Rappels sur le calcul des probabilités

1 - Définition formelle ou axiomatique

On appelle expérience aléatoire une expérience dont le résultat est le choix “ au hasard ” d'un élément d'un ensemble S, l'ensemble des épreuves.

On note E la classe des événements qui sont des sous-ensembles de S, enfin, on assigne à chaque événement A de E une mesure de probabilité P qui vérifie :

(1) $P(\Omega) = 1$

(2) $0 \leq P(A) \leq 1$

(3) $P(A + B) = P(A) + (B)$

si A et B sont mutuellement exclusifs dans E.

Ces trois axiomes constituent une définition implicite d'une probabilité - Le triplet (S, E, P) est appelé espace probabilisé.

2. Propriétés

- $P(\overline{A}) = 1 - P(A)$
- $P(A + B) = P(A) + P(B) - P(AB)$

Indépendance :

Deux événements sont indépendants si

$P(A.B) = P(A).P(B)$

Les événements A_1, A_2, ..., A_n sont indépendants dans leur ensemble si :

$P(A_1, ..., A_n) = P(A_1),...,P(A_n)$

Rq : des événements indépendants 2 à 2 ne le sont pas nécessairement dans leur ensemble.

Deux événements indépendants ne sont pas incompatibles.

En effet :

P(A+B) = P(A) + P(B)	incompatibles
P(A+B) = P(A) + P(B) - P(A).P(B)	indépendants

Probabilités conditionnelles :

$$P(AB) = P(A|B)\,P(B) = P(B|A)\,P(A)$$

P(A|B) est la probabilité conditionnelle de A si B est réalisé. Pour des événements indépendants, P(A|B) = P(A). La formule

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

est la formule de BAYES.

Exemple : canal binaire symétrique

Il s'agit d'un canal de transmission sur lequel on cherche à transmettre des 0 et des 1. En raison du bruit d'observation, on fait occasionnellement une erreur de décision en sortie, c'est-à-dire qu'on décide un 1 alors qu'un 0 a été envoyé, et vice-versa. On suppose le canal sans mémoire, c'est-à-dire que la sortie ne dépend que de l'entrée à un instant donné. On le suppose symétrique, c'est-à-dire que les erreurs interviennent avec la même probabilité, q'un 1 ou un 0 ait été émis.

On note $P_0 = P(X = 0)$ et $P_1 = P(X=1)$. On a nécessairement $P_0 + P_1 = 1$.
La probabilité d'erreur est notée P :

$$P = P(Y = 1 \mid X = 0) = P(Y = 0 \mid X = 1)$$

La probabilité d'obtenir un 1 en sortie vaut alors

$$\begin{aligned} P(Y = 1) &= P(Y = 1 \mid X = 0)\, P(X = 0) + P(Y = 1 \mid X = 1)\, P(X = 1) \\ &= P \;.\; P_0 + P(Y = 1 \mid X = 1)\;.\;P_1 \end{aligned}$$

Comme $P(Y = 1 \mid X = 1) + P(Y = 0 \mid X = 1) = 1$
on a $P(Y = 1 \mid X = 1) = 1 - P$
Enfin, $P(Y = 1) = P\,P_0 + (1 - P)P_1$
et de la même manière $P(Y = 0) = P_0\,(1 - P) + PP_1$

On résume ceci par le diagramme de transition

Quelle est maintenant que l'entrée soit à 1 si la sortie est à 1 ? La solution est donnée par la formule de BAYES :

$$P(X = 1 \mid Y = 1) = \frac{P(Y = 1 \mid X = 1).\,P(X = 1)}{P(Y = 1)} = \frac{(1 - P).\,P_0}{PP_0(1 - P)P_1}$$

puis $P(X = 0 \mid Y = 1) = 1 - P(X = 1 \mid Y = 1)$
et similairement pour les autres probabilités.
Les probabilités de l'entrée <u>compte tenu des observations</u>, sont appelées <u>probabilités à posteriori.</u>

VARIABLES ALEATOIRES

On appelle Variable aléatoire la fonction qui associe à l'espace d'épreuve S (ou à un sous-ensemble) un ensemble de nombres réels (éventuellement complexes) :

$\forall w \in \Omega \xrightarrow{X} x = X(w) \in \Re$

Si X peut prendre n'importe quelle valeur sur un intervalle de ú, ou sur tout ú, la variable aléatoire est dite continue.

Pour obtenir une description probabilité de la variable aléatoire X, on s'intéresse à l'événement X # x et on définit la fonction de répartition.

$$F_x(x) = P(X \le x)$$

Comme il s'agit d'une probabilité, on a

(1) $0 < F_X(x) \le 1$

(2) $F_X(-\infty) = 0 \qquad F_X(+\infty) = 1$

(3) $F_X(x_1) < F_X(x_2)$ si $x_2 > x_1$

(fonction monotone non décroissante).

