# DENOMBREMENT
# CALCUL DE PROBABILITES

## Exercice 1. ARRANGEMENTS.

Une télévision privée décide d'opter pour le système de « programmes à péage » en utilisant des décodeurs commandés par des codes à huit chiffres.

a) Donner le nombre d'abonnés potentiels puis le nombre d'abonnés avec code composés de huit chiffres différents.
b) Calculer le nombre de codes à 2 chiffres différents, l'un étant utilisé 1 fois et l'autre 7 fois.
c) Même question avec 3 chiffres différents, dont 2 sont utilisés une fois et le troisième 6 fois.

## SOLUTION.

### 1. Nombre d'arrangements avec répétitions de dix éléments huit à huit.

Un abonné potentiel correspond à un code possible, c'est-à-dire à une suite de huit chiffres, chacun étant pris parmi les 10 chiffres possibles { 0 , 1 , 2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 }. Il y a 10 façons de choisir le premier chiffre, 10 façons de choisir le deuxième chiffre, etc., 10 façons de choisir le huitième chiffre, soit au total $10 \times 10 \times ... \times 10 = 10^8$ façons de choisir un code. Il y a donc ***$10^8$* abonnés potentiels** ayant chacun un code différent.

### 2. Nombre d'arrangements sans répétition de dix éléments huit à huit.

Un abonné avec code composé de huit chiffres différents correspond à une suite de huit chiffres différents pris parmi les dix chiffres possibles { 0 , 1 , 2 , 3 , 4 , 5 , 6 , 7 , 8 , 9 }. Il y a 10 façons de choisir le premier chiffre, 9 façons de choisir le deuxième chiffre, etc., 3 façons de choisir le huitième chiffre, soit au total $10 \times 9 \times ... \times 3$ = 1 814 400 façons de choisir un code. Il y a donc ***1 814 400*** = $\dfrac{10\,!}{(10-8)\,!}$ ***abonnés potentiels*** ayant chacun un code différent composé de huit chiffres différents.

### 3. Nombre de codes de deux chiffres différents, l'un utilisé une fois, l'autre sept fois.

Il y a $A_{10}^2 = \dfrac{10\,!}{(10-2)\,!} = 90$ façons de choisir deux chiffres différents parmi les dix chiffres possibles, lorsque le premier chiffre n'apparaîtra qu'une fois et le deuxième chiffre sept fois. Pour chaque choix de deux chiffres différents, le premier chiffre peut occuper l'une quelconque des $C_8^1 = 8$ places dans le code, d'où le nombre de codes de deux chiffres différents, l'un utilisé une seule fois et l'autre sept fois :

$$A_{10}^2 \times C_8^1 = \frac{10!}{8!} \times 8 = 720$$

### 4. Nombre de codes de 3 chiffres différents, dont 2 utilisés une fois et le 3$^e$ six fois.

Il y a $A_{10}^3 = 720$ façons de choisir le chiffre qui sera utilisé 6 fois et les deux chiffres qui ne seront utilisés qu'une fois. Pour chaque choix de trois chiffres, le deux chiffres qui ne sont utilisés qu'une fois peuvent être placés de $C_8^2 = 28$ façons dans le code. Il y a donc au total $A_{10}^3 \times C_8^2$ = ***20 160 codes possibles*** de ce type.

# Exercice 2. COMBINAISONS.

Neuf personnes se présentent à la médecine du travail pour passer la visite annuelle. Deux médecins les reçoivent. Le premier verra 5 personnes, le second 4.

1. De combien de façons différentes les neuf personnes peuvent-elles être réparties entre chaque médecin ?
2. Il y a 4 personnes portant des lunettes. De combien de façons différentes peut-on réaliser cette répartition, sachant que chaque médecin verra 2 personnes portant des lunettes ?
3. De plus, on veut que M. Durand qui porte des lunettes et M. Dupond qui n'en porte pas, soient examinés par le même médecin. Combien de répartitions sont possibles ?

# SOLUTION.

## 1. Nombre de façons de répartir 4 personnes pour un médecin et 5 pour l'autre.

Il y a $C_9^4$ = 126 façons de choisir 4 personnes parmi les 9, et $C_2^1$ = 2 façons de choisir le médecin qui recevra 4 personnes, soit, au total, $C_9^4 \times C_2^1$ = ***252 façons*** de répartir les 9 personnes en groupes de 4 et de 5 entre les médecins.

## 2. Nombre de façons dont chaque médecin peut voir 2 personnes à lunettes.

Il y a $C_2^1$ = 2 façons de choisir le médecin qui recevra 4 personnes. Parmi les 4 personnes à lunettes, il y a $C_4^2$ = 6 façons d'en choisir 2 pour le médecin qui recevra 4 personnes. Parmi les 5 personnes sans lunette, il y a $C_5^2$ = 10 façons de choisir les 2 autres personnes qui seront vues par le médecin qui recevra 4 personnes. Il y a donc au total :

$C_2^1 \times C_4^2 \times C_5^2 = 2 \times 6 \times 10$ = ***120 façons*** dont les 2 médecins peuvent voir chacun deux personnes à lunettes.

## 3. Nombre de façons dont MM. Durand et Dupond peuvent être vus par le même médecin.

Si MM Durand, à lunettes, et Dupond, sans lunettes, sont vus par le médecin qui voit 4 malades, dont 2 à lunettes, il y a $C_3^1$ = 3 façons de choisir un autre patient à lunettes et $C_4^1$ = 4 façons de choisir un autre patient sans lunettes, soit $C_3^1 \times C_4^1$ = 12 façons de choisir un groupe de 4 personnes dont 2 à lunettes, contenant MM Durand et Dupond.

Si MM Durand, à lunettes, et Dupond, sans lunettes, sont vus par le médecin qui voit 5 malades, dont 2 à lunettes, il y a $C_3^1$ = 3 façons de choisir un autre patient à lunettes et $C_4^2$ = 6 façons de choisir un autre patient sans lunettes, soit $C_3^1 \times C_4^2$ = 18 façons de choisir un groupe de 5 personnes dont 2 à lunettes, contenant MM Durand et Dupond.

Comme, par ailleurs, il y a 2! = 2 façons, pour les médecins de se répartir les groupes de 4 et 5 personnes, cela fait au total : 2! ( 12 + 18 ) = ***60 façons*** dont MM Durand et Dupond peuvent être examinés ensemble.

# Exercice 3. COMBINAISONS.

L'épreuve orale de statistique et probabilités d'un DEUG est organisée en lots de 3 sujets tirés au sort parmi 80 sujets portant sur ce cours. L'étudiant doit traiter un des sujets de son choix.

1°/ Combien d'épreuves différentes peut-on organiser ?

2°/ Un candidat se présente en n'ayant révisé que 50 sujets. Quelle est la probabilité pour qu'il puisse traiter :

a) les 3 sujets,
b) deux sujets,
c) un sujet,
d) aucun sujet.

3°/ Combien de sujets un étudiant doit-il réviser pour avoir une probabilité au moins égale à 0,99 de répondre au moins à un sujet ?

# SOLUTION.

## 1.Nombre possible d'épreuves différentes.

C'est le nombre de façons de choisir 3 sujets parmi 80, sans tenir compte de l'ordre :

$$C_{80}^{3} = 82\ 160$$

## 2.Probabilités relatives à un étudiant n'ayant révisé que 50 sujets.

Le nombre de cas possibles est $C_{80}^{3}$ = 82 160.

a) Le nombre de cas favorables est le nombre des combinaisons des 50 sujets révisés, 3 à 3 : $C_{50}^{3}$ = 19 600. La probabilité pour que l'étudiant tombe sur 3 sujets qu'il a révisé est donc :

$$P(3) = \frac{C_{50}^{3}}{C_{80}^{3}} = \frac{19600}{82160} = \frac{245}{1027} \approx 0{,}24$$

b) Le nombre de cas favorables est $C_{50}^{2} \times C_{30}^{1}$ = 1225 × 30 = 36750. La probabilité pour que l'étudiant tombe sur 2 sujets qu'il a révisés est donc :

$$P(2) = \frac{C_{50}^{2} \times C_{30}^{1}}{C_{80}^{3}} = \frac{36750}{82160} = \frac{3675}{8216} \approx 0{,}45$$

c) Le nombre de cas favorables est $C_{50}^{1} \times C_{30}^{2}$ = 50 × 435 = 21750. La probabilité pour que l'étudiant tombe sur un sujet, et un seul, qu'il a révisé, est :

$$P(1) = \frac{C_{50}^{1} \times C_{30}^{2}}{C_{80}^{3}} = \frac{21750}{82160} = \frac{2175}{8216} \approx 0{,}26$$

d) Le nombre de cas favorables est $C_{30}^{3}$ = 4060. La probabilité pour que l'étudiant tombe sur aucun des sujets qu'il a révisés est :

$$P(0) = \frac{C_{30}^{3}}{C_{80}^{3}} = \frac{4060}{82160} = \frac{203}{4108} \approx 0{,}05$$

## 3.Nombre de sujets à réviser pour avoir une probabilité 0,99 de tomber sur un sujet révisé.

Soit *x* le nombre de sujets à réviser : on doit avoir

$$P(0) = \frac{C_{80-x}^{3}}{C_{80}^{3}} \approx 0{,}01, \text{ donc } C_{80-x}^{3} \approx 821{,}60, \text{ soit } \boldsymbol{x = 62}. \ (\ C_{18}^{3} = 816\ ).$$

# Exercice 4. PROBABILITES.

Une étude de la population d'une grande ville de province a fait apparaître que pendant un mois :

35 % des personnes sont allées au cinéma,
12 % des personnes sont allées au musée,
6 % des personnes sont allées aux deux.

Calculer la probabilité que, pendant ce mois, une personne ait fait les choix suivants :

a) Aller au cinéma ou au musée,
b) Ne pas aller au cinéma,
c) N'aller ni au cinéma, ni au musée,
d) Aller au cinéma mais pas au musée.

# SOLUTION.

Les résultats de l'étude peuvent être présentés sous forme de tableau :

| | Sont allées au musée | Ne sont pas allées au musée | Total |
|---|---|---|---|
| Sont allées au cinéma | 6 % | | 35 % |
| Ne sont pas allées au cinéma | | | |
| Total | 12 % | | |

Les cases vides du tableau peuvent être complétées par différence :

| | Sont allées au musée | Ne sont pas allées au musée | Total |
|---|---|---|---|
| Sont allées au cinéma | 6 % | 29 % | 35 % |
| Ne sont pas allées au cinéma | 6 % | 59 % | 65 % |
| Total | 12 % | 88 % | 100 % |

Ce tableau montre que :

59 % des personnes ne sont allées ni au cinéma, ni au musée, donc, par différence,
41 % sont allées au cinéma ou au musée.
65 % des personnes ne sont pas allées au cinéma.
29 % sont allées au cinéma mais ne sont pas allées au musée.

# Exercice 5. PROBABILITES.

Dans un laboratoire se trouve une cage avec 100 souris présentant deux caractères : sexe (mâle ou femelle ), couleur (blanche ou noire) ; 87 sont mâles, 57 sont blanches et 55 sont mâles et blanches.

1°/ Donner l'effectif par catégorie.

2°/ Une assistante prend une souris au hasard. Calculer la probabilité pour qu'elle obtienne une souris blanche ou une souris mâle.

3°/ Elle décide de choisir 6 souris. Calculer la probabilité qu'elle obtienne 6 souris blanches si les prélèvements sont réalisés :

a) avec remise

b) sans remise

# SOLUTION.

## 1.Tableau des effectifs.

Les quatre données de l'énoncé permettent de remplir le tableau des effectifs, suivant les caractères.

| | Mâle | Femelle | Total |
|---|---|---|---|
| Blanche | **55** | 2 | **57** |
| Noire | 32 | 11 | 43 |
| Total | **87** | 13 | **100** |

## 2.Probabilité de tirer une souris blanche ou mâle.

Soit B = « La souris est blanche »

M = « La souris est mâle »

$$P(B \cup M) = P(B) + P(M) - P(B \cap M) = 0{,}57 + 0{,}87 - 0{,}55 = 0{,}87 + 0{,}02 = 0{,}89.$$

## 3.Probabilité de tirer 6 souris blanches dans un tirage avec remise.

Dans un tirage, la probabilité de tirer une souris blanche est $P(B) = 0{,}57$.
Dans 6 tirages indépendants, la probabilité de tirer 6 souris blanches est

$$P(6\,B) = (0{,}57)^6 = 0{,}034296447249 \approx 3{,}43 \times 10^{-2}.$$

## 4.Probabilité de tirer 6 souris blanches en 6 tirages sans remise.

Pour n = 1 à 6, quand on a déjà tiré n - 1 souris blanches, la probabilité de tirer une souris blanche au n-ième tirage est $\frac{57-(n-1)}{100-(n-1)}$. La probabilité de tirer 6 souris blanches en 6 tirages successifs sans remise est alors le produit des probabilités :

$$P(6\,B) = \frac{57}{100}\times\frac{56}{99}\times\frac{55}{98}\times\frac{54}{97}\times\frac{53}{96}\times\frac{52}{95} \approx 3{,}04 \times 10^{-2}.$$

Cette probabilité peut être calculée en faisant le rapport du nombre de cas favorables $C_{57}^{6}$ au nombre de cas possibles $C_{100}^{6}$ :

$$P(6\,B) = \frac{C_{57}^{6}}{C_{100}^{6}} = \frac{36288252}{1192052400} = \frac{2067}{67900} = 0{,}030441826215 \approx 3{,}04 \times 10^{-2}.$$

# Exercice 6.

On veut étudier l'influence de l'ordre de la prise de trois médicaments sur l'efficacité d'un traitement constitué par ces trois produits. De combien de façons possibles peut-on organiser ce traitement ?

# SOLUTION.

Le nombre de traitements possibles est le nombre de ***permutations*** d'un ensemble à trois éléments, qui est aussi le nombre de ***bijections*** d'un ensemble à trois éléments : c'est 3 !, factorielle 3, qui vaut $3 \times 2 \times 1 = 6$.

Il y a 6 façons possibles d'organiser le traitement.

On peut énumérer facilement les six façons possibles d'organiser le traitement avec trois produits $A$, $B$, $C$ :

$(A, B, C)$
$(A, C, B)$
$(B, C, A)$
$(B, A, C)$
$(C, A, B)$
$(C, B, A)$.

# Exercice 7.

Soit l'ensemble des chiffres de 0 à 9 inclus. Combien peut-on former de nombres de cinq chiffres, tous distincts, avec les dix chiffres de l'ensemble ? (On rappelle qu'un nombre de cinq chiffres commençant par un 0 est considéré comme un nombre de quatre chiffres).

# SOLUTION.

## 1^e méthode.

Il y a $A_9^1 = 9$ façons de choisir le premier chiffre parmi les chiffres de 1 à 9. Une fois le premier chiffre choisi, il y a $A_9^4 = 9 \times 8 \times 7 \times 6 = 3\,024$ façons de choisir les 4 chiffres suivants. Au total, il y a $9 \times 3\,024 = 27\,216$ nombres de 5 chiffres différents ne commençant pas par un 0.

27 216 nombres de 5 chiffres différents ne commencent pas par un 0.

## 2^e méthode.

Il y a 9 façons de choisir le premier chiffre différent de 0.
Il y a 9 façons de choisir le deuxième chiffre parmi les 9 chiffres restant après le choix du premier chiffre.
Il y a 8 façons de choisir le 3^e chiffre parmi les 8 chiffres restant après le choix des deux premiers chiffres.
Il y a 7 façons de choisir le 4^e chiffre parmi les 7 chiffres restant après le choix des trois premiers chiffres.
Il y a 6 façons de choisir le 5^e chiffre parmi les 6 chiffres restant après le choix des quatre premiers chiffres.
Au total, nous obtenons $9 \times 9 \times 8 \times 7 \times 6 = 27\,216$ nombres de 5 chiffres différents ne commençant pas par un 0.

27 216 nombres de 5 chiffres différents ne commencent pas par un 0.

## 3^e méthode.

Il y a $A_{10}^5 = 10 \times 9 \times 8 \times 7 \times 6 = 30\,240$ façons de choisir 5 chiffres différents parmi les 10 chiffres de 0 à 9, et de les ordonner pour faire un nombre de 5 chiffres.

Il y a $A_9^4 = 9 \times 8 \times 7 \times 6 = 3\,024$ nombres ayant quatre chiffres différents de 0 et différents entre eux. Ces $A_9^4$ sont en correspondance biunivoque avec les nombres de 5 chiffres différents commençant par un 0.

Il y a donc $A_{10}^5 - A_9^4 = 30\,240 - 3\,024 = 27\,216$ nombres de 5 chiffres différents ne commençant pas par un 0.

27 216 nombres de 5 chiffres différents ne commencent pas par un 0.

## 4^e méthode.

Il y a $A_{10}^5 = 10 \times 9 \times 8 \times 7 \times 6 = 30\,240$ façons de choisir 5 chiffres différents parmi les 10 chiffres de 0 à 9, et de les ordonner pour faire un nombre de 5 chiffres. Pour chacun de ces 30 240 nombres possibles, il y a une chance sur 2 pour qu'il y ait un 0, puisque l'on prend 5 chiffres sur 10. Parmi les nombres contenant un 0, il y a une chance sur 5 pour que le 0 soit soit placé au début, puisque les nombres ont 5 chiffres. Donc, parmi les 30 240 nombres possibles, il y a une chance sur 10 ($\frac{1}{2} \times \frac{1}{5} = \frac{1}{10}$) pour que le nombre commence par un 0, et 9 chances sur 10 pour qu'il ne commence pas par un 0. Il y a donc $0{,}9 \times 30\,240 = 27\,216$ nombres de 5 chiffres différents ne commençant pas par un 0.

27 216 nombres de 5 chiffres différents ne commencent pas par un 0.

# Exercice 8.

On dispose de cinq antibiotiques efficaces pour traiter une maladie infectieuse. On a vérifié au laboratoire que les cinq produits sont également actifs in vitro sur le microbe, mais on ne peut pas donner plus de deux antibiotiques à la fois. Combien y a-t-il de traitements possibles en associant deux antibiotiques ?

# SOLUTION.

Il y a $C_5^2 = \dfrac{5 \times 4}{1 \times 2} = 10$ traitements utilisant 2 antibiotiques si l'ordre de la prise des médicaments n'entre pas en ligne de compte.

