Mathématiques propédeutiques
Statistique

J.P. Gabriel et C. Mazza

Table des matières

1 Introduction

La statistique peut être définie comme la science du dépouillement de données issues de l'observation de phénomènes naturels. Par exemple, si un fabricant de médicaments désire créer un nouveau médicament destiné à traiter la migraine, il va effectuer un sondage dans la population pour estimer la proportion $0 < p < 1$ des personnes de la population qui souffrent de ce trouble afin d'obtenir une idée sur le nombre de potentiels acheteurs. Pour ce faire, un institut de sondage va choisir au hasard un nombre n de personnes dans la population totale, et calculer la proportion $\hat{p} = m/n$ de ces personnes qui souffrent de migraine. Cette proportion $\hat{p}$ fournit une estimation de la vraie proportion (inconue) p. Pourquoi prendre un échantillon ? Tout simplement parce qu'il est impossible pratiquement de poser la question à tous les membres d'une population. Quelle est la qualité de cette estimation ? On verra dans ce cours comment il est possible de quantifier la marge d'erreur commise par une telle estimation. Sans trop entrer dans les détails, nous verrons que si l'échantillon aléatoire est assez grand, nous pouvons affirmer avec 95 % de confiance que

$$p = \hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Par exemple, si la taille de l'échantillon vaut $n = 1000$, et si $m = 210$ personnes de cet échantillon souffrent de migraine, on trouve que

$$p = 0.21 \pm 1.96\sqrt{\frac{0.21(1-0.21)}{1000}} = 0.21 \pm 0.025.$$

On peut estimer avec une confiance au niveau de 95% que la proportion de personnes souffrant de migraines dans la population est comprise entre 0.185 et 0.235. L'intervalle ainsi obtenu $[0.185, 0.235]$ est un **intervalle de confiance** pour la proportion inconnue p de niveau de confiance 95 %.

Dans de nombreuses situations pratiques, le traitement statistique des données est précédé d'une phase exploratoire lors de laquelle le scientifique examinera les donnée afin d'en extraire le plus d'information possible, et de déduire de ceci diverses conjectures.

Nous allons utiliser très souvent la notion de **variable aléatoire**, qui associe une valeur numérique au résultat d'une expérience, qui permet ainsi de créer une fonction des valeurs expérimentales. Cette fonction est déterminée par le résultat de l'expérience et est génériquement notée X.

Si l'on revient à l'exemple du sondage, on peut poser par exemple que $X = 1$ si la personne souffre de migraines et poser $X = 0$ sinon. La question étant posée à toutes les personnes de l'échantillon, on doit considérer une suite de variables aléatoires X_i, $i = 1, \cdots, n$ correspondant aux réponses de tous les membres de l'échantillon. Cette formalisation mathématique du sondage permet de décrire mathématiquement diverses quantités naturelles, comme la proportion estimée $\hat{p}$ qui devient

$$\hat{p} = \frac{m}{n} = \frac{\sum_{i=1}^{n} X_i}{n},$$

où on utilise le fait que $m = \sum_{i=1}^{n} X_i$.

Dans l'exemple précédent, le résultat de l'exprience est binaire ; dans de nombreuses situations, le réstultat d'une expérience X est un nombre réel, i.e. $X \in \mathbb{R}$. Un exemple simple consiste mesurer la taille d'une personne en cm.

Dans le premier cas, on parle de donnée **discrète, et dans le deuxième cas, de donnée continue**.

2 Analyse exploratoire

Les données sont des informations quantitatives ou qualitatives.

Ex : $\{1, 0, 0, 1, 1, 1, 1, 0, 1, 0\}$ sont des données de jets d'une pièce de monnaie.

Modélisées comme réalisations d'une v.a., les données peuvent être :

- Discrètes : catégorielles (ex : H/F, P/F) ou ordinales (ex : dé)
- Continues (ex : poids)
- **Univariées** quand on ne mesure qu'un phénomène à la fois.
- Multivariées quand on mesure plusieurs phénomènes *conjointement*.

	Red	Green	Blue	Orange	Yellow	Brown	Weight
1	15	9	3	NA	9	19	49.79
2	9	17	190	3	3	8	48.98
⋮							

Pour être utiles, les données doivent être :

- vérifiées : données manquantes, aberrantes ?
- **analysées** :
  - résumées avec des chiffres
  - visualisées graphiquement
  - disséquées pour en comprendre la structure et proposer des modèles.
- modélisées : trouver un modèle probabiliste le plus simple possible qui est le plus en adéquation avec la réalité et proche des données.

2.1 Données univariées discrètes

Un casino embauche un statisticien pour trouver de potentiels fraudeurs.

Un jeu consiste à lancer une pièce de monnaie 2 fois et à parier sur le nombre T de Piles.

Des données sont collectées avec :
- Une pièce du casino. Un employé est embauché et récolte $N_1 = 1000$ données (2h de travail) : $t_1 = 0, t_2 = 1, t_3 = 2, t_4 = 0, \ldots$
- Une pièce aamenée par un joueur. Observé plus rarement on a $N_2 = 392$ données lors d'une semaine de jeu en 2006 : $t_1 = 0, t_2 = 2, \ldots$

On compte les nombres de fois n_0, n_1, n_2 où $T = 0, 1, 2$.

Comment feriez-vous pour savoir si la pièce du joueur ressemble à celui du casino ou s'il est truqué ?

La table des fréquences des données est :

CASINO

T	0	1	2
n_i	231	517	252
$\hat{p}_i = n_i/N_1$	0.231	0.517	0.252

JOUEUR

T	0	1	2
n_i	209	155	28
$\hat{p}_i = n_i/N_2$	0.533	0.395	0.071

La pièce du casino est-elle équilibrés ?

Modélisation probabiliste : soit les **variables aléatoires** :

– $X_1 \in \{\text{Pile}, \text{Face}\}$ et $X_2 \in \{\text{Pile}, \text{Face}\}$ pour les résultats au premier lancé et au deuxième lancé.

– $T =$ nombre de Pile dans $\{X_1, X_2\}$

On dénote par p la probabilité de Pile :

– Quelles sont les valeurs possibles de $\{X_1, X_2\}$?

– Quelles sont les valeurs possibles de T ?

– Quelles sont les probabilités de réalisation de ces valeurs ?

– Quelles sont les probabilités de réalisation de ces valeurs si la pièce est équilibrée ?
Le même joueur et le même dé sont observés lors d'un tournoi en 2007, ce qui amène le statisticien à mesurer $N_3 = 114$ lancés.

JOUEUR 2006

T	0	1	2
n_i	209	155	28
Probabilités estimées $\hat{p}_i$	0.533	0.395	0.071

JOUEUR 2007

T	0	1	2
n_i	46	56	12
Probablités estimées $\hat{p}_i$	0.404	0.491	0.105

DE EQUILIBRE

T	0	1	2
Espérance $\mathrm{E}(n_i)$	$N\frac{1}{4}$	$N\frac{1}{2}$	$N\frac{1}{4}$
Probabilités p_i si $p = \frac{1}{2}$	0.25	0.50	0.25

Le rôle des probabilités et des statistiques est de :

– Faire une analyse exploratoire des données pour proposer un modèle probabiliste

– Estimer ce modèle à partir de données

– Vérifier que le modèle colle bien aux données ; sinon, proposer un autre modèle

– Faire de l'inférence, par exemple tester si, pour la pièce du joueur, la probabilité d'un Pile est bien $p = 0.5$.

Couleur de M&M's

Le nombre X de M&M's Rouge est mesuré dans $n = 30$ paquets :

15 9 14 15 10 12 6 14 4 9 9 8 12 9 6
4 3 14 5 8 8 9 20 12 8 4 10 5 15 11

Pour les Verts on mesure :

9 17 8 7 3 7 7 11 2 9 11 8 9 7 6
6 5 5 5 9 7 8 2 6 9 6 12 4 11 6

Voyez-vous une différence entre Rouge et Vert ?

L'ensemble fondamental est $\Omega = \{0, 1, 2, \ldots\}$.

Table des fréquences pour les Rouge :

X	0,1,2	3	4	5	6	7	8	9	...
n_i	0	1	3	2	2	0	4	5	
$\hat{p}_i$	0	0.03	0.10	0.07	0.07	0	0.13	0.17	
$\sum_{j\leq i}\hat{p}_j$	0	0.03	0.13	0.20	0.27	0.27	0.40	0.57	

X	10	11	12	13	14	15	16,17,18,19	20
n_i	2	1	3	0	3	3	0	1
$\hat{p}_i$	0.07	0.03	0.10	0	0.10	0.10	0	0.03
$\sum_{j\leq i}\hat{p}_j$	0.63	0.67	0.77	0.77	0.87	0.97	0.97	1.00

où :
– n_i sont les comptages/fréquences
– $\hat{p}_i = n_i/n$ sont les fréquences relatives/probabilités estimées
– $\sum_{j\leq i}\hat{p}_j$ sont les fréquences relatives cumulées.

```
> summary(as.factor(Red))
 3  4  5  6  8  9 10 11 12 14 15 20
 1  3  2  2  4  5  2  1  3  3  3  1
```

```
> round(summary(as.factor(Red))/30,2)
```

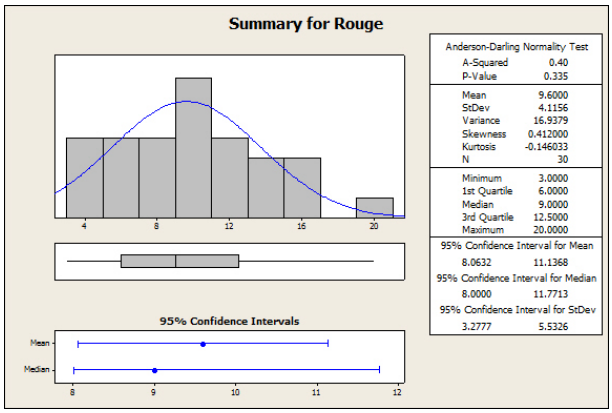


FIGURE 1 – Les nombres de M&M's trouvés dans les 30 paquets sont résumés graphiquement

```
   3    4    5    6    8    9   10   11   12   14   15   20
0.03 0.10 0.07 0.07 0.13 0.17 0.07 0.03 0.10 0.10 0.10 0.03
```

```
> round(cumsum(summary(as.factor(Red))/30),2)
   3    4    5    6    8    9   10   11   12   14   15   20
0.03 0.13 0.20 0.27 0.40 0.57 0.63 0.67 0.77 0.87 0.97 1.00
```

Les logiciels statistiques permettent de résumer les données en utilisant divers types de méthodes de statistique exploratoire. La figure **??** nous donne un résumé du nombre de M&M's rouges trouvés dans 30 paquets.

Le **mode** est 9 : valeur la plus fréquente.

2.2 Données univariées continues

Certaines mesures ne sont par discrètes ou dénombrables.

Exemple 2.1 *Le poids de chaque paquet de M&M's est une variable aléatoire continue. Données arrondies au centième :*

49.79 48.98 50.40 49.16 47.61 49.80 50.23 51.68 48.45 46.22 50.43 49.80 46.94 47.98 48.49 48.33 48.72 49.69 48.95 51.71 51.53 50.97 50.01 48.28 48.74 46.72 47.67 47.70 49.40 52.06

2.2.1 L'histogramme

L'histogramme est l'équivalent du diagramme à bâtons pour les variables/données continues :

– Diagramme à bâtons = estimateur des probabilités d'une variable aléatoire discrète

– Histogramme = estimateur d'une fonction de densité d'une variable aléatoire continue.
Basé sur une partition subjective de l'ensemble fondamental

$$\Omega = (0, \infty) = \bigcup_i (b_i, b_{i+1}],$$

l'histogramme est le graphique des **densités** dans chaque intervalle de la partition

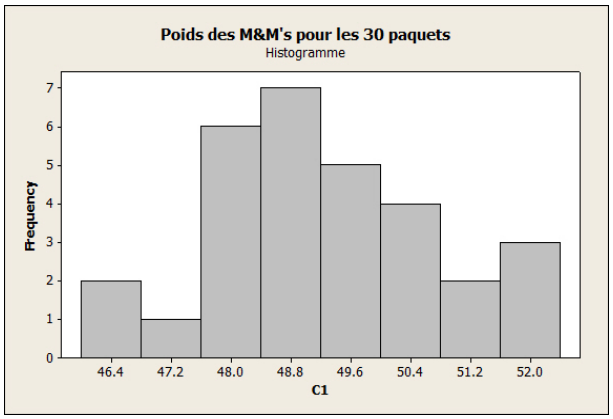


FIGURE 2 – Histogramme des poids des paquets de MM's

	(b_1, b_2)	(b_2, b_3)	(b_3, b_4)	(b_4, b_5)	(b_5, b_6)	(b_6, b_7)
	(0,46)	[46,48)	[48,50)	[50,52)	[52,54)	[54 ,∞)
n_i	0	7	14	8	1	0
$\hat{f}_i$	0	0.12	0.23	0.13	0.02	0

où $\hat{f}_i = \frac{n_i}{n(b_{i+1}-b_i)}$ est la densité estimée.

2.2.2 Statistiques de centralité

Définition : une statistique est une fonction des données $x_1, \ldots, x_n$.

La moyenne : (données discrètes ordinales et continues)

$$\bar{x} = \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Le mode : (données discrètes) réalisation/donnée la plus fréquente (pas forcément unique).

Le mode : (données continues) valeur où la densité a un maximum local (pas forcément unique).

Définition : les statistiques d'ordre $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ sont les données ordonnées, c'est-à-dire

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}.$$

La médiane : (données discrètes ordinales et continues) Valeur telle que 50% des données sont plus petites (et donc que 50% des données sont plus grandes).

$$x_{.5} = \begin{cases} x_{(\frac{n+1}{2})} & \text{si } n \text{ est impaire,} \\ \frac{1}{2}(x_{(n/2)} + x_{(1+n/2)}) & \text{si } n \text{ est paire.} \end{cases}$$

Propriété de la médiane : elle est robuste. La robustesse est la propriété d'une statistique à ne pas être influencée de façon trop forte par une 'mauvaise' donnée. C'est à la fois un avantage et un inconvénient.

	Mode	Moyenne	Médiane
Red	9	9.6	9
Green	{6,7,9}	7.4	7
Blue	7	7.2	6.5
Orange	6	6.6	6
Yellow	7	13.8	13.5
Brown	8	12.5	12.5

2.2.3 Statistiques de dispersion

Etendue : $x_{(n)} - x_{(1)}$, la différence entre valeurs maximum et minimum.

Définition 2.2 *Les deux quartiles inférieur* $\hat{q}(25\%) = x_{.25}$ *et supérieur* $\hat{q}(75\%) = x_{.75}$ *sont les statistiques telles qu'environ* 25% *des données sont plus petites que* $\hat{q}(25\%)$ *et* 25% *des données sont plus grandes que* $\hat{q}(75\%)$.

Soit $m = \lfloor (n+1)/2 \rfloor$ (partie entière inférieure). On trouve les deux quartiles en comptant $(m+1)/2$ valeurs des deux extrêmes des statistiques d'ordre :

$$\hat{q}(25\%) = x_{(\frac{m+1}{2})} \quad \text{et} \quad \hat{q}(75\%) = x_{(n+1-\frac{m+1}{2})}.$$

Note : si $m+1$ est impaire, alors prendre la moyenne des deux quantiles gauche et droite. Ex. : $n = 15$, alors $m = 8$ donc $x_{.25} = (x_{(4)} + x_{(5)})/2$.

Etendue interquartile : EIQ $= x_{.75} - x_{.25}$, différence entre quartiles supérieurs et inférieurs. L'intervalle interquartile $[x_{.25}, x_{.75}]$ contient 50% des données.

Définition 2.3 (Variance empirique) *La variance empirique est* $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Ecart-type empirique *:* $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

Ecart absolu médian *:* mad $=$ median$(|x - x_{.5}|)$.

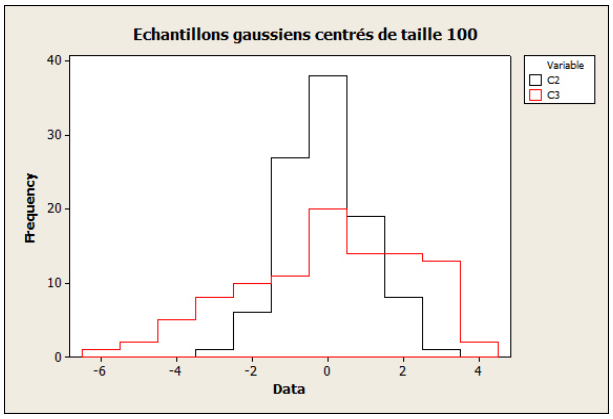


FIGURE 3 – Histogrammes associés à deux échantillons de taille 100 d'erreurs normales de précisions $\sigma = 1$ et $\sigma = 2$.

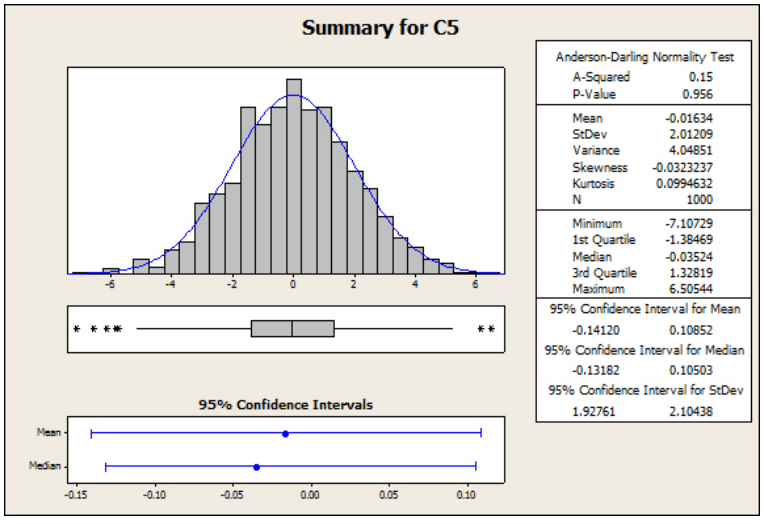


FIGURE 4 – Résumé statistique d'un échantillon gaussien centré $\varepsilon_1, \cdots, \varepsilon_{1000}$ de taille $n = 1000$ et d'écart type $\sigma = 2$. Dans ce résumé, l'écart type empirique s (=StDev) vaut 2.01209 est proche de la vraie valeur $\sigma = 2$.

