Chapitre 4 - Probabilités

Julien GREPAT (cours) - Mica MURPHY (note) - Antoine SAGET (note)

Vendredi 15 février 2019

1 - Introduction

On considère une expérience aléatoire. L'ensemble des futurs possibles est noté Ω. Un événement est un sous ensemble A de Ω. L'ensemble des événements de Ω que l'on peut considérer est la tribu $\mathcal{F}$

$A, B \in \mathcal{F}$ alors $A \cup B \in \mathcal{F}$, $A \cap B \in \mathcal{F}$, $\overline{A} \in \mathcal{F}$

La probabilité $\mathcal{P} : \left\{ \begin{array}{ccc} \mathcal{F} & \rightarrow & [0,1] \\ A & \mapsto & \mathcal{P}(A) \end{array} \right.$ est la fréquence théorique de l'évènement A dans Ω.

$\mathcal{P}$ est une mesure :

- Soit $A, B \in \mathcal{F}$ et $A \cap B = \emptyset$ alors $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B)$
- $\mathcal{P}(\emptyset) = 0$ et $\mathcal{P}(\Omega) = 1$

Propriétés. - $\mathcal{P}(A \cup B) = \mathcal{P}(A) + \mathcal{P}(B) - \mathcal{P}(A \cap B)$ - $\mathcal{P}(A \cup B) \leq \mathcal{P}(A) + \mathcal{P}(B)$ - $\mathcal{P}(A \backslash B) = \mathcal{P}(A) - \mathcal{P}(A \cap B)$ - etc.

Définition. Probabilités conditionnelles : soient $A, B \in \mathcal{F}$. Supposons que A se réalise. On mesure ce qu'il "reste" de B dans A.

$$\mathcal{P}(B|A) = P_A(B) = \text{"}\frac{\text{mesure } (A \cap B)}{\text{mesure } (A)} \times \frac{\text{mesure } \Omega}{\text{mesure } \Omega}\text{"} = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(A)}$$

Définition. $A, B \in \mathcal{F}$ sont indépendants si $\mathcal{P}_A(B) = \mathcal{P}(B)$

On a $\mathcal{P}_A(B) = \frac{\mathcal{P}(A \cap B)}{\mathcal{P}(A)} \Leftrightarrow \mathcal{P}(A \cap B) = \mathcal{P}(A) \times \mathcal{P}_A(B)$

Si A et B sont indépendants : $\mathcal{P}(A \cap B) = \mathcal{P}(A) \times \mathcal{P}_A(B) = \mathcal{P}(A) \times \mathcal{P}(B)$

$(\mathcal{B})_{i=1}^n$ est un système complet d'évènements : $\mathcal{B}_i \cap \mathcal{B}_j = \emptyset$ et $\bigcup\limits_{i=1}^{n} \mathcal{B}_i = \Omega$

Formule de probabilités totales. $\mathcal{P}(A) = \sum\limits_{i=1}^{n} \mathcal{P}(A \cap \mathcal{B}_i)$

2 - Variables aléatoires

On peut (presque toujours) quantifier les résultats d'une expérience.

$$\mathcal{P}(X \in [a,b]) = \mathcal{P}(\{\omega \in \Omega : X(\omega) \in [a,b]\})$$

Loi de Bernouilli. Soit $X \sim \mathcal{B}(p)$, X vaut 1 si réussite avec probabilité p et 0 sinon (échec).

$$\mathcal{P}(X = 1) = p \; ; \; \mathcal{P}(X = 0) = 1 - p$$

Loi binomiale. Soit $X \sim \mathcal{B}(n,p)$, X compte le nombre de succès à n expériences de Bernouilli identiques et indépendantes avec probabilité de succès p.

$$\mathcal{P}(X=k) = \binom{n}{p} p^k (1-p)^{n-k}$$

Rappel. $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Loi géométrique. $X \sim \mathcal{G}(p)$

X donne le rang de la première réussite à une succession d'expériences de Bernouilli indépendantes, identiques et de paramètre p.

$$\mathcal{P}(X=k) = (1-p)^{k-1}p$$

Loi de Poisson. Soit $X \sim \mathcal{P}(X)$ on a :

$$\mathcal{P}(X=\lambda) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Remarque. $e^{\lambda} = \sum \frac{\lambda^k}{k!}$

On s'intéresse, dans un premier temps aux variables prenant un nombre dénombrable de valeurs ($\mathbb{N}$)

Loi de probabilité, fonction de répartition. La loi de probabilité est la donnée de toute les probabilités $P(X = x_k)$.

Définition. La fonction de répartition est définie par

$$F_X(x) = P(X \leq x)$$

Théorème. La fonction de répartition détermine entièrement une loi de probabilité.

