Modèles probabilistes pour l'informatique

Responsable et auteur : Massih-Reza AMINI
Intervenante : Myriam TAMI

Université Grenoble Alpes (UGA)
Laboratoire d'Informatique de Grenoble (LIG)

A quoi nous servent les probabilités?

Probabilités et Informatique

Les probabilités interviennent aujourd'hui dans tous les secteurs d'activité d'un informaticien qui les utilise aussi bien

- En Intelligence Artificielle,
- qu'en Décision,
- qu'en Réseaux informatique,
- qu'en traitement de l'image ou en bioinformatique, ...

L'objectif de ces cours

Ces cours ont pour objectif de vous familiariser avec les notions de bases en probabilité et en statistique.

A quoi nous servent les statistiques?

Statistique d'après E. Universalis

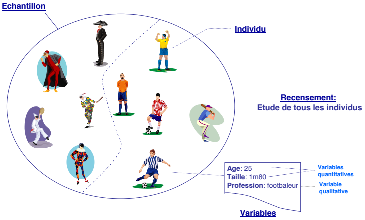
Le mot *statistique* désinge à la fois un ensemble de données d'observation et l'activité qui consiste dans leur recueil, leur traitement et leur intérpretation.

Exemples

Nous étudions par exemple, des fiches concernant 100000 sportifs et gens de spectacles. Ces fiches décrivent entre autre, les gains individuels, leur performance, ... Ces fihces constituent une **statistique**.
Faire de **statistique** sur ces données consiste par exemple, à:

- Calculer la moyenne des gains d'un individu,
- Prévoir ses performances pour une saison donnée, etc.

Notions de Bases



Définitions

- L'analyse statistique trouve sa justification dans la variabilité,
- On cherche à étudier cette variabilité (description), à la prévoir (estimation) et à l'expliquer (modelisation)

Organisation

Infos pratiques

- Le site du cours

  http://ama.liglab.fr/~amini/Cours/L3/

- Les intervenants:
  - TAMI, Myriam (https://myriamtami.github.io/)
  - GAST, Nicolas (http://mescal.imag.fr/membres/nicolas.gast/)

Validation

- 2 projets (30%)
- un examen final (70%)

Introduction

Généralités

- Le terme *probabilité* (du latin `probare`) signifie *prouver*,
- Ce terme est passé dans le langage courant et a parfois une connotation liée à la chance,
- D'autres termes du calcul des probabilités montrent leur lien avec les jeux de hasard
  - Par exemple, celui **d'espérance mathématiques** qui correspond à **l'espérance du gain**.

Historique

- Les premières traces d'une théorie de la probabilité remontent au 17^{eme} siècle en liaison avec des jeux de hasard,
- Les premiers résultats mathématiques sont obtenus par Pascal et Fermat,
- Huyghens, Bernoulli, de Moivre, Bayes, Laplace, Borel, ... apportent de nouveaux concepts,
- Ce cours est basé sur les axiomatiques de Kolmogorov.

Expérience aléatoire, évènements aléatoires

- **Une expérience** est qualifiée d'**aléatoire** si on ne peut pas prévoir par avance son résultat et si répétée dans les mêmes conditions, elles donnent des résultats différents.
- Exemples
  - Lancé de dé,
  - La roue de la fortune,
- Les résultats possibles de cette expérience constituent **l'ensemble fondamental** Ω appelé aussi **univers des possibles**.
- Un **événement aléatoire** est un résultat parmi d'autre de l'ensemble Ω

Exemple

Dans le lancer simultané de deux dés, les résultats obtenus sur les faces supérieures, l'ensemble fondamental Ω est

$$\Omega = \{(1,1),(1,2),(1,3),...(5,6),(6,6)\}$$

Algèbre des événements

Nous allons nous référer à des propriétés ensemblistes usuelles:

- À tout événement E, on associe son contraire $\bar{E}$,
- Aux événements E et F, on associe $E \cup F$,ou $E \cap F$,
- L'événement certain est représenté par Ω,
- L'événement impossible est représenté par $\varnothing$

Ce qui nous amène aux définitions suivantes:

- On appelle **tribu de parties d'un ensemble** Ω un ensemble C de parties de Ω vérifiant:
  - $\Omega \in C$,
  - $\forall E \in C, \bar{E} \in C$,
  - Pour tout ensemble fini ou dénombrable de parties E_i de C, $\bigcup_{i \in I} E_i \in C$
- On appelle **espace probabilisable** un couple (Ω, C) où C est une **tribu** de parties de l'ensemble Ω,
- E et F sont deux **événements incompatibles** si la réalisation de l'un exclut celle de l'autre, (i.e. $E \cap F = \varnothing$)
- $E_1, ..., E_n$ forment un **système complets d'événements** si elles constituent une partition de Ω

Probabilité

On appelle **probabilité sur l'espace probabilisable** (Ω, C), une application P de C dans $[0, 1]$ telle que:

- $P(\Omega) = 1$
- Pour tout ensemble dénombrable $E_1, .., E_n$ d'événements incompatibles:

$$P(\bigcup_i E_i) = \sum_i P(E_i)$$

- Le triplet (Ω, C, P) est appelé **espace probabilisé**

Propriétés

- $P(\varnothing) = 0$,
- $P(\bar{E}) = 1 - P(E)$
- $E \subset F \Rightarrow P(E) \leq P(F)$
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$
- $P(\bigcup_i E_i) \leq \sum_i P(E_i)$

