Eléments de statistiques pour les data sciences

Cours 2 : probabilités discrètes

Dr. Ph. Müllhaupt

IGM - EPFL

—

Plan

l'espace $(\Omega, \mathcal{F}, P)$

- Ω est la collection de tous les résultats possibles ω de l'expérience. C'est un ensemble.
- $\mathcal{F}$ est la famille des sous-ensembles de Ω. C'est l'ensemble des évènements $A \subseteq \Omega$.
- P est une fonction de $\mathcal{F}$ dans l'intervalle $[0, 1]$ qui à tout élément A de $\mathcal{F}$ assigne la probabilité $P(A)$.

la logique des évènements

- Si un ensemble A est contenu dans B, noté $A \subset B$, cela exprime que l'évènement A implique l'évènement B.
- Deux évènements A et B sont incompatibles lorsque $A \cap B = \emptyset$.
- Soit une famille d'évènements $A_1, \ldots, A_k$, l'égalité $\Omega = \cup_{i=1}^{k} A_i$ signifie qu'au moins un évènement est réalisé.
- Lorsque qu'un évènement A n'est pas réalisé, cela signifie que l'évènement complémentaire $\bar{A}$ est réalisé.
- $\emptyset$ représente un évènement impossible.
- Ω représente un évènement certain.

axiomes concernant les évènements

(i) $\Omega \in \mathcal{F}$

(ii) Si $A \in \mathcal{F}$ alors $\bar{A} \in \mathcal{F}$

(iii) Si la séquence $(A_n, n \geq 1)$ a tout ses membres qui appartiennent à $\mathcal{F}$, alors l'union $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

axiomes concernant la fonction probabilité P

I. $P(\Omega) = 1$

II. Pour toute séquence $(A_n, n \geq 1)$ d'évènements disjoints de $\mathcal{F}$

$$\boxed{A_i \bigcap A_j = \emptyset \qquad \forall i \neq j}$$

la propriété d'additivité (appelée σ addititivité) est vraie :

$$\boxed{P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)}$$

propriétés découlant des axiomes

1. $P(\bar{A}) = 1 - P(A)$
2. $P(\emptyset) = 0 \leq P(A) \leq 1 = P(\Omega)$
3. $A \subset B \Rightarrow P(B - A) = P(B) - P(A)$
4. $A \subset B \Rightarrow P(A) \leq P(B)$
5. $$P\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} P(A_n)$$

exemple 1 : points dans un carré

Un point (x, y) dans un carré continu dans le plan avec $x \in [0, 1]$ $y \in [0, 1]$.

- $\Omega = [0, 1] \times [0, 1]$
- $\mathcal{F}$ tout sous ensemble de Ω qui a une surface $S(A)$ par exemple $[x_1, x_2] \times [y_1, y_2]$
- $P(A) = S(A)$

exemple 2 : deux dès

- $\Omega = \{(a, b) | a, b \in \{1, 2, 3, 4, 5, 6\}\}$
- $\mathcal{F}$ est l'ensemble des sous ensembles de Ω
- $P(A) = (1/6)^2 \times \mathrm{Card} A$

variable aléatoire

intuition et exemples

Lors de nombreuses expériences stochastiques, on est plus intéressé par des nombres associés à chaque résultat possible que par le résultat de l'expérience elle-même.

- <u>Jet de pièce de monnaie</u>. On jette un pièce de monnaie 2 fois. On a l'ensemble de résultats possibles $\Omega = \{PP, PF, FP, FF\}$ avec F face, et P pile. Ω est le produit cartésien $\Omega = \{P, F\} \times \{P, F\}$. Si on s'intéresse au nombre de faces lors des deux jets, ceci revient à associer à chaque résultat ω un nombre naturel $X(\omega)$. Si par exemple $\omega = \{FP\}$ on a $X = 1$ et par constre si $\omega = \{PP\}$, $X = 0$ et si $\omega = \{FF\}$ on a $X = 2$. On constate que l'on partitionne l'ensemble Ω en fonction des valeurs de X possibles.
- <u>Durée de bon fonctionnement</u>. On prélève une machine et on teste au hasard son bon fonctionnement. L'ensemble Ω est alors formé de tous les dispositifs produits et $X(\omega)$ est la durée de bon fonctionnement de la machine.
- <u>Flux de véhicules</u>. On peut associer la variable X au type de véhicule, voiture, camion, moto. La variable X peut alors, par exemple, prendre les valeurs $X = 1, 2, 3$ en fonction du type de véhicule.

