SVM 201
Rappels de Probabilités

Alain Yves Le Roux

1 Lois de probabilités

Soit $\Omega \neq \emptyset$ un ensemble donné, et $\mathcal{P}(\Omega)$ l'ensemble des parties de Ω. Une loi de probabilité sur Ω correspond à la donnée d'une application p définie sur $\mathcal{P}(\Omega)$, à valeurs dans $\mathbb{R}$, telle que

$$\begin{gathered} p(\Omega) = 1 \ , \\ \forall A \subset \Omega \ \ p(A) \geq 0 \ , \\ \forall A_1, A_2, ..., A_n, .. \subset \Omega \ tels \ que \ A_i \cap A_j \ si \ i \neq j, \ \forall K \subset \mathbb{N} : \ p(\textstyle\bigcup_{i \in K} A_i) \ = \sum_{i \in K} p(A_i) \end{gathered}$$

Les parties de Ω (c'est à dire les éléments de $\mathcal{P}(\Omega)$) sont appelés des évènements. Un élément ω de Ω correspond à une partie $\{\omega\} \subset \Omega$, appelée évènement élémentaire. Ils correspondent à différents résultats d'une épreuve bien déterminée. La valeur $p(A)$, pour $A \subset \Omega$, est appelée probabilité de A. Si tous les évènements élémentaires ont la même probabilité, ils sont dits équiprobables, et la loi de probabilité est la loi uniforme.

Par exemple, considérons l'épreuve consistant à tirer "au hasard" une carte d'un jeu de 52 cartes (rectangulaires), et les évènements suivants :

$$\begin{array}{ll} A_1 = \{carte\ rectangle\} & alors\ p(A_1) = 1 \\ A_2 = \{carte\ jaune\} & alors\ p(A_2) = 0 \\ A_3 = \{carte\ rouge\} & alors\ p(A_3) = \frac{1}{2} \\ A_4 = \{dame\ rouge\} & alors\ p(A_4) = \frac{1}{26} \\ A_5 = \{dame\ de\ coeur\} & alors\ p(A_5) = \frac{1}{52} \end{array}$$

L'évènement A_5 est un évènement élémentaire.

Deux évènements A et B sont indépendants lorsque

$$p(A \cap B) \ = \ p(A)p(B) \tag{1.1}$$

On définit la probabilité conditionnelle de "A *sachant* B", noté A/B par

$$p(A/B) \ = \ \frac{p(A \cap B)}{p(B)} \ , \tag{1.2}$$

lorsque $p(B) > 0$.

Typeset by the $\TeX$-preprocessor PreTeX

2 Variables Aléatoires

Une variable aléatoire (souvent notée V.A.) est une application définie sur Ω, à valeurs dans $\mathbb{R}$. Elle correspond à la valeur quantitative du résultat d'une épreuve. Dans l'exemple précédent du jeu de 52 cartes, les résultats ne sont pas des nombres réels, mais ils peuvent le devenir si on numérote les cartes de 1 à 52. Le numéro de la carte sortie est la valeur de la V.A. associée à cette épreuve.

Dans un épreuve de type "pile ou face", on peut associer la valeur 1 à pile, et 0 à face, par exemple, et définir ainsi une V.A. à valeurs dans $\{0,1\}$.

Soit X une V.A. correspondant à une loi de probabilité p définie sur $\mathcal{P}(\Omega)$. De façon "abusive", on notera, pour toute partie $B \subset \mathbb{R}$,

$$p(B) \;=\; p(X^{-1}(B)) \;=\; p(\{\omega \in \Omega | X(\omega) \in B\}) \;\;,$$

ce qui revient à identifier la loi de probabilité sur Ω et celle induite par X, sur $\mathbb{R}$.

La fonction de répartition F d'une V.A. X est définie sur $\mathbb{R}$ par

$$F(x) \;=\; p(X \leq x)\;. \tag{2.1}$$

Il s'agit d'une fonction non décroissante, de limite 0 lorsque $x \longrightarrow -\infty$, et de limite 1 lorsque $x \longrightarrow +\infty$.

Deux V.A. X et Y sont indépendantes si pour tout x, y réels, on a

$$p(X \leq x, Y \leq y) \;=\; p(X \leq x)\, p(Y \leq y) \tag{2.2}$$

ce qui revient à dire que les évènements $\{X \leq x\}$ et $\{Y \leq y\}$ sont indépendants.

En pratique, des épreuves très variées peuvent correspondre à la même V.A. On distingue essentiellement deux classes de V.A., les variables discrètes et les variables continues.