2.2.4 La loi normale

On peut décrire statistiquement la mesure X d'une valeur μ, comme par exemple la longueur d'une barre ou le poids d'un individu, en posant

$$X = \mu + \varepsilon,$$

où ε modélise l'erreur de mesure. On dispose d'un échantillon $X_1, \cdots, X_n$ de n mesures, où

$$X_i = \mu + \varepsilon_i.$$

On suppose qu'il n'y a pas d'erreur systématique, ce qui fait que en moyenne l'erreur est nulle. La **précision** d'une mesure est décrite par le paramètre

$$\frac{1}{\sigma},$$

où σ est l'écart type. La précision est grande si σ est petit.

La fréquence des erreurs ε_i tombant dans un intervalle $[a, b]$ est approximativement celle de l'aire

$$\int_a^b f(x)\mathrm{d}x, \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}},$$

qui est la **densité normale** de moyenne nulle et de variance σ^2. Comme $X = \mu + \varepsilon$, on verra dans le cours de probabilité que X est alors normale de moyenne μ et de variance σ^2, de densité

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

X est une variable aléatoire normale $\mathrm{N}(\mu, \sigma^2)$; le lecteur peut voir plusieurs de ces densités dans la figure **??**. La fréquence des données étant plus petites que le nombre a est donn'ee par la **fonction de répartition**

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

On peut voir que sous certaines hypothèses la variance empirique s^2 est une bonne approximation de σ^2, i.e.,

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

De même, on a l'approximation

$$\bar{x} \approx \mu.$$

Nous verrons dans la suite du cours que s^2 et $\bar{x}$ sont effectivement de bons estimateurs de σ^2 et μ (voir par exemple la figure **??**).

2.2.5 Le boxplot

Construction du boxplot :

- la hauteur du rectangle est l'EIQ, le bord bas est à $x_{.25}$ et le bord haut à $x_{.75}$.
- le trait épais au centre du rectangle est la médiane.
- la "moustache" supérieure est la valeur de l'observation la plus proche en deçà de BS $= x_{.75} + 1.5 \times$ EIQ.
- la "moustache" inférieure est la valeur de l'observation la plus proche au delà de BI $= x_{.25} - 1.5 \times$ EIQ.

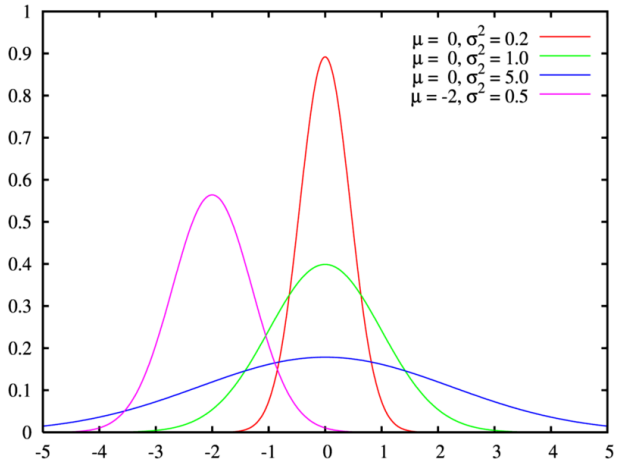


FIGURE 5 – Graphes de la densité normale pour différentes valeurs de μ et σ

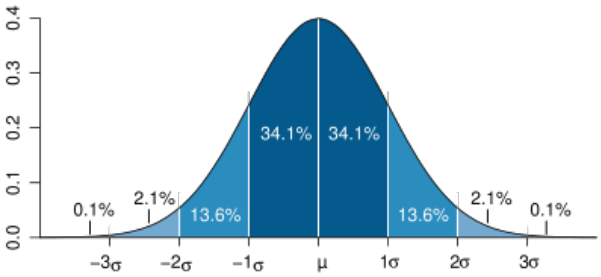


FIGURE 6 – Probabilités associées à certains secteurs caractéristiques de la densité normale $\mathrm{N}(\mu, \sigma^2)$ de moyenne μ et d'écart type σ.

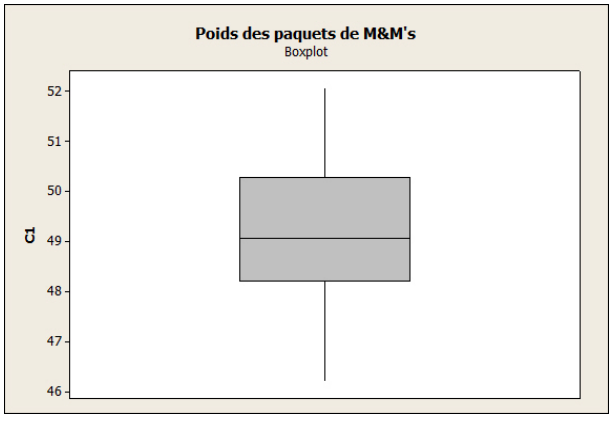


FIGURE 7 – Un boxplot associé aux poids des paquets de MM's

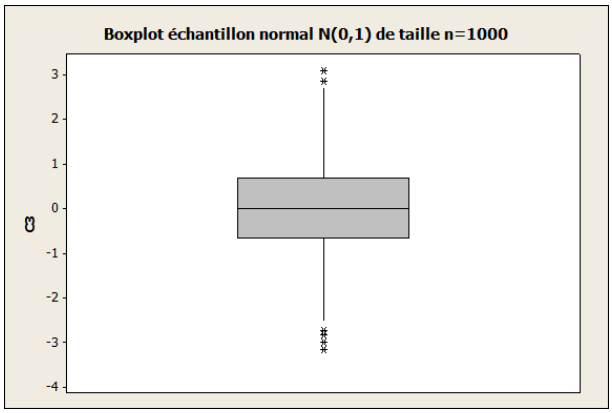


FIGURE 8 – Boxplot associé à un échantillon normal N(0,1) de taille 1000. En moyenne, on a 7 données excentriques.

– les points au delà de ces moustaches sont considérés comme des observations extrêmes, peut-être aberrantes, à regarder de plus près.

Pour qu'une valeur soit excentrique, il faut la comparer avec un standard, qui est la loi normale ; Pour la loi normale,

$$x_{.5} = \mu,\ x_{.25} = \mu - 0.6745\sigma,\ x_{.75} = \mu + 0.6745\sigma.$$

Il s'ensuit que l'étendue interquartile et les moustaches sont données par

$$\text{EIQ} = 1.349\sigma,\ \text{BI} = \mu - 2.698\sigma,\ \text{BS} = \mu + 2.698\sigma.$$

Si F_X désigne la fonction de répartition associée à la loi normale $\text{N}(\mu, \sigma^2)$, on a

$$F_X(\text{BI}) = 0.0035 = 1 - F_X(\text{BS}).$$

Ainsi, sur 1000 observations normales, il y en a en moyenne 7 qui sont excentriques (c.f. figure **??**). La statistique accepte un pourcentage (faible) d'erreurs dans le but de pouvoir contrôler correctement et fréquemment la normalité.

2.3 Données multivariées

Nous illustrons ici un exemple bivarié : Le biologiste T. Carlson a étudié une population de levures (*saccharomyce*). Les mesures décrivent l'évolution de la population lorsque le temps t est mesuré en heure [h] et $N(t)$ donne un nombre proportionnel au nombre de levures vivant en t. Les données obtenures sont de la forme $(t_i, N(t_i))$, $i = 0, \cdots, n$, où $n = 19$:

$$(0, 9.6),\ (1, 18.3),\ (2, 29),\ (3, 47.2),\ (4, 71.1),\ (5, 119.1),\ (6, 174.6),\ (7, 257.33),\ (8, 350.7),\ (9, 441),\ (10, 513.3),$$

$$(11, 559.7),\ (12, 594.8),\ (13, 629.4),\ (14, 640.8),\ (15, 651.1),\ (16, 655.9),\ (17, 659.6),\ (18, 661.8).$$

La première chose à faire consiste à représenter les données graphiquement (scatter plot), comme dans la figure (**??**) Le scatter plot présente une allure sigmoïdale typique dans ce contexte expérimental. Le modèle standard en croissance de population est la **courbe logistique**

$$N(t) = \frac{K}{1 + Ce^{-lt}},$$

où K, C et l sont des paramètres positifs que l'on peut ajuster (ou estimer) à partir des données. Cette courbe est un grand classique, et est solution de l'équation différentielle de Verhulst

$$\frac{\mathrm{d}N}{\mathrm{d}t} = lN(K - N).$$

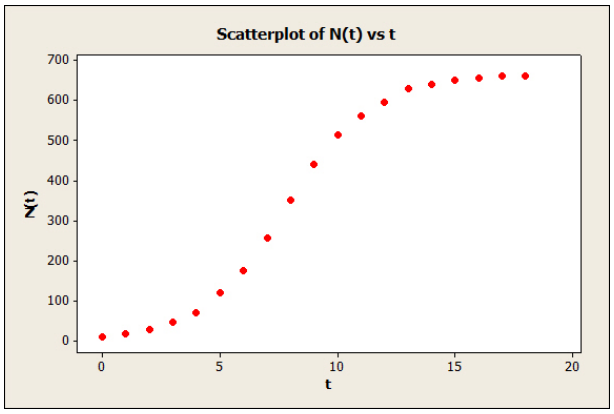


FIGURE 9 – Scatter plot de $N(t_i)$ versus t_i. On voit émerger l'allure sigmoïdale typique en croissance de population.

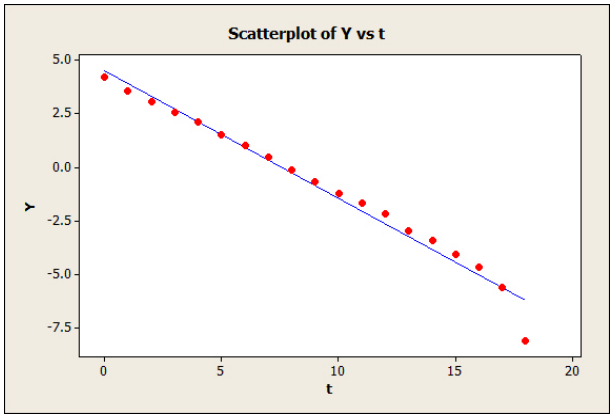


FIGURE 10 – Le graphe de $Y_i = Y(t_i)$ versus t_i. La droite est obtenue en appliquant la méthode des moindres carrés (régression linéaire)

Une telle courbe commence par augmenter exponentiellement vite comme fonction de t puis entre dans une phase de saturation pour t assez grand, la valeur du niveau de saturation étant K. Le problème statistique consiste à estimer les paramètres K, C et l de manière à ce que la courbe explique le mieux possible les données.

Une méthode courante en analyse exploratoire consiste à appliquer des transformations sur les données, typiquement en prenant le log ou... Cette approche est fructueurse dans notre situation : Posons

$$Y = \ln\left(\frac{K-N}{N}\right).$$

On remarque alors que

$$Y = \ln(C) - lt,$$

qui est une fonction affine de t ! La transformation utilise le paramètre inconnu K que l'on doit estimer à partir des données. Une méthode simple consiste à faire varier K ; Pour chaque valeur de K, on cherche les paramètres l et C qui mènent au meilleur ajustement. On pose pour illustrer la méthode $K = 662$. Le figure (**??**) donne la représentation des données transformées $(t_i, Y(t_i))$; on voit apparaître le graphe d'une fonction affine, ce qui nous indique que le modèle de croissance logistique est bien adapté aux données. Dans le cadre du modèle statistique donné ci-dessus $Y = \ln(C) - lt$, qui possède la forme classique en mathématique

$$Y = at + b,$$

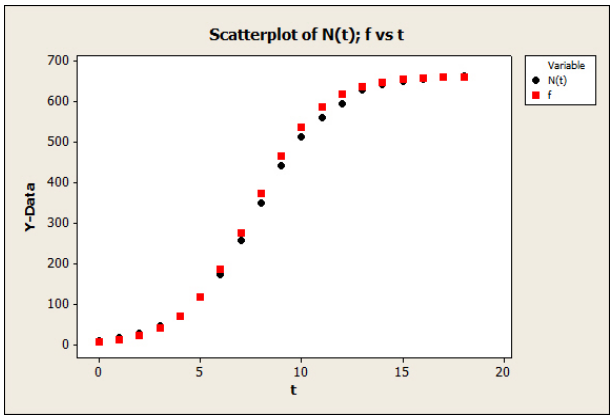


FIGURE 11 – Comparaison des graphes $(t_i, N(t_i))$ et $(t_i, f(t_i))$ où la première courbe est basée sur les données expérimentales et la seconde est obtenue par la méthode des moindres carrés

où la pente de la droite vaut $a = -l$ et l'ordonnée à l'origine est $b = \ln(C)$.

On cherche ensuite la droite qui passe le mieux au travers du nuage de point $(t_i, Y(t_i))$, en utilisant la méthode des moindres carrés, qui a été inventée par Gauss. Posons

$$\bar{t} = \frac{\sum_{i=0}^{18} t_i}{n}, \quad \bar{Y} = \frac{\sum_{i=0}^{18} Y_i}{n}, \quad Y_i = Y(t_i),$$

$$S_{tt} = \sum_{i=0}^{18} (t_i - \bar{t})^2, \quad S_{YY} = \sum_{i=0}^{18} (Y_i - \bar{Y})^2,$$

et

$$S_{tY} = \sum_{i=0}^{18} (t_i - \bar{t})(Y_i - \bar{Y}).$$

On verra dans la suite du cours que les paramètres optimaux sont donnés par

$$\begin{aligned} a &= \frac{S_{tY}}{S_{tt}}, \\ b &= \bar{Y} - a\bar{X}. \end{aligned}$$

Ceci nous indique que le point constitué des moyennes arithmétiques $(\bar{t}, \bar{Y})$ appartient à la droite de régression. La droite est représentée dans la figure (**??**). On revient ensuite aux paramètres l et C à l'aide des relations $a = -l$ et $b = \ln(C)$. La figure (**??**) compare les données expérimentales à la courbe obtenue à l'aide de la méthode des moindres carrés ; On voit sans peine que le modèle logistique est particulièrement bien adapté aux données expérimentales.

2.4 Conclusions

L'analyse exploratoire prend du temps. Quand on présente ses résultats.

– Les graphiques doivent rester simples et clairs.

– Tout graphique présenté doit être décrit avec précision : quels sont les axes et les unités, quel est le but du graphique, etc.

– Tout tableau de statistiques doit être décrit avec précision : quels sont les unités, arrondir les statistiques à la décimale reflétant la précision de la statistique.

– Tirer des conclusions de chaque graphique et tableau de statistiques présentés.

– Quand le but est de comparer plusieurs graphiques, garder la même échelle pour tous.

3 Probabilités

Nous considérons ici des épreuves dites aléatoires, c'est-à-dire des épreuves dont les issues dépendent du hasard. En voici quelques exemples :

(1) Jet d'une pièce de monnaie ; issues : pile, face.
(2) Jet d'un dé à 6 faces ; issues : il y en a 6.
(3) n jets consécutifs d'une pièce de monnaie ; issues : il y en a 2^n, chacune étant formée d'une suite de longueur n constituée de pile ou face.
(4) Choix aléatoire d'un individu dans une population ; issues : chaque individu de la population.
(5) Choix aléatoire d'un nombre dans l'intervalle $[0, 1]$; issues : chaque nombre compris entre 0 et 1.

Dans un premier temps, nous n'envisageons que des épreuves finies, c'est-à-dire des épreuves comportant un nombre fini d'issues.

3.1 Modèle d'une épreuve finie

Considérons une épreuve aléatoire comportant N issues. Il est d'usage de désigner celles-ci par $\omega_1, \omega_2, \ldots, \omega_N$ et de former l'ensemble $\Omega := \{\omega_1, \omega_2, \ldots, \omega_N\}$. Nous convenons qu'un sous-ensemble
$A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_p}\} \subset \Omega$, $1 \leq i_1 < i_2 < \ldots < i_i \leq N$, représente l'événement (associé à l'épreuve) qui se réalise si et seulement si ω_{i_1} ou $\omega_{i_2} \ldots$ ou ω_{i_p} se réalise.

Exemple :
On jette un dé à 6 faces

ω_1 correspond à : la face n^o 1 est réalisée
$\vdots$
ω_6 correspond à : la face n^o 6 est réalisée.

$A = \{\omega_2, \omega_4, \omega_6\}$ représente donc l'événement : une face portant un nombre pair est réalisée.

L'ensemble Ω correspond ainsi à l'événement certain (celui qui est toujours réalisé) tandis que le sous-ensemble vide, noté $\emptyset$, représente l'événement impossible.

Remarques :

– Les issues d'une épreuve sont aussi appelées événements élémentaires par opposition à un événement composé dont la réalisation est impliquée par plusieurs issues (ex. $A = \{\omega_2, \omega_4, \omega_6\}$ ci-dessus est un événement composé).
– Dans ces notations, les issues seront notées $\{\omega_i\}$ afin de les comprendre comme sous-ensemble de Ω.

Si $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ est l'événement certain d'une épreuve aléatoire comportant N issues, alors la famille de tous les événements associés à cette épreuve est donnée par la famille de tous les sous-ensembles de Ω. Cette famille sera notée $\mathcal{P}(\Omega)$.