variables aléatoires

discrètes

Soit

- $\mathcal{X}$ un ensemble dénombrable
- $(\Omega, \mathcal{F}, P)$ un espace probabilisé

définition

Toute fonction $X : \Omega :\rightarrow \mathcal{X}$ telle que pour tout $x \in \mathcal{X}$ on ait L'ensemble

$$\{\omega | X(\omega) = x\} \in \mathcal{F}$$

est appelée une variable aléatoire discrète.

dénombrement — combinatoire

- Un grand nombre de cas en probabilité se réduisent à dénombrer les éléments d'un ensemble fini, i.e. l'ensemble des résultats de l'expérience Ω est fini.
- Tous les résultats $\omega \in \Omega$ ont la même probabilité

$$\boxed{p(\omega) = P(\{\omega\}) = \frac{1}{\text{card } \Omega}}$$

comptage — dénombrement

$$\boxed{P(A) = \frac{\text{card } A}{\text{card } \Omega}}$$

distribution de Bernouilli (expérience de Bernouilli)

L'expérience stochastique la plus simple est celle qui donne deux résultats possibles. Traditionnelement on les appellent *succès* A et *échec* $\bar{A}$. Soient

$$\begin{aligned} p &\triangleq P(A) \\ q &\triangleq 1 - p = P(\bar{A}) \end{aligned}$$

En associant à ces deux évènements les valeurs 1 et 0 on a

définition

Un variable aléatoire de Bernouilli X dont la loi de probabilité est donnée par

$$P(X) = \begin{cases} p & X = 1 \\ q = 1 - p & X = 0 \end{cases}$$

- jet de pièce de monnaie pile ou face
- tirage d'une urne avec deux boules, une boule noire et une boule blanche
- contrôle de qualité, pièce défectueuse ou pièce bonne

distribution binomiale

La loi binomiale donne la probabilité d'un nombre de réussites à un jeu de hasard avec deux résultats possibles $X = 0$ ou $X = 1$ (par exemple pile ou face dans le jet d'une pièce de monnaie). La pièce de monnaie n'est pas forcément équilibrée et la probabilité de réussite est p et la probabilté de l'échec est $q = 1 - p$. On considère la somme de plusieurs lancés $X_1 \dots X_n$

$$S_n = X_1 + \cdots + X_n$$

La probabilité que

$$\boxed{P(S_n = k) = p_k \qquad 0 \leq k \leq n}$$

est

$$\boxed{\boxed{p_k = \frac{n!}{k!(n-k)!} p^k q^{(n-k)}} \qquad 0 \leq k \leq n}$$

loi binomiale

La loi binomiale décrit également le tirage <u>avec remise</u> de deux classes de boules en nombre de N, avec n boules noires et $N - n$ boules blanches de telle sorte que $p = n/N$ et $q = 1 - p$.

$$\boxed{p_k = \binom{n}{k} p^k q^{(n-k)} = C_k^n p^k q^{(n-k)}}$$

notation

$$
\begin{aligned}
\text{arrangements : } A_n^k &= n \cdot (n-1) \cdots (n-k+1) = \frac{n\,!}{(n-k)\,!} \\
\text{permutations : } P_n &= n\,! \\
\text{combinaisons : } C_k^n &= \binom{n}{k} = \frac{n!}{k!(n-k)!}
\end{aligned}
$$

- Arrangements ordonné de k objets parmi n sans répétitions : A_n^k (<u>Attention</u> : sur la calculatrice c'est la touche P_k^n)
- Permutations de n objets : P_n
- Comptage, dénombrement de sous-ensembles ou d'éléments. Soit F un ensemble de card $(F) = n$ éléments, et soit $k \leq n$. Le nombre de sous-ensemble de F avec k éléments : C_k^n

