Probabilité et Statistiques

I – Introduction

Richard Wilson

(richard.wilson@sorbonne-universite.fr)

LATMOS/IPSL
Sorbonne Université

13 octobre 2020

Sommaire

Introduction

Définitions

- Le mot **statistique** dérive du latin **status** qui signifie «état».
- Le mot «statistique» est employé à double sens :
  - **LA** ou **LES** statistique(s) recouvre un ensemble de méthodes destinées à l'analyse, à l'interprétation et à la présentation de données.
  - **UNE** statistique est un nombre calculé à partir d'observations. Par exemple :
    - √ la moyenne ou la fréquence d'occurrence d'un évènement sont des statistiques.
    - √ On parle fréquemment des statistiques du chômage
- La statistique est pour les uns un domaine des mathématiques, pour les autres (en particulier les anglo-saxons) une discipline à part entière hors des mathématiques.
- De plus en plus, elle fait partie de ce que l'on appelle la science des données (Data Science)

Historique I

- Introduit au dix-huitième siècle (vers 1785) par l'économiste allemand Gottfried Achenwall, le mot «Statistik» est dérivé de l'italien «statista» (« homme d'État »),
- la statistique représentait pour cet auteur l'ensemble des connaissances que doit posséder un homme d'état (ressources, évolution des prix, démographie).
- Si le mot est récent, la pratique est aussi ancienne que les sociétés humaines.
  - √ Les premiers textes écrits retrouvés (Sumériens) étaient des recensements du bétail et des informations sur les prix en cours, bref des listes d'objets et des nombres.
  - √ On a ainsi trace de recensements des récoltes en Chine au XXIII^e siècle av. J.C. (empereur Yao), et de la population en Égypte au XVIII^e siècle av. J.C. (pharaon Amasis) (`http://www.statistix.fr/IMG/pdf/Une_approche_historique_de_la_statistique_v3.pdf`)

Historique II

- La statistique mathématique moderne, basée sur le calcul des probabilités, est née au XIX^e siècle suite aux travaux de Laplace, Gauss et Moivre.
- La statistique connaît un développement fulgurant au tournant du vingtième siècle en Angleterre sous l'impulsion décisive de Francis Galton, Karl et Egon Pearson (le père et le fils), Ronald Fisher et William Gosset.

William Sealy Gosset (13 juin 1876 – 16 octobre 1937) connu sous le pseudonyme Student est un statisticien anglais. Embauché par la brasserie Guinness pour stabiliser le goût de la bière, il a le premier décrit la distribution dite de Student. Le test d'hypothèse de Student, construit sur cette distribution, est aujourd'hui utilisé de manière standard pour mettre en évidence une éventuelle différence entre deux échantillons.

Historique III

- Au vingtième siècle, les techniques et méthodes statistiques se sont propagées et imposées dans tous les domaines scientifiques :
  - √ sciences «dures» (physique, astronomie, géophysique),
  - √ sciences du vivant (biologie, agronomie, médecine),
  - √ sciences économiques et sociales (sociologie, psychologie, histoire, économétrie),
- et bien au delà des sciences
  - √ publicité, marketing ;
  - √ production : sécurité (pannes, risques), contrôle de qualité ;
  - √ politique.
- Les méthodes statistiques sont devenues des outils d'aide à la décision (investissement financier, prévention des risques,...) voire des moyens opérationnels de gestion (files d'attente, circulation automobile, distribution d'électricité,...).
- Les progrès récents de l'informatique ont permis le développement de méthodes statistiques nouvelles, permettant la manipulation de très grandes quantités de données (exploration/fouille de données, data mining).

Les deux branches de la statistique

Il y a deux grandes branches de l'analyse statistique :

1. la **statistique descriptive** consistant au traitement, à la classification et à mise en forme des données. Un échantillon sera résumé en quelques statistiques, un nuage de points sera projeté sur un système d'axes mettant en évidence des sous-groupes, etc...
2. la **statistique inférentielle** dont l'objet est la caractérisation des propriétés d'une population à partir d'échantillons de celle-ci.

Les méthodes de la statistique inférentielle reposent sur la théorie des probabilités, pas celles de la statistique descriptive.

La statistique descriptive

Le but de la statistique **descriptive** (ou **exploratoire**) est de structurer les données issues de l'échantillonnage afin d'en extraire des informations pertinentes.
C'est **l'analyse des données** dont les principales méthodes sont :

√ Analyse univariée

- L'objectif sera de décrire la distribution des valeurs prises par une variable caractérisant un échantillon (un sondage, des mesures physiques, des caractères qualitatifs).
- Une telle description repose sur l'estimation des quelques statistiques résumant les données,
  - ◇ tendance centrale (moyenne, médiane, mode)
  - ◇ variabilité (déviation standard, étendue, asymétrie (skewness) aplatissement (kurtosis))
  - ◇ distribution des valeurs (quantiles, étendue).
- et sur des représentations graphiques (histogrammes).

Si les individus d'un échantillon sont caractérisés par plus d'une variable, la statistique descriptive s'attache à décrire les relations entre variables.

√ Analyse multivariée

- Les tableaux de contingence ;
- Les graphiques de dispersion (scatter plots) visant à synthétiser les propriétés d'un échantillon ou d'une population :
- La mesure quantitative d'un lien de dépendance entre variables (corrélation, axe principal)
- Les méthodes factorielles ayant pour but de réduire le nombre de variables à l'aide d'un petit nombre de facteurs, comme par exemple l'analyse en composantes principales.

Le calcul des probabilité ne joue ici aucun rôle.

La statistique inférentielle

La statistique inférentielle

Le propos de la **statistique inférentielle** est de déduire les propriétés d'un ensemble, on dira (pour des raisons historiques) d'une population, à partir de la connaissance d'échantillons de cet ensemble.

L'échantillon peut résulter d'une série de mesures, d'un prélèvement aléatoire, d'un sondage.

- série temporelle de mesures météorologiques au parc Montsouris (1872–)
- prélèvement aléatoire de pièces manufacturées (contrôle de qualité)
- sondage d'opinion.

Les outils de base sont ici les méthodes d'estimation et les tests d'hypothèse.

La théorie des probabilités joue ici un rôle central.

Voici quelques exemples :

Exemple 1 : Estimation d'une grandeur à partir d'une série de mesures

- Une grandeur physique ξ est mesurée n fois. À cause des imprécisions de mesure ou à cause de la variabilité de facteurs ayant une influence sur la quantité mesurée, on recueille une suite de résultats $x_1, x_2, \ldots, x_n$, à priori différents.
- La mesure est une variable aléatoire (v.a.), X. L'objectif est d'estimer une valeur pertinente de la grandeur mesurée ξ à partir des $x_1, \ldots, x_n$.