10 traitements possibles comprenant deux des cinq antibiotiques utilisables.

Il est facile d'énumérer les traitements possibles en notant $A$, $B$, $C$, $D$, $E$ les antibiotiques :
$\{ A, B \}$ $\{ A, C \}$ $\{ A, D \}$ $\{ A, E \}$
$\{ B, C \}$ $\{ B, D \}$ $\{ B, E \}$
$\{ C, D \}$ $\{ C, E \}$
$\{ D, E \}$.

# Exercice 9.

Une urne contient dix boules (6 blanches et 4 rouges).
On tire au hasard et successivement deux boules de cette urne.
Calculer, dans le cas où le tirage est effectué avec remise, puis dans le cas où le tirage est effectué sans remise, les probabilités suivantes :
— probabilité pour que les deux boules soient blanches,
— probabilité pour que les deux boules soient de même couleur,
— probabilité pour que l'une au moins des boules tirées soit blanche.

# SOLUTION.

## A. Tirage avec remise.

Dans ce cas, les épreuves sont indépendantes et les probabilités sont les mêmes dans les deux tirages successifs. Dans un tirage, la probabilité d'une boule blanche est égale au nombre de cas favorables (6), divisé par le nombre de cas possibles (10), c'est donc 0,6.

### 1. Probabilité pour que les deux boules soient blanches.

Notons $B_1$ l'évènement : " La boule tirée dans le premier tirage est blanche ".
Notons $B_2$ l'évènement : " La boule tirée dans le deuxième tirage est blanche ".
On a :
$P(B_1) = 0{,}6$
$P(B_2) = 0{,}6$
$P(B_1 \cap B_2) = P(B_1)\, P(B_2)$ (relation traduisant l'indépendance en probabilité).
On en déduit la valeur de la probabilité de l'évènement $B_1 \cap B_2$ : " La boule tirée dans le premier tirage est blanche et la boule tirée dans le deuxième tirage est blanche ".
$P(B_1 \cap B_2) = P(B_1)\, P(B_2) = 0{,}6^2 = 0{,}36$.

La probabilité pour que les deux boules tirées soient blanches est 0,36.

### 2. Probabilité pour que les deux boules soient de la même couleur.

Soient
$B_1$ l'évènement : " La boule tirée dans le premier tirage est blanche ",
$B_2$ l'évènement : " La boule tirée dans le deuxième tirage est blanche ",
l'évènement : " La boule tirée dans le premier tirage est rouge ",
l'évènement : " La boule tirée dans le deuxième tirage est rouge ".
On a :
$P(B_1) = 0{,}6$
$P(B_2) = 0{,}6$
$P(\overline{B_1}) = 1 - P(B_1) = 0{,}4$
$P(\overline{B_2}) = 1 - P(B_2) = 0{,}4$
L'évènement $C$ : " Les boules tirées lors des deux tirages successifs sont de même couleur " peut être écrit :

$$C = (B_1 \cap B_2) \cup (\overline{B_1} \cap \overline{B_2})$$

Or les évènements $B_1 \cap B_2$ et $\overline{B_1} \cap \overline{B_2}$ sont incompatibles, leur intersection est vide. L'axiome des probabilités totales indique alors que l'on a :

$$P(C) = P(B_1 \cap B_2) + P(\overline{B_1} \cap \overline{B_2})$$

$$P(C) = P(B_1 \cap B_2) + P(\overline{B_1 \cup B_2}) = P(B_1 \cap B_2) + 1 - P(B_1 \cup B_2)$$
$$P(C) = P(B_1)\, P(B_2) + 1 - (P(B_1) + P(B_1) - P(B_1 \cap B_2)) = 1 + 2\, P(B_1)\, P(B_2) - P(B_1) - P(B_2)$$
$$P(C) = 1 + 2 \times 0{,}6 \times 0{,}6 - 0{,}6 - 0{,}6 = 1{,}72 - 1{,}20 = 0{,}52.$$

La probabilité pour que les deux boules tirées soient de même couleur est 0,52.

On aurait pu écrire aussi $\overline{C} = (B_1 \cap \overline{B_2}) \cup (\overline{B_1} \cap B_2)$, qui est aussi la réunion de deux évènements incompatibles. L'axiome des probabilités totales donne :

$$P(\overline{C}) = P(B_1 \cap \overline{B_2}) + P(\overline{B_1} \cap B_2) = P(B_1)\, P(\overline{B_2}) + P(\overline{B_1})\, P(B_2) = 0{,}6 \times 0{,}4 + 0{,}4 \times 0{,}6 = 0{,}48$$
$$P(C) = 1 - P(\overline{C}) = 0{,}52.$$

### 3. Probabilité pour que l'une au moins des boules tirées soit blanche.

Soit $B$ l'évènement : " L'une au moins des boules tirées est blanche ". L'évènement complémentaire $\overline{B}$ est : " Les deux boules tirées sont rouges ",

$$\overline{B} = \overline{B_1} \cap \overline{B_2}\,.$$
$$P(\overline{B}) = P(\overline{B_1} \cap \overline{B_2}) = P(\overline{B_1}) \times P(\overline{B_2}) = (1 - P(B_1))(1 - P(B_2)) = (1 - 0{,}6)(1 - 0{,}6) = 0{,}16$$
$$P(B) = 1 - P(\overline{B}) = 1 - 0{,}16 = 0{,}84$$

La probabilité pour que l'une au moins des boules tirées soit blanche est 0,84.

## B. Tirage exhaustif (sans remise).

Dans ce cas, la probabilité de tirer une boule blanche dans le premier tirage est $P(B_1) = 0{,}6$. Mais la probabilité de tirer une boule blanche dans le deuxième tirage dépend de la boule tirée au premier tirage :
Si la première boule tirée est blanche : $P(B_2 \mid B_1) = \frac{5}{9}$, sinon $P(B_2 \mid \overline{B_1}) = \frac{6}{9}$.

### 1. Probabilité pour que les deux boules tirées soient blanches.

La définition des probabilité conditionnelles donne :

$$P(B_1 \cap B_2) = P(B_1) \times P(B_2 \mid B_1) = \frac{6}{10} \times \frac{5}{9} = \frac{1}{3}$$

La probabilité pour que les deux boules tirées soient blanches est $\frac{1}{3}$

### 2. Probabilité pour que les deux boules tirées soient de la même couleur.

Comme précédemment, l'évènement $C$ : " Les boules tirées lors des deux tirages successifs sont de même couleur " peut être écrit :

$$C = (B_1 \cap B_2) \cup (\overline{B_1} \cap \overline{B_2})$$

Or les évènements $B_1 \cap B_2$ et $\overline{B_1} \cap \overline{B_2}$ sont incompatibles, leur intersection est vide. L'axiome des probabilités totales indique alors que l'on a :

$$P(C) = P(B_1 \cap B_2) + P(\overline{B_1} \cap \overline{B_2})$$

La différence vient de l'expression des probabilités du deuxième membre puisque les tirages ne sont pas indépendants :

$$P(B_1 \cap B_2) = P(B_1) \times P(B_2 \mid B_1) = \frac{1}{3}$$

$$P(\overline{B_1} \cap \overline{B_2}) = P(\overline{B_1}) \times P(\overline{B_2} \mid \overline{B_1}) = (1 - P(B_1)) \times (1 - P(B_2 \mid \overline{B_1})) = (1 - 0{,}6)\left(1 - \frac{6}{9}\right) = \frac{2}{15}$$

$$P(C) = P(B_1 \cap B_2) + P(\overline{B_1} \cap \overline{B_2}) = \frac{1}{3} + \frac{2}{15} = \frac{7}{15}$$

La probabilité pour que les deux boules soient de même couleur est $\frac{7}{15}$.

## 3. Probabilité pour que l'une au moins des boules soient blanches.

Soit $B$ l'évènement : " L'une au moins des boules tirées est blanche ". L'évènement complémentaire $\overline{B}$ est : " Les deux boules tirées sont rouges ",

$$\overline{B} = \overline{B_1} \cap \overline{B_2}.$$

$$P(\overline{B}) = P(\overline{B_1} \cap \overline{B_2}) = \frac{2}{15}$$

$$P(\mathrm{B}) = 1 - P(\overline{B}) = 1 - \frac{2}{15} = \frac{13}{15}$$

La probabilité pour que l'une au moins des boules tirées soit blanche est $\frac{13}{15}$.

# Exercice 10.

| |
|---|
| Les cultures de tissus végétaux peuvent être infectées soit par des champignons, soit par des bactéries.<br>La probabilité d'une infection par un champignon est 15 %.<br>La probabilité d'infection par une bactérie est 8 %.<br>1°/ Quelle est la probabilité d'une infection simultanée par champignons et bactéries,<br>— dans le cas où les infections sont indépendantes,<br>— dans le cas où les infections n'étant pas indépendantes, la probabilité d'infection par les bactéries quand on a une infection par les champignons est égale à 4 %.<br>2°/ Calculer la probabilité d'infection quelle qu'en soit l'origine (dans les deux cas proposés ci-dessus). |

# SOLUTION.

Nous noterons :
$B$ l'évènement : " La culture est infectée de bactéries ",
$C$ l'évènement : " La culture est infectée de champignons ".
**Données :**

| % | $B$ | $\overline{B}$ | *Total* |
|---|---|---|---|
| $C$ | $P(B \cap C)$ | | 15 |
| $\overline{C}$ | | | 85 |
| *Total* | 8 | 92 | 100 |

## 1. Probabilité d'infection simultanée par champignons et bactéries.

### 1.1. Infections indépendantes.

On cherche la probabilité de l'évènement $B \cap C$. Par définition de l'indépendance en probabilité, on a :
$P(B \cap C) = P(B) \times P(C) = 0{,}08 \times 0{,}15 = 0{,}012$

$$\boxed{\text{La probabilité d'une infection simultanée par champignons et bactéries est 1,2 \%.}}$$

### 1.2. Champignons sécrétant un antibiotique.

$$P(B \cap C) = P(C) \times P(B \mid C) = 0{,}15 \times 0{,}04 = 0{,}006$$

$$\boxed{\text{La probabilité d'une infection simultanée par bactéries et champignons est 0,6 \%.}}$$

## 2. Probabilité d'une infection.

On cherche la probabilité de l'évènement $B \cup C$.

### 2.1. Infections indépendantes.

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0{,}080 + 0{,}150 - 0{,}012 = 0{,}218$$

$$\boxed{\text{La probabilité d'une infection par bactéries ou champignons est 0, 218.}}$$

### 2.2. Champignons sécrétant un antibiotique.

$$P(B \cup C) = P(B) + P(C) - P(B \cap C) = 0{,}080 + 0{,}150 - 0{,}006 = 0{,}224$$

$$\boxed{\text{La probabilité d'une infection par bactéries ou champignons est 0, 224.}}$$

# Exercice 11.

Pour garantir l'anonymat dans certaines enquêtes par sondage, on introduit le hasard dans les réponses possibles.
Admettons que l'on cherche à savoir quelle est la proportion de médecins ayant pratiqué illégalement l'avortement. On demande alors à chaque médecin interrogé de se retirer dans son cabinet et de jouer à pile ou face de la façon suivante :
— S'il obtient pile, il doit répondre, sincèrement, à la question : " Avez-vous un jour pratiqué un avortement illégal ?
— S'il obtient face, il doit rejouer et répondre, toujours aussi sincèrement, à la question : " Avez-vous obtenu face au deuxième tirage ?
Ainsi le médecin donne une seule réponse, " Oui " ou " Non ", mais l'enquêteur ne sait pas s'il a répondu à la première ou à la deuxième question.
Supposons que la proportion de " Oui " ait été de 41 % . Quelle est la proportion de médecins ayant pratiqué un avortement illégal ?

# SOLUTION.

Notons
$F_1$ l'évènement : " On obtient face au premier tirage du jeu de pile ou face ",
$F_2$ l'évènement : " On obtient face au deuxième tirage du jeu de pile ou face ",
$A$ l'évènement : " Le médecin a pratiqué un avortement illégal ",
*Oui* l'évènement : " Le médecin a répondu Oui à l'enquête ".
On connaît : $P(F_1) = P(F_2) = 0{,}5$ et les événements $F_1$ et $F_2$ sont indépendants.
$P(Oui) = 0{,}41$
On cherche à calculer $P(A) = P(Oui \mid \overline{F_1})$

$$Oui = (Oui \cap F_1) \cup (Oui \cap \overline{F_1})$$

Comme les évènement $(Oui \cap F_1)$ et $(Oui \cap \overline{F_1})$ sont incompatibles, l'axiome des probabilités totales donne :

$$P(Oui) = P(Oui \cap F_1) + P(Oui \cap \overline{F_1})$$
$$P(Oui \cap F_1) = P(F_1) \times P(Oui \mid F_1) = P(F_1) \times P(F_2 \mid F_1) = P(F_1) \times P(F_2)$$
$$P(Oui \cap \overline{F_1}) = P(\overline{F_1}) \times P(Oui \mid \overline{F_1}) = (1 - P(F_1)) \times P(A)$$
$$P(Oui) = P(F_1) \times P(F_2) + (1 - P(F_1)) \times P(A)$$
$$P(A) = \frac{P(Oui) - P(F_1)\,P(F_2)}{1 - P(F_1)} = \frac{0{,}41 - 0{,}5 \times 0{,}5}{1 - 0{,}5} = 0{,}32$$

La proportion de médecins ayant pratiqué un avortement illégal est de 32 %.

# PROBABILITES CONDITIONNELLES INDEPENDANCE EN PROBABILITE

# Exercice 12.

Un nouveau vaccin a été testé sur 12500 personnes ; 75 d'entre elles, dont 35 femmes enceintes, ont eu des réactions secondaires nécessitant une hospitalisation.

1°/ Sachant que ce vaccin a été administré à 680 femmes enceintes, quelle est la probabilité qu'une femme enceinte ait eu ne réaction secondaire si elle reçoit le vaccin ?

2°/ Quelle est la probabilité qu'une personne non enceinte ait une réaction secondaire ?

# SOLUTION.

## 1. Probabilité de réaction secondaire pour une femme enceinte vaccinée.

Les données de l'énoncé peuvent être présentées dans un tableau :

| | Réactions secondaires | Pas de réaction secondaire | Total |
|---|---|---|---|
| Femme enceinte | 35 | | 680 |
| Personne non enceinte | | | |
| Total | 75 | | 12500 |

et ce tableau peut être complété par différences :

| | Réactions secondaires | Pas de réaction secondaire | Total |
|---|---|---|---|
| Femme enceinte | **35** | 645 | **680** |
| Personne non enceinte | 40 | 11780 | 11820 |
| Total | **75** | 12425 | **12500** |

35 femmes enceintes sur 680 présentent une réaction secondaire nécessitant une hospitalisation. La probabilité de réaction secondaire pour une femme enceinte est donc $\frac{35}{680} = \frac{7}{136} \approx 5{,}1 \times 10^{-2}$.

## 2. Probabilité de réaction secondaire pour une personne non enceinte vaccinée.

Le tableau précédent montre que 40 personnes non enceintes vaccinées sur 11 820 présentent une réaction secondaire nécessitant une hospitalisation. La probabilité de réaction secondaire pour une personne non enceinte est donc $\frac{40}{11820} = \frac{2}{591} \approx 3{,}4 \times 10^{-3}$.

# Exercice 13.

La population des USA en 1980, classée par région, à répondu de la façon suivante à un sondage sur son attitude face à la législation de la marijuana : (noter que la somme des proportions est égale à 100 %)

| | *Pour* | *Contre* |
|---|---|---|
| Région Est | 7,8 % | 22,2 % |
| Autres régions | 18,2 % | 51,8 % |

1°/ Quelle est la probabilité pour qu'un individu choisi au hasard soit pour la législation ?
2°/ Quelle est la probabilité pour qu'un individu de la Région Est soit pour la législation ?
3°/ Peut-on dire que les évènements « Appartenir à la Région Est » et « Etre pour la législation de la marijuana » sont indépendants ?

# SOLUTION.

## 1. Probabilité pour qu'un individu choisi au hasard soit pour la législation.

On peut compléter le tableau des données par les proportions marginales :

| | Pour | Contre | Total |
|---|---|---|---|
| Région Est | 7,8 % | 22,2 % | 30 % |
| Autres Régions | 18,2 % | 51,8 % | 70,0 % |
| Total | 26,0 % | 74,0 % | 100,0 % |

Dans ce tableau, on voit que 26 % des individus sont pour la législation de la marijuana : la probabilité pour qu'un individu tiré au hasard dans la population soit pour la législation de la marijuana est donc

$$\boxed{P\ (Pour) = 0{,}26}$$

## 2. Probabilité pour qu'un individu de la région Est soit pour la législation.

Par définition de la probabilité conditionnelle :

$$P\ (Pour \mid Est) = \frac{P(\ Pour \cap Est\ )}{P(\ Est\ )} = \frac{7{,}8\ \%}{30{,}0\ \%} = 0,\ 26$$

$$\boxed{P\ (Pour \mid Est) = 0{,}26}$$

## 3. Indépendance en probabilité de deux évènements.

Les évènements « *Est* » et « *Pour* » sont indépendants en probabilité si la relation suivante est vérifiée :

$$P(Pour \cap Est) = P(Pour) \times P(Est)$$

ce qui est équivalent à : $P(Pour|Est) = P(Pour)$. Or on vient de montrer que ces deux probabilités sont égales à 0,26. Donc les deux évènements sont indépendants.

« *Appartenir à la Région Est* » et « *Etre pour la législation de la marijuana* » sont des évènements indépendants.

# Exercice 14.

Dans le cadre de la mise au point d'un alcootest, on a observé les résultats suivants :

| | Sujets ivres | Sujets sains |
|---|---|---|
| Alcootest positif | 150 | 18 |
| Alcootest négatif | 10 | 322 |

les sujets ayant été classés en sujets ivres et sains d'après leur alcoolémie, c'est-à-dire leur concentration sanguine en alcool.

-1- Donner la fréquence des réponses « faux positifs » et des réponses « faux négatifs ».

-2- Calculer la sensibilité et la spécificité de ce test.

-3- Calculer les valeurs prédictives positive et négative ainsi que l'efficacité du test lorsqu'il est appliqué :

- à une population où la probabilité *a priori* de trouver un sujet ivre est P = 0,01 ;
- un soir de réveillon, à une population de sujets dont la probabilité *a priori* d'être ivres est P = 0,20.