graphiques de la loi binomiale selon p et n

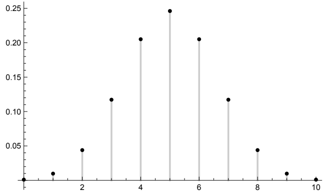

Figure – $p = 0.5, \ n = 10$

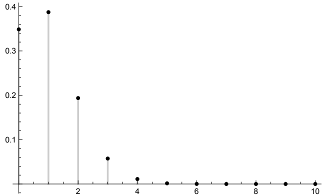

Figure – $p = 0.1, \ n = 10$

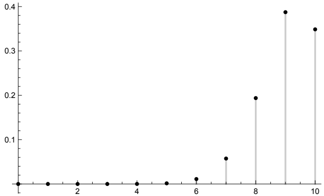

Figure – $p = 0.9, \ n = 10$

graphiques de la loi binomiale selon p et n

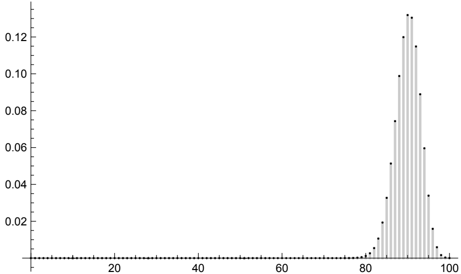

Figure – $p = 0.9, \; n = 100$

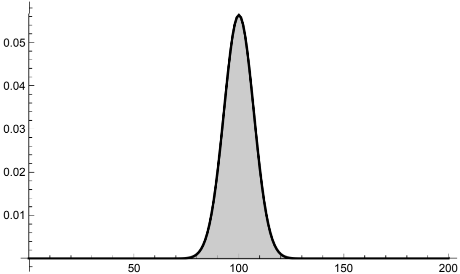

Figure – $p = 0.5, \; n = 200$

loi géométrique

loi de Pascal

Supposons l'expérience de tester un pièce en la percutant jusqu'à ce qu'elle casse. On peut donc donner une probabilité au nombre d'essais nécessaires jusqu'à rupture. On peut donc associer le 'succès' à la première destruction de la pièce. La variable de Bernouilli associée prend alors la valeur 1 avec une probabilité $0 < p < 1$. La variable aléatoire X égale alors au nombre d'expériences nécessaires jusqu'à l'apparition du premier succès :

$$\boxed{P(X = k) = (1 - p)^{k-1} p \qquad \text{où } k = 1, 2, 3, \ldots, n, \ldots}$$

on peut également mettre sous la forme

$$\boxed{P(X = n) = (1 - p)^{k} \, p \qquad \text{où } k = 0, 1, 2, \ldots, n, \ldots}$$

loi multinomiale — tirage avec remise

Cette distribution est importante car elle décrit n tirages avec remise de m possibilités exclusives au lieu de deux possibilités comme pour la loi binomiale.

évènement	A_1	A_2	$\dots$	A_m	total
probabilité	p_1	p_2	$\dots$	p_m	1
fréquence [1]	X_1	X_2	$\dots$	X_m	n

$$
\begin{aligned}
p_1 + p_2 + \cdots + p_m &= 1 \\
X_1 + X_2 + \cdots + X_m &= n
\end{aligned}
$$

1. au sens du nombre des évènements constatés

loi multinomiale — tirage avec remise

Soit p_1, p_2, ..., p_m la probabilité de chaque évènement.

$$\boxed{\left(\begin{array}{c} n \\ k_1 k_2 \ldots k_m \end{array} \right) \triangleq \frac{n!}{k_1!\, k_2! \cdots k_m!}}$$

La probabilité de tirer en tout n tirages avec remise, avec k_i le nombre de tirages de type $i = 1$ et i allant de 1 à m, est donnée par

$$\boxed{P(k_1, k_2, \ldots, k_m) = \left(\begin{array}{c} n \\ k_1 k_2 \ldots k_m \end{array} \right) p_1^{k_1} p_2^{k_2} \cdots p_m^{k_m}}$$

Remarque :

Lorsqu'un sondage est effectué avec m possibilités et que les probabilités sont connues de chaque possibilité, alors la distribution multinomiale peut-être utilisée comme un modèle à condition que la population soit grande vis-à-vis de n, de telle sorte que n tirages n'affectent pas la probabilité de chaque évènement $p_i = 1, 2, \ldots, m$.

loi hypergéométrique — tirage sans remise

Lorsqu'on effectue un tirage de n boules sans remise d'une urne contenant des boules blanches ou noires (N en tout et r boules blanches), les n expériences partielles dont est composé ce processus, à savoir les prélèvements des n boules, ne sont ni identiques, ni indépendantes l'une de l'autre. En effet, la probabilité de prélever une boule blanche varie constamment au cours du tirage et dépend des résultats déjà obtenus.