2.1 Variables aléatoires discrètes

Les valeurs de la V.A. constituent un ensemble fini ou dénombrable; on les note x_j pour $j \in J$, $J \subset \mathbb{N}$. En notant X la V.A., on pose, pour tout $j \in J$,

$$p(X = x_j) \;=\; p_j \;\;, \tag{2.3}$$

et comme les évènements élémentaires $\{X = x_j\}$ sont disjoints, on aura

$$\sum_{j \in J} p_j \;=\; 1 \;\;. \tag{2.4}$$

La fonction de répartition F est une fonction constante par morceaux, et les discontinuités sont situées aux points x_j, $j \in J$, avec un saut (croissant) de longueur p_j.

La moyenne de la V.A. X est définie par

$$E(X) \;=\; m \;=\; \sum_{j \in J} x_j\, p_j\;, \tag{2.5}$$

lorsque cette quantité existe (il faut que la série converge absolument si J est infini). Lorsque $m = 0$, on dit que la V.A. est centrée.

La variance de la V.A. X est définie par

$$v(X) \;=\; \sigma^2 \;=\; E((X-m)^2) \;=\; \sum_{j\in J} (x_j-m)^2\, p_j \quad , \tag{2.6}$$

lorsque cette quantité existe (il faut que la série converge si J est infini). Lorsque $v(X) = 1$, et $m = 0$, on dit que la V.A. est réduite. La quantité (positive) $\sigma = \sqrt{v(X)}$ est appelée écart type de la V.A. X.

Pour $k \in \mathbb{N}$, on définit le moment d'ordre k de la V.A. X par

$$M_k(X) \;=\; \sum_{j\in J} x_j^k\, p_j \;,$$

toujours lorsque cette quantité existe (il faut que la série converge absolument si J est infini). Le moment centré d'ordre k est égal à $M_k(X-m)$. Notons que $M_0(X) = 1,\ M_1(X) = m,\ M_2(X) = m^2 + \sigma^2$.

2.2 Exemples de lois discrètes

La loi uniforme correspond à un ensemble J fini, de taille N, et à la loi de probabilité

$$\forall j \in J \qquad p(x_j) \;=\; \frac{1}{N} \;. \tag{2.7}$$

On peut décider, en reprenant éventuellement la numérotation, d'avoir $J = \{1, 2, .., N\}$. La moyenne et la variance sont alors données par

$$m \;=\; \frac{x_1 + x_2 + \ldots x_N}{N} \quad , \quad \sigma^2 \;=\; \frac{(x_1-m)^2 + (x_2-m)^2 + \ldots + (x_N-m)^2}{N} \;.$$

La loi de Bernoulli correspond à la donnée d'un paramètre $p \in [0,1]$, et la V.A. ne prend que les valeurs 0 et 1, de telle façon que

$$p(X=1) \;=\; p \quad ; \quad p(X=0) \;=\; 1-p \;. \tag{2.8}$$

La moyenne vaut $m = p$, et la variance vaut $\sigma^2 = p(1-p)$, donc l'écart type vaut $\sigma = \sqrt{p(1-p}$.

La loi binomiale $\mathcal{B}(n,p)$ correspond à la somme de n V.A. de Bernoulli indépendantes et de même paramètre p. Elle prend des valeurs entières, entre 0 et n, avec

$$p(X=k) \;=\; \binom{n}{k}\, p^k (1-p)^{n-k} \;, \quad \textit{où}\; \binom{n}{k} = \frac{n!}{k!(n-k)!} \;. \tag{2.9}$$

La moyenne vaut $m = np$ et la variance $\sigma^2 = np(1-p)$, donc $\sigma = \sqrt{np(1-p)}$.

La loi de Poisson $\mathcal{P}(\lambda)$ correspond à la donnée d'un paramètre $\lambda > 0$. La V.A. prend des valeurs entières (y compris 0) avec

$$p(X=n) \;=\; \frac{\lambda^n}{n!}\, e^{-\lambda} \;. \tag{2.10}$$

La moyenne vaut $m = \lambda$, et la variance vaut $\sigma^2 = \lambda$ $(= m)$, et donc $\sigma = \sqrt{\lambda}$.

La loi géométrique correspond à la donnée d'un paramètre $a \in]0, 1[$. La V.A. prend des valeurs entières (à partir de 1) avec

$$p(X = n) \;=\; a^{n-1}\,(1 - a). \tag{2.11}$$

La moyenne vaut $m = \dfrac{1}{1-a}$, et la variance vaut $\sigma^2 = \dfrac{a}{(1-a)^2}$, et donc $\sigma = \dfrac{\sqrt{a}}{1-a}$.