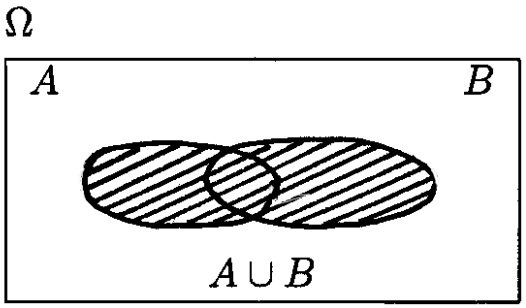


FIGURE 12 –

3.2 Plusieurs jets consécutifs d'une pièce de monnaie

On considère l'épreuve consistant à jeter n fois une pièce de monnaie. Cette épreuve joue un rôle fondamental dans cette théorie. Il s'agit de n répétitions de l'épreuve aléatoire la plus simple puisqu'elle admet 2 issues, à savoir pile ou face, lors de chaque jet. (Une épreuve avec une seule issue perd bien sûr tout caractère aléatoire !)

Afin de simplifier les notations, pile sera noté 1 et face 0. Ainsi une issue de l'épreuve consistant à jeter n fois une pièce de monnaie est représentée par une suite de longueur n formée de 0 et de 1 en convenant que le $i^{ième}$ élément donne le résultat du $i^{ième}$ jet.

Exemple : $n = 5$
$(0, 1, 1, 0, 1)$ est l'issue correspondant à face au premier jet, pile au second, pile au troisième, face au quatrième et pile au cinquième.

Exercice : Vérifier que le nombre d'issues de l'épreuve ci-dessus est 2^n.

Ainsi $\Omega = \{\omega_1, \omega_2, \ldots, \omega_{2^n}\}$ = ensemble de toutes les suites de longueur n formées de 0 et de 1. (Cet ensemble est souvent noté $\{0, 1\}^n$.) Par conséquent le nombre de sous-ensembles de Ω est $2^{(2^n)}$ et donc le nombre d'événements associé à cette épreuve est $2^{(2^n)}$.

Exemple : Si $n = 6$, alors $2^{(2^6)} = 2^{64}$.

En résumé, une épreuve finie est représentée par $\Big(\Omega, \mathcal{P}(\Omega)\Big)$ où Ω est l'événement certain dont les points correspondent aux issues de l'épreuve et $\mathcal{P}(\Omega)$ est la famille de tous les événements.

3.3 Opérations sur les événements

Soit $\Big(\Omega, \mathcal{P}(\Omega)\Big)$ une épreuve aléatoire finie et A, B, deux événements associées i.e. $A, B \in \mathcal{P}(\Omega)$ ($\Longleftrightarrow$ A, B sous-ensembles de Ω).

- $A \cup B$ est l'événement réalisé lorsque A ou B est réalisé
- $A \cap B$ est l'événement réalisé lorsque A et B sont réalisés
- $\overline{A} = A^c$ est l'événement contraire de A, à savoir celui qui est réalisé lorsque A ne l'est pas. Ainsi $\overline{\Omega} = \emptyset$.

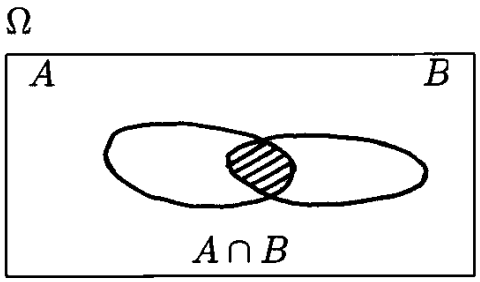


FIGURE 13 –

Définition 3.1 *Deux événements A et B associés à une épreuve sont dits* incompatibles *si $A \cap B = \emptyset$, c'est-à-dire si leur réalisation simultanée est impossible.*

Exemple :
Epreuve : jet d'un dé à 6 faces, $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.

$$A = \{\omega_1, \omega_3, \omega_5\} = \text{ une face impaire est réalisée.}$$

$$B = \{\omega_2, \omega_4, \omega_6\} = \text{ une face paire est réalisée.}$$

Alors $A \cap B = \emptyset$.

3.4 La notion de probabilité

Nous nous appuyons ici sur l'interprétation de la probabilité comme limite d'une fréquence. Soit A un événement lié à une épreuve. Supposons que celle-ci est répétée n fois en prenant garde que ces répétitions n'interfèrent pas entre elles. On note n_A le nombre de realisations de A dans ces n répétitions. Le nombre $\frac{n_A}{n}$ est compris entre 0 et 1 et est appelé fréquence relative de réalisation de A dans ces n répétitions.

Credo : lorsque $n \to \infty$, $\frac{n_A}{n}$ se rapproche d'un nombre noté $P(A)$ et appelé probabilité de A :

$$\frac{n_A}{n} \underset{n\to\infty}{\rightsquigarrow} P(A)$$

! Il ne s'agit pas de la convergence usuelle d'une suite de nombres réels. En effet, si l'on jette une infinité de fois une pièce symétrique, des événements élémentaires tels que pile n'est jamais réalisé (idem pour face) sont possibles. Si A est l'événement "pile est réalisé lors d'un jet" alors, pour tout n, $n_A = 0$ ($n_A = n$) et $\frac{n_A}{n}$ ne tend pas vers $\frac{1}{2}$ comme on pourrait l'espérer pour une pièce symétrique.

3.5 Propriétés d'une probabilité (épreuve finie)

La probabilité d'un événement associé à une épreuve est un nombre compris entre 0 et 1 qui mesure sa chance de réalisation lors de l'épreuve. Si cette dernière est représentée par $\Big(\Omega(\mathcal{P}(\Omega)\Big)$, alors une probabilité P associe un nombre compris entre 0 et 1 à tout sous-ensemble A de Ω :

$$A \subset \Omega \longmapsto P(A) \in [0, 1].$$

Autre notation équivalente :

$$A \in \mathcal{P}(\Omega) \longmapsto P(A) \in [0,1].$$

Puisque Ω représente l'événement certain il est logique de poser $P(\Omega) = 1$. De plus, soient A et B deux événements incompatibles ($A \cap B = \emptyset$) liés à l'épreuve. Supposons que lors de n répétitions de l'épreuve, A a été realisé n_A fois et B, n_B fois. Puisque A et B sont incompatibles, on a $n_{A\cup B} = n_A + n_B$ et donc

$$\frac{n_{A\cup B}}{n} = \frac{n_A + n_B}{n} = \frac{n_A}{n} + \frac{n_B}{n}\,.$$

En faisant tendre $n \to \infty$, notre credo suggère que

$$P(A \cup B) = P(A) + P(B).$$

Nous réunissons les propriétés précédentes dans la définition suivante :

Définition 3.2 *Si $\Big(\Omega, \mathcal{P}(\Omega)\Big)$ représente une épreuve finie (Ω est fini) alors une probabilité P associée à cette dernière vérifie :*

1) $A \in \mathcal{P}(\Omega) \mapsto P(A) \in [0,1]$

2) $P(\Omega) = 1$

3) Si $A, B \in \mathcal{P}(\Omega)$, $A \cap B = \emptyset$, alors $P(A \cup B) = P(A) + P(B)$.

Remarque : La propriété 3) porte le nom d'additivité.

Par induction finie on en déduit que, pour n événements $A_1, A_2, \ldots, A_n \in \mathcal{P}(\Omega)$ incompatibles deux à deux, c'est-à-dire $A_i \cap A_j = \emptyset$ si $i \neq j$, $1 \leq i, j \leq n$, on a

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

où $\bigcup\limits_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n$.

Conséquences :
Les propriétés ci-dessous découlent toutes de 1), 2) et 3) :

- $P(\overline{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- Si $A \subset B$, alors $P(A) \leq P(B)$ et $P(B \setminus A) = P(B) - P(A)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Remarque : Une probabilité P associée à une épreuve finie est complètement déterminée par les probabilités des événements élémentaires. En effet, si $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ et $A \subset \Omega$, alors $A = \{\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_\ell}\}$ avec $1 \leq i_1 < i_2 < \ldots < i_\ell \leq N$, et

$$\begin{aligned} P(A) &= P\left(\bigcup_{k=1}^{\ell} \{\omega_{i_k}\}\right) \quad \Big(\text{ les } \{\omega_{i_k}\} \text{ sont incompatibles 2 à 2}\Big) \\ &= \sum_{k=1}^{\ell} P(\{\omega_{i_k}\}). \end{aligned}$$

On déduit donc la valeur de $P(A)$ de celles des $P(\{\omega_i\})$. Il est clair que $0 \leq P(\{\omega_i\}) \leq 1$ et $\sum\limits_{i=1}^{N} P(\{\omega_i\}) = P(\Omega) = 1$.

Pour modéliser une épreuve infinie, on doit en général renoncer à $\mathcal{P}(\Omega)$ et remplacer cette famille par une "σ–algèbre".

Un cas particulier important :
La probabilité uniforme dans une épreuve finie associe la même valeur à tous les événements élémentaires. Plus précisément, si $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ alors $P(\{\omega_i\}) = \frac{1}{N}$ pour tout $1 \leq i \leq N$. Si $A = \{\omega_{i_1}, \ldots, \omega_{i_\ell}\}$, alors

$$P(A) = \sum_{k=1}^{\ell} P(\{\omega_{i_k}\}) = \sum_{k=1}^{\ell} \frac{1}{N} = \frac{\ell}{N} = \frac{\#(A)}{N} = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} .$$

3.5.1 Le problème des anniversaires

Nous supposons que les années comptent 365 jours et nous réunissons n ($n \leq 365$) personnes choisies au hasard dans une population. Nous nous intéressons à l'événement

$A_n =$ 2 personnes au moins parmi les n ont un anniversaire commun
(jours identiques mais années éventuellement différentes)

Pour calculer $P(A_n)$, il est judicieux de passer par l'événement contraire

$$P(A_n) = 1 - P(\overline{A}_n)$$

où

$\overline{A}_n =$ aucun anniversaire commun parmi les n personnes.

En admettant que les jours de naissance sont répartis uniformément dans l'année nous avons :

$$\begin{aligned} P(\overline{A}_n) &= \frac{\#\text{ cas favorables}}{\#\text{ cas possibles}} = \frac{365 \cdot 364 \cdots (365-(n-1))}{365 \cdot 365 \cdots 365} \\ &= \frac{365 \cdot 364 \cdots (365-(n-1))}{365^n} . \end{aligned}$$

$P(A_n)$ dépasse la valeur $\frac{1}{2}$ dès que $n \geq 23$.

3.6 Notion de probabilité conditionnelle

Soient A et B deux événements liés à une épreuve avec $P(A) > 0$ et $P(B) > 0$. Nous nous intéressons à $P(A \mid B) =$ probabilité pour que A se réalise sachant que B est réalisé.

Exemple : On jette un dé symétrique à 6 faces (i.e. les faces sont équiprobables).

$A =$ une face portant un nombre pair est réalisée
$B =$ une face portant un nombre plus grand ou égal à 4 est réalisée.

$P(A \mid B) = ?$

Revenons à la situation générale et à notre credo. Supposons que l'épreuve a été répétée n fois et que B et $A \cap B$ ont été réalisés respectivement n_B et $n_{A\cap B}$ fois. Nous ne tenons pas compte des situations dans lesquelles B n'est pas réalisé. Ainsi la fréquence relative qui nous intéresse ici est $\dfrac{n_{A\cap B}}{n_B}$ et sa "limite" lorsque $n \to \infty$ doit fournir $P(A)B$). Or

$$\frac{n_{A\cap B}}{n_B} = \frac{\frac{n_{A\cap B}}{n}}{\frac{n_B}{n}} \underset{n\to\infty}{\rightsquigarrow} \frac{P(A \cap B)}{P(B)} = P(A \mid B).$$

Ainsi, on peut définir

Définition 3.3

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \textit{ou de façon équivalente } P(A \cap B) = P(A \mid B)P(B)\,.$$

Dans notre exemple $\Omega = \{1,2,3,4,5,6\}$ et $P(\{i\}) = \frac{1}{6}$, $1 \leq i \leq 6$.

$$A = \{2,4,6\}, \quad B = \{4,5,6\}, \quad A \cap B = \{4,6\}$$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}\,.$$

3.6.1 Le jeu des trois boîtes

Avant le début du jeu, le présentateur introduit une boule dans une des trois boîtes dont il dispose. Le joueur doit deviner la boîte qui contient la boule. Il désigne donc une boîte et le présentateur ouvre alors une des deux autres et lui montre qu'elle est vide. Il laisse au joueur la possibilité de modifier son choix initial. Que doit faire ce dernier ?

La notion de probabilité conditionnelle sera utilisée pour construire un modèle probabiliste de ce jeu. Numérotons les boîtes de 1 à 3 et posons :

X = numéro de la boîte contenant la boule
Y = numéro de la boîte désignée par le joueur
Z = numéro de la boîte vide ouverte par le présentateur.

Nous recensons d'abord les événements élémentaires :

X	Y	Z	X	Y	Z	X	Y	Z
1	1	2	2	1	3	3	1	2
1	1	3	2	2	1	3	2	1
1	2	3	2	2	3	3	3	1
1	3	2	2	3	1	3	3	2

Afin d'alléger l'écriture nous introduisons la notation suivante

$$\{X = i, Y = j, Z = k\} = \{X = i\} \cap \{Y = j\} \cap \{Z = k\}$$

où i, j, k prennent les valeurs précises dans le tableau précédent. Il est clair que

$\{X = Y\} =$ le joueur gagne en maintenant son choix initial
$\{X \neq Y\} =$ le joueur gagne en modifiant son choix initial
et $P\{X \neq Y\} = 1 - P\{X = Y\}$.

Nous devons discuter deux types d'événements élémentaires, à savoir ceux qui contiennent deux fois un même numéro et les autres. Ainsi, en utilisant la probabilité conditionnelle :

$$P(Z = 2, Y = 1, X = 1) = P(Z = 2 \mid Y = 1, X = 1)P(Y = 1, X = 1)$$

$$\Big(P(A \cap B) = P(A \mid B)P(B) \text{ avec } A = \{Z = 2\} \text{ et } B = \{Y = 1, X = 1\}\Big).$$

Si $X = 1$, et $Y = 1$, alors le présentateur peut choisir d'ouvrir les boites 2 ou 3. Nous poserons donc

$$P(Z = 2 \mid Y = 1, X = 1) = \frac{1}{2}\,.$$

De même nous aurons

$$P(Y = 1, X = 1) = P(Y = 1 \mid X = 1)P(X = 1).$$

Le joueur ne sachant pas où se trouve la boule, il est logique de poser $P(Y = 1 \mid X = 1) = P(Y = 1) = \frac{1}{3}$. Finalement nous admetterons que $P(X = 1) = \frac{1}{3}$. Par conséquent :

$$P(Z = 2, Y = 1, X = 1) = \frac{1}{2} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{18}\,.$$

A l'aide du même raisonnement nous obtenons :

$$\begin{aligned} \frac{1}{18} &= P(Z = 3, Y = 1, X = 1) = P(Z = 1, Y = 2, X = 2) \\ &= P(Z = 3, Y = 2, X = 2) = P(Z = 1, Y = 3, X = 3) \\ &= P(Z = 2, Y = 3, X = 3). \end{aligned}$$

Par ailleurs,

$$P(Z = 3, Y = 2, X = 1) = P(Z = 3 \mid Y = 2, X = 1)P(Y = 2 \mid X = 1)P(X = 1).$$

A l'évidence nous poserons :

$$\begin{aligned} &P(Z = 3 \mid Y = 2, X = 1) = 1 \\ &P(Y = 2 \mid X = 1) = P(Y = 2) = \frac{1}{3} \\ &P(X = 1) = \frac{1}{3}\,. \end{aligned}$$

Par conséquent nous aurons :

$$\begin{aligned} \frac{1}{9} &= P(Z = 3, Y = 2, X = 1) = P(Z = 2, Y = 3, X = 1) \\ &= P(Z = 3, Y = 1, X = 2) = P(Z = 1, Y = 3, X = 2) \\ &= P(Z = 2, Y = 1, X = 3) = P(Z = 1, Y = 2, X = 3). \end{aligned}$$

Il suffit maintenant de sommer les probabilités des événements élémentaires constituant les événements qui nous intéressent :

$$P(X = Y) = \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} + \frac{1}{18} = \frac{6}{18} = \frac{1}{3}.$$

et donc

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

Conclusion : le joueur a intérêt à modifier son choix initial.

3.7 Théorème des probabilités totales et formule de Bayes

Soit Ω l'événement certain associé à une épreuve et A, $B_1, B_2, \ldots, B_n$ des événements liés à celle-ci mais tels que :

- $B_i \cap B_j = \emptyset$ si $i \neq j$, $1 \leq i, j \leq n$ (les B_i sont incompatibles 2 à 2)
- $\bigcup_{i=1}^{n} B_i = \Omega$.

Alors on a :

(a) Théorème des probabilités totales :

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i).$$

(b) Formule de Bayes :

$$\text{Pour } 1 \leq j \leq n, \quad P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}\,.$$

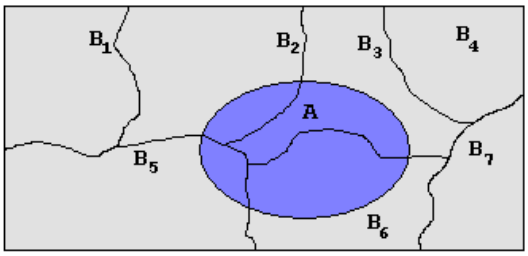


FIGURE 14 – La décomposition de A à l'aide d'une partition $B_1, \cdots, B_n$

Démonstration :

(a) $A = A \cap \left(\bigcup_{i=1}^{n} B_i \right) = \bigcup_{i=1}^{n} A \cap B_i$.

Les événements B_i étant incompatibles 2 à 2, il en va de même des $A \cap B_i$. L'additivité de la probabilité fournit :

$$\begin{aligned} P(A) &= \sum_{i=1}^{n} P(A \cap B_i) \\ &= \sum_{i=1}^{n} \frac{P(A \cap B_i)}{P(B_i)} P(B_i) \\ &= \sum_{i=1}^{n} P(A \mid B_i) P(B_i). \end{aligned}$$

(b)

$$\begin{aligned} P(B_j \mid A) &= \frac{P(B_j \cap A)}{P(A)} \\ &= \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j) P(B_j)}{P(A)} \\ &\overset{(a)}{=} \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i) P(B_i)} . \end{aligned}$$

Applications : On tire consécutivement (sans remise) deux billets d'un lot de n billets parmi lesquels m sont gagnants.

