Université de Liège
2ème Bloc Ingénieur civil
Année 2019-2020

MATH0062-1 - Eléments du calcul des probabilités

Résumé du cours

Basé sur les slides du cours d'éléments du calcul des probabilités
Professeur : Pierre Sacré

DEHARENG Simon

LIÈGE
Mai 2020

Table des matières

Chapitre 1

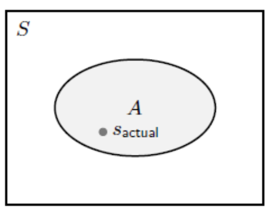
Probability and counting

1 Introduction - Pourquoi étudier les probabilités ?

Les probabilités sont étudiées dans plusieurs domaines comme les statistiques, la physique (pour la mécanique quantique, par exemple), en biologie pour la génétique, en sciences des données avec l'intelligence artifielle et le "machine learning" ou encore pour décrire la météo, les stocks financiers et les essais cliniques.
La vie est incertaine et les probabilités régissent cette logique.

2 Echantillon spatial et événement

l' **échantillon spatial** (ou *sample space* en anglais) S d'une expérience est l'ensemble de tous les résultats possibles de l'expérience. Un événement A est un sous-ensemble de l'échantillon S. On peut dire que A s'est réalisé si le résultat réel est dans A.



- sample space S
- event A
- event A occurred if $s_{\text{actual}} \in A$

3 Définition naïve des probabilités

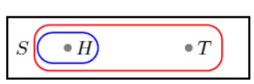
De manière naïve et la plus générale possible, une probabilité s'exprime mathématiquement comme,

$$P_{\text{naive}}(A) = \frac{\text{nombre de possibilités favorables à A}}{\text{nombre total de possibilités}} = \frac{|A|}{|S|}$$

Où $|A|$ correspond au **cardinal** de A

example: coin flip

- experiment: flip a coin once
- sample space: $S = \{H, T\}$

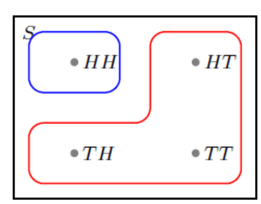


- $P(H) = \frac{1}{2}$, $P(H \text{ or } T) = 1$, $P(H \text{ and } T) = 0$, $P(\text{neither } H \text{ nor } T) = 0$

example: coin flips

- experiment: flip a coin twice
- sample space: $S = \{HH, HT, TH, TT\}$

- $P(HH) = \frac{1}{4}$, $P(\text{at least one } T) = \frac{3}{4}$



3.1 Hypothèses fortes

- Tous les résultats sont équiprobables,
- Espace d'échantillonage fini.

3.2 fausses applications de la définition naïve

Cette définition est souvent appliquée à tort par des gens pensant que les événements sont tous équiprobables et se basant sur la conclusion suivante : *" Soit cela se passe, soit cela ne se passe pas, donc la probabilité est de 50/50 pour chaque événement "*. Or ceci est totalement faux pour la plupart des cas.

example: life on Mars

- $P(\text{life on Mars}) = 1/2$ — S life • • no life
- $P(\textit{intelligent}\text{ life on Mars}) = 1/2$ — S int. life • • no int. life

intuitively, we should have $P(\textit{intelligent}\text{ life on Mars}) < P(\text{life on Mars})$

Ce qui nous mène alors à la section suivante.

4 Comment compter les probabilités ?

4.1 Théorème de multiplication

Considérons une expérience constituée de 2 sous-expériences, la A et la B. Supposons que l'expérience A possède a résultats possibles et l'expérience B, b résultats possibles. L'expérience est alors composée de $(a \cdot b)$ résultats possibles.

4.2 Ajustement au comptage excessif (overcounting)

Pour tout intégrants non-négatifs k et n, le coefficient binômial $\binom{n}{k}$, lu "k parmi n", est le nombre de sous-ensembles de taille k pour un ensemble de taille n.

Formule des coefficients binômiaux

Pour $k \leq n$, on a

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

Pour $k > n$, on a $\binom{n}{k} = 0$

Exemple. Full house en poker
Problème : Une main de 5 cartes est distribuée à partir d'un deck de 52 cartes mélangées. Trouver la probabilité d'obtenir un full (ex : trois 7 et deux 10).
Solution :

$$P_{\text{full house}} = \frac{13\binom{4}{3} \cdot 12\binom{4}{2}}{\binom{52}{5}} = \frac{3744}{2598960} \approx 0.00144$$

4.3 Table d'échantillonage : Choisir k parmi n

	order matters	order doesn't matter
replace	n^k	$\binom{n+k-1}{k}$
don't replace	$n(n-1)\cdots(n-k+1)$ [0 for $k>n$]	$\binom{n}{k}$ [0 for $k>n$]

Voir exemples de Bose-Einstein et le problème d'anniversaire

4.4 Définition générale des probabilités

Un *espace de probabilité* consiste en un espace d'échantillonage S et en une *fonction de probabilité* P qui prend en entrée un événement $A \subseteq S$ et retourne $P(A) \in [0,1]$. La fonction P doit sastisfaire les axiomes suivants :

1. $P(\emptyset) = 0, P(S) = 1$;
2. if $A_1, A_2, \ldots$ sont des événements disjoints, alors

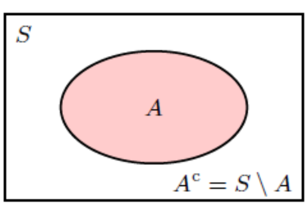
$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P\left(A_j\right)$$

2 événements sont disjoints si ils sont mutuellement exclusifs : $A_i \cap A_j = \emptyset$ for $i \neq j$.