3 - Espérance, variance, écart-type

Définition. La valeur moyenne des résultats de l'expérience X est nommée **espérance** de X, notée $E[X]$ ou EX

$$EX = \sum_{i\geq 1} x_i \times P(X = x_i)$$

Remarque. EX est la valeur que prendrait X si l'expérience n'était pas aléatoire. Si $a \in \mathbb{R}$ est une constante $Ea = a$.

Propriété. La moyenne est **linéaire** : X, Y variables aléatoire, λ, β, $\gamma \in \mathbb{R}$

$$E[\lambda X + \beta Y + \gamma]\lambda EX + \beta EX + \gamma$$

Définition. L'écart-type est la distance moyenne entre les valeurs de X et son espérance.

$$\sigma_x = \sqrt{\sum_{i\geq 1}(x_i - EX)^2 P(X = x_i)}$$

Remarque. Formule de la distance dans $\mathbb{R}^3$: $\sqrt{(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2}$

$$VarX = \sigma_x^2$$

Théorème du transfert. $E\varphi(X) = \sum_{i\geq 1} \varphi(x_i)P(X = x_i)$

Remarque. $VarX = E(X - EX)^2 = E(X^2) - (EX)^2$

Propriété. $Var(\alpha X + \beta) = \alpha^2 VarX$ Si X et Y sont indépendantes : $Var(X + Y) = VarX + VarY$

Inégalité de Bienaymé-Tchebychev. $P(|X - EX| > t) \leq \frac{VarX}{t^2}$

Posons $t = a\sigma$

$$P(|X - EX| > a\sigma) \leq \frac{VarX}{a^2\sigma^2} = \frac{1}{a^2}$$

Inégalité loin d'être optimale

4 - Retour sur les lois usuelles

Loi de Bernouilli : $X \sim \mathcal{B}(p)$

$$\begin{array}{rcl} F_X(x) & = & P(X \leq x) \\ & = & \left\{ \begin{array}{lll} 0 & \text{si} & x < 0 \\ P & \text{si} & 0 \leq x < 1 \\ 1 & \text{si} & x \geq 1 \end{array} \right. \\ & = & p1_{|[0,1[}(x) + 1_{|[1,+\infty[}(x) \end{array}$$

Remarque. $1_{|A}(x) = 1$ si $x \in A$, 0 sinon.

$$EX = 1 \times P(X = 1) + 0 \times P(X = 0) = P$$

$$\begin{array}{rcl} VarX & = & E[(X - EX)^2] \\ & = & (1-p)^2 P(X = 1) + (0 - p)^2 P(X = 0) \\ & = & (1-p)^2 p + p^2(1-p) \\ & = & p - p^2 \\ & = & p(1-p) \end{array}$$

Loi Binomiale : $X \sim \mathcal{B}(n, p)$

$X_1, \ldots, X_n$ lois de Bernouilli de paramètre p *indépendantes*.

$$X = \sum_{i=1}^{n} X_i$$

$$EX = E\sum_{i=1}^{n} X_i = \sum_{i=1}^{n} EX_i = np \text{ (par linéarité)}$$

$$\begin{array}{rcl} VarX & = & Var \sum_{i=1}^{n} X_i \\ & = & \sum_{i=1}^{n} VarX_i = np(1-p) \end{array}$$

Théorème. Soit $X \sim \mathcal{B}(n, p)$ et $V \sim \mathcal{B}(m, p)$ indépendantes :

$$X+Y \sim \mathcal{B}(n+m,p)$$

Loi géométrique : $X \sim \mathcal{G}(p)$

$$P(X=k)=p(1-p)^{k-1}$$

$$EX=\frac{1}{p}$$

$$VarX=\frac{1-p}{p}$$

$$F_x(k)=1-(1-p)^k$$

Loi de Poisson : $X \sim \mathcal{P}(\lambda)$

$$P(X=k)=e^{-\lambda}\frac{\lambda^k}{k!}, k \geq 0$$

$$\begin{aligned} EX &= \sum_{k\geq 0} k e^{-\lambda}\tfrac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k\geq 1}\tfrac{k\lambda^k}{k!} \end{aligned}$$

$$e^{\lambda}=\sum_{k\geq 0}\frac{\lambda^k}{k!}$$

$$\begin{aligned} \tfrac{\partial e^{\lambda}}{\partial\lambda} &= \sum_{k\geq 0}\tfrac{\partial\lambda(k/k!)}{\partial\lambda} \\ &= \sum_{k\geq 0} k\tfrac{\lambda^{k-1}}{k!} \end{aligned}$$

$$e^{\lambda}=\sum_{k\geq 1} k\frac{\lambda^{k-1}}{k!}$$

$$\begin{aligned} \lambda e^{\lambda} &= \sum_{k\geq 1} k\tfrac{\lambda^{k-1}}{k!} \\ &= \sum k\tfrac{\lambda\lambda^{k-1}}{k!} \\ &= \sum \tfrac{k\lambda^k}{k!} \end{aligned}$$

$$EX=e^{-\lambda}\times\lambda e^{\lambda}=\lambda$$

Tous calculs faits

$$VarX=\lambda$$

En pratique, une loi discrète dont l'espérance est très proche de sa variance ($EX=VarX=\lambda$) se modélise par une loi de Poisson.