$G_1 =$ obtenir un billet gagnant lors du premier tirage,
$G_2 =$ obtenir un billet gagnant lors du second tirage.

Il est clair que $P(G_1) = \frac{m}{n}$. Pour calculer $P(G_2)$ nous pouvons utiliser le théorème des probabilités totales en posant :

$$A = G_2, \quad n = 2, \quad B_1 = G_1, \quad B_2 = \overline{G}_1 \quad (G_1 \cap \overline{G}_1 = \emptyset \text{ et } G_1 \cup \overline{G}_1 = \Omega).$$

Ainsi :

$$\begin{aligned} P(G_2) &= P(G_2 \mid G_1) P(G_1) + P(G_2 \mid \overline{G}_1) P(\overline{G}_1) \\ &= \frac{m-1}{n-1} \frac{m}{n} + \frac{m}{n-1} \left(1 - \frac{m}{n}\right) \\ &= \frac{m}{n-1} \left(\frac{m-1}{n} + 1 - \frac{m}{n} \right) \\ &= \frac{m}{n-1} \left(1 - \frac{1}{n} \right) = \frac{m}{n-1} \frac{n-1}{n} = \frac{m}{n} = P(G_1). \end{aligned}$$

Il est possible de démontrer que cette probabilité reste la même pour tous les tirages successifs. (Problème des billets de loterie discuté dans le cours.)

3.7.1 Applications de la formule de Bayes

On considère 10 pièces de monnaie dont l'une est truquée car ses deux côtés sont des piles. On choisit une pièce au hasard et on la jette. Sachant que le résultat du jet est pile, calculer la probabilité pour que la pièce en question soit la pièce truquée.

Numérotons les pièces de 1 à 10 en convenant que la première est la pièce truquée.

B_i = on tire la pièce n^o i, $1 \leq i \leq 10$
A = pile est réalisé en jetant la pièce choisie

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(A \mid B_1)P(B_1)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_{10})P(B_{10})} \\ &= \frac{1 \cdot \frac{1}{10}}{1 \cdot \frac{1}{10} + 9\frac{1}{2} \cdot \frac{1}{10}} = \frac{1}{1 + \frac{9}{2}} = \frac{1}{\frac{11}{2}} = \frac{2}{11} \, . \end{aligned}$$

3.7.2 Détection d'une maladie

Nous considérons un test pour la détection d'une maladie dans une population donnée. Nous introduisons les notations :

$$\begin{aligned} M &= \text{un individu, choisi au hasard dans la population, est malade,} \\ A &= \text{le test, appliqué à un individu, est positif.} \end{aligned}$$

Sachant que

$$P(M) = 0,001, \quad P(A \mid M) = 0,95 \quad \text{et} \quad P(\overline{A} \mid \overline{M}) = 0,95,$$

peut-on conclure que le test est de bonne qualité ?

La grandeur qui nous intéresse est en fait $P(M \mid A)$ et la formule de Bayes nous fournit :

$$P(M \mid A) = \frac{P(A \mid M)P(M)}{P(A \mid M)P(M) + P(A \mid \overline{M})P(\overline{M})} \, .$$

Remarquons que $P(A \mid \overline{M}) = 1 - P(\overline{A} \mid \overline{M})$. En effet, si A et B sont deux événements, alors

$$\begin{aligned} P(\overline{A} \mid B) &= \frac{P(\overline{A} \cap B)}{P(B)} = \frac{P\Big(B \smallsetminus (A \cap B)\Big)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} \\ &= 1 - \frac{P(A \cap B)}{P(B)} = 1 - P(A \mid B). \end{aligned}$$

Par conséquent :

$$P(M \mid A) = \frac{0,95 \cdot 0,001}{0,95 \cdot 0,001 + 0,05 \cdot 0,999} \cong 0,0187 \, .$$

Seulement $1,8$ % des malades sont détectés par le test !

3.8 Evénements indépendants :

Considérons deux événements A_1 et A_2 liés à une épreuve et tels que $P(A_1) > 0$ et $P(A_2) > 0$. Comment pouvons-nous traduire l'idée "A_1 est indépendant de A_2" ? Une façon de procéder consiste à exiger :

$$P(A_1 \mid A_2) = P(A_1)$$

pour signifier que la réalisation de A_2 n'influence pas celle de A_1. Dans ce cas nous avons les équivalences suivantes :

$$\begin{aligned} P(A_1 \mid A_2) = P(A_1) &\iff \frac{P(A_1 \cap A_2)}{P(A_2)} = P(A_1) \iff P(A_1 \cap A_2) = P(A_1)P(A_2) \\ &\iff \frac{P(A_2 \cap A_1)}{P(A_1)} = P(A_2) \iff P(A_2 \mid A_1) = P(A_2). \end{aligned}$$

Ainsi "A_1 est indépendant de A_2" équivaut à "A_2 est indépendant de A_1" entraînant la symétrie de cette notion. Afin d'inclure les cas où $P(A_1)$ et $P(A_2)$ peuvent être nuls, nous travaillerons avec la définition suivante :

Définition 3.4 *Deux événements A_1 et A_2 liés à une épreuve sont dits indépendants si $P(A_1 \cap A_2) = P(A_1)P(A_2)$.*

Exemple : On tire consécutivement deux billets d'un lot de n billets parmi lesquels m sont gagnants. Désignons par G_1 et G_2 les événements qui consistent respectivement à tirer un billet gagnant en première et seconde position. Nous avons déjà vu que $P(G_1) = P(G_2) = \frac{m}{n}$. Ces deux événements sont-ils indépendants ?

$$P(G_2 \cap G_1) = P(G_2 \mid G_1)P(G_1) = \frac{m-1}{n-1}\frac{m}{n} \neq \frac{m}{n} \cdot \frac{m}{n} = P(G_2)P(G_1).$$

En conclusion G_1 et G_2 ne sont pas indépendants. Si par contre les tirages s'effectuent *avec remise*, alors G_1 et G_2 sont indépendants car $P(G_2 \mid G_1) = P(G_2) = \frac{m}{n}$ et donc

$$P(G_2 \cap G_1) = P(G_2 \mid G_1)P(G_1) = \frac{m}{n}\frac{m}{n} = P(G_2)P(G_1).$$

Comment définir l'indépendance de 3 événements ou plus ? Considérons d'abord A_1, A_2 et A_3 liés à une épreuve. Une façon raisonnable (sous forme conditionnelle) de définir l'indépendance de ces 3 événements est d'exiger :

$$\begin{aligned} &P(A_1 \mid A_2 \cap A_3) = P(A_1 \mid A_2) = P(A_1 \mid A_3) = P(A_1) \\ &P(A_2 \mid A_1 \cap A_3) = P(A_2 \mid A_1) = P(A_2 \mid A_3) = P(A_2) \\ &P(A_3 \mid A_1 \cap A_2) = P(A_3 \mid A_1) = P(A_3 \mid A_2) = P(A_3). \end{aligned}$$

Un calcul direct montre que cet ensemble de propriétés équivaut à (forme "produit")

$$P(A_1 \cap A_2) = P(A_1)P(A_2), \quad P(A_1 \cap A_3) = P(A_1)P(A_3), \quad P(A_2 \cap A_3) = P(A_2)P(A_3),$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

Les trois premières propriétés reflètent l'indépendance 2 à 2 de A_1, A_2 et A_3 tandis que la dernière est l'indépendance 3 à 3. Malheureusement ces deux notions ne s'impliquent pas mutuellement. Voici un exemple de 3 événements indépendants 2 à 2 mais pas 3 à 3.

$$\Omega = \{1,2,3,4\}, \quad P(\{i\}) = \frac{1}{4}, \quad 1 \leq i \leq 4$$

$$A_1 = \{1,2\}, \quad A_2 = \{1,3\}, \quad A_3 = \{1,4\}, \quad P(A_1) = P(A_2) = P(A_3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}\,.$$

$$P(A_1 \cap A_2) = P(\{1\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_2)$$

$$P(A_1 \cap A_3) = P(\{1\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_1)P(A_3)$$

$$P(A_2 \cap A_3) = P(\{1\}) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A_2)P(A_3)$$

et ainsi A_1, A_2 et A_3 sont indépendants 2 à 2. Par contre

$$P(A_1 \cap A_2 \cap A_3) = P(\{1\}) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}\,.$$

Nos événements ne sont pas indépendants 3 à 3.

La notion d'indépendance utilisée en théorie des probabilités est :

Définition 3.5 *Les événements $A_1, A_2, \ldots, A_n$ sont dits indépendants si*

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_m}) = P(A_{i_1})P(A_{i_2}) \ldots P(A_{i_m})$$

pour tout sous-ensemble $\{i_1, i_2, \ldots, i_m\} \subset \{1, 2, \ldots, n\}$. Ainsi n événements sont indépendants s'ils le sont 2 à 2, 3 à 3, ..., n à n.

4 La notion de variable aléatoire

Dans le modèle d'une épreuve aléatoire, les éléments de l'événement certain Ω représentent les issues (= événements élémentaires) de l'épreuve. Une fonction $X\colon \Omega \to \mathbb{R}$ associe donc à chaque issue $\omega \in \Omega$ un nombre $X(\omega)$ qui dépend du hasard puisque tel est est le cas de l'argument ω. Une telle fonction porte le nom de variable aléatoire (v.a.). Dans le cas d'une épreuve infinie (non dénombrable) une condition technique supplémentaire est exigée.

Exemples :

1) Dans la population humaine du Canton de Fribourg, on choisit au hasard un individu et on mesure son poids. Si Y désigne ce dernier, alors Y est une variable aléatoire à valeurs dans $\mathbb{R}_+$ (réels positifs).
2) On jette n fois une pièce de monnaie et on désigne par X le nombre de réalisations de pile dans les n jets. X est une variable aléatoire à valeurs dans $\{0, 1, \ldots, n\}$.
3) On considère un événement A lié à une épreuve aléatoire. On désigne par T le nombre de répétitions de l'épreuve pour obtenir la première apparition de A. T est une variable aléatoire à valeurs dans $N^* = \{1, 2, \ldots\}$.
4) Pierre et Paul jouent à un jeu de hasard. Ils disposent chacun d'une même fortune initiale et à chaque coup le gagnant reçoit 1 franc du perdant. Le jeu s'arrête lorsque la fortune d'un joueur atteint 0. La durée du jeu est une variable aléatoire ; la fortune de Pierre, tant que le jeu dure, est une variable aléatoire.

Nous considérons deux familles importantes de variables aléatoires.

4.1 Les variables aléatoires à valeurs entières

Soit X une variable aléatoire à valeurs dans $\mathbb{N} = \{0, 1, \ldots\}$. L'information stochastique d'une telle variable aléatoire est contenue dans la fonction

$$k \in \mathbb{N} \longmapsto P(X = k).$$

A l'aide de celle-ci il est en effet possible de calculer $P(n_1 \leq N \leq n_2)$, pour $n_1 \leq n_2$ quelconques :

$$\begin{aligned} P(n_1 \leq N \leq n_2) &= P\left(\bigcup_{k=n_1}^{n_2} \{X = k\}\right) \\ &= \sum_{k=n_1}^{n_2} P(X = k) \qquad \text{par additivité de } P. \end{aligned}$$

Soit $f_k = P(X = k)$. On a $0 \leq f_k \leq 1$ et on peut également montrer que $\sum_{k=0}^{\infty} f_k = 1$.

4.1.1 Les variables de Bernoulli

On jette une pièce de monnaie. A la réalisation de pile on associe la valeur 1 et 0 s'il s'agit de face. Si p désigne la probabilité de réalisation de pile et si X désigne le résultat du jet, on aura :

$$X = \begin{cases} 1 & \text{avec probabilité } p \\ 0 & \text{avec probabilité } q = 1 - p \end{cases} .$$

Une telle variable aléatoire est appelée variable aléatoire de Bernoulli de paramètre p et nous noterons $X = \text{Ber}(p)$.

4.1.2 Variable binômiale

On jette n fois une pièce de monnaie. On suppose les jets indépendants et on désigne par p la probabilité de réalisation de pile lors d'un jet. Soit

$$X_i = \begin{cases} 1 & \text{avec probabilité } p \\ \\ 0 & \text{avec probabilité } q = 1 - p \end{cases} , \qquad 1 \leq i \leq n,$$

le résultat du $i^{\text{ème}}$ jet avec la convention : 1 pour pile et 0 pour face. L'hypothèse d'indépendance des jets est traduite par l'indépendance des variables aléatoires
$X_1, X_2, \ldots, X_n$. Elles sont de plus identiquement distribuées, $X_i = \text{Ber}(p)$, $1 \leq i \leq n$.

Nous nous intéressons au nombre de réalisations de pile dans les n jets. Cette variable aléatoire, que nous noterons S_n, est donnée par :

$$S_n = \sum_{i=1}^{n} X_i \ .$$

S_n prend ses valeurs dans $\{0, 1, \ldots, n\}$ et nous calculons maintenant $P(S_n = k)$.
L'événement $\{S_n = k\}$ correspond à la réalisation de k fois pile dans n jets. Calculons d'abord la probabilité pour que, lors de n jets, les k premiers fournissent pile et les $n - k$ derniers face, i.e. :

$$\begin{aligned} &P(X_1 = 1, X_2 = 1, \ldots, X_k = 1, X_{k+1} = 0, \ldots, X_n = 0) \\ &\overset{\text{indépendance}}{=} P(X_1 = 1)P(X_2 = 1) \ldots P(X_k = 1)P(X_{k+1} = 0) \ldots P(X_n = 0) \\ &= p \cdot p \cdot \ldots \cdot p\, q \cdots q \\ &= p^k q^{n-k} . \end{aligned}$$

Chaque façon de réaliser exactement k fois pile dans n jets a la probabilité $p^k q^{n-k}$ d'être réalisée. L'additivité de la probabilité nous assure alors que $P(S_n = k)$ est donné par le nombre de façons de réaliser k fois pile dans n jets que l'on multipliera par $p^k q^{n-k}$. Le nombre cherché est identique au nombre de sous-ensembles de taille k que possède un ensemble de taille n. On peut démontrer que le nombre cherché est donné par

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad \text{où} \quad n! = n(n-1) \ldots 2 \cdot 1. \qquad \text{(exercices)}$$

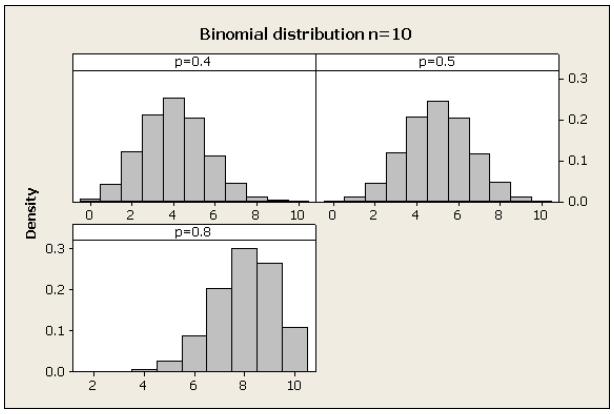


FIGURE 15 – Représentation graphique des valeurs prises par la loi binômiale lorsque $n = 10$, pour $p = 0.4$, $p = 0.5$ et $p = 0.8$.

Ainsi

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k}\,.$$

La figure **??** donne trois représentations graphiques des valeurs $P(S_n = k)$, $k = 0, \cdots, n$, lorsque $n = 10$, pour $p = 0.4$, $p = 0.5$ et $p = 0.8$. Une telle variable aléatoire est dite binômiale de paramètre n et p. On notera $S_n = \text{Bin}(n, p)$. Nous avons ainsi montré que la somme de n variables aléatoires $\text{Ber}(p)$ indépendantes est une variable $\text{Bin}(n, p)$. Les coefficients $\binom{n}{k}$ sont ceux du binôme de Newton :

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}\,,$$

et on peut ainsi en déduire que :

$$\sum_{k=0}^{n} P(S_n = k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1.$$

4.1.3 Variable géométrique

Considérons un événement A lié à une épreuve aléatoire dont la probabilité de réalisation est p. Nous effectuons des répétitions indépendantes de cette épreuve et nous désignons par T le nombre nécessaire à faire apparaître A. Pour $k \geq 1$, nous avons

$$P(T = k) = q^{k-1} p \quad \text{où} \quad q = 1 - p.$$

Une telle variable aléatoire est appelée géométrique de paramètre p et nous noterons $T = G(p)$. En utilisant l'égalité (série géométrie)

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}\ ,\ |x| < 1,$$

on obtient

$$\begin{aligned} \sum_{k=1}^{\infty} P(T = k) &= \sum_{k=1}^{\infty} q^{k-1} p = p \sum_{k=1}^{\infty} q^{k-1} \\ &= p \sum_{k=0}^{\infty} q^k = \frac{p}{1-q} = \frac{p}{p} = 1. \end{aligned}$$

4.1.4 Variable de Poisson

Considérons une suite de variables aléatoires $Y_n = \text{Bin}(n, p_n)$ telle que $\lim_{n\to\infty} np_n = \lambda > 0$. Rappelons que

$$P(Y_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k}, \quad 0 \le k \le n.$$

Un calcul direct montre que, pour tout k fixé,

$$\lim_{n\to\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = e^{-\lambda}\frac{\lambda^k}{k!}$$

où $\lambda = \lim_{n\to\infty} np_n$. Il est clair que

$$\begin{aligned} \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} \quad \text{(série de la fonction exponentielle)} \\ &= 1. \end{aligned}$$

Une variable aléatoire X vérifiant $P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k \in \mathbf{N}$ est dite variable aléatoire de Poisson de paramètre λ et nous noterons $X = \text{Poi}(\lambda)$. La condition $\lim_{n\to\infty} np_n = \lambda$ suggère que pour n grand et p petit, une variable aléatoire $\text{Bin}(n, p)$ peut être approchée par une variable aléatoire $\text{Poi}(\lambda)$ où $\lambda = np$. La règle $n > 10$ et $p < 0.05$ garantit une approximation convenable (voir les exercices).