5 Propriétés des probabilités

5.1 Théorème : $P(A^c) = 1 - P(A)$

intuition:



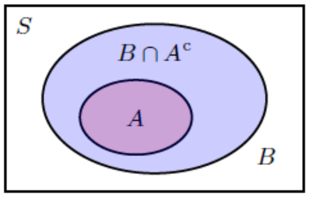
proof:

$$1 = P(S) = P(A \cup A^c) = P(A) + P(A^c),$$

since $A \cap A^c = \emptyset$.

5.2 Théorème : $A \subseteq B \Longrightarrow P(A) \leq P(B)$

intuition:

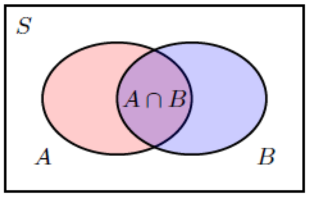


proof:

$$P(B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c) \geq P(A).$$

5.3 Inclusion/Exclusion pour 2 événements : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

intuition:

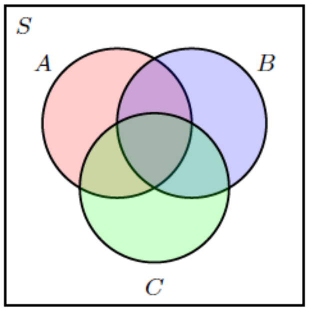


proof:

$$\begin{aligned} P(A \cup B) = P(A \cup (B \cap A^c)) &= P(A) + P(B \cap A^c) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

since $A \cap B$ and $B \cap A^c$ are disjoint and their union is B.

5.4 Inclusion/Exclusion pour 3 événements



$$\begin{aligned} P(A \cup B \cup C) = \ & P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C). \end{aligned}$$

5.5 Inclusion/Exclusion pour n événements

$$P(\textstyle\bigcup_{i=1}^n A_i) = \sum_i P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) + \sum_{i<j<k} P\left(A_i \cap A_j \cap A_k\right)$$
$$- \ldots + (-1)^{n+1} P(A_1 \cap \cdots \cap A_n)$$

Chapitre 2

Probabilités conditionnelles et règle de Bayes

1 Définition

Si A et B sont des événements avec $P(B) > 0$, alors la probabilité d'avoir A sous la condition de B est donnée par

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

$\Longrightarrow$ Exemple : 2 cartes sont tirées aléatoirement, une à la fois et sans remise. Soit A l'événement où la première carte est un coeur et B l'événement où la seconde est rouge. On a que,

$$P(A \cap B) = \frac{13 \cdot 25}{52 \cdot 51} = \frac{25}{204}$$

et

$$P(A) = \frac{1}{4} \quad ; \quad P(B) = \frac{26 \cdot 51}{52 \cdot 51} = \frac{1}{2}$$

Nous avons alors,

$$P(B|A) = \tfrac{P(A \cap B)}{P(A)} = \tfrac{25/204}{1/4} = \tfrac{25}{51}$$

$$P(A|B) = \tfrac{P(A \cap B)}{P(B)} = \tfrac{25/204}{1/2} = \tfrac{25}{102}$$

1.1 Propriétés des probabilités conditionnelles

1. Pour tous événements A et B probables,

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

2. Pour tous événements $A_1, A_2, ..., A_n$ probables,

$$P(A_1, A_2, ..., A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)...P(A_n|A_1, ..., A_{n-1})$$

2 Loi de Bayes et loi de probabilité totale

Loi de Bayes

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

Loi de probabilité totale (LOTP)

Soit $A_1, A_2, ..., A_n$ une partition de S (càd que A_i sont disjoints et leur union est S), avec $P(A_i) > 0$ pour tout i, alors

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

3 Indépendance des événements

3.1 Indépendance de 2 événements

2 événements A et B sont indépendants si

$$P(A \cap B) = P(A)P(B)$$

Si $P(A) > 0$ et $P(B) > 0$, alors ceci est équivalent à

$$P(A|B) = P(A)$$

NB : Indépendance ≠ disjonction. Si A et B sont disjoints, alors $P(A \cap B) = 0$, donc des événements disjoints ne peuvent être indépendants que si $P(A) = 0$ ou si $P(B) = 0$.

3.2 Indépendance de 3 événements

Les événements A, B et C sont dits indépendants si les équations suivantes sont respectées,

$$P(A \cap B) = P(A)P(B)$$
$$P(A \cap C) = P(A)P(C)$$
$$P(B \cap C) = P(B)P(C)$$
$$P(A \cap B \cap C) = P(A)F(B)P(C)$$

Dans le cas où les 3 premières conditions sont respectées, A, B et C sont dits *indépendants paire à paire*.
NB : Encore une fois, l'indépendance paire à paire ne signifie pas indépendance. En effet,

consider two fair, independent coin tosses.