La v.a. X suit une loi de probabilité dépendant de paramètres que l'on cherche à estimer.

La question centrale de l'estimation est d'estimer des paramètres de la loi suivie par la v.a. X à partir de la connaissance d'un échantillon issu d'un processus aléatoire (mesures, tirage).

On s'attache donc à décrire les propriétés de la v.a. (tendance centrale, variabilité,...) à partir d'estimateurs déduits de l'échantillon.

Exemple 1

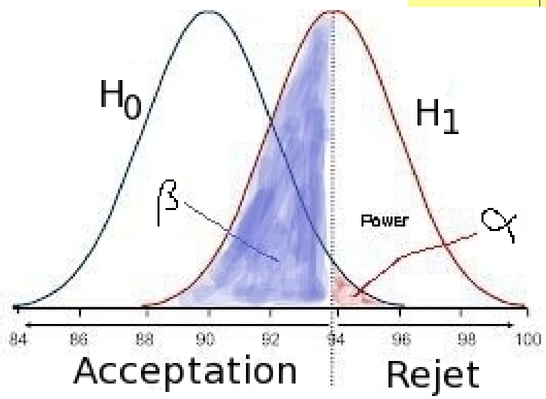
- Une estimation intuitive xi est la moyenne arithmétique $\overline{X}$ des X, i.e. $\boxed{\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$. La moyenne arithmétique est un estimateur de μ_X.
- Chaque mesure constitue une réalisation du processus de mesure. Une autre série de mesure donnera, à priori, une autre moyenne empirique $\overline{x}$, c.à.d. une autre estimation.
- L'estimateur $\overline{X}$ est donc une variable aléatoire dont on cherchera à préciser les propriétés.
- En particulier on cherchera un intervalle de confiance pour μ_X : $[\overline{X} - \Delta\overline{X}; \overline{X} + \Delta\overline{X}]$ avec un risque d'erreur que l'on quantifiera.

Exemple 2 : Test d'hypothèse

- Il s'agit ici de confirmer ou d'infirmer une hypothèse à partir de l'observation d'un échantillon constitué de v.a..
- Considérons par exemple un processus dépendant du temps (une série de mesure de température par exemple). On cherche à savoir si le processus est stationnaire, c'est à dire si sa valeur moyenne et sa variance dépendent ou non de l'intervalle de temps considéré.
- Une méthode consiste à estimer des moyennes et des variances sur des intervalles de temps consécutifs. On évaluera la pertinence de l'hypothèse de stationnarité en comparant les distributions des moyennes/variances obtenues.
- Le test d'hypothèse consiste à confronter deux hypothèses,
  - une hypothèse dite nulle, l'hypothèse H_0, sous laquelle les propriétés de l'échantillon sont connues, ici l'hypothèse stationnaire ;
  - l'hypothèse alternative non-H_0, notée H_1.
- Un test d'hypothèse permettra de prendre une décision relative à H_0, i.e. rejeter ou non H_0, avec un risque quantifié de se tromper.

Exemple 2

réel \ décision	Ho acceptée	Ho rejetée
Ho vraie	1−α	α
Ho fausse	β	1-β (puissance)



Principe du test d'hypothèse

Quelques rappels de probabilités

Probabilité : définition

On définit successivement :

- Une expérience ou un processus aléatoire : une expérience, un processus dont l'issue est imprévisible (mesure, tirage, marche aléatoire).
- L'univers des possibles : l'ensemble Ω contient tous les résultats ω possibles d'une expérience aléatoire.
- Un évènement, ou une réalisation, A est un sous-ensemble de Ω
- Une probabilité est une application associant un scalaire $\in [0, 1]$ à tout évènement A. La probabilité de l'évènement A est notée $\Pr[A]$.
- **exemple :** Le jet d'un dé à six faces. L'univers des possibles est $\Omega = \{1, 2, 3, 4, 5, 6\}$. Un évènement est, par exemple, $A = \{\text{Le résultat d'un jet de dé est pair}\}$, soit $A = \{2, 4, 6\}$.

Probabilité : propriétés élémentaires

- $\Pr[\Omega] = 1$
- $\Pr[\varnothing] = 0$
- Soient A et $B \in \Omega$: $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$
- Si $A \subseteq B$: $\Pr[A] \leqslant \Pr[B]$

Définition : système complet d'évènements

Un ensemble A_i ($i \in \mathbb{N}$) d'évènements constitue un **système complet** si les A_i sont deux à deux disjoints et si leur réunion est Ω.

1. $A_i \cap A_j = \varnothing$ si $i \neq j$
2. $\displaystyle\bigcup_{i \in \mathbb{N}} A_i = \Omega$

Probabilités conditionnelles

On dit que la probabilité de l'évènement A est conditionnée à l'évènement B si la probabilité de A est modifiée selon que l'évènement B est ou non réalisé. La probabilité de A conditionné à B est notée $\Pr[A|B]$.

Considérons par exemple deux évènements A et B issus d'une expérience aléatoire, chacun des évènements et son complémentaire formant un système complet. On considère les résultats du couple (A, B). Les probabilités de chaque occurrence (A_i, B_j) sont indiquées sur le tableau.

	B	$\overline{B}$
A	p_{11}	p_{12}
$\overline{A}$	p_{21}	p_{22}

Les quatre évènements de probabilité p_{ij} forment un système complet d'évènements.

Il vient :
$$\begin{cases} \Pr[A \cap B] \equiv \Pr[AB] \equiv \Pr[A\&B] = p_{11} \\ \Pr[A] = \Pr[A\&B] + \Pr[A\&\overline{B}] = p_{11} + p_{12} \\ \Pr[A|B] = \dfrac{\Pr[A\&B]}{\Pr[B]} = \dfrac{p_{11}}{p_{11}+p_{21}} \end{cases}$$

$$\Rightarrow \quad \boxed{\Pr[A\&B] = \Pr[A|B] \times \Pr[B] = \Pr[B|A] \times \Pr[A]}$$

Théorème de Bayes

En conséquence :

$$\Pr[A|B] = \frac{\Pr[B|A]\Pr[A]}{\Pr[B]} = \frac{\Pr[B|A]\Pr[A]}{\Pr[B|A]\Pr[A] + \Pr[B|\overline{A}]\Pr[\overline{A}]}$$

Ce résultat se généralise à un ensemble complet d'évènements A_i.

