

Résumé de Stats & Probas (MATHH-2002)

Premier quadrimestre 2015-2016

Contents

1	Statistique descriptive	1
1.1	Introduction	1
1.2	Notations et définitions	1
1.3	Représentations graphiques	1
1.4	Valeurs typiques d'une variable aléatoire	1
1.4.1	Paramètres de position	2
1.4.2	Paramètres de dispersion	2
1.5	Statistique descriptive à deux dimensions	3
1.5.1	Bornes de la covariance	3
1.5.2	Coefficient de corrélation	4
1.5.3	Régression linéaire, erreur et variance résiduelle	4
2	Les probabilités	7
2.1	Introduction	7
2.2	Notations et notions	7
2.3	Lien avec les statistiques	7
2.4	Axiomes de Kolmogorov	7
2.5	Evenements stochastiquement indépendants	8
2.6	Probabilité conditionnelle	8
2.7	Théorème de Bayes	8
3	Variables aléatoires	9
3.1	Variables aléatoires discrètes	9
3.1.1	Fonction de répartition	9
3.1.2	Lien avec les statistiques descriptives	10
3.2	Variables aléatoires continues	10
3.2.1	Vérification des axiomes	10
3.2.2	Lien avec les statistiques descriptives	11
3.3	Opérations sur des variables aléatoires	11
3.3.1	Variables aléatoires à deux dimensions	11
3.3.2	Transformée d'une variable aléatoire	13
3.3.3	Transformée affine	13
3.3.4	Opérations arithmétiques	14
3.4	Manipulation de transformées	15
3.4.1	Espérance d'une somme	16
3.4.2	Espérance d'un produit	16
3.4.3	Calcul de l'écart-type	16
3.5	Fonction génératrice des moments	17
3.5.1	Fonction génératrice d'une somme	18
4	Variables aléatoires particulières	18
4.1	Binomiales et variables de Bernoulli	18
4.1.1	Les axiomes de Kolmogorov	18

4.1.2	Valeurs particulières	18
4.1.3	stabilité des binomiales	20
4.2	Variables de Poisson	20
4.2.1	Distribution	20
4.2.2	Vérification des axiomes	21
4.2.3	Valeurs particulières	21
4.2.4	Stabilité des Poissons	22
4.3	Les exponentielles négatives	22
4.3.1	Axiomes	22
4.3.2	Valeurs particulières	23
4.3.3	Stabilité	23
4.3.4	Propriété d'oubli	24
4.4	Variable normale	24
4.4.1	Axiomes de Kolmogorov	24
4.4.2	Intérêts de la normale $\mathcal{N}(\mu, \sigma)$	24
4.4.3	Valeurs particulières	25
4.4.4	Stabilité des normales	26
4.4.5	La normale $\mathcal{N}(0, 1)$	26
5	Théorèmes importants	26
5.1	Théorème central limite	26
5.2	Approximation par une normale	28
5.2.1	Approximer une binomiale	28
5.2.2	Approximation d'une Poisson	28
5.3	Théorème de Binaymé-Tchebycheff	29
5.4	Théorème de Bernoulli	29
6	Inférence statistique	30
6.1	Biais des valeurs particulières	30
6.2	Estimer un paramètre	31
6.3	Intervalle de confiance	32
6.4	Test d'hypothèse	32
6.5	Student et χ^2	33
6.6	Analyse de deux populations distinctes	34

1 Statistique descriptive

1.1 Introduction

Il existe différents types de variables qui sont manipulées en statistiques (et en probabilités) : soit une variable est *quantitative* (numérique) (en quel cas elle est soit discrète soit continue) soit elle est *qualitative*. Si elle est qualitative, elle ne peut être représentée par un nombre car il serait alors sous-entendu que toutes les opérations applicables aux nombres le seraient également sur ces variables particulières, ce qui insérerait un **biais**.

Les variables qualitatives sont regroupées en quatre sous-parties : *nominales*, *ordinales*, *intervalles*, *ratios*. Ce qui différencie ces sous-parties est l'ensemble des opérations que l'on peut appliquer. Deux variables nominales ne peuvent qu'être comparées $=, \neq$, deux variables ordinales peuvent **en plus** être ordonnées, deux variables intervalles acceptent **en plus** les opérations additives et les variables ratios peuvent servir d'opérandes pour des opérations multiplicatives.

1.2 Notations et définitions

Il est important de commencer par déterminer les notations et les notions qui vont suivre afin de ne pas les confondre :

- une variable aléatoire est notée par une lettre **latine majuscule** : X ;
- les valeurs possibles pour une telle variable aléatoire (discrète) sont notées par la même lettre latine mais **minuscule** avec un indice tel que $\forall i \in \{1, \dots, p\} : x_i \leq x_{i+1}$;
- l'*effectif* est le nombre d'observations réalisées et est habituellement noté n ;
- la *fréquence absolue* d'une valeur x_i est le nombre d'observations de cette valeur et est notée n_i ;
- la *fréquence relative* d'une valeur x_i est la proportion d'observations correspondant à cette valeur et est notée $f_i = \frac{n_i}{n}$;
- le mode est la valeur ayant la plus grande fréquence (donc la plus observée) ;
- la fréquence cumulée (absolue ou relative) d'une valeur x_i est la somme des fréquences $f_i + f_{i-1} + \dots + f_1$ (dans le cas continu, la somme devient une intégrale, à voir plus tard) ;
- la distribution de fréquences est l'ensemble des couples (x_i, f_i) (donc une fonction $f_X(x_i)$ dans le cas d'une variable continue).

1.3 Représentations graphiques

Une distribution de fréquences peut être totalement aléatoire mais il est assez fréquent de voir quatre différentes classes de graphiques :

- les graphiques en cloche (moyenne d'une distribution aléatoire de taille, de nombre, etc.) ;
- les graphiques en U (interventions/jour sur la vie d'un objet) ;
- les graphiques en i décroissants de gauche à droite (nombre de gagnants du Lotto en fonction du montant) ;
- les graphiques en j croissant de gauche à droite (risque de panne en fonction de l'âge d'une machine).

1.4 Valeurs typiques d'une variable aléatoire

Une variable aléatoire peut être représentée et manipulée au travers de certaines valeurs particulières. Ces valeurs sont réparties en *paramètres de position* et *paramètres de dispersion*.

1.4.1 Paramètres de position

La moyenne (arithmétique) est définie par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p x_i n_i = \sum_{i=1}^p x_i f_i.$$

La médiane est définie par :

$$\tilde{x} \text{ t.q. } = F^*(\tilde{x}) = \frac{1}{2}.$$

où F^* est la fréquence cumulée relative. Le mode est défini comme les maxima locaux de la distribution.

Ces valeurs ne sont cependant pas suffisantes pour bien comprendre l'**entièrement** de la distribution. Par exemple, la moyenne est très sensible aux valeurs aberrantes alors que la médiane l'est beaucoup moins. De plus, une moyenne n'est pas unique : il existe plusieurs possibilités pour calculer une moyenne (représentant une valeur différente) : arithmétique, géométrique, harmonique, etc.

1.4.2 Paramètres de dispersion

L'écart moyen est défini par :

$$\frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x}).$$

Mais cette valeur n'est pas intéressante car la moyenne est définie comme l'*exact* centre de toutes les valeurs. Dès lors, les valeurs $< \bar{x}$ sont compensées par les valeurs $> \bar{x}$ de manière à ce que l'écart moyen soit toujours nul :

$$\frac{1}{n} \sum_{i=1}^p n_i (x_i - \bar{x}) = \frac{1}{n} \sum_{i=1}^p n_i x_i - \frac{1}{n} \sum_{i=1}^p n_i \bar{x} = \bar{x} - \bar{x} \frac{1}{n} \sum_{i=1}^p n_i = 0.$$

Il est alors possible de parler d'écart moyen absolu :

$$\frac{1}{n} \sum_{i=1}^p |n_i (x_i - \bar{x})|.$$

Mais la notion de variance est beaucoup plus agréable à manipuler et est construite de manière à pénaliser les gros écarts à la moyenne plus fort que les petits écarts :

$$s^2 = \frac{1}{n} \sum_{i=1}^p (x_i - \bar{x})^2 \geq 0.$$

Sur base de cette notion de variance, l'écart-type s est défini par $s = \sqrt{s^2}$.

L'amplitude est la différence entre la plus petite et la plus grande valeur. Cependant, telle la moyenne, l'amplitude est très sensible aux valeurs dégénérées.

Les quartiles quant à eux représentent une division des valeurs en 4 grâce à trois séparateurs :

$$q_1 \text{ t.q. } F^*(q_1) = \frac{1}{4}, \quad q_2 = \tilde{x}, \quad q_3 \text{ t.q. } F^*(q_3) = \frac{3}{4}.$$

Les quartiles q_1 et q_3 servent à définir l'intervalle *interquartile* $IQ = q_3 - q_1$. Cet intervalle donne une manière de repérer les valeurs aberrantes : si $x < q_1 - 2IQ$ ou si $x > q_3 + 2IQ$, alors on considère x comme dégénéré.

1.5 Statistique descriptive à deux dimensions

Soient $X(x_1, \dots, x_p)$ et $Y(y_1, \dots, y_q)$ avec un effectif de n observations. On définit $n_{ij} :=$ la fréquence absolue du couple (x_i, y_j) . On définit également les fréquences absolues restreintes $n_{i.} = \sum_{j=1}^q n_{ij}$ et $n_{.j} = \sum_{i=1}^p n_{ij}$.

Les valeurs typiques que l'on utilise pour caractériser une telle situation sont :

- les moyennes marginales définies par :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p n_{i.} x_i, \bar{y} = \frac{1}{n} \sum_{j=1}^q n_{.j} y_j;$$

- les variances marginales définies par :

$$s_1^2 = \frac{1}{n} \sum_{i=1}^p n_{i.} (x_i - \bar{x})^2, s_2^2 = \frac{1}{n} \sum_{j=1}^q n_{.j} (y_j - \bar{y})^2;$$

- les moyennes conditionnelles définies par :

$$\bar{x}_{.k} = \frac{1}{n_{.k}} \sum_{i=1}^p n_{ik} x_i, \bar{y}_{k.} = \frac{1}{n_{k.}} \sum_{j=1}^q n_{kj} y_j.$$

Afin de déterminer s'il existe une relation linéaire entre ces deux variables aléatoires, on utilise la **covariance** :

$$m_{11} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}).$$

En effet, si ce paramètre est positif, c'est qu'il y a une majorité d'observations qui se situent dans les cadrans inférieur gauche et supérieur droit¹ et donc que la droite relative à la relation linéaire est croissante. De manière similaire, si m_{11} est négatif, c'est qu'il y a majorité d'observations dans les cadrans supérieur gauche et inférieur droit. Attention cependant à ne pas trouver une relation linéaire là où il n'y en a pas.

1.5.1 Bornes de la covariance

Il existe une propriété disant que $|m_{11}| \leq s_1 s_2$. De manière plus précise, si $|m_{11}| = s_1 s_2$, c'est que tous les points se situent sur un axe non parallèle aux axes Ox et Oy .

Démonstration :

$$\begin{aligned} \forall u \in \mathbb{R} : & \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (u(x_i - \bar{x}) + (y_j - \bar{y}))^2 && \geq 0 \\ & \frac{u^2}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})^2 + \frac{2u}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}) + \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (y_j - \bar{y})^2 && \geq 0 \\ & \frac{u^2}{n} \sum_{i=1}^p n_{i.} (x_i - \bar{x})^2 + 2um_{11} + \sum_{j=1}^q n_{.j} (y_j - \bar{y})^2 && \geq 0 \\ & u^2 s_1^2 + 2um_{11} + s_2^2 && \geq 0 \end{aligned}$$

On a donc une équation du second degré en u de discriminant $\Delta = (2m_{11})^2 - 4s_1^2 s_2^2$. Or du fait que cette équation du second degré soit ≥ 0 , il y a soit 1 soit 0 racine. Il faut donc $\Delta \leq 0$ ou encore $4m_{11}^2 - 4s_1^2 s_2^2 \leq 0$ dès lors, on a $m_{11}^2 \leq s_1^2 s_2^2$. En prenant de part et d'autre la racine carrée, on a $|m_{11}| \leq s_1 s_2$.

¹En considération une découpe du plan en 4 : en traçant un axe $y = \bar{y}$ et un autre $x = \bar{x}$.

Plus précisément, si l'équation vaut 0, c'est que le discriminant est nul et qu'il y a une seule racine. Il faut donc avoir $\Delta = 0$, ou encore $m_{11} = s_1 s_2$. \square

1.5.2 Coefficient de corrélation

Vu que m_{11} et $s_1 s_2$ sont liés, on définit :

$$r := \frac{m_{11}}{s_1 s_2},$$

le coefficient de corrélation de X et Y . Ce coefficient montre à quel point les deux variables aléatoires X et Y sont en relation linéaire. En effet, si tous les points sont placés sur une droite, alors on a $|r| = 1$ et dans les autres cas, on a $|r| < 1$.