- A: event that the first is Heads ($P(A) = 1/2$)
- B: event that the second is Heads ($P(B) = 1/2$)
- C: event that both tosses have the same result ($P(C) = 1/2$)

then A, B, and C are pairwise independent but not independent, since

$$\frac{1}{4} = P(A \cap B \cap C) \neq P(A)\,P(B)\,P(C) = \frac{1}{8}$$

just knowing about A or just knowing about B tells us nothing about C,
but knowing what happened with *both* A and B gives us information about C
(in fact, in this case it gives us perfect information about C).

3.3 Indépendance de plusieurs événements

Pour que n événements $A_1, A_2, ..., A_n$ soient indépendants, chaque paire doit statisfaire $P(A_i \cap A_j) = P(A_i)P(A_j)$ $(i \neq j)$, chaque triplet doit satisfaire $P(A_i \cap A_j \cap A_k) = P(A_i)P(A_j)P(A_k)$ $(i \neq j \neq k)$ et ainsi de suite pour les quadruplets, quintuplets, etc...

4 Indépendance conditionnelle

Les événements A et B sont dits *conditionnellement indépendants* si pour tout événement E, $P(A \cap B|E) = P(A|E)P(B|E)$
NB :

- 2 événements peuvent être conditionnellement indépendants pour E, mais pas indépendants pour E^c.
- 2 événements peuvent être conditionnellement indépendants pour E, mais pas indépendants.
- 2 événements peuvent être indépendants, mais pas conditionnellement indépendants pour E.
- Conditionnellement indépendant n'implique pas l'indépendance et inversement.

NB : Voir l'exemple de Monty Hall et le paradoxe de Simpson.

Chapitre 3

Variables discrètes aléatoires

1 Variables aléatoires (r.v.s)

Une variable alétoire est une fonction réelle de l'espace d'échantillonage S.
Par comparaison,

- **Variable aléatoire discrète :** Peut prendre n'importe quelle valeur entière dans une intervalle donné.
- **Variable aléatoire continue :** Peut prendre n'importe quelle valeur réelle dans un intervalle donné.

1.1 Distribution de Bernouilli

Une variable aléatoire X suit une distribution de Bernouilli avec le paramètre p si $P(X=1)=p$ et $P(X=0)=1-p$, où $0<p<1$. Celle-ci s'écrit

$$X \sim \text{Bern}(p)$$

Où " $\sim$ " se lit "est distribué comme"

- Toute variable alétoire dont les valeurs possibles sont 0 et 1 possède une distribution Bern(p), avec p la probabilité telle que r.v. = 1.
- Le nombre p dans la distribution de Bernouilli est appelé *paramètre de distribution.*
- Il existe une famille de distributions de Bernouilli, indexée par p.
- Chaque événement possède une variable aléatoire qui lui est associée, égale à 1 si l'événement se produit et égale à 0 sinon.

On définit la variable aléatoire égale à 1 si l'événement A se produit et égale à 0 sinon comme la *variable aléatoire indicatrice*. On dénotera cette variable de A par I_A ou $I(A)$.

$$I_A \sim \text{Bern}(p)$$

avec $p = P(A)$

1.2 Distribution Binômiale

Dans le cas où $X \sim \text{Bin}(n,p)$, alors la *fonction de masse de* X est

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{pour } k = 0,1,\ldots,n$$

2 Distributions et fonctions de masse

Une variable aléatoire X est dite *discrète* si il existe une liste finie de valeurs $a_1, a_2, ..., a_n$ ou une liste infinie telle que $P(X=a_j)=1$. Si X est une variable discrète aléatoire, alors la liste de valeurs x telle que $P(X=x)>0$ est appelée *support* de X.

Attention

- Souvent, une variable aléatoire discrète est un entier.
- Une variable continue peut prendre n'importe quelle valeur réelle.
- Une variable aléatoire peut être à la fois continue et discrète.

2.1 Fonction de masse pour une variable discrète aléatoire

La *fonction de masse* (PMF = Probability Mass Function) pour une variable aléatoire discrète X est la fonction p_X donnée par

$$p_X = P(X = x)$$

Qui est positive si x est dans le support de X, et nul sinon. Ici, $X = x$ représente un événement composé de tous les résultats s auxquels X assigne le nombre x. Formellement, $\{X = x\}$ est défini comme $\{s \in S : X(s) = x\}$

2.2 Propriétés des PMF

Soit X une variable alétoire discrète. La PMF p_X de X doit satisfaire les 2 critères suivants :

- $p_X(x) > 0$ si x appartient au support de X, et $p_X(x) = 0$ sinon ;
- $\sum_x p_X(x) = 1$, la somme s'effectuant sur le support de X.

Une fois la PMF de X connue, on peut calculer la probabilité que X tombe dans un sous-ensemble donné de R en additionnant les valeurs appropriées de x.