4.1.5 La loi des séries

En 2005, 5 avions civils se sont crashés sur une période de 22 jours (Toronto, Palerme, Athènes, Venezuela, Amazonie). Comment peut-on expliquer cette série noire ? Est-ce dû au hasard ou alors traduisent-ils une baisse du niveau de sécurité dans les transports aériens ?

On peut faire quelques calculs afin de voir si le hasard peut expliquer cette série noire. La fréquence moyenne des crashs sur la période 1995-2004 était de 1/500 000. On en déduit que le probabilité que ces 5 avions se crashent vaut

$$P(\text{ces 5 avions se crashent}) = \left(\frac{1}{500000}\right)^5 \sim 32\ 10^{-30},$$

qui est donc très petite. On revient à notre question et on essaie de calculer la probabilité qu'au moins 5 avions se crashent sur une période donnée de 22 jours, soit

$$P(\text{au moins 5 avions se crashent sur une période donnée de 22 jours}).$$

On va utiliser le fait que le nombre moyen quotidien de décollages vaut environ 20 000 ; les accidents étant indépendants les uns des autres, on suppose par ailleurs que le nombre de crashs sur les 22 jours suit une loi binômiale avec $n = 22\ 20000$ et $p = 1/500000$. La probabilité d'avoir k crashs sur 22 jours vaut donc

$$\binom{n}{k} p^k (1-p)^{n-k}.$$

Les paramètres étant grands, on observe que

$$\lambda = np = 440000\ \frac{1}{500000} = 0.88,$$

ce qui nous permet d'utiliser l'approximation de la loi binômiale par la loi de Poisson. On trouve que

$$P(\text{au moins 5 avions se crashent sur une période donnée de 22 jours})$$

$$\sim 1 - P(\{0\}) - P(\{1\}) - P(\{2\}) - P(\{3\}) - P(\{4\}),$$

où

$$P(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

Cette approximation nous donne que $P(\{0\}) = 0.415$, $P(\{1\}) = 0.365$, $P(\{2\}) = 0.161$, $P(\{3\}) = 0.047$, $P(\{4\}) = 0.01$, et donc,

$$P(\text{au moins 5 avions se crashent sur une période donnée de 22 jours}) \approx 1 - 0.998 = 0.002,$$

soit une chance sur 500.

Une année comprenant 16 périodes consécutives de 22 jours, on en déduit que la probabilité qu'il n'y ait aucune série noire sur chacune de ces 16 périodes vaut $(998/1000)^{16}$, soit environ 97 chances sur 100.

Quelle est la probabilité qu'au moins 5 avions se crashent en 22 jours **sur une année entière ?** On ne peut plus calculer simplement une telle probabilité, ceci à cause des chevauchements des périodes de 22 jours sur une année (les variables aléatoires ne sont plus indépendantes). Un calcul montre que

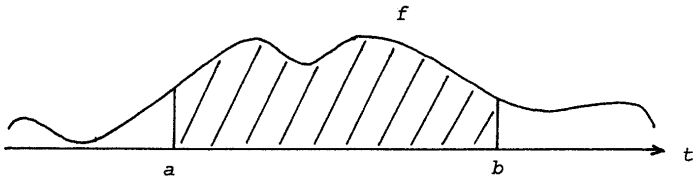
$$P(\text{ au moins 5 avions se crashent en 22 jours sur une année}) \approx 0.11,$$

soit plus d'une chance sur 10 ! !

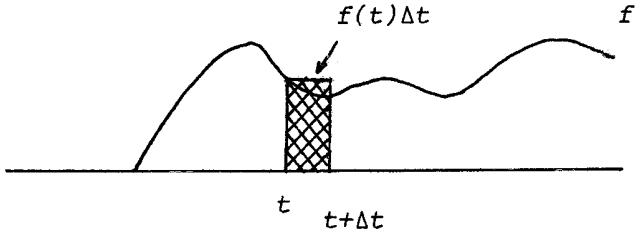
4.2 Les variables aléatoires réelles avec densité

On dit qu'une variable aléatoire X à valeurs dans $\mathbb{R}$ possède une densité f si pour $a \leq b$ quelconques on a

$$P(a \leq X \leq b) = \int_a^b f(t)dt.$$



La densité f est une fonction non-négative $\Big(f(t) \geq 0\Big)$ qui doit vérifier la condition $\int_{-\infty}^{+\infty} f(t)dt = 1$. Remarquons que $P(a \leq X \leq b)$ est donnée par l'aire de la surface indiquée ci-dessus. On constate ainsi que pour des Δt très petits, $f(t)\Delta t$ fournit une approximation de la probabilité pour que X prenne ses valeurs entre t et $t + \Delta t$.



4.2.1 Variable normale ou gaussienne

La famille la plus célèbre est celle des variables aléatoires dites normales ou gaussiennes et dont les densités sont de la forme

$$t \in \mathbb{R} \longmapsto f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$$

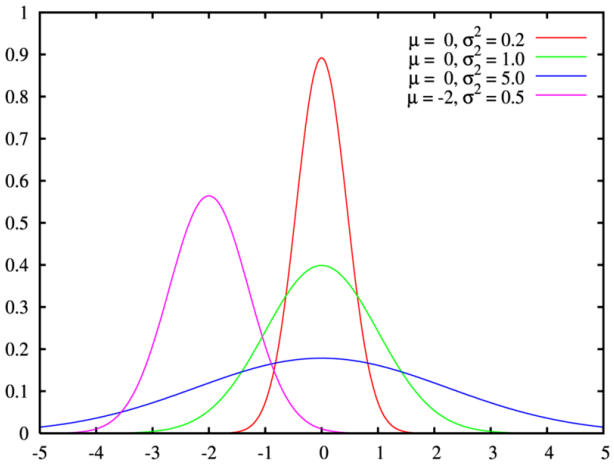


FIGURE 16 – Graphes de la densité normale pour différentes valeurs de μ et σ

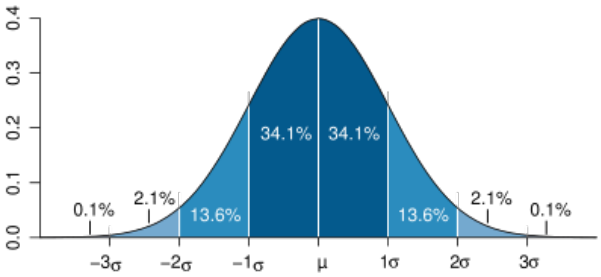
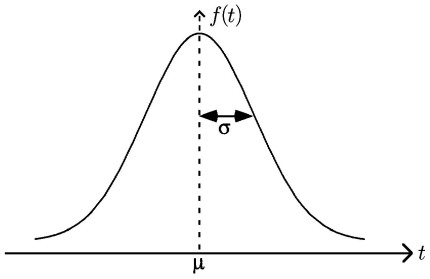


FIGURE 17 – Probabilités associées à certains secteurs caractéristiques

avec $\mu \in \mathbb{R}$ et $\sigma > 0$. Une variable aléatoire X dont la densité est f sera dite normale (ou gaussienne) de paramètres μ et σ et nous noterons $X = N(\mu, \sigma)$. Un cas particulier important est donné par $\mu = 0$ et $\sigma = 1$. Une variable aléatoire $U = N(0, 1)$ est dite normale standard (ou standardisée) ou encore centrée réduite.

L'allure de la densité ci-cessus est une courbe en forme de cloche :



Les variables aléatoires normales sont sorties des travaux de Gauss consacrés à la théorie des erreurs. Elles jouent un rôle fondamental notamment à cause des propriétés asymptotiques décrites dans le théorème limite central (voir après). Il est intéressant de remarquer que l'intégrale

$$\int_a^t e^{-s^2}\, ds$$

ne se laisse pas exprimer de façon simple à l'aide des fonctions dites élémentaires (théorème difficile). Par conséquent, certains calculs faisant intervenir les densités de variables aléatoires normales devront être effectués

numériquement ou à l'aide d'une table. Nous verrons en fait qu'il suffira de disposer d'une table pour la fonction de répartition d'une variable aléatoire $N(0,1)$, c'est-à-dire :

$$t \in \mathbb{R} \longmapsto F_{N(0,1)}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{s^2}{2}}\, ds.$$

Nous utiliserons dorénavant la notation

$$\Phi(t) = F_{N(0,1)}(t).$$

4.2.2 Variable exponentielle

Une famile importante de variables aléatoires sont celles dont la densité est :

$$f(t) = \begin{cases} 0 & \text{si } t < 0 \\ \\ \lambda e^{-\lambda t} & \text{si } t \geq 0 \end{cases} \qquad \text{où } \lambda > 0.$$

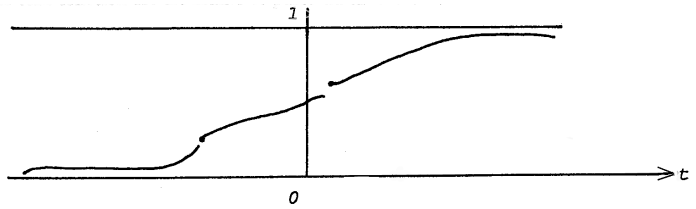
On vérifie facilement que $\int_{-\infty}^{+\infty} \lambda e^{\lambda t}\ dt = 1$. Une variable aléatoire X admettant cette densité est appelée exponentielle de paramètre λ. Nous noterons alors $X = E(\lambda)$. Ce type de variables aléatoires intervient dans la modélisation du temps de vie d'un système.

4.3 Fonction de répartition d'une variable aléatoire :

A chaque variable aléatoire X on peut associer sa fonction de répartition définie par :

$$t \in \mathbb{R} \longmapsto F_X(t) = P(X \leq t).$$

Une telle fonction est croissante et passe du niveau 0 au niveau 1.



Définition 4.1 *On dit que deux variables aléatoires X_1 et X_2 sont* identiquement distribuées *(ou ont même répartition) si elles ont la même fonction de répartition ($F_{X_1} \equiv F_{X_2}$).*

On dit que deux variables aléatoires X_1 et X_2 sont indépendantes *si pour t_1 et t_2 quelconques on a*

$$P(X_1 \leq t_1, X_2 \leq t_2) = P(X_1 \leq t_1) P(X_2 \leq t_2).$$

Cette propriété signifie que les événements associés à X_1 et ceux associés à X_2 sont indépendants. La généralisation de la définition à une famille quelconque de variables aléatoires se fait de façon identique à celle des événements indépendants. Le hasard peut donc être responsable de la série de crashs.

4.4 Les notions d'espérance et de variance d'une variable aléatoire :

Considérons les trois jeux dont les gains respectifs X_1, X_2 et X_3 sont donnés par :

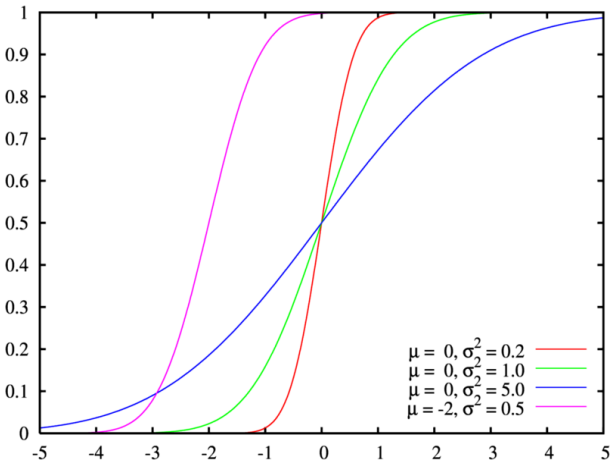


FIGURE 18 – Graphes de fonctions de répartition normales pour plusieurs valeurs de μ et σ

$$\text{Jeu 1} : X_1 = \begin{cases} \text{gagne} & 10 \text{ francs avec probabilité } \frac{1}{2} \\ \text{gagne} & 0 \text{ franc avec probabilité } \frac{1}{2} \end{cases}$$

$$\text{Jeu 2} : X_2 = \begin{cases} \text{gagne} & 20 \text{ francs avec probabilité } \frac{1}{4} \\ \text{perd} & 1 \text{ franc avec probabilité } \frac{3}{4} \end{cases}$$

$$\text{Jeu 3} : X_3 = \begin{cases} \text{gagne} & 20 \text{ francs avec probabilité } \frac{1}{5} \\ \text{gagne} & 0 \text{ franc avec probabilité } \frac{4}{5} \end{cases}$$

Quel est le jeu le plus avantageux ? Un critère pour les comparer est le gain espéré définit comme suit :

$$\begin{aligned} E(X_1) &= 10 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{2} = 5 \\ E(X_2) &= 20 \cdot \tfrac{1}{4} + (-1)\tfrac{3}{4} = \tfrac{17}{4} \\ E(X_3) &= 20 \cdot \tfrac{1}{5} + 0 \cdot \tfrac{4}{5} = 4 \end{aligned}$$

Selon ce critère le premier jeu est le plus avantageux des trois. La notion d'espérance d'une variable aléatoire est une généralisation de l'idée qui précède.

Définition 4.2 *Si X est une variable aléatoire à valeurs dans* $\mathbf{N}$*, son espérance notée $E(X)$ est définie par*

$$E(X) = \sum_{k=0}^{\infty} k\, P(X = k).$$

Si X est une variable aléatoire réelle avec densité f, son espérance notée $E(X)$ est définie par

$$E(X) = \int_{-\infty}^{+\infty} t\, f(t) dt.$$

L'espérance d'une variable aléatoire est donc sa moyenne (théorique) ou encore la position de son "centre de gravité" si l'on interprète les probabilités comme des masses pesantes. En ce sens l'espérance est un paramètre de position de la répartition de masse = probabilité.

4.4.1 Propriétés de l'espérance :

On peut démontrer que l'espérance possède les propriétés suivantes :

- E (constante) = constante
- $\alpha \in \mathbb{R}$, X v.a. : $E(\alpha X) = \alpha E(X)$
- X, Y v.a. : $E(X+Y) = E(X) + E(Y)$
- $X \geq 0$ v.a. : $E(X) \geq 0$
- X, Y v.a., $X \leq Y$: $E(X) \leq E(Y)$
- si g est une fonction alors : $E\Big(g(X)\Big) = \sum_{k=0}^{\infty} g(k)P(X=k)$ ou $E\Big(g(X)\Big) = \int_{-\infty}^{+\infty} g(t)f(t)dt$ suivant que X prend ses valeurs dans N ou dans $\mathbb{R}$ avec densité f.

Exemples :

- $X = \text{Ber}(p)$: $E(X) = 1 \cdot p + 0 \cdot q = p$
- $X = \text{Bin}(n,p)$: nous avons vu que X est alors une somme $X = X_1 + X_2 + \ldots + X_n$ de n variables aléatoires (indépendantes) Ber(p). On conclut que

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \ldots + E(X_n) \\ &= p + p \ldots + p \\ &= np. \end{aligned}$$

- $X = G(p)$: le calcul montre que $E(X) = \sum_{k=1}^{\infty} k\, q^{k-1} p = \frac{1}{p}$.
- $X = \text{Poi}(\lambda)$: le calcul montre que $E(X) = \sum_{k=0}^{\infty} k\, e^{-\lambda} \frac{\lambda^k}{k!} = \lambda$
- $X = N(\mu, \sigma)$: le calcul montre que $E(X) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} t\, e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt = \mu$
- $X = E(\lambda)$: le calcul montre que $E(X) = \int_{0}^{+\infty} t\, \lambda\, e^{-\lambda t} dt = \frac{1}{\lambda}$.

L'espérance est un paramètre de position qui ne nous indique pas si la probabilité est peu ou beaucoup dispersée autour de son centre de gravité. Pour mesurer cette dispersion on peut faire appel à la variance définie par :

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

Il s'agit de l'écart quadratique moyen autour de l'espérance. Il est clair que

$$\begin{aligned} \text{Var}(X) &= E\bigg(X^2 + \Big(E(X)\Big)^2 - 2XE(X)\bigg) \\ &= E(X^2) + \Big(E(X)\Big)^2 - 2\Big(E(X)\Big)^2 \\ &= E(X^2) - \Big(E(X)\Big)^2. \end{aligned}$$

Pour le calcul de $E(X^2)$ nous avons, suivant le type de variables aléatoires :

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X=k) \quad \text{ou} \quad E(X^2) = \int_{-\infty}^{+\infty} t^2 f(t) dt.$$

Exemples :

$$X = \text{Ber}(p) \qquad E(X^2) = 1^2 \cdot p + 0^2 \cdot q = p, \qquad E(X) = p$$

$$\text{Var}(X) = E(X^2) - \Big(E(X)\Big)^2 = p - p^2 = p(1-p) = pq.$$

Par calcul direct on obtient les résultats suivants :

$$\begin{array}{ll} X = \text{Bin}(n,p) & \text{Var}(X) = npq \\ X = \text{Poi}(\lambda) & \text{Var}(X) = \lambda \\ X = N(\mu,\sigma) & \text{Var}(X) = \sigma^2\,. \end{array}$$

4.4.2 Propriétés de la variance :

- X v.a. $\text{Var}(X) \geq 0$ et $\text{Var}(X) = 0 \Longleftrightarrow X =$ constante
- X v.a. $\alpha \in \mathbb{R}$, $\text{Var}(X + \alpha) = \text{Var}(X)$
- X v.a. $\alpha \in \mathbb{R}$, $\text{Var}(\alpha X) = \alpha^2 \text{Var}(X)$.

Variance d'une somme de variables aléatoires :

Si X_1 et X_2 sont deux variables aléatoires, que peut-on dire de $\text{Var}(X_1 + X_2)$? En général rien sans hypothèse supplémentaire sur X_1 et X_2.