-4- En supposant que la spécificité prenne sa valeur maximale et que P = 0,01 :

- quelle valeur aurait VPP ?
- quelle valeur devrait avoir $S_p$ pour que VPP > 0,50 ?

# SOLUTION.

Le tableau des données peut être complété par les effectifs marginaux :

| | Sujets ivres | Sujets sains | Total |
|---|---|---|---|
| Alcootest positif | 150 | **18** | 168 |
| Alcootest négatif | 10 | 322 | 332 |
| Total | 160 | **340** | 500 |

## 1.Fréquence des faux positifs et des faux négatifs.

Il y a, d'après le tableau précédent, 18 ***alcootests positifs*** sur 340 ***sujets sains***, soit une fréquence de « faux positifs » de :

$$\mathbf{FP} = \frac{18}{340} = \frac{9}{170} \approx \mathbf{5{,}29 \times 10^{-2}}.$$

De même, le tableau précédent montre qu'il y a 10 ***alcootests négatifs*** sur 160 ***sujets ivres***, soit une fréquence de « faux négatifs » de :

$$\mathrm{FN} = \frac{10}{160} = \frac{1}{16} = 6{,}25 \times 10^{-2}.$$

## 2.Sensibilité et spécificité du test.

La ***sensibilité*** $S_e$ est définie comme la probabilité d'obtenir un alcotest positif chez un sujet ivre :

$$S_e = \mathrm{P}\ (\text{Alcootest positif} \mid \text{Sujet ivre}) = 1 - \mathrm{FN} = 0{,}9375.$$

La ***spécificité*** $S_p$ est définie comme la probabilité d'obtenir un alcootest négatif chez un sujet sain :

$$S_p = \mathrm{P}\ (\text{Alcootest négatif} \mid \text{Sujet sain}) = 1 - \mathrm{FP} \approx 0{,}9471.$$

# 3. Valeurs prédictives positives et négatives du test dans deux cas.

La *valeur prédictive positive VPP* est, par définition, la probabilité d'être ivre lorsque le test est positif :

$$VPP = P\,(\,Sujet\ ivre \mid Alcootest\ positif\,) = \frac{P(Sujet\ ivre \cap Alcootest\ positif)}{P(Alcootest\ positif)}$$

Cette probabilité ne se calcule pas à partir du tableau directement car, dans le tableau, la probabilité d'être ivre est de $\frac{160}{500}$ = 0,32, différente des valeurs indiquées de 0,01 et de 0,20. Il faut la calculer en fonction de la valeur de $P$ dans la population, compte tenu des valeurs de sensibilité et de spécificité du test.

$$VPP = \frac{P(Alcootest\ positif \mid Sujet\ ivre)\,P(Sujet\ ivre)}{P(Alcootest\ positif \cap Sujet\ ivre) + P(Alcootest\ positif \cap Sujet\ sain)}$$

$$VPP = \frac{P(Alcootest\ positif \mid Sujet\ ivre)\,P(Sujet\ ivre)}{P(Alcootest\ positif \mid Sujet\ ivre)\,P(Sujet\ ivre) + P(Alcootest\ positif \mid Sujet\ sain)\,P(Sujet\ sain)}$$

$$VPP = \frac{P(Alcootest\ positif \mid Sujet\ ivre)\,P(Sujet\ ivre)}{P(Alcootest\ positif \mid Sujet\ ivre)\,P(Sujet\ ivre) + [1 - P(Alcootest\ négatif \mid Sujet\ sain)][1 - P(Sujet\ ivre)]}$$

$$\boxed{VPP = \frac{S_e \times P}{S_e \times P + (1 - S_p) \times (1 - P)}}$$

La *valeur prédictive négative VPN* est la probabilité de ne pas être ivre lorsque l'alcootest est négatif :

$$VPN = P\,(\,Sujet\ sain \mid Alcootest\ négatif\,) = \frac{P(Sujet\ sain \cap Alcootest\ négatif)}{P(Alcootest\ négatif)}$$

$$VPN = \frac{P(Alcootest\ négatif \mid Sujet\ sain) \times P(Sujet\ sain)}{P(Alcootest\ négatif \cap Sujet\ sain) + P(Alcootest\ négatif \cap Sujet\ ivre)}$$

$$VPN = \frac{P(Alcootest\ négatif \mid Sujet\ sain) \times P(Sujet\ sain)}{P(Alcootest\ négatif \mid Sujet\ sain) \times P(Sujet\ sain) + P(Alcootest\ négatif \mid Sujet\ ivre) \times P(Sujet\ ivre)}$$

$$VPN = \frac{P(Alcootest\ négatif \mid Sujet\ sain) \times P(Sujet\ sain)}{P(Alcootest\ négatif \mid Sujet\ sain) \times [1 - P(Sujet\ ivre)] + [1 - P(Alcootest\ positif \mid Sujet\ ivre)] \times P(Sujet\ ivre)}$$

$$\boxed{VPN = \frac{S_p \times (1 - P)}{S_p \times (1 - P) + (1 - S_e) \times P}}$$

Les valeurs de sensibilité $S_e$ et spécificité $S_p$ sont connues pour le test considéré.
La valeur de la prévalence $P$ de l'ivresse dépend de la population étudiée.
L'*efficacité Ef* du test est la probabilité d'obtenir une réponse correcte du test :

$$Ef = P\,(\,Alcootest\ positif \cap Sujet\ ivre\,) + P\,(\,Alcootest\ négatif \cap Sujet\ sain\,)$$

$$Ef = P\,(Alcootest\ positif \mid Sujet\ ivre) \times P\,(Sujet\ ivre) + P\,(Alcootest\ négatif \mid Sujet\ sain) \times P\,(Sujet\ sain)$$

$$Ef = S_e \times P + S_p \times (\,1 - P\,) = (\,S_e - S_p\,)\,P + S_p = \frac{1288 - 13 \times P}{1360}$$

$$\boxed{Ef = (\,S_e - S_p\,)\,P + S_p}$$

## 3.1. Pour P = 0,01 :

On trouve :

$$VPP = \frac{S_e \times P}{S_e \times P + (1 - S_p) \times (1 - P)} = \frac{\frac{15}{16} \times 0{,}01}{\frac{15}{16} \times 0{,}01 + \frac{9}{170} \times 0{,}99} = \frac{425}{2801} \approx 0{,}1517$$

$$VPN = \frac{S_p \times (1 - P)}{S_p \times (1 - P) + (1 - S_e) \times P} = \frac{\frac{161}{170} \times 0{,}99}{\frac{161}{170} \times 0{,}99 + \frac{1}{16} \times 0{,}01} = \frac{127512}{127597} \approx 0{,}9993$$

$Ef = (S_e - S_p)P + S_p = \frac{1288 - 13\times 0,01}{1360}$ = 0,9470

### 3.2. Pour P = 0,20 :

On trouve :

$$VPP = \frac{S_e \times P}{S_e \times P + (1 - S_p)\times(1-P)} = \frac{\frac{15}{16}\times 0,20}{\frac{15}{16}\times 0,20 + \frac{9}{170}\times 0,80} = \frac{425}{521} \approx 0,8157$$

$$VPN = \frac{S_p \times (1-P)}{S_p \times (1-P) + (1 - S_e)\times P} = \frac{\frac{161}{170}\times 0,80}{\frac{161}{170}\times 0,80 + \frac{1}{16}\times 0,20} = \frac{5152}{5237} \approx 0,9838$$

$$Ef = (S_e - S_p)P + S_p = \frac{1288 - 13\times 0,20}{1360} = 0,9451$$

### 3.3. Remarques.

* Dans une population où la probabilité *a priori* $P$ d'être ivre est faible, le test n'a pas grande valeur pour détecter les sujets ivres (moins d'une chance sur 6 que le sujet soit ivre si le test est positif), bien que sensibilité et spécificité du test soient élevées. Le contrôle de l'alcoolémie reste donc important.
* Par contre, la probabilité de laisser passer un sujet ivre est très faible ( $1 - VPN = 0,0007$ si $P = 0,01$).
* $VPN$ et $VPP$ varient en sens inverse en fonction de $P$.

## 4. Cas où la spécificité est maximale.

La valeur maximum de $S_e$ est $S_e = 1$. Dans ce cas, les formules établies plus haut deviennent :

$$VPP = \frac{P}{P + (1 - S_p)\times(1-P)} = \frac{P}{(1 - S_p) + P\times S_p}; \; VPN = 1 \; ; \; Ef = P + (1 - P)\, S_p$$

Pour $P = 0,01$, on obtient :
$VPP = \frac{170}{1061} \approx 0,1602$ ; $VPN = 1$ ; $Ef = \frac{16109}{17000} \approx 0,9476$
Pour que $VPP$ soit plus grand que 0,50, il faudrait que l'on ait : $P > 0,50 - (P + (1 - S_p) \times (1 - P))$, soit $P > (1 - S_p) \times (1 - P)$
$S_p > 1 - \frac{P}{1-P} = \frac{1-2P}{1-P} = \frac{98}{99} \approx 0,9899$.

# Exercice 15.

Dans une population présentant une douleur abdominale, 30 % des patients ont une appendicite aiguë. Parmi ces derniers, 70 % ont une température corporelle supérieure à 37, 5 °C alors que chez des patients sans appendicite, une température supérieure à 37, 5 °C est retrouvée dans 40 % des cas. Pour un patient de cette population présentant une douleur abdominale, on définit les évènements suivants :

$A$ : le patient présente une appendicite aiguë

$T$ : le patient a une température corporelle supérieure à 37, 5 °C.

1.On demande de calculer les probabilités suivantes :

$P(A) =$ Réponse 1

$P(T \mid A) =$ Réponse 2

$P(T \mid \overline{A}) =$ Réponse 3

$P(\overline{T} \mid A) =$ Réponse 4

$P(T) =$ Réponse 5

2. Pour un patient ayant une température supérieure à 37, 5 °C, calculer la probabilité "$P_1$" qu'il ait une appendicite aiguë.

$P_1 =$ Réponse 6

3. Pour un patient n'ayant pas une température supérieure à 37, 5 °C, calculer la probabilité "$P_2$" qu'il n'ait pas une appendicite aiguë.

$P_2 =$ Réponse 7

*Tableau de réponses de l'exercice 15 (Réponses 1 à 7)*

| | 0,01 | 0,03 | 0,05 | 0,07 | 0,09 | 0,13 | 0,20 | 0,25 | 0,33 | 0,40 | 0,49 | 0,60 | 0,65 | 0,73 | 0,83 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0,02 | 0,04 | 0,06 | 0,08 | 0,10 | 0,18 | 0,23 | 0,30 | 0,35 | 0,43 | 0,55 | 0,63 | 0,70 | 0,80 | 0,90 |

| | B | D | AB | AD | BC | BE | CE | ABC | ABE | ACE | BCD | BDE | ABCD | ABDE | BCDE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | C | E | AC | AE | BD | CD | DE | ABD | ACD | ADE | BCE | CDE | ABCE | ACDE | ABCDE |

# SOLUTION.

## 1°/ Probabilités conditionnelles.

| | T | $\overline{T}$ | Total |
|---|---|---|---|
| **A** | 0,30 × 0,70 | | 0,30 |
| **$\overline{A}$** | 0,70 × 0,40 | | 0,70 |
| **Total** | | | 1 |

| | T | $\overline{T}$ | Total |
|---|---|---|---|
| **A** | 0,21 | 0,09 | 0,30 |
| **$\overline{A}$** | 0,28 | 0,42 | 0,70 |
| **Total** | 0,49 | 0,51 | 1 |

Les données de l'énoncé permettent de construire le ***<u>tableau des probabilités conjointes</u>***.

Ce tableau peut être complété par soustraction et addition.

Sur ce tableau, on lit alors :

$P(A) = 0,30$ **Réponse 1 : ABD**

$P(T \mid A) = \dfrac{0,21}{0,30} = 0,70$ **Réponse 2 : ABCE**

$P(T \mid \overline{A}) = \dfrac{0,28}{0,70} = 0,40$ **Réponse 3 : ACE**

$P(\overline{T} \mid A) = \dfrac{0,09}{0,30} = 0,30$ **Réponse 4 : ABD**

$P(T) = 0,49$ **Réponse 5 : BCD**

## 2°/ Probabilité $P_1$ .

$P_1 = P(A \mid T) = \dfrac{0,21}{0,49} = \dfrac{3}{7} = 0,43$ **Réponse 6 : ADE**

## 3°/ Probabilité $P_2$.

$P_1 = P(\overline{A} \mid \overline{T}) = \dfrac{0,42}{0,51} = \dfrac{14}{17} = 0,82$ **Réponse 7 : BCDE**

# Exercice 16.

Une boîte de médicament $B_1$ contient 10 comprimés jaunes et 0 rouge.
Une boîte de médicament $B_2$ contient 4 comprimés jaunes et 1 rouge.
Une boîte de médicament $B_3$ contient 5 comprimés jaunes et 5 rouges.
1°/ On prend une boîte au hasard et de cette boîte, on tire un comprimé au hasard : le comprimé est rouge. Calculer la probabilité pour que le comprimé ainsi tiré provienne de la boîte $B_1$ ? de la boîte $B_2$ ? de la boîte $B_3$ ?
2°/ On réunit les 25 comprimés des boîtes $B_1$, $B_2$, $B_3$. On tire au hasard un comprimé et le comprimé obtenu est rouge. Calculer la probabilité pour que le comprimé ainsi tiré provienne de la boîte $B_1$ ? de la boîte $B_2$ ? de la boîte $B_3$ ?

# Solution.

La ***formule de Bayes*** donne :

$$\boxed{P(B_i \mid R) = \frac{P(R|B_i)P(B_i)}{P(R|B_1)P(B_1)+P(R|B_2)P(B_2)+P(R|B_3)P(B_3)}}$$

avec

$$P(R \mid B_1) = 0$$
$$P(R \mid B_2) = \frac{1}{5} = 0{,}2$$
$$P(R \mid B_3) = \frac{5}{10} = 0{,}5$$

## 1°/ Tirage en deux étapes.

Les boîtes sont équiprobables : $P(B_1) = P(B_2) = P(B_3) = \frac{1}{3}$.
La formule de Bayes ci-dessus se réduit à :

$$P(B_i \mid R) = \frac{P(R|B_i)}{P(R|B_2)+P(R|B_3)} = \frac{10\times P(R|B_i)}{7}$$
$$\boxed{P(B_i \mid R) = \frac{10\times P(R|B_i)}{7}}$$

D'où :

$$\boxed{\begin{array}{c} P(B_1 \mid R) = 0 \\ P(B_2 \mid R) = \frac{2}{7} \\ P(B_3 \mid R) = \frac{5}{7} \end{array}}$$

## 2°/ Tirage en une étape.

La probabilité $P(B_i)$ pour que le comprimé tiré provienne de la boîte $B_i$ est égale au nombre de cas favorables, nombre de comprimés de la boîte $B_i$, divisé par le nombre de cas possibles, nombre total de comprimés, soit 25.

$P(B_1) = \frac{10}{25} = 0{,}4$

$P(B_2) = \frac{5}{25} = 0{,}2$

$P(B_3) = \frac{10}{25} = 0{,}4$

La formule de Bayes devient :

$$P(B_i \mid R) = \frac{P(R|B_i)P(B_i)}{0{,}2\times 0{,}2 + 0{,}5\times 0{,}4} = \frac{P(R|B_i)P(B_i)}{0{,}24} = 25 \times P(B_i) \times \frac{P(R|B_i)}{6}$$

$$= (\textit{nombre de comprimés de la boîte } B_i) \times \frac{P(R|B_i)}{6}$$

$$\boxed{P(B_i \mid R) = \frac{\textit{nombre de comprimés rouges de la boîte } B_i}{\textit{nombre total de comprimés rouges}}}$$

On obtient donc :

$$\boxed{\begin{array}{c} P(B_1 \mid R) = 0 \\ P(B_2 \mid R) = \dfrac{1}{6} \\ P(B_3 \mid R) = \dfrac{5}{6} \end{array}}$$

# Exercice 17.

Dans une population contenant 0,1 % de malades atteints d'une maladie " M ", on choisit une personne au hasard. Cette personne subit un test donné " T ". On suppose que l'aptitude du test à déceler la maladie " M " est définie ainsi :

— sur 100 personnes atteintes de la maladie " M ", 90 auront un test positif.

— sur 100 personnes non atteintes de la maladie " M ", 1 personne aura un test positif.

On définit les évènements aléatoires suivants :

$T$ : " Le test considéré est positif "

$\overline{T}$ : " Le test considéré est négatif "

$M$ : " La personne choisie au hasard est atteinte de la maladie M "

$\overline{M}$ : " La personne choisie au hasard n'est pas atteinte de la maladie M " .

1°/ Donner les valeurs des probabilités suivantes : $P(M)$, $P(\overline{M})$, $P(T \mid M)$, $P(T \mid \overline{M})$.

2°/ Calculer la spécificité du test.

3°/ En remarquant que l'on peut écrire $T = (M \cap T) \cup (\overline{M} \cap T)$, calculer la probabilité pour que le test " T " soit positif.

4°/ En déduire la valeur prédictive positive du test, c'est-à-dire la valeur de la probabilité pour que, le test étant positif, la personne choisie soit réellement atteinte de la maladie " M ".

5°/ Calculer la valeur prédictive négative et l'efficacité du test.

6°/ Refaire l'exercice en supposant que la prévalence de la maladie est 10 %.