Théorème. Si $X \sim \text{Bin}(n, p)$, $Y \sim \text{Bin}(m, p)$ et X est indépendant de Y, alors,

$$X + Y \sim \text{Bin}(n + m, p)$$

2.3 distribution hypergéométrique

Si $X \sim \text{HGeom}(w, b, n)$, alors la PMF de X est

$$P(X = k) = \frac{\begin{pmatrix} w \\ k \end{pmatrix}\begin{pmatrix} b \\ n-k \end{pmatrix}}{\begin{pmatrix} w+b \\ n \end{pmatrix}}$$

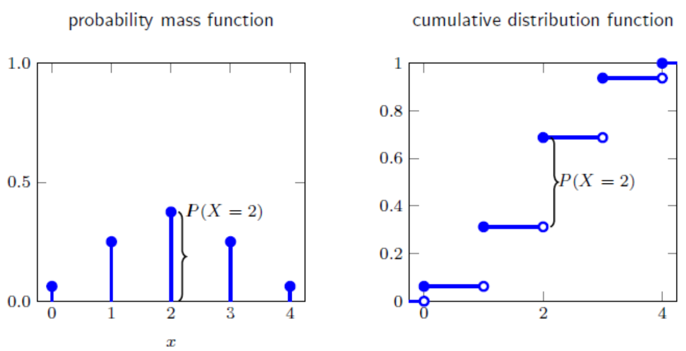
Pour des entiers k satisfaisant $0 \leq k \leq w$ et $0 \leq n - k \leq b$

3 Fonctions de distribution cumulative (CDF)

La fonction de distribution cumulative d'une variable discrète X est la fonction F_X donnée par $F_X(x) = P(X \leq x)$. Si il n'y a pas de risque d'ambiguïté, la CDF peut s'écrire simplement F.

- La CDF est définie sur les réels même si X prend des valeurs entières.
- la CDF est définie pour toutes les variables discrètes aléatoires (discrètes, continues ou hybrides).

Example: $X \sim \mathrm{Bin}(4, 1/2)$



Example: $X \sim \mathrm{Bin}(4, 1/2)$

problem: find $P(1 < X \leq 3)$ using F_X.

solution: we have by the second axiom

$$P(X \leq 3) = P(X \leq 1) + P(1 < X \leq 3)$$

thus we can write

$$P(1 < X \leq 3) = P(X \leq 3) - P(X \leq 1) = F_X(3) - F_X(1)$$

in general: we have $P(a < X \leq b) = F_X(b) - F_X(a)$ for all $a, b \in \mathbf{R}$

3.1 Propriétés des CDF

- Si $x_1 \leq x_2$, alors $F(x_1) \leq F(x_2)$
- La CDF est continue sauf où il y possiblement quelques sauts. Partout où il y a un saut, la CDF est continue à droite.

$$F(a) = \lim_{x \to a^+} F(x)$$

- Convergence à l'infini

$$\lim_{x \to -\infty} F(x) = 0 \quad \text{et} \quad \lim_{x \to \infty} F(x) = 1$$

4 Indépendance des variables aléatoires

Les variables aléatoires X et Y sont dites *indépendantes* si

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

pour tout $x, y \in \mathscr{R}$

Dans le cas discret, ceci est équivalent à la condition

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

pour tout x, y appartenant respectivement au support de X et au support de Y.

5 Espérance et attentes

Avant de commencer une expérience, il peut être intéressant de dire ce qu'il va se produire potentiellement et comment s'étend la distribution.

5.1 Espérance d'une variable aléatoire discrète

L'espérance d'une r.v.s est donnée par la formule,

$$E(X) = \sum_x xP(X = x),$$

où la somme s'effectue sur le support de X.

Example: Bernoulli

problem: let $X \sim \mathrm{Bern}(p)$. find the expected value of X.

solution:

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = p$$

5.2 Indicateur et pont fondamental

On a défini précédemment la variable aléatoire indicatrice,

$$I_A = \begin{cases} 1, & \text{si } A \text{ se réalise} \\ 0, & \text{sinon} \end{cases}$$

Théorème (pont fondamental entre probabilité et espérance). Les événements correspondent un-à-un avec les variables aléatoires discrètes et la probabilité d'un événement A est la valeur attendue de son indicateur I_A

$$P(A) = E(I_A)$$

6 Linéarité de l'espérance

Pour toute r.v.s X, Y et toute constante c,

$$E(X+Y) = E(X) + E(Y)$$
$$E(cX) = cE(X)$$

Example: Binomial expectation

problem: let $X \sim \mathrm{Bin}(n, p)$. find the expected value of X.

solution:

$$\begin{aligned} E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} &= n \sum_{k=0}^{n} \binom{n-1}{k-1} p^k q^{n-k} \qquad [k\tbinom{n}{k} = n\tbinom{n-1}{k-1}] \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{n-k} \\ &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-j-1} \qquad [j = k-1] \\ &= np \end{aligned}$$

6.1 Distribution géométrique

Si $X \sim \mathrm{Geom}(p)$, alors la PMF de X est

$$P(X = k) = q^k p, \quad \text{pour } k = 0, 1, 2, \ldots, \text{ où } q = 1 - p$$

problem: 5-card hand. $X = (\#\text{ aces})$. find the expected value of X.

solution: let I_j be indicator of jth card being an ace $(1 \le j \le 5)$.

$$\begin{aligned} E(X) &= E(I_1 + \cdots + I_5) && \text{[by indicator r.v.]} \\ &= E(I_1) + \cdots + E(I_5) && \text{[by linearity]} \\ &= 5 \cdot E(I_1) && \text{[by symmetry]} \\ &= 5 \cdot P(\text{1st card is ace}) && \text{[by fundamental bridge]} \\ &= 5 \cdot \frac{4}{52} = \frac{5}{13} \end{aligned}$$