1.5.3 Régression linéaire, erreur et variance résiduelle

Lorsque deux variables X et Y sont en relation linéaire, on détermine l'équation de la droite $d : ax + b$. a, b sont ici les valeurs qui caractérisent notre droite. Cependant, la droite ne convient pas **exactement** à **tous** les points. Il y a donc une erreur que l'on peut calculer :

$$\epsilon(a, b) = \sum_{i=1}^p \sum_{j=1}^q n_{ij} (y_j - (ax_i + b))^2.$$

Cette fonction doit être minimisée afin de trouver les paramètres optimaux. Pour minimiser la fonction, il faut déterminer son gradient et l'annuler :

$$\begin{aligned} \frac{\partial}{\partial a} \epsilon(a, b) &= \frac{\partial}{\partial a} \sum_{i=1}^p \sum_{j=1}^q n_{ij} [y_j - (ax_i + b)]^2 = \sum_{i=1}^p \sum_{j=1}^q n_{ij} \frac{\partial}{\partial a} [y_j - (ax_i + b)]^2 \\ &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} 2[y_j - (ax_i + b)](-x_i) \\ &= 2 \sum_{i=1}^p \sum_{j=1}^q n_{ij} ax_i^2 + 2 \sum_{i=1}^p \sum_{j=1}^q n_{ij} bx_i - 2 \sum_{i=1}^p \sum_{j=1}^q n_{ij} y_j x_i. \end{aligned}$$

Or cette dérivée partielle doit être nulle pour trouver le minimum :

$$\begin{aligned} a \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i^2 + b \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j \\ a \sum_{i=1}^p x_i^2 \sum_{j=1}^q n_{ij} + b \sum_{i=1}^p x_i \sum_{j=1}^q n_{ij} &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j \\ a \sum_{i=1}^p n_{i.} x_i^2 + b \sum_{i=1}^p n_{i.} x_i &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j. \end{aligned}$$

De même pour b :

$$\begin{aligned}
\frac{\partial}{\partial b} \epsilon(a, b) &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} \frac{\partial}{\partial b} [y_j - (ax_i + b)]^2 \\
&= \sum_{i=1}^p \sum_{j=1}^q n_{ij} 2[y_j - (ax_i + b)](-1) \\
&= 2 \sum_{i=1}^p \sum_{j=1}^q n_{ij} [(ax_i + b) - y_j].
\end{aligned}$$

Or à nouveau, cette dérivée partielle doit s'annuler pour trouver le minimum :

$$\begin{aligned}
2 \sum_{i=1}^p \sum_{j=1}^q n_{ij} [(ax_i + b) - y_j] &= 0 \\
\sum_{i=1}^p \sum_{j=1}^q n_{ij} ax_i + \sum_{i=1}^p \sum_{j=1}^q n_{ij} b &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} y_j \\
a \sum_{i=1}^p x_i \sum_{j=1}^q n_{ij} + b \sum_{i=1}^p \sum_{j=1}^q n_{ij} &= \sum_{j=1}^q y_j \sum_{i=1}^p n_{ij} \\
a \sum_{i=1}^p n_{i.} x_i + bn &= \sum_{j=1}^q n_{.j} y_j \\
an\bar{x} + bn &= n\bar{y}.
\end{aligned}$$

On a donc deux équations à deux inconnues :

$$\begin{cases} a \sum_{i=1}^p n_{i.} x_i^2 + bn\bar{x} = \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j & (1) \\ an\bar{x} + bn = n\bar{y} & (2) \end{cases}$$

En développant (1) – \bar{x} (2), on obtient ceci :

$$\begin{aligned}
\left[a \sum_{i=1}^p n_{i.} x_i^2 + b n \bar{x} \right] - \left[a n \bar{x}^2 + b n \bar{x} \right] &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j - n \bar{x} \bar{y} \\
a \left[\sum_{i=1}^p n_{i.} x_i^2 - n \bar{x}^2 \right] &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j - n \bar{x} \bar{y} \\
a \left[\sum_{i=1}^p n_{i.} x_i^2 + n \bar{x}^2 - 2 n \bar{x}^2 \right] &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j + n \bar{x} \bar{y} - 2 n \bar{x} \bar{y} \\
a \left[\sum_{i=1}^p n_{i.} (x_i^2 + \bar{x}^2 - 2 \bar{x} x_i) \right] &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} x_i y_j - \bar{x} \sum_{j=1}^q y_j \sum_{i=1}^p n_{ij} - \bar{y} \sum_{i=1}^p x_i \sum_{j=1}^q n_{ij} + n \bar{x} \bar{y} \\
a \sum_{i=1}^p n_{i.} (x_i - \bar{x})^2 &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i y_j - \bar{x} y_j - x_i \bar{y} + \bar{x} \bar{y}) \\
a n s_1^2 &= \sum_{i=1}^p \sum_{j=1}^q n_{ij} (x_i - \bar{x})(y_j - \bar{y}) \\
a n s_1^2 &= n m_{11} \\
a &= \frac{m_{11}}{s_1^2}.
\end{aligned}$$

Et comme l'équation (2) nous dit que $b = \bar{y} - a \bar{x}$, on a $b = \bar{y} - \frac{m_{11}}{s_1^2} \bar{x}$. Dès lors, l'équation de la droite est :

$$D : y = \frac{m_{11}}{s_1^2} (x - \bar{x}) + \bar{y}.$$

Il est également possible de faire une régression linéaire pour trouver x en fonction de y . Dans ce cas, l'équation de droite devient :

$$D : x = \frac{m_{11}}{s_1^2} (y - \bar{y}) + \bar{x}.$$

La *variance résiduelle* de y en x est donnée par :

$$\begin{aligned}
s_{yx}^2 &= \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} [y_j - a x_i - b]^2 \\
&= \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} \left[y_j - \frac{m_{11}}{s_1^2} x_i - \bar{y} + \frac{m_{11}}{s_1^2} \bar{x} \right]^2 \\
&= \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (y_j - \bar{y})^2 + \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} \left(\frac{m_{11}}{s_1^2} \right)^2 (x_i - \bar{x})^2 - \frac{2}{n} \sum_{i=1}^p \sum_{j=1}^q n_{ij} (y_j - \bar{y}) \frac{m_{11}}{s_1^2} (x_i - \bar{x}) \\
&= s_2^2 + \frac{m_{11}^2}{s_1^4} s_1^2 - 2 \frac{m_{11}}{s_1^2} m_{11} = s_2^2 - \frac{m_{11}^2}{s_1^2} = s_2^2 - \frac{m_{11}^2 s_2^2}{s_1^2 s_2^2} = s_2^2 \left(1 - \frac{m_{11}^2}{s_1^2 s_2^2} \right) = s_2^2 (1 - r^2).
\end{aligned}$$

On peut comprendre cette valeur comme étant la proportion de la variance n'étant pas expliquée par le modèle linéaire.

2 Les probabilités

2.1 Introduction

Suite à des questions telles que *qu'est-ce que l'aléatoire ?* ou *comment le quantifier ?*, on arrive vite à la notion de probabilité. Le domaine des probabilités est très fortement lié au domaine des statistiques comme ce sera vu plus loin.

2.2 Notations et notions

Tout comme pour les statistiques, il est important de commencer par définir **précisément** les termes et conceptions qui vont suivre :

- une *expérience aléatoire* est une expérience dont l'issue ne peut être définie précisément (ou avec certitude) ;
- l'univers Ω (ou E) est l'ensemble des issues possibles d'une expérience aléatoire ;
- un résultat est un élément $e \in E$ qui représente l'issue d'une variable aléatoire ;
- un événement $A \subseteq E$ est un sous-ensemble de possibilités sur lequel la probabilité d'occurrence peut être évaluée ;
- l'événement certain est l'événement $A = E$ de probabilité 1 (donc dont l'occurrence est sûre) ;
- l'événement impossible $\emptyset \subset E$ est l'événement de probabilité 0 (donc dont la non-occurrence est sûre).

2.3 Lien avec les statistiques

Il existe plusieurs manières de lier les statistiques et les probabilités. La première est l'*approche fréquentiste*. Cette théorie dit que quand $n \rightarrow \infty$, la fréquence relative $\frac{n_i}{n}$ converge en une valeur $\in [0, 1]$. C'est cette valeur que représente la probabilité.

A partir de là, on peut définir la probabilité d'un événement $A \subseteq E$ comme suit² :

$$P(A) = \frac{\#A}{\#E}.$$

2.4 Axiomes de Kolmogorov

Comme toute théorie mathématique, la théorie des probabilités est basée sur certains axiomes. Ces axiomes sont appelés *axiomes de Kolmogorov*. Ils sont au nombre de trois et peuvent être exprimés comme suit :

1. toute probabilité est évaluée entre 0 et 1 compris : $\forall A \subseteq E : P(A) \in [0, 1]$;
2. La probabilité de l'événement certain est de 1 : $P(E) = 1$;
3. Si $A \subseteq E$ et $B \subseteq E$ sont disjoints (incompatibles), la probabilité de leur union est la somme des probabilités :

$$\forall A, B \subseteq E : A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B).$$

De là, une définition plus précise de la fonction de probabilité peut être exprimée :

$$P : 2^E \rightarrow [0, 1] : A \mapsto P(A).$$

Remarque : Attention, il existe des événements possibles mais de **stochastiquement** impossibles. Cela veut dire qu'il existe des événements qui peuvent théoriquement apparaître mais dont la fréquence relative tend vers 0 quand n tend vers $+\infty$. Par exemple, lors du lancer d'une pièce de monnaie, l'événement relatif au fait que la pièce ne tombe ni sur **pile**, ni sur **face** mais *pile sur la tranche* est possible mais de probabilité nulle.

²Quant le comptage n'est plus trivial, il faut passer par l'analyse combinatoire.

2.5 Evenements stochastiquement indépendants

Soient $A, B \subseteq E$, deux événements. Notons $P(A|B)$ la probabilité de l'événement A en sachant que l'événement B s'est produit. Si $P(A|B) = P(A)$, on dit que A et B sont (stochastiquement) indépendants (ce que l'on note $A \sqcup B$).

2.6 Probabilité conditionnelle

Dans le cas précédent, la notation $P(A|B) = \frac{P(A \cap B)}{P(B)}$ représente la probabilité que l'événement A arrive **en sachant que** l'événement B est arrivé. Pour pouvoir appeler cette quantité *probabilité*, il faut montrer qu'elle vérifie les axiomes de Kolmogorov :

1. $P(A|B) = \frac{P(A \cap B)}{P(B)} \geq 0$ car $P(A \cap B), P(B) \geq 0$;
2. $P(E|B) = 1$ car $P(E \cap B) = P(B)$, donc $\frac{P(B)}{P(B)} = 1$;
3. $A \cap C = \emptyset \Rightarrow P(A \cup C|B) = P(A|B) + P(C|B)$ car $P(A \cup C|B) = \frac{P((A \cap B) \cup (C \cap B))}{P(B)} = \frac{P(A \cap B)}{P(B)} + \frac{P(C \cap B)}{P(B)} = P(A|B) + P(C|B)$.

Les axiomes de Kolmogorov ayant été démontrés, il est acceptable d'utiliser cette notion afin de représenter une probabilité.

Remarque : Il y a une **très forte** corrélation entre *événements incompatibles* et *événements indépendants*. En effet, deux événements incompatibles ne peuvent pas apparaître en même temps. Dès lors la probabilité d'un événement en sachant que l'autre est apparu est facilement calculable et vaut 0. On peut donc dire que deux événements incompatibles sont très fort dépendants.

De plus, si deux événements sont indépendants, alors $P(A|B) = P(A) = \frac{P(A \cap B)}{P(B)}$ ou encore $P(A)P(B) = P(A \cap B)$.

2.7 Théorème de Bayes

Le théorème de Bayes est un moyen d'*inverser* les probabilités conditionnelles. Il sert en effet à exprimer $P(A|B)$ en connaissant $P(B|A)$. Soient m événements deux à deux incompatibles :

$$\left\{ \begin{array}{l} B \subseteq \bigcup_{i=1}^m A_i \\ \forall i \neq j : A_i \cap A_j = \emptyset \\ P(B) \neq 0 \end{array} \right.$$

En disposant des informations $\forall i \in \{1, \dots, m\} : P(A_i), P(B|A_i)$, on peut retrouver les informations $P(A_k|B)$:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^m P(B|A_i)P(A_i)}.$$

Par exemple, en connaissant les données suivantes :

- la probabilité qu'une personne soit un homme ou une femme (respectivement $P(H)$ et $P(F)$) ;
- la probabilité qu'un gène s'exprime chez une femme ($P(G|F)$) ;
- la probabilité que ce gène s'exprime chez un homme ($P(G|H)$),

on peut déterminer la probabilité qu'une personne quelconque ayant le gène exprimé soit un homme ou une femme. On sait, par le théorème de Bayes, exprimer les égalités suivantes :

$$P(H|G) = \frac{P(G|H)P(H)}{P(G|H)P(H) + P(G|F)P(F)} \quad (3)$$

$$P(F|G) = \frac{P(G|F)P(F)}{P(G|H)P(H) + P(G|F)P(F)} \quad (4)$$

Démonstration du théorème de Bayes : En partant de l'égalité suivante (corollaire direct de la définition de probabilité conditionnelle) :

$$P(B)P(A_k | B) = P(B | A_k)P(A_k),$$

on obtient :

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{P(B)}. \quad (5)$$

Or on sait que $A_i \cap A_j = \emptyset \forall i \neq j$. Dès lors $(B \cap A_j) \cap (B \cap A_i) = \emptyset$ également. De plus, $\{A_i\}_i$ forme une partition de E . On a donc :

$$P(B) = P(B \cap E) = P\left(B \cap \left[\bigcup_{i=1}^m A_i\right]\right) = P\left(\bigcup_{i=1}^m [B \cap A_i]\right).$$

Et comme les couples $((B \cap A_i), (B \cap A_j))$ sont disjoints, on peut utiliser le troisième axiome de Kolmogorov :

$$P(B) = \sum_{i=1}^m P(B \cap A_i).$$

Par définition de la probabilité conditionnelle, on obtient :

$$P(B) = \sum_{i=1}^m P(B | A_i)P(A_i). \quad (6)$$

En remettant (5) et (6) ensemble, on obtient le théorème de Bayes :

$$P(A_k | B) = \frac{P(B | A_k)P(A_k)}{\sum_{i=1}^m P(B | A_i)P(A_i)}. \square$$

3 Variables aléatoires

3.1 Variables aléatoires discrètes

Une variable aléatoire (notée par une lettre latine majuscule par convention) est *discrète* si le nombre de résultats possibles est dénombrable. Les valeurs qu'elle peut prendre sont notées x_1, \dots, x_n où $\forall i : x_i \leq x_{i+1}$. Ce sont ces valeurs qui définissent la variable aléatoire. On note également P_i la probabilité que la variable aléatoire X prenne la valeur x_i . Dès lors, $P_i := P(X = x_i)$.