NB : Cette fonction est continue à gauche. Elle peut ne pas être continue à droite (probabilité non nulle en un point - masse de probabilité).

$$\begin{aligned} P(x_1 \le X \le x_2) &= P(X \le x_2) - P(X \le x_1) \\ &= F_X(x_2) - F_X(x_1) \qquad (4) \\ &= \int_{x_1}^{x_2} dF_X(x) \end{aligned}$$

A partir de la propriété (4), en prenant x_1 = - 4 et x_2 = x, on a

$$F_x(x) = P(X \le x) = \int_{-\infty}^{x} dF_x(x)$$

Si F_X (x) est continue et différentielle pour tout x, on pose dF_X (x) = f_X (x) d_x , avec f_X (x) $ 0 [puisqu'il s'agit de la dérivée d'une fonction non décroissante]. F_X (x) est appelée densité de probabilité, mais CE N'EST PAS UNE PROBABILITE.

On a en fait

$$\begin{aligned} dF_X(x) &= F_X(x) - F_X(x + dx) \\ &= P(x < X \le x + dx) \\ &= f_X(x)dx \end{aligned}$$

C'est alors f_X (x)dx qui est une probabilité.

La fonction de répartition peut maintenant s'écrire :

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx,$$

et puisque F_X (+4) = 1, on a

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

EXEMPLE - Loi uniforme

Plusieurs variables aléatoires

La fonction de répartition pour 2 variables est définie par

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

On définit la densité de probabilité conjoint par

$$dF_{X,Y}(x,y) = f_{X,Y}(x,y)dx\ \ dy,$$

en supposant que la fonction de répartition est partout différentiable.

On a ainsi $\displaystyle\int_{-\infty}^{y}\int^{x} f_{X,Y}(x,y)dx\ \ dy = F_{X,Y}(x,y)$

On peut obtenir la fonction de répartition d'une seule des variables en notant que

$$\begin{aligned} F_{X,Y}(x,+\infty) &= P(X \leq x, Y \leq +\infty) \\ &= P(X \leq x \mid Y \leq +\infty).\qquad \underbrace{P(Y \leq +\infty)}_{\text{évènement toujours réalisé,de proba=1}} \\ &= P(X \leq x) \\ &= F_X(x) \end{aligned}$$

D'où $\displaystyle F_X(x) = \int_{-\infty}^{x}\int_{-\infty}^{+\infty} f_{X,Y}(\zeta,\eta)d\zeta\, d\eta$

Et on identifie $\displaystyle f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,\eta)d\eta$

Cette opération s'appelle opération de marginalisation. Elle permet d'obtenir la densité de probabilité d'une des variables en intégrant la densité de probabilité conjointe pour toutes les valeurs de la variable non désirée.

On peut donc noter que la densité conjointe contient toute l'information sur les 2 variables, puisque l'on peut obtenir la densité de probabilité de chacune des variables par marginalisation.

On définit la densité de probabilité conditionnelle par

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

C'est la densité de Y avec X = x donné.
La fonction $f_{Y|X}(y|x)$ est une fonction de la variable y, avec x arbitraire, mais <u>fixé</u>. $f_{Y|X}(y|x)$ est une densité de probabilité, et, en particulier

$$\int_{-\infty}^{+\infty} f_{Y|X}(y|x)dy = 1$$

Lorsque X et Y sont indépendants, on a

$$f_{Y|X}(y|x) = f_Y(y)$$

et

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Tout ceci se généralise à un nombre quelconque de variables aléatoires.

MOYENNES STATISTIQUES

On peut chercher à caractériser une variable aléatoire par des comportements " moyens " des résultats des expériences aléatoires.

Espérance mathématique :
Si X(w) est intégrable par rapport à la mesure de probabilité P, on montre que

$$E[X] = \int_{\Omega} X(w)dP = \int_{\Re} x dF_X(x)$$

$$= \int_{\Re} x f_X(x)dx \qquad \text{si la loi est à densité}$$

E[-] note l'opérateur de moyenne statistique.
Pour une variable aléatoire discrète, on a E[X] = 3x_i p_i
On peut généraliser ceci à toute fonction g(x) intégrale : Y = g(X) est une variable aléatoire, et

$$E[g(X)] = \int_{\Re} g(x)f_X(x)\, dx$$

Exemple :
Soit Y = cos (X), ou X est distribuée uniformément sur [0,2 B].

$$E[Y] = \int_0^{2\Pi} \cos(x).\frac{1}{2\Pi}dx = -\frac{1}{2\Pi}\left[\sin(x)\right]_0^{2\Pi} = 0$$

Deux cas particuliers sont importants :

- si g(x) = x^n, on obtient les <u>moments</u> de la loi

  $$E[X^n] = \int_{\Re} x^n f_X(x)dx$$

  n = 1 : moyenne

n = 2 : moment d'ordre 2

On s'intéresse aussi souvent aux moments centrés, c'est-à-dire aux moments de la variable aléatoire centrée