# Solution

## 1°/ Valeurs de probabilités.

Les données peuvent être résumées dans le tableau suivant :

| | $T$ | $\overline{T}$ | |
|---|---|---|---|
| $M$ | 90 | 10 | 100 |
| $\overline{M}$ | 999 | 98 901 | 99 900 |
| | 1 089 | 98 911 | 100 000 |

Les probabilités demandées se lisent sur le tableau :

$$P(M) = \frac{100}{100.000} = 0{,}1\ \%$$

$$P(\overline{M}) = \frac{99.900}{100.000} = 99{,}9\ \%$$

$$P(T \mid M) = \frac{90}{100} = 90\ \%$$

$$P(T \mid \overline{M}) = \frac{999}{99.900} = 1\ \%$$

## 2°/ Spécificité.

Par définition, la *spécificité* du test est l'aptitude à éliminer les faux malades : c'est la probabilité pour que le test soit négatif lorsque la personne n'est pas malade :

$$S_p = P(\overline{T} \mid \overline{M})$$

| | $T$ | $\overline{T}$ | |
|---|---|---|---|
| $M$ | 90 | 10 | 100 |
| $\overline{M}$ | 999 | 98 901<br>*Spécificité* | 99 900 |
| | 1 089 | 98 911 | 100 000 |

La spécificité est ici : $S_p = \frac{98.901}{99.900} = \frac{99}{100}$.

$$\boxed{S_p = 99\ \%}$$

# 3°/ Probabilité d'un test positif.

Le tableau :

| | $T$ | $\overline{T}$ | |
|---|---|---|---|
| $M$ | 90 | 10 | 100 |
| $\overline{M}$ | 999 | 98 901 | 99 900 |
| | 1 089 | 98 911 | 100 000 |

donne la réponse :

$$\boxed{P(T) = \frac{1.089}{100.000} = 1{,}089 \times 10^{-2}}$$

# 4°/ Valeur prédictive positive du test.

La valeur prédictive positive se déduit du tableau :

| | $T$ | $\overline{T}$ | |
|---|---|---|---|
| $M$ | 90 | 10 | 100 |
| $\overline{M}$ | 999 | 98 901 | 99 900 |
| | 1 089 | 98 911 | 100 000 |

$$\boxed{VPP = P(M \mid T) = \frac{90}{1089} = \frac{10}{121} = 8{,}3\ \%}$$

# 5°/ Valeur prédictive négative du test.

La valeur prédictive négative se déduit du tableau :

| | $T$ | $\overline{T}$ | |
|---|---|---|---|
| $M$ | 90 | 10 | 100 |
| $\overline{M}$ | 999 | 98 901 | 99 900 |
| | 1 089 | 98 911 | 100 000 |

$VPN = P(\overline{M}|\overline{T}) = \frac{98.901}{98.911} = 99{,}99\ \%$

# 6°/ Cas d'une prévalence de 10 %.

Le tableau est alors le suivant :

| | $T$ | $\overline{T}$ | | |
|---|---|---|---|---|
| $M$ | 9 000 | 1 000 | 10 000 | Sensibilité |
| $\overline{M}$ | 900 | 89 100 | 90 000 | Spécificité |
| | 9 900 | 90 100 | 100 000 | |
| | Valeur Prédictive Positive | Valeur Prédictive Négative | | |

Les nouvelles valeurs de probabilité se déduisent du tableau :

$P(M) = 10\ \%$ $P(\overline{M}) = 90\ \%$ $P(T\,|\,M) = 90\ \%$ (Sensibilité) $P(T\,|\overline{M}) = 1\ \%$

*Sensibilité* ($S_e$) = $P(T\,|\,M) = 90{,}00\ \%$

*Spécificité* ($S_p$) = $P(\overline{T}|\overline{M}) = \frac{89.100}{90.000} = 99{,}00\ \%$

*Valeur prédictive positive* (*VPP*) = $P(M\,|\,T) = \frac{9.000}{9.900} = 90{,}91\ \%$

*Valeur prédictive négative* (*VPN*) = $P(\overline{M}|\overline{T}) = \frac{89.100}{90.100} = 98{,}89\ \%$.

On peut calculer aussi :

· Le *rapport de vraisemblance* du test : $k = \frac{S_e}{1 - S_p} = \frac{90}{100 - 99} = 90$.

· Le *risque relatif* : $\frac{VPP}{1 - VPN} = \frac{90{,}91}{100 - 98{,}89} = 91{,}82\ \%$.

On voit donc que c'est surtout la valeur prédictive positive du test qui est augmentée, ainsi que le risque relatif, lorsque la prévalence augmente.

# VARIABLES ALEATOIRES DISCRETES

# Exercice 18.

On place un hamster dans une cage. Il se trouve face à 5 portillons dont un seul lui permet de sortir de la cage. A chaque essai infructueux, il reçoit une décharge électrique et on le replace à l'endroit initial.

1 – En supposant que le hamster ne soit pas doué d'apprentissage et qu'il choisisse donc de façon équiprobable entre les 5 solutions à chaque nouvel essai, déterminer la probabilité des évènements :

a) le hamster sort au premier essai,
b) le hamster sort au troisième essai,
c) le hamster sort au septième essai.

2 – Le hamster mémorise maintenant les essais infructueux et choisit de façon équiprobable entre les portillons qu'il n'a pas encore essayés. On désigne par X la variable aléatoire égale au nombre d'essais effectués.

a) Quelles valeurs peut prendre X ? Déterminer sa loi de probabilité.
b) Déterminer l'espérance mathématique E(X) : interpréter le résultat.
c) Déterminer la variance V(X).

# SOLUTION.

## 1°/ Hamster non doué d'apprentissage.

***Les essais sont indépendants*** et, pour chaque essai, ***la probabilité p de tomber sur le bon portillon est la même*** :

$$\text{P ( succès )} = p = \frac{1}{5} = 0{,}20.$$

A chaque essai, la probabilité qu'a le hamster de se tromper est la même : P ( échec ) = $q = 1 - p = \frac{4}{5} = 0{,}80$.

a) ***P ( un essai )*** $= \frac{1}{5} = 0{,}20$.

b) ***P ( trois essais )*** = P ( échec au premier essai $\cap$ échec au deuxième essai $\cap$ succès au troisième essai )
= P ( échec au premier essai ) $\times$ P ( échec au deuxième essai ) $\times$ P ( succès au troisième essai )
= P ( échec ) $\times$ P ( échec ) $\times$ P ( succès ) = $q^2\, p$
$= \frac{4}{5} \times \frac{4}{5} \times \frac{1}{5} = \frac{16}{125} = 0{,}128$

c) ***P (sept essais)*** $= q^6\, p = \left(\frac{4}{5}\right)^6 \times \frac{1}{5} = 0{,}0524288 \approx 5{,}24 \times 10^{-2}$.

Si le hamster dispose d'un nombre infini d'essais, sa probabilité de sortir de la cage est :

$$\text{P ( sortie )} = p \times \sum_{i=0}^{\infty} q^i = p \times \frac{1}{1-q} = 1.$$

A force d'essayer au hasard, le hamster est sûr de sortir de sa cage ! Mais il peut mettre longtemps.

## 2°/ Hamster doué de mémoire des échecs.

### a) Valeurs de X et loi de probabilité.

Comme il y a 5 portillons, le hamster est sûr de pouvoir sortir de sa cage en 5 essais au maximum, puisqu'il n'essaie pas deux fois la même mauvaise porte. Le nombre d'essais X que doit faire le hamster pour trouver le bon portillon de sortie peut prendre les valeurs 1, 2, 3, 4 et 5. C'est une variable aléatoire discrète. Sa loi de probabilité s'obtient en calculant les probabilités pour que X soit égal à 1, 2, 3, 4 ou 5.

Au départ, le hamster a une chance sur 5 de trouver le bon portillon du premier coup. P (X = 1) = $\frac{1}{5}$ = 0,20.

Si le hamster n'a pas trouvé du premier coup le bon portillon (probabilité q = $\frac{4}{5}$), il tente à nouveau sa chance avec une chance sur 4 de réussite : P (X = 2 ) = $\frac{4}{5} \times \frac{1}{4} = \frac{1}{5}$ = 0,20.

Si le hamster n'a toujours pas trouvé le bon portillon au deuxième essai (probabilité $\frac{4}{5} \times \frac{3}{4} = \frac{3}{5}$), il tente à nouveau sa chance avec une chance sur 3 de réussite : P ( X = 3 ) = $\frac{3}{5} \times \frac{1}{3} = \frac{1}{5}$ = 0,20.

Si le hamster n'a toujours pas trouvé le bon portillon au troisième essai (probabilité $\frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} = \frac{2}{5}$), il tente à nouveau sa chance avec une chance sur 2 de réussite : P ( X = 4 ) = $\frac{2}{5} \times \frac{1}{2} = \frac{1}{5}$ = 0,20.

Si le hamster n'a toujours pas trouvé le bon portillon au quatrième essai (probabilité $\frac{4}{5} \times \frac{3}{4} \times \frac{2}{3} \times \frac{1}{2} = \frac{1}{5}$), il est assuré de trouver le bon portillon au dernier essai puisqu'il n'en reste qu'un à essayer : P ( X = 5 ) = $\frac{1}{5}$ = 0,20.

Nous pouvons donc dresser le tableau suivant de la loi de probabilité de X :

| Valeurs de X = nombre d'essais $x_i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| P ( X = $x_i$ ) | 0,20 | 0,20 | 0,20 | 0,20 | 0,20 |

Le hamster a la même probabilité de pouvoir sortir de la cage au bout de 1, 2, 3, 4 ou 5 essais.

La somme des probabilités des valeurs de X est bien égale à 1 : $\sum_{i=1}^{i=5} P\ (\ X = x_i\ ) = 1$.

## b) Espérance mathématique de X.

$$E(X) = \sum_{i=1}^{i=5} x_i\, P\ (\ X = x_i\ ) = (\ 1 + 2 + 3 + 4 + 5\ ) \times 0{,}2 = 3$$

En moyenne, le hamster trouvera la bonne porte de sortie en trois essais.

## c) Variance de X.

$$V(X) = E\ (X^2) - [\ E(X)\ ]^2 = (\ 1^2 + 2^2 + 3^2 + 4^2 + 5^2\ ) \times 0{,}2 - 3^2 = 2$$

# Exercice 19.

Un certain jeu dépend du nombre de points X obtenus en jetant un dé équilibré.
1°/ Calculer la loi de probabilité de X. Calculer l'espérance mathématique E(X) et la variance V(X).
2°/ Le gain de ce jeu est donné par la fonction linéaire suivante G = 2 X + 8.

- Tabuler la distribution de G ;
- Calculer le gain espéré E(G) et la variance V(G).

# SOLUTION.

## 1°/ Loi de probabilité de la variable aléatoire X.

Chaque valeur de X est équiprobable, donc, avec un dé à six faces bien équilibré, les probabilités des valeurs de X sont données dans le tableau suivant :

| $X = x_i$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(X = x_i)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

Ce tableau permet de calculer les paramètres de la distribution de X :

$$E(X) = \sum_{i=1}^{i=6} x_i \ P(X = x_i) = (1+2+3+4+5+6) \times \frac{1}{6} = \frac{7}{2} = 3{,}5$$

$$V(X) = E(X^2) - [E(X)]^2 = (1^2+2^2+3^2+4^2+5^2+6^2) \times \frac{1}{6} - \left(\frac{7}{2}\right)^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12} \approx 2{,}9167$$

L'écart-type de X est la racine carrée de la variance :

$$\sigma_X = \left(\frac{35}{12}\right)^{\frac{1}{2}} = \frac{\sqrt{105}}{6} \approx 1{,}708$$

## 2°/ Loi de probabilité de la variable aléatoire G.

| $G = g_i = 2\,x_i + 8$ | 10 | 12 | 14 | 16 | 18 | 20 |
|---|---|---|---|---|---|---|
| $P(G = g_i)$ $= P(X = x_i)$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ | $\frac{1}{6}$ |

La relation $G = 2\,X + 8$ permet de calculer les pramètres de la distribution de $G$.

$$E(G) = 2\,E(X) + 8 = 2 \times \frac{7}{2} + 8 = 15.$$

$$V(G) = 2^2\,V(X) = 4 \times \frac{35}{12} = \frac{35}{3} \approx 11{,}67$$

$$\sigma_G = 2\,\sigma_X = \frac{\sqrt{105}}{3} \approx 3{,}416$$

# Exercice 20.

La probabilité d'observer une maladie dans une population est $P$ = 0,1 et la maladie peut être détectée sans erreur par un dosage sanguin. On souhaite déterminer par ce dosage le nombre de personnes malades sur un échantillon de 100 personnes.
Mais au lieu de tester le sérum de chaque individu, on partitionne au hasard l'échantillon en 10 groupes de 10 personnes dont on mélange les sérums.
Si le test est négatif, sur l'un de ces mélanges, on considère que les 10 personnes correspondantes sont toutes négatives et l'on est ainsi dispensé des 10 tests individuels.
Si, au contraire, le test est positif, c'est qu'alors au moins une personne est atteinte de la maladie et il faut alors tester séparément chacun des 10 sérums ayant participé au mélange, on doit donc, dans ce cas, effectuer 11 tests.

1°/ Trouver les probabilités pour que, dans un groupe, on observe :
- aucune personne malade,
- une et une seule personne malade,
- au moins une personne malade.

2°/ En désignant par $N$ le nombre total de tests à effectuer avec cette méthode de partition pour un échantillon de 100 personnes, trouver les probabilités suivantes :
- $P ( N = 110 )$
- $P ( N = 100 )$

3°/ Calculer le nombre moyen de tests $E(N)$ et la variance du nombre des tests $Var (N)$.

# SOLUTION.

## 1°/ Loi de probabilité du nombre X de personnes malades dans un groupe de 10 personnes.

Pour une personne du groupe, de deux choses l'une : ou bien elle est atteinte de la maladie (probabilité $P$ = 0,1), ou bien elle n'est pas atteinte de la maladie (probabilité $q$ = 0,9). Le fait de regarder si la personne est, ou non, atteinte de la maldie constitue une ***épreuve de Bernoulli***. Si l'on répète pour les 10 personnes du groupe cette épreuve de Bernoulli, le nombre $X$ de succès, c'est-à-dire le nombre de personnes atteintes de la maladie suit une ***loi binomiale*** $B$ ( 10 ; 0,1 ) :

$$P ( X = k ) = C_{10}^{k} ( 0,1 )^{k} ( 0,9 )^{10-k}.$$

De cette loi binomiale, on tire :

$$P ( X = 0 ) = ( 0,9 )^{10} \approx 0,35$$
$$P ( X = 1 ) = 10 \times 0,1 \times ( 0,9 )^{9} \approx 0,39$$
$$P ( X \geq 1 ) = 1 - P ( X = 0 ) \approx 1 - 0,35 = 0,65$$

## 2°/ Loi de probabilité du nombre total N de tests à effectuer.

Pour un groupe, de deux choses l'une : ou bien il contient au moins un malade (probabilité $p = 1 - 0,9^{10} = 0,65$ ) et, dans ce cas, l'on refait 10 tests, ou bien il ne comporte aucun malade (probabilité $q = 0,9^{10} = 0,35$ ) et, dans ce cas, l'on ne refait aucun test.

Le fait de **regarder, dans un groupe**, s'il y a ou non au moins un malade, c'est-à-dire le fait de regarder, pour un groupe, **si le test est positif**, constitue une ***épreuve de Bernoulli***. Lorsque l'on répète cette épreuve de Bernoulli pour les 10 groupes, le nombre $X$ de succès, c'est-à-dire le nombre de groupes pour lesquels il faut refaire 10 tests, suit une ***loi binomiale*** $B ( 10 ; 1 - 0,9^{10} )$ :

$$P ( X = k ) = C_{10}^{k} ( 1 - 0,9^{10} )^{k} 0,9^{10(10-k)}.$$

Lorsque l'on a $X = k$, le nombre $N$ de tests à faire est $N = 10 + 10\ k = 10\ (k + 1)$. Inversement si l'on pratique au total $N = 10\ (k + 1)$ tests, c'est que l'on a refait $k$ fois 10 tests, donc qu'il y avait $X = k$ groupes à test positif. Ainsi :

$$X = k \Leftrightarrow N = 10\ (k + 1).$$

La loi de probabilité de $X$ donne alors la ***loi de probabilité de* N** :

$$\boxed{P\ [\ N = 10\ (k + 1)\ ] = C_{10}^{k}\ (1 - 0{,}9^{10})^{k}\ 0{,}9^{10(10 - k)}.}$$

Pour $k = 10$ (10 groupes ayant au moins un malade), cette loi de probabilité donne :

$$\boldsymbol{P\ (N = 110) =} (1 - 0{,}9^{10})^{10} = \mathbf{1{,}37 \times 10^{-2}}.$$

Pour k = 9 (9 groupes ayant au moins un malade), cette loi de probabilité donne :

$$\boldsymbol{P\ (N = 100)} = 10 \times (1 - 0{,}9^{10})^{9} \times 0{,}9^{10} = \mathbf{7{,}36 \times 10^{-2}}.$$

## 3°/ Paramètres statistiques de la loi de probabilité de *N*.

Le nombre $X$ de groupes dans lesquels il faut refaire 10 tests parce le test est positif, suit une loi binomiale

$$B\ (10\ ;\ 1 - 0{,}9^{10})$$

La valeur moyenne de $X$ est donnée par $\boldsymbol{E\ (X)} = 10 \times (1 - 0{,}9^{10}) \approx 6{,}51$.

La variance de $X$ est donnée par $\boldsymbol{Var\ (X)} = 10 \times (1 - 0{,}9^{10}) \times 0{,}9^{10} \approx 2{,}27$.

Les paramètres statistiques de $X$ permettent de calculer les paramètres statistiques de $\boldsymbol{N = 10\ (X + 1)}$ :

$\boldsymbol{E\ (N) =} 10\ (\mathrm{E}\ (\mathrm{X}) + 1) = 100 \times (1 - 0{,}9^{10}) + 10 = 110 - 100 \times 0{,}9^{10} \approx \mathbf{75{,}132}$.

$\boldsymbol{Var\ (N) =} 100\ \mathrm{Var}\ (\mathrm{X}) = 10^{3} \times (1 - 0{,}9^{10}) \times 0{,}9^{10} \approx \mathbf{227{,}1018}$.

## 4°/ Cas de *G* groupes de taille *n*.

Si l'on considère un groupe de taille $n$, la probabilité pour qu'il contienne au moins un malade est $1 - 0{,}9^{n}$ . Si l'on considère $G$ groupes de taille $n$, le nombre $x$ de groupes comportant au moins un malade suit une loi binomiale

$$B\ (G\ ;\ 1 - 0{,}9^{n})$$

La valeur moyenne de $X$ est donnée par $\boldsymbol{E\ (X)} = G \times (1 - 0{,}9^{n})$.

La variance de $X$ est donnée par $\boldsymbol{Var\ (X)} = G \times (1 - 0{,}9^{n}) \times 0{,}9^{n}$ .