Remarque : Il faut que ces valeurs P_i respectent les propriétés des probabilités. Il faut donc $P_i \geq 0$ et $P(E) = 1$. Or $E = \bigcup_{i=1}^n x_i$, donc $1 = P(E) = P(\bigcup_{i=1}^n X = x_i) = \sum_{i=1}^n P_i$.

3.1.1 Fonction de répartition

Afin de déterminer le comportement d'une variable aléatoire, on utilise la notion de *fonction de répartition*. On la définit comme suit :

$$F_X : \mathbb{R} \rightarrow [0, 1] : x \mapsto P(\{e \in E \text{ t.q. } X(e) \leq x\}).$$

Dans le cas d'une variable discrète, on a donc :

$$F_X(x) = \sum_{x_i \leq x} P(X = x_i) \leq 1.$$

3.1.2 Lien avec les statistiques descriptives

En statistique descriptive, on avait plusieurs variables types pour désigner une situation. Par exemple, la moyenne était définie comme suit :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^p n_i x_i = \sum_{i=1}^p \frac{n_i}{n} x_i.$$

Or dans le cas des statistiques, on sait que la valeur $\frac{n_i}{n}$ converge vers p_i . La notion de moyenne est donc appelée μ (en stats on utilisait des lettres latines, en probabilités, on utilise les équivalents grecs) et définie comme suit :

$$\mu_X = E(X) := \sum_{i=1}^p p_i x_i.$$

De même pour l'écart-type défini en statistique descriptive a un équivalent en probabilités :

$$\sigma_X^2 = D^2(X) := \sum_{i=1}^p p_i (x_i - \mu_X)^2.$$

3.2 Variables aléatoires continues

Abrégées VAC, les variables aléatoires continues sont des variables aléatoires dont l'ensemble des valeurs possibles n'est pas dénombrable (tel \mathbb{R}). Comme le nombre de possibilités pour cette variable est infini, on ne peut pas définir la probabilité comme la convergence $\frac{n_i}{n}$ pour $n \rightarrow \infty$ vu que ces valeurs convergeraient toutes vers 0.

Dès lors, on définit une autre notion : la *densité de probabilité*. Cette fonction est définie sur base de la fonction de répartition F_X . Plus précisément,

$$f_X(x) = \frac{\partial}{\partial x} F_X(x). \quad (7)$$

Le fait de prendre la dérivée permet de définir la fonction de densité de probabilité comme étant la pente de la tangente de la fonction de répartition. Cette définition est pertinente car au plus F_X est croissante pour une valeur x , au plus il est probable que la valeur de la variable aléatoire lors d'une expérience aléatoire soit proche de x .

3.2.1 Vérification des axiomes

A nouveau, le fait de définir une nouvelle notation de probabilité implique de vérifier que l'appellation *probabilité* est cohérente et acceptable. Il faut donc vérifier les axiomes de Kolmogorov. Le premier axiome implique $P(A) \geq 0$ ou encore $P(a \leq X \leq b) \geq 0$. Cela implique :

$$\int_a^b f_X(x) dx \geq 0.$$

La seule manière de garantir cela est d'imposer $f_X(x) \geq 0 \forall x$. Le deuxième axiome impose $P(E) = 1$. Donc :

$$\int_{-\infty}^{\infty} f_X(x) dx = 1.$$

Par définition, on a :

$$F_X(x) = \int_{-\infty}^x f_X(t) dt. \quad (8)$$

On peut donc en déduire que $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$. Ce qui donne lieu à une remarque importante : il a été mentionné dans l'introduction que la probabilité que la variable aléatoire prenne une valeur exacte est nulle car la convergence du quotient $\frac{n_i}{n}$ vaut 0. Donc la valeur $P(X = a) = P(a \leq X \leq a) = F_X(a) - F_X(a) = 0$. On a donc bien en réalité une intégrale sur un point donc un résultat nul. Un corollaire de ceci est le fait que $\forall x : P(X \leq x) = P((X < x) \cup (X = x)) = P(X < x)$. Le fait de prendre - ou pas - la valeur x dans l'évaluation de la probabilité n'intervient donc pas dans le résultat final. Ce corollaire sera souvent réutilisé sans être explicité.

De plus, on peut justifier l'apparition d'une dérivée par le fait que l'aire sous la courbe est donnée par :

$$A := \lim_{\Delta t \rightarrow 0} \frac{F_X(t + \Delta t) - F_X(t)}{\Delta t} = \frac{\partial}{\partial t} F_X(t).$$

3.2.2 Lien avec les statistiques descriptives

Tout comme pour les variables aléatoires discrètes, les variables aléatoires continues peuvent être reliées aux valeurs particulières vues en statistiques descriptives. La notion de moyenne et celle d'écart-type sont données par :

$$\mu_X = E(X) := \int_{-\infty}^{\infty} x f_X(x) dx \quad (9)$$

$$\sigma_X^2 = D^2(X) := \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx. \quad (10)$$

3.3 Opérations sur des variables aléatoires

Lorsque l'on a deux variables aléatoires X et Y définies par leur fonction de densité respective (f_X et f_Y), il serait intéressant de pouvoir définir une nouvelle variable Z définie par une opération sur les deux autres (par exemple $Z = X + Y$).

3.3.1 Variables aléatoires à deux dimensions

Lorsque l'on passe à deux dimensions, on définit la fonction de répartition comme suit :

$$F_{(V,W)}(x, y). \quad (11)$$

Avec une seule variable aléatoire, l'indépendance est définie par :

$$A \sqcup B \iff P(A \cap B) = P(A)P(B).$$

Cette définition doit être adaptée pour le cas de dimension 2. Il y a donc un théorème pour caractériser l'indépendance sur base de la fonction de répartition à deux dimensions :

Théorème : $\forall x_1, x_2, y_1, y_2 : V \sqcup W$

1. $\iff P(V \in [x_1, x_2] \wedge W \in [y_1, y_2]) = P(V \in [x_1, x_2])P(W \in [y_1, y_2])$
2. $\iff F_{(V,W)}(x_2, y_2) = F_V(x_2)F_W(y_2)$
3. disc. $\iff P_{ij} = P_i P_j \forall i, j$
- cont. $\iff f_{(V,W)}(x_2, y_2) = f_V(x_2)f_W(y_2)$

Remarque : le cas discret et continu ont bien été différenciés pour le point 3. car il est question de la fonction de densité ce qui n'est pas pertinent dans le cas d'une variable discrète.

Démonstration : Pour démontrer toutes les équivalences, il faut démontrer $(i) \rightarrow (ii)$, $(ii) \rightarrow (iii)$ et $(iii) \rightarrow (i)$.

$(i) \rightarrow (ii)$ Si la propriété (i) est vraie pour tous x_1, x_2, y_1, y_2 , alors plus précisément avec $x_1, x_2 = -\infty$:

$$P(V \in]-\infty, x_2] \wedge W \in]-\infty, y_2]) = P(V \in]-\infty, x_2])P(W \in]-\infty, y_2]) = P(V \leq x_2)P(W \leq y_2) = F_V(x_2)F_W(y_2).$$

$(ii) \rightarrow (iii)$ En utilisant la propriété (ii) , on a :

$$\begin{aligned} \frac{\partial^2}{\partial x \partial y} F_{(V,W)}(x, y) &= \frac{\partial}{\partial x} \left[\frac{\partial}{\partial y} F_{(V,W)}(x, y) \right] = \frac{\partial}{\partial x} \left[\frac{\partial}{\partial y} (F_V(x)F_W(y)) \right] \\ &= \frac{\partial}{\partial x} \left[F_V(x) \frac{\partial}{\partial y} F_W(y) \right] = \frac{\partial}{\partial x} [F_V(x)f_W(y)] = f_W(y) \frac{\partial}{\partial x} F_V(x) = f_V(x)f_W(y). \end{aligned}$$

$(iii) \rightarrow (i)$ En intégrant doublement l'égalité (iii) de part et d'autre, on obtient :

$$\begin{aligned} \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_V(x)f_W(y) dx dy &= \int_{x_1}^{x_2} f_V(x) \left(\int_{y_1}^{y_2} f_W(y) dy \right) dx = \int_{x_1}^{x_2} f_V(x)P(W \in [y_1, y_2]) dx \\ &= P(W \in [y_1, y_2]) \int_{x_1}^{x_2} f_V(x) dx = P(V \in [x_1, x_2])P(W \in [y_1, y_2]) \\ &= \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{(V,W)}(x, y) dx dy = P(V \in [x_1, x_2] \wedge W \in [y_1, y_2]). \square \end{aligned}$$

En statistique descriptive à deux dimensions, il y avait le paramètre m_{11} qui permettait d'évaluer la corrélation (linéaire) entre deux événements. En probabilités, il existe la même notion définie comme suit :

$$\mu_{11} := \sum_{i=1}^p \sum_{j=1}^q p_{ij}(x_i - \mu_X)(y_j - \mu_Y) \quad \text{si VAD} \quad (12)$$

$$\mu_{11} := \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_{(X,Y)}(x, y) dx dy \quad \text{si VAC.} \quad (13)$$

On peut donc définir $X \sqcup Y \Rightarrow \mu_{11} = 0$. En effet :

$$\begin{aligned} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_{(X,Y)}(x, y) dx dy &\stackrel{X \sqcup Y}{=} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^{\infty} (x - \mu_X)f_X(x) \left(\int_{-\infty}^{\infty} (y - \mu_Y)f_Y(y) dy \right) dx \\ &= \int_{-\infty}^{\infty} (x - \mu_X)f_X(x) dx \int_{-\infty}^{\infty} (y - \mu_Y)f_Y(y) dy \\ &= \left(\int_{-\infty}^{\infty} y f_Y(y) dy - \int_{-\infty}^{\infty} \mu_Y f_Y(y) dy \right) \left(\int_{-\infty}^{\infty} x f_X(x) dx - \int_{-\infty}^{\infty} \mu_X f_X(x) dx \right) \\ &= (E(Y) - \mu_Y 1)(E(X) - \mu_X 1) = (\mu_Y - \mu_Y)(\mu_X - \mu_X) = 0. \square \end{aligned}$$

3.3.2 Transformée d'une variable aléatoire

Soit X , une VAC avec une fonction de densité $f_X(x)$. Soit une nouvelle variable $Z := G(X)$, une transformée croissante monotone de X . Comment peut-on déterminer la fonction de densité $f_Z(z)$?

Commençons par déterminer ce que vaut $F_Z(z)$:

$$F_Z(x) := P(Z \leq x) = P(G(X) \leq x) = P(X \leq G^{-1}(x)) = F_X(G^{-1}(x)) = (F_X \circ G^{-1})(x).$$

Or on sait que $f_Z(x) = \frac{\partial}{\partial x} F_Z(x)$. Dès lors :

$$f_Z(x) = \frac{\partial}{\partial x} (F_X \circ G^{-1})(x) = f_X(G^{-1}(x)) \frac{\partial}{\partial x} G^{-1}(x)$$

Posons $y(x) := G^{-1}(x)$, donc $x = G(y)$, ou encore $\frac{\partial}{\partial x} x = \frac{\partial}{\partial x} G(y) \iff 1 = G'(y)y'$. Ce qui veut dire que $y' = \frac{1}{G'(y)}$. Et par la définition de y , on a : $y' = \frac{1}{G'(G^{-1}(x))}$.

Dès lors :

$$f_Z(x) = f_X(G^{-1}(x))y'(x) = f_X(G^{-1}(x)) \frac{1}{G'(G^{-1}(x))}.$$

Attention, ce résultat n'est valable que si G est une transformée croissante. Si G est décroissante, le raisonnement est le même à l'exception près que lors de la probabilité suivante :

$$P(G(X) \leq x),$$

la prise de la fonction inverse $G^{-1}(x)$ impose de changer le sens de l'inégalité (car la fonction est décroissante). On a donc :

$$F_Z(x) = P(X \geq G^{-1}(x)) = 1 - P(X \leq G^{-1}(x)) = 1 - F_X(G^{-1}(x)).$$

Dès lors, la dérivée donne :

$$f_Z(x) = -f_X(G^{-1}(x)) \frac{1}{G'(G^{-1}(x))}.$$

Le résultat est donc presque similaire si ce n'est un signe – supplémentaire.