Théorème de Bayes

Soit A_i , $i = (1, \ldots, n)$ un système complet de Ω tel que $\forall i, \Pr[A_i] \neq 0$. Alors, pour tout $B \in \Omega$:

$$\boxed{\Pr[A_i|B] = \frac{\Pr[B|A_i]\Pr[A_i]}{\Pr[B]} = \frac{\Pr[B|A_i]\Pr[A_i]}{\sum_{i=1}^{n}\Pr[B|A_i]\Pr[A_i]}}$$

Exemple : Test de dépistage

Un test servant à dépister une maladie vient d'être mis au point. Ses résultats, annonce le fabriquant, sont très fiables :

- si le sujet est malade, le test est positif dans 95% des cas ;
- si le sujet n'est pas malade, le test donne un résultat négatif dans 9 cas sur 10.

Sachant que la maladie en question ne frappe que 1% de la population, quelle est la probabilité pour un patient dont le test est positif d'être effectivement atteint ? Posons :

- T : le test est positif,
- M : le sujet est malade.

On connaît $\Pr[M] = 0,01$ et on sait que $\Pr[T|M] = 0.95$ et $\Pr[\overline{T}|\overline{M}] = 0,9$.

On cherche $\Pr[M|T]$: $\boxed{\Pr[M|T] = \dfrac{\Pr[T|M]\Pr[M]}{\Pr[T]}}$

Or : $\Pr[T] = \Pr[T|M]\Pr[M] + \Pr[T|\overline{M}]\Pr[\overline{M}]$, où $\Pr[T|\overline{M}] = (1 - \Pr[\overline{T}|\overline{M}])$

Finalement : $\boxed{\Pr[M|T] = \dfrac{\Pr[T|M]\Pr[M]}{\Pr[T|M]\Pr[M] + (1-\Pr[\overline{T}|\overline{M}])\Pr[\overline{M}]} = 0,088}$

La probabilité d'être malade sachant le test positif est moins de 9% ! !

Évènements indépendants

Deux évènements A et B sont indépendants si la probabilité d'occurrence de A ne dépend pas de celle de B :

$$\Pr[A|B] = \Pr[A], \quad \Longrightarrow \quad \Pr[A \cap B] = \Pr[A|B]\Pr[B] = \Pr[A]\Pr[B]$$

De même, la probabilité de B ne dépend pas de A ;

$$\Pr[B|A] = \Pr[B], \quad \Longrightarrow \quad \Pr[A \cap B] = \Pr[A]\Pr[B]$$

En résumé, deux évènements sont indépendants si et seulement si :

$$\Pr[A \cap B] = \Pr[A]\Pr[B]$$

Cette notion d'indépendance se généralise à un nombre quelconque d'évènements.

Théorème : Indépendance d'évènements aléatoires

Soit $A_1, \ldots, A_N$ une suite de N évènements aléatoires. Les A_i sont indépendants si et seulement si :

$$\Pr\left[\bigcap_{i=1}^{N} A_i\right] = \prod_{i=1}^{N}\Pr[A_i] \qquad (1)$$

Variable aléatoire

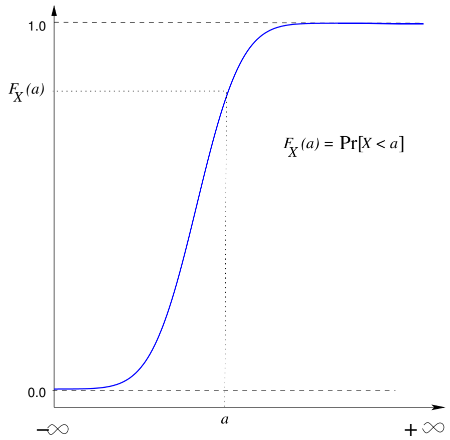
On considère ici des variables aléatoires (v.a.) quantitatives.

- **Notations** – Par convention, on note en majuscule une variable aléatoire, et en minuscule une valeur prise par cette variable.
  Ainsi, la notation : $\Pr[X = x] = \alpha$ signifie : la probabilité que la variable aléatoire X prenne la valeur particulière x est α.
- Une v.a. peut être discrète (l'ensemble des valeurs prise par X est dénombrable) ou continue.
- Les propriétés d'une v.a. sont décrites par sa Fonction de répartition

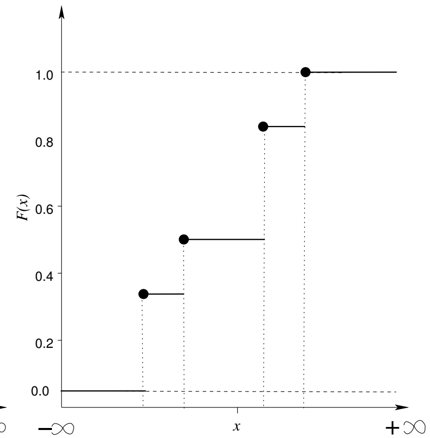
La **fonction de répartition** d'une v.a. X, notée F_X est une fonction $\mathbb{R} \to [0, 1]$ ayant pour valeur en $x \in \mathbb{R}$ la probabilité que $X \leqslant x$.

$$\boxed{F_X(x) \stackrel{\text{def}}{=} \Pr[X \leqslant x]}$$

Fonction de répartition



Fonctions de répartition, v.a. continue



Fonction de répartition, v.a. discrète

Densité de probabilité

La densité de probabilité (ddp ou pdf pour probability density function) d'une variable aléatoire X décrit la probabilité pour la v.a. d'être dans un intervalle donné (v.a. continue) ou de prendre une valeur particulière (v.a. discrètes).
La ddp s'obtient en dérivant la fonction de répartition F_X.
Il convient de distinguer les cas ou la v.a. est continue ou discrète.