Théorème. Si $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes :

$$X + Y \sim \mathcal{P}(\lambda + \mu)$$

Rappel. $X \sim \mathcal{B}(n,p)$, $EX = np$, $VarX = np(1-p)$

Si p est petit (n grand) alors $(1-p) \approx 1$. $EX \approx VarX$

D'où le théorème suivant :

Théorème. Si n est grand (en pratique $n \geq 50$) et p petit (en pratique $p < 0,1$) alors :

$$\mathcal{B}(n,p) \simeq \mathcal{P}(np)$$

> *Note.* On considère une loi $\mathcal{B}(n,p)$ avec p grand ($p \geq 0,1$), on peut plutôt compter les échecs par une loi Binomiale $\mathcal{B}(n, 1-p)$ avec $1-p$ petit. On utilise alors le théorème précédent.

5 - Variables (absolument) continue

Une varibale X est dite à densité s'il existe une fonction f_x telle que

$$P(a \leq X \leq b) = \int_a^b f_x(x)dx \forall a, b \in \mathbb{R}$$

- $P(X = a) = 0$
- $f_X \geq 0$
- $P(\Omega) = 1 = \int_{-\infty}^{+\infty} f_X(x)\, dx$

Fonction de répartition

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_x(x)dx$$

C'est la primitive de f_x qui a une limite nulle en $-\infty$.

$$(F_X(x))' = f_X(x)$$

Espérance et variance

$$E\ X = \int_{-\infty}^{+\infty} f_X(x)\, dx$$

> Si X discrète : $E\ X = \sum x_i P(X = x)$

Théorème. Transfert. Soit φ continue bornée,

$$E\ \phi(X) = \int_{-\infty}^{+\infty} \varphi(x) f_x(x) dx$$

$$Var\ X = E[(X - EX)^2]$$

$$\sigma_x = \sqrt{VarX}$$

Toutes les propriétés énoncées au paragraphe précédent restent vraies dans le cadre des variables continues.

1) Variable uniforme sur $[a,b]$: $X \sim \mathcal{U}([a,b])$

Tout intervalle I de longueur l totalement inclus dans $[a,b]$ à la même probabilité.

$$P(X \in I) = \frac{l}{b-a}$$

$b-a$ longueur de $[a,b]$ et l longueur de I.

Fonction de répartition

$$F_X(x) = P(X \le x) = \begin{cases} 0 & \text{si} \quad x \le a \\ P(a \le X \le x) = \frac{x-a}{b-a} & \text{si} \quad a \le x \le b \\ 1 & \text{si} \quad b \le x \end{cases}$$

$$f_X(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$$

Espérance et variance

$$\begin{aligned} E\ X &= \int_a^b x f_X(x) dx \\ &= \int_a^b \tfrac{x}{b-a} dx \\ &= \left[\tfrac{x^2}{2(b-a)}\right]_a^b \\ &= \tfrac{a+b}{2} \end{aligned}$$

$$\begin{aligned} Var\ X &= E(X^2) - E(X)^2 \\ &= \int x^2 f_X(x)dx - (\tfrac{a+b}{2})^2 \\ &= \int_a^b \tfrac{x^2}{b-a} dx - (\tfrac{a+b}{2})^2 \\ &= [\tfrac{x^3}{3(b-a)}]_a^b - (\tfrac{a+b}{2})^2 \\ &= \tfrac{(b-a)^2}{12} \end{aligned}$$

Remarque. La loi uniforme correspond à tous les générateurs “random” des outils numériques

2) Loi exponentielle ! $T \approx \varepsilon(\lambda)$

T est le temps avant défaillance (électronique / atomes ...) sur des dispositifs sans veillisement.

$$f_T(t) = \lambda e^{-\lambda t}\mathbb{1}_{\mathbb{R}+}(t)$$

Fonction de répartition

Si $t \ge 0$

$$\begin{aligned} F_T(t) &= P(T \le t) \\ &= \int_{-\infty}^t f_T(t)\ dt \\ &= \int_0^t \lambda e^{-\lambda t}\ dt \\ &= \big[-e^{-\lambda t}\big]_0^t \\ &= 1 - e^{-\lambda t} \end{aligned}$$

Probabilité de défaillance avant $t \ge 0$.

Comme on peut assimiler proba et fréquence sur un échantillon, $F_T(t)$ est aussi la proportion d'appareils défaillants avant T.