Ajoutons à cette liste la V.A. nulle, telle que $p(X = 0) = 1$, de moyenne $m = 0$, et de variance $\sigma^2 = 0$, que l'on utilise dans les théorèmes de convergence, lorsqu'on dit qu'une V.A. tend vers zéro.

2.3 Variables Aléatoires continues

Les valeurs de la V.A. constituent un ensemble continu ; on suppose que la fonction de répartition F est la primitive d'une fonction f, appelée densité de probabilité. La moyenne de la V.A. X est définie par

$$E(X) \;=\; m \;=\; \int_{-\infty}^{+\infty} x\, f(x)\, dx\ , \tag{2.12}$$

lorsque cette quantité existe (il faut que l'intégrale converge absolument). Lorsque $m = 0$, on dit que la V.A. est centrée.

La variance de la V.A. X est définie par

$$v(X) \;=\; \sigma^2 \;=\; E((X-m)^2) \;=\; \int_{-\infty}^{+\infty} (x-m)^2\, f(x)\, dx\ , \tag{2.13}$$

lorsque cette quantité existe (il faut que l'intégrale converge). Lorsque $v(X) = 1$, et $m = 0$, on dit que la V.A. est réduite. La quantité (positive) $\sigma = \sqrt{v(X)}$ est appelée écart type de la V.A. X.

Pour $k \in \mathbb{N}$, on définit le moment d'ordre k de la V.A. X par

$$M_k(X) \;=\; \int_{-\infty}^{+\infty} x^k\, f(x)\, dx\ ,$$

toujours lorsque cette quantité existe (il faut que l'intégrale converge absolument). Le moment centré d'ordre k est égal à $M_k(X-m)$. Notons que $M_0(X) = 1,\ M_1(X) = m,\ M_2(X) = m^2 + \sigma^2$.

2.4 Exemples de Variables Aléatoires continues

La loi uniforme sur un intervalle $]a, b[$ correspond à la densité de probabilités

$$f(x) \;=\; \begin{cases} \frac{1}{b-a} & si\ \ a < x < b\ , \\ 0 & sinon \end{cases} \tag{2.14}$$

La moyenne vaut $m = \dfrac{a+b}{2}$ et la variance $\sigma^2 = \dfrac{(b-a)^2}{12}$, donc $\sigma = \dfrac{b-a}{2\sqrt{3}}$. Notons que la fonction de répartition est donnée par

$$F(x) \;=\; \begin{cases} 1 & si\ x \geq b\ , \\ \frac{x-a}{b-a} & si\ a < x < b\ , \\ 0 & si\ x \leq a\ \ . \end{cases}$$

et qu'elle tend vers la fonction de répartition de la V.A. nulle lorsque b et a tendent vers zéro.

La loi normale depend de deux paramètres m et $\sigma > 0$, respectivement sa moyenne et son écart type, et a pour densité

$$f(x) \; = \; \frac{1}{\sigma\sqrt{2\pi}} \; e^{-\frac{(x-m)^2}{2\sigma^2}} \; . \tag{2.15}$$

Lorsque $m = 0$ et $\sigma = 1$, il s'agit de la loi normale réduite.

La loi exponentielle dépend d'un paramètre $\lambda > 0$, et a pour densité

$$f(x) \; = \begin{cases} \lambda \, e^{-\lambda x} & si \; x > 0 \\ 0 & si \; x \leq 0. \end{cases} \tag{2.16}$$

La moyenne vaut $m = \frac{1}{\lambda}$ et de variance $\sigma^2 = \frac{1}{\lambda^2}$, donc $\sigma = \frac{1}{\lambda}$.

La loi Gamma est construite à partir de la fonction

$$\Gamma(q) \; = \; \int_0^{+\infty} \; t^{q-1} \; e^{-t} \; dt \; ,$$

définie pour $q > 0$, et qui vaut $(q-1)!$ lorsque q est un entier ≥ 1. La loi dépend de deux paramètres $\lambda > 0$ et $q > 0$, et a pour densité

$$f(x) \; = \; \begin{cases} \frac{\lambda^q x^{q-1}}{\Gamma(q)} \; e^{-\lambda x} & si \; x > 0, \\ 0 & si \; x \leq 0. \end{cases} \tag{2.17}$$

Pour $q = 1$, on retrouve la loi exponentielle. La moyenne vaut $m = \frac{q}{\lambda}$ et la variance $\sigma^2 = \frac{q}{\lambda^2}$, donc $\sigma = \frac{\sqrt{q}}{\lambda}$. Lorsque $0 < q < 1$, la densité n'est pas bornée (en x proche de 0).