- **densité de probabilité v.a. continue**
  La densité de probabilité est une fonction, notée f_X, définie par :

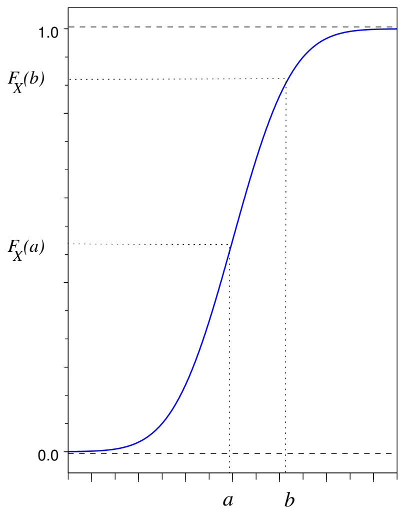
$$\Pr[X \leqslant x] = F_X(x) = \int_{-\infty}^{x} f_X(t)\,\mathrm{d}t$$

Autrement dit, la fonction f_X est (presque partout) égale à la dérivée de F_X :

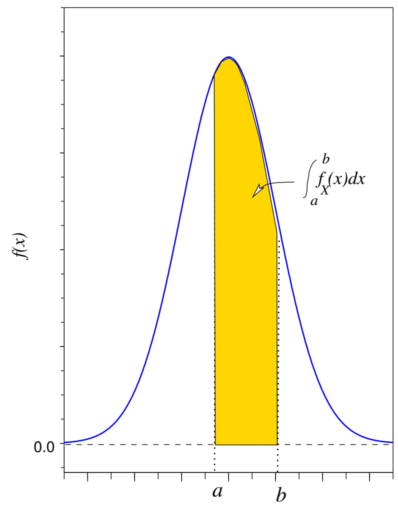
$$\boxed{f_X(x)\mathrm{d}x = \mathrm{d}F_X \iff f_X(x) = \frac{\mathrm{d}F_X}{\mathrm{d}x}}$$

Densité de probabilité

Fonction de répartition



Densité de probabilité



Densité de probabilité

Si la v.a. X est discrète, la fonction de répartition F_X est en escalier : elle est donc constante presque partout sauf aux points de discontinuité. Une telle fonction n'est pas dérivable au sens usuel des fonctions. Néanmoins, on définit une densité de probabilité à l'aide de la distribution (ou impulsion) de Dirac.

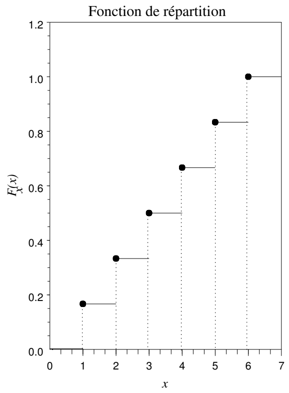
- **densité de probabilité v.a. discrètes**

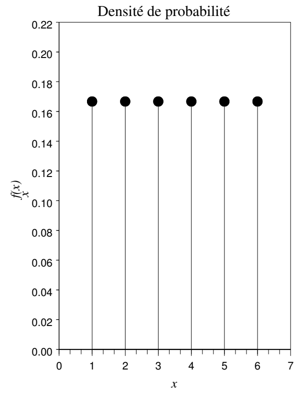
  La densité de probabilité f_X d'une v.a. discrète est nulle presque partout sauf aux points x_i, discontinuités de f : formellement, on écrit une telle densité de probabilité à l'aide d'un ensemble d'impulsions de Dirac $\delta(x - x_i)_{i \in \mathbb{N}}$:

$$f_X(x_i) = p_i \delta(x - x_i) \qquad \Longrightarrow \qquad \boxed{f_X(x) = \sum_{i=1}^{N} p_i \delta(x - x_i)} \qquad (2)$$

Densité de probabilité d'une v.a. discrète

La fonction de répartition (à gauche) ainsi que la densité de probabilité (à droite) de la v.a. discrète associée à un jet de dé non-pipé (chaque résultat est équiprobable). $X = (1, 2, \cdots, 6)$ et $\Pr[X = x_i] = 1/6$. Les impulsions de Dirac ($\times 1/6$) sont représentées par des bâtons de hauteur $1/6$.





Condition de normalisation

$$\boxed{\int_{\Omega} f_X(x)\mathrm{d}x = 1}$$

v.a. continue

$$\boxed{\sum_{i=1}^{N} p_i = 1 \qquad (1 \leqslant i \leqslant N)}$$

v.a. discrète

Moments d'une variable aléatoire

La fonction de répartition et la densité de probabilité décrivent la distribution des probabilités de la variable aléatoire X sur l'ensemble des valeurs que celle-ci peut atteindre.

On peut aussi résumer les propriétés d'une v.a. à l'aide de quelques statistiques : les **moments** dont les plus importants sont le premier et le second moments.

Le premier moment décrit la moyenne statistique de X, le second moment (centré) caractérisant la dispersion autour de cette moyenne.

- **Espérance**

  Le premier moment, appelé espérance mathématique de X, ou plus simplement **espérance de** X, caractérise la moyenne statistique :

$$\boxed{\mathrm{E}[X] \equiv m_1 \stackrel{\text{def}}{=} \int x f_X(x)\mathrm{d}x} \qquad \boxed{\mathrm{E}[X] = \sum_{i=1}^{N} x_i p_i}$$

Moments d'une variable aléatoire

D'une façon générale, le **moment d'ordre n ($n \in N^*$)** est défini par :

$$\boxed{\mathrm{E}\left[X^n\right] \stackrel{\mathrm{def}}{=} m_n = \int x^n f_X(x)\mathrm{d}x}$$

Le moment d'ordre n ($n \geq 1$) n'est pas toujours défini, (l'intégrale ci-dessus ne convergeant pas nécessairement)

- **Moments centrés**

  On définit les moments centrés par :

$$\boxed{\mu_n \stackrel{\mathrm{def}}{=} \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^n\right] = \int (x - m_1)^n f_X(x)\mathrm{d}x}$$

- **La variance**

  La variance de la v.a. X, notée $\mathrm{Var}\left[X\right]$ ou σ_X^2, est le moment centré d'ordre 2 :

$$\boxed{\mathrm{Var}\left[X\right] \equiv \mu_2 \stackrel{\mathrm{def}}{=} \mathrm{E}\left[(X - \mathrm{E}\left[X\right])^2\right] = \int (x - m_1)^2 f_X(x)\mathrm{d}x}$$

Écart type, v.a. centrées réduites

- Écart type

  La dispersion des valeurs d'une v.a. autour de l'espérance est évaluée par l'**écart type**, σ_X, défini comme la racine carrée de la variance :

$$\sigma_X = \sqrt{\mathrm{Var}\left[X\right]}$$

- Variables centrées réduites

  On est fréquemment conduit à utiliser des variables centrées réduites.

  Une **variable aléatoire réduite** est une v.a. d'espérance nulle et d'écart type 1.

- Ainsi, si X est une v.a. de moyenne m et de variance σ_X^2, on définit la v.a. centrée-réduite Y par :

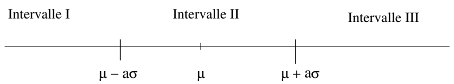
$$\boxed{Y = \frac{X - m}{\sigma_X}}$$

Inégalité de Bienaymé-Chebychev

Les amis Bienaymé et Chebychev (ils entretenaient une correspondance) nous ont légué un important résultat.