Théorème des probabilités totales

Théorème des probabilités totales

Soit $(F_i)_i$ un système complet d'événements de Ω, alors:

$$\forall E \in \Omega, P(E) = \sum_i P(E \cap F_i)$$

Espaces probabilisés élémentaires

- Considérons un espace Ω fini ou dénombrable, A tout élément e de Ω, on associe un nombre réel positif ou nul $P(e)$ tel que

$$\sum_e P(e) = 1$$

- À toute partie E de Ω on associe alors le nombre $P(E) = \sum_{e \in E} P(e)$
- Si on attribue le même poids à chaque événement élémentaire de Ω: $P(e) = \frac{1}{card(\Omega)}$
- Avec une probabilité uniforme sur Ω, la probabilité d'une partie $E \subset \Omega : P(E) = \frac{card(E)}{card(\Omega)}$

Espaces probabilisés élémentaires

Exemple

Avec le lancer des deux dés, La probabilité de l'événement
E=`La somme des deux chiffres est inférieure ou égale à` 5 est

$$P_E = \frac{10}{36}$$

L'ensemble correspondant à l'événement E est en effet:

$$E = \{(1,1),(1,2),(1,3),(1,4),(2,1),(2,2),(2,3),(3,1),(3,2),(4,1)\}$$

Exemple

Toujours avec l'exemple précédent
F=`Au moins un des chiffres est supérieur ou égal à` 5 est

$$P_F = \frac{26}{36}$$

Rappels élémentaires d'analyse combinatoire

Soient E et F deux ensembles finis, avec $card(E) = k$ et $card(F) = n$. Le dénombrement de certaines applications de $E \to F$ donnent

- Nombre d'applications quelconques: n^k,
- Nombre d'applications injectives: $A_n^k = (n-k+1)...(n-1)n$,
- Nombre de bijections (cas $n = k$): $n! = 1...(n-1)n$

Suite ...

- soit E un ensemble fini de cardinal n, Le nombre de sous-ensembles distincts de cardinal k contenus dans E:

$$C_n^k = \frac{n!}{k!(n-k)!}$$

- Remarque

$$C_n^k = C_n^{n-k} \Rightarrow C_n^n = C_n^0 = 1 \Rightarrow 0! = 1$$

- Formule du binôme de Newton

$$(x+y)^n = \sum_{k=0}^{n} C_n^k x^{n-k} y^k$$

Rappels élémentaires d'analyse combinatoire

Conséquence

Si E est un ensemble de cardinal n, on a
$card\ \mathcal{P}(E) = 2^n$

Exemple

Tirer deux cartes, sans remise, dans un jeu de 52 cartes.
L'ensemble de tous les événements:

$$\Omega = \{\{a, b\} \mid a \text{ et } b \text{ sont deux cartes du jeu}\}$$

Tous les sous-ensembles sont de *card* 2 sont équiprobables

$$P(\{a, b\}) = \frac{1}{1326}, \forall \{a, b\} \in \Omega$$

E=`La probabilité d'obtenir une dame au moins`

$$P(E) = 1 - \frac{C_{48}^2}{1326}$$

Problème du Prince de Toscane

Exemple

Pourquoi en lançant trois dés, obtient-on plus souvent un total de 10 points qu'un total de 9 points, alors qu'il y a 6 façons d'obtenir ces deux totaux?

$$\Omega = \{(i,j,k) \mid i \in \{1,..,6\}, j \in \{1,..,6\}, k \in \{1,..,6\}\}$$

Le cardinal de Ω, $card(\Omega) = 6^3 = 216$ et comme tous les résultats sont équiprobables

$$P((i,j,k)) = \frac{1}{216}, \forall (i,j,k) \in \Omega$$

Pour un lancé de $\{i,j,k\}$ donné on a les cas suivants:

- $i \neq j \neq k \neq i$ dans ce cas, $P(\{i,j,k\}) = \frac{6}{216}$
- $i = j \neq k$, dans ce cas, $P(\{i,j,k\} = \frac{3}{216}$
- $i = j = k$, dans ce cas, $P\{i,j,k\}) = \frac{1}{216}$

On a alors $P(\{\text{total}9\}) = \frac{25}{216}$ et $P(\{\text{total}10\}) = \frac{27}{216}$

Lois Conditionnelles

Considérons deux événements E et F, Supposons qu'on ne s'intéresse à la réalisation de E, étant donné la réalisation de F. Cela revient à estimer la réalisation de $E \cap F$ par rapport à F

Définition

Soit (Ω, C, P) un espace probabilisé et F un événement de probabilité non nulle. On appelle **probabilité conditionnelle sachant** F l'application:

$$P(. \mid F) : C \to [0,1]$$

définie par

$$\forall E \in C, P(E \mid F) = \frac{P(E \cap F)}{P(F)}$$

Cette application définit bien une probabilité sur le même espace probabilisé.

Lois Conditionnelles (2)

Remarques

- Pour tout événement E, nous avons

$$P(E \mid \Omega) = P(E) \text{ et } P(E \mid E) = 1$$

Formule de Bayes

événements réalisables, nous avons alors:
$P(E \cap F) = P(F \mid E) \times P(E) = P(E \mid F) \times P(F)$, soit

$$P(E \mid F) = \frac{P(F|E)P(E)}{P(F)}$$

La formule précédente s'étend à un nombre quelconque d'événements: Soient $(E_i)_{i \in \{1,..,n\}}$, n événements aléatoires de Ω, on a alors

$$P(E_1 \cap E_2 \cap \ldots \cap E_n) = P(E_1) \prod_{i=2}^{n} P(E_i \mid E_1, \ldots, E_{i-1})$$

Lois Conditionnelles (3)

Exemple

Quelle est la probabilité de tirer trois boules de la même couleur dans une urne contenant 7 boules rouges et 5 boules bleues, en tirant les trois boules l'une après l'autre et sans remise?

