Objectifs : Maîtriser la probabilité et la statistique (Séries statistiques, probabilité, variables aléatoires)

Contenu de la matière : Les Chapitres		
Statistiques	**Probabilités**	
1- *Notion et définitions de base (1S)*	**1-** *Analyse combinatoire (1S)*	**4-** *Variables aléatoires (1S)*
2- *Séries statistiques à une variable (3S)*	**2-** *Introduction aux probabilités (2S)*	**5-** *Lois de probabilité discrètes (1S)*
3- *Séries statistiques à deux variables (3S)*	**3-** *Conditionnement et indépendance (1S)*	**6-** *Lois de probabilité continues(2S)*

Références : **1** : Bernard Verlant_Statistique et probabilités BTS Industriel, **2** : Daniel Fredon_Mathématiques Statistique et probabilités en 30 fiches **3** : Murray R. spiegel_Theory and Problems of Statistics 3edition.

Partie A : Statistiques Chapitre 1: Définitions de base 1 semaine

Définition du problème→ Collecte→Dépouillement et Tri→Tableaux Statistiques→
→Représentations Graphiques→Indicateurs Statistiques→Conclusions

Statistiques : Descriptive **(الوصفي),** Inférentielle **(الإستنتاجي)**

Population : C'est un ensemble d'éléments à étudier (C'est l'univers de référence, finie de nombre n)

Individus : (unités statistiques) : éléments de la population

Echantillon : C'est est un sous-ensemble de la population de taille $n_e \leq n$ individus

Caractère Les caractéristiques étudiées (Variables,) qualitatives (non mesurables), quantitatives (mesurables),

Modalités : Attributs ou valeurs possibles du caractère (exhaustives (شاملة), incompatibles غير متوافق)

Variables statistiques (quantitative), discrètes (fini ou dénombrable), continues (n'est pas dénombrable).

Statistique à une dimension : étudier un seul caractère série deux dimensions deux caractères sont étudiés

Chapitre 2: Séries statistiques à une variable 3 semaines

Soit *X* variable statistique discrète avec comme ensemble des valeurs $M = \{x_1, x_2, ..., x_k\}$

Effectif :n_i, le nombre d'apparitions associé à la valeur x_i de la variable X, L'effectif total $n = \sum_{i=1}^{k} n_i$

Fréquencef_i, associée à la valeur x_i de la variable X,,avec $f_i = (n_i/n)$ et $\sum_{i=1}^{k} f_i = 1$

Pourcentage p_i, associée à la valeur x_i de la variable X, le produit $p_i = f_i \times 100\%$

Effectif Cumulé Croissant ECC : N_i, associé à x_i, de la variable X, le nombre $N_i = \sum_{j=1}^{i} n_j$.

Effectif cumulé décroissant ECD associé à la valeur x_i de la variable X, le nombre $\left(n - \sum_{j=1}^{i-1} n_j\right)$

Fréquence cumulée croissante Fi, associée à la valeur x_i de la variable X, le nombre $F_i = \sum_{j=1}^{i} f_j$

Fréquence cumulée décroissante, associée à la valeur x_i de la variable X, le nombre $\left(1 - \sum_{j=1}^{i-1} f_j\right)$

Remarque : Une série statistique n'est autre que l'ensemble des couples $(x_i, n_i)_{\ i = 1,...,k}$

Soit X Variables statistique continues on définit des classes ou intervalles $[e_i, e_{i+1}[$de valeurs possibles de x_i .

Amplitude a_i de la classe : C'est la longueur de l'intervalle $[e_i, e_{i+1}[$ tel que la différence $a_i = e_{i+1} - e_i$.

Centre $c_i$de la classe : C'est le milieu de l'intervalle $[e_i, e_{i+1}[$, c'est la quantité $c_i = (e_{i+1} + e_i)/2$.

Effectif et fréquence de la classe x_i, le nombre n_i de valeurs prises dans $[e_i, e_{i+1}[$. La fréquence est $f_i = (n_i/n)$

Effectif cumulé en x_i, le nombre N_i de valeurs prises dans l'intervalle $]-\infty, e_{i+1}]$ fréquence cumulée $F_i = \sum_{j=1}^{i+1} f_j$

Représentations graphiques

Diagrammes différentiels : Basés sur les n_i, il visualise la différence d'effectifs (fréquences) entre les modalités.

Diagrammes cumulatifs : Basés sur les effectifs cumulés ou les fréquences cumulées. visualiser l'évolution des effectifs (ou des fréquences) cumulés.

Diagrammes différentiels (graphiques de distribution توزيع): Proportionnelle à l'effectif n_i (ou fréquence f_i)

Données qualitatives **(**diagramme à bande –barre-, ou circulaire - camembert-$\alpha_i = (n_i \times 360^{\circ})/n$

Données quantitatives : Les valeurs discrètes x_i (ou classe) prises par la variable sont placées sur les abscisses

Variable discrète (Diagramme en bâtons : La hauteur du bâton est proportionnelle à l'effectif (ou la fréquence).

Le polygone : A partir du diagramme en bâtons en joignant par un segment les sommets des bâtons.