Définition 4.3 *On dit que deux variables aléatoires X_1 et X_2 sont* non-corrélées *si $E(X_1X_2) = E(X_1)E(X_2)$.*

Si X_1 et X_2 sont non-corrélées, alors $\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)$ (exercice). On peut démontrer que si X_1 et X_2 sont indépendantes, alors elles sont non-corrélées l'inverse étant faux. Ainsi la variance d'une somme de variables aléatoires indépendantes est égale à la somme des variances. On en déduit facilement que si $X = \text{Bin}(n,p)$, alors $\text{Var}(X) = npq$ puisque X est alors somme de n variables aléatoires indépendantes $\text{Ber}(p)$ dont la variance vaut pq.

Il est d'usage de noter σ^2 la variance d'une variable aléatoire. L'unité de $\text{Var}(X)$ est le carré de celle de X. Pour cette raison on préférera certaines fois travailler avec $\sigma = \sqrt{\text{Var}(X)}$.

Définition 4.4 *On appelle écart-type d'une variable aléatoire X le nombre*

$$\sigma = \sqrt{Var(X)}.$$

4.4.3 Utilisation d'une table de loi normale :

Si $X = N(\mu,\sigma)$, alors $\frac{X-\mu}{\sigma}$ est une variable aléatoire d'espérance 0 et d'écart-type 1. On peut de plus vérifier que $\frac{X-\mu}{\sigma}$ est encore une variable aléatoire normale donc $\frac{X-\mu}{\sigma} = N(0,1)$. Cette propriété a d'importantes conséquences pratiques. Supposons en effet que l'on désire calculer

$$P(a \leq X \leq b)$$

où $X = N(\mu,\sigma)$.

$$\begin{aligned} P(a \leq X \leq b) &= P(a - \mu \leq X - \mu \leq b - \mu) \\ &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \end{aligned}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{\frac{a-\mu}{\sigma}}^{\frac{b-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt$$
$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

puisque $\frac{X-\mu}{\sigma} = N(0,1)$. Il suffit ainsi de disposer d'une table donnant la fonction de répartition $\Phi(t)$ d'une variable aléatoire $N(0,1)$. On appelle standardisée une variable aléatoire dont l'espérance est 0 et l'écart-type 1.

Exemple : Calculer $P(29 \leq X \leq 32)$ où $X = N(30,2)$.

$$\begin{aligned} P(29 \leq X \leq 32) &= P\left(\frac{29-30}{2} \leq \frac{X-30}{2} \leq \frac{32-30}{2}\right) \\ &= P\left(-\frac{1}{2} \leq \frac{X-30}{2} \leq 1\right) \\ &= \Phi(1) - \Phi(-\frac{1}{2}) \overset{\text{table}}{=} 0.8413 - 0.3085 = 0.5328. \end{aligned}$$

4.5 Modèle des observations d'une variable aléatoire :

Nous nous intéressons à la répartition du poids des personnes dans la population d'une région donnée (ville, canton, pays, ...). Désignons par X le poids d'une personne choisie au hasard dans cette population. X est donc une variable aléatoire et nous pouvons par exemple nous intéresser à sa moyenne, c'est-à-dire son espérance $\mu = E(X)$. Si la population ne peut pas être observée dans son intégralité (ce qui est le cas dans la pratique), alors μ est une grandeur qui ne sera jamais connue exactement. On peut cependant essayer de l'estimer sur la base d'observations de X.

Considérons n observations successives de X, c'est-à-dire n personnes choisies successivement au hasard dont on mesure le poids. Désignons par $X_1, X_2, \ldots, X_n$ les résultats obtenus i.e. X_i = poids de la $i^{\text{ième}}$ personne. Dans un modèle des observations, $X_1, X_2, \ldots, X_n$ sont des variables aléatoires ; certains auteurs désignent les valeurs effectivement obtenues par des lettres minuscules $x_1, x_2, \ldots, x_n$.

Quelles sont les propriétés de $X_1, X_2, \ldots, X_n$? Si l'on s'arrange pour éviter des influences mutuelles entre les observations (tirage avec remise ou taille de la population très grande), alors on peut supposer l'indépendance des variables aléatoires $X_1, X_2, \ldots, X_n$. De plus, chacune d'elle représentant une observation de X, elles ont même loi que X. Ainsi

$$X_1, X_2, \ldots, X_n \quad \text{sont i.i.d. comme } X$$

où i.i.d. = indépendantes et identiquement distribuées. Une telle famille est appelée *n–échantillon issu de X*. Il est important de noter que les X_i ayant même loi que X, elles ont la même espérance μ.

Comment estimer μ avec $X_1, X_2, \ldots, X_n$? La réponse usuelle à cette question est : à l'aide de la moyenne arithmétique

$$\frac{X_1 + X_2 + \ldots + X_2}{n}.$$

Il est légitime de se demander pourquoi et l'une des réponses possibles est fournie par le théorème appelé "loi des grands nombres" et qui s'énonce ainsi :

5 Théorèmes limites

5.1 Théorème (Loi des grands nombres)

Si $X_1, X_2, \ldots, X_n, \ldots$ est une suite infinie de variables aléatoires i.i.d. comme la variable aléatoire X, alors

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = E(X)\right) = 1.$$

Dans notre contexte nous interprétons les variables aléatoires comme des observations indépendantes de X. Ainsi

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

est la moyenne des n premières observations qu'on l'appellera aussi moyenne empirique. La loi des grands nombres nous assure alors que la suite des moyennes empiriques converge vers l'espérance $\mu = E(X)$ (moyenne théorique) avec probabilité égale à 1 lorsque $n \to \infty$. Il est donc pertinent d'estimer $\mu = E(X)$ avec $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. La question de l'estimation de σ^2 (ou σ) se pose de la même façon. Par analogie, en remplaçant $E(\ \ )$ par moyenne arithmétique dans $\mathrm{Var}(X) = E\Big((X - E(X))^2\Big)$, on obtient :

$$\frac{1}{n}\sum_{i=1}^{n}\Big(X_i - \overline{X}\Big)^2.$$

Pour des raisons théoriques (point peu important si n est grand), les statisticiens préfèrent l'expression

$$S^2 := \frac{1}{n-1}\sum_{i=1}^{n}\Big(X_i - \overline{X}\Big)^2,$$

car $E(S^2) = \sigma^2$, alors que la première expression ne possède pas cette propriété.

On appelle *variance empirique* la grandeur S^2 et

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\Big(X_i - \overline{X}\Big)^2},$$

l'*écart-type empirique* du n–échantillon $X_1, \ldots, X_n$ issu de X. A l'aide de la loi des grands nombres, on peut démontrer que S^2 (resp. S) converge avec probabilité 1 vers σ^2 (resp. σ) lorsque $n \to \infty$.

Nous avons dégagé trois fonctions des observations, à savoir $\overline{X}$, S^2 et S qui sont utilisées pour estimer respectivement μ, σ^2 et σ. De telles fonctions sont appelées des *estimateurs* pour les grandeurs inconnues correspondantes.

Voici le second grand théorème asymptotique de la théorie des probabilités.

5.2 Théorème limite-central

Soient $X_1, X_2, \ldots, X_n, \ldots$ une suite de variables aléatoires i.i.d. comme la variable aléatoire X, $\mu = E(X)$, $\sigma^2 = \mathrm{Var}(X)$ et $S_n = \sum_{i=1}^{n} X_i$.
Alors, pour tout nombre réel t, on a

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \le t\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-\frac{s^2}{2}}\, ds.$$

Remarque : en transformant S_n en $\frac{S_n - n\mu}{\sigma\sqrt{n}}$, la nouvelle variable aléatoire a une espérance nulle et un écart-type égal à 1. En effet :

$$\begin{aligned}
E(S_n) &= E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i) = n\mu \quad \text{car} \quad E(X_i) = E(X) = \mu, \\
E\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) &= \frac{1}{\sigma\sqrt{n}} E(S_n - n\mu) = \frac{1}{\sigma\sqrt{n}}\Big(E(S_n) - n\mu\Big) = \frac{1}{\sigma\sqrt{n}}(n\mu - n\mu) = 0 \\
\mathrm{Var}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}}\right) &= \frac{1}{\sigma^2 n}\mathrm{Var}(S_n - n\mu) = \frac{1}{\sigma^2 n}\mathrm{Var}(S_n) \\
&= \frac{1}{\sigma^2 n}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{n\sigma^2}{n\sigma^2} = 1
\end{aligned}$$

car les variables aléatoires sont indépendantes et $\mathrm{Var}(X_i) = \mathrm{Var}(X) = \sigma^2$. Ainsi l'écart-type de $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ est égal $\sqrt{1} = 1$. Le théorème limite central peut être formulé en terme de fonctions de répartition. Rappelons que si Y est une variable aléatoire alors sa fonction de répartition est :

$$F_Y(t) = P(Y \leq t).$$

Ainsi le théorème limite central affirme que, sous les hypothèses précédentes,

$$F_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t) \longmapsto F_{n(0,1)}(t) = \Phi(t)$$

pour tout nombre réel t. Ce résultat suggère que pour n suffisamment grand, $F_{\frac{S_n - n\mu}{\sigma\sqrt{n}}}(t)$ peut être approché par $\Phi(t)$.

5.3 Approximation d'une loi binômiale par une loi normale

Nous avons que si $S_n = \mathrm{Bin}(n, p)$, alors S_n est somme de n variables aléatoires $X_1, X_2, \ldots, X_n$ indépendantes $\mathrm{Ber}(p)$ i.e. $S_n = \sum_{i=1}^{n} X_i$. Le théorème limite central affirme alors que

$$P\left(\frac{S_n - np}{\sqrt{npq}} \leq t\right) \underset{n\to\infty}{\longrightarrow} \Phi(t)$$

car $E(X_i) = p$ et $\mathrm{Var}(X_i) = pq$. Il est possible de montrer que si $npq > 9$, alors $P\left(\frac{S_n - np}{\sqrt{npq}} \leq t\right)$ peut être correctement approché par $\Phi(t)$.

Application : On considère 1000 jets indépendants d'une pièce de monnaie dont la probabilité de pile est $p = \frac{1}{4}$. Soit S_{1000} le nombre de réalisations de pile dans les 1000 jets. Calculer $P(230 \leq S_{1000} \leq 270)$ à l'aide d'une approximation normale ($npq = 1000 \cdot \frac{1}{4} \cdot \frac{3}{4} = 187.5 > 9$).

$$\begin{aligned}
P(230 \leq S_{1000} \leq 270) &= P\left(\frac{230 - 1000 \cdot \frac{1}{4}}{\sqrt{1000 \cdot \frac{1}{4} \cdot \frac{3}{4}}} \leq \frac{S_{1000} - 1000 \cdot \frac{1}{4}}{\sqrt{1000 \cdot \frac{1}{4} \cdot \frac{3}{4}}} \leq \frac{270 - 1000 \cdot \frac{1}{4}}{\sqrt{1000 \cdot \frac{1}{4} \cdot \frac{3}{4}}}\right) \\
&\cong P(-1.460 \leq N(0,1) \leq 1.460) \\
&= \Phi(1.460) - \Phi(-1.460) = 0.855
\end{aligned}$$

car $\Phi(1.460) = 0.9279$ et $\Phi(-1.460) = 1 - \Phi(1.460)$.

5.4 Somme de variables aléatoires normales indépendantes

Le calcul montre que la somme de variables aléatoires normales indépendantes est encore une variable aléatoire normale. Si $X_1 = N(\mu_1, \sigma_1)$ et $X_2 = N(\mu_2, \sigma_2)$ sont indépendantes alors $X_1 + X_2 = N(\mu, \sigma)$. Que valent μ et σ ?

$$\begin{array}{rccl} & \mu & = & E(X_1 + X_2) = E(X_1) + E(X_2) = \mu_1 + \mu_2 \\ & \sigma^2 & = & \mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}X_2) = \sigma_1^2 + \sigma_2^2 \\ \text{donc} & \sigma & = & \sqrt{\sigma_1^2 + \sigma_2^2} \quad \text{et} \quad X_1 + X_2 = N(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}). \end{array}$$

Considérons un n–échantillon $X_1, X_2, \ldots, X_n$ issu de $X = N(\mu, \sigma)$. Alors $S_n = \sum_{i=1}^{n} X_i = N(n\mu, \sigma\sqrt{n})$ en vertu du calcul précédent et donc $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} = N(0, 1)$.

5.5 Intervalle de confiance pour l'espérance

Nous estimons l'espérance μ d'une variable aléatoire $X = N(\mu, \sigma)$ à l'aide de la moyenne empirique $\overline{X} = \dfrac{S_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ d'un n–échantillon $X_1, X_2, \ldots, X_n$ issu de X. Que peut-on dire de l'erreur commise ? Supposons d'abord que σ est connu. Le résultat ci-dessus nous affirme que $\dfrac{S_n - n\mu}{\sigma\sqrt{n}} = N(0, 1)$. En divisant numérateur et dénominateur par n, on obtient :

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\frac{S_n}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = N(0, 1).$$

Puisque $\Phi(-1.96) = 2.5\,\%$, nous avons

$$\begin{array}{l} P\left(-1.96 \leq \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) = 95\,\% \\ = P\left(-\frac{1.96\sigma}{\sqrt{n}} \leq \overline{X} - \mu \leq \frac{1.96\sigma}{\sqrt{n}}\right) \\ = P\left(\overline{X} - \frac{1.96\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{1.96\sigma}{\sqrt{n}}\right). \end{array}$$

Par conséquent, la probabilité pour que la valeur cherchée μ soit dans l'intervalle aléatoire

$$\left[\overline{X} - \frac{1.96\sigma}{\sqrt{n}}\,,\ \overline{X} + \frac{1.96\sigma}{\sqrt{n}}\right]$$

centré en $\overline{X}$ vaut 95 %. Ceci signifie qu'en répétant cette construction, la grandeur inconnue μ appartiendra à un tel intervalle environ 95 fois sur 100 mais nous ne savons bien sûr pas lesquels.

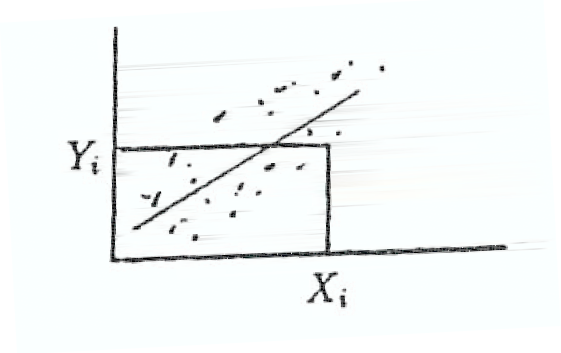
L'intervalle aléatoire ci-dessus est appelé intervalle de confiance de μ au coefficient de risque 5 %. Un aspect important réside dans le fait que la longueur de l'intervalle décroît comme $\dfrac{\text{constante}}{\sqrt{n}}$ avec la taille n de l'échantillon.

Par contre, si σ est inconnu, on l'estimera à l'aide de l'écart-type empirique $S = \sqrt{\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$. L'expression $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ est donc remplacée par $\dfrac{S_n - n\mu}{S\sqrt{n}} = \dfrac{\frac{S_n}{n} - \mu}{\frac{S}{\sqrt{n}}} = \dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}}$. Le calcul montre que cette dernière variable aléatoire est de type Student à $n - 1$ degrés de liberté. Il suffit alors d'utiliser une table de Student à $n - 1$ degrés de liberté à la place d'une table de loi normale. De toute façon, pour $n \geq 30$, $\dfrac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \cong N(0, 1)$.

Dans le cas où les observations ne sont pas issues d'une variable aléatoire normale, le théorème limite central nous assure que l'on peut asymptotiquement s'y ramener lorsque la taille de l'échantillon est très grande.

5.6 Droite de régression, coefficient de corrélation

Un problème important est la discussion d'une éventuelle relation entre deux variables aléatoires X et Y. Une première approche de la question est la recherche d'une relation affine entre X et Y, c'est-à-dire une relation de la forme $Y = aX + b$. Supposons que l'on ait observé n fois le couple (X, Y) et que les valeurs obtenues soient



Les couples précédents, représentés dans le plan (X, Y), ne seront en général pas alignés sur une droite. Nous allons donc écrire la relation en corrigeant avec une erreur stochastique ε_i :

$$Y_i = aX_i + b + \varepsilon_i\,, \quad 1 \leq i \leq n.$$

Nous proposons de chercher a et b, c'est-à-dire une droite de pente a et d'ordonnée à l'origine b, qui "approche au mieux" le nuage de points définis par les couples observés. Il faut évidemment préciser dans quel sens l'approximation est mesurée. L'usage veut que l'on travaille avec l'erreur quadratique totale définie par :

$$E^2(a,b) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (aX_i + b - Y_i)^2.$$

Pour trouver les valeurs de a et b qui minimisent $E^2(a,b)$ nous imposons

$$\frac{\partial}{\partial a} E^2(a,b) = 0, \quad \frac{\partial}{\partial b} E^2(a,b) = 0.$$

La seconde condition fournit

$$\sum_{i=1}^{n} 2(aX_i + b - Y_i) = 0$$

$$\sum_{i=1}^{n} Y_i = a \sum_{i=1}^{n} X_i + nb$$

$$\tfrac{1}{n} \sum_{i=1}^{n} Y_i = a \tfrac{1}{n} \sum_{i=1}^{n} X_i + b$$

$$\overline{Y} = a\overline{X} + b.$$

Nous constatons donc que la droite optimale passe par le point $(\overline{X}, \overline{Y})$ où $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ et $\overline{Y} = \dfrac{1}{n}\sum_{i=1}^{n} Y_i$. Cette condition implique $b_{\text{opt}} = \overline{Y} - a\overline{X}$ et en remplaçant dans $E^2(a,b)$, on trouve

$$\begin{aligned} E^2\Big(a, b_{\text{opt}}(a)\Big) &= \sum_{i=1}^{n} (aX_i - a\overline{X} + \overline{Y} - Y_i)^2 \\ &= \sum_{i=1}^{n} \Big(a(X_i - \overline{X}) - (Y_i - \overline{Y})\Big)^2. \end{aligned}$$

En dérivant par rapport à a nous obtenons

$$\sum_{i=1}^{n} 2\Big(a(X_i-\overline{X})-(Y_i-\overline{Y})\Big)(X_i-\overline{X})=0,$$

et par conséquent

$$a_{\text{opt}}=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^{n}(X_i-\overline{X})^2}\,.$$

Il est judicieux de diviser numérateur et dénominateur par $n-1$ pour faire apparaître la variance empirique S_x^2 de $X_1, X_2, \ldots, X_n$:

$$\begin{aligned} a_{\text{opt}} &= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} \\ &= \frac{C_{xy}}{S_x^2} \end{aligned}$$

où $C_{xy}=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})$. Cette dernière expression, est appelée *covariance empirique* de l'échantillon $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ et estime la grandeur théorique

$$\text{Cov}(X,Y)=E\bigg(\Big(X-E(X)\Big)\Big(Y-E(X)\Big)\bigg)$$

appelée covariance du couple (X,Y). Un calcul direct montre que

$$C_{xy}=\frac{1}{n-1}\left(\sum_{i=1}^{n}X_iY_i-n\overline{XY}\right)\quad\text{et}\quad\text{Cov}(X,Y)=E(XY)-E(X)E(Y).$$

Ces égalités permettent de faciliter les calculs et la dernière montre que $\text{Cov}(X,Y)=0$ si et seulement si X et Y sont non-corrélées.