Posons

- R_i=La i^{eme} boule tirée est rouge, $i \in \{1,2,3\}$
- B_i=La i^{eme} boule tirée est bleue, $i \in \{1,2,3\}$

On a alors,

$$P(R_1 \cap R_2 \cap R_3) = \frac{7}{12} \times \frac{6}{11} \times \frac{5}{10} \text{ et } P(B_1 \cap B_2 \cap B_3) = \frac{5}{12} \times \frac{4}{11} \times \frac{3}{10}$$

Formule de Bayes

Soient (Ω, C, P) un espace probabilisé et B un événement réalisable de C. Soit $\{A_j\}_{j \in \mathbb{I} \subset \mathbb{N}}$ un système complet d'événements, on a alors :

$$\forall j \in \mathbb{I}, P(A_j \mid B) = \frac{P(B \mid A_j) \times P(A_j)}{\sum_{i \in \mathbb{I}} P(B \mid A_i) \times P(A_i)}$$

Exemple

Il manque une carte dans un jeu de 52 cartes et on ignore laquelle. On tire au hasard une carte dans ce jeu incomplet et c'est un coeur. Quelle est alors la probabilité pour que la carte perdue soit un coeur?
Soient les événements élémentaires suivants:

- CP: La carte perdue est un coeur, TC: Tirer un coeur du jeu incomplet

Nous avons alors $P(CP) = \frac{1}{4}$ et $P(TC \mid CP) = \frac{12}{51}$
TC peut s'écrire comme: $TC = (TC \cap CP) \cup (TC \cap \bar{CP})$ et

$$P(CP \mid TC) = \frac{P(TC \mid CP) \times P(CP)}{P(TC \mid CP) \times P(CP) + P(TC \mid \bar{CP}) \times P(\bar{CP})} = \frac{\frac{12}{51} \times \frac{1}{4}}{\frac{12}{51} \times \frac{1}{4} + \frac{13}{51} \times \frac{3}{4}} = \frac{12}{51}$$

Indépendance

Définition

- L'événement E est **indépendant** de l'événement F si:

$$P(E \mid F) = P(E)$$

- Conséquence:
  - E et F sont **indépendants** si et seulement si

$$P(E \cap F) = P(E) \times P(F)$$

Indépendance mutuelle

Les événements $E_1, ..., E_n$ sont dits **mutuellement indépendants** si, pour toute partie $\mathbb{I}$ de l'ensemble $\{1, .., n\}$:

$$P(\bigcap_{i \in \mathbb{I}} E) = \prod_{i \in \mathbb{I}} P(E_i)$$

Indépendance

Exemple

On lance un dé jusqu'à ce qu'on obtienne le résultat $\{1\}$ pour la première fois.

$$\Omega = \{(\underbrace{\bar{1}, \bar{1}, \ldots, \bar{1}}_{k-1}, 1) \cup (\bar{1}, \ldots, \bar{1}, \ldots) \mid \forall k \in \mathbb{N}^*\}$$

Les probabilités des événements suivants sont:

- E_k=Obtenir le résultat 1 pour la 1^{ere} fois au k^{ieme} lancer

$$P(E_k) = \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6}$$

- E=Obtenir le résultat 1

$$P(E) = \sum_{k\geq 1} P(E_k) = \lim_{n\to\infty} \sum_{k=1}^{n} \left(\frac{5}{6}\right)^{k-1} \times \frac{1}{6} = 1$$

- J Ne jamais obtenir 1

$$P(J) = 1 - P(E) = 0$$

Variables aléatoires, Inégalités et Lois de probabilité

Responsable et auteur : Massih-Reza AMINI
Intervenante : Myriam TAMI

Université Grenoble Alpes (UGA)
Laboratoire d'Informatique de Grenoble (LIG)

Intuition

C'est quoi une *variable aléatoire*?

Lorsqu'on est devant une épreuve aléatoire, on s'intéresse plus souvent à la *valeur* attribuée à l'épreuve qu'à son résultat.

Exemple

Par exemple, lorsque l'on joue à un jeu de hasard on s'intéresse plus au gain qu'on aura avec notre jeu qu'à son résultat.

Définition

Une variable aléatoire (*v.a.*) X est un nombre réel que l'on associe à chaque élément e de l'ensemble Ω. X est donc une application définie comme:

$$\begin{aligned} X:\Omega &\rightarrow \mathbb{R} \\ e &\mapsto X(e) = x \end{aligned}$$

X est la variable aléatoire réelle et x est une réalisation de cette *v.a.*

Exemple

Exemple du lancé de dé

On lance un dé après avoir misé 1€. Si le résultat est un 5 ou un 6 on double la mise sinon la perd. Dans ce cas:

- $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Card $\Omega = 6$ et Card $\mathcal{P}(\Omega) = 2^6$
- Soit X la *v.a.* qui associe à tout résultat du dé un *gain*:
  $X(1) = X(2) = X(3) = X(4) = -1$ et $X(5) = X(6) = (2 - 1) = 1$
  L'ensemble des valeurs possibles de X noté $\mathcal{X}$ est
  $$\{-1, 1\} \subset \mathbb{R}$$
- Question: Comment calculer la probabilité de gagner 1€?
- Réponse: Définir une probabilité sur $\mathcal{X}$, notée $\mathbb{P}$, en retournant dans l'espace probabilisé (Ω, C, P)
  i.e. utiliser $P(\textit{Le résultat du dé est } 5 \textit{ ou } 6)$ pour estimer $\mathbb{P}(1)$.

Définition formelle d'une *v.a.*

But

- On a un espace probabilisé (Ω, C, P) et l'application $X : \Omega \to \mathbb{R}$
- On veut pouvoir probabiliser des événements sur l'espace d'arrivé: $\mathbb{R}$

Solution

Construire un espace probabilisé sur $\mathbb{R}$ muni de sa tribu usuelle $\mathcal{B}(\mathbb{R})$. On notera $\mathbb{P}$ la mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

La tribu $\mathcal{B}(\mathbb{R})$ est appelée *tribu des boréliens* - en hommage à Émil Borel. Cette tribu est engendrée par les intervalles de la forme

$$]-\infty, a[, a \in \mathbb{R}$$

Schématiquement parlant ...

$$
\begin{array}{c}
(\Omega, C, P) \\
\Big\downarrow X \\
(\mathbb{R}, \mathcal{B}, \mathbb{P})
\end{array}
$$

$$
\begin{array}{rcl}
C & \xrightarrow{P} & [0,1] \\
X^{-1}(B) = A & \longmapsto & P(A)
\end{array}
$$

$$
\begin{array}{rcl}
\mathcal{B}(\mathbb{R}) & \xrightarrow{\mathbb{P}} & [0,1] \\
B & \longmapsto & \mathbb{P}(B)
\end{array}
$$

Définition formelle d'une *v.a.*

Suite ...