Variable continue (Histogramme) : Rectangle dont la surface est proportionnelle à l'effectif (ou fréquence) des classes. L'ordonnée c'est la densité d_i d'une classe calculer par la formule suivante $d_i = n_i/a_i$.

Le Polygone des effectifs **(**ou des fréquences**) :** En joignant, par des segments les milieux des côtés supérieurs.

Diagrammes cumulatifs (ou graphiques de répartition تقسيم) (fonction de distribution cumulative)

Le diagramme cumulatif d'une variable statistique est appelé polygone des effectifs (ou des fréquences) cumulés.

Variable discrète : Diagramme en escalier un saut égal à l'effectif (ou fréquence) cumulé croissant ou décroissant.

Variable continue **:** *(cumulés croissants)* abscisses la limite supérieure e_{i+1}des classes et pour ordonnées les effectifs cumulés croissants N_i. Il donne le nombre d'observations inférieures à une valeur quelconque de la série.

(Cumulés décroissants) : abscisses la limite inférieure $e_i$des classes et pour ordonnées les effectifs cumulés décroissants. Ce diagramme donne le nombre d'observations supérieures à une valeur quelconque de la série.

INDICATEURS STATISTIQUES

Caractéristiques de tendance centrale et de position :

La moyenne arithmétique pondérée

Variable discrète : $\bar{X} = \frac{1}{n}\sum_{i=1}^{k}(n_i \times x_i) = \sum_{i=1}^{k}(f_i \times x_i)$,

Variable continue $\bar{X} = \frac{1}{n}\sum_{i=1}^{k}(n_i \times c_i) = \sum_{i=1}^{k}(f_i \times c_i)$

Le mode ou la dominante : Le mode M_o est la valeur de la variable qui correspond au plus grand effectif

Cas Unimodale La série a un seul mode, *Cas plurimodale* Elle a plusieurs modes. *Cas pas de mode :* pas de mode

Variable continue (classe modale) : la classe modale est la classe correspondant à la densité la plus élevée. Graphiquement. C'est l'abscisse d'un point d'ordonnée maximum d'un diagramme de dispersion.

La médiane M_e : La valeur qui sépare la moitié inférieure de la moitié supérieure d'un ensemble d'effectif n

Variable discrete (Si $n = 2p + 1$ est impair, M_e est la valeur de rang $p + 1$ lorsque l'on a ordonné les valeurs

Si $n = 2p$ est pair, la médiane $x_p \leq M_e \leq x_{p+1}$. Il est appelé intervalle médian et toute valeur de cet intervalle est une valeur médiane. On prend généralement $M_e = (x_{p+1} + x_p)/2$,

Variable continue Elle est la valeur qui correspond à l'effectif cumulé $n/2$.

$$M_e = e_i + (e_{i+1} - e_i)\frac{(n/2) - N_{i-1}}{N_i - N_{i-1}}$$

Les quantiles sont des caractéristiques de position qui partagent la série ordonnée en k parties égales

Quartiles $(k = 4)$: Il y a 3 quartiles Q_1, Q_2, Q_3 pour $n/4, 2n/4$ *et* $3n/4$ (fréquences cumulées $1/4, 2/4$ et $3/4$)

Déciles $(k = 10)$ **,** Il y a 9 déciles $D_1, D_2, \cdots, D_9$ pour $n/10, 2n/10, \cdots, 9n/10$.

Centiles $(k = 100)$. il y'a 99. Centiles, $C_1, C_2, \cdots, C_{99}$ pour $n/100, 2n/100, \cdots, 99n/100$.

Caractéristiques de dispersion

Ils mesurer les fluctuations (تقلب) des valeurs autour de la valeur centrale et d'apprécier l'étalement de la série.

La variance V d'une série $(x_i, n_i)_{i=1,\ldots,k}$ est la moyenne des carrés des écarts à la moyenne arithmétique $\bar{x}$.

Variable/Variance	Forme générale	Forme allégée (théorème de Koenig)
Variable discrète	$V = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^2$	$V = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \bar{x}^2$
Variable continue	$V = \frac{1}{n}\sum_{i=1}^{k} n_i(c_i - \bar{x})^2$	$V = \frac{1}{n}\sum_{i=1}^{k} n_i c_i^2 - \bar{x}^2$

L'écart-type : Caractérise la dispersion $\sigma = \sqrt{V}$, il a la même unité que les observations. Plus σ est petit, plus les données sont regroupées autour de la moyenne arithmétique $\bar{x}$ et plus la population est homogène.

Le coefficient de variation $CV = \sigma / \bar{x}$ sans dimension. il est exprimé en pourcentage.

Plus le coefficient de variation est élevé, plus la dispersion autour de la moyenne arithmétique est élevée.

L'étendue $e = x_{max} - x_{min}$ différence entre les valeurs extrêmes observées (plus petite et plus grande)

L'écart interquartile noté $EIQ = Q_3 - Q_1$ est la différence entre le troisième et le premier quartile : Il donne une idée sur l'étalement de la série autour de la médiane. Il permet d'écarter les valeurs extrêmes accidentelles.