Théorème

La probabilité d'avoir exactement k boules blanches parmi les n boules tirées, d'une urne qui contient r boules blanches et $N-r$ boules noires, est donnée par

$$\boxed{P(A_k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} \qquad k = 0, 1, 2, \ldots, \min(r, n)}$$

loi hypergéométrique — tirage sans remise

- L'ensemble des résultats possibles est une famille de tous les sous-ensembles ω de n boules parmi les N boules en tout card $\Omega = \begin{pmatrix} N \\ n \end{pmatrix}$.
- Il faut dénombrer les ensembles de n boules qui contient k boules blanches.
- On commence par dénombrer le nombre de choix possibles de k blanches parmi les r possibles $\begin{pmatrix} r \\ k \end{pmatrix}$.
- On dénombre le nombre de $n-k$ boules noires du tirage parmi les $N-r$ boules noires possibles $\begin{pmatrix} N-r \\ n-k \end{pmatrix}$ ce qui multiplie les possibilités.
- card $(A_k) = \begin{pmatrix} r \\ k \end{pmatrix} \begin{pmatrix} N-r \\ n-k \end{pmatrix}$

$$\boxed{P(A_k) = \frac{\text{card } (A_k)}{\text{card } \Omega}}$$

définition

Une variable aléatoire X qui prend ses valeurs dans l'ensemble dénombrable $X = \mathbb{N}$ et qui admet une distribution

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

où λ est un nombre réel non négatif est appélée une variable aléatoire de Poisson de paramètre λ. Ceci est noté

$$X \sim \mathscr{P}ois(\lambda)$$

Par convention, on rappel que $0! = 1$. La loi de Poisson est utilisée pour modéliser le nombre d'évènements qui se produisent dans un intervalle de temps ou d'espace donné, en supposant que ces <u>évènements sont indépendants</u> et surviennent à un <u>taux moyen constant</u>.

loi de Poisson

domaines d'application

Théorie des files d'attente et télécommunications

- Modélisation du nombre d'appels reçus par un centre d'appels par minute.
- Arrivée de paquets dans un réseau informatique.

Fiabilité et analyse des risques

- Estimation du nombre de pannes d'un système sur une période donnée.
- Nombre d'accidents de la route à un carrefour par an.

Biologie et épidémiologie

- Nombre de mutations dans une séquence d'ADN.
- Nombre de cas d'une maladie dans une région pendant une épidémie.

loi de Poisson

domaines d'application (suite)

Finance et assurance

- Modélisation du nombre de défauts de paiement dans un portefeuille de prêts.
- Prévision du nombre de réclamations d'assurance par mois.

Astronomie et physique

- Comptage du nombre de photons détectés par un télescope.
- Désintégrations radioactives d'un élément instable.

Sport et journalisme

- Nombre de buts marqués dans un match de football.
- Nombre de fautes d'impression dans un livre.

espérance mathématique

définition

Soit X une variable aléatoire prenant des valeurs dans $\mathcal{X}$ avec une distribution $p(x), x \in \mathcal{X}$ et supposons une fonction

$$f : \mathcal{X} \to \mathbb{R}$$

et supposons de plus que

$$\sum_{x \in \mathcal{X}} |f(x)| \, p(x) < \infty$$

définition :
On définit alors l'espérance mathématique de $f(X)$, notée $\mathbb{E}[f(X)]$, comme

$$\boxed{\mathbb{E}[f(X)] \triangleq \sum_{x \in \mathcal{X}} f(x)p(x)}$$

espérance mathématique

exemple

Soit X une variable aléatoire de Poisson de paramètre λ et soit la fonction f qui est l'identité $f : x \to x$ de $\mathbb{N}$ dans $\mathbb{R}$ d'une part, et, d'autre part $f : x \to x^2$. On a

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = \lambda \\ \mathbb{E}[X^2] &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda + \lambda^2\end{aligned}$$

Démonstration : exercice

moyenne et variance

définition

définition

La moyenne, notée μ, est définie par l'espérance mathématique correspondant à la fonction f qui est l'identité

$$\mu \triangleq \mathbb{E}[X]$$

définition

La variance, notée var ou mieux σ^2 est définie par

$$\sigma^2 \triangleq \mathbb{E}[(X - \mu)^2]$$

moyenne et variance

remarques

- La moyenne permet de mesurer le centrage de la distribution. (Analogie mécanique : centre de masse.)
- La variance permet de quantifié l'ecartement de la distribution par rapport à la moyenne. (Analogie mécanique : inertie autour du centre de masse.)