- $\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}(B) = P(X^{-1}(B)) = P(\{e_i \in \Omega \mid X(e_i) \in B\})$
- On dit que $\mathbb{P}(B)$ est la mesure par P de l'image réciproque de B par X, i.e. l'ensemble des antécédents de B. Cette image réciproque doit être un élément de c pour qu'on puisse lui appliquer la mesure de probabilité P

Définition

Soient (Ω, C) un espace probabilisable et X une application de Ω vers $\mathbb{R}$.
X est dite *mesurable* si et seulement si:

$$\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) \in \Omega$$

$$\begin{tikzcd}
C \arrow[dr, "P"] & \\
& {[0,1]} \\
\mathcal{B}(\mathbb{R}) \arrow[uu] \arrow[ur, "\mathbb{P}"'] &
\end{tikzcd}
\qquad
\begin{tikzcd}
A = X^{-1}(B) \arrow[dr] & \\
& \mathbb{P}(B) = P(A) \\
B \arrow[uu] \arrow[ur] &
\end{tikzcd}$$

Définition formelle d'une *v.a.*

Suite ...

- Une variable aléatoire réelle est une *application mesurable* d'un espace probabilisable quelconque dans $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$

- Pour alléger les notations:
  - L'événement $X^{-1}(B) = \{e \in \Omega \mid X(e) \in B\}$ sera noté par $\{X \in B\}$
  - L'événement $X \in]-\infty, a]$ sera noté par $\{X < a\}$
  - L'événement $X \in [a, b[$ sera noté par $\{a \leq X < b\}$
  - L'événement $X \in \{a\}$ sera noté par $\{X = a\}$

- On a donc $\mathbb{P}(B) = P(X^{-1}(B)) = P(\{X \in B\})$
- On appellera *loi de X* la donnée de la probabilité

$$\mathbb{P} \text{ sur } (\mathbb{R}, \mathcal{B}(\mathbb{R}))$$

V.A.

Exemple: Lancé de dé

Avec l'exemple du lancé de dé précédent nous avons $\Omega = \{1,2,3,4,5,6\}$ et

$$\forall e \in \Omega, P(e) = \frac{1}{6}$$

Posons: $X : \Omega \to \mathbb{R}$ la *v.a. gain* définie par

$$\begin{aligned} X(e_i) &= -1 \text{ si } e_i \in \{1,2,3,4\} = \mathrm{P} \\ X(e_i) &= +1 \text{ si } e_i \in \{5,6\} = \mathrm{G} \end{aligned}$$

Il est équivalent d'écrire

$$\begin{aligned} (X = -1) &= X^{-1}(\mathrm{P}) \\ (X = 1) &= X^{-1}(\mathrm{G}) \end{aligned}$$

Nous avons

$$P(X = -1) = \frac{2}{3}, P(X = 1) = \frac{1}{3}$$

$$\sum_{x \in \{-1,1\}} P(X = x) = \sum_{x \in \{-1,1\}} \mathbb{P}(\{x\}) = 1$$

Fonction de répartition

Définition

- Comme le tribu des boréliens $\mathcal{B}(\mathbb{R})$ est engendrée par des intervalles de la forme $]-\infty,a]$, tout événement de $\mathcal{B}(\mathbb{R})$ peut s'exprimer à partir de ceux-ci par des opérations ensemblistes. autrement dit:
  La connaissance des probabilité $\mathbb{P}(]-\infty,a[) = P(X<a)$ *suffit à probabiliser* $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ *par* $\mathbb{P}$.

- Soit X une *v.a.* on appelle **fonction de répartition** de X, notée F l'application:

$$\begin{aligned} F:\mathbb{R} &\rightarrow [0,1] \\ x &\mapsto F(x) = P(X<x) \end{aligned}$$

Fonction de répartition

Remarque

Pour définir la loi de X, il faut savoir calculer $\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{P}(B)$. Or $\mathbb{P}$ est une fonction d'ensemble.

En remarquant que la loi de X est équivalent à la donnée de sa fonction de répartition, il est possible de calculer la probabilité d'un intervalle à l'aide de $F(x)$ qui elle est une fonction d'une seule variable.

Propriétés

- $P(a \leq X \leq b) = F(b) - F(a)$
- F est croissante bornée:
  - $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$
  - $F(+\infty) = \lim_{x \to +\infty} F(x) = 1$

Fonction de répartition

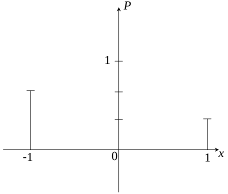
Exemple: Lancé de dé

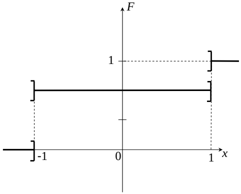
Avec l'exemple du lancé de dé précédent nous avons $\Omega = \{1,2,3,4,5,6\}$, et $X : \Omega \rightarrow \mathbb{R}$ la *v.a. gain* définie par

$$X(e_i) = -1 \text{ si } e_i \in \{1,2,3,4\} = \mathrm{P}, X(e_i) = +1 \text{ si } e_i \in \{5,6\} = \mathrm{G}$$