L'écart inter-décile, noté $= D_9 - D_1$, est la différence entre le neuvième et le premier décile : Lui aussi donne une idée de la façon dont la série est étalée autour de la médiane.

Caractéristiques de forme

Ils donnent une idée sur la forme d'une distribution statistique par la mesurer de son asymétrie et aplatissement.

Les moments centrées d'ordre r

	Variable discrète	Variable continue
Le moment centré	$\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i(x_i - \bar{x})^r$	$\mu_r = \frac{1}{n}\sum_{i=1}^{k} n_i(c_i - \bar{x})^r$

Remarque : Le moment centré μ_2 n'est autre que la variance V.

L'asymétrie : Le coefficient d'asymétrie de Fisher, noté $\gamma_1 = \mu_3 / \sigma^3$ ainsi que la comparaison de la moyenne arithmétique, de la médiane et du mode permet de caractériser l'asymétrie d'une distribution. Tel que :

- Pour une distribution *étalée vers la gauche* : $\gamma_1 < 0 \Leftrightarrow Moyenne < Médiane < Mode$.
- Pour une distribution *symétrique* : $\gamma_1 = 0 \Leftrightarrow Moyenne = Médiane = Mode$.
- Pour une distribution *étalée vers la droite* : $\gamma_1 > 0 \Leftrightarrow Moyenne > Médiane > Mode$.

L'aplatissement : Les coefficients d'aplatissement mesurent l'aplatissement d'une distribution. Le coefficient d'aplatissement de Fisher, noté γ2, est défini par : $\gamma_2 = (\mu_4 / \sigma^4) - 3$.

- Si $\gamma_2 < 0$, la courbe est dite platicurtique. C'est-à-dire plus plate que la loi normale
- Si $\gamma_2 = 0$, la courbe est plus proche d'une distribution normale (loi de Gauss).
- Si $\gamma_2 > 0$, la courbe est dite leptocurtique. C'est-à-dire plus pointue que la loi normale

Chapitre 3 : Séries statistiques à deux variables 3 semaines

Quand on étudie deux caractères X et Y sur une population donnée, c'est en général parce qu'on cherche à savoir s'il existe un lien entre eux et quelle est l'intensité de ce lien.

Tableau de contingence, Nuage de points et modèle

Le tableau de contingence : Soit (X, Y) un couple de caractères défini sur une population d'effectif n. Notons $x_1, x_2, \cdots, x_p$ et $y_1, y_2, \cdots, y_q$ les valeurs distinctes observées pour X et Y et ordonnées dans l'ordre croissant. *Le tableau de contingence* (croisés) est un tableau qui croise les différentes modalités des deux caractères. Dans chaque case du tableau, on écrit l'effectif partiel n_{ij}du couple (x_i, y_j) de l'échantillon,

$X \setminus Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_q	l'effectif marginal de x_i $n_{i\cdot} = \sum_{i=1}^{q} n_{ij}$
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1q}	$n_{1\cdot}$
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2q}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{iq}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
x_p	n_{p1}	n_{p2}	$\cdots$	n_{pj}	$\cdots$	n_{pq}	$n_{p\cdot}$
l'effectif marginal de y_j $n_{\cdot j} = \sum_{i=1}^{p} n_{ij}$	$n_{\cdot 1}$	$n_{\cdot 2}$	$\cdots$	$n_{\cdot j}$	$\cdots$	$n_{\cdot q}$	l'effectif total $n = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij} = \sum_{i=1}^{p} n_{i\cdot} = \sum_{j=1}^{q} n_{\cdot j}$

Cas particulier (couple (x_i, y_j) isolées) : Les effectifs partiels n_{ij} prennent l'une des deux valeurs $\{0,1\}$

Remarque : Tous qui s'applique à l'effectif s'applique à la fréquence $f_{ij} = \frac{n_{ij}}{n}$ et $\sum_{i=1}^{p}\sum_{j=1}^{q} f_{ij} = 1$

Distributions marginales

les p couples $(x_i,\ n_{i\cdot})$ sont des distributions marginales des effectifs de X, (les modalités de X quel que soit Y).

les q couples $(y_j,\ n_{\cdot j})$ sont des distributions marginales des effectifs de Y, (les modalités de Y quel que soit X).

Remarque : Dans le cas continu on remplace les valeurs de x_i et y_j par les centre de classes $c_{x,i}$ et les $c_{y,j}$

Les moyennes marginales $\bar{x}$ et $\bar{y}$ sont les moyennes arithmétiques de X et de Y sont données par les formules :

	Variable discrète	Variable continue
Moyennes marginales	$\bar{x} = \frac{1}{n}\sum_{i=1}^{p} n_{i\cdot} \cdot x_i$ et $\bar{y} = \frac{1}{n}\sum_{j=1}^{q} n_{\cdot j} \cdot y_j$	$\bar{x} = \frac{1}{n}\sum_{i=1}^{p} n_{i\cdot} \cdot c_{x,i}$ et $\bar{y} = \frac{1}{n}\sum_{j=1}^{q} n_{\cdot j} \cdot c_{y,j}$