$X_c = X - E[X]$

En notant $m_X = E[X]$, les moments centrés sont

$$\int_{\Re} (x-m_X)^n f_X(x)dx$$

Pour n = 1, on a bien sûr zéro. Pour n = 2, on obtient la variance :

$$Var[X] \hat{=} \sigma^2 = \int_{\Re} (x-m_X)^2 f_X(x)dx$$

En développant, on notera aussi que

$$\sigma^2 = E\left[(X-m_X)^2\right] = E\left[X^2 - 2m_X \times + m_X^2\right]$$
$$= E[X^2] - {m_X}^2$$

La variance mesure, en un certain sens, le degré d'aléatoire de X. Il s'agit d'une mesure des fluctuations autour de la moyenne m_X. L'inégalité de Chebyshev rend compte de ceci :
pour tout $\varepsilon > 0$,

$$P\left(\left|X-m_X\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2}$$

- Comme second cas particulier important de la moyenne E[g(X)], on a la définition de la fonction caractéristique, notée N_X (v),

$$\phi_X(v) = E\left[e^{jvX}\right]$$
$$= \int_{-\infty}^{+\infty} f_X(x) e^{jvx} dx$$

A un signe près, et au facteur 2B près, la fonction caractéristique est la transformée de Fourier de la densité de probabilité f_x (x) ...

En dérivant K fois la fonction caractéristique, on obtient

$$\frac{d^K P_X(v)}{dv^K} = \frac{j^K}{K!} E\left[X^K\right] + \frac{j^{K+1} v}{(K+1)!} E\left[X^{K+1}\right] + \ldots$$

et en v = 0,

$$\frac{d^K \phi_X(0)}{dv^K} = \frac{j^K}{K!} E\left[X^K\right]$$

On notera pour mémoire que la seconde fonction caractéristique est définie par
$\psi_X(v) = \log \phi_X(v)$,

et que les coefficients de son développement sont appelés cumulants.

Exemple : variable gaussienne

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{(x - m_X)^2}{2\sigma^2}$$

$$\phi_X(v) = \exp\left(jm_X v - \frac{1}{2} v^2 \sigma^2 \right)$$

Comme la TF est inversible, on peut bien-entendu retrouver la densité de proba à partir de la fonction caractéristique :

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \phi_X(v) e^{-jvx} dv$$

La fonction caractéristique porte donc autant d'information que la densité. En particulier, si on considère le développement en série de

$$e^{jvx} = 1 + j^x + j\frac{x^2}{2} + \ldots + j\frac{x^n}{n!} + \ldots,$$

$$E\left[e^{jvx}\right] = 1 + jvE[X] + \frac{(jv)^2}{2} E[X^2] + \ldots + \frac{(jv)^n}{n!} [X^n]$$

les moments sont les coefficients du développement en série de $\phi_X(v)$.

Moments joints

Si on considère une paire de variables aléatoires X et Y, on définit

$$E\left[X^i Yj\right] = \iint x^i y^j f_{X,Y}(x, y) dx\, dy$$

Un moment joint particulier est la corrélation E[X,Y].
Dans le cas des variables centrées, la corrélation

$$E[(X - m_X)(Y - m_Y)] = E[XY] - m_X m_Y$$

est appelée covariance, notée cov(X,Y).
On dit que deux variables aléatoires sont décorrélées si leur covariance st nulle :

$$\operatorname{cov}(X, Y) = E[X_c Y_c] = 0$$

Elles sont orthogonales si leur corrélation est nulle.
Notons que si X et Y sont indépendantes, alors elles sont décorrélées (orthogonales). Par contre, l'inverse n'est pas vrai.

Transformations des variables aléatoires

Il peut être utile de pouvoir déterminer la densité de probabilité d'une variable aléatoire obtenue comme transformée d'une première va. dont on connaît la densité. C'est typiquement le problème de la détermination de la densité de probabilité de
Y = g(X)

On considère le cas d'une transformation bijective.
On a

$P(y < Y \leq y + dy) = P(x < X \leq x + dx),$

ou $\quad y = g(x)$ et $dy = \dfrac{d_g}{d_x}.dx$

On en déduit que $\quad f_Y(y)dy = f_X(x)dx$

si $\dfrac{d_g}{d_x} > 0$ (fonction monotone croissante), et

si $\dfrac{d_g}{d_x} < 0 \quad f_Y(y)dy = -f_X(x)dx$

En combinant ces deux relations, il vient

$$f_Y(y) = f_X(x)\left|\frac{d_x}{d_y}\right|$$

Enfin, en utilisant $x = g^{-1}(y)$, on obtient

$$f_Y(y) = f_X\left(g^{-1}(y)\right).\frac{1}{\frac{d_g}{d_x}\Big|x = g^{-1}(y)}$$

Dans le cas ou plusieurs valeurs x_k fournissent la même observation y, i.e $y = g(x_k)$ $k = 1 \ldots K$, l'expression de la densité est

$$f_Y(y) = \sum_K f_X(x_K).\frac{1}{\left|\frac{d_g}{dx_K}\right|}\Big|x_K = g^{-1}(y).$$