On a l'ensemble des valeurs possibles $\mathcal{X} = \{-1, 1\}$. On peut définir la distribution de probabilité de la *v.a.* par sa fonction de répartition $F : F(x) = P(X < x)$

- si $x \leq 1, \mathcal{X} \cap (]-\infty, x[) = \varnothing \Rightarrow F(x) = 0$
- si $x \in]-1, 1], \mathcal{X} \cap (]-\infty, x[) = \{-1\} \Rightarrow F(x) = \frac{2}{3}$
- si $x \in]1, \infty], \mathcal{X} \cap (]-\infty, x[) = \{-1, 1\} \Rightarrow F(x) = \frac{2}{3} + \frac{1}{3} = 1$





Fonction de répartition

Exemple: Tapis vert

On dispose d'un bulletin comportant six tables de jeu correspondant à la mise de départ, 2€, 5€, 10€, 20€, 50€ ou 100€. Une table de jeu représente un jeu de 32 cartes réparties en 4 colonnes (pique, coeur, trèfle, carreau). Chaque colonne, 8 lignes de 7 à l'as. On coche une case de chaque colonne, le jeu gagnant est tiré au hasard. Pour 4 résultats exacts on gagne 1000 fois la mise, 3, 30 fois, pour 2, 2 fois, pour moins d'un 1 la mise est perdue. Nous nous intéressons au *gain* pour une mise de 1€.

- $\Omega = \{(e_p, e_{xo}, e_{ca}, e_t) \mid e_i \in \{7, ..., as\}, i \in \{p, co, ca, t\}\}$, card $\Omega = 8^4 = 4096$,
- À tout tirage est associé un *gain* noté $X(e)$ qui est une $v.a. \in \{-1, 1, 29, 999\}$
- $\mathbb{P}(29) = P(X = 29) = P(\textit{trois cartes gagnantes})$

Fonction de répartition

Exemple: Tapis vert (suite)

- La loi de probabilité de la *v.a.* se déduit de la mesure de probabilité P sur (Ω, C):
  - $P(X=-1)=\frac{3773}{4096}, P(X=1)=\frac{294}{4096}, P(X=29)=\frac{28}{4096}, P(X=999)=\frac{1}{4096}$
- Si $x\in]-\infty,-1], F(x)=0$, pas de gain possible
- Si $x\in]-1,1], F(x)=\frac{3773}{4096}$, Le seul gain possible $=-1$
- Si $x\in]1,29], F(x)=\frac{3773+294}{4096}$, Les gains possibles $\in\{-1,1\}$
- Si $x\in]29,999], F(x)=\frac{3773+294+28}{4096}$, Les gains possibles $\in\{-1,1,29\}$
- Si $x\in]999,\infty[, F(x)=1$, tous les gains sont possibles

Massih-Reza.Amini@imag.fr

v.a. et Fonction de répartition

Définition

- Toute *v.a.* dont la fonction de répartition est une fonction en escalier, est une *v.a. discrète*
- Dans ce cas, définir la *v.a.* discrète X est de se donner un ensemble $\mathcal{X}$ de valeurs possibles, au plus dénombrable, et des $p(x) = P(X = x)$ définie par

$$\forall x \in \mathcal{X}, p(x) \geq 0 \text{ et } \sum_{x \in \mathcal{X}} p(x) = 1$$

  On dit que la famille $(p(x))_{x \in \mathcal{X}}$ définit une loi de X.
- Toute *v.a.* dont la fonction de répartition F est une fonction continue est appelée variable *absolument continue*
- On appelle <u>densité</u> d'une *v.a.*, X la dérivée p de sa fonction de répartition F

$$F(x) = \int_{-\infty}^{x} p(u) du$$

v.a. et Fonction de répartition

Définition (suite)

- Toute fonction réelle $p(x)$ telle que

$$\forall x \in \mathbb{R}, p(x) \geq 0 \text{ et } \int_{-\infty}^{\infty} p(x)dx = 1$$

est une densité de probabilité. i.e. Il existe un espace probabilisé (Ω, C, P) et une *v.a.* réelle X, application mesurable de $(\Omega, C) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, dont p est la densité de probabilité.

- On a ainsi

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

v.a. et Fonction de répartition

Définition (suite)

- Définir la *v.a.* absolument continue X c'est de se donner sa densité de probabilité $p(x)$ telle que

$$\forall x \in \mathbb{R}, p(x) \geq 0 \text{ et } \int_{-\infty}^{\infty} p(x)dx = 1$$

On assimile souvent la loi de X à la donnée de sa densité de probabilité p ou de sa fonction de répartition F.

Espérance mathématique

Définition

Soit X une *v.a.* on appelle *espérance mathématique* de X, que l'on note $\mathbb{E}(X)$ la quantité suivante:

- Si X est discrète

$$\mathbb{E}(X) = \sum_i x_i P(X = x_i)$$

- si X est continue

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} xp(x)dx = \int_{\mathbb{R}} xp(x)dx$$

Propriétés

Pour tout réel $a \in \mathbb{R}$ et tout couple de *v.a.*(X, Y)

- $\mathbb{E}(a) = a$
- $\mathbb{E}(aX) = a\mathbb{E}(X)$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Espérance mathématique

Espérance d'une *v.a.* centrée

Une variable aléatoire X est dite centrée si et seulement si $\mathbb{E}(X) = 0$. Pour toute *v.a.* X, la variable $X - \mathbb{E}(X)$ est alors centrée.

Espérance du produit de deux *v.a.*

Pour le produit nous avons:

$$\mathbb{E}(X \times Y) = \int_{\mathcal{X}} xp(x)dx \int_{\mathcal{Y}} yp(y \mid x)dy$$

Cas particulier

Si X et Y sont deux v.a. indépendantes on a alors

$$\mathbb{E}(X \times Y) = \mathbb{E}(X)\mathbb{E}(Y)$$

La réciproque est fausse: l'égalité $\mathbb{E}(X \times Y) = \mathbb{E}(X)\mathbb{E}(Y)$ n'entraîne pas l'indépendance des variables X et Y.