Le nombre $N$ de tests à effectuer, qui est donné par $\boldsymbol{N = G + n\ X}$, a pour paramètres statistiques :

$$\boxed{\begin{array}{c} \boldsymbol{E\ (N) =} n\ E\ (X) + G = n\ G \times (1 - 0{,}9^{n}) + G \\ \boldsymbol{Var\ (N) =} n^{2}\ Var\ (X) = n^{2}\ G \times (1 - 0{,}9^{n}) \times 0{,}9^{n} \end{array}}$$

### a) Pour G = 5 groupes de n = 20 personnes :

**E ( N ) =** $n\ G \times (1 - 0{,}9^{n}) + G = 100 \times (1 - 0{,}9^{20}) + 5 =$ **92,8**42

**Var ( N )** $= n^{2}\ G \times (1 - 0{,}9^{n}) \times 0{,}9^{n} = 2000 \times (1 - 0{,}9^{20}) \times 0{,}9^{20} =$ **213,59**2

### b) Pour G = 20 groupes de n = 5 personnes :

**E ( N ) =** $n\ G \times (1 - 0{,}9^{n}) + G = 100 \times (1 - 0{,}9^{5}) + 20 =$ **60,95**1

**Var ( N )** $= n^{2}\ G \times (1 - 0{,}9^{n}) \times 0{,}9^{n} = 500 \times (1 - 0{,}9^{5}) \times 0{,}9^{5} =$ **120,9**06

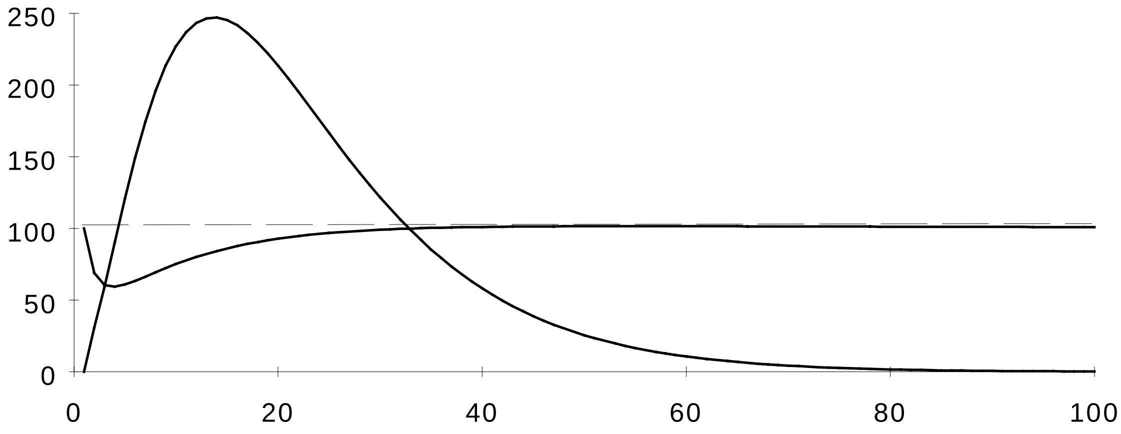
### c) Remarques.

- Lorsqu'on prend n G = 100, les formules donnant E ( N ) et Var ( N ) peuvent s'écrire en fonction du seul paramètre n (taille des groupes), pour n > 1 (pour n = 1 : E ( N ) = 100, Var ( N ) = 0) :

  $$\mathbf{E\ (N) =} 100 \times (1 - 0{,}9^{n}) + \frac{100}{n}$$

  $$\mathbf{Var\ (N) =} 100\ n \times (1 - 0{,}9^{n}) \times 0{,}9^{n}$$

- Ces formules permettent d'étudier la variation de E ( N ) et de Var ( N ) en fonction de la taille des groupes.




*Evolution du nombre moyen de tests E(N) et de la variance Var(N) du nombre de tests à effectuer en fonction de la taille des groupes constitués dans une population de 100 tests.*

# Exercice 21.

On sait que la probabilité pour qu'une personne soit allergique à un certain médicament est égale à $10^{-3}$. On s'intéresse à un échantillon de 1000 personnes. On appelle $X$ la variable aléatoire dont la valeur est le nombre de personnes allergiques dans l'échantillon.
1 - Déterminer, en la justifiant, la loi de probabilité de $X$.
2 - En utilisant une approximation que l'on justifiera, calculer les probabilités des évènements suivants :
a) Il y a exactement deux personnes allergiques dans l'échantillon.
b) Il y a au moins deux personnes allergiques dans l'échantillon.

# SOLUTION.

## 1°/ Loi de probabilité de $X$.

Pour une personne de l'échantillon, de deux choses l'une : ou bien elle est allergique (probabilité $p = 10^{-3}$ ), ou bien elle ne l'est pas (probabilité $q = 1 - p = 0{,}999$ ). L'épreuve qui consiste à regarder, pour une personne de l'échantillon si elle allergique ou non, constitue une ***épreuve de Bernoulli***. Lorsqu'on répète cette épreuve de Bernoulli pour les n = 1000 individus de l'échantillon, le nombre X de succès, c'est-à-dire le nombre X de personnes allergiques, suit une ***loi binomiale*** B ( 1000 ; $10^{-3}$ ) :

$$P(X = k) = C_{1000}^{k}\ 10^{-3k}\,(0{,}999)^{1000-3k}.$$

## 2°/ Approximation de la loi de probabilité de $X$ par une loi de Poisson.

Les trois conditions pour que l'approximation par la loi de Poisson soit valable sont réalisées :

- $n = 1000$ est grand
- $p = 10^{-3}$ est petit
- $np = 1$ est de l'ordre de grandeur de l'unité, ni trop grand, ni trop petit.

On peut donc remplacer la loi binomiale $B$ ( 1000 ; $10^{-3}$ ) par une loi de Poisson de paramètre $\lambda = n\,p = 1$.

$$P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!} = e^{-1} \times \frac{1}{k!}$$

a) $P(X = 2) = \dfrac{1}{2e} = 0{,}1839$

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-1} - e^{-1} = 1 - \dfrac{2}{e} = 0{,}2642.$

# Exercice 22.

Le temps $T$ nécessaire à un rat pour parcourir un labyrinthe est une variable aléatoire dont la distribution de probabilité est donnée par :

| | | | | | | |
|---|---|---|---|---|---|---|
| $T$ (en secondes) | 2 | 3 | 4 | 5 | 6 | 7 |
| $P( T = t)$ | 0,1 | 0,1 | 0,3 | $p_5$ | 0,2 | 0,1 |

1°/ Compléter le tableau en calculant $p_5 = P ( T = 5 )$.

2°/ Calculer le temps moyen, la variance, l'écart-type.

3°/ Le rat est récompensé à l'aide d'un biscuit pour chaque seconde économisée sur un temps de parcours de 6 secondes (par exemple, s'il met 4 secondes, il reçoit 2 biscuits, mais, s'il met 6 secondes ou plus, il ne reçoit rien). Tabuler la loi de probabilité de la variable « Récompense » $R$ égale au nombre de biscuits reçus. Calculer la récompense moyenne du rat, ainsi que la variance et l'écart-type de $R$.

4°/ Supposons que le rat soit puni par un choc électrique dont la puissance augmente fortement avec le temps mis à parcourir le labyrinthe, soit un choc de $T^2$ volts pour un temps de $T$ secondes. Calculer la punition moyenne du rat.

5°/ Reprendre la question précédente en supposant que la punition soit un choc électrique dont la puissance varie avec le temps selon la formule $P = 10\ T + 5$ (la puissance du choc électrique est exprimée en volts). Calculer de deux façons différentes la punition moyenne ainsi que $Var ( P)$.

# SOLUTION.

## 1°/ Condition de normalisation.

La somme des probabilités est toujours égale à 1 :

$$P(T = 2) + P(T = 3) + P(T = 4) + P(T = 5) + P(T = 6)+ P(T = 7)= 1$$

C'est la « condition de normalisation ».
Il en résulte :

$$\begin{aligned} P(T = 5) &= 1 - ( P(T = 2) + P(T = 3) + P(T = 4) + P(T = 6)+ P(T = 7) ) \\ &= 1 - 0,8 \\ &= 0,2 \end{aligned}$$

$$\boxed{P(T = 5) = 0,2}$$

## 2°/ Espérance mathématique, variance, écart-type.

Par définition, $E(T) = \sum_i P(T = t_i)\ t_i$ . On a donc ici :

$$\begin{aligned} E(T) &= 2 \times 0,1 + 3 \times 0,1 + 4 \times 0,3 + 5 \times 0,2 + 6 \times 0,2 + 7 \times 0,1 \\ &= 0,2 + 0,3 + 1,2 + 1,0 + 1,2 + 0,7 \\ &= 4,6 \end{aligned}$$

$$\boxed{E(T) = 4,6 \text{ s}}$$

$Var (T) = \sum_i P(T = t_i ) \times (t_i - E(T))^2 = E(T^2 ) - (E(T))^2$ (formule de la variance). On en déduit :

$Var(T) = \sum_i P(T = t_i)\, t_i^2 - (4,6)^2$

$= 4 \times 0,1 + 9 \times 0,1 + 16 \times 0,3 + 25 \times 0,2 + 36 \times 0,2 + 49 \times 0,1 - 21,16$

$= 0,4 + 0,9 + 4,8 + 5,0 + 7,2 + 4,9 - 21,16$

$= 23,2 - 21,16$

$= 2,04$

$$\boxed{Var(T) = 2,04 \text{ s}^2}$$

L'écart-type est la racine carrée de la variance :

$$\boxed{\sigma_T = 1,43 \text{ s}}$$

## 3°/ Récompense : variable *R*.

**Valeurs de *R* en fonction des valeurs de *T* :**

| *T* | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| *R* | 4 | 3 | 2 | 1 | 0 | 0 |
| *P ( T )* | 0,1 | 0,1 | 0,3 | 0,2 | 0,2 | 0,1 |

**Loi de probabilité de la variable *R* :**

| *R* | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| *P ( R )* | 0,3 | 0,2 | 0,3 | 0,1 | 0,1 |

Cette loi de probabilité permet de calculer les paramètres de la variable *R* :

$E(R) = 0 \times 0,3 + 1 \times 0,2 + 2 \times 0,3 + 3 \times 0,1 + 4 \times 0,1$

$= 0,2 + 0,6 + 0,3 + 0,4$

$= 1,5$

$$\boxed{\text{Le rat reçoit, en moyenne, 1,5 biscuits.}}$$

$Var(R) = E(R^2) - (E(R))^2$

$= 0 \times 0,3 + 1 \times 0,2 + 4 \times 0,3 + 9 \times 0,1 + 16 \times 0,1 - (1,5)^2$

$= 0,2 + 1,2 + 0,9 + 1,6 - 2,25$

$= 3,90 - 2,25$

$= 1,65$

$$\boxed{Var(R) = 1,65 \text{ biscuits}^2}$$

$\sigma_R = \sqrt{Var(R)}$

$= \sqrt{1,65}$

$= 1,28$ biscuits

$$\boxed{\sigma_R = 1,28 \text{ biscuits}}$$

## 4°/ Punition : variable $P_1$.

**Valeurs de $P_1$ en fonction des valeurs de *T* :**

| *T* | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| $P_1 = T^2$ | 4 | 9 | 16 | 25 | 36 | 49 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $P(T)$ | 0,1 | 0,1 | 0,3 | 0,2 | 0,2 | 0,1 |

Ce tableau donne aussi la **Loi de probabilité de la variable $P_1$ .**
Cette loi de probabilité permet de calculer les paramètres de la variable $P_1$ :
$E(P_1) = E(T^2)$
$= 23{,}2$ (déjà calculé dans la première question)

$$\boxed{E(P_1) = 23{,}2 \text{ volts}}$$

# 5°/ Punition : variable $P_2$.

**Valeurs de $P_2$ en fonction des valeurs de $T$ :**

| $T$ | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| $P_2 = 10\,T + 5$ | 25 | 35 | 45 | 55 | 65 | 75 |
| $P(T)$ | 0,1 | 0,1 | 0,3 | 0,2 | 0,2 | 0,1 |

Ce tableau donne aussi la **Loi de probabilité de la variable $P_2$ .**
Cette loi de probabilité permet de calculer les paramètres de la variable $P_2$ :

$$\begin{aligned} E(P_2) &= 25 \times 0{,}1 + 35 \times 0{,}1 + 45 \times 0{,}3 + 55 \times 0{,}2 + 65 \times 0{,}2 + 75 \times 0{,}1 \\ &= 2{,}5 + 3{,}5 + 13{,}5 + 11{,}0 + 13{,}0 + 7{,}5 \\ &= 51{,}0 \text{ volts} \end{aligned}$$

$$\boxed{E(P_2) = 51 \text{ V}}$$

$$\begin{aligned} Var(P_2) &= E(P_2^2) - (E(P_2))^2 \\ &= 25^2 \times 0{,}1 + 35^2 \times 0{,}1 + 45^2 \times 0{,}3 + 55^2 \times 0{,}2 + 65^2 \times 0{,}2 + 75^2 \times 0{,}1 - 51^2 \\ &= 625 \times 0{,}1 + 1225 \times 0{,}1 + 2025 \times 0{,}3 + 3025 \times 0{,}2 + 4225 \times 0{,}2 + 5625 \times 0{,}1 - 2601 \\ &= 62{,}5 + 122{,}5 + 607{,}5 + 605{,}0 + 845{,}0 + 562{,}5 - 2601 \\ &= 2805 - 2601 \\ &= 204 \end{aligned}$$

$$\boxed{Var(P_2) = 204 \text{ V}^2}$$

Comme la variable aléatoire $P_2$ est une fonction linéaire de la variable aléatoire $T$ :

$$P_2 = 10\,T + 5$$

on peut calculer les paramètres de $P_2$ en fonction des paramètres, déjà calculés de $T$. Pour cela on utilise les formules, faciles à démontrer :
$E(a\,T + b) = a\,E(T) + b$
$Var(a\,T + b) = a^2\,Var(T)$
Ces formules donnent ici :

$$\begin{aligned} E(P_2) &= E(10\,T + 5) \\ &= 10\,E(T) + 5 \\ &= 10 \times 4{,}6 + 5 \\ &= 51 \text{ V} \end{aligned}$$

$$\begin{aligned} Var(P_2) &= Var(10\,T + 5) \\ &= 10^2\,Var(T) \\ &= 100\,Var(T) \\ &= 100 \times 2{,}04 \\ &= 204 \end{aligned}$$

On retrouve, bien plus facilement, les résultats obtenus directement à partir de la loi de probabilité.

# Exercice 23.

Les centres de transfusion sanguine diffusent un tableau donnant la répartition en France des principaux groupes sanguins :

| | O | A | B | AB |
|---|---|---|---|---|
| Rhésus + | 37,0 % | 38,1 % | 6,2 % | 2,8 % |
| Rhésus – | 7,0 % | 7,2 % | 1,2 % | 0,5 % |

1°/ Dix personnes prises au hasard en France donnent leur sang. Soit $X$ la variable aléatoire qui prend pour valeur le nombre de personnes appartenant au groupe A. On demande de calculer :

$E\,(\,X\,)$ ; $Var\,(\,X\,)$ ; $P\,(\,X = 4\,)$ ; $P\,(\,X = 0\,)$ ; $P\,(\,X > 0\,)$.

2°/ Pour une intervention chirurgicale, on doit avoir au moins 3 donneurs de groupe O et de rhésus +. Dix personnes ignorant leur groupe sanguin sont disposées à ce don. Calculer la probabilité d'avoir au moins les donneurs nécessaires parmi les dix volontaires.

# SOLUTION.

## 1°/ Loi de probabilité de *X*.

Pour une personne choisie au hasard en France, de deux choses l'une :

- ou bien elle est de groupe A (probabilité 0,381 + 0,072 = 0,453)
- ou bien elle n'est pas de groupe A (probabilité 1 − 0,453 = 0,547)

Ainsi l'épreuve qui consiste à regarder si une personne choisie au hasard en France est de groupe A ou n'est pas de groupe A est une épreuve de Bernoulli, avec deux résultats possibles : être ou ne pas être de groupe A, les probabilités étant toujours les mêmes.

Quand on répète cette épreuve de Bernoulli 10 fois, le nombre de sussès de l'épreuve, c'est-à-dire le nombre de personnes de groupe A parmi les 10 personnes choisies au hasard, suit une loi binomiale de paramètres $n = 10$ et $p = 0,453$. La loi de probabilité de $X$ est donc donnée par la formule de la **loi binomiale** :

$$P\,(\,X = k\,) = B_{(\,n\,;\,p\,)}\,(\,k\,) = C_n^k\,p^k\,(\,1-p\,)^{n-k}$$

$$\boxed{P\,(\,X = k\,) = C_{10}^k\,(0,453)^k \times (1-0,453)^{10-k}}$$

où $C_{10}^k$ est le coefficient binomial d'indices 10 et $k$, donné par :

$$C_{10}^k = \frac{10!}{k\,!(10-k\,)!}$$

Pour la loi binomiale d'indices $n$ et $p$, le paramètres de la variable aléatoire sont :

$$\boxed{\begin{array}{c} E\,(\,X\,) = n\,p \\ Var\,(\,X\,) = n\,p\,(\,1-p\,). \end{array}}$$

Nous obtenons donc ici :

$E\,(\,X\,) = 10 \times 0,453 = 4,53$

$Var\,(\,X\,) = 4,53 \times 0,547 = 2,48$

$P\,(\,X = 4\,) = C_{10}^4 \times 0,453^4 \times 0,547^6$

$= 210 \times 0,042 \times 0,027$

$= 0,237$

$P\,(\,X = 0\,) = 0,547^{10}$

$= 2,4 \times 10^{-3}$

$P\,(\,X > 0\,) = 1 - P\,(\,X = 0\,)$

$= 1 - 2,4 \times 10^{-3}$

$= 0,998$

## 2°/ Probabilité d'avoir 3 O+ dans un groupe de 10.

Pour une personne choisie au hasard, la probabilité d'être du groupe O et de rhésus positif est, d'après le tableau, de 0,37. Nous supposerons que les dix volontaires constituent un échantillon aléatoire, c'est-à-dire constitué au hasard. Cette hypothèse ne va pas de soi, même si les volontaires ignorent leur groupe sanguin, car les volontaires viennent peut-être de la famille.