L'inégalité de Bienaymé-Chebychev permet de quantifier la probabilité qu'a une v.a. de s'écarter de l'espérance de plus d'une quantité donnée.

Considérons une v.a. X de moyenne μ et d'écart type σ.
On distingue 3 intervalles :



I : $X < \mu - a\sigma$

II : $|X - \mu| \leqslant a\sigma$

III : $X > \mu + a\sigma$

où $a \in \mathbb{R}^{*+}$. Notons que dans les régions I et III, $|X - \mu| \geq a\sigma$.

Inégalité de Bienaymé-Chebychev

Par définition de la variance :

$$V[X] = \underbrace{\int_{x<\mu-a\sigma} (x-\mu)^2 f_X(x)\mathrm{d}x}_{\text{Région I}} + \underbrace{\int_{|x-\mu|\leqslant a\sigma} (x-\mu)^2 f_X(x)\mathrm{d}x}_{\text{Région II}} + \underbrace{\int_{x>\mu+a\sigma} (x-\mu)^2 f_X(x)\mathrm{d}x}_{\text{Région III}}$$

Donc :

$$V[X] > \int_{x<\mu-a\sigma} (x-\mu)^2 f_X(x)\mathrm{d}x + \int_{x>\mu+a\sigma} (x-\mu)^2 f_X(x)\mathrm{d}x$$

Comme $|x - \mu| > a\sigma$ dans les régions I et III :

$$V[X] > \int_{x<\mu-a\sigma} a^2\sigma^2 f_X(x)\mathrm{d}x + \int_{x>\mu+a\sigma} a^2\sigma^2 f_X(x)\mathrm{d}x = a^2\sigma^2 \int_{|x-\mu|>a\sigma} f_X(x)\mathrm{d}x$$

La dernière intégrale est la probabilité que X soit dans les régions I ou III, c.-à-d. $\Pr[|X - \mu| > a\sigma]$.

Inégalité de Bienaymé-Chebychev

Il vient : $\sigma^2 > a^2\sigma^2 \Pr[|X-\mu| > a\sigma]$ On en déduit :

Théorème : Inégalité de Bienaymé-Chebychev

$$\boxed{\Pr[|X-\mu| > a\sigma] < \frac{1}{a^2}}$$

On peut écrire l'inégalité de Bienaymé-Chebychev sous une forme légèrement différente. Posons $\beta = a\sigma$. Il vient :

$$\boxed{\Pr[|X-\mu| > \beta] < \frac{\sigma^2}{\beta^2}}$$

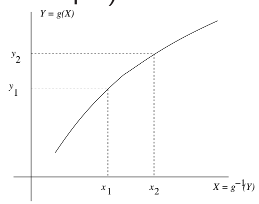
Fonction d'une variable aléatoire

- Étant donnée une v.a. X de densité f_X, on est souvent amené à considérer la v.a. Y liée à X par une fonction connue, g :

$$Y = g(X)$$

  Comment exprimer la densité de probabilité (d.d.p.) de Y connaissant celle de X ?

- Si g est une fonction monotone (croissante par exemple) :



  $\Pr[Y \leqslant y_1] = \Pr[X \leqslant x_1]$ soit : $F_Y(y_1) = F_X(x_1)$

  où $y_1 = g(x_1)$, F_Y et F_X sont les fonctions de répartition de Y et de X.

  On déduit :

$$\left(\frac{\mathrm{d}F_Y}{\mathrm{d}y}\right)_{y=y_1}\mathrm{d}y = \left(\frac{\mathrm{d}F_X}{\mathrm{d}x}\right)_{x=x_1}\mathrm{d}x \quad \text{soit :} \quad \boxed{f_Y(y_1)\mathrm{d}y = f_X(x_1)\mathrm{d}x}$$

Fonction d'une variable aléatoire

- Par conséquent, la d.d.p. de la v.a. $Y = g(X)$, g étant une fonction monotone, s'écrit :

$$\boxed{f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|}$$

- Espérance d'une fonction d'une variable aléatoire

  Comme $f_Y(y)\mathrm{d}y = f_X(x)\mathrm{d}x$, $E[Y]$ s'écrit :

$$\boxed{E[Y] = \int y f_Y(y)\mathrm{d}y = \int g(x) f_X(x)\mathrm{d}x}$$

  Ce résultat est important car il permet de calculer l'espérance de $\mathrm{E}[Y] = \mathrm{E}[g(X)]$ sans avoir à évaluer la d.d.p. de Y.

Lois de probabilité

Introduisons maintenant les lois de probabilité les plus couramment rencontrées. C'est une véritable ménagerie ! Nous ne décrirons que les plus importantes d'entre elles.

- Une loi de probabilité décrit la probabilité qu'une v.a. se trouve dans un intervalle donné (v.a. continue) ou prenne une valeur particulière (v.a. discrète).
- Elle est décrite par une fonction définie pour l'ensemble des valeurs qu'une variable aléatoire peut atteindre, soit
  - la fonction de répartition
  - la densité de probabilité, dérivée de la fonction de répartition.

  Cette fonction (distribution) dépend d'un ou plusieurs paramètres de forme.
- **Notations**

  Soit X une v.a. suivant une loi $\mathcal{L}$, cette loi dépendant des paramètres α et $\beta \ldots$. On notera :

$$\boxed{X \sim \mathcal{L}(\alpha, \beta, \ldots)}$$

Loi de Bernoulli

- Loi de Bernoulli
  La loi de Bernoulli correspond à un évènement aléatoire dont l'issue est binaire, par exemple un lancer de pile ou face, ou plus généralement toute expérience aléatoire se concluant par un succès ou un échec.
- La v.a. X prend la valeur 1 en cas de succès, 0 en cas d'échec.
- Si une v.a. suit une loi de Bernoulli, on note $X \sim \mathcal{B}(1,p)$. La loi de Bernoulli dépend du seul paramètre p, probabilité de succès.

$$\begin{aligned} \Pr[X=1] &= p && \text{si succès} \\ \Pr[X=0] &= (1-p) = q && \text{si échec.} \end{aligned}$$

$$\boxed{X \sim \mathcal{B}(1,p) \iff f_X(x) = p\delta(x-1) + (1-p)\delta(x)}$$

- Moments

$$\boxed{\begin{aligned} \mathrm{E}[X] &= p \\ \mathrm{Var}[X] &= p(1-p) = pq \end{aligned}}$$