Espérance mathématique

Exemple du jeu Tapis Vert

Rappel

- $\Omega = \{(e_p, e_{xo}, e_{ca}, e_t) \mid e_i \in \{7, ..., \text{as}\}, i \in \{p, co, ca, t\}\}$, card $\Omega = 8^4 = 4096$,
- À tout tirage est associé un *gain* dans $\{-1, 1, 29, 999\}$
- $P(X = -1) = \frac{3773}{4096}, P(X = 1) = \frac{294}{4096}, P(X = 29) = \frac{28}{4096}, P(X = 999) = \frac{1}{4096}$

L'espérance de gain est pour ce jeu ... *négatif*

$$\begin{aligned}\mathbb{E}(X) &= -1 \times \frac{3773}{4096} + 1 \times \frac{294}{4096} + 29 \times \frac{28}{4096} + 999 \times \frac{1}{4096} \\ &= -\frac{1668}{4096} \approx -0.4\end{aligned}$$

Variance d'une *v.a.*

Définition

Soit X une *v.a.* d'espérance m, On appelle <u>variance</u> de X, la quantité:

$$V(X) = \sigma_X^2 = \mathbb{E}\left[(X - \mathbb{E}(X))^2\right] = \int_{\mathbb{R}} (x - m)^2 p(x) dx$$

Remarques

- La variance est le carré de l'écart-type, σ_X
- La variance est une mesure de la dispersion de X autour de m.

Propriétés

- $V(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$,
- $\forall a \in \mathbb{R}, V(a) = 0$,
- $\forall a \in \mathbb{R}, V(X + a) = V(X)$,
- $\forall a \in \mathbb{R}, V(aX + b) = a^2 V(X)$

Variance

Exemple du jeu `Tapis Vert`

Rappel

- $\Omega = \{(e_p, e_{xo}, e_{ca}, e_t) \mid e_i \in \{7, ..., as\}, i \in \{p, co, ca, t\}\}$, card $\Omega = 8^4 = 4096$,
- À tout tirage est associé un *gain* dans $\{-1, 1, 29, 999\}$
- $P(X = -1) = \frac{3773}{4096}, P(X = 1) = \frac{294}{4096}, P(X = 29) = \frac{28}{4096}, P(X = 999) = \frac{1}{4096}$
- $\mathbb{E}(X) = -\frac{1668}{4096}$

La dispersion des valeurs des *v.a.* autour du gain

$$\begin{aligned} V(X) &= 1 \times \frac{3773}{4096} + 1 \times \frac{294}{4096} + 29^2 \times \frac{28}{4096} + 999^2 \times \frac{1}{4096} - \left(\frac{1668}{4096}\right)^2 \\ &= 239.23 \end{aligned}$$

Autres moments

Définition

On définit les moments centrés d'ordre k d'une *v.a.* X d'espérance m:

$$\mu_k = \mathbb{E}\left[(X - m)^k\right]$$

Remarques

- $\mu_1 = 0$,
- $\mu_2 = V(X)$,
- μ_3 caractérise l'asymétrie de la distribution de X,
- μ_4 caractérise l'aplatissement de la distribution de X

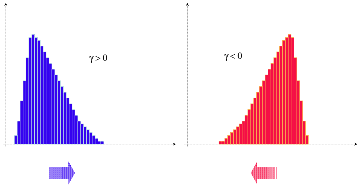
Skewness

Définition

Le coefficient d'asymétrie normalisé ou skewness est défini par

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

Un coefficient γ_3 positif, indique une queue de distribution étalée vers la droite, alors qu'un coefficient négatif indique un étalement vers la gauche.

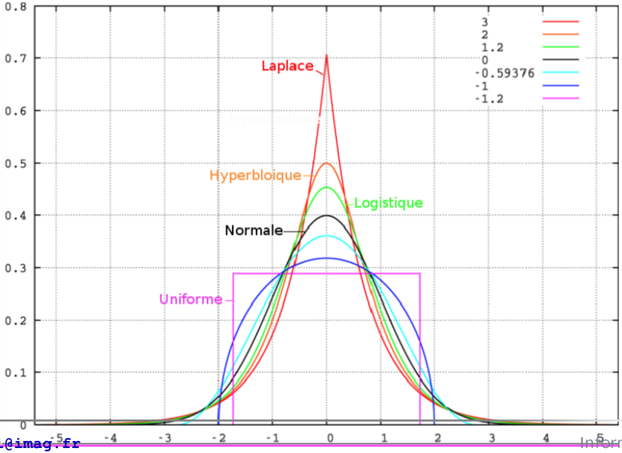


Kurtosis

Définition

Le coefficient d'aplatissement normalisé ou kurtosis est défini par

$$\gamma_2 = \frac{\mu_4}{\sigma^4}$$



Inégalité de Markov

Énoncé

Soit X une *v.a.* non négative, et a un réel strictement positif, on a alors

$$P(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Démonstration

$$aP(X \geq a) = a\int_a^\infty p(x)dx \leq \int_a^\infty xp(x)dx \leq \mathbb{E}(X)$$

Interprétation

Si $\mathbb{E}(X)$ est petit et que la *v.a.* X est positive alors, cette variable est proche de 0 avec une grande probabilité.