Les variances marginales $V(x)$ et de $V(y)$ se calculent à partir des distributions marginales selon les formules :

Les variances marginales Dans le cas continu on utilise $c_{x,i}$ et $c_{y,j}$	$V(x) = \frac{1}{n}\sum_{i=1}^{p} n_{i\cdot}(x_i - \bar{x})^2$	$V(y) = \frac{1}{n}\sum_{j=1}^{q} n_{\cdot j}(y_j - \bar{y})^2$

Distributions conditionnelles

La distribution de Y quand on a fixé la modalité x_i pour le caractère X est appelée ***distribution conditionnelle***

La distribution de X quand on a fixé la modalité y_j pour le caractère Y est appelée ***distribution conditionnelle***

La fréquence conditionnelle de la valeur $y_j$sachant x_i est représentée par la fraction $f_{j/i} = \frac{n_{ij}}{n_{i\cdot}} = \frac{f_{ij}}{f_{i\cdot}}$

<u>La fréquence conditionnelle</u> de la valeur $x_i$sachant y_j est représentée par la fraction $f_{i/j} = \frac{n_{ij}}{n_{\cdot j}} = \frac{f_{ij}}{f_{\cdot j}}$

La distribution $\left(y_j, \frac{n_{ij}}{n_{i\cdot}}\right) = \left(y_j, \frac{f_{ij}}{f_{i\cdot}}\right)$ est la *distribution conditionnelle des fréquences* de Y sachant $= x_i$.

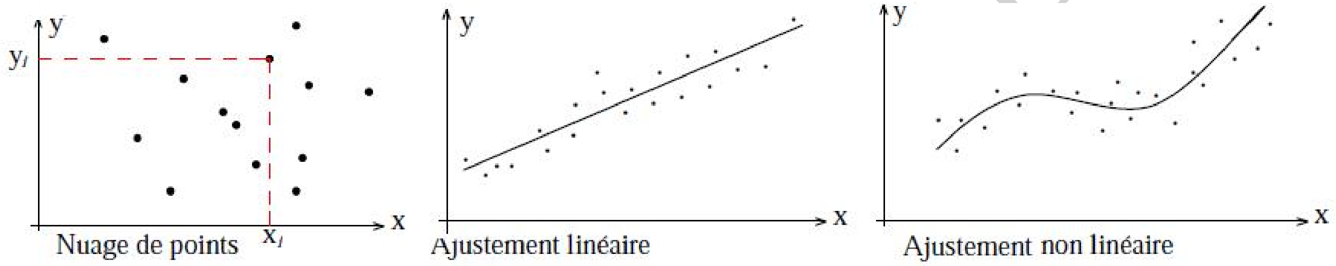
<u>Covariance du couple (X, Y) :</u>

Covariance du couple (X, Y) $\begin{cases} Cov(X,Y) > 0: x, y \text{ varient (même sens)} \\ Cov(X,Y) < 0: x, y \text{ varient (sens contraire)} \end{cases}$ Dans le cas continu on utilise $c_{x,i}$ et $c_{y,j}$	$Cov(X,Y) = \frac{1}{n}\sum_{i,j}^{p,q} n_{ij}(x_i - \bar{x})(y_j - \bar{y})$ $Ou \quad Cov(X,Y) = \overline{X \cdot Y} - \bar{X} \cdot \bar{Y} \; Avec \; \overline{X \cdot Y} = \frac{1}{n}\sum_{i,j}^{p,q} n_{ij} x_i y_j$

<u>Coefficient de corrélation linéaire</u> : Permet de donner le degré de ressemblance (liaison linéaire entre X et Y)

$$-1 \leq R(X,Y) = \frac{Cov(X,Y)}{\sigma(X) \cdot \sigma(Y)} \leq 1 \quad cas\ particuliers \begin{cases} R(X,Y) = 0 \Rightarrow \text{X et Y sont indépendants} \\ |R(X,Y)| = 1 \Rightarrow \text{X et Y sont confondues} \end{cases}$$

<u>Ajustement d'un nuage de points et construction du modèle :</u> Soit deux variables X et Y. L'observation (x_i, y_j) représente les coordonnées d'un point M_{ij}. L'ensemble des points ainsi obtenu est appelé nuage de points.



- les points sont dispersés n'importe où dans le plan : on dira alors que X et Y sont indépendants ;
- les points semblent disposés autour d'une certaine courbe. La recherche d'une telle courbe est ce que l'on appelle l'ajustement d'un nuage de points. Cela permet d'avoir une relation linéaire ou non-linéaire.