Loi binomiale

- Loi binomiale
  La v.a. X décrivant le nombre de succès après n épreuves de Bernoulli suit une **loi binomiale** : X est la somme de n v.a. binaires (0/1) indépendantes.
- La loi binomiale dépend de deux paramètres : n, le nombre d'épreuves binaires, et p, la probabilité de succès lors d'une épreuve.
- On note $X \sim \mathcal{B}(n,p)$
  La loi binomiale est décrite par :

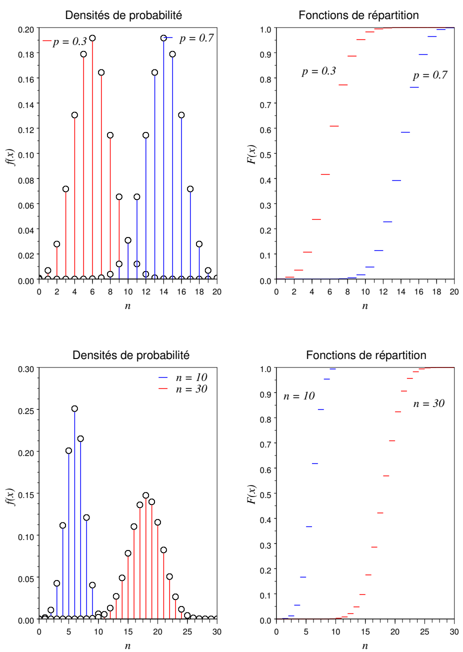
$$\Pr[X=k] = C_n^k\, p^k (1-p)^{n-k} \quad \forall k \in 0, 1, \ldots, n$$

- La densité de probabilité d'une v.a suivant une loi binomiale de paramètre (n,p) s'écrit :

$$\boxed{X \sim \mathcal{B}(n,p) \iff f_X(x) = \sum_{k=0}^{n} C_n^k\, p^k (1-p)^{n-k} \delta(x-k)}$$

Rappelons que : $C_n^k = \dfrac{n!}{k!(n-k)!}$

Loi binomiale



La d.d.p. d'une v.a. $X \sim \mathcal{B}(n,p)$.

En haut : $n = 20$, et deux valeurs de p, $p = 0{,}3$ et $p = 0{,}7$.

En bas : $p = 0{,}6$ et deux valeurs de n : $n = 10$ et $n = 30$.

Moments

$$\boxed{\begin{aligned} \mathrm{E}[X] &= np \\ \mathrm{Var}[X] &= np(1-p) = npq \end{aligned}}$$

Loi uniforme

- Une variable aléatoire discrète X comportant n issues x_i, toutes équiprobables suit une loi uniforme.

$$\Pr[X = x_i] = \frac{1}{n}$$

On note $X \sim \mathcal{U}(n)$.

- Densité de probabilité

$$\boxed{X \sim \mathcal{U}(n) \implies f_X(x) = \sum_{i=1}^{n} \frac{1}{n}\delta(x - x_i)}$$

- Moments

$$\boxed{\begin{aligned} \mathrm{E}[X] &= \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i \equiv m \\ \mathrm{Var}[X] &= \mathrm{E}[X^2] - E[X]^2 = \tfrac{1}{n}\textstyle\sum_{i=1}^n x_i^2 - m^2 \end{aligned}}$$

m est la moyenne arithmétique des x_i.

Loi de Poisson

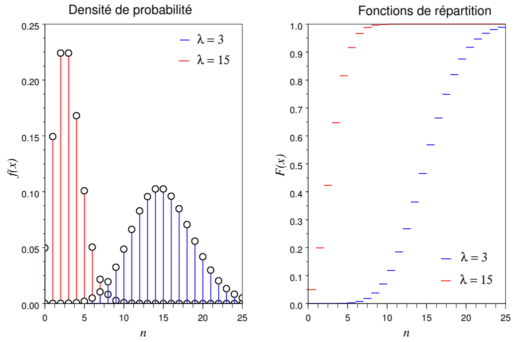
- La loi de Poisson est associée à des évènements «rares». Elle décrit la probabilité pour qu'un nombre donné d'évènements se produisent pendant un intervalle de temps ou dans une région de l'espace limitée.
- Une des premières applications de la loi de Poisson a permis d'estimer la probabilité du nombre de soldats tués par ruade de cheval dans la cavalerie Prussienne entre 1875 et 1894 !
- La loi suivie par la v.a. X dépend d'un paramètre λ, réel sans unité :

$$\boxed{\Pr[X=k] = \frac{\lambda^k}{k!}\exp(-\lambda) \qquad k \in \mathbb{N}, \qquad \lambda \in \mathbb{R}^{*+}}$$

- On dit que la v.a. X suit une loi de Poisson de paramètre λ, et on note :

$$\boxed{X \sim \mathcal{P}(\lambda)}$$

Loi de Poisson



La loi de Poisson pour $\lambda = 3$ et $\lambda = 15$

moments :

$$\boxed{\begin{aligned} \mathrm{E}[X] &= \lambda \\ \mathrm{Var}[X] &= \lambda \end{aligned}}$$

Loi uniforme

- Une variable aléatoire X suit une loi uniforme si la densité de probabilité est constante sur un intervalle de largeur donné.

$$\Pr[x \leqslant X \leqslant x + \mathrm{d}x] = C\mathrm{d}x \quad \text{où } C = \text{ Constante}$$

- Une variable de loi uniforme est définie sur un intervalle borné, c'est à dire :

$$X \in [a, b]$$

- La loi uniforme dépend alors de deux paramètres a et b, qui sont la plus petite et la plus grande valeur que la variable aléatoire peut prendre.
- Cette loi est notée $\mathcal{U}(a, b)$.

Loi uniforme (continue)

- Densité de probabilité

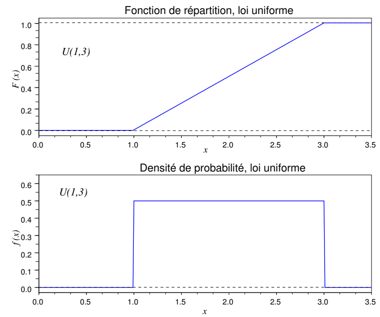
$$\boxed{X \sim \mathcal{U}(a,b) \quad \Longleftrightarrow \quad f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad a \leqslant x \leqslant b \\ 0 & \text{si} \quad x \notin [a,b] \end{cases}}$$

- Fonction de répartition

$$\boxed{F(x) = \begin{cases} 0 & \text{si} \quad x < a \\ \dfrac{x-a}{b-a} & \text{si} \quad a \leqslant x \leqslant b \\ 1 & \text{si} \quad x > b \end{cases}}$$

- Moments

$$\boxed{\begin{aligned} \mathrm{E}[X] &= \frac{(a+b)}{2} \\ \mathrm{Var}[X] &= \frac{(b-a)^2}{12} \end{aligned}}$$



Fonction de répartition (haut) et densité de probabilité (bas) d'une v.a. $X \sim \mathcal{U}(1,3)$.