Sous cette hypothèse, le nombre $X$ de O+ dans un groupe de 10 suit une loi binomiale de paramètres $n = 10$ et $p = 0{,}37$. Et l'on a :

$$\begin{aligned} P(X \geq 3) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - 0{,}63^{10} - 10 \times 0{,}37 \times 0{,}63^{9} - 45 \times 0{,}37^{2} \times 0{,}63^{8} \\ &= 0{,}706 \end{aligned}$$

$$\boxed{P(X \geq 3) = 0{,}706}$$

Si les volontaires ont été recrutés parmi les membres de la famille, on peut penser que la probabilité d'avoir des gens de groupe O+ sera encore plus forte.

# Exercice 24.

On considère deux portées de lapereaux d'une race donnée : la première portée est de 6 lapereaux et la deuxième de 8 lapereaux. On regroupe tous les lapereaux et on suppose que chacun a la même probabilité d'être choisi.

1°/ On prélève simultanément 4 lapereaux. Calculer les probabilités d'obtenir :

- 3 lapereaux de la première portée ;
- 1 lapereau de la deuxième portée ;
- au plus un lapereau de la deuxième portée.

2°/ Donner le nombre moyen attendu de lapereaux de la première portée.

# SOLUTION.

Soit $X$ le nombre de lapereaux de la première portée, tirés dans un échantillon de $n = 4$ lapereaux tirés simultanément. Soit $Y$ le nombre de lapereaux de la deuxième portée, tirés dans un échantillon de $n = 4$ lapereaux tirés simultanément.

## 1°/ Loi hypergéométrique.

La probablité pour que $X = k$ est le rapport du nombre de cas favorables au nombre de cas possibles. Le nombre de cas possibles est le nombre d'échantillons de $n = 4$ lapereaux que l'on peut constituer avec

$$N = n_1 + n_2 = 6 + 8 = 14 \text{ lapereaux :}$$

c'est $C_N^n = C_{14}^4 = 1\,001$.

Il y a $C_{n_1}^k = C_6^3$ façons de choisir $k = 3$ lapereaux de la première portée, et $C_{n_2}^{n-k} = C_8^1$ façons de choisir le 4e lapereau dans la 2e portée. Le nombre de cas favorables est donc :

$$C_{n_1}^k \times C_{n_2}^{n-k} = C_6^3 \times C_8^1 = 20 \times 8 = 160$$

et la probabilité d'avoir exactement $k = 3$ lapereaux de la première portée dans le tirage effectué est :

$$P\,(\,X = k\,) = \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n} = \frac{C_6^3 \times C_8^1}{C_{14}^4} = \frac{160}{1001} = 0{,}1598$$

$$\boxed{P\,(\,X = 3\,) = 0{,}1598}$$

L'évènement $Y = 1$ est le même que l'évènement $X = 3$. Les probabilités sont donc les mêmes :

$$\boxed{P\,(\,Y = 1\,) = P\,(\,X = 3\,) = 0{,}1598}$$

Enfin l'évènement « Au plus un lapereau de la deuxième portée » est la réunion des deux évènements disjoints « 3 lapereaux de la 1e portée et 1 lapereau de la 2e portée » et « 4 lapereaux de la 1e portée et 0 lapereau de la 2e portée ». La probabilité de cette réunion est la somme des probabilités (axiome des probabilités totales) :

$$\begin{aligned}
P\,(\,Y \le 1) &= P\,(\,Y = 1\,) + P\,(\,Y = 0\,)\\
&= P\,(\,X = 3\,) + P\,(\,X = 4\,)\\
&= \frac{C_6^3 \times C_8^1}{C_{14}^4} + \frac{C_6^4 \times C_8^0}{C_{14}^4}\\
&= \frac{160}{1001} + \frac{15}{1001}\\
&= \frac{175}{1001}\\
&= 0{,}1748
\end{aligned}$$

$$\boxed{P\,(\,Y \le 1\,) = 0{,}1748}$$

## 2°/ Paramètres de la loi hypergéométrique.

$$E(X) = \sum_{k=0}^{k=n} k \times \frac{C_{n_1}^{k} C_{n_2}^{n-k}}{C_{n_1+n_2}^{n}}$$

On démontre que cette quantité est donnée par la formule :

$$E(X) = n\,p = n \times \frac{n_1}{n_1 + n_2}$$

où $p = \dfrac{n_1}{n_1 + n_2}$ est la probabilité de tirer un lapereau de la première portée quand on tire au hasard un lapereau dans le lot des 14 lapereaux. Ici, nous obtenons numériquement :

$$E(X) = 4 \times \frac{6}{14} = \frac{12}{7} = 1{,}7143$$

$$\boxed{E(X) = 1{,}7143 \text{ lapereaux}}$$

La variance de $X$ est donnée par :

$$Var(X) = \frac{N-n}{N-1}\, n\,p\,q$$

où $N = n_1 + n_2$ et $q = 1 - p$. Ici, elle vaut :

$$Var(X) = \frac{14-4}{14-1} \times 4 \times \frac{6}{14} \times \frac{8}{14} = \frac{10}{13} \times \frac{48}{49} = \frac{480}{637} = 0{,}7535$$

On considère que cette variance est pratiquement égale à $n\,p\,q$ dès que le « taux de sondage » $n/N$ est inférieur à 0,1. Dans ce cas, on peut approcher la loi hypergéométrique par une loi binomiale de paramètres $n$ et $p$. Cette approximation revient en fait à considérer que le tirage se fait avec remise et que la probabilité de tirage reste la même pour tous.

# Exercice 25.

On considère qu'il apparaît, en moyenne, chaque année 20 nouveaux cas de poliomyélite pour 100 000 habitants. On appelle $X$ le nombre de nouveaux cas de poliomyélite dans l'année, dans une ville de 50 000 habitants (respectivement 5 000 habitants).

1°/ Montrer que la variable X suit une loi binomiale.

2°/ Montrer que l'on peut faire ici l'approximation de la loi binomiale par une loi de Poisson.

3°/ Calculer la probabilité pour qu'il apparaisse au plus quatre nouveaux cas de poliomyélite :

a) dans une ville de 50 000 habitants.

b) dans une ville de 5 000 habitants.

# SOLUTION.

## 1°/ Loi de probabilité de la variable aléatoire *X*.

Pour un individu, la probabilité de contracter la poliomyélite dans un délai d'un an est

$$p = \frac{20}{100.000} = 2,0 \times 10^{-4}.$$

Considérons l'épreuve suivante : on regarde si un individu a contracté la poliomyélite dans l'année qui précède l'observation. Cette épreuve a deux résultats possibles

– le succès : probabilité $p = 2,0 \times 10^{-4}$

– l'échec : probabilité $q = 1 - p$.

C'est une épreuve de Bernoulli. Lorsqu'on recommence $n$ fois cette épreuve de Bernoulli, pour tous les habitants d'une ville, le nombre de succès est le nombre $X$ de cas de poliomyélite apparus dans la ville dans l'année précédant l'observation. La variable aléatoire $X$ suit une loi binomiale de paramètres $n$ et $p$.

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

## 2°/ Approximation par la loi de Poisson.

Pour qu'on puisse approcher une loi binomiale de paramètres $n$ et $p$ par une loi de Poisson de paramètre μ, il faut que le paramètre μ représente à la fois l'espérance mathématique $n\,p$ et la variance $n\,p\,q$. Il faut donc déjà que $q$ soit voisin de 1. C'est la première condition : $p$ doit être petit.

De plus, pour que les probabilités

$$C_n^k\, p^k\, q^{n-k} = p^k \times C_n^k\, q^{n-k} = \frac{p^k}{k!} \times \frac{n!}{(n-k)!} \times (1-p)^{n-k} = \frac{p^k}{k!} \times \frac{n!}{(n-k)!} \times \left(1 - \frac{\mu}{n}\right)^{n-k}$$

$$\text{et } e^{-\mu}\,\frac{\mu^k}{k!} = p^k \times e^{-\mu}\,\frac{n^k}{k!} = \frac{p^k}{k!} \times n^k \times e^{-\mu}$$

soient voisines, il faut que $k$ soit négligeable devant $n$, au moins pour les termes intéressants, c'est-à-dire pour les termes entourant la moyenne $\mu = n\,p$. C'est le cas si $n$ est grand et si le produit $n\,p$ reste de l'ordre d'au plus quelques unités, car, dans ce cas, on a :

$$\frac{n!}{(n-k)!} \approx n^k \quad \text{et} \quad \left(1 - \frac{\mu}{n}\right)^{n-k} \approx e^{-\mu}$$

En résumé, l'approximation de la loi binomiale par la loi de Poisson est justifée si :

- $p$ est petit
- $n$ est grand
- $n\,p$ est de l'ordre de grandeur de l'unité

En pratique, on considère que l'approximation est justifiée pour

- $p < 0,1$
- $n > 30$
- $n\,p$ de l'ordre de quelques unités

Dans le cas présent, nous avons :

$p = 2 \times 10^{-4} < 0{,}1$
$n$ = 50 000 ou $n$ = 5 000. Les deux valeurs sont supérieures à 30
$n\,p$ = 10 ou 1.
On peut donc considérer que, dans les deux cas, l'approximation par la loi de Poisson de paramètre $\mu = n\,p$ est justifiée.

## 3°/ Ville de 50 000 habitants.

La probabilité pour qu'il apparaisse au plus 4 cas de poliomyélite dans la ville dans l'année à venir est :
$P(X \le 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$$\approx e^{-10} + e^{-10} \times 10 + e^{-10} \times \frac{100}{2} + e^{-10} \times \frac{1000}{6} + e^{-10} \times \frac{10000}{24}$$

$$\approx 4{,}5 \times 10^{-5} + 4{,}5 \times 10^{-4} + 2{,}3 \times 10^{-3} + 7{,}6 \times 10^{-3} + 1{,}9 \times 10^{-2} = 2{,}9 \times 10^{-2}$$

$$\boxed{P(X \le 4) \approx 2{,}9 \times 10^{-2}}$$

## 2°/ Ville de 5 000 habitants.

La probabilité pour qu'il apparaisse au plus 4 cas de poliomyélite dans la ville dans l'année à venir est :
$P(X \le 4) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$

$$\approx e^{-1} + e^{-1} + e^{-1} \times \frac{1}{2} + e^{-1} \times \frac{1}{6} + e^{-1} \times \frac{1}{24}$$

$$\approx \frac{65}{24 \times e} = 0{,}9963$$

$$\boxed{P(X \le 4) = 0{,}9963}$$

# DISTRIBUTIONS CONJOINTES

# Exercice 26.

Soient $X$ et $Y$ deux variables aléatoires ***indépendantes*** de lois :

| $x$ | 1 | 2 |
|---|---|---|
| $P(X = x)$ | $\frac{1}{2}$ | $\frac{1}{2}$ |

| $y$ | 1 | 2 | 3 |
|---|---|---|---|
| $P(Y = y)$ | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

1°/ Calculer $E(X)$ ; $Var(X)$ ; $E(Y)$ ; $Var(Y)$.
2°/ Donner la loi de probabilité conjointe de $X$ et $Y$.
3°/ Etablir la loi de probabilité de $T = X + Y$. Calculer $E(T)$ et $Var(T)$.
4°/ Etablir la loi de probabilité de $Z = X\,Y$. Calculer $E(Z)$ et $Var(Z)$.
5°/ Vérifier par le calcul que $Cov(X,Y) = 0$.

# SOLUTION.

## 1°/ Paramètres des lois de probabilités de X et de Y.

$E(X) = \sum_i x_i \, P(X = x_i) = (1 + 2) \times \frac{1}{2} = \frac{3}{2} = 1{,}5.$

$V(X) = E(X^2) - [E(X)]^2 = (1^2 + 2^2) \times \frac{1}{2} - \left(\frac{3}{2}\right)^2 = \frac{1}{4}.$

$E(Y) = \sum_i y_i \, P(Y = y_i) = 1 \times \frac{1}{4} + 2 \times \frac{1}{2} + 3 \times \frac{1}{4} = 2.$

$V(Y) = E(Y^2) - [E(Y)]^2 = 1^2 \times \frac{1}{4} + 2^2 \times \frac{1}{2} + 3^2 \times \frac{1}{4} - 2^2 = \frac{1}{2}.$

## 2°/ Loi de probabilité conjointe de X et Y.

Comme les variables X et Y sont supposées indépendantes, pour chaque modalité $x_i$ et chaque modalité $y_j$ , la probabilité de l'évènement « $X = x_i$ et $Y = y_j$ » est égale au produit des probabilités des évènements « $X = x_i$ » et « $Y = y_j$ » : $\mathbf{P(X = x_i \text{ et } Y = y_j) = P(X = x_i) \times P(Y = y_j)}$. La probabilité conjointe est le produit des probabilités marginales de X et Y. On obtient ainsi le tableau suivant :

| X \ Y | 1 | 2 | 3 | P ( X = x ) |
|---|---|---|---|---|
| 1 | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{2}$ |
| 2 | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{8}$ | $\frac{1}{2}$ |
| P ( Y = y ) | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ | 1 |

## 3°/ Loi de probabilité de la somme T = X + Y.

La somme T peut prendre les valeurs contenues dans le tableau suivant :

| X \ Y | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 2 | 3 | 4 | 5 |

$P ( T = 2 ) = P ( X = 1 \text{ et } Y = 1 ) = \frac{1}{8}$

$P ( T = 3) = P ( X = 1 \text{ et } Y = 2 ) + P ( X = 2 \text{ et } Y = 1 ) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$

$P ( T = 4 ) = P ( X = 1 \text{ et } Y = 3 ) + P ( X = 2 \text{ et } Y = 2 ) = \frac{1}{8} + \frac{1}{4} = \frac{3}{8}$

$P ( T = 5 ) = P ( X = 2 \text{ et } Y = 3 ) = \frac{1}{8}$

Ces valeurs donnent le tableau suivant de la loi de probabilité de T :

| t | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| P ( T = t ) | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |

La loi de probabilité de T permet de calculer les paramètres :

$$E(T) = \sum_i t_i \; P (T = t_i ) = 2 \times \frac{1}{8} + 3 \times \frac{3}{8} + 4 \times \frac{3}{8} + 5 \times \frac{1}{8} = \frac{28}{8} = \frac{7}{2} = 3{,}5$$

$$V(T) = E(T^2) - [ E(T) ]^2 = 2^2 \times \frac{1}{8} + 3^2 \times \frac{3}{8} + 4^2 \times \frac{3}{8} + 5^2 \times \frac{1}{8} - \left(\frac{7}{2}\right)^2 = \frac{3}{4} = 0{,}75.$$

***Remarques :***

- $E(T) = E(X + Y) = E(X ) + E(Y) = \frac{3}{2} + 2 = \frac{7}{2}$
- La loi de probabilité de T est symétrique, donc la moyenne est la valeur centrale de la distribution.
- Comme les variables X et Y sont indépendantes, on a $V(T) = V(X + Y) = V(X) + V(Y) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$.

## 4°/ Loi de probabilité du produit Z = X Y.

On peut mettre dans un tableau les évènements élémentaires avec leurs probabilités et les valeurs de Z correspondantes :

| X \ Y | 1 | 2 | 3 |
|---|---|---|---|
| 1 | Z = 1<br>$P ( 1 ; 1 ) = \frac{1}{8}$ | Z = 2<br>$P ( 1 ; 2 ) = \frac{1}{4}$ | Z = 3<br>$P ( 1 ; 3 ) = \frac{1}{8}$ |
| 2 | Z = 2<br>$P ( 2 ; 1 ) = \frac{1}{8}$ | Z = 4<br>$P ( 2 ; 2 ) = \frac{1}{4}$ | Z = 6<br>$P ( 2 ; 3 ) = \frac{1}{8}$ |

On déduit de ce tableau la loi de probabilité de Z en recensant les valeurs de Z et les probabilités élémentaires correspondantes :

| z | 1 | 2 | 3 | 4 | 6 |
|---|---|---|---|---|---|
| P ( Z = z ) | $\frac{1}{8}$ | $\frac{1}{8} + \frac{1}{4} = \frac{3}{8}$ | $\frac{1}{8}$ | $\frac{1}{4}$ | $\frac{1}{8}$ |

Ce tableau permet de calculer les paramètres de la loi :

$$E(Z) = 1 \times \frac{1}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} + 4 \times \frac{1}{4} + 6 \times \frac{1}{8} = 3$$

$$V(Z) = E(Z^2) - [\, E(Z)\, ]^2 = 1^2 \times \frac{1}{8} + 2^2 \times \frac{3}{8} + 3^2 \times \frac{1}{8} + 4^2 \times \frac{1}{4} + 6^2 \times \frac{1}{8} - 3^2 = \frac{9}{4} = 2{,}25.$$

## 5°/ Calcul de la covariance de X et Y.

### a) Première méthode.

La formule

$$V\,(\,X + Y\,) = V\,(\,X\,) + V\,(\,Y\,) + 2\,\mathrm{Cov}\,(\,X\,,\,Y\,),$$

jointe à la remarque $V\,(\,X + Y\,) = V\,(\,X\,) + V\,(\,Y\,)$ du 3°/, montre que l'on a :

$$\mathrm{Cov}\,(\,X\,,\,Y\,) = 0$$

### b) Deuxième méthode.

La formule

$$\mathrm{Cov}\,(\,X\,,\,Y\,) = E\,(\,X\,Y\,) - E\,(\,X\,) \times E\,(\,Y\,)$$

donne, avec les valeurs numériques obtenues :

$$\mathrm{Cov}\,(\,X\,,\,Y\,) = 3 - \frac{3}{2} \times 2 = 0$$

# Exercice 27.

La loi du couple ( X , Y ) est définie par le tableau suivant :

| X \ Y | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 0 | $\frac{1}{2}$ | 0 |
| 2 | $\frac{1}{4}$ | 0 | $\frac{1}{4}$ |

1°/ Trouver les lois marginales de X et Y.
2°/ Montrer que X et Y ne sont pas indépendantes.
3°/ Calculer la covariance de X et Y et le coefficient de corrélation linéaire ρ.

