Simulation des réalisations d'une loi discrète $\begin{cases} P(X=1) = P(U \in [0, p_1[) \\ P(X=2) = P(U \in [p_1, p_1+p_2[) \\ P(X=3) = P(U \in [p_1+p_2, p_1+p_2+p_3)[\end{cases}$

On fait un tirage u suivant $\mathcal{U}_{[0,1]}$ $\begin{array}{l} x = 1 \text{ si } u \in [0, p_1[\\ x = 2 \text{ si } u \in [p_1, p_1+p_2[) \\ x = 3 \text{ si } u \in [p_1+p_2, p_1+p_2+p_3)[\end{array}$

c'est-à-dire $x = \inf\{a, F_x(a) > u\} = F^{-1}(u)$
On a simulé une réalisation de X en "inversant" la fdr.



Méthode de la fonction de répartition inverse On a $F_X(X) = U$ qui suit une loi uniforme sur $[0,1]$. En inversant la fonction, on peut simuler la réalisation de x de X par $x = F_X^{-1}(u)$

4 Générer des lois gaussiennes multivariées

Soient U_1 et U_2 deux v.a. indépendantes uniformes sur $[0,1]$. Alors les v.a. $X = \sqrt{-2ln(U_1}cos(2\pi U_2)$ et $Y = \sqrt{-2ln(U_2}sin(2\pi U_2)$ sont normales, centrées, réduites et indépendantes. On peut donc simuler un vecteur gaussien centré et réduit, donc simuler tous les vecteurs gaussiens.

5 Méthode du rejet

Hypothèses Soient f, g deux densités de probabilités $\mathbb{R}$

- On sait simuler des v.a. de densité g
- $\exists c > 1,\ f(x) \leq cg(x)$ et $h : \{ \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ x & \mapsto & \frac{f(x)}{cg(x)} \end{array}$
- $Y_n, n \geq 1$ suite i.i.d. de v.a. de densité g
- $U_n, n \geq 1$ suite i.i.d. de v.a. de dloi $\mathcal{U}_{[0,1]}$

Alors : on pose

- $N = inf\{n, U_n \leq h(Y_n)\}$
- $X = Y_n$

On a X de densité f et $E[N] = c$.

6 Générer des réalisations d'un mélange de lois

Plus généralement, si g_y et h_z sont les densités de probabilité respectives de v.a. Y et Z et que $p \in]0,1[$ Pour générer des réalisations d'une v.a. X de densité $f_X = pg_Y + (1-p)h_z$ et si U suit une loi uniforme sur $[0,1]$: $X = Y\mathbf{1}_{\{U \leq p\}} + Z\mathbf{1}_{\{U > p\}}$

Pour simuler une réalisation x de X On simule une réalisation u de U, si $u > p$ on simule une réalisation z de Z et $x = z$, sinon on simule une réalisation y de Y et $x = y$

Part XI
Régression

1 Régression linéaire

Erreur quadratique moyenne $EQM(a,b) = E[(Y - (aX+b)^2)]$ (est fonction polynomiale de a et b)

Systeme d'équations pour minimiser l'erreur $\left\{\begin{array}{l}\dfrac{dEQM(a,b)}{db} = 0\\ \dfrac{dEQM(a,b)}{da} = 0\end{array}\right. \Longleftrightarrow \left\{\begin{array}{l} a = \dfrac{Cov(X,Y)}{Var(X)}\\ b = E[Y] - E[X]\cdot\dfrac{Cov(X,Y)}{Var(X)}\end{array}\right.$

L'erreur minimum s'écrit donc $EQM(a,b) = Var(Y) - \frac{Cov(X,Y)^2}{Var(X)}$

L'équation de la droite $Y = aX + b$ s'écrit donc aussi $\underbrace{\dfrac{Y-E[Y]}{\sigma_Y}}_{} = \dfrac{Cov(X,Y)}{\sigma_X\sigma_Y}\cdot\underbrace{\dfrac{Y-E[Y]}{\sigma_Y}}_{}$
v.a. centrées et réduites

On pose $Y' = \rho_{xy}\cdot X'$. La droite passe par $(E[X], E[y])$.
$EQM_{min}(a,b) = Var(Y)(1-\rho_{xy}^2)$

Résultats $\rho \in [-1,1]$, EQM est minimun pour $\rho_{xy} = -1$ ou 1, maximum pour 0. Dans ce cas, l'équation est $Y = \mu_y$ (ne dépend pas de X) et donc $Cov(X,Y) = 0$
Remarque: Pour $\rho \in]0,1[$ la relation entre X et Y est stochastique et non deterministe.

2 Régression non linéaire

Cette fois, sans présager de la forme de la fonction : $EQM(g) = E[(Y-g(X))^2] = E[(Y-E[Y/X])^2] + E[(E[Y/X]-g(X))^2]$. On en déduit $min_g EQM(g) = E[(Y-E[Y/X])^2]$. La meilleure approximation au sens de l'erreur quadratique moyenne est l'espérance conditionnelle.