théorème, propriété

$$\sigma^2 = \mathbb{E}[(X-\mu)^2] = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Analogie : Thérème de Steiner : $I_G = I_O - M\|OG\|^2$. Comme il s'agit d'une probabilité $M = 1$ dans cette analogie.

moyenne et variance

tableau récapitulatif de quelques distributions discrètes

distribution	μ	σ^2
uniforme	$(n+1)/2$	$(n^2-1)/12$
de Bernouilli $\mathscr{B}er(1,p)$	p	$p(1-p)$
géométrique	$1/p$	$(1-p)/p^2$
binomiale $\mathscr{B}in(n,p)$	$n\,p$	$n\,p\,(1-p)$
de Poisson $\mathscr{P}ois$	λ	λ

probabilité conditionnelle

Soit B un évènement de probabilité strictement positive.

définition

Quel que soit l'évènement A, on définit la probabilité conditionnelle notée $P(A|B)$ par la formule

$$\boxed{P(A|B) \triangleq \frac{P(A \cap B)}{P(B)}}$$

La quantité $P(A|B)$ est appelée la probabilité de A sachant B (étant donné B).

probabilité conditionnelle

interprétation

Supposons un grand nombre d'expériences n a été effectué et les résultats des évènements A et B ont été enregistrés sous la forme suivante : n_A expériences parmi les n expériences correspondent à l'évènement A, et n_B expériences correspondent à l'évènement B. On note $n_{A\cap B}$ le nombre d'expériences qui correspondent aux évènements qui ont lieu simultanément A et B. L'interprétation fréquentiste de la probabilité suggère que si les expériences ont été réalisées de manière indépendante, la fréquence $\frac{n_A}{n}$ sera proche de la probabilité $P(A)$ et la fréquence $\frac{n_B}{n}$ sera proche de la probabilité $P(B)$. Maintenant,

$$\frac{P(A\cap B)}{P(B)} \approx \frac{\frac{n_{A\cap B}}{n}}{\frac{n_B}{n}} = \frac{n_{A\cap B}}{n_B} \approx P(A|B)$$

Ceci mesure l'expectative d'avoir l'évènement A se produire lorsque l'évènement B est observé.

indépendance des évènements

considération générale

Supposons deux évènements, le nom d'un personne commence par la lettre 'E' (évènement A) et la couleur des cheveux de la personne est noire (évènement B). Soit n la population totale et n_A et n_B le nombre de personnes avec la propriété respective. On s'attend à ce qu'il y ait une proportion identique de personnes avec les cheveux noirs dans la population complète que ceux qui ont les cheveux noirs et dont le nom commence par 'E'. Autrement dit

$$\frac{n_B}{n} = \frac{n_{A\cap B}}{n_A} = \frac{\frac{n_{A\cap B}}{n}}{\frac{n_A}{n}}$$

ce qui peut s'écrire sous la forme en supposant un grand nombre d'individus

$$P(B) = P(B|A) = \frac{P(B\cap A)}{P(A)}$$

indépendance des évènements

définition

définition

Deux évènements A et B sont dits indépendants si, et seulement si,

$$\boxed{P(A \cap B) = P(A)\, P(B)}$$

trucs & astuces

Théorème

$0 < a < 1$

$$\sum_{k=1}^{\infty} k a^{k-1} = \frac{1}{(1-a)^2}$$

$$\sum_{k=1}^{\infty} k^2 a^{k-1} = \frac{(1+a)}{(1-a)^3}$$

trucs & astuces

Démonstration

$$
\begin{aligned}
\sum_{k=1}^{\infty} k\, a^{k-1} &= \frac{d}{da}\sum_{k=0}^{\infty} a_k = \frac{d}{da}\left(\frac{1}{1-a}\right) = \frac{1}{(1-a)^2} \\
\sum_{k=1}^{\infty} k^2 a^{k-1} &= a\sum_{k=1}^{\infty} k(k-1)\, a^{k-2} + \sum_{k=1}^{\infty} k\, a^{k-1} \\
&= a\frac{d^2}{da^2}\sum_{k=0}^{\infty} a^k + \frac{1}{(1-a)^2} = \frac{1+a}{(1-a)^3}
\end{aligned}
$$