Modèle par la méthode de Mayer : Elle permet d'avoir l'équation de la droite $Y = a \cdot X + b$ qui représente le plus fidèlement le nuage. Consiste à partager le nuage en deux rangés dans l'ordre croissant des abscisses en deux sous-groupes SG_1 et SG_2 de même effectif (à une unité près). Ensuite déterminer l'équation de la droite d'ajustement (de Mayer) qui passe par les deux points des moyennes arithmétiques G_1 et G_2 des deux sous-groupes

La méthode des moindres carrés régression linéaire : elle permet d'avoir l'équation de la droite $y = ax + b$

$$a = \frac{cov(x,y)}{V(x)} \quad et \quad b = \bar{y} - a\bar{x}$$

Ajustement à une fonction logarithme : Elle permet d'avoir la relation non-linéaire $Y = a \cdot \ln(X) + b$

$$y = a \cdot ln(x) + b \quad Alors \quad a = \frac{cov(ln(x), y)}{V(ln(x))} \quad et \quad b = \bar{y} - a \cdot \overline{ln(x)}$$

Ajustement à une fonction exponentielle : elle permet d'avoir la relation non-linéaire $y = b \cdot e^{ax}$

$$y = be^{ax} \Rightarrow ln(y) = ax + ln(b) \; Alors \; a = \frac{cov(x, ln(y))}{V(x)} \; et \; ln(b) = \overline{ln(y)} - a\bar{x} \Rightarrow b = \exp(\overline{ln(y)} - a\bar{x})$$

Ajustement à une fonction puissance : Elle permet d'avoir la relation non-linéaire $y = b \cdot x^a \; Avec \; a > 0$

$$y = b \cdot x^a \Rightarrow ln(y) = a\,ln(x) + ln(b) \quad Alors \quad a = \frac{cov(ln(x), ln(y))}{V(ln(x))} \quad et \quad ln(b) = \overline{ln(y)} - a \cdot \overline{ln(x)}$$

- **Partie B : Probabilités Chapitre 1 : Analyse combinatoire 1 Semaine**

Analyse combinatoire : Principe fondamental de comptage

Si une expérience comporte k étapes offrant respectivement $n_1, n_2, \cdots, n_k$ résultats possibles alors le nombre total d'issues de cette expérience est :$n = n_1 \times n_2 \times \cdots \times n_k$

Diagrammes Arborescents : C'est un moyen pratique pour dénombrer les résultats (ou les issues) possibles d'une suite d'expériences ; elle consiste à utiliser un *diagramme arborescent* (pratique pour petit nombre de n).

Dénombrement : Soit un ensemble U de n éléments, on veut les mettre en partie ordonnée constituée de p éléments $(1 < p < n)$. Le nombre de parties possibles selon l'importance de l'ordre et de la répétition est :

	Arrangements avec répétition	Arrangements	Permutations	Combinaisons
Répétition	Oui	Non	Non	Non
Ordre	Oui	Oui	Oui	Non
Formule	$\alpha_n^p = n^p$	$A_n^p = \dfrac{n!}{(n-p)!}$	$P_n = n!$	$C_n^p = \dfrac{n!}{(n-p)! \times p!}$

Chapitre 2 : Introduction aux probabilités 2 semaines

Opérations sur les ensembles : Soit *A*, *B* et *C* trois parties d'un ensemble *U*

Réunion (ou Union) : La réunion *A* ∪ *B* est définie par : $x \in A \cup B \Leftrightarrow (x \in A \; ou \; x \in B)$.

Intersection. L'intersection *A* ∩ *B* est définie par : *x* ∈ *A* ∩ *B* ⇔ *(x* ∈ *A* et *x* ∈ *B).* *A* et *B* disjoints si *A* ∩ *B* = ∅.

Partition. Une famille $(A_i)_{i \in I}$ est une partition de Ω si $\underset{i \in I}{\cup} A_i = \Omega$ et $\forall (i,j) \in I^2, \left(i \neq j \Rightarrow A_i \cap A_j = \phi\right)$

Complémentaire. Le complémentaire $\bar{A}$ de *A* dans *U* est défini par : $x \in \bar{A} \Leftrightarrow (x \in U \; et \; x \notin A)$.

Différence. La différence *A* \ *B* est défini par : $x \in A \setminus B \Leftrightarrow (x \in A \; et \; x \notin B)$. alors l'égalité $A \setminus B = A \cap \bar{B}$.

Différence symétrique. Notée $A\Delta B$ l'ensemble défini par : $x \in A\Delta B \Leftrightarrow (x \in (A \; ou \; B) \; et \; x \notin (A \; et \; B))$

Par conséquent, nous avons l'égalité $A\Delta B = (A \cup B) \setminus (A \cap B)$

Produit cartésien. Le produit cartésien *E* × *F* d est défini par :$E \times F = \{(x,y) / x \in E \; et \; y \in F\}$

Ensemble des parties d'un ensemble *U*. Noté $\mathcal{P}(U)$, C'est l'ensemble de tous les sous-ensembles de U. Avec $A \in \mathcal{P}(U) \Leftrightarrow A \subseteq U$ Si U contient n éléments, $\mathcal{P}(U)$ contient 2^n éléments. Avec l'ensemble vide $\emptyset \in \mathcal{P}(U)$

Règles de calcul entre les ensembles : Le tableau suivant donne les opérations les plus utilisées

$C \cap (A \cup B) = (C \cap A) \cup (C \cap B)$	$\overline{(A \cup B)} = \bar{A} \cap \bar{B}$	$\emptyset = A \cap \bar{A}$	$\bar{\bar{A}} = A$
$C \cup (A \cap B) = (C \cup A) \cap (C \cup B)$	$\overline{(A \cap B)} = \bar{A} \cup \bar{B}$	$U = A \cup \bar{A}$	

Le calcul des probabilités Il permet de fournir un modèle mathématique pour décrire un phénomène aléatoire.