Loi normale (Laplace-Gauss)

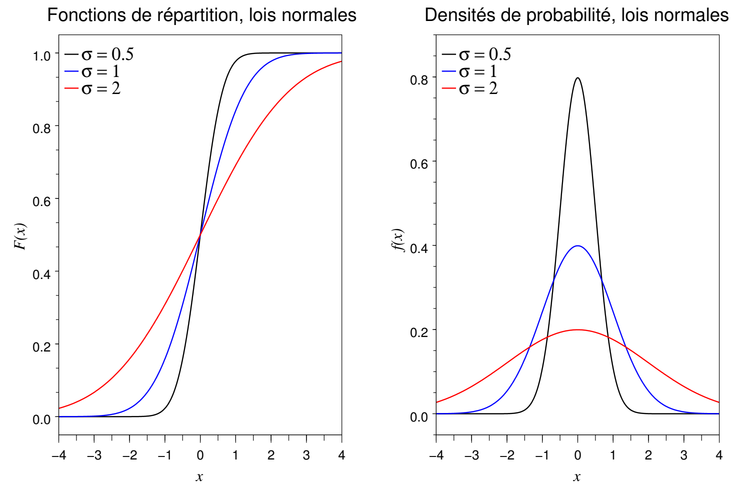
- La loi normale (ou loi de Laplace-Gauss, ou encore loi de Gauss) est sans doute la plus importante de toutes car de nombreuses variables aléatoires suivent (au moins approximativement) une loi normale (théorème de la limite centrale)
- Elle dépend de deux paramètres, l'espérance μ et l'écart type σ.
- Une v.a. suivant une loi normale de moyenne μ et d'écart type σ est notée $X \sim \mathcal{N}(\mu, \sigma^2)$.
- Densité de probabilité

$$\boxed{X \sim \mathcal{N}(\mu, \sigma^2) \iff f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

- Fonction de répartition

$$\boxed{F_X(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-\mu}{\sqrt{2}\sigma}\right)\right)}$$

Loi normale (Laplace-Gauss)



Les fonctions de répartition et les densités de probabilité de v.a. $X \sim \mathcal{N}(0, \sigma^2)$ pour trois valeurs de l'écart type σ.

Loi normale (Laplace-Gauss)

- Moments

$$\boxed{\begin{aligned} \mathrm{E}[X] &= \mu \\ \mathrm{Var}[X] &= \sigma^2 \end{aligned}}$$

- Les moments centrés d'ordre supérieur sont aisés à calculer pour une loi normale.
  - Les moments centrés d'ordre **impair** sont tous nuls.

$$\mu_{2k+1} = 0 \quad \forall k \in \mathbb{N}$$

  - Les moments centrés **pairs** s'expriment tous en fonction de la variance μ_2 :

$$\mu_{2k} = \frac{(2k)!}{2^k k!}\mu_2^k$$

cette dernière relation s'obtenant par récurrence.

Le fait que les moments centrés impairs soient tous nuls est vérifié pour toutes les lois symétriques par rapport à la moyenne, c'est à dire les lois telles que :

$$Pr[X \leqslant m_1 - x] = \Pr[X \geq m_1 + x]$$

les densités de probabilité des variables centrées sont alors des fonctions paires.

Loi normale centrée-réduite

- Une forme particulièrement utile de la loi normale est celle d'une v.a. centrée réduite.
- Soit $X \sim \mathcal{N}(\mu, \sigma^2)$.

  Alors la variable centrée réduite $\widetilde{X} = \dfrac{X-\mu}{\sigma}$ suit une loi $\mathcal{N}(0,1)$.

$$\boxed{X \sim \mathcal{N}(\mu, \sigma^2) \implies \widetilde{X} = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)}$$

- Table de la loi normale

  Les probabilités de la loi normale pour une v.a. centré-réduite sont tabulées. La table ci-dessous contient les valeurs de la fonction de répartition, à savoir les valeurs de :

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\mathrm{d}u$$

où f_X est la d.d.p. d'une v.a. centrée réduite :

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

- Il est aisé de déduire de cette table les probabilités pour n'importe quelle v.a. normalement distribuée. Considérons une v.a. $Y \sim \mathcal{N}(m, \sigma^2)$.
- La probabilité $Pr[Y \leqslant y] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-m)^2}{2\sigma^2}} du$

  s'obtient en effectuant le changement de variable $X = \dfrac{Y-m}{\sigma}$

  En effet,

$$Pr[Y \leqslant y] = \Pr[X \leqslant x] = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

  où $x = (y-m)/\sigma$

exemple : Soit $Y \sim \mathcal{N}(5,4)$. Quelle est la probabilité $\Pr[Y \leqslant 6{,}5]$? Il suffit de rechercher la probabilité de la v.a. X centrée-réduite :

$$\Pr[Y \leqslant 6,5] = \Pr\left[X \leqslant \frac{6,5-5}{2}\right] = \Pr[X \leqslant 0,75] = 0,7734$$

La table ci-dessous donne les valeurs de la fonction de répartition de la loi normale d'une v.a. centrée-réduite, à savoir les valeurs de : $F_X(x) = \int_{-\infty}^{x} f_X(x')dx'$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133

- Soit T une variable aléatoire décrivant la **durée de vie** d'un phénomène (ampoule électrique ou noyau radioactif). Par conséquent, $T \geq 0$.
- Certains phénomènes possèdent la caractéristique remarquable de ne pas vieillir, c'est le cas par exemple d'un noyau radioactif.
- Le phénomène est sans vieillissement si la durée de vie au-delà de l'instant t est indépendante de l'instant t, c-à-d :

$$\forall (s,t) \in \mathbb{R}^+ \times \mathbb{R}^+, \quad \Pr[\,T > s+t \,|\, T > t\,] = \Pr[\,T > s\,]$$

- Dans ce cas, la v.a. T suit une loi exponentielle. Elle dépend d'un seul paramètre α.
- La loi exponentielle est notée $\boxed{T \sim \mathcal{E}(\alpha)}$.