3.3.3 Transformée affine

Soit X une VAC définie par f_X . Si on définit $Z := aX + b$ avec $(a, b) \in (\mathbb{R}_0^+, \mathbb{R})$, on peut trouver f_Z de deux manières : soit on passe par la formule démontrée ci-dessus, soit on repart de la définition et on évite des calculs fastidieux :

$$f_Z(x) = \frac{\partial}{\partial x} F_Z(x) = \frac{\partial}{\partial x} P(Z \leq x) = \frac{\partial}{\partial x} P(aX + b \leq x) = \frac{\partial}{\partial x} P\left(X \leq \frac{x-b}{a}\right) = \frac{\partial}{\partial x} F_X\left(\frac{x-b}{a}\right) = f_X\left(\frac{x-b}{a}\right) \frac{1}{a}.$$

3.3.4 Opérations arithmétiques

Soient X, Y deux VAC définies par f_X et f_Y . La variable $Z := X + Y$ est également définie par une fonction f_Z . Or cette fonction ne peut être $f_Z(x) = f_X(x) + f_Y(x)$ car il faut que l'intégrale sur \mathbb{R} soit égale à 1. Or :

$$\int_{-\infty}^{\infty} f_Z(x) dx = \int_{-\infty}^{\infty} (f_X(x) + f_Y(x)) dx = 1 + 1 = 2$$

ne convient pas du tout à cet axiome ! Cette définition de f_Z n'est donc logiquement pas exacte. Il faut à nouveau partir de la définition de F_Z :

$$F_Z(t) := P(Z \leq t) = P(X + Y \leq t) = \iint_{X+Y \leq t} f_{(X,Y)}(x, y) dx dy$$

Pour avancer, il nous faut changer de variables :

$$\begin{cases} x = \beta \\ y = \alpha - \beta \end{cases}$$

On a donc :

$$F_Z(t) = \int_{-\infty}^t \left(\int_{-\infty}^{\infty} f_{(X,Y)}(\beta, \alpha - \beta) \left| \begin{array}{cc} \frac{\partial}{\partial \alpha} x & \frac{\partial}{\partial \beta} x \\ \frac{\partial}{\partial \alpha} y & \frac{\partial}{\partial \beta} y \end{array} \right| d\beta \right) d\alpha$$

Où

$$\left| \begin{array}{cc} \frac{\partial}{\partial \alpha} x & \frac{\partial}{\partial \beta} x \\ \frac{\partial}{\partial \alpha} y & \frac{\partial}{\partial \beta} y \end{array} \right| = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix}$$

est le déterminant du jacobien (à prendre en valeur absolue).

Donc on a au final :

$$F_Z(t) = \int_{-\infty}^t \left(\int_{-\infty}^{\infty} f_{(X,Y)}(\beta, \alpha - \beta) d\beta \right) d\alpha.$$

Et comme $F_Z(t) = \int_{-\infty}^t f_Z(x) dx$, on a au final :

$$f_Z(\alpha) = \int_{-\infty}^{\infty} f_{(X,Y)}(\beta, \alpha - \beta) d\beta$$

Et si au lieu de faire une somme des deux variables X et Y , nous avions voulu les multiplier, le procédé aurait été le même :

$$F_Z(t) = P(XY \leq t),$$

et les variables à poser auraient été :

$$\begin{cases} x = \beta \\ y = \frac{\alpha}{\beta} \end{cases}$$

et donc le déterminant du jacobien aurait donné :

$$\left| \begin{array}{cc} \frac{\partial}{\partial \alpha} x & \frac{\partial}{\partial \beta} x \\ \frac{\partial}{\partial \alpha} y & \frac{\partial}{\partial \beta} y \end{array} \right| = \left| \begin{array}{cc} 0 & 1 \\ \frac{1}{\beta} & -\frac{\alpha}{\beta^2} \end{array} \right| = \frac{1}{|\beta|}$$

de manière à obtenir :

$$F_Z(t) = \int_{-\infty}^t \left(\int_{-\infty}^{\infty} f_{(X,Y)} \left(\beta, \frac{\alpha}{\beta} \right) \frac{d\beta}{|\beta|} \right) d\alpha.$$

Et donc on peut déterminer f_Z par le même principe :

$$f_Z(\alpha) = \int_{-\infty}^{\infty} f_{(X,Y)} \left(\beta, \frac{\alpha}{\beta} \right) \frac{d\beta}{|\beta|}.$$

3.4 Manipulation de transformées

Nous avons trouvé comment déterminer les fonctions de densité pour de nouvelles variables créées par opérations sur d'autres. Maintenant tentons de les manipuler pour déterminer l'espérance ou l'écart-type :

$$E(G(V)) = \int_{-\infty}^{\infty} G(v) f_V(v) dv \quad (14)$$

Démonstration : On sait par définition :

$$E(W = G(V)) = \int_{-\infty}^{\infty} w f_W(w) dw,$$

qui peut être partitionné en \overline{G} et \underline{G} , les sous-ensemble où $G(v)$ est croissante (respectivement décroissante). On a dès lors :

$$E(W) = \int_{\overline{G}} w f_W(w) dw + \int_{\underline{G}} w f_W(w) dw.$$

De là, on applique la formule de la transformée (exprimer f_W en fonction de f_V) :

$$E(W) = \int_{\overline{G}} w f_V(G^{-1}(w)) \frac{dw}{G'(G^{-1}(w))} - \int_{\underline{G}} w f_V(G^{-1}(w)) \frac{dw}{G'(G^{-1}(w))}.$$

Remarque : Effectivement, le + devient - lors du changement de variable pour la raison vue précédemment (fonction décroissante). En réitérant un changement de variable ($w = G(y)$), on rechange le signe - en + et on a donc une somme de deux intégrales de la même fonction sur un partitionnement de \mathbb{R} :

$$E(W) = \int_{-\infty}^{\infty} G(y) f_V(y) dy. \quad \square$$

On peut également montrer la linéarité de l'opérateur $E(V)$:

$$E(aV + b) = \int_{-\infty}^{\infty} (av + b) f_V(v) dv = a \int_{-\infty}^{\infty} v f_V(v) dv + b \int_{-\infty}^{\infty} f_V(v) dv = aE(V) + b. \quad \square$$

3.4.1 Espérance d'une somme

Montrons que l'espérance d'une somme de variables aléatoires correspond à la somme des espérances. En effet, soient X et Y deux VAC. Définissons $Z := X + Y$. On peut déterminer l'espérance de cette variable comme suit (par définition) :

$$E(Z) = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} z f_{X+Y}(z) dz = \int_{-\infty}^{\infty} z \left(\int_{-\infty}^{\infty} f_{(X,Y)}(u, z-u) du \right) dz.$$

Posons donc $z = \alpha + \beta$ et $u = \alpha$ (le déterminant du jacobien vaut 1 en valeur absolue).

$$\begin{aligned} E(Z) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha + \beta) f_{(X,Y)}(\alpha, \beta) d\alpha d\beta = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \alpha f_{(X,Y)}(\alpha, \beta) d\alpha d\beta + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \beta f_{(X,Y)}(\alpha, \beta) d\alpha d\beta \\ &= \int_{-\infty}^{\infty} \alpha \left(\int_{-\infty}^{\infty} f_{(X,Y)}(\alpha, \beta) d\beta \right) d\alpha + \int_{-\infty}^{\infty} \beta \left(\int_{-\infty}^{\infty} f_{(X,Y)}(\alpha, \beta) d\alpha \right) d\beta \\ &= \int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha + \int_{-\infty}^{\infty} \beta f_Y(\beta) d\beta = E(X) + E(Y). \quad \square \end{aligned}$$

3.4.2 Espérance d'un produit

Tout comme pour la somme, montrons que l'espérance d'un produit correspond au produit des espérances. Soient X , Y et $Z := XY$.

$$E(Z) = \int_{-\infty}^{\infty} z f_Z(z) dz = \int_{-\infty}^{\infty} z \left(\int_{-\infty}^{\infty} f_{(X,Y)} \left(z, \frac{z}{u} \right) \frac{du}{|u|} \right) dz$$

Posons donc à nouveau $u = \alpha$ et $z = \alpha\beta$ (le déterminant du jacobien vaut α en valeur absolue). On a donc :

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\alpha\beta) f_{(X,Y)}(\alpha, \beta) d\alpha d\beta.$$

Prenons pour hypothèse $X \sqcup Y$. Dès lors,

$$E(Z) = \int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha \int_{-\infty}^{\infty} \beta f_Y(\beta) d\beta = E(X)E(Y). \quad \square$$

3.4.3 Calcul de l'écart-type

Par la définition de l'écart-type, on a :

$$D^2(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f_X(x) dx.$$

Donc en définissant une fonction quelconque $G : x \mapsto (x - E(X))^2$, on a alors :

$$D^2(X) = \int_{-\infty}^{\infty} G(x) f_X(x) dx = E[G(X)] = E[(x - E(X))^2].$$

On a donc pu, en réalité, définir le calcul de l'écart-type sur base du calcul d'espérance. L'avantage est le fait que toutes les propriétés de l'espérance sont alors utilisables pour l'écart-type. Par exemple, pour déterminer l'écart-type d'une transformée linéaire, on a :

$$\begin{aligned} D^2(aX + b) &= E(((ax + b) - (E(ax + b)))^2) = E((ax + b - (aE(X) + b))^2) = E((ax - aE(X))^2) \\ &= E(a^2(x - E(X))^2) = a^2 E((x - E(X))^2) = a^2 D^2(X). \quad \square \end{aligned}$$

Concernant une somme, on a :

$$\begin{aligned} D^2(X + Y) &= E((x + y - E(X) - E(Y))^2) = E(((x - E(X)) + (y - E(Y))))^2) \\ &= E((x - E(X))^2) + E((y - E(Y))^2) + 2E((x - E(X))(y - E(Y))) \\ &= D^2(X) + D^2(Y) + 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y)) f_{(X,Y)}(x, y) dx dy = D^2(X) + D^2(Y) + 2\mu_{11}. \end{aligned}$$

Ce résultat est intéressant car μ_{11} peut être négatif. Dans le cas où effectivement le coefficient de corrélation linéaire est négatif, on a $D^2(X + Y) < D^2(X) + D^2(Y)$. Cela a des applications entre autres en économie : dans ce cas, investir dans deux actions X et Y permet de *limiter la variance* donc avoir plus d'uniformité dans les gains. Ce peut être intuitif dans le sens qu'il est possible d'investir uniquement dans une action ayant des gains entre janvier et juillet compris ou dans une action ayant des gains uniquement entre août et décembre. Dans ce cas, investir **simultanément** dans les deux actions permet d'avoir des gains pendant toute l'année (donc limiter la variance).

3.5 Fonction génératrice des moments

On définit une fonction $\Psi_X(t) : t \mapsto E(\exp(tX))$ que l'on appelle *fonction génératrice des moments*. On la nomme ainsi car elle permet de déterminer les moments d'ordre k d'une variable aléatoire :

$$\left. \frac{\partial^k}{\partial t^k} \Psi_X(t) \right|_{t=0} = \frac{\partial^k}{\partial t^k} \int_{-\infty}^{\infty} \exp(tx) f_X(x) dx = \int_{-\infty}^{\infty} x^k f_X(x) dx = E(X^k).$$

Cette fonction permet donc effectivement de déterminer les moments d'ordre k en dérivant et en évaluant en $t = 0$. Ce résultat peut se montrer par le développement en série de l'exponentielle :

$$\exp(v) = \sum_{i=0}^{\infty} \frac{v^i}{i!}.$$

Dès lors, en réinjectant cela dans la définition de Ψ_X , on obtient :

$$\Psi_X(t) = E \left[\sum_{i=0}^{\infty} \frac{(tX)^i}{i!} \right] = \sum_{i=0}^{\infty} \frac{t^i}{i!} E(X^i).$$

Donc en dérivant k fois, les k premiers termes s'annulent (par dérivation du t). On a donc :

$$\frac{\partial^k}{\partial t^k} \Psi_X(t) = \sum_{i=k}^{\infty} E(X^i) \frac{t^{i-k}}{(i-k)!}.$$

Dès lors, en évaluant la k ème dérivée en $t = 0$, tous les termes sauf le premier s'annulent afin qu'il ne reste que le moment d'ordre k :

$$\left. \frac{\partial^k}{\partial t^k} \Psi_X(t) \right|_{t=0} = E(X^k). \quad \square \tag{15}$$

3.5.1 Fonction génératrice d'une somme

Il faut savoir qu'une variable aléatoire peut être définie soit par sa fonction de répartition, soit par sa fonction de densité, soit par sa fonction génératrice des moments (il est possible de passer de l'un à l'autre de manière univoque). De là, il est aussi important de regarder ce que donnent les opérations classiques sur la fonction génératrice que sur les autres.

Dans le cas d'une somme, on a $\Psi_Z(t) = \Psi_{X+Y}(t) = E[\exp(t(X+Y))] = E[\exp(tY)\exp(tX)]$. En prenant pour hypothèse l'indépendance de X et Y ($X \sqcup Y$), on a $\Psi_Z(t) = E[\exp(tX)]E[\exp(tY)] = \Psi_X(t)\Psi_Y(t)$.

4 Variables aléatoires particulières

4.1 Binomiales et variables de Bernoulli

Dans le cas où une expérience aléatoire a pour seuls résultats possibles *la réussite* et *l'échec* avec une probabilité respective de p et $(1-p)$, on appelle **variable de Bernoulli** une variable représentant le résultat de cette expérience. La probabilité d'une telle variable est donc exprimée par :

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

Comme une telle situation est réductrice, on définit une distribution **normale** (ou **binomiale**) lorsqu'une variable X est définie comme n répétitions successives **et indépendantes** d'une expérience de Bernoulli. On note cela $X \sim \mathcal{B}(n, p)$, ce qui veut dire : « X suit une distribution binomiale représentant une répétition de n expériences de Bernoulli ayant pour probabilité de réussite p ». X peut donc prendre comme différentes valeurs le nombre d'expériences se soldant en réussite. Dès lors, on a :

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

4.1.1 Les axiomes de Kolmogorov

À nouveau, pour s'assurer de l'appellation *probabilité*, il faut vérifier que les axiomes de Kolmogorov soient vérifiés :

$P(X = k) \geq 0 \forall k$. En effet, $P(X = k)$ est donné par un produit de valeurs toutes positives. Dès lors, la probabilité en elle-même est positive.