Expérience aléatoire ε Expérience dans les mêmes conditions donne le résultat est imprévisible et dû au hasard.

Univers noté U, l'ensemble de tous les résultats possibles d'une expérience aléatoire ε. u tout élément de U.

Un Événement : lié à l'expérience ε est un sous-ensemble $A \subset \mathcal{P}(U)$ des résultats pour lesquels il est réalisé.

Ecriture logique	Ecriture ensembliste
Le contraire de A s'est réalisé	$\bar{A}$
A et B se sont réalisés	$A \cap B$
A ou B s'est réalisé	$A \cup B$
A s'est réalisé mais pas B	$A \setminus B$
Impossible que *A* et *B* se réalisent simultanément (événements incompatibles ou disjoints)	$A \cap B = \emptyset$

Terminologie des événements
∅ : événement impossible
U : événement certain
Tout singleton $\{u\}$, où *u* ∈*U*, est appelé événement élémentaire

Espace probabilisable lié à ε est(U, A), U : Ensemble des résultats possibles et A :Ensemble des événements

Une Probabilité **:** sur (U, A) une application p de A dans $[0,1]$ vérifiant $p(U) = 1$ et pour tout ensemble dénombrable A_i d'événements deux à deux incompatibles $p\left(\underset{i\in I}{\cup} A_i\right) = \sum_{i\in I} p(A_i)$. Avec $0 \le p(A_i) \le 1$

Il y a équiprobabilité lorsque les probabilités de tous les événements élémentaires sont égales. Par conséquence la probabilité uniforme pour tout événement A, est : $p(A) = \frac{Card(A)}{Card(U)} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}$

Quelques propriétés de la probabilité : Pour tous événements A, B et C de $\mathcal{P}(U)$, nous avons

$p(\bar{A}) = 1 - p(A)$ d'où $p(\emptyset) = 0$	$p(A \cup B) = p(B) + p(B) - p(A \cap B)$
$p(A \backslash B) = p(A) - p(A \cap B)$	$si\ A\ et\ B\ incompatible \Rightarrow\ p(A \cup B) = p(A) + p(B)$
$si\ A \subset B \Rightarrow p(A) \le p(B)$	$si\ les\ A_i\ incompatible\ deux\ \grave{a}\ deux \Rightarrow p\left(\underset{i\in I}{\cup} A_i\right) = \sum_{i\in I} p(A_i)$
$p(A) = p(A \cap B) + p(A \cap \bar{B})$	

Probabilité conditionnelle : Sur un espace probabilisé, B un événement tel $p(B) > 0$. Pour tout évènement A, la probabilité que A soit réalisé sachant que B l'est déjà (la probabilité conditionnelle de A étant donné B):

$$p(A/B) = \frac{p(A \cap B)}{p(B)} = \frac{Card(A \cap B)/Card(U)}{Card(B)/Card(U)} = \frac{Card(A \cap B)}{Card(B)} = \frac{\text{nombre de façons de réaliser A et B}}{\text{nombre de façons de réaliser B}}$$

$p(U/C) = 1$	$p(\emptyset/C) = 0$
$p(\bar{A}/C) = 1 - p(A/C)$	$si\ A \subseteq B \Rightarrow p(A/C) \le p(B/C)$
$p(A/C) = p(A \cap B/C) + p(A \cap \bar{B}/C)$	$p(A \cup B/C) = p(A/C) + p(B/C) - p(A \cap B/C)$

Indépendance : Deux événements A et B sont indépendants alors $p(A \cap B) = p(A) \cdot p(B)$; cela signifie que la réalisation de A n'intervient pas dans celle de B. Avec $p(A) = p(A/B) \Leftrightarrow p(B) = p(B/A)$.

Si A et B sont indépendants il est de même pour $A\ et\ \bar{B}, \bar{A}\ et\ B, \bar{A}\ et\ \bar{B}$

n événements $A_1, A_2, \dots, A_i, \dots, A_n$ sont *mutuellement indépendants* si $p\left(\bigcap_{j=1}^{n} A_j\right) = \prod_{j=1}^{n} p(A_j)$

Ensemble complet d'événements et Formule de Bayes (appelé aussi probabilité des causes): Soit un événement B qui peut dépendre de n causes $A_1, A_2, \dots, A_i, \dots, A_n$ formant un ensemble complet d'événements défini par : $\begin{pmatrix} \forall(i,j)\ si\ i \ne j\ alors\ A_i \cap\ A_j = \emptyset \\ et\ B\ = \bigcup_{i=1}^{n}(B \cap A_i) \end{pmatrix}$. Etant donnée la réalisation de l'événement B, on cherche la probabilité pour que l'événement A_i en soit la cause. La formule de Bayes suivante permet de donner cette probabilité :

$$p(A_i/B) = \frac{p(A_i) \cdot p(B/A_i)}{\sum_{j=1}^{n} p(A_j) \cdot p(B/A_j)}$$

Variables aléatoire 1 Semaine

Variable aléatoire : C'est une variable X dont la valeur x est déterminée après une expérience aléatoire ε

La fonction de répartition (loi de probabilité) : notée $F(x) = p(X \le x) \in [0,1]$, c-à-dire la probabilité pour que X prenne une valeur inférieure à x. En d'autre terme le cumul des probabilités individuelles de la série (x_i, p_i).