Inégalité de Chebychev

Énoncé

Soit X une *v.a.* d'espérance $\mathbb{E}(X)$ et de variance $V(X)$ et soit $\lambda \in \mathbb{R}_+$, on a alors

$$P(|X - \mathbb{E}(X)| \geq \lambda) \leq \frac{V(X)}{\lambda^2}$$

Démonstration

$\forall a \in \mathbb{R}_+$, l'événement $\{|X - \mathbb{E}(X)| \geq a\sigma(X)\}$ est équivalent à $\{(X - \mathbb{E}(X))^2 \geq a^2 V(X)\}$, et donc

$$P(|X - \mathbb{E}(X)| \geq a\sigma(X)) = P((X - \mathbb{E}(X))^2 \geq a^2 V(X)) \leq \frac{1}{a^2}$$

Interprétation

Si la variance de X est petite, la *v.a.* X est proche de son espérance avec une forte probabilité.

Inégalité de Chebychev

Exemple du jeu `Tapis Vert` *des pauvres*

Rappel

- $\Omega = \{(e_p, e_{xo}, e_{ca}, e_t) \mid e_i \in \{7, ..., as\}, i \in \{p, co, ca, t\}\}$, card $\Omega = 8^4 = 4096$,
- Si on change la donne: À tout tirage est associé un *gain* dans $\{-1, 1, 9, 99\}$
- $P(X = -1) = \frac{3773}{4096}, P(X = 1) = \frac{294}{4096}, P(X = 29) = \frac{28}{4096}, P(X = 999) = \frac{1}{4096}$
- $\mathbb{E}(X) = -\frac{1668}{4096} \approx -0.40$
- $Var(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X) \approx 251.22$

La probabilité d'avoir au moins trois cartes gagnantes est majorée par

$$P(X \geq 29) = P(X - \mathbb{E}(X) \geq 29.4) \leq \frac{251.22}{29.4^2} \approx 0.29$$

Le calcul direct donne

$$P(X \geq 9) = P(X = 29) + P(X = 999) = \frac{29}{4096} \approx 0.007$$

Covariance de deux *v.a.*

Définitions

On appelle <u>covariance</u> de deux variables *v.a.* X et Y, l'expression:

$$Cov(X,Y) = \mathbb{E}\left[(X-\mathbb{E}(X)) \times (Y-\mathbb{E}(Y))\right] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Propriétés

- $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$,
- Si deux variables X et Y sont centrées on a alors $Cov(X,Y) = \mathbb{E}(XY)$
- Si deux variables sont indépendants alors:
  - $Cov(X,Y) = 0$, la réciproque est fausse
  - $V(X+Y) = V(X) + V(Y)$

Coefficient de corrélation

Propriété

Soient deux variables *v.a.* X et Y définies sur un même espace probabilisé (Ω, C, P), on a

$$|Cov(X,Y)| \leq \sqrt{V(X)V(Y)}$$

De plus **l'égalité** a lieu si et seulement si Y et X sont linéairement dépendantes, ou si X est une constante

Définition

Soient deux variables *v.a.* X et Y définies sur un même espace probabilisé (Ω, C, P), on appelle le coefficient de corrélation la version normalisée de la covariance entre X et Y

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{V(X) \times V(Y)}}$$

On a dans ce cas

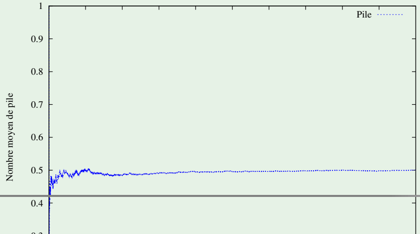
$$-1 \leq \rho(X,Y) \leq +1$$

Loi faible des grands nombres

Historique - Généralité

- Cette loi a été formalisée au $XVII^e$ siècle lors de la définition du cadre des probabilités,
- Cette loi indique que les *caractéristiques statistiques* d'un grand nombre de tirages aléatoires tendent vers les vrais *caractéristiques statistiques définissant l'échantillon*

Exemple: Pile ou Face avec une pièce non-truquée



Loi faible des grands nombres - Théorème de Khinchin

Énoncé

Soit n variables aléatoires indépendantes $X_1, \ldots, X_n$ même espérance $\mathbb{E}(X)$ et une variance finie $V(X)$,

$$\forall \epsilon > 0, \lim_{n \to +\infty} P\left(\left| \frac{X_1 + X_2 + \ldots + X_n}{n} - \mathbb{E}(X) \right| \geq \epsilon \right) = 0$$

Démonstration

L'espérance et la variance de la *v.a.* Y définie par $Y = \frac{X_1 + X_2 + \ldots + X_n}{n}$ sont $\mathbb{E}(Y) = \mathbb{E}(X)$ et $V(Y) = \frac{V(X)}{n}$. Le résultat du théorème est alors une application de l'inégalité de Chebychev appliqué à Y:

$$\forall \epsilon > 0, P(|Y - \mathbb{E}(X)| \geq \epsilon) \leq \frac{V(X)}{n\epsilon^2}$$

Loi faible des grands nombres

Utilisation

- C'est sur cette loi que reposent le fondement des *sondages*: la tendance générale est bien approximée par la tendance exprimée par un sous-ensemble assez grands des individus tirés aléatoirement dans la population
- De même sans la formulation de cette loi, les assurances n'auraient connu un tel essor: cette loi permet en effet d'approximer les probabilités des sinistres assurés par les ces compagnies
- En outre, cette loi soulève une question métaphysique si les éléments du *hasards* se comportent individuellement d'une manière chaotique, dans l'ensemble ils ont un comportement stable (i.e. le chaos généralisé est impossible!)