$P(E) = 1$. En effet :

$$P(E) = P\left(\bigcup_{k=0}^n X = k\right) = \sum_{k=0}^n P(X = k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = [p + (1-p)]^n = 1.$$

4.1.2 Valeurs particulières

L'intérêt de définir des variables *spéciales* comme les binomiales ou d'autres (plus loin) est de pouvoir trouver des règles correspondantes à toute variable de ce type. En effet, une variable binomiale est paramétrée par les valeurs n et p , respectivement le nombre de répétitions de l'expérience et la probabilité de réussite. Cependant, on peut tout de même exprimer $E(X)$ et $D^2(X)$ pour une variable de ce type :

$$\begin{aligned}
E(X) &= \sum_{i=0}^n iP(X=i) = \sum_{i=1}^n iP(X=i) = \sum_{i=1}^n i \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=1}^n \frac{in!}{i!(n-i)!} p^i (1-p)^{n-i} \\
&= \sum_{i=1}^n \frac{n!}{(i-1)!(n-i)!} p^i (1-p)^{n-i} = np \sum_{i=1}^n \frac{(n-1)!}{(i-1)!(n-i)!} p^{i-1} (1-p)^{n-i} = np 1^{n-1} = np.
\end{aligned}$$

De même manière, on peut déterminer la variance. Mais commençons par faire une remarque applicable à **toute** VAC :

$$D^2(X) = E((X - E(X))^2) = E(X^2) + E(E(X)^2) - 2E(XE(X)) = E(X^2) - 2E(X)E(E(X)) + E(X^2) = E(X^2) - E(X)^2.$$

Appliquons donc cette formule pour déterminer la variance d'une variable binomiale :

$$\begin{aligned}
D^2(X) &= E(X^2) - E(X)^2 = \sum_{i=0}^n i^2 P(X=i) - (np)^2 = \sum_{i=1}^n i^2 P(X=i) - (np)^2 = \sum_{i=1}^n (i(i-1) + i) P(X=i) - (np)^2 \\
&= \sum_{i=2}^n \frac{i(i-1)n!}{i!(n-i)!} p^i (1-p)^{n-i} + \sum_{i=1}^n iP(X=i) - (np)^2 = \sum_{i=2}^n \frac{n!}{(i-2)!(n-i)!} p^i (1-p)^{n-i} + np - (np)^2 \\
&= n(n-1)p^2 \sum_{i=2}^n \frac{(n-2)!}{(i-2)!(n-i)!} p^{i-2} (1-p)^{n-i} + np - (np)^2 = n(n-1)1^{n-2} + np - (np)^2 \\
&= n^2p^2 - np^2 + np - (np)^2 = np(1-p).
\end{aligned}$$

On pouvait également utiliser la fonction génératrice des moments pour déterminer $E(X^2)$. Commençons par expliciter cette fonction dans le cas d'une binomiale :

$$\begin{aligned}
\Psi(t) &= E(\exp(tX)) = \sum_{i=0}^n \exp(ti) P(X=i) = \sum_{i=0}^n \exp(ti) \binom{n}{i} p^i (1-p)^{n-i} = \sum_{i=0}^n \binom{n}{i} (\exp(t)p)^i (1-p)^{n-i} \\
&= (\exp(t)p + (1-p))^n.
\end{aligned}$$

Dès lors, pour trouver $E(X)$, le moment d'ordre 1, on a :

$$\left. \frac{\partial}{\partial t} \Psi(t) \right|_{t=0} = \left. \frac{\partial}{\partial t} [(1-p + \exp(t)p)^n] \right|_{t=0} = \left[n(1-p + \exp(t)p)^{n-1} p \exp(t) \right] \Big|_{t=0} = n(1)^{n-1} p = np.$$

Ensuite, pour $E(X^2)$, le moment d'ordre 2, on a :

$$\begin{aligned}
\left. \frac{\partial^2}{\partial t^2} \Psi(t) \right|_{t=0} &= \left. \frac{\partial}{\partial t} [n(1-p + \exp(t)p)^{n-1} p \exp(t)] \right|_{t=0} \\
&= \left[n(n-1)(1-p + \exp(t)p)^{n-2} p \exp(t) p \exp(t) + n(1-p + \exp(t)p)^{n-1} p \exp(t) \right] \Big|_{t=0} \\
&= \left[n(n-1)(1-p + \exp(t)p)^{n-2} p^2 \exp(2t) + n(1-p + \exp(t)p)^{n-1} p \exp(t) \right] \Big|_{t=0} \\
&= n(n-1)(1)^{n-2} p^2 + n(1)^{n-1} p = n^2 p^2 - np^2 + np.
\end{aligned}$$

Et donc $D^2(X) = E(X^2) - E(X)^2 = -np^2 + np = np(1-p)$.

4.1.3 stabilité des binomiales

On a vu comment appliquer des opérations sur des variables aléatoires en les additionnant ou multipliant. Il est cependant intéressant de voir si des opérations sur deux variables de même type donnent une troisième variable de ce même type. Une bonne manière de procéder à cela est de regarder la fonction génératrice des moments et de voir si on ré-obtient une fonction similaire. Soient $X \sim \mathcal{B}(n_1, p)$, $Y \sim \mathcal{B}(n_2, p)$ tels que $X \sqcup Y$. Soit $Z = X + Y$.

$$\Psi_Z(t) = E(\exp(tx))E(\exp(ty)) = (1 - p + \exp(t)p)^{n_1}(1 - p + \exp(t)p)^{n_2} = (1 - p + \exp(t)p)^{n_1+n_2}.$$

Donc Effectivement, on a bien $Z \sim \mathcal{B}(n_1 + n_2, p)$. Cependant, si on définit $X \sim \mathcal{B}(n, p_1)$ et $Y \sim \mathcal{B}(n, p_2)$, alors on obtient :

$$\Psi_Z(t) = \Psi_X(t)\Psi_Y(t) = (1 - p_1 + \exp(t)p_1)^n(1 - p_2 + \exp(t)p_2)^n = ((1 - p_1 + \exp(t)p_1)(1 - p_2 + \exp(t)p_2))^n.$$

Et dans ce cas là, on ne peut pas faire ressembler cela à une fonction génératrice des moments d'une binomiale.

Donc pour résumer, en sommant deux binomiales de même probabilité p mais avec chacune leur nombre de répétitions n_1 et n_2 , on obtient une troisième binomiale ayant cette même probabilité p et un nombre de répétition $n_1 + n_2$. Ce qui peut sembler intuitif, les répétitions étant indépendantes, les faire dans une première variable ou dans la seconde ne change rien au final et il y a bien $n_1 + n_2$ répétitions d'une expérience aléatoire (*a priori* identique ou assimilable) alors que si les binomiales n'ont pas la même probabilité, les sommer n'a pas de sens car elles représentent des choses trop différentes.

4.2 Variables de Poisson

Une variable de Poisson est une particularisation d'une binomiale soumise à des conditions particulières. Elles servent typiquement à représenter le nombre d'occurrences d'un événement dans un intervalle de temps T donné. On peut visualiser cela comme une binomiale

$$\mathcal{B}\left(\frac{T}{\Delta t}, a\Delta t\right).$$

Or, pour réellement parler de temps, il faut approcher le continu par le discret. On fait donc tendre $\Delta t \rightarrow 0$. Dès lors, on a :

- $n \rightarrow \infty$;
- $p \rightarrow 0$;
- $np \rightarrow \frac{T}{\Delta t}a\Delta t = aT = \lambda$.

Ici, la valeur λ est le paramètre de la distribution et est défini par $\lambda := np$.

4.2.1 Distribution

Pour exprimer la fonction de distribution d'une variable de Poisson, partons de la distribution binomiale :

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

Et le quotient binomial $\binom{n}{k}$ peut se réécrire sous la forme :

$$\frac{n!}{k!(n-k)!} = \frac{1}{k!} \prod_{i=n-k+1}^n i = \frac{1}{k!} n(n-1) \dots (n-k+2)(n-k+1) = \frac{1}{k!} n^k \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(n - \frac{k-1}{n}\right).$$

De plus, la quantité $(1 - p)^{n-k}$ peut se réécrire sous la forme :

$$(1-p)^{n-k} = \frac{(1-p)^n}{(1-p)^k} = \frac{(1-\frac{np}{n})^n}{(1-p)^k} = \frac{(1-\frac{\lambda}{n})^n}{(1-p)^k}.$$

Dès lors, pour déterminer la distribution d'une variable de Poisson, on soumet la formule de la binomiale aux conditions décidées :

$$\begin{aligned} P_\lambda(X = k) &= \lim_{p \rightarrow 0} \lim_{n \rightarrow \infty} P_{\mathcal{B}(n,p)}(X = k) = \lim_{p \rightarrow 0} \lim_{n \rightarrow \infty} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \lim_{p \rightarrow 0} \lim_{n \rightarrow \infty} \frac{n^k (1 - \frac{1}{n}) \dots (1 - \frac{k+1}{n})}{k!} p^k \frac{(1 - \frac{\lambda}{n})^n}{(1-p)^k} \\ &= \exp(-\lambda) \lim_{p \rightarrow 0} \lim_{n \rightarrow \infty} p^k n^k \frac{1}{k!} = \frac{\lambda^k}{k!} \exp(-\lambda). \end{aligned}$$

4.2.2 Vérification des axiomes

Une fois de plus, une nouvelle distribution implique une nouvelle vérification de l'application des axiomes de Kolmogorov. De plus, il faut vérifier que la fonction de distribution est bien définie sur tout son domaine. Le domaine de P_λ est \mathbb{N} , donc $k!$ est définie partout. De plus, \exp a pour ensemble d'arrivée \mathbb{R}_0^+ . Donc $P_\lambda(X = k)$ est un produit de valeurs positives, ce qui satisfait l'axiome 1. Pour le second, il faut :

$$P(E) = P\left(\bigcup_{k=0}^{\infty} X = k\right) = \sum_{k=0}^{\infty} P(X = k) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \exp(-\lambda) = \exp(-\lambda) \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \exp(-\lambda) \exp(\lambda) = 1.$$

4.2.3 Valeurs particulières

Nous avons trouvé pour une binomiale que $E(\mathcal{B}(n,p)) = np$. Faisons de même pour une variable de Poisson. Soit $X \sim P_\lambda$.

$$E(X) = \sum_{k=0}^{\infty} k P(X = k) = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} \exp(-\lambda) = \exp(-\lambda) \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda \exp(-\lambda) \exp(\lambda) = \lambda.$$

Ce résultat est cohérent : dans une binomiale, $E(X) = np$ or une Poisson est un cas particulier de binomiale. Dès lors, il est cohérent d'obtenir $E(X) = \lambda$.

Pour la variance, utilisons $D^2(X) = E(X^2) - E(X)^2$. Déterminons $E(X^2)$ à l'aide de la fonction génératrice des moments :

$$\Psi_\lambda(t) = E(\exp(tX)) = \sum_{k=0}^{\infty} \exp(tk) \frac{\lambda^k}{k!} \exp(-\lambda) = \sum_{k=0}^{\infty} \frac{(\exp(t)\lambda)^k}{k!} \exp(-\lambda) = \exp(-\lambda) \exp(\exp(t)\lambda) = \exp(\lambda(\exp(t) - 1)).$$

On peut vérifier l'espérance en dérivant une fois et en évaluant en $t = 0$, maintenant regardons $E(X^2)$:

$$\begin{aligned} \left. \frac{\partial^2}{\partial t^2} \Psi_X(t) \right|_{t=0} &= \left. \frac{\partial}{\partial t} [\exp(\lambda(\exp(t) - 1)) \lambda \exp(t)] \right|_{t=0} = \lambda \left. \frac{\partial}{\partial t} [\exp(\lambda(\exp(t) - 1) + t)] \right|_{t=0} \\ &= \lambda [\exp(\lambda(\exp(t) - 1) + t) (\lambda \exp(t) + 1)] \Big|_{t=0} = \lambda(\lambda + 1). \end{aligned}$$

Ce que l'on peut conclure en :

$$D^2(X) = E(X^2) - E(X)^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

À nouveau, ce résultat est cohérent par rapport à l'origine de la variable de Poisson qui vient de la binomiale. En effet, la variance d'une binomiale est donnée par $np(1-p)$ et dans une variable de Poisson, np tend vers λ alors que p tend vers 0. C'est donc logique que $np(1-p)$ tende vers $\lambda(1-0) = \lambda$.

4.2.4 Stabilité des Poissons

Tout comme pour les binomiales regardons si la famille des variables de Poisson est stable. Commençons par l'addition. Soient $X \sim P_{\lambda_1}$ et $Y \sim P_{\lambda_2}$ tels que $X \sqcup Y$. Soit $Z = X + Y$.

$$\begin{aligned}\Psi_Z(t) &= \Psi_X(t)\Psi_Y(t) = \exp(\lambda_1(\exp(t) - 1))\exp(\lambda_2(\exp(t) - 1)) \\ &= \exp(\lambda_1(\exp(t) - 1) + \lambda_2(\exp(t) - 1)) = \exp((\exp(t) - 1)(\lambda_1 + \lambda_2)).\end{aligned}$$

Z est donc défini comme étant également une Poisson de paramètre $\lambda = \lambda_1 + \lambda_2$. Les Poissons sont donc stables pour l'addition.