6.2 Distribution binômiale négative (voir paradoxe de St. Petersburg

Si $X \sim \text{Nbin}(r,p)$, alors la PMF de X est

$$P(X=n) = \binom{n+r-1}{r-1} r^r q^n \quad \text{for } n = 0,1,2,\ldots, \text{ where } q = 1-p$$

problem: let $X \sim \text{NBin}(r,p)$. find $E(X)$.

solution: let X_j be the number of failures between $(j-1)$th and jth success

$$\begin{aligned} E(X) &= E(X_1 + \cdots + X_r) && [\text{as } X_j \overset{\text{i.i.d.}}{\sim} \text{Geom}(p)] \\ &= E(X_1) + \cdots + E(X_r) && \text{[by linearity]} \\ &= r \cdot \frac{q}{p} \end{aligned}$$

6.3 Distribution de Poisson

Une variable aléatoire discrète possède une distribution de Poisson avec le paramètre $\lambda > 0$ si la PMF de X est

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0,1,2,\ldots$$

On écrit $X \sim \text{Pois}(\lambda)$

Poisson paradigm (= law of rare events)

let $A_1, \ldots, A_n$ be events with $p_j = P(A_j)$, where n is large, the p_j are small, and the A_j are independent or weakly dependent. let

$$X = \sum_{j=1}^{n} I(A_j)$$

count how many of the A_j occur. then X is approximately distributed as $\text{Pois}(\lambda)$, with $\lambda = \sum_{j=1}^{n} p_j$.

Example: Poisson approximation to Binomial

problem: let $X \sim \text{Bin}(n,p)$. let $n \to \infty$, $p \to 0$, $\lambda = np$ is held constant. find what happens to $P(X=k)$.

solution:

$$\begin{aligned} P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\ &= \frac{\lambda^k}{k!} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k} \\ &\to \frac{e^{-\lambda}\lambda^k}{k!} \qquad \text{as } \frac{n!/k!}{n^k} \to 1,\ \left(1-\frac{\lambda}{n}\right)^n \to e^{-\lambda},\ \left(1-\frac{\lambda}{n}\right)^{-k} \to 1 \end{aligned}$$

Chapitre 4

Variables aléatoires continues

Une variable aléatoire possède une *distribution continue* si sa CDF est différentiable. Il est permis que ses extrémités soient non continues, tant que la CDF est continue partout ailleurs. Une variable aléatoire continue est donc une variable aléatoire distribuée de manière continue.
Pour une v.a. continue, $P(X = x) = 0$, ∀x. Vu que sa PMF est partout égale à 0, on travaille avec une fonction de densité de probabilité.

1 Fonctions de densité de probabilité

Une probabilité est une quantité adimensionnelle, l'unité d'une densité de probabilité $f(x)$ correspond donc à (unité de X)$^{-1}$.

1.1 Définition

Pour une v.a. continue X de CDF F, la *densité de probabilité* (PDF) de X est la dérivée f de la CDF,

$$f(x) = F'(x)$$

Le *support de X*, est l'ensemble des x tels que $f(x) > 0$.

⚠ $f(x)$ n'est **pas** une probabilité,

⚠ Il est possible d'avoir $f(x) > 1$ pour certains x.

1.2 Théorèmes

PDF → CDF. Soit X une v.a continue avec une PDF f,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

Par le théorème fondamental du calcul intégral,

$$P(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)dx$$

Validité de la PDF. La PDF f d'une v.a. continue doit satisfaire les critères suivants,

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

1.3 Espérance d'une variable aléatoire continue

L'espérance d'une v.a. continue avec une PDF f est donnée par,

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

2 Variance

2.1 Définition

La variance d'une v.a. X est donnée par,

$$\text{Var}(X) = E(X - E(X))^2 = E(X)^2 - (E(X))^2$$

On appelle *déviation standard* (SD),

$$SD(X) = \sqrt{\text{Var}(X)}$$

2.2 Propriétés de la variance

1. Contrairement à l'espérance, la variance **n'est pas linéaire**, le théorème de linéarité de l'espérance ne s'applique donc pas.
2. $\text{Var}(X + c) = \text{Var}(X)$ pour n'importe quelle constante c,
3. $\text{Var}(cX) = c^2\text{Var}(X)$ pour n'importe quelle constante c,
4. Si X et Y sont indépendants, alors $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$,
5. $\text{Var}(X) \geq 0$, l'égalité étant vraie si et seulement si $P(X = a) = 1$ où a est une constante.

2.3 Distribution uniforme : $U \sim$ **Unif**(a, b)

Une variable aléatoire U est distribuée uniformément sur l'intervalle (a, b) si sa PDF est,

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad a < x < b \\ 0 & \text{sinon} \end{cases}$$

La CDF correspond à l'aire en dessous de la PDF,

$$F(x) = \int_{-\infty}^{x} f(t) dt = \int_{a}^{x} f(t) dt = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } a < x < b, \\ 1, & \text{si } x \geq b \end{cases}$$