Fonction de survie

$$P(T \geq t) = 1 - F_T(t) = e^{-\lambda t}$$

Temps de demi-vie $\mathscr{C}$

$$P(R \leq \mathscr{C}) = 0.5$$

$$F_T\mathscr{C} = 1 - e^{-\lambda\mathscr{C}} = 0.5$$

$$e^{-\lambda\mathscr{C}} = 1/2 \Leftrightarrow -\lambda\mathscr{C} = \ln(1/2) \Leftrightarrow \mathscr{C} = \frac{\ln(2)}{\lambda}$$

Espérance et variance

$$EX = \frac{1}{\lambda}$$

$$Var\ X = \frac{1}{\lambda^2}$$

Loi sans vieillissement

$$\begin{array}{rcl} P_{T\geq t}(T \geq t+S) & = & \frac{P(\{T\geq t+S\}\cap\{T\geq t\})}{P(T\geq t)} \\ & = & \frac{P(T\geq t+S)}{P(T\geq t)} \\ & = & \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} \\ & = & e^{-\lambda S} \\ & = & P(T \geq S) \end{array}$$

Théorème. Si $T_1 \approx \varepsilon(\lambda)$ et $T_2 \approx \varepsilon(\mu)$ avec T_1 et T_2 indépendantes, $\boxed{\min(T_1 + T_2) \approx \varepsilon(\lambda + \mu)}$

3) Lois normales

Loi centrée réduite

Densité

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2} \text{ (peu utile)}$$

En pratique, on utilise une table de la fonction de répartition.

Théorème. Soit $a, b \in \mathbb{R}$, si X suit une loi normale, $aX + b$ aussi.

$$X \sim \mathcal{N}(m, \sigma)$$

$$aX + b \sim \mathcal{N}(am + b, a\sigma)$$

Loi normale $\mathcal{N}(m, \sigma)$

$$\mathcal{N}(m, \sigma) = \sigma \mathcal{N}(0,1) + m$$

Si $X \sim \mathcal{N}(m, \sigma)$

$$\frac{X - m}{\sigma} \sim \mathcal{N}(0,1)$$

Densité (anecdotique)

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

Théorème. La somme de plusieurs lois normales est une loi normale.

Si $X_1 \sim \mathcal{N}(m, \sigma)$ et $X_2 \sim \mathcal{N}(m', \sigma')$ indépendantes, alors $X_1 + X_2 \sim \mathcal{N}(m + m', \sqrt{\sigma^2 + \sigma'^2})$

Remarque. Si X_1 et X_2 indépendantes :$Var(X_1 + X_2) = Var\ X_1 + Var\ X_2$

$$P(m - \sigma \leq X \leq m + \sigma) \approx 0.68$$

$$P(m - 2\sigma \leq X \leq m + 2\sigma) \approx 0.95$$

$$P(m - 3\sigma \leq X \leq m + 3\sigma) \approx 0.995$$

À comparer avec l'inégalité de Bienaymé-Tchebichev : on voit que BT est loin d'être optimal.

6 - Théorèmes limites

Théorème. Loi des grands nombres. Soit $X_1, \dots, X_n$ identiquement distribués et indépendantes (IDD) :

$$\frac{X_1 + \cdots + X_n}{n} \to EX_1$$

(base des méthodes de Monte-Carlo)

Théorème Central limite (TCL).

$X_1, \dots, X_n$ variables IDD.

$$EX_i = \mu$$

$$\sigma_{X_i} = \sigma$$

$$\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow[n \to \infty]{} \mathcal{N}(0,1)$$

Une loi normale résulte de la conséquence d'un grand nombre d'évènements indépendants (nature, manufacture).

Propriété. On a vu que la loi binomiale $\mathscr{B}(n,p)$ est une somme de lois de Bernouilli IID. D'après le TCL : "$\mathscr{B}(n,p) \xrightarrow[n\to\infty]{} \mathcal{N}$". Pour n grand ($n \geq 50$) et p pas trop extrême ($0,4 \leq p \leq 0,6$ ou vérifier que $npq \geq 18$)

$$\mathscr{B}(n,p) \approx \mathcal{N}(np, \sqrt{np(1-p)})$$

$$X \sim \mathscr{B}(n,p)$$

$$Y \sim \mathcal{N}(np, \sqrt{np(1-p)})$$

$$X \approx Y$$

$P(X = 8) \neq 0$ mais $P(Y = 8) + 0$. Par contre, vrai en terme de fonctions de répartitions :

$$P(X \leq 8) \approx P(Y \leq 8)$$

On applique une "correction de continuité"

$$P(X = 8) = P(X \leq 8.5) - P(X \leq 7.5) \approx P(Y \leq 8.5) - P(Y \leq 7.5) \approx P(7.5 \leq Y \leq 8.5)$$