4.3 Les exponentielles négatives

En considérant un événement pouvant se produire pendant un intervalle de temps Δt avec une probabilité de $a\Delta t$ tel que $P(X \leq T + t | X > T) = P(X \geq t)$, on a :

$$F(t + \Delta t) - F(t) = (1 - F(t))a\Delta t.$$

En faisant tendre $\Delta t \rightarrow 0$ et en divisant de chaque côté par Δt , on obtient une équation différentielle :

$$F'(t) = (1 - F(t))a.$$

Qui se résout en :

$$F(t) = 1 - C \exp(-at).$$

En sachant qu'il y a pour condition initiale $F(0) = 0$, on trouve $C = 1$, ou encore $F(t) = 1 - \exp(-at)$ pour $t \geq 0$. Une telle variable aléatoire est appelée variable exponentielle négative. Si on a :

$$F(t) = \begin{cases} 1 - \exp(-at) & \text{si } t \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

Et par définition de la fonction de densité, on a également :

$$f(t) = \begin{cases} a \exp(-at) & \text{si } t \geq 0, \\ 0 & \text{sinon.} \end{cases}$$

4.3.1 Axiomes

La vérification des axiomes se fait à nouveau : le premier axiome est vérifié car $a > 0$ et \exp a \mathbb{R}_0^+ pour domaine d'arrivée. Pour le second axiome, il faut :

$$\int_{-\infty}^{\infty} f(t) dt = \int_0^{\infty} a \exp(-at) dt = [-\exp(-at)]_0^{\infty} = 0 - (-1) = 1.$$

4.3.2 Valeurs particulières

Les valeurs particulières d'une exponentielle négative (notée $T \sim \text{Exp}_a$) est donnée par :

$$E(T) = \int_{-\infty}^{\infty} t f_T(t) dt = \int_0^{\infty} t f_T(t) dt = \int_0^{\infty} t(-a \exp(-at)) dt.$$

Posons donc $f(t) = t$ et $g'(t) = -a \exp(-at)$, d'où $f'(t) = 1$ et $g(t) = \exp(-at)$. Dès lors :

$$E(T) = [t \exp(-at)]_0^{\infty} - \int_0^{\infty} \exp(-at) dt = 0 - \left[-\frac{1}{a} \exp(-at)\right]_0^{\infty} = -\left(-\frac{1}{a}\right) = \frac{1}{a}.$$

Pareillement pour la variance :

$$D^2(T) = E(T^2) - E(T)^2.$$

Passons donc par la fonction génératrice des moments (supposons $a > t$) :

$$\Psi_X(t) = E(\exp(tX)) = \int_{-\infty}^{\infty} \exp(tx)(-a \exp(-ax)) dx = -a \int_0^{\infty} \exp((t-a)x) dx.$$

Posons $\omega = (t-a)x$ donc $d\omega = (t-a) dx$. Dès lors :

$$-\frac{a}{t-a} \int_{-\infty}^0 \exp(\omega) d\omega = \frac{a}{a-t} [1-0] = \frac{a}{a-t}.$$

On peut donc calculer les moments d'ordre 1 et 2 :

$$E(X) = \frac{\partial}{\partial t} \Psi(t) \Big|_{t=0} = \frac{\partial}{\partial t} \left[\frac{a}{a-t} \right] \Big|_{t=0} = \left[\frac{a}{(a-t)^2} \right] \Big|_{t=0} = \frac{a}{a^2} = \frac{1}{a}.$$

Pour le moment d'ordre 2 :

$$E(X^2) = \frac{\partial}{\partial t} \left[\frac{a}{(a-t)^2} \right] \Big|_{t=0} = \left[\frac{2a}{(a-t)^3} \right] \Big|_{t=0} = \frac{2}{a^2}.$$

Et comme on a $D^2(X) = E(X^2) - E(X)^2$, on sait : $D^2(X) = \frac{2}{a^2} - \left(\frac{1}{a}\right)^2 = \frac{1}{a^2}$.

4.3.3 Stabilité

Comme nous allons le voir, contrairement aux familles vues précédemment, les exponentielles négatives ne sont pas stables même pour l'addition. Soient X, Y deux variables indépendantes telles que $X \sim \text{Exp}_a$, $Y \sim \text{Exp}_b$. Soit $Z = X + Y$. On sait $\Psi_Z(t) = \Psi_X(t)\Psi_Y(t) = \frac{a}{a-t} \frac{b}{b-t}$. Il n'existe pas de $c \in \mathbb{R}^+$ tel que $\Psi_Z(t) = \frac{c}{c-t}$. Ce qui montre que les exponentielles négatives ne sont pas stables pour l'addition.

4.3.4 Propriété d'oubli

Comme mentionné plus haut, pour une telle variable, on a :

$$P(T \leq t + \Delta t |) = P(T \geq \Delta t).$$

Pour le *démontrer*, il faut utiliser la définition de probabilité conditionnelle.

$$\begin{aligned} P(T \leq t + \Delta t | T \geq t) &= \frac{P((T \leq t + \Delta t) \cap (T \geq t))}{P(T \geq t)} = \frac{P(t \leq T \leq t + \Delta t)}{1 - P(T \leq t)} \\ &= \frac{F(t + \Delta t) - F(t)}{1 - F(t)} = \frac{1 - \exp(-a(t + \Delta t)) - (1 - \exp(-at))}{1 - (1 - \exp(-at))} = \frac{-\exp(-a(t + \Delta t)) + \exp(-at)}{\exp(-at)} \\ &= \frac{\exp(-at) - \exp(-a(t + \Delta t))}{\exp(-at)} = 1 - \frac{\exp(-a(t + \Delta t))}{\exp(-at)} = 1 - \exp(-a\Delta t) = F(\Delta t). \end{aligned}$$

Ce résultat veut dire que peu importe quand tombe t , la probabilité que l'événement apparaisse dans l'intervalle Δt suivant est constante.

4.4 Variable normale

Une variable aléatoire est soit continue soit discrète. SI elle est continue, elle peut être une exponentielle négative comme vu ci-dessus, mais elle peut également être une *variable normale*. Ce type de variable représente une répartition totalement aléatoire et se note $\mathcal{N}(\mu, \sigma)$ où μ et σ sont respectivement la moyenne et l'écart type. Une normale suit la distribution suivante :

$$f_{\mathcal{N}(\mu, \sigma)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

4.4.1 Axiomes de Kolmogorov

Une fois de plus, vérifions les axiomes de Kolmogorov afin de vérifier l'appellation *fonction de densité* :

$$\int_{-\infty}^{+\infty} f_{\mathcal{N}(\mu, \sigma)}(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx.$$

Posons $y := \frac{x - \mu}{\sqrt{2}\sigma}$, dès lors $dy = \frac{dx}{\sqrt{2}\sigma}$.

$$\int_{-\infty}^{+\infty} f_{\mathcal{N}(\mu, \sigma)}(x) dx = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \exp(-y^2) dy.$$

Or l'intégrale sur \mathbb{R} de $\exp(-x^2)$ est une intégrale bien connue et vaut $\sqrt{\pi}$. Dès lors, l'intégrale sur \mathbb{R} de la fonction de distribution d'une normale est bien égale à 1.

De plus, $f_{\mathcal{N}(\mu, \sigma)}$ est un produit de valeurs positives pour tout x et est donc positive partout.

4.4.2 Intérêts de la normale $\mathcal{N}(\mu, \sigma)$

En plus de pouvoir modéliser une répartition aléatoire, les variables aléatoires ont pour avantage d'être très simples à calculer : un ordinateur n'a aucun mal à évaluer $\exp(x)$ pour n'importe quel x , contrairement à une factorielle $k!$ qui devient difficile à calculer au dessus de valeurs de k valant $\simeq 40$. On va donc utiliser les variables normales afin d'approximer d'autres distributions en les soumettant à des caractéristiques particulières.

De plus, les normales sont centrales en probabilités dû à un théorème (le théorème central limite qui sera vu plus loin).

4.4.3 Valeurs particulières

La définition d'une variables $\mathcal{N}(\mu, \sigma)$ fait intervenir deux paramètres : μ et σ . Prouvons que ces deux variables correspondent bien à la moyenne et à l'écart type de la distribution. Soit $X \sim \mathcal{N}(\mu, \sigma)$.

$$E(X) = \int_{-\infty}^{+\infty} x f_{\mathcal{N}(\mu, \sigma)}(x) dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx.$$

Posons $y := \frac{x-\mu}{\sigma}$. Dès lors $dy = \frac{dx}{\sigma}$. On a donc :

$$\begin{aligned} &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} (\sqrt{2\sigma^2}y + \mu) \exp(-y^2) \sqrt{2\sigma^2} dy \\ &= \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{+\infty} (\sqrt{2\sigma^2}y + \mu) \exp(-y^2) dy \right] \\ &= \frac{1}{\sqrt{\pi}} \left[\int_{-\infty}^{+\infty} \sqrt{2\sigma^2}y \exp(-y^2) dy + \int_{-\infty}^{+\infty} \mu \exp(-y^2) dy \right] \\ &= \frac{1}{\sqrt{\pi}} \left(0 + \mu \int_{-\infty}^{+\infty} \exp(-y^2) dy \right) = \frac{1}{\sqrt{\pi}} \mu \sqrt{\pi} = \mu. \end{aligned}$$

En effet, vu que la première intégrale est une fonction impaire sur des bornes opposées, elle vaut 0. On a donc bien $E(\mathcal{N}(\mu, \sigma)) = \mu$. Il reste à prouver que $D^2(\mathcal{N}(\mu, \sigma)) = \sigma^2$. Pour cela, utilisons la propriété disant $D^2(X) = E(X^2) - E(X)^2$. Donc afin de trouver $E(X^2)$, utilisons la fonction génératrice des moments donnée par $\Psi_X(t) = E(\exp(tx))$:

$$\begin{aligned} \Psi_X(t) &= \int_{-\infty}^{+\infty} \exp(tx) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\left[\frac{(x-\mu)^2}{2\sigma^2} - tx\right]\right) dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\left[\frac{(x-\mu)^2}{2\sigma^2} - tx + \mu t + \frac{(\sigma t)^2}{2} - \mu t - \frac{(\sigma t)^2}{2}\right]\right) dx \\ &= \frac{\exp(\mu t + \frac{(\sigma t)^2}{2})}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp\left(-\left(\frac{x-\mu}{\sqrt{2}\sigma} - \frac{\sigma t}{\sqrt{2}}\right)^2\right) dx. \end{aligned}$$

Posons $y := \frac{x-\mu}{\sqrt{2}\sigma} - \frac{\sigma t}{\sqrt{2}}$, donc $dy = \frac{dx}{\sqrt{2}\sigma}$. Dès lors :

$$\begin{aligned} \Psi_X(t) &= \frac{\exp(\mu t + \frac{(\sigma t)^2}{2})}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \exp(-y^2) \sqrt{2}\sigma dy \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\mu t + \frac{t^2\sigma^2}{2}) \sqrt{2}\sqrt{\pi}\sigma \\ &= \exp(\mu t + \frac{t^2\sigma^2}{2}). \end{aligned}$$

Dérivons la maintenant deux fois et instancions-la en $t = 0$ afin de déterminer $E(X^2)$:

$$\frac{\partial}{\partial x} \Psi_X(t) = \Psi_X(t)(\mu + t\sigma^2).$$

Et donc :

$$\frac{\partial}{\partial x} \frac{\partial}{\partial x} \Psi_X(t) = \Psi_X(t)(\mu + t\sigma^2)^2 + \sigma^2 \Psi_X(t).$$

Qui, instancié en 0 donne : $1(\mu)^2 + \sigma^2 1$. Dès lors, on sait trouver $D^2(X) = E(X^2) - E(X)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$.

Nous avons bien trouvé, ici, que μ et σ représentent respectivement la moyenne et l'écart type de la distribution normale.

4.4.4 Stabilité des normales

Avec la fonction génératrice des moments, il nous est possible de déterminer si la famille des normales est une famille stable ou non. Soient $X \sim \mathcal{N}(\mu_1, \sigma_1)$ et $Y \sim \mathcal{N}(\mu_2, \sigma_2)$, $Z = X + Y$. On a :

$$\begin{aligned} \Psi_Z(t) &= \Psi_X(t) \Psi_Y(t) \exp\left(\mu_1 t + \frac{(t\sigma_1)^2}{2}\right) \exp\left(\mu_2 t + \frac{(t\sigma_2)^2}{2}\right) \\ &= \exp\left((\mu_1 + \mu_2)t + t^2 \frac{\sigma_1^2 + \sigma_2^2}{2}\right). \end{aligned}$$

Ce qui montre donc que la somme de deux normales est à nouveau une normale, telle que $Z \sim \mathcal{N}(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

4.4.5 La normale $\mathcal{N}(0, 1)$

Soient $X \sim \mathcal{N}(\mu, \sigma)$, $Z = \frac{X - \mu}{\sigma}$. On sait exprimer :

$$\begin{aligned} \Psi_Z(t) &= E\left(\exp\left(t \frac{X - \mu}{\sigma}\right)\right) = E\left(\exp\left(\frac{tx}{\sigma}\right) \exp\left(\frac{-t\mu}{\sigma}\right)\right) = \exp\left(\frac{-t\mu}{\sigma}\right) E\left(\exp\left(\frac{tx}{\sigma}\right)\right) \\ &= \exp\left(\frac{-t\mu}{\sigma}\right) \Psi_X\left(\frac{t}{\sigma}\right) = \exp\left(\frac{-t\mu}{\sigma}\right) \exp\left(\mu \frac{t}{\sigma} + \frac{t^2}{2} \frac{\sigma^2}{\sigma^2}\right) = \exp\left(\frac{-\mu t}{\sigma} + \frac{\mu t}{\sigma} + \frac{t^2}{2}\right) = \exp\left(\frac{t^2}{2}\right) = \Psi_{\mathcal{N}(0,1)}(t). \end{aligned}$$

La fonction génératrice des moments de Z est la fonction génératrice des moments d'une normale de moyenne 0 et d'écart-type 1. Cette propriété veut dire que toute distribution normale peut être ramenée à une normale $\mathcal{N}(0, 1)$. Ce qui est extrêmement intéressant : pour calculer $P(\mathcal{N}(\mu, \sigma) \leq \alpha)$ pour toute distribution normale, il faudrait calculer une infinité de tables de valeurs. Maintenant, on sait que $P(\mathcal{N}(\mu, \sigma) \leq \alpha) = P(\mathcal{N}(0, 1) \leq \frac{\alpha - \mu}{\sigma})$. De plus, comme une variable $\mathcal{N}(0, 1)$ est symétrique, $\forall \delta \geq 0$, $P(\mathcal{N}(0, 1) \leq -\delta) = 1 - P(\mathcal{N}(0, 1) \leq \delta)$. Les variables normales sont donc finalement très simples à calculer : il existe énormément de tables avec des précisions différentes.