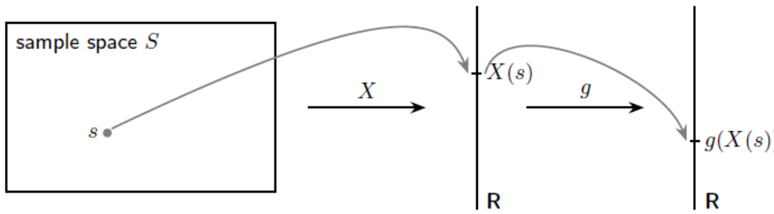
2.4 Variances particulières

- Soit $X \sim \text{Bern}(p)$, $\text{Var}(X) = pq$
- Soit $X \sim \text{Bin}(n, p)$, $\text{Var}(X) = npq$
- Soit $X \sim \text{Geom}(p)$, $\text{Var}(X) = \frac{q}{p^2}$
- Soit $X \sim \text{NBin}(r, p)$, $\text{Var}(X) = r^2 \frac{q}{p^2}$
- Soit $X \sim \text{Pois}(\lambda)$, $\text{Var}(X) = \lambda$

3 "Law of the unconscious statistician (LOTUS)"

3.1 Définition

Pour un espace d'échantillonage S, une v.a. X et une fonction $g : \mathbf{R} \to \mathbf{R}$, $g(X)$ est la variable aléatoire qui lie s à $g(X(s))$ pour tout $s \in S$.



3.2 Théorèmes

1. Si X est une variable aléatoire avec une PDF f et g est une fonction s'étendant de $\mathbf{R}$ dans $\mathbf{R}$, alors

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$$

2. Si X est une variable aléatoire discrète et g une fonction s'étendant de $\mathbf{R}$ dans $\mathbf{R}$, alors,

$$E(g(X)) = \sum_x g(x)P(X = x)$$

4 Universalité de la distribution uniforme

Soit F une CDF continue et croissante sur son support. Son inverse F_{-1} existe et est définie de $(0,1)$ dans $\mathbf{R}$. On a les résultats suivants,

1. Soit $U \sim \text{Unif}(0,1)$ et $X = F_{-1}(U)$, alors X est une variable aléatoire avec une CDF F,
2. Soit X une variable aléatoire avec une CDF F, $F(X) = \text{Unif}(0,1)$

5 Distribution normale

5.1 Distribution normale standard : $Z \sim \mathcal{N}(0,1)$

Une v.a. continue Z possède une distribution normale standard si sa PDF ϕ est donnée par,

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \quad -\infty < z \infty$$

On écrit $Z \sim \mathcal{N}(0,1)$ puisque Z a une moyenne de 0 et une variance de 1.
En effet,

- **Espérance** :

$$E(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz = 0$$

- **Variance** :

$$\text{Var}(Z) = E(Z^2) - (E(Z))^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz = 1$$

Propriétés de symétrie

- *Symétrie de la PDF* : $\phi(z) = \phi(-z)$, ϕ est paire,
- *Symétrie des extrémités* :$\Phi(z) = 1 - \Phi(-z)$,
- *Symétrie de Z et $-Z$* : Si $Z \sim \mathcal{N}(0,1)$ alors $-Z \sim \mathcal{N}(0,1)$ et,

$$P(-Z \leq z) = P(Z \geq -z) = 1 - \Phi(-z) = \Phi(z)$$

5.2 Distribution normale : $X \sim \mathcal{N}(\mu, \sigma^2)$

Si $Z \sim \mathcal{N}(0,1)$, alors

$$X = \mu + \sigma Z$$

avec μ comme moyenne et σ^2 comme variance.

Distribution normale non-standard

Pour $X \sim \mathcal{N}(\mu, \sigma^2)$, la version standardisée de X est,

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

Théorème 1. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, la CDF de X est,

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

et la PDF de X est

$$f(x) = \phi\left(\frac{x-\mu}{\sigma}\right)\frac{1}{\sigma}$$

Théorème 2. Si $X \sim \mathcal{N}(\mu, \sigma^2)$, alors

$$\begin{aligned} P(|X-\mu| < \sigma) &\approx 0.68 \\ P(|X-\mu| < 2\sigma) &\approx 0.95 \\ P(|X-\mu| < 3\sigma) &\approx 0.997 \end{aligned}$$

6 Distribution exponentielle : $X \sim$ Expo(λ)

Une v.a. continue X possède une distribution exponentielle avec le paramètre λ, où $\lambda > 0$, si sa PDF correspond à,

$$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$

Sa CDF est,

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0$$

6.1 "Memoryless property"

"No matter how long you have waited so far, your additional waiting time is still Exponential (with the same parameter)"

Une distribution continue "ne possède pas de mémoire" si une variable aléatoire X provenant de cette distribution valide

$$P(X \geq s + t | X \geq s) = P(X \geq t)$$

avec s le temps déjà passé à attendre et t le temps additionnel.

Chapitre 5

Distributions jointes

1 Distributions jointes, marginales et conditionnelles

- la loi *jointe* de 2 v.a. X et Y donne des informations sur la probabilité que le vecteur (X,Y) tombe dans un sous-ensemble de $\mathbb{R}^2$.
- La loi *marginale* de X correspond à la probabilité de X, ignorant Y.
- La loi *conditionnelle* de X sachant $Y=y$ est la probabilité de X après observation de Y.