- Fonction de répartition

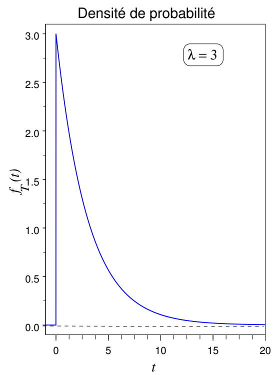
$$\boxed{F_T(t) = 1 - e^{-\alpha t} \qquad (t>0)}$$

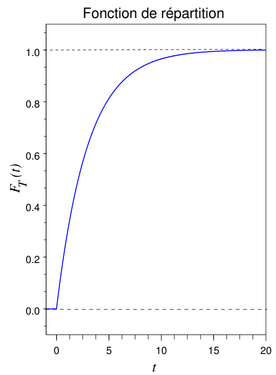
- Densité de probabilité

$$\boxed{T \sim \mathcal{E}(\alpha) \iff f_T(t) = \alpha e^{-\alpha t}}$$

Moments

$$\boxed{\begin{aligned} \mathrm{E}\,[T] &= 1/\alpha \\ \mathrm{Var}\,[T] &= 1/\alpha^2 \end{aligned}}$$





Loi du χ^2

- Soit $X_1, \ldots, X_k$ k variables aléatoires indépendantes de loi normale, centrées et réduites.
- Définissons une nouvelle variable Z définie par :

$$Z = \sum_{i=1}^{k} X_i^2$$

La v.a. Z suit une loi du χ^2 à k degrés de liberté.

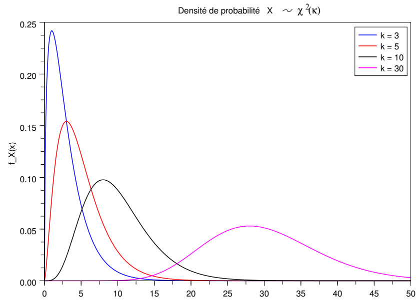
- La loi du χ^2 est notée $Z \sim \chi^2(k)$.

Loi du χ^2

Densité de probabilité

$$\boxed{Z \sim \chi^2(k) \iff f_Z(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}$$

où Γ est la fonction gamma.



Densité de probabilité de v.a. $X \sim \chi^2(k)$ ($k = 3, 5, 10, 30$).

Moments

$$\boxed{\begin{aligned} \mathrm{E}[Z] &= k \\ \mathrm{Var}[Z] &= 2k \end{aligned}}$$

Lorsque k est «grand» ($k > 30$), la loi du χ^2 peut être approchée par une loi normale d'espérance k et de variance $2k$.

Loi de Student

- La loi de Student est construite à partir de deux v.a., l'une suivant une loi normale centrée-réduite, l'autre une loi du χ^2.
- Soient $X \sim \mathcal{N}(0, 1)$ et $K \sim \chi^2_n$, X et K étant indépendants.
- Alors la v.a. T_n définie par :

$$T_n = \frac{X}{\sqrt{K/n}}$$

suit une loi de Student à n degrés de liberté.

- La loi de Student est notée $T \sim \mathrm{St}(n)$.
- La loi de Student est tabulée.
- Pour n grand ($n \geq 30$) la loi de Student peut être approchée par une loi normale.

Loi de Fisher

- La loi de Fisher est construite à partir de deux v.a., suivant des lois du χ^2.
- Soient K_1 et K_2 deux v.a. de lois χ^2_n et χ^2_p.
- Alors la v.a. $X_{n,p}$ définie par

$$X_{n,p} = \frac{K_1/n}{K_2/p}$$

suit une loi de Fisher à n et p degrés de liberté.

- La loi de Fisher est notée $X \sim \mathrm{F}(n, p)$.
- La loi de Fisher est tabulée.

Somme de variables aléatoires

La loi des grands nombres affirme que, si le nombre d'épreuves est très grand, la fréquence d'occurrence d'une éventualité s'écarte «peu» de la probabilité de cette éventualité.

Théorème : Loi faible des grands nombres

Soit (X_i) $(1 \leqslant i \leqslant n)$ une suite de n variables aléatoires indépendantes et de même loi. Posons :

$$Y_n = \frac{X_1 + \cdots + X_n}{n}$$

Alors la v.a. Y_n converge en probabilité vers l'espérance des X_i. En d'autres termes :

$$\boxed{\Pr[|Y_n - \mu| \geq \epsilon] \xrightarrow[n\to\infty]{} 0}$$

On démontre ce théorème à partir de l'inégalité de Bienaymé-Chebychev.

Somme de variables aléatoires discrètes

Dans un premier temps, considérons la somme de deux variables aléatoires indépendantes X et Y de densités f_X et f_Y. Quelle est la densité de la somme $Z = X + Y$?
Soient X et Y deux variables aléatoires discrètes. La probabilité d'avoir $Z = X + Y = z$ est :

$$\Pr[Z = z] = \sum_x \Pr[X = x \;\&\; Y = z - x] = \sum_y \Pr[X = z - y \;\&\; Y = y]$$

Si les variables sont indépendantes :

$$\Pr[Z = z] = \sum_x \Pr[X = x]\Pr[Y = z - x] = \sum_y \Pr[X = z - y]\Pr[Y = y]$$

Somme de variables aléatoires continues

Si X et Y sont continues, alors Z est continue et :

$$\Pr[Z \leqslant z] = \int \Pr[X \leqslant z - y | y \leqslant Y \leqslant y + \mathrm{d}y]\mathrm{d}F_Y[y]$$

soit :

$$F_Z(z) = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-y} f_{X|Y}(x, y)\mathrm{d}x \right) f_Y(y)\mathrm{d}y$$

soit en dérivant par rapport à z

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X|Y}(z - y, y) f_Y(y)\mathrm{d}y$$

où $F_{X|Y}$ et $f_{X|Y}$ sont les fonctions de répartition et densité de probabilité conditionnelles.
Si les variables X et Y sont indépendantes :

$$f_{X|Y}(z - y, y) = f_X(z - y)$$

Somme de variables aléatoires continues

En conséquence :

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\mathrm{d}y = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)\mathrm{d}x$$

On déduit le théorème fondamental :

Théorème : Somme de variables aléatoires indépendantes

La densité de probabilité f_Z de la v.a. $Z = X + Y$ est égale au produit de convolution des densités f_X et f_Y des deux variables X et Y si celles-ci sont indépendantes.

$$f_Z(z) = \int f_X(t) f_Y(z - t)\,\mathrm{d}t = (f_X * f_Y)(z)$$

Cette dernière intégrale définit le produit de convolution des densités f_X et f_Y.