5 Théorèmes importants

5.1 Théorème central limite

Soient V_1, V_2, \dots, V_n , des variables aléatoires quelconques³, $V = \sum_{i=1}^n V_i$ telles que $V_i \perp V_j \forall i \neq j$. Alors :

$$\frac{V - \mu_V}{\sigma_V} \xrightarrow{n \rightarrow +\infty} \mathcal{N}(0, 1).$$

³Possiblement de distributions totalement différentes.

Démonstration : Soient V_1, \dots, V_n des variables aléatoires iid⁴. Soient $V = \sum_{i=1}^n V_i$, $Z = \frac{V-E(V)}{D(V)}$. Supposons que $\forall i : E(V_i) = \bar{\mu}$, $D(V_i) = \bar{\sigma}$. On sait que :

$$\Psi_Z(t) = E \left[\exp \left(t \frac{V - E(V)}{D(V)} \right) \right].$$

On sait également que :

$$E(V) = E \left(\sum_{i=1}^n V_i \right) = \sum_{i=1}^n E(V_i) = \sum_{i=1}^n \bar{\mu} = n\bar{\mu}.$$

Posons $W_i := V_i - \bar{\mu}$. Dès lors, on sait que $E(W_i) = \bar{\mu} - \bar{\mu} = 0$. Donc :

$$\begin{aligned} \Psi_Z(t) &= E \left[\exp \left(\frac{t}{\sigma} (V - E(V)) \right) \right] = E \left[\exp \left(\frac{t}{\sigma} \left(\sum_{i=1}^n V_i - \sum_{i=1}^n \bar{\mu} \right) \right) \right] \\ &= E \left[\exp \left(\frac{t}{\sigma} \sum_{i=1}^n (V_i - \bar{\mu}) \right) \right] = E \left[\exp \left(\frac{t}{\sigma} \sum_{i=1}^n W_i \right) \right]. \end{aligned}$$

On sait également que :

$$\sigma^2 = D(V) = D \left(\sum_{i=1}^n V_i \right) = \sum_{i=1}^n D^2(V_i) = \sum_{i=1}^n \bar{\sigma} = n\bar{\sigma}.$$

Or en reprenant l'égalité précédente, on a :

$$\begin{aligned} \Psi_Z(t) &= E \left[\exp \left(\frac{t}{\sigma} \sum_{i=1}^n W_i \right) \right] = E \left[\prod_{i=1}^n \exp \left(\frac{t}{\sigma} W_i \right) \right] = \prod_{i=1}^n E \left[\exp \left(\frac{t}{\sigma} W_i \right) \right] = \prod_{i=1}^n \Psi_{W_i} \left(\frac{t}{\sigma} \right) \\ &\stackrel{iid}{=} \left(\Psi_{W_j} \left(\frac{t}{\sigma} \right) \right)^n \quad \text{où } j \in \{1, 2, \dots, n\} \text{ est choisi sans importance.} \end{aligned}$$

En exprimant $\Psi_{W_j}(t)$ sous la forme suivante :

$$\Psi_{W_j}(t) = E[\exp(tW_j)] = E \left[\sum_{k=0}^{+\infty} \frac{(tW_j)^k}{k!} \right] = \sum_{k=0}^{+\infty} E(W_j^k) \frac{t^k}{k!} = 1 + E(W_j)t + E \left(\frac{W_j^2}{2} \right) t^2 + R(t^3),$$

où $R(t^3)$ est un reste en $O(t^3)$ et $E(W_j) = 0$, on peut déterminer $D^2(W_j) = E(W_j^2) - E(W_j)^2 = E(W_j^2)$ et $D^2(W_j) = D^2(V_j - \bar{\mu}) = D^2(V_j) = \bar{\sigma}^2$. Dès lors, on a :

$$\Psi_Z(t) = \left(\Psi_{W_j} \left(\frac{t}{\sigma} \right) \right)^n = \left(\Psi_{W_j} \left(\frac{t}{\sqrt{n\bar{\sigma}}} \right) \right)^n = \left(1 + \frac{\bar{\sigma}^2}{2} \frac{t^2}{n\bar{\sigma}^2} + R \left(\left(\frac{t}{\sqrt{n\bar{\sigma}}} \right)^3 \right) \right)^n.$$

⁴Indépendantes et Identiquement Distribuées.

Le reste en $\frac{1}{\sqrt{n^3}}$ devient négligeable quand $n \rightarrow +\infty$. Dès lors, on a :

$$\Psi_Z(t) = \left(1 + \frac{t^2}{2n}\right)^n \xrightarrow{n \rightarrow +\infty} \exp\left(\frac{t^2}{2}\right) = \Psi_{\mathcal{N}(0,1)}(t).$$

Cela prouve bien qu'une somme infinie de variables aléatoires iid⁵ est une variable normale. Les variables normales sont donc effectivement totalement **centrales** dans le domaine des probabilités.ⁱ

5.2 Approximation par une normale

5.2.1 Approximer une binomiale

Une variable binomiale peut vite devenir difficile à évaluer à cause des factorielles qui deviennent vite élevées et qu'il faut sommer bon nombre de fois. Il est donc appréciable de trouver un moyen d'approximer la valeur d'une binomiale. Comme on sait que la binomiale est stable pour l'addition, on peut écrire :

$$B(n, p) = \sum_{i=1}^n B(1, p).$$

Par le théorème central limite, on sait que plus n est grand, plus l'approximation de $\frac{B(n,p)-np}{\sqrt{np(1-p)}}$ par une $\mathcal{N}(0, 1)$ est précise. Dès lors, pour des grandes valeurs de n , on peut dire que :

$$P(X \geq \beta) = P\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{\beta - np}{\sqrt{np(1-p)}}\right).$$

En posant $\alpha := \frac{\beta - np}{\sqrt{np(1-p)}}$, on a $P(X \geq \beta) = 1 - F_{\mathcal{N}(0,1)}(\alpha)$.

On a donc deux moyens différents d'approximer une binomiale. Le premier étant une poisson :

$$\begin{array}{ccc} n \rightarrow +\infty \\ p \rightarrow 0 \\ np \rightarrow \lambda \\ B(n, p) \longrightarrow P_\lambda. \end{array}$$

Le second étant une normale :

$$B(n, p) \xrightarrow{n \rightarrow +\infty} (np, \sqrt{np(1-p)}).$$

Pour déterminer quelle est la meilleure approximation à choisir, il faut déterminer celle qui correspond le mieux aux hypothèses posées⁶.

5.2.2 Approximation d'une Poisson

Tout comme une binomiale, une Poisson peut être approximée par une normale selon une certaine hypothèse. En effet, si $\lambda \rightarrow +\infty$, alors $P_\lambda \rightarrow \mathcal{N}(\lambda, \sqrt{\lambda})$.

Pour le montrer, il faut à nouveau passer par la stabilité des Poissons : $P_\lambda = \lambda P_1 + P_{\lambda - \lfloor \lambda \rfloor}$. Par le théorème central limite, on a la même propriété, à savoir : une normale $\mathcal{N}(\lambda, \sqrt{\lambda})$ devient une bonne approximation d'une Poisson P_λ quand $n \rightarrow +\infty$.

⁵hypothèse pour cette démonstration-ci qui peut être supprimée à l'aide d'une autre démonstration.

⁶Sur n , p et np .

5.3 Théorème de Binaymé-Tchebycheff

Soient X une variable aléatoire et $k \in \mathbb{R}_0^+$. Alors :

$$P\left(|X - E(X)| \geq kD(X)\right) \leq k^{-2}.$$

Démonstration : Soient X une variable aléatoire quelconque et k un réel positif. On sait que :

$$\begin{aligned}\sigma_X^2 &= D^2(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x) dx \\ &= \int_{-\infty}^{\mu_X - k\sigma_X} (x - \mu_X)^2 f_X(x) dx + \int_{\mu_X - k\sigma_X}^{\mu_X + k\sigma_X} (x - \mu_X)^2 f_X(x) dx + \int_{\mu_X + k\sigma_X}^{+\infty} (x - \mu_X)^2 f_X(x) dx \\ &\geq \int_{-\infty}^{\mu_X - k\sigma_X} (x - \mu_X)^2 f_X(x) dx + \int_{\mu_X + k\sigma_X}^{+\infty} (x - \mu_X)^2 f_X(x) dx.\end{aligned}$$

Soient $P_1 =]-\infty, \mu_X - k\sigma_X]$, $P_2 = [\mu_X + k\sigma_X, +\infty[$. On a :

$$\begin{aligned}\forall x \in P_1 : \exists \epsilon \geq 0 \text{ t.q. } x &= \mu_X - (k\sigma_X + \epsilon) \\ (x - \mu_X)^2 &= (\mu_X - (k\sigma_X + \epsilon) - \mu_X)^2 \geq k^2\sigma_X^2, \\ \forall x \in P_2 : \exists \epsilon \geq 0 \text{ t.q. } x &= \mu_X + (k\sigma_X + \epsilon) \\ (x - \mu_X)^2 &= (\mu_X + (k\sigma_X + \epsilon) - \mu_X)^2 \geq k^2\sigma_X^2.\end{aligned}$$

Dès lors :

$$\begin{aligned}\sigma_X^2 &\geq k^2\sigma_X^2 \int_{-\infty}^{\mu_X - k\sigma_X} f_X(x) dx + k^2\sigma_X^2 \int_{\mu_X + k\sigma_X}^{+\infty} f_X(x) dx \\ &= k^2\sigma_X^2 \left[\int_{-\infty}^{+\infty} f_X(x) dx - \int_{\mu_X - k\sigma_X}^{\mu_X + k\sigma_X} f_X(x) dx \right] \\ &= k^2\sigma_X^2 \left(1 - P(|X - \mu_X| \leq k\sigma_X) \right) \\ &= k^2\sigma_X^2 \left(P(|X - \mu_X| \geq k\sigma_X) \right).\end{aligned}$$

En divisant de part et d'autre par $k^2\sigma_X^2$, on obtient :

$$\frac{1}{k^2} \geq P(|X - \mu_X| \geq k\sigma_X). \quad \square$$

5.4 Théorème de Bernoulli

Si $F \sim B(n, p)$, alors :

$$\forall \epsilon \geq 0 : P\left(\left|\frac{F}{n} - p\right| \geq \epsilon\right) \xrightarrow{n \rightarrow +\infty} 0.$$

Démonstration : Soient $F \sim B(n, p)$, $V = \frac{F}{n}$, $\mu_V = E(V) = \frac{1}{n}E(F) = \frac{1}{n}np = p$. On sait que :

$$P\left(\left|\frac{F}{n} - p\right| \geq \epsilon\right) = P(|V - \mu_V| \geq \epsilon).$$

Or $D^2(V) = D^2\left(\frac{F}{n}\right) = \frac{1}{n^2} D^2(F) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$. Posons $k := \frac{\epsilon}{\sqrt{p(1-p)n}}$. Dès lors :

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{F}{n} - p\right| \geq \epsilon\right) = \lim_{n \rightarrow +\infty} P(|V - \mu_V| \geq k\sigma_V) \leq \lim_{n \rightarrow +\infty} \frac{1}{\frac{\epsilon^2}{\sqrt{\frac{p(1-p)}{n}}^2}} = \frac{p(1-p)}{n\epsilon^2} \lim_{n \rightarrow +\infty} \frac{1}{n} = 0.$$

6 Inférence statistique

La population est l'ensemble des objets étudiés et n'est pas observable (soit pas en entier, soit pas directement). **Un échantillon** est un sous-ensemble (aléatoire et simple) réellement observé de la population. Le sens de l'inférence statistique est de voir en quelle mesure il est possible inférer des conclusions venant de l'échantillon sur la population.

Si l'échantillon n'est pas aléatoire simple, il risque d'y avoir un biais qui peut être important.

6.1 Biais des valeurs particulières

Soit une population P . Soit A l'expérience correspondant au prélèvement d'un échantillon d'effectif n dans P . Chaque échantillon est sous la forme x_1, x_2, \dots, x_n . Soient X_1, X_2, \dots, X_n des expériences aléatoires telles que X_i est l'expérience relative à la valeur de x_i . Du fait que l'échantillon est aléatoire et simple, on sait que la distribution de probabilité de X_i est la même que la distribution de probabilité de la population pour tout i . De plus, chaque échantillon a une moyenne $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$. Notons \bar{X} l'expérience aléatoire associant chaque échantillon à sa moyenne.