1.1 Distribution jointe

CDF. La CDF jointe de deux v.a. X et Y est la fonction $F_{X,Y}$,

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

PMF. La PMF jointe de deux v.a. discrètes X et Y est la fonction $p_{X,Y}$,

$$p_{X,Y}(x,y) = P(X = x, Y = y)$$

CDF. Si X et Y sont continues avec une CDF jointe $F_{X,Y}$,

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$$

1.2 Distribution marginale

— Pour des v.a. **discrètes** X et Y, la PMF marginale de X est,

$$P(X = x) = \sum_y P(X = x, Y = y)$$

— Pour des v.a. **continues** X et Y avec une PDF conjointe $f_{X,Y}$, la PDF marginale de X est,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Indépendance des événements

2 variables aléatoire sont indépendantes si pour tout x et y,

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

— Si X et Y sont discrètes : $P(X = x, Y = y) = P(X = x)P(Y = y)$

— Si X et Y sont continues : $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

1.3 Distribution conditionnelle

PMF. Pour 2 v.a. discrètes X et Y, la PMF de Y sachant $X = x$ est donnée par

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

PDF. Pour 2 v.a. continues X et Y avec une PDF jointe $f_{X,Y}$, la PDF de Y sachant $X = x$ est donnée par

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

Formule de Bayes

Pour des v.a. continues X et Y,

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)} \qquad \text{Loi de Bayes}$$

$$f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|y)f_Y(y)dy \qquad \text{LOTP}$$

Indépendance des événements

— Si X et Y sont discrètes, l'indépendance entre les deux est vérifiée si,

$$P(Y = y|X = x) = P(Y = y)$$

— Si X et Y sont continues, l'indépendance entre les deux est vérifiée si,

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

LOTUS 2D

Soit g une fonction de $\mathbb{R}^2$ dans $\mathbb{R}$. Si X et Y sont discrètes, alors,

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

Si X et Y sont continues, alors,

$$E(g(X, Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$$

2 Distribution polynômiale

La distribution polynomiale est une généralisation de la loi binômiale. En effet, cette loi permet d'étendre la binômiale à plus que 2 catégories. Là où la binômiale classait en succès ou en échecs, la polynômiale peut classer en catégories "excellente", "correcte", "médiocre", par exemple.

Distribution polynômiale jointe

Si $\mathbf{X} \sim \text{Mult}_k(n, \mathbf{p})$, alors la PMF jointe de X est,

$$P(X_1 = n_1, ..., X_k = n_k) = \frac{n!}{n_1!n_2!...n_k!} \cdot p_1^{n_1} p_2^{n_2} ... p_k^{n_k},$$

Pour $n_1, ..., n_k$ satisfaisant $n_1 + ... + n_k = n$

Distribution polynômiale marginale

Si $\mathbf{X}\sim$ $\text{Mult}_k(n,\mathbf{p})$, alors $X_j \sim \text{Bin}(n, p_j)$

Distribution polynômiale "lumping" (regroupement)

Si $\mathbf{X}\sim$ $\text{Mult}_k(n,\mathbf{p})$ pour tout i et j distincts, $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$. Le vecteur aléatoire obtenu de la fusion des catégories i et j est toujours multipolynômial. Par exemple, la fusion des catégories 1 et 2 donne,

$$(X_1 + X_2, X_3, ..., X_k) \sim \text{Mult}_{k-1}(n, (p_1 + p_2, p_3, ..., p_n))$$

Distribution polynômiale conditionnelle

Si $\mathbf{X}\sim$ $\text{Mult}_k(n,\mathbf{p})$, alors

$$(X_2, ..., X_k)|X_1 = n_1 \sim \text{Mult}_{k-1}(n - n_1, (p'_2, ..., p'_k)),$$

Où $p'_j = p_j/(p_2 + ... + p_k)$

3 Covariance et corrélation

3.1 Covariance

La covariance correspond à un résumé en seul nombre de la distribution jointe de deux variables aléatoires.
En gros, elle mesure la tendance des deux variables à augmenter ou diminuer ensemble par rapport à leurs valeurs attendues,

— Une covariance positive indique que quand X "augmente", Y a tendance également à s'élever.

— Une covariance négative indique que quand X "augmente", Y a tendance à "diminuer".

Définition

La covariance entre X et Y est donnée par,

$$\text{Cov(X,Y)} = E((X - EX)(Y - EY)) = E(XY) - E(X)E(Y)$$

Propriétés de la covariance

1. $\text{Cov}(X, Y) = \text{Var}(X)$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3. $\text{Cov}(X, c) = 0$ pour toute constante c
4. $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$ pour toute constante a
5. $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
6. $\text{Cov}(X + Y, Z + W) = \text{Cov}(X, Z) + \text{Cov}(X, W) + \text{Cov}(Y, Z) + \text{Cov}(Y, W)$ et sa version généralisée,

$$\text{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \,\text{Cov}\big(X_i, Y_j\big)$$

7. $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$ et sa version généralisée,

$$\text{Var}\,(X_1 + \cdots + X_n) = \text{Var}\,(X_1) + \cdots + \text{Var}\,(X_n) + 2\sum_{i<j} \text{Cov}\big(X_i, X_j\big)$$

3.2 Corrélation et dépendance

Si X et Y sont indépendantes, elles sont dites **non-corrélées**.
En effet, si X et Y sont indépendantes, leur covariance est égale à 0.

⚠ l'absence de corrélation n'implique pas l'indépendance!