Lois de probabilités discrètes

Loi Uniforme - définition

Soit X une *v.a.* prenant chaque valeur de l'ensemble $\{1, ..., n\}$ la loi de X est dite uniforme sur $[\![1, n]\!]$ si

$$\forall k \in [\![1, n]\!], P(X = k) = \frac{1}{n}$$

Moments

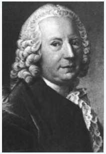
- $\mathbb{E}(X) = \frac{n+1}{2}$
- $V(X) = \frac{n^2-1}{12}$

Domaine d'application

Cette loi intervient dans l'étude de divers jeux (pile ou face, dès, loterie, ..), dans les sondages → théorème de Khinchin

Lois de probabilités discrètes

Loi de Bernoulli - définition



C'est la loi d'une *v.a.* binaire $X \in \{0,1\}$
Si on appelle q la probabilité que X soit égale à 1

$$\forall x \in \{0,1\}, P(X = x) = q^x(1-q)^{1-x}$$

Moments

- $\mathbb{E}(X) = q$
- $V(X) = q(1-q)$

Domaine d'application

Cette loi décrit les phénomènes où il y a que deux valeurs possibles à prendre, par exemple en apprentissage automatique pour modéliser l'erreur d'un modèle.

Lois de probabilités discrètes

Loi binomiale - définition

Soit X la somme de n variables X_i indépendantes, suivant toutes la loi de Bernoulli de paramètre q. La loi suivie par X est la loi binomiale de paramètres n et q, notée $\mathcal{B}(n,q)$

$$\forall k \in \{1, ..., n\} P(X = k) = C_n^k q^k (1-q)^{n-k}$$

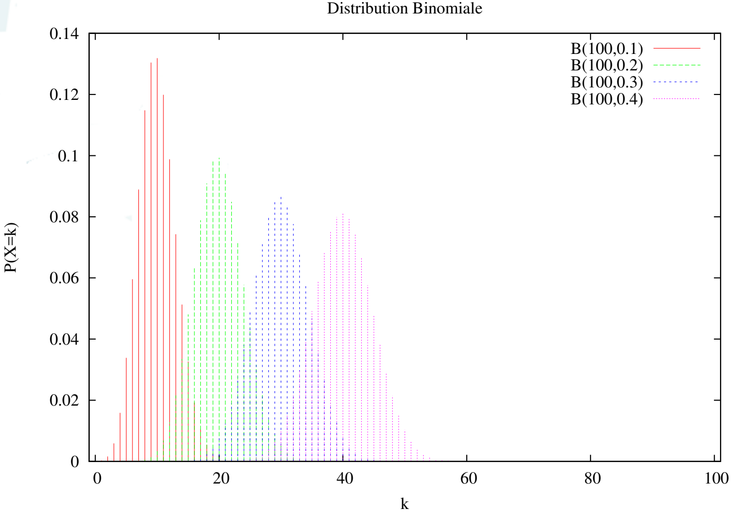
Moments

- $\mathbb{E}(X) = nq$
- $V(X) = nq(1-q)$

Domaine d'application

Cette loi décrit des phénomènes formés de répétitions d'une même épreuve ayant deux possibilités s'excluant mutuellement

Loi de Binomiale



Loi de Binomiale

Exemple

Soit une étude portant sur des familles de 10 enfants. On sait que la probabilité d'avoir une fille ou un garçon est identique.

- La probabilité que 2 des enfants d'une famille soient des filles

$$P(X=2) = C_{10}^{2}\frac{1}{2^{10}} \approx 0.0439$$

- Qu'une famille ait au moins deux filles

$$P(X \geq 2) = \frac{1}{2^{10}}\sum_{k=2}^{10} C_{10}^{k} = 1 - P(X<2) = 1 - P(X=0) - P(X=1)$$

- Qu'une famille n'ait pas plus de deux garçons

$$P(X \geq 8) = P(X=8) + P(X=9) + P(X=10) \approx 0.0547$$

Lois de probabilités discrètes

Loi de Poisson - définition

Sur une période T, un événement arrive en moyenne λ fois. La loi suivit par la *v.a.* déterminant le nombre de fois où l'événement se produit dans la période T est:

$$\forall k \in \mathbb{N}, P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Moments

- $\mathbb{E}(X) = \lambda$
- $V(X) = \lambda$

La loi de Poisson est la seule loi discrète vérifiant $\mathbb{E}(X) = V(X)$

Domaine d'application

La loi de Poisson est la loi des petites probabilités ou loi des événements rares. On l'utilise dans les réseaux informatiques ou pour décrire les accidents dans un atelier de grande taille.

Loi de Poisson

Exemple

Le tableau ci-dessous répertorie le nombre d'accidents du personnel d'une usine sur une période de 200 jours

# accidents	0	1	2	3	4	5
# de jours	86	82	22	7	2	1

Les accidents sont survenus indépendamment les uns des autres, on peut supposer qu'ils suivent une loi de Poisson de paramètre $\mathbb{E}(X)$. Dans ce cas, l'estimation du nombre de jours où qu'il s'est produit moins de 3 accidents est:

$$200 * P(X < 3) \approx 190$$

Lois de probabilités continues

Loi uniforme

Une *v.a.* réelle X suit une loi uniforme sur l'intervalle $[0,a]$ si sa loi de probabilité admet pour densité

$$p(x) = \begin{cases} \frac{1}{a} & \text{si } x \in [0,a], \\ 0 & \text{si } x \in]-\infty, 0[\cup]a, \infty[\end{cases}$$

Moments

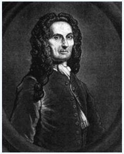
- $\mathbb{E}(X) = \frac{a}{2}$
- $V(X) = \frac{a^2}{12}$

Domaine d'application

La loi uniforme traduit l'hypothèse d'équipartition sur l'intervalle

Lois de probabilités continues

Loi de Laplace-Gauss ou Loi normale



Une *v.a.* réelle X suit une loi de Laplace-Gauss de paramètres m et σ si sa densité a pour expression

$$\forall x \in \mathbb{R}, p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

Moments

- $\mathbb{E}(X) = m$
- $V(X) = \sigma^2$

Domaine d'application

Cette loi joue un rôle fondamental en statistique et elle apparaît comme loi limite de certaines caractéristique d'échantillons de grandes tailles. Son domaines d'application est variée: apprentissage automatique, médecine, ...

Loi de Laplace-Gauss ou Loi normale