# SOLUTION.

## 1°/ Lois marginales de X et Y.

On obtient les lois de probabilités marginales de X et de Y en faisant les sommes par lignes ou par colonnes des probabilités élémentaires du tableau de probabilités conjointes :

| x | 1 | 2 |
|---|---|---|
| P ( X = x ) | $\frac{1}{2}$ | $\frac{1}{2}$ |

| y | 1 | 2 | 3 |
|---|---|---|---|
| P ( Y = y ) | $\frac{1}{4}$ | $\frac{1}{2}$ | $\frac{1}{4}$ |

Ces tableaux permettent de calculer les paramètres des distributions de X et Y :

$E ( X) = \frac{3}{2}$ (loi symétrique) ; $Var ( X ) = \frac{1}{4}$ ; $E ( Y ) = 2$ (loi symétrique) ; $Var ( Y ) = \frac{1}{2}$.

## 2°/ Les variables X et Y ne sont pas indépendantes.

On a, par exemple : $P ( X = 1 \text{ et } Y = 1 ) = 0 \neq P ( X = 1 ) \times P ( Y = 1 ) = \frac{1}{8}$

## 3°/ Covariance et coefficient de corrélation linéaire de X et Y.

La covariance Cov ( X , Y ) de X et de Y se calcule par la formule $Cov ( X , Y ) = E ( X Y ) – E ( X ) E ( Y )$. Comme on connaît déjà E ( X ) et E ( Y ), il suffit de calculer E ( X Y ). Pour cela, étudions la distribution du produit Z = X Y.

| X \ Y | 1 | 2 | 3 |
|---|---|---|---|
| 1 | Z = 1<br>P ( 1 ; 1 ) = 0 | Z = 2<br>P ( 1 ; 2 ) = $\frac{1}{2}$ | Z = 3<br>P ( 1 ; 3 ) = 0 |
| 2 | Z = 2<br>P ( 2 ; 1 ) = $\frac{1}{4}$ | Z = 4<br>P ( 2 ; 2 ) = 0 | Z = 6<br>P ( 2 ; 3 ) = $\frac{1}{4}$ |

| Z | 1 | 2 | 3 | 4 | 6 |
|---|---|---|---|---|---|
| P ( Z = z ) | 0 | $\frac{1}{4} + \frac{1}{2} = \frac{3}{4}$ | 0 | 0 | $\frac{1}{4}$ |

Ce tableau permet le calcul des paramètres de la distribution de Z, notamment E ( Z ) = E ( X Y ) = 3.

La formule Cov ( X , Y ) = E ( X Y ) – E ( X ) E ( Y ) montre alors que l'on a Cov ( X , Y ) = $3 - \frac{3}{2} \times 2 = 0$.
Et cela permet de vérifier que X et Y peuvent ne pas être indépendants, même si leur covariance est nulle :

**Cov ( X , Y ) = 0 n'entraîne pas « X et Y indépendants ».**

Le **coefficient de corrélation linéaire ρ** est lié à la covariance par la formule :

$$\rho = \frac{Cov\ (X, Y)}{\sigma_X\ \sigma_Y} = 0$$

# Exercice 28.

Pour essayer de savoir si le « bonheur » dépend du niveau d'instruction, on a demandé aux individus d'une population donnée de se classer eux-mêmes en fonction de l' *indice bonheur* '$X$' suivant :

$X = -1$ s'ils s'estiment malheureux
$X = 0$ s'ils s'estiment moyennement heureux
$X = 1$ s'ils s'estiment heureux
$X = 2$ s'ils s'estiment très heureux

En même temps, on a relevé, pour chaque individu, le *degré d'instruction* '$Y$' :

$Y = -1$ s'ils n'ont aucun diplôme
$Y = 0$ s'ils ont le niveau du baccalauréat
$Y = 1$ s'ils ont un diplôme supérieur

La loi de probabilité conjointe des deux variables aléatoires $X$ et $Y$ dans la population considérée est la suivante :

| $X$ : indice bonheur / $Y$ : degré d'instruction | $-1$ | $0$ | $1$ | $2$ |
|---|---|---|---|---|
| $-1$ | 0, 01 | 0, 07 | 0, 01 | 0, 00 |
| $0$ | 0, 00 | 0, 12 | 0, 30 | 0, 10 |
| $1$ | 0, 35 | 0, 04 | 0, 00 | 0, 00 |

1°/ Calculer les probabilités suivantes :

$P ( X = 1 )$ **Réponse : 1**

$P ( X = 0 \text{ et } Y \geq 0 )$ **Réponse : 2**

$P ( X = 0 \mid Y \geq 0 )$ **Réponse : 3**

2°/ Calculer : $E(X)$ **Réponse : 4** $Var (X)$ **Réponse : 5**

$E ( X Y )$ **Réponse : 6** $Cov (X , Y)$ **Réponse : 7**

3°/ Cocher les propositions exactes :

**A** : L'indice bonheur '$X$' est indépendant du degré d'instruction '$Y$'.
**B** : L'indice bonheur '$X$' et le degré d'instruction '$Y$' ne sont pas indépendants.
**C** : $Cov ( X , Y ) = 0 \Leftrightarrow$ Les deux variables $X$ et $Y$ sont indépendantes
**D** : $Cov ( X , Y ) = 0 \Rightarrow$ Les deux variables $X$ et $Y$ sont liées.

**Réponse : 8**

( *Tableau des réponses de l'Exercice* 3 : Réponses 1 à 7)

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | – 0,99 | – 0,81 | – 0,55 | – 0,40 | – 0,29 | – 0,10 | 0,10 | 0,15 | 0,18 | 0,31 | 0,39 | 0,51 | 0,63 | 0,87 | 0,99 |
| – 1,00 | – 0,85 | – 0,62 | – 0,49 | – 0,35 | – 0,15 | 0,00 | 0,12 | 0,16 | 0,25 | 0,35 | 0,44 | 0,59 | 0,75 | 0,91 | 1,00 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | D | AB | AD | BC | BE | CE | ABC | ABE | ACE | BCD | BDE | ABCD | ABDE | BCDE |
| A | C | E | AC | AE | BD | CD | DE | ABD | ACD | ADE | BCE | CDE | ABCE | ACDE | ABCDE |

## 1°/ Calcul de diverses probabilités.

Commençons par compléter le tableau de données par les probabilités marginales.

| $X$ : indice bonheur / $Y$ : degré d'instruction | – 1 | 0 | 1 | 2 | Total |
|---|---|---|---|---|---|
| – 1 | 0, 01 | 0, 07 | 0, 01 | 0, 00 | 0, 09 |
| 0 | 0, 00 | 0, 12 | 0, 30 | 0, 10 | 0, 52 |
| 1 | 0, 35 | 0, 04 | 0, 00 | 0, 00 | 0, 39 |
| Total | 0, 36 | 0, 23 | 0, 31 | 0, 10 | 1, 00 |

Sur ce tableau, on voit apparaître la probabilité marginale $P\ (X = 1) = 0,31$ **Réponse 1 : ACE**

Dans la colonne $X = 0$, la somme des probabilités pour que $Y = 0$ ou $Y = 1$ est 0,16 **Réponse 2 : ABD**

$$P\ (X = 0 \mid Y \geq 0) = \frac{P(X = 0\ et\ Y \geq 0)}{P(Y \geq 0)} = \frac{P(X = 0\ et\ Y = 0) + P(X = 0\ et\ Y = 1)}{P(Y = 0) + P(Y = 1)} = \frac{0{,}12 + 0{,}04}{0{,}52 + 0{,}39} = \frac{16}{91} = 0{,}176$$

**Réponse 3 : ABE**

## 2°/ Espérance mathématique, variance, covariance.

$$E(X) = \sum_i x_i\ P(X = x_i) = -0{,}36 + 0{,}31 + 2 \times 0{,}10 = 0{,}15$$

**Réponse 4 : ABC**

$$Var(Y) = E(Y^2) - (E(Y))^2 = \sum_i y_i^2\ P(Y = y_i) - \left(\sum_i y_i\ P(Y = y_i)\right)^2$$

$$= 0{,}09 + 0{,}39 - (-0{,}09 + 0{,}39)^2 = 0{,}39$$

**Réponse 5 : BCD**

$$E(X\,Y) = \sum_i x_i\ y_i\ P(X = x_i \text{ et } Y = y_i)$$

$$= 0{,}01 - 0{,}35 - 0{,}01 = -0{,}35$$

**Réponse 6 : AE**

$$Cov(X,Y) = E(XY) - E(X)\,E(Y) = -0{,}35 - 0{,}15 \times 0{,}30 = -0{,}405$$

**Réponse 7 : AD**

## 3°/ Indépendance en probabilité des deux variables.

La relation $0{,}36 \times 0{,}09 \neq 0{,}01$ montre que l'on a : $P\ (X = -1) \times P\ (Y = -1) \neq P\ (X = -1 \text{ et } Y = -1)$ et ceci suffit pour montrer que les deux variables $X$ et $Y$ ne sont pas indépendantes : **A** est faux, **B** est vrai.
On sait aussi que lorsque deux variables sont indépendantes, leur covariance est nulle, mais que la réciproque est fausse (voir, par exemple, l'exercice 2, précédent). Il en résulte que **C** est faux et **D** est vrai.

**Réponse 8 : BD**

# Exercice 29.

Soient $X$ et $Y$ deux variables aléatoires ***indépendantes*** de lois :

| $X$ | 20 | 40 |
|---|---|---|
| $P(X = x_i)$ | 0,25 | 0,75 |

| $Y$ | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| $P(Y = y_i)$ | 0,16 | 0,32 | 0,32 | 0,20 |

1°/ Calculer $E(X)$, $Var(X)$, $E(Y)$, $Var(Y)$.
2°/ Donner la loi de probabilité conjointe de $X$ et $Y$.
3°/ Etablir la loi de probabilité de la variable aléatoire $T = X + Y$ et calculer $E(T)$ et $Var(T)$.
4°/ Etablir la loi de probabilité de la variable aléatoire $Z = X.Y$ et calculer $E(Z)$ et $Var(Z)$.
5°/ Vérifier par le calcul que $Cov(X,Y) = 0$.

# SOLUTION.

## 1°/ Espérance mathématique et variance de *X* et *Y*.

$E(X) = 0,25 \times 20 + 0,75 \times 40 = 0,25 \times ( 20 + 3 \times 40 ) = 0,25 \times 140 = \dfrac{140}{4} = 35$

$Var(X) = 0,25 \times (20 - 35)^2 + 0,75 \times (40 - 35)^2 = 0,25 \times 125 + 0,75 \times 25 = 0,25 \times 25 \times (5 + 3) = 2 \times 25 = 50$

$E(Y) = 0,16 \times 10 + 0,32 \times 20 + 0,32 \times 40 + 0,20 \times 80 = 1,6 + 6,4 + 12,8 + 16 = 36,8$

$Var(Y) = 0,16 \times (10–36,8)^2 + 0,32 \times (20–36,8)^2 + 0,32 \times (40–36,8)^2 + 0,20 \times (80–36,8)^2$
$= 114,9184 + 90,3168 + 3,2768 + 373,2480 = 581,7600$

$$\boxed{E(X) = 35 \quad Var(X) = 50 \quad E(Y) = 36,8 \quad Var(Y) = 581,76}$$

## 2°/ Loi de probabilité conjointe de *X* et *Y*.

**Probabilités associées aux valeurs de *X* et *Y* :**

| X \ Y | 10 | 20 | 40 | 80 | Probabilités |
|---|---|---|---|---|---|
| 20 | 0,04 | 0,08 | 0,08 | 0,05 | 0,25 |
| 40 | 0,12 | 0,24 | 0,24 | 0,15 | 0,75 |
| Probabilités | 0,16 | 0,32 | 0,32 | 0,20 | 1,00 |

Ce tableau donne la loi conjointe de $X$ et $Y$. Comme les variables aléatoires sont indépendantes par hypothèse, chaque probabilité conjointe s'obtient par ***produit des probabilités marginales*** :

$$\boxed{P(X = x_i \text{ et } Y = y_j) = P(X = x_i) \times P(Y = y_j)}$$

## 3°/ Loi de probabilité de la variable aléatoire *T* = *X* + *Y*.

**Valeurs de *T* = *X* + *Y*.**

| X \ Y | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| 20 | 30 | 40 | 60 | 100 |
| 40 | 50 | 60 | 80 | 120 |

**Probabilités associées aux évènements élémentaires :**

| X \ Y | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| 20 | 0,04 | 0,08 | 0,08 | 0,05 |
| 40 | 0,12 | 0,24 | 0,24 | 0,15 |

**Loi de probabilité de *T* = *X* + *Y*.**

| $T$ | 30 | 40 | 50 | 60 | 80 | 100 | 120 |
|---|---|---|---|---|---|---|---|
| $P(T=t_i)$ | 0,04 | 0,08 | 0,12 | 0,32 | 0,24 | 0,05 | 0,15 |

$E(T) = E(X) + E(Y) = 35 + 36{,}8 = 71{,}8$

$Var(T) = Var(X) + Var(Y) = 50 + 581{,}76 = 631{,}76$

## 4°/ Loi de probabilité de la variable aléatoire *Z* = *X Y*.

**Valeurs de *Z* = *X Y*.**

| X \ Y | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| 20 | 200 | 400 | 800 | 1600 |
| 40 | 400 | 800 | 1600 | 3200 |

**Probabilités associées aux évènements élémentaires :**

| X \ Y | 10 | 20 | 40 | 80 |
|---|---|---|---|---|
| 20 | 0,04 | 0,08 | 0,08 | 0,05 |
| 40 | 0,12 | 0,24 | 0,24 | 0,15 |

**Loi de probabilité de *Z* = *X Y*.**

| $Z$ | 200 | 400 | 800 | 1600 | 3200 |
|---|---|---|---|---|---|
| $P(Z=z_i)$ | 0,04 | 0,20 | 0,32 | 0,29 | 0,15 |

$$E(Z) = 0{,}04 \times 200 + 0{,}20 \times 400 + 0{,}32 \times 800 + 0{,}29 \times 1600 + 0{,}15 \times 3200$$
$$= 8 + 80 + 256 + 464 + 480 = 1288$$

$$Var(Z) = 0{,}04\times(200-1288)^2 + 0{,}20\times(400-1288)^2 + 0{,}32\times(800-1288)^2 + 0{,}29\times(1600-1288)^2 + 0{,}15\times(3200-1288)^2$$
$$= 47349{,}76 + 157708{,}80 + 76206{,}08 + 28229{,}76 + 548361{,}60 = 857856$$

## 5°/ Calcul de la covariance de *X* et *Y*.

$$Cov(X,Y) = E(X\,Y) - E(X)\,E(Y) = E(Z) - E(X)\,E(Y) = 1288 - 35 \times 36{,}8 = 0$$

# Exercice 30.

Dans certains accidents de la route, les chocs peuvent être latéraux ou frontaux : ces deux états sont résumés par la variable aléatoire $X$ à deux valeurs 0 et 1 ; choc latéral : $X = 0$ ; choc frontal : $X = 1$.
La gravité de l'accident est décrite par la variable aléatoire $Y$ à trois valeurs 1, 2, 3.

Une enquête réalisée sur un grand nombre d'accidents conduit au tableau ci-contre donnant la loi de probabilité conjointe du couple $( X , Y )$ :

| Y \ X | 0 | 1 |
|---|---|---|
| 1 | 0,10 | 0,15 |
| 2 | 0,08 | 0,25 |
| 3 | 0,02 | 0,40 |

1°/ Calculer les probabilités suivantes :

$P ( X = 1 )$ $P ( X = 0 \text{ et } Y \geq 2 )$ $P ( X = 0 \mid Y \geq 2 )$

2°/ Calculer les lois de probabilité marginales de $X$ et de $Y$, ainsi que leurs espérances mathématiques et leurs variances.
3°/ Les variables aléatoires $X$ et $Y$ sont-elles indépendantes ?
4°/ Etablir la loi de probabilité de la variable aléatoire $Z = X.\ Y$ et calculer $E ( Z )$ et $Var ( Z )$.
5°/ Calculer la covariance $Cov ( X , Y )$ et le coefficient de corrélation linéaire $\rho$. Conclusion ?
6°/ Calculer l'espérance mathématique et la variance de la variable aléatoire $T = X + Y$.
7°/ On définit une fonction de gravité $G = a X + b Y$ où $a = 0,2$ et $b = 0,8$. Calculer $E ( G )$ et $Var ( G )$.

# SOLUTION.

## 1°/ Probabilités marginales.

Le tableau des données peut être compété par une ligne de totaux et une colonne de totaux :

| Y \ X | 0 | 1 | Total |
|---|---|---|---|
| 1 | 0,10 | 0,15 | 0,25 |
| 2 | 0,08 | 0,25 | 0,33 |
| 3 | 0,02 | 0,40 | 0,42 |
| Total | 0,20 | 0,80 | 1,00 |

Ce tableau permet de répondre aux deux premières questions :
$P ( X = 1 ) = 0,80$
$P ( X = 0 \text{ et } Y \geq 2 ) = P ( X = 0 \text{ et } Y = 2 ) + P ( X = 0 \text{ et } Y = 3 ) = 0,08 + 0,02 = 0,10$

$$P ( X = 0 \mid Y \geq 2 ) = \frac{P( X = 0 \text{ et } Y \geq 2 )}{P(Y \geq 2 )} = \frac{P( X = 0 \text{ et } Y = 2 ) + P( X = 0 \text{ et } Y = 3 )}{P(Y = 2 ) + P(Y = 3 )} = \frac{0,08 + 0,02}{0,33 + 0,42} = \frac{0,10}{0,75} = \frac{2}{15}$$

La loi de probabilité marginales de la variable aléatoire $X$ se lit sur la dernière ligne :

| $X$ | 0 | 1 |
|---|---|---|
| $P ( X = x_i )$ | 0,20 | 0,80 |

La loi de probabilité marginale de la variable aléatoire $Y$ se lit dans la dernière colonne :

| $Y$ | 1 | 2 | 3 |
|---|---|---|---|
| $P ( Y = y_j )$ | 0,25 | 0,33 | 0,42 |

De la loi de probabilité de la variable aléatoire $X$, on tire les paramètres de $X$ :
$E ( X ) = 0,20 \times 0 + 0,80 \times 1 = 0,80$
$Var ( X ) = 0,20 \times ( 0 - 0,80 )^2 + 0,80 \times ( 1 - 0,80 )^2 = 0,20 \times 0,64 + 0,80 \times 0,04 = 0,128 + 0,032 = 0,1600$
De la loi de probabilité de la variable aléatoire $Y$, on tire les paramètres de $Y$ :
$E ( Y ) = 0,25 \times 1 + 0,33 \times 2 + 0,42 \times 3 = 2,17$
$Var ( Y ) = E ( Y^2 ) - ( E ( Y ) )^2 = 0,25 \times 1^2 + 0,33 \times 2^2 + 0,42 \times 3^2 - 2,17^2 = 5,35 - 4,7089 = 0,6411$

## 2°/ Indépendance des variables aléatoires *X* et *Y*.

Dire que les deux variables aléatoires $X$ et $Y$ sont indépendantes, c'est dire que, pour tout couple $(x_i, y_j)$ de valeurs de $X$ et $Y$, on a :

$$P(X = x_i \text{ et } Y = y_j) = P(X = x_i) \times P(Y = y_j)$$

Or on a, par exemple :

$$P(X = 0 \text{ et } Y = 1) = 0{,}10$$

et

$$P(X = 0) \times P(Y = 1) = 0{,}20 \times 0{,}25 = 0{,}05 \neq 0{,}10 = P(X = 0 \text{ et } Y = 1)$$

Les deux variables aléatoires $X$ et $Y$ ne sont donc pas indépendantes en probabilité.