Définition

La corrélation entre X et Y est donnée par,

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

Bien entendue, celle-ci est indéfinie dans les cas dégénérés $\mathrm{Var}(X) = 0$ ou $\mathrm{Var}(Y) = 0$.
Le domaine de définition de la corrélation est,

$$-1 \leq \mathrm{Corr}(X,Y) \leq 1$$

4 Loi normale multidimensionnelle

La loi normale multidimensionnelle est une distribution continue qui généralise la distribution normale dans des dimensions supérieurs.

Définition

Un vecteur aléatoire **X**$= (X_1, ..., X_k)$ suit une loi normale multidimensionnelle (MVN) si chaque combinaison linéaire des X_j suivent une loi de distribution normale.

$$t_1 X_1 + ... + t_k X_k$$

doit alors suivre une loi de distribution normale pour n'importe quel choix de constantes $t_1, ..., t_k$.

Propriétés

- Si $t_1 X_1 + ... + t_k X_k$ est une constante (par exemple, $t_i = 0$), on la considère comme suivant une distribution normale, bien que la variance étant 0, on puisse l'attribuer à une distribution normale dégénérée.
- Pour $k = 2$, cette distribution est appellée la *Normale Bivariée* (BVN)
- Si $(X_1, ..., X_k)$ est MVN, la distribution marginale de X_j est normale, puisque $t_j = 1$ et tous les autres $t_i = 0$ pour $i \neq j$ (contraire impossible).
- Pour un vecteur MVN aléatoire, la non-corrélation implique l'indépendance.

Chapitre 6

Espérance conditionnelle

1 Espérance compte tenu d'un événement $E(Y|A)$

1.1 Définition

Soit A un événement à probabilité positive. Si Y est une v.a. discrète, alors l' *espérance conditionnelle de Y étant donné A* est donnée par,

$$E(Y|A) = \sum_y yP(Y = y|A),$$

Où la somme s'étend sur le support de Y. Si Y est une v.a. continue avec une PDF f, alors,

$$E(Y|A) = \int_{-\infty}^{\infty} yf(y|A)dy,$$

où la PDF conditionnelle $f(y|A)$ est définie comme la dérivée de la CDF conditionnelle $F(y|A) = P(Y \leq y|A)$

1.2 Loi de l'espérance totale

Soit $A_1, ..., A_n$ une partie de l'espace d'échantillonage, avec $P(A_i) > 0$ pour tout i, et soit Y une variable aléatoire de cet espace. Alors,

$$E(Y) = \sum_{i=1}^{n} E(Y|A_i)P(A_i)$$

Et puisque toutes les probabilités sont des espérances compte tenu du pont fondamental, la loi de probabilité totale est un cas spécial de la loi de l'espérance totale,

$$P(B) = E(I_B) = \sum_{i=1}^{n} E(I_B|A_i)P(A_i) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

2 Espérance compte tenu d'une variable aléatoire $E(Y|X)$

Soit $g(x) = E(Y|X = x)$. L'espérance conditionnelle notée $E(Y|X)$ est définie par la variable aléatoire $g(X)$. En d'autres termes, si après l'expérience, X "cristallise" en x, alors $E(Y|X)$ "cristallise" en $g(x)$.

Mauvaise interprétation : Cela ne veut *pas* dire que "$g(x) = E(Y|X = x)$, donc $g(X) = E(Y|X = X)$, qui est égal à $E(Y)$ parceque $X = X$ est toujours vrai".

Interprétation correcte : On doit d'abord calculer la fonction $g(x)$, *ensuite* on insère x à la place de X. Par exemple, si $g(x) = x^2$, alors $g(X) = X^2$.

3 Loi d'Adam et propriétés

Loi d'Adam

Pour toutes v.a. X et Y,

$$E(E(Y|X)) = E(Y)$$

Propriétés

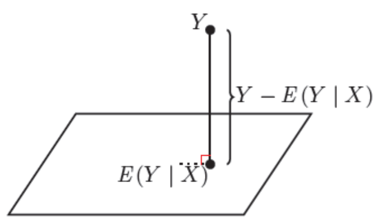
- Si X et Y sont indépendants, alors $E(Y|X) = E(Y)$
- Pour n'importe quelle fonction h,
$$E(h(X)Y|X) = h(X)E(Y|X)$$
- Par linéarité,
$$E(Y_1 + Y_2|X) = E(Y_1|X) + E(Y_2|X)$$

Interprétation de projection

Pour toute fonction h, la variable aléatoire $Y - E(Y|X)$ est non-corrélée avec $h(X)$.

$$E((Y - E(Y|X))h(X) = 0$$

Puisque par linéarité et par la loi d'Adam, $E(Y - E(Y|X)) = 0$.



- Le plan représente l'ensemble des fonctions de X
- $E(Y|X)$ est la fonction de X la plus proche de Y, on l'appelle la *projection* de Y sur l'espace des fonctions de X.
- Le *résidu* $Y - E(Y|X)$ est orthogonal au plan, il est perpendiculaire à toute fonction de X.

4 Loi d'Eye et variance conditionnelle

Variance conditionnelle

La variance conditionnelle de Y sachant X est donnée par,

$$\text{Var}(Y|X) = E((Y - E(Y|X))^2|X) = E(Y^2|X) - (E(Y|X))^2$$

Loi d'Eye

Pour toutes variables aléatoires X et Y,

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$