L'équation de la droite optimale (i.e. celle qui minimise l'erreur quadratique) est donc

$$y-\overline{Y}=\frac{C_{xy}}{S_x^2}(x-\overline{X}).$$

Nous pouvons maintenant calculer l'erreur minimale commise si l'on "remplace" le nuage de points observés par la droite optimale :

$$\begin{aligned} E^2_{\text{min}} &= E^2(a_{\text{opt}},b_{\text{opt}})=\sum_{i=1}^{n}\left(\frac{C_{xy}}{S_x^2}(X_i-\overline{X})-(Y_i-\overline{Y})\right)^2 \\ &= \frac{C_{xy}^2}{S_x^4}\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{i=1}^{n}(Y_i-\overline{Y})^2-2\frac{C_{xy}}{S_x^2}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y}). \end{aligned}$$

En posant $S_y^2=\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(Y_i-\overline{Y})^2$ (variance empirique de $Y_1,Y_2,\ldots,Y_n$) on obtient :

$$\begin{aligned} E^2_{\text{min}} &= (n-1)S_Y^2\left(\frac{C_{xy}^2}{S_x^4S_y^2}S_x^2+1-2\frac{C_{xy}^2}{S_x^2S_y^2}\right) \\ &= (n-1)S_y^2\left(1-\left(\frac{C_{xy}}{S_xS_y}\right)^2\right). \end{aligned}$$

Sachant que $E_{\min} \geq 0$, nous en déduisons que $1 - \left(\frac{C_{xy}}{S_x S_y}\right)^2 \geq 0$ et donc $-1 \leq \frac{C_{xy}}{S_x S_y} \leq 1$. Cette dernière grandeur nous renseigne sur la valeur de l'erreur minimale lorsque l'on essaie d'expliquer les points observés par une droite. La droite optimale obtenue précédemment est appelée *droite de régression*.

Définition 5.1 *Le nombre* $r_{xy} = \frac{C_{xy}}{S_x S_y}$ *est appelé* coefficient de corrélation empirique *de l'échantillon* $(X_1, Y_1), \ldots, (X_n, Y_n)$.

Le nombre r_{xy} *permet d'estimer le* coefficient de corrélation *entre* X *et* Y *défini par :*

$$\rho = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

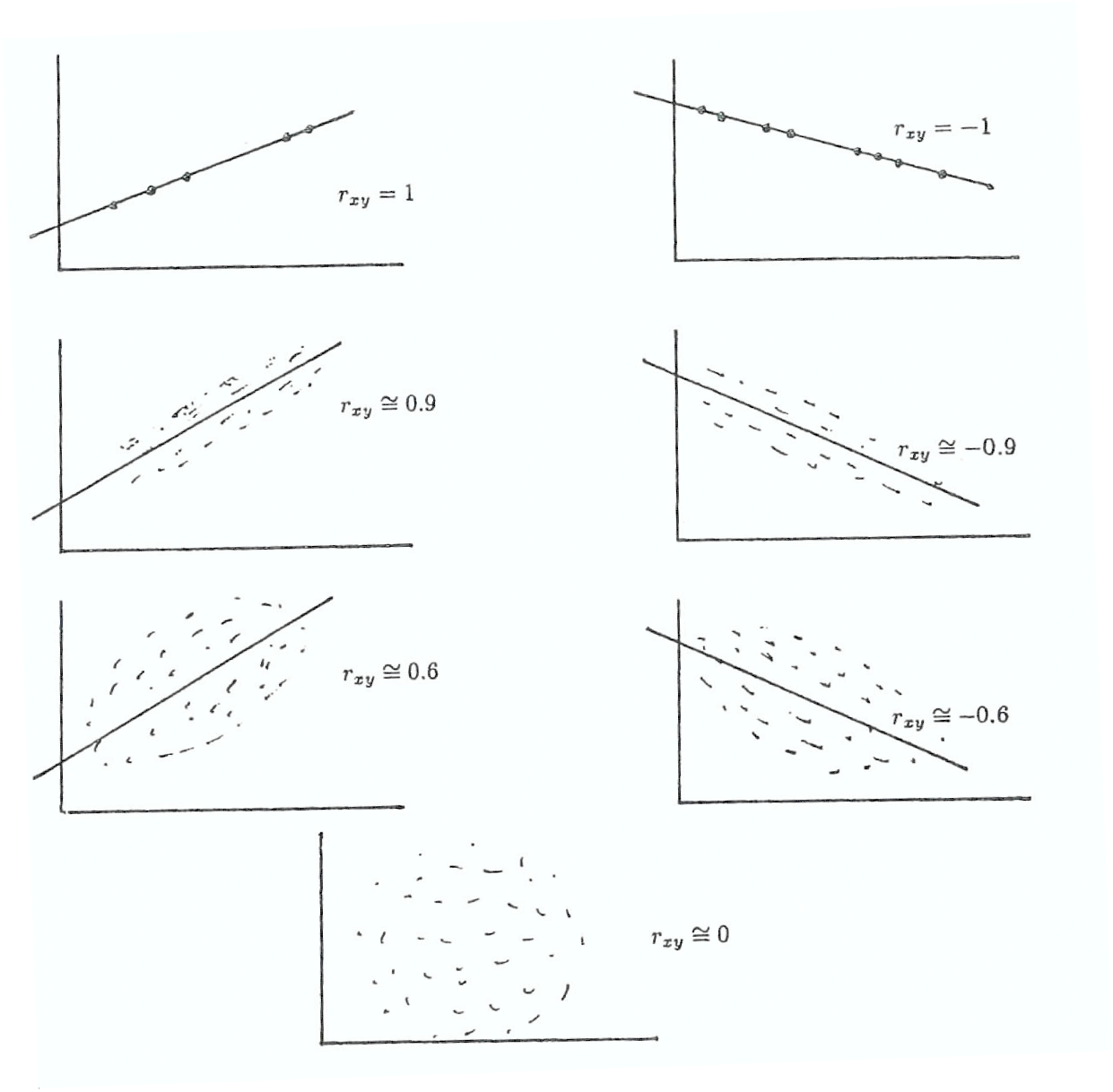
où σ_x *et* σ_y *sont respectivement les écarts-type de* X *et* Y.

Nous constatons que $E^2_{\min} = 0$ équivaut à $r^2_{xy} = 1$ et donc à $r_{xy} = \pm 1$. Dans ces deux cas, tous les points observés sont alignés sur la droite de régression et $+1$ ou -1 nous indiquent le signe de sa pente. De façon générale, plus $|r_{xy}|$ est proche 1, plus $E^2_{\min}$ est petite et inversement, plus r_{xy} est proche de 0, plus $E^2_{\min}$ est grande. $E^2_{\min}$ est maximale pour $r_{xy} = 0$ et la seule conclusion que l'on peut en tirer est qu'une droite ne représente pas les observations de manière satisfaisante. Il ne faut toutefois pas en conclure que X et Y ne sont liées par aucune relation. On peut facilement donner des exemples de variables aléatoires X et Y qui sont non-corrélées et telles que $X^2 + Y^2 = R^2$ (voir exercices). Inversement, r_{xy} proche de 1 ne signifie pas un lien causal entre X et Y. Voici l'exemple célèbre tiré de "Ornithologischen Monatsberichten 44, Nr 2 (1936) et 48, Nr 1 (1940)" qui traite l'évolution de la population humaine et du nombre de couples de cigognes de ville d'Oldenburg entre 1930 et 1936 :

	1930	1931	1932	1933	1934	1935	1936
# couples de cigognes	132	142	166	188	240	250	252
Habitants	55'400	55'400	65'000	67'700	69'800	72'300	76'000

Exercice : Faire une représentation graphique des points et déterminer la droite de régression pour $X = \#$ couples de cigognes et $Y = \#$ habitants. Le coefficient de corrélation $r_{xy} = 0,945$ est très proche de 1 mais il serait dangereux de conclure que l'accroissement de la population humaine provient de la présence des cigognes !

Voici quelques exemples de nuages de points avec les coefficients de corrélation (approximatifs) correspondants :



6 Introduction aux tests statistiques

Beaucoup de situations pratiques conduisent à opposer des hypothèses invérifiables de façon directe. Par exemple, un nouveau médicament est-il meilleur que l'ancien ? Une nouvelle méthode d'enseignement est-elle supérieure à l'ancienne ?

Il faut comparer des observations menées sur des malades dans le premier cas et sur des étudiants dans le second. Cependant, les résultats dépendront aussi de fluctuations d'échantillonage car aussi bien un médicament qu'une méthode d'enseignement agissent de façons différentes sur des sujets différents. En effet, certains patients sont plus ou moins réceptifs que d'autres à un médicament et il en va de même avec les étudiants et une méthode d'enseignement.

La littérature scientifique propose un grand nombre de tests pour traiter des situations différentes. Notre but vise ici la compréhension d'un test typique car le mécanisme de base sera le même pour tous.

Un test est constitué de plusieurs éléments à savoir :

- une hypothèse et une alternative
- une fonction des observations
- un niveau
- un domaine de rejet.

Il est plus facile de refuter une hypothèse que de la démontrer. Il est d'usage dans notre contexte de formuler l'hypothèse qui nous intéresse comme contre-hypothèse susceptible d'être rejetée. Dans les situations précédentes, l'hypothèse sera le nouveau médicament (respectivement la nouvelle méthode) et l'ancien (resp. l'ancienne méthode) sont équivalents (resp. équivalentes). Il est d'usage de la qualifier d'hypothèse nulle notée H_0 qui sera opposée à une alternative H_1 qui sera ici le nouveau médicament (nouvelle méthode) est supérieur(e) à l'ancien(ne).

Afin de simplifier l'exposé, nous discutons d'abord le problème de la "chute d'une tartine". Une affirmation fréquente prétend qu'une tartine a tendance à tomber du mauvais côté, c'est-à-dire du côté confiture. Désignons par p la probabilité pour que, lors d'une chute, une tartine donnée tombe du côté confiture. Pour hypothèse nulle nous choisissons :

$$H_0: \quad p = 0.50\,.$$

Dans un but didactique nous supposons que p peut admettre seulement les valeurs 0.50 et 0.55. L'alternative sera donc

$$H_1: \quad p = 0.55\,.$$

En fait H_1 peut être $p \neq \frac{1}{2}$ ou $p > \frac{1}{2}$.
Supposons que n chutes de tartines aient été observées et que m fois celle-ci soit tombée du côté confiture. Pouvons-nous trancher entre H_0 et H_1 sur la base de ces observations ?

Les deux décisions possibles dans un tel test sont "rejeter H_0 ou ne pas rejeter H_0". Dans chaque cas une erreur peut être commise et nous résumons la situation dans le tableau suivant :

Statut de H_0 inconnu ! Décision	H_0 fausse (H_1 vraie)	H_0 vraie (H_1 fausse)
rejette H_0	ok	erreur de type I
ne rejette pas H_0	erreur de type II	ok

Il est d'usage de désigner par α (respectivement β) la probabilité de commettre une erreur de type I (respectivement de type II). L'idéal consisterait à pouvoir réduire simultanément les valeurs de α et β pour qu'elles soient proches de 0. Malheureusement, la réduction de α entraîne en général une augmentation de β et réciproquement. Il faut donc se résoudre à ne contrôler qu'un des deux nombres et l'usage veut que cela soit α. Ce cernier est alors appelé le *niveau du test* et les praticiens utilisent des valeurs telles que $\alpha = 5\,\%$, $\alpha = 2\,\%$, $\alpha = 1\,\%$ etc. ... Ainsi la bonne configuration dans un test est celle qui amène le rejet de H_0 car, dans ce cas, le risque est sous contrôle puisqu'il est de type I. On dit alors que le test est significatif et que H_0 est rejetée au niveau α. Dans le cas de non rejet de H_0, nous dirons que H_0 n'est pas rejetée au niveau α et que le test est donc non significatif. La situation est plus délicate car nous ne contrôlons pas l'erreur de type II. La valeur de β peut en fait être très voisine de 1. Certains auteurs comme Neyman et Pearson proposent alors d'agir comme si H_0 était vraie. D'autres statisticiens tels que Fisher proposent au contraire de suspendre tout jugement en attendant de nouvelles données ou informations.

Pour illustrer le fonctionnement d'un test nous revenons au problème de la tartine. Supposons que n chutes aient

été observées et introduisons les variables aléatoires suivantes :

$$X_i = \begin{cases} 1 & \text{si la tartine tombe du côté confiture lors de la } i\text{–ème chute} \quad \text{probabilité } p \\ 0 & \text{sinon} \quad \text{probabilité } q = 1-p \end{cases}$$

$1 \leq i \leq n$. Ainsi $S_n = \sum_{i=1}^{n} X_i$ nous fournit le nombre de fois que la tartine est tombée du côté confiture lors des n chutes. En admettant que les chutes (donc les X_i) soient indépendantes, nous savons que $S_n = \text{Bin}(n,p)$. De plus si $npq > 9$, alors $\dfrac{S_n - np}{\sqrt{npq}}$ peut être approchée par une variable aléatoire $N(0,1)$. Pour une valeur quelconque de p, la loi des grands nombres suggère que les valeurs de S_n auront tendance à se concentrer autour de $E(S_n) = np$. Pour $n = 500$, nous aurons $E(S_{500}) \Big|_{H_0 \text{ vraie}} = 500 \cdot 0.50 = 250$ tandis que $E(S_{500}) \Big|_{H_1 \text{ vraie}} = 500 \cdot 0.55 = 275$. Il est donc pertinent de rejeter H_0 si la valeur de S_n est "trop" grande par rapport à la valeur espérée qui vaut ici 250. Nous cherchons donc un nombre n_α qui aura la fonction suivante :

si $S_n > n_\alpha$ alors on rejette H_0
si $S_n \leq n_\alpha$ alors on ne rejette pas H_0

En choisissant un niveau α nous déterminons $n(\alpha)$ avec la condition

$$P\Big(S_n > n(\alpha)\Big) \Big|_{H_0 \text{ vraie}} = \alpha\,\%.$$

Puisque $n > n(\alpha)$ correspond au rejet de H_0, si celle-ci est vraie, nous commettons une erreur de type I avec un risque de $\alpha\,\%$.

Supposons que $n = 500$ et $\alpha = 5\,\%$. H_0 étant supposée vraie, nous avons $p = 0.50$ et $npq = 500 \cdot \frac{1}{2} \cdot \frac{1}{2} = 125 > 9$. Nous pouvons donc approcher

$$\frac{S_{500} - 500 \cdot \frac{1}{2}}{\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}}}$$

par une variable aléatoire $N(0,1)$ et donc

$$P(S_{500} > n_{0.05}) = P\left(\frac{S_{500} - 500 \cdot \frac{1}{2}}{\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}}} > \frac{n_{0.05} - 500 \cdot \frac{1}{2}}{\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}}}\right) = 5\,\%.$$

Voici une petite table de loi normale :

α	5 %	2.5 %	1 %	0.1 %
$u_{1-\alpha}$	1.645	1.960	2.326	3.090
$u_{1-\frac{\alpha}{2}}$	1.960	2.241	2.576	3.291

L'équation $\dfrac{n_{0.05} - 500 \cdot \frac{1}{2}}{\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = u_{0.95} = 1.645$ nous fournit

$$n_{0.05} = 500 \cdot \frac{1}{2} + 1.645\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 268.391 \cong 268.$$

En conclusion, si H_0 est vraie (i.e. $p = 0.50$), alors celle-ci est rejetée lorsque $S_{500} > 268$ et cet événement, qui correspond à l'erreur de type I, survient avec une probabilité de 5 %. Inversement, si H_0 est fausse (i.e. H_1 est vraie et donc $p = 0.55$), l'erreur de type II correspond à $S_{500} \leq 268$ et la probabilité de cet événement peut être calculée de la façon suivante :

$$\begin{aligned} \beta = P(S_{500} < 268) \Big|_{H_1 \text{ vraie}} &= P\left(\frac{S_{500} - 500 \cdot 0.55}{\sqrt{500 \cdot 0.55 \cdot 0.45}} < \frac{268 - 500 \cdot 0.55}{\sqrt{500 \cdot 0.55 \cdot 0.45}}\right) \\ &\cong \Phi\left(\frac{268 - 500 \cdot 0.55}{\sqrt{500 \cdot 0.55 \cdot 0.45}}\right) \\ &= \Phi(-0.629) = 0.264\,. \end{aligned}$$

Par conséquent $\beta = 26.4\,\%$ dans cette situation.