Propriétés : $\lim_{x\to-\infty} F(x) = 0$ | $\lim_{x\to+\infty} F(x) = 1$ | $p\left(x_i \begin{Bmatrix} < \\ \le \end{Bmatrix} X_i \begin{Bmatrix} < \\ \le \end{Bmatrix} x_j\right) = F(x_j) - F(x_i)$ | $p(X \le x) = 1 - p(X > x)$ |

Variable discrète : calcul de Espérance mathématique $E(X)$, la Variance $V(X)$ et l'écart type $\sigma(X)$

$$E(X) = m = \sum_{i=1}^{+\infty} x_i\, p(x_i) \qquad V(X) = \sum_{i=1}^{+\infty} (x_i - m)^2\, p(x_i) = E(X^2) - E(X)^2 \qquad \sigma(X) = \sqrt{V(X)}$$

Variable continue : $F(x) = p(X \leq x)$ et on parle de fonction de densité (distribution) de probabilité $f(x)$ tel que:

$$F(x) = \int_{-\infty}^{x} f(t)dt \quad et \quad \int_{-\infty}^{+\infty} f(t)dt = 1 \quad E(X) = m = \int_{-\infty}^{+\infty} xf(x)dx \quad et \quad V(X) = \int_{-\infty}^{+\infty} x^2 \cdot f(x) \cdot dx - m^2$$

Lois de probabilité discrètes usuelles (finie et infinie) **1 Semaine**

Loi uniforme discrète (finie) :

Soit U un univers comportant $n \in N^*$ issues équiprobables. Une variable aléatoire X suit une loi uniforme discrète si :

$$\forall x_i \ p(X = x_i) = p(x_i) = p_i = \frac{1}{n} \ et \ i = 1,2,\cdots,n \Rightarrow \begin{cases} E(X) = m = \sum_{i=1}^{n} x_i p(x_i) = \frac{1}{n}\sum_{i=1}^{n} x_i \\ V(X) = E(X^2) - E(X)^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2 \end{cases}$$

Loi Bernoulli (finie) :

Une variable aléatoire X suit une loi de Bernoulli de paramètre p, notée $\mathcal{B}(1,p)$, si $X = 1$ en cas de succès avec $p(X = 1) = p$ et $X = 0$ en cas d'échec avec $p(X = 0) = q = 1 - p$

$$E(X) = m = \sum_{i=1}^{n} x_i p(x_i) = 1 \times p + 0 \times q = p \qquad V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p) = pq$$

Loi binomiale de paramètres n et p (finie) :

La loi de probabilité est considérée aléatoire binomiale notée $B(n,p)$ si elle satisfait les critères suivants :

1. L'expérience est constituée d'une suite de $n \in N^*$ épreuves indépendantes.
2. Deux résultats possibles sont associés à chaque épreuve : « succès » ou « échec » (Loi de Bernouli).
   A chaque épreuve, la probabilité d'un succès, notée $p \in [0,1]$, la probabilité d'un échec, notée $q = 1 - p$.
3. La variable aléatoire X correspond au nombre de succès en n épreuves. $\Rightarrow X = k \leq n$ le nombre de succès.

$$p(X = k) = C_n^k \, p^k (1-p)^{n-k} \quad Avec \ \ 0 \leq k \leq n \Rightarrow \begin{cases} E(X) = \sum_{k=0}^{n} k C_n^k \, p^k (1-p)^{n-k} = n\,p \\ V(X) = \sum_{k=0}^{n} (k - np)^2 p(k) = np(1-p) = npq \end{cases}$$

Loi de Poisson de paramètre λ (Lois discrètes usuelles infinies) : Une variable aléatoire X suit une loi de Poisson de paramètre λ, notée $P(\lambda)$, si la variable aléatoire X prend la valeur $k \in N$ avec la probabilité égale à

$$f(x) = p(X = x = k) = \frac{\lambda^k e^{(-\lambda)}}{k!} \qquad x = k = 0,1,2,\cdots \Rightarrow \begin{cases} E(X) = \sum_{k=X=0}^{+\infty} k\frac{\lambda^k}{k!} e^{(-\lambda)} = \lambda \\ V(X) = \sum_{k=X=0}^{+\infty} k^2 \frac{\lambda^k}{k!} e^{(-\lambda)} - \lambda^2 = \lambda \end{cases}$$

Il est possible d'approximer la loi binomiale par la loi de Poisson si $n > 20$, $p \leq 0.1$ et $np < 5$.