En supposant que la population admet pour moyenne μ , on peut déterminer l'espérance de la variable aléatoire représentant la moyenne de l'échantillon :

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = n \frac{\mu}{n} = \mu.$$

Dès lors, on peut affirmer que *la moyenne de la moyenne est la moyenne*, ce qui veut dire que la valeur moyenne de la moyenne de l'échantillon est égale à la moyenne de la population. Ou encore : en moyenne, l'évaluation de la moyenne de l'échantillon donnera la moyenne de la population. Cela veut dire que la prise d'échantillon est un **estimateur non biaisé**.

En supposant $X_i \sqcup X_j \forall i \neq j$ et en supposant que la population est d'écart type σ , on peut déterminer l'écart type de l'échantillon (dispersion par rapport à la moyenne) :

$$D^2(\bar{X}) = D^2\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n D^2(X_i).$$

Comme tous les X_i sont de même distribution que la population, $D^2(X_i) = \sigma^2 \forall i$. Dès lors, $D^2(\bar{X}) = \frac{1}{n^2} n \sigma^2 = \frac{1}{n} \sigma^2$. Quand n est grand, $D^2(\bar{X})$ est petit et donc \bar{X} représente bien la moyenne de la population. De plus, si n est *très grand*, alors le théorème central-limite dit que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Soit S l'événement aléatoire relatif à la variance de l'échantillon. Dès lors :

$$E(S^2) = E\left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) - E(\bar{X}^2).$$

Or $D^2(X_i) = E(X_i^2) - E(X_i)^2$, donc $E(X_i^2) = D^2(X_i) + E(X_i)^2$. Et de manière similaire, $E(\bar{X}^2) = D^2(\bar{X}) + E(\bar{X})^2$. Donc :

$$E(S^2) = \frac{1}{n} \sum_{i=1}^n (D^2(X_i) + E(X_i)^2) - (D^2(\bar{X}) + E(\bar{X})^2) = \frac{1}{n} n(\sigma^2 + \mu^2) - \left(\frac{\sigma^2}{n} + \mu^2\right) = \frac{n-1}{n} \sigma^2.$$

Contrairement à \bar{X} , S est un opérateur biaisé : en moyenne la variance de l'échantillon correspond à $\frac{n-1}{n}$ fois la variance de la population. Soit $\tilde{S}^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$. On a donc :

$$E(\tilde{S}) = \frac{1}{n-1} E\left(\sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n}{n-1} E\left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \frac{n}{n-1} E(S^2) = \sigma^2.$$

\tilde{S} est donc un meilleur estimateur que S^2 pour l'écart-type et la variance.

6.2 Estimer un paramètre

Quand il faut pouvoir estimer le paramètre d'une distribution de probabilité, il faut soit partir de sa définition (ex : le paramètre λ d'une Poisson est l'espérance de la distribution) et la retrouver comme cela ($\lambda = \frac{1}{n} \sum_{i=1}^n x_i$ pour la Poisson), soit utiliser le maximum de vraisemblance.

Le maximum de vraisemblance (maximum likelihood en anglais) d'un paramètre θ est la valeur de θ qui maximise la probabilité d'observation de l'échantillon.

On note $L_\theta(x_1, \dots, x_n) := P(X_1 = x_1, \dots, X_n = x_n)$ la probabilité que l'échantillon observé soit x_1, x_2, \dots, x_n en ayant le paramètre θ (la fonction de vraisemblance). On a donc :

$$L_\theta(x_1, \dots, x_n) = \prod_{i=1}^n P(X_i = x_i; \theta).$$

Afin de maximiser la probabilité selon le paramètre θ , il faut trouver pour quelle valeur la dérivée voit 0 :

$$\frac{\partial}{\partial \theta} L_\theta(x_1, \dots, x_n) = 0.$$

Cependant, afin de ne pas devoir dériver un produit, on passe par la vraisemblance logarithmique (log-likelihood en anglais) :

$$\frac{\partial}{\partial \theta} \log L_\theta(x_1, \dots, x_n) = \frac{\partial}{\partial \theta} \log \prod_{i=1}^n P(X_i = x_i; \theta) = \sum_{i=1}^n \frac{\partial}{\partial \theta} \log P(X_i = x_i).$$

Par exemple, dans le cas d'une variable de Poisson, trouver le paramètre λ sur base de l'échantillon x_1, \dots, x_n se fait de la manière suivante :

$$\begin{aligned} \sum_{i=1}^n \frac{\partial}{\partial \lambda} \log P_\lambda(X_i = x_i) &= \sum_{i=1}^n \frac{\partial}{\partial \lambda} \log \left(\frac{\lambda^{x_i}}{x_i!} \exp(-\lambda) \right) = \sum_{i=1}^n \frac{\partial}{\partial \lambda} (\log \lambda^{x_i} - \log x_i! + \log \exp - \lambda) \\ &= \sum_{i=1}^n \frac{\partial}{\partial \lambda} (x_i \log \lambda - \log x_i! - \lambda) = \sum_{i=1}^n \left(x_i \frac{1}{\lambda} - 1 \right) = \frac{1}{\lambda} \sum_{i=1}^n x_i - n. \end{aligned}$$

Or cette valeur vaut 0, donc :

$$\frac{1}{\lambda} \sum_{i=1}^n x_i - n = 0 \iff \lambda = \frac{1}{n} \sum_{i=1}^n x_i.$$

Ce qui dit que pour estimer le paramètre λ d'une poisson il faut faire la moyenne des éléments. On retombe bien sur la définition.

6.3 Intervalle de confiance

L'idée derrière les intervalles de confiance est de trouver un intervalle de valeurs où le paramètre inconnu apparaît avec une haute probabilité.

Soit une population de moyenne μ inconnue et d'écart-type σ connu. $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ car n est suffisamment grand. Soit $\epsilon \in [0, 1[$. On cherche un intervalle $[-z_\epsilon, z_\epsilon]$ tel que :

$$P\left(-z_\epsilon \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z_\epsilon\right) = 1 - \epsilon.$$

Ou encore :

$$P\left(\mu \in \left[\bar{X} \pm z_\epsilon \frac{\sigma}{\sqrt{n}}\right]\right) = 1 - \epsilon$$

L'intervalle dépend de z_ϵ qui dépend d' ϵ . Donc plus ϵ est grand, plus la probabilité attendue est basse donc plus l'intervalle est réduit (plus $|z_\epsilon|$ est petit). La longueur de l'intervalle est donc $2z_\epsilon \frac{\sigma}{\sqrt{n}}$. En imposant une limite d telle que $d \geq 2z_\epsilon \frac{\sigma}{\sqrt{n}}$, on sait déterminer le nombre minimum d'effectifs à avoir dans l'intervalle pour obtenir une telle précision : $n \geq \left(\frac{2z_\epsilon \sigma}{d}\right)^2$.

6.4 Test d'hypothèse

Si \bar{X} est la variable aléatoire correspondant à la moyenne de l'échantillon, on sait que $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Posons $T := \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$. Soient deux hypothèses $H0$ et $H1$ telles que $H0 \equiv \neg H1$.

Dès lors, en reprenant la notion d'intervalle de confiance, il faut déterminer laquelle des deux hypothèses $H0$ ou $H1$ est admissible. L'intervalle de confiance permet de délimiter *jusqu'à quel point l'hypothèse peut s'écarter de la réalité pour qu'elle soit admissible*. Par exemple, si \bar{X} correspond à la moyenne des notes à un examen (la valeur observée est 5), si l'échantillon est d'effectif 9, que l'écart-type est 1, prenons pour hypothèses $H0 \equiv \mu = 12$ donc $H1 \equiv \mu \neq 12$. Pour $H0$, on a :

$$T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{5 - 12}{\frac{1}{\sqrt{9}}} = -21.$$

Or en prenant $\epsilon = 0.05$, on veut trouver un intervalle contenant μ avec une probabilité de 0.95. La valeur de z_ϵ pour $\epsilon = 0.05$ est 1.96. Dès lors, on voit bien que $-21 \notin [-1.96, 1.96]$. L'hypothèse $H0$ est donc à rejeter. En faisant le même calcul si \bar{X} avait pour valeur 12.5, on aurait eu :

$$T = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{12.5 - 12}{\frac{1}{\sqrt{9}}} = \frac{3}{2}.$$

En prenant le même ϵ , on peut admettre l'hypothèse $H0$ car $\frac{3}{2} \in [-1.96, 1.96]$.

Le lien *formel* entre test d'hypothèse en intervalle de confiance est que l'hypothèse $H0$ est rejetée si et seulement si l'hypothèse n'entre pas dans l'intervalle de confiance :

$$RHO \iff \mu \notin \left[\bar{X} \pm z_\epsilon \frac{\sigma}{\sqrt{n}}\right].$$

Le lien peut également être formulé comme suit : *la probabilité que l'hypothèse soit rejetée en faisant un faux négatif est ϵ* . En effet, un test d'hypothèse est concluant si l'hypothèse H_0 est rejetée et qu'elle n'est pas valable ou si l'hypothèse H_1 est rejetée et qu'elle n'est pas valable. Dans le cas où H_0 est rejetée alors qu'elle est valable (idem pour H_1) est appelé faux négatif. On a donc :

$$P(RH_0 | H_0) = \epsilon.$$

On y voit que en prenant ϵ de plus en plus petit, on s'attend à une probabilité de plus en plus grande de trouver H_0 et donc pour $\epsilon \rightarrow 0$, H_0 ne sera jamais rejetée. Dès lors, $P(RH_0 | H_0) \rightarrow 0 = \epsilon$. À l'inverse, quand $\epsilon \rightarrow 1$, l'intervalle est réduit jusqu'à devenir le singleton $\{\mu\}$ et donc H_0 sera toujours rejeté. Donc $P(RH_0 | H_0) \rightarrow 1 = \epsilon$.

6.5 Student et χ^2

Soient $X_1, X_2, \dots, X_n \sim \mathcal{N}(0, 1)$ des variables aléatoires indépendantes. Alors :

$$X := \sum_{i=1}^n X_i^2 \sim \chi_{(n)}^2.$$

Soient $V_0, V_1, V_2, \dots, V_n \sim \mathcal{N}(0, \sigma)$ des variables aléatoires indépendantes. Alors :

$$V := \frac{V_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n V_i^2}} \sim t_{(n)}.$$

X est appelée χ^2 (*chi carrée*) et V est appelée student. Toutes deux sont de degré de liberté n .

Si $\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ est la variable correspondant à la moyenne de l'échantillon, n est l'effectif de l'échantillon et S son écart type, alors :

$$\frac{nS^2}{\sigma^2} \sim \chi_{(n-1)}^2.$$

De plus, pour une student d'écart type $\sigma = 1$, on a :

$$t_{(n)} = \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{1}{n} \sum_{i=1}^n \mathcal{N}(0, 1)^2}} = \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{1}{n} \chi_{(n)}^2}}.$$

Dès lors :

$$t_{(n-1)} = \frac{\mathcal{N}(0, 1)}{\sqrt{\frac{1}{n-1} \chi_{(n-1)}^2}} = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{1}{(n-1)} \frac{nS^2}{\sigma^2}}} = \frac{(\bar{X} - \mu) \sqrt{n} \sqrt{n-1} \sigma}{\sigma \sqrt{n} S} = \frac{(\bar{X} - \mu) \sqrt{n-1}}{S}$$

Or, si $\frac{(\bar{X} - \mu) \sqrt{n-1}}{S} \sim t_{(n-1)}$, alors :

$$P\left(-t_{(n-1), \epsilon} \leq \frac{(\bar{X} - \mu) \sqrt{n-1}}{S} \leq t_{(n-1), \epsilon}\right) = 1 - \epsilon.$$

Où $\pm t_{(n-1), \epsilon}$ sont les quantiles d'une student à $(n-1)$ degrés de liberté. De là :

$$P\left(\mu \in \left[\bar{X} \pm t_{(n-1), \epsilon} \frac{S}{\sqrt{n-1}}\right]\right) = 1 - \epsilon.$$

Ce résultat permet de se débarrasser de l'hypothèse disant que l'on connaît σ , l'écart type de la population. En effet, il est possible de faire un intervalle de confiance et des tests d'hypothèses sans avoir à connaître σ .

6.6 Analyse de deux populations distinctes

Soient deux populations P_1, P_2 suivant des distributions normales $\mathcal{N}(\mu_1, \sigma_1)$ et $\mathcal{N}(\mu_2, \sigma_2)$. Soient deux échantillons $E_1 \subset P_1, E_2 \subset P_2$ d'effectif respectif n_1 et n_2 . Soient \bar{X}_1 et \bar{X}_2 les variables aléatoires des moyennes de E_1 et E_2 . On sait que $\bar{X}_1 \sim \mathcal{N}\left(\mu_1, \frac{\sigma_1}{\sqrt{n_1}}\right)$ et $\bar{X}_2 \sim \mathcal{N}\left(\mu_2, \frac{\sigma_2}{\sqrt{n_2}}\right)$

Pour évaluer si les deux populations ont la même moyenne, on procède à nouveau à un test d'hypothèse. Posons :

$$T := \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}.$$

Si les moyennes sont égales ($\mu_1 = \mu_2$), alors on procède au test d'hypothèse suivant :

$$\mu_1 = \mu_2 \iff 0 \in \left[\bar{X}_1 - \bar{X}_2 \pm z_\epsilon \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right] \iff \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_\epsilon.$$