Que se passe-t-il si l'on abaisse le niveau α à 1 % ? En remplaçant $u_{0.95}$ par $u_{0.99} = 2.326$, nous obtenons $n_{0.01} = 500 \cdot \frac{1}{2} + 2.326\sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}} = 276.005$ et

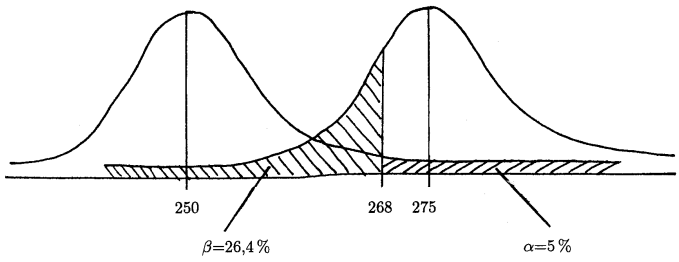
$$\beta = P(S_{500} < 276) \Big|_{H_1 \text{ vraie}} \cong \Phi(0.089) = 0.535\,.$$

Ainsi en abaissant α de 5 % à 1 %, β passe de 26.4 % à 53.5 %. Nous avons donc ici une illustration de l'impossibilité du contrôle simultané de α et β.

Nous pouvons illustrer graphiquement les deux cas précédents en remarquant que $\dfrac{S_n - np}{\sqrt{npq}} \cong N(0,1)$ équivaut à $S_n \cong N(np, \sqrt{npq})$:

$$H_0 \text{ vraie } (p = 0.50): \qquad S_{500} \cong \left(500 \cdot \frac{1}{2}, \sqrt{500 \cdot \frac{1}{2} \cdot \frac{1}{2}}\right) = N(250, 11.18)$$

$$H_1 \text{ vraie } (p = 0.55): \qquad S_{500} \cong \left(500 \cdot 0.55, \sqrt{500 \cdot 0.55 \cdot 0.45}\right) = N(275, 11.12).$$



Remarques : Il est clair que si un test rejette H_0 au niveau α, il rejettera H_0 à tout niveau $\alpha' > \alpha$. Dans la mesure où cela est possible, on peut chercher le niveau le plus petit auquel le test rejette H_0. Ce nombre est appelée "*p–value*" (terminologie anglaise) du test. Il donne une meilleure information sur la situation de H_0 qu'un niveau imposé à priori.

En général l'alternative H_1 sera plus compliquée que $p = 0.55$. H_0 pourra être opposée par exemple à $H_1 : p > 0.50$ et dans de tels cas il est plus difficile de calculer β. Il faut donc rester prudent lorsqu'un test ne rejette pas H_0 car β peut être très proche de 1.

Lorsque l'on teste une probabilité $H_0: p = 0.50$ versus une alternative H_1, cette dernière peut être $p > \frac{1}{2}$ ou $p < \frac{1}{2}$ ou $p \neq \frac{1}{2}$. Dans les deux premiers cas on parlera d'un test unilatéral et dans le troisième d'un test bilatéral.

1712 chutes d'une tartine ont été observées et 1506 fois celle-ci est tombée du côté confiture. Que pouvons-nous conclure ? Nous testons $H_0: p = 0.50$ contre $H_1: p > 0.50$ avec le test décrit précédemment et nous nous proposons de calculer sa p–value, c'est-à-dire $\alpha_{\min}$ tel que H_0 est rejetée.

$$\begin{aligned} \frac{S_n - np}{\sqrt{npq}} &= \frac{S_{1712} - 1712 \cdot \frac{1}{2}}{\sqrt{1712 \cdot \frac{1}{2} \cdot \frac{1}{2}}} \\ &= \frac{1506 - 1712 \cdot \frac{1}{2}}{\sqrt{1712 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = 31.41\,. \end{aligned}$$

Par conséquent

$$\alpha_{\min} = 1 - \Phi(31.41) = 5.57 \cdot 10^{-217}$$

et l'affirmation "H_0 est fausse" peut être considérée ici comme une quasi-certitude.

6.1 Test portant sur une probabilité

Parité des nombres dans un lotto.

Il est légitime de se demander si les parités des nombres choisis au hasard dans un lotto sont équiprobables. On a observé $n = 306$ nombres parmi lesquels 147 étaient impairs et 159 pairs. Les derniers sont-ils plus probables ? Désignons par p la probabilité pour qu'un tel nombre soit pair. Nous avons $H_0: p = 0.50$ et $H_1: p \neq 0.50$. Il s'agit d'un test bilatéral et nous rejetons H_0 si la quantité de nombres pairs observés est soit trop petite, soit trop grande relativement à $n \cdot \frac{1}{2}$. Si S_n désigne le nombre d'entiers pairs, alors $\dfrac{S_n - np}{\sqrt{npq}}$ est proche d'une variable aléatoire normale $N(0,1)$ ($306 \cdot \frac{1}{2} \cdot \frac{1}{2} = 76.5 > 9$). Nous rejetons H_0 au niveau α si

$$\left| \frac{S_n^{\text{obs}} - np}{\sqrt{npq}} \right| > u_{1-\frac{\alpha}{2}}$$

où $\Phi(u_{1-\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$. Pour $\alpha = 5\,\%$, $u_{1-\frac{\alpha}{2}} = 1.960$ et

$$\frac{S_{306}^{\text{obs}} - 306 \cdot \frac{1}{2}}{\sqrt{306 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = -0.686\,.$$

Puisque $-1.960 < -0.686 < 1.960$, nous ne rejetons pas H_0.

Probabilité du sexe à la naissance.

Nous nous proposons de tester l'équiprobabilité des sexes à la naissance dans la population humaine. Parmi $n = 91'342$ naissances en Suisse en 1972, $47'179$ étaient des garçons. Si p désigne la probabilité d'avoir un garçon lors d'une naissance, nous posons $H_0: p = \frac{1}{2}$ et $H_1: p \neq \frac{1}{2}$. Soit S_n le nombre de garçons dans n naissances observées. Nous rejetons H_0 au niveau α si

$$\left| \frac{S_n^{\text{obs}} - np}{\sqrt{npq}} \right| > u_{1-\frac{\alpha}{2}}\,.$$

Dans notre cas :

$$\frac{S_n^{\text{obs}} - np}{\sqrt{npq}} = \frac{47'179 - 91'342 \cdot \frac{1}{2}}{\sqrt{91'342 \cdot \frac{1}{2} \cdot \frac{1}{2}}} = 9.979\,.$$

Il est clair que H_0 est rejetée à $5\,\%$ ($u_{0.975} = 1.960$) et aussi à $1\,\%$ ($u_{0.995} = 2.576$). En fait la p–value du test est de l'ordre de 10^{-24} donc $p \neq \frac{1}{2}$ est une quasi-certitude.

6.2 Le test d'ajustement de χ^2

Nous considérons une variable aléatoire X et un n–échantillon $X_1, X_2, \ldots, X_n$ issu de X constituant n observations indépendantes de X. Nous divisons la droite réelle en r intervalles disjoints

$$\begin{aligned} \mathbf{R} &= (-\infty, t_1] \cup (t_1, t_2] \cup \ldots \cup (t_{\nu-2}, t_{\nu-1}] \cup (t_{\nu-1}, +\infty) \\ &= I_1 \cup I_2 \cup \ldots \cup I_\nu \end{aligned}$$

et nous notons $p_k = P(X \in I_k)$, $1 \leq k \leq \nu$.

Introduisons les variables aléatoires

$$U_i^{(k)} = \begin{cases} 1 & \text{si } X_i \in I_k\,, \\ & \qquad\qquad 1 \leq i \leq n,\ 1 \leq k \leq \nu\,. \\ 0 & \text{sinon} \end{cases}$$

Alors $P(U_i^{(k)} = 1) = P(X_i \in I_k) = P(X \in I_k) = p_k$ et $n_k = \sum_{i=1}^{n} U_i^{(k)}$ donne le nombre de points du n–échantillon qui appartiennent à l'intervalle I_k. Le nombre espéré de points dans I_k est donné par

$$\begin{aligned} E(n_k) &= E\left(\sum_{i=1}^{n} U_i^{(k)}\right) = \sum_{i=1}^{n} E(U_i^{(k)}) = \sum_{i=1}^{n} 1 \cdot p_k \\ &= np_k\,. \end{aligned}$$

Pearson a démontré que la variable aléatoire

$$\sum_{k=1}^{\nu} \frac{(n_k - np_k)^2}{np_k}$$

(le caractère aléatoire provenant de n_k) converge, lorsque $n \to \infty$ (au sens des fonctions de répartition comme dans le théorème limite central), vers une variable aléatoire dite de χ^2 à $\nu - 1$ degrés de liberté notée $\chi^2_{\nu-1}$ dont les fonctions de répartition sont données dans les tables. Nous notons

$$\sum_{k=1}^{\nu} \frac{(n_k - np_k)^2}{np_k} \underset{n\to\infty}{\longrightarrow} \chi^2_{\nu-1}.$$

A nouveau, on tentera d'approcher $\sum_{k=1}^{\nu} \frac{(n_k - np_k)^2}{np_k}$ par $\chi^2_{\nu-1}$ pour n suffisamment grand. On peut montrer que si $np_k \geq 5$ pour $1 \leq k \leq \nu$, alors l'approximation précédente est justifiée. Remarquons que n_k est une fréquence observée tandis que np_k est une fréquence espérée donc théorique et que la somme $\sum_{k=1}^{\nu} \frac{(n_k - np_k)^2}{np_k}$ permet de mesurer l'écart entre fréquences oberservées et fréquences théoriques. Cette grandeur est à la base du test d'ajustement du χ^2. En effet, si l'écart précédent est trop grand, il convient d'admettre que les observations ne proviennnent pas d'une variable aléatoire ayant même distribution que X.

Exemple : Dans une de ses expériences, Mendel a observé 556 petits pois parmi lesquels 315 étaient ronds et jaunes, 108 ronds et verts, 101 ridés et jaunes et 32 ridés et verts. Ces observations sont-elles compatibles avec la théorie de Mendel qui prévoit les probabilités respectives $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$ et $\frac{1}{16}$ pour ces événements ?

fréquences observées	315	108	101	32
fréquences théoriques	312.75	104.25	104.25	34.75

En effet : $556 \cdot \frac{9}{16} = 312.75$, $556 \cdot \frac{3}{16} = 104.25$ et $556 \cdot \frac{1}{16} = 34.75$. Nous constatons que $np_k \geq 5$, $1 \leq k \leq 4$ et donc

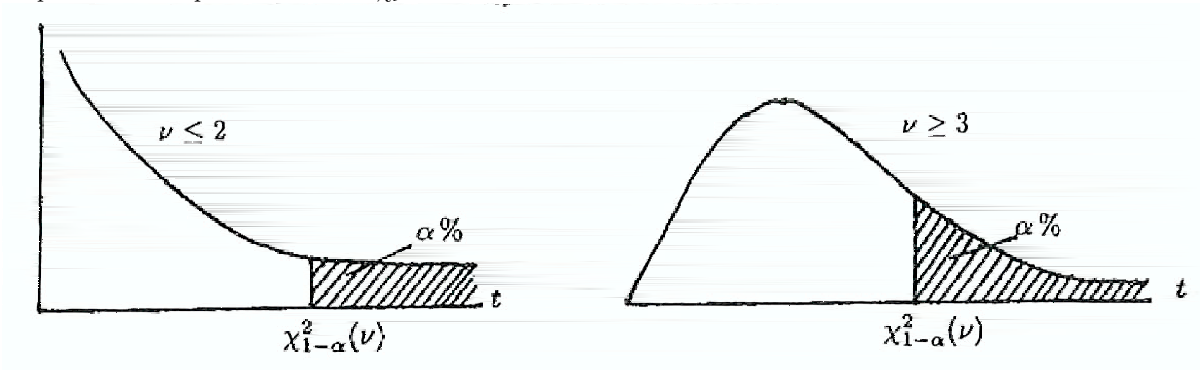
$$\chi^2_{\text{obs}} = \frac{(315 - 312.75)^2}{312.75} + \frac{(108 - 104.25)^2}{104.25} + \frac{(101 - 104.25)^2}{104.25} + \frac{(32 - 34.75)^2}{34.75} = 0.470$$

est (approximativement) une observation d'une loi de χ^2 à $4-1=3$ degrés de liberté.

Posons :

H_0 : les observations sont compatibles avec la théorie de Mendel
H_1 : négation de H_0

On peut montrer que la densité de χ^2_ν a la forme suivante :



Nous introduisons $\chi^2_{1-\alpha}(\nu)$ dont le sens est donné par les graphiques ci-dessus. Ainsi nous rejetons H_0 au niveau α si $\chi^2_{\text{obs}} \geq \chi^2_{1-\alpha}(\nu)$. Dans notre cas $\nu = 4-1=3$ et une table nous fournit $\chi^2_{0,95}(3) = 7.81$. Par conséquent, puisque $\chi^2_{\text{obs}} = 0.470 < 7.81$, nous ne rejetons pas H_0 au niveau 5 %. Il en va évidemment de même à tout niveau $\alpha < 5\,\%$.

6.3 Test d'indépendance d'événements

Les yeux bleus et les cheveux blonds sont-ils des événements indépendants dans la population humaine ? 50 personnes choisies au hasard ont été observées et les résultats sont présentés dans le tableau suivant :

A : avoir les yeux bleus $\quad$ $\overline{A}$: ne pas avoir les yeux bleus
B : avoir les cheveux blonds $\quad$ $\overline{B}$: ne pas avoir les cheveux blonds

	A	$\overline{A}$	Total en ligne
B	12	6	18
$\overline{B}$	12	20	32
Total en colonne	24	26	50

Un tel tableau est appelé tableau de contingence. Nous allons tester

H_0 : indépendance de A et B

contre

H_1 : négation de H_0 .

Il faut remarquer que l'indépendance de A et B, c'est-à-dire $P(A \cap B) = P(A)P(B)$, entraîne celles de A et $\overline{B}$, celle de $\overline{A}$ et B et celle de $\overline{A}$ et $\overline{B}$ (exercice).

Nous estimons d'abord les probabilités de A, $\overline{A}$, B et $\overline{B}$, à l'aide des observations, par les proportions :

$$P(A) = \frac{24}{50} = \frac{12}{25}, \quad P(\overline{A}) = 1 - P(A) = \frac{13}{25}$$

$$P(B) = \frac{18}{5} = \frac{9}{25}, \quad P(\overline{B}) = 1 - P(B) = \frac{16}{25}\,.$$

Sous H_0 nous avons $P(A \cap B) = P(A)P(B)$, $P(A \cap \overline{B}) = P(A)P(\overline{B})$, $P(\overline{A} \cap B) = P(\overline{A})P(B)$ et $P(\overline{A} \cap \overline{B}) = P(\overline{A})P(\overline{B})$ et donc

Fréquence théorique

$$P(A \cap B) = \frac{12}{25} \cdot \frac{9}{25} \quad n_{A\cap B} = nP(A)P(B) = 50 \cdot \frac{12}{25} \cdot \frac{9}{25} = \ 8.64$$

$$P(A \cap \overline{B}) = \frac{12}{25} \cdot \frac{16}{25} \quad n_{A\cap \overline{B}} = nP(A)P(\overline{B}) = 50 \cdot \frac{12}{25} \cdot \frac{16}{25} = 15.36$$

$$P(\overline{A} \cap B) = \frac{13}{25} \cdot \frac{9}{25} \quad n_{\overline{A}\cap B} = nP(\overline{A})P(B) = 50 \cdot \frac{13}{25} \cdot \frac{9}{25} = \ 9.36$$

$$P(\overline{A} \cap \overline{B}) = \frac{13}{25} \cdot \frac{16}{25} \quad n_{\overline{A}\cap \overline{B}} = nP(\overline{A})P(\overline{B}) = 50 \cdot \frac{13}{25} \cdot \frac{16}{25} = 16.64\,.$$

Nous constatons que chaque fréquence théorique est ≥ 5. On compare les fréquences observées aux fréquences théoriques (sous H_0) :

$$\begin{aligned} \chi^2_{\text{obs}} &= \frac{(12-8.64)^2}{8.64} + \frac{(6-9.36)^2}{9.36} + \frac{(12-15.36)^2}{15.36} + \frac{(20-16.64)^2}{16.64} \\ &= 3.93\,. \end{aligned}$$

On peut montrer que χ^2_{obs} provient (approximation) d'une loi de χ^2 à $\nu = 1$ degré de liberté.

Si $\alpha = 5\,\%$ alors $\chi^2_{0.95}(1) = 3.84$. Puisque $\chi^2_{\text{obs}} = 3.93 > 3.84$, le teste rejette H_0 au niveau de $5\,\%$. Par contre, $\chi^2_{0.99}(1) = 6.63$ et $3.93 < 6.63$ montre que H_0 n'est pas rejetée au niveau $1\,\%$. Certains auteurs disent alors que le test est significatif sans être hautement significatif. Les tables montrent que la p–value dans ce cas vaut 0.047.

Remarque : Pour un tableau de contingence

	A	$\overline{A}$
B	a	b
$\overline{B}$	c	d

un calcul direct montre que la valeur associée de χ^2_{obs} est

$$\chi^2_{\text{obs}} = \frac{n(ad-bc)^2}{(a+b)(a+c)(b+d)(c+d)}$$

où n est le nombre d'observations. Cette formule permet de simplifier les calculs.

7 Bibliographie

R. Ineichen, Hj. Stocker : Stochastik, Raeber Verlag.

R. Ineichen : Elementare Beispiele zum Testen statistischer Hypothesen, Orell Fssli.

A. Engel : Les certitudes du hasard, Aleas Editeur.

M.R. Spiegel : Theory and problems of statistics, Schaum Publishing Co.

Y. Dodge : Premiers pas en statistique, Springer.

G. Smith : Statistical reasoning, Allyn and Bacon.

hasard :	Zufall
épreuve aléatoire :	Zufallsexperiment
événement :	Ereignis
événement élémentaire, issue :	Elementarereignis, Ergebnis
événement certain :	sicheres Ereignis
événement impossible :	unmögliches Ereignis
événement contraire :	entgegengesetztes Ereignis
événements incompatibles :	unvereinbare Ereignisse
événements indépendants :	unabhängiges Ereignisse
sous-ensemble, partie :	Teilmenge
espace de probabilité :	Stichprobenraum, Ergebnismenge, Resultatmenge
partition :	Zerlegung