Loi à densité uniforme : Une variable aléatoire X continue suit une loi uniforme continue sur l'intervalle$[a, b]$, notée $U[a, b]$ Si X admet pour densité de probabilité la fonction $f(x)$ suivante :

$$f(x) = \begin{cases} \frac{1}{b-a} & si \quad X \in [a,b] \\ 0 & ailleurs \end{cases} \qquad F(x) = \begin{cases} 0 & si \quad x < a \\ \frac{x-a}{b-a} & si \quad x \in [a,b] \\ 1 & si \quad x > b \end{cases} \Rightarrow \begin{cases} E(X) = m = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b x \frac{1}{b-a} dx = \frac{b+a}{2} \\ V(X) = E(X^2) - E(X)^2 = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{b+a}{2}\right)^2 = \frac{(a-b)^2}{12} \end{cases}$$

Loi exponentielle : X variable aléatoire à valeurs dans $[0, +\infty[$ suit une loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si X est une variable continue et admet pour densité de probabilité la fonction $f(x)$ suivante :

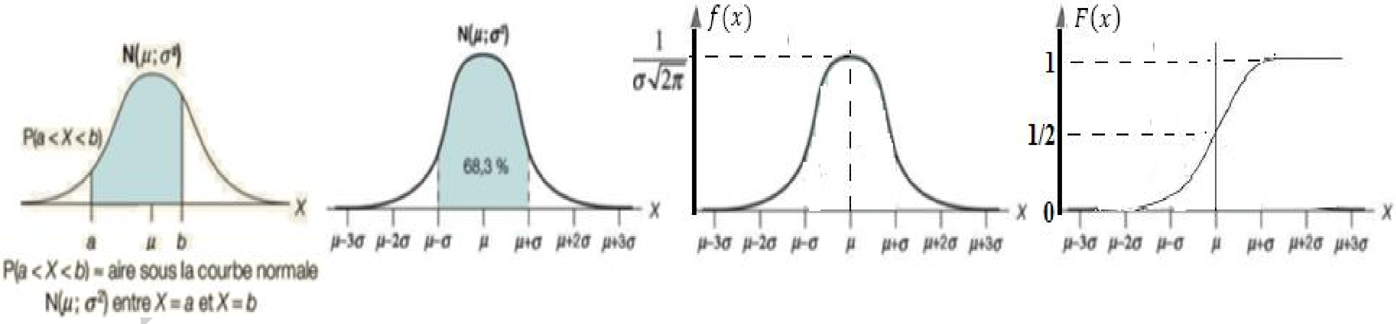
$$f(x) = \begin{cases} \lambda e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases} \qquad F(x) = \begin{cases} 1 - e^{-\lambda x} & si \quad x \geq 0 \\ 0 & si \quad x < 0 \end{cases} \Rightarrow \begin{cases} E(X) = m = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b x\lambda e^{-\lambda x} dx = \frac{1}{\lambda} \\ V(X) = E(X^2) - E(X)^2 = \int_a^b x^2 \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2} \end{cases}$$

Loi normale : La variable aléatoire X continue suit une loi normale de paramètres $\mu \in R$ et $\sigma > 0$ (ou *loi de Gauss ou Gaussienne)*, notée $\mathcal{N}(\mu, \sigma)$ si elle admet pour densité de probabilité la fonction $f(x)$ suivante :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2}\left(\frac{(x-\mu)}{\sigma}\right)^2\right) \quad pour\ tout\ x \in R \quad avec \quad \int_{-\infty}^{+\infty} f(x)dx = 1$$

$$E(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \mu \quad et \quad V(X) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx - \mu^2 = \sigma^2$$

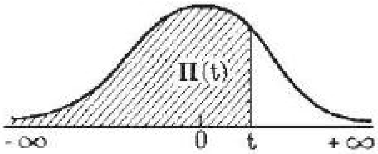
L'histogramme qui représente une distribution d'une variable aléatoire continue qui suit une loi normale a une forme d'une cloche dont le nom est courbe normale, ou courbe de Gauss. elle dépend des paramètres μ et σ.



Les observations : $68.3\% \in [\mu - \sigma, \mu + \sigma]$, $95\% \in [\mu - 2\sigma, \mu + 2\sigma]$ et $99.7\% \in [\mu - 3\sigma, \mu + 3\sigma]$

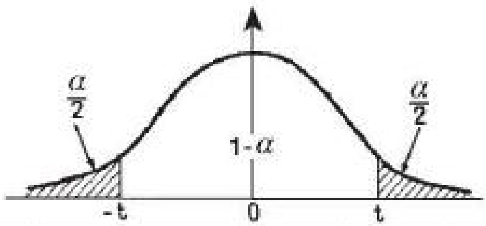
Avec le changement de variable $Z = (x - \mu)/\sigma$, nous avons la fonction de répartition $F(x)$: $\mathbb{R} \rightarrow \mathbb{R}^+$ pour une loi normale centrée réduite $\mathcal{N}(0,1)$ donnée par la formule suivante :

$$F(x) = \Phi(x) = p(X < x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$



$$\pi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2}\, dx \ .$$

t	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	0,7290	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	0,7704	0,7734	0,7764	0,7794	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
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,243	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013