La loi du χ^2 (ou khi-deux) est un cas particulier de la loi Gamma, lorsque $\lambda = \frac{1}{2}$ et $q = \frac{n}{2}$, avec $n \geq 1$, entier. Il n'y a plus qu'un seul paramètre n, appelé "nombre de degrés de liberté". On peut évaluer le coefficient $\Gamma(\frac{n}{2})$ en utilisant la formule de récurrence $\Gamma(q+1) = q\Gamma(q)$, et $\Gamma(\frac{1}{2}) = \sqrt{\pi}$,

$$\Gamma(\frac{n}{2}) \; = \; \begin{cases} \frac{(n-1)!\sqrt{\pi}}{2^{n-1} \; (\frac{n-1}{2})!} & si \; n \; impair, \\ (\frac{n}{2} - 1)! & si \; n \; pair. \end{cases}$$

3 Deux théorèmes de convergence

On dit qu'une suite de V.A. X_n converge en loi vers une V.A. X, lorsqu'il y a convergence de la suite des fonctions de répartition F_n de X_n, vers la fonction de répartion F de X, en tout point de continuité de F.

On dit que la suite de V.A. X_n converge en probabilité vers la V.A. X lorsque pour tout $\epsilon > 0$, la quantité $p(|X_n - X| > \epsilon)$ tend vers zéro lorsque n tend vers l'infini.

La convergence en probabilité implique la convergence en loi.

Théorème 3.1 (Théorème Central Limite). *Soit X_n une suite de V.A. indépendantes, toutes de même loi admettant m comme moyenne et σ comme écart type. Alors la V.A.*

$$S_n \; = \; \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} (X_i - m) \tag{3.1}$$

converge en loi vers une V.A. suivant une loi normale réduite.

Remarque : En posant $\overline{X}_n = \frac{1}{n}(X_1 + X_2 + ... + X_n)$, on obtient $S_n = \frac{\sqrt{n}}{\sigma}(\overline{X}_n - m)$, qui, sous les hypothèses du théorème, converge donc en loi vers une loi normale réduite.

Théorème 3.2 (Loi faible des grands nombres) *Soit X_n une suite de V.A. indépendantes, centrées, telles que leurs variances $v(X_n) = \sigma_n^2$ vérifient*

$$\lim_{n \longrightarrow +\infty} \frac{1}{n^2} \sum_{k=1}^{n} \sigma_k^2 \quad = \quad 0 \quad , \tag{3.2}$$

Alors $\overline{X}_n$ converge en probabilité vers zéro (la V.A. nulle).

Remarque : Si la suite X_n est de variance constante, égale à σ^2, on a $\frac{1}{n^2}\sum_{k=1}^{n} \sigma_k^2 \;=\; \frac{\sigma^2}{n}$ qui tend bien vers zéro. Ce résultat est encore acquis si les variances σ_n^2 restent bornées.

4 D'autres résultats utiles

Si X et Y sont deux V.A., λ et μ deux nombres réels, alors $\lambda X + \mu Y$ est une V.A. et on a

$$E(\lambda X + \mu Y) \;=\; \lambda E(X) \;+\; \mu E(Y) \,. \tag{4.1}$$

De plus, si X et Y sont indépendantes, on a

$$E(XY) \;=\; E(X)\; E(Y) \tag{4.2}$$

et

$$v(\lambda X + \mu Y) \;=\; \lambda^2\, v(X) \;+\; \mu^2\, v(Y) \,. \tag{4.3}$$

La moyenne est en général différente de la médiane d'une V.A., définie par la valeur a déterminée par les relations

$$\forall \epsilon > 0 \quad F(a-\epsilon) \;\leq\; \frac{1}{2} \quad , \quad F(a+\epsilon) \;\geq\; \frac{1}{2} \;.$$

Dans le cas d'une V.A. dicrète, en notant x_k le mode, c'est à dire parmi les valeurs x_j celle qui réalise $p_k = \sup_j(p_j)$, on a la formule empirique de Pearson : $3a \approx x_k + 2m$.

Dans le cas d'une V.A. continue, on a exactement $F(a) = \frac{1}{2}$.