## 3°/ Loi de probabilité de la variable aléatoire *Z = X Y*.

Tableau des valeurs de $Z$ :

| Y \ X | 0 | 1 |
|---|---|---|
| 1 | 0 | 1 |
| 2 | 0 | 2 |
| 3 | 0 | 3 |

En comparant ce tableau avec la loi de probabilité conjointe de $X$ et $Y$, on déduit la loi de probabilité de $Z$ :

| $Z$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(Z=z_i)$ | 0,20 | 0,15 | 0,25 | 0,40 |

Cette loi de probabilité permet de calculer les paramètres de la variable aléatoire $Z$.

$E(Z) = 0{,}20 \times 0 + 0{,}15 \times 1 + 0{,}25 \times 2 + 0{,}40 \times 3 = 1{,}85$

$Var(Z) = E(Z^2) - (E(Z))^2 = 0{,}20 \times 0^2 + 0{,}15 \times 1^2 + 0{,}25 \times 2^2 + 0{,}40 \times 3^2 - 1{,}85^2 = 4{,}75 - 3{,}4225 = 1{,}3275$

## 4°/ Covariance et coefficient de corrélation linéaire.

La covariance de $X$ et $Y$, définie comme l'espérance mathématique du produit des aléatoires centrées :

$$Cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

peut se calculer par la formule de la covariance :

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

(« moyenne du produit, moins produit des moyennes »). Il vient ainsi :

$$Cov(X, Y) = E(Z) - E(X)E(Y) = 1{,}85 - 0{,}80 \times 2{,}17 = 0{,}1140$$

Le coefficient de corrélation linéaire ρ est le rapport de la covariance au produit des écarts-types marginaux :

$$\rho = \frac{Cov(X, Y)}{\sigma_X \ \sigma_Y}$$

On obtient ici :

$$\rho = \frac{0{,}1140}{\sqrt{0{,}1600} \times \sqrt{0{,}6411}} = 0{,}3559$$

**Conclusion :**

Le fait que la covariance et le coefficient de corrélation linéaire ne soient pas nuls confirme ce que l'on savait déjà d'autre part : les variables aléatoires $X$ et $Y$ ne sont pas indépendantes en probabilité.

## 5°/ Variable aléatoire *T = X + Y*.

Par combinaison linéaire, on obtient :

$E(T) = E(X + Y) = E(X) + E(Y) = 0{,}80 + 2{,}17 = 2{,}97$

$Var(T) = Var(X + Y) = Var(X) + Var(Y) + 2 \times Cov(X, Y) = 0{,}16 + 0{,}6411 + 2 \times 0{,}114 = 1{,}029$

## 6°/ Fonction de gravité *G*.

$$G = 0{,}2\ X + 0{,}3\ Y$$

$E\ (\ G\ )$ $= 0{,}2\ E\ (\ X\ ) + 0{,}3\ E\ (\ Y\ ) = 0{,}2 \times 0{,}8 + 0{,}3 \times 2{,}17 = 0{,}811$

$Var\ (\ G\ )$ $= Var(\ 0{,}2\ X + 0{,}3\ Y\ ) = 0{,}2^2 \times Var\ (\ X\ ) + 0{,}3^2 \times Var\ (\ Y\ ) + 2 \times 0{,}2 \times 0{,}3 \times Cov\ (\ X\ ,\ Y\ )$

$= 0{,}04 \times 0{,}16 + 0{,}09 \times 0{,}6411 + 2 \times 0{,}06 \times 0{,}114 = 0{,}077779$

# Exercice 31.

1°/ Soit $X$ le nombre de filles dans une famille de trois enfants :
- Etablir la loi de probabilité de $X$. (On suppose que la probabilité de naissance d'une fille ou d'un garçon est la même à chaque naissance, soit $p = 0,5$).
- Quelle est la probabilité d'avoir au moins deux garçons ?

2°/ Soit $Y$ le nombre de séquences. On appelle « *séquence* » une suite d'enfants de même sexe (exemples : $FFF \to Y = 1$ ; $GFF \to Y = 2$ ; $GFG \to Y = 3$).
- Donner la loi de probabilité conjointe de $X$ et $Y$.
- Etablir la loi marginale de $Y$.

# SOLUTION.

## 1°/ Loi de probabilité du nombre de filles dans une famille de trois enfants.

Pour une naissance, l'épreuve qui consiste à regarder de quel sexe est le nouveau-né, est une épreuve de Bernoulli avec deux résultats possibles :
- le succès, naissance d'une fille, avec une probabilité $p = 0,5$ ;
- l'échec, avec naissance d'un garçon, avec une probabilité $q = 0,5$.

Lorsqu'on répète cette épreuve pour les trois enfants de la famille, le nombre de succès est le nombre $X$ de filles dans la famille ; cette variable aléatoire suit une loi binomiale de paramètres $n = 3$ et $p = 0,5$.

$$P( X = k) = C_n^k \, p^k \, (1-p)^{n-k} = \frac{C_3^k}{8}$$

La loi de probabilité de la variable aléatoire $X$ est donnée par le tableau suivant :

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $P(X = x_i)$ | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ |

La probabilité d'avoir au moins deux garçons est la probabilité d'avoir au plus une fille, c'est donc la somme :

$$P(X = 0) + P(X = 1) = \frac{1}{8} + \frac{3}{8} = 0,5$$

Il y a une chance sur deux d'avoir au moins deux garçons.

## 2°/ Loi de probabilité du nombre de séquences dans une famille de trois enfants.

Le nombre de séquences $Y$ n'est pas le même quelle que soit la valeur de $X$ :
Si $X = 0$, il n'y a que des garçons, il y a donc une seule séquence, la valeur de $Y$ est 1. La probabilité de cette valeur est 1. $P(Y=1|X=0) = 1$.
Si $X = 1$, il y a une seule fille : si elle est l'aînée, elle forme à elle seule une séquence, puis, après elle, les garçons forment une deuxième séquence et $Y = 2$. Si la fille est la cadette, on a aussi $Y = 2$. Enfin si la fille est le deuxième enfant de la famille, les sexes sont alternés, il y a trois séquences, $Y = 3$. Chaque possibilité a la même probabilité : la fille peut être l'aînée, la cadette ou la puînée avec une égale probabilité de 1/3.
De façon symétrique, si $X = 2$ ou si $X = 3$, on peut échanger les rôles des filles et des garçons : les nombres de séquences sont les mêmes.
En définitive, on obtient le tableau suivant de probabilités conjointes, qu'on peut compléter par les probabilités marginales :

| Y \ X | 0 | 1 | 2 | 3 | *P(Y = $y_j$ )* |
|---|---|---|---|---|---|
| 1 | 0,125 | 0 | 0 | 0,125 | *0,250* |
| 2 | 0 | 0,250 | 0,250 | 0 | *0,500* |
| 3 | 0 | 0,125 | 0,125 | 0 | *0,250* |
| *P(X = $x_i$ )* | *0,125* | *0,375* | *0,375* | *0,125* | *1,000* |

La loi de probabilité marginale de la variable aléatoire $Y$ est donnée dans la dernière colonne :

| *Y* | 1 | 2 | 3 |
|---|---|---|---|
| *P ( Y = $y_j$ )* | 0,25 | 0,50 | 0,25 |

Cette loi de probabilité est symétrique : son espérance mathématique est donc la valeur centrale 2, qui est aussi mode et médiane.

# VARIABLES ALEATOIRES CONTINUES

# Exercice 32.

Soit la fonction $f(x) = a x^2 + b x + c$ définie et positive sur l'intervalle [ 0 ; 1 ].

1°/ Ecrire la relation de normalisation existant entre les coefficients a, b, et c pour que la fonction f(x) puisse être considérée comme la fonction densité de probabilité d'une variable aléatoire continue X prenant ses valeurs dans l'intervalle [ 0 ; 1 ] de $\mathbb{R}$.

2°/ Parmi les fonctions densité de probabilité f(x) précédentes, trouver celle qui s'annule pour x = 0 et x = 1. Donner l'expression de sa fonction de répartition et représenter graphiquement ces deux fonctions.

3°/ Calculer la moyenne μ, la variance σ² et l'écart-type σ de la variable aléatoire X.

4°/ Calculer la médiane et le mode de la variable aléatoire X.

5°/ Calculer les probabilités suivantes : $\Pr\left(X = \frac{1}{3}\right)$ ; $\Pr\left(X < \frac{1}{3}\right)$ et $\Pr\left(X > \frac{2}{3}\right)$.

6°/ Calculer les probabilités des intervalles de fluctuation définis par la relation $|X - \mu| < 2\sigma$
Comparer les résutats avec ceux qui auraient été fournis par l'inégalité de Bienaymé-Tchebychev dans le cas où la loi de probabilité de la variable aléatoire est inconnue.

# SOLUTION.

## 1°/ Relation liant les coefficients a, b, c, de la densité de probabilité f(x).

Comme la fonction f(x) définie par $f(x) = a x^2 + b x + c$ sur [ 0 ; 1 ] et 0 ailleurs, est continue sur 'intervalle ouvert ] 0 ; 1 [, elle définit une densité de probabilité si :

- Elle vérifie la « ***condition de normalisation*** » : $\int_{-\infty}^{+\infty} f(x)\, dx = 1$.
- Elle est ***positive*** sur l'intervalle [ 0 ; 1 ]

### a) Condition de normalisation.

Puisque f(x) est nulle en dehors de l'intervalle [ 0 ; 1 ], la condition de normalisation se réduit à :

$\int_0^1 (a x^2 + b x + c)\, dx = 1$, soit $\left[ a\frac{x^3}{3} + b\frac{x^2}{2} + c\,x \right]_{x=0}^{x=1} = 1$, ou encore : $\frac{a}{3} + \frac{b}{2} + c = 1$,

c'est l'équation d'un plan dans l'espace ( a , b c ), qui s'écrit aussi :

$$\boxed{2a + 3b + 6c = 6}$$

***Cette condition n'est pas suffisante***, comme on le voit en prenant, par exemple : a = 0, c = – 1 , b = 4. Avec ces valeurs, la relation précédente est vérifiée, mais f(x), qui est représentée par la droite y = 4 x - 1, est négative entre 0 et $\frac{1}{4}$ : ce ne peut donc pas être une densité de probabilité. Tous les points du plan 2a + 3b + 6c = 6 ne conviennent pas.

### b) Positivité sur [ 0 ; 1 ].

L'étude de la positivité de f(x) est ici hors du sujet de l'exercice. La positivité sur l'intervalle [ 0 ; 1 ] se traduit par :

- $f(0) \geq 0$ et $f(1) \geq 0$ ;
- $f(x) = 0 \Rightarrow x \notin \,]0 ; 1[$.

$f(0) \geq 0$ donne $c \geq 0$.

$f(1) \geq 0$ donne $a + b + c \geq 0$.

Or la condition de normalisation 2 a + 3 b + 6 c = 6, s'écrit aussi : 3 b = 6 (1 – c ) – 2 a = 2 ( 3 ( 1 – c ) – a ), soit

$$b = -\frac{2}{3} [ a + 3 ( c - 1) ]$$

donc la condition a + b + c ≥ 0 s'écrit aussi 3 a – 2 [ a + 3 ( c – 1 ) ] + 3 c ≥ 0, soit a – 3 c + 6 ≥ 0,

$$a \geq 3 ( c - 2 )$$

Dans le plan ( c , a ), ou b = 0, les conditions c ≥ 0 et a ≥ 3 ( c – 2 ) définissent l'angle compris entre les droite c = 0 et a = 3 ( c – 2 ). Donc seuls les points du secteur du plan 2 a + 3 b + 6 c = 6 qui se projettent parallèlement à l'axe b dans l'angle du plan b = 0 que l'on vient de définir peuvent éventuellement donner des fonctions f(x) = a x² + b x + c densités de probabilité.

Le discriminant Δ = b² – 4 a c de f(x) s'écrit, par élimination de b, uniquement en fonction de a et c sous la forme :

$$\Delta = \frac{4}{9} [ a^2 + 6 a ( c - 1 ) + 9 ( c - 1 )^2 ] - 4 a c$$

de sorte que l'on a :

$$\frac{9}{4} \Delta = a^2 + 6 a ( c - 1 ) + 9 ( c - 1 )^2 - 9 a c$$

$$\boxed{\frac{9}{4} \Delta = a^2 - 3 a c + 9 c^2 - 6 a - 18 c + 9}$$

On voit ainsi que, pour discuter des valeurs de a, b et c qui peuvent éventuellement donner des densités de probabilité, il faut étudier la conique d'équation :

$$Y^2 - 3 X Y + 9 X^2 - 6 Y - 18 X + 9 = 0$$

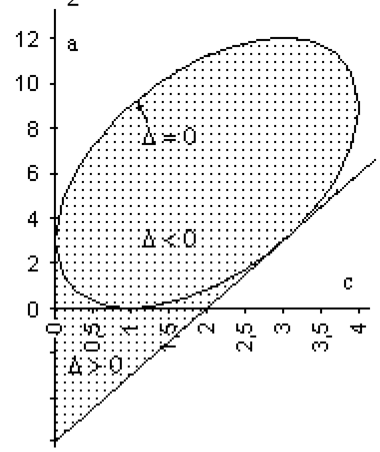
équation qui est de la forme :

$$A Y^2 + B X Y + C X^2 + D Y + C X + E = 0$$

Si l'on ordonne l'équation par rapport à Y, on obtient : Y² - 3 ( X + 2 ) Y + 9 ( X² - 2 X + 1) = 0, d'où l'on tire Y en fonction de X :

$$Y = \frac{1}{2}\{3 ( X + 2 ) \pm 3 [ ( X + 2 )^2 - 4 ( X - 1 )^2 ]^{½} \} = \frac{3}{2}\{ X + 2 \pm [ ( X + 2 + 2 ( X - 1 ) ) ( X + 2 - 2 ( X - 1 ) ) ]^{½} \}$$

$$Y = \frac{3}{2} \{ X + 2 \pm [ 3 X ( 4 - X) ]^{½} \}$$




Toute la conique est contenue entre les droites X = 0 et X = 4 : c'est une ellipse ( $E_0$ ). Cette ellipse ( $E_0$ ) est la projection parallèlement à l'axe b d'une ***ellipse ( E )*** du plan d'équation 2 a + 3 b + 6 c = 6. Le centre de l'ellipse ( $E_0$ ) est le point X = 2, Y = 6. En ce point, le polynôme qui définit l'ellipse vaut 36–36+36–36–36+9 = –25. Il est négatif. Donc :

- le discriminant Δ = b² – 4 a c de f(x) est négatif à l'intérieur de l'ellipse ( E ),
- Δ est nul sur l'ellipse ( E ).
- Δ est positif à l'extérieur de l'ellipse ( E ).

***1er cas : Intérieur et frontière de l'ellipse ( $E_0$ ).***

C'est la région (dite ***adhérence***) définie par :

$$0 \leq c \leq 4$$

$$\frac{3}{2} \{c + 2 - [ 3 c ( 4 - c) ]^{½} \} \leq a \leq \frac{3}{2} \{ c + 2 + [ 3 c ( 4 - c) ]^{½} \}$$

Dans ce domaine, $\Delta$ est plus petit que zéro, f(x) garde toujours le signe de a. Or

$$\{ c + 2 - [ 3 c ( 4 - c) ]^{½} \} \geq 0$$

puisque $3 c ( 4 - c ) = ( c + 2 )^2 - 4 ( c - 1 )^2$, qui est positif, est plus petit que $( c + 2 )^2$, avec un c + 2 compris entre 2 et 6, donc a est positif et f(x) est positif quel que soit x. Ainsi, lorsque le point ( a , c ) est adhérent à l'ellipse, f(x) définit toujours une densité de probabilité.

On peut résumer ce raisonnement de la façon suivante :

> Pour $\frac{a}{3} + \frac{b}{2} + c = 1$ , $0 \leq c \leq 4$ et $\frac{3}{2}\{c + 2 - [ 3 c ( 4 - c) ]^{½} \} \leq a \leq \frac{3}{2}\{ c + 2 + [ 3 c ( 4 - c) ]^{½} \}$,
> $f(x) = a x^2 + b x + c$ définit toujours une densité de probabilité sur [ 0 ; 1 ].

***2ème cas : Extérieur de l'ellipse.***

L'extérieur de l'ellipse est défini par :

$$\{[c < 0 \text{ ou } c > 4] \textbf{ et } a \text{ quelconque}\} \text{ ou}$$
$$\{0 \leq c \leq 4 \textbf{ et } [a < \frac{3}{2}\{c + 2 - [3c(4 - c)]^{½}\} \text{ ou } a > \frac{3}{2}\{c + 2 + [3c( 4 - c) ]^{½} \}]\}$$

Dans ce domaine, le discriminant $\Delta$ de f(x) est strictement positif, f(x) a deux racines réelles distinctes :

$$x_1 = \frac{1}{2a} ( - b - \sqrt{\Delta} ) = -\frac{1}{3a} \{ a + 3 ( c - 1) + [ a^2 - 3 a c + 9 c^2 - 6 a - 18 c + 9 ]^{½} \}$$
$$x_2 = \frac{1}{2a} ( - b + \sqrt{\Delta} ) = -\frac{1}{3a} \{ a + 3 ( c - 1) - [ a^2 - 3 a c + 9 c^2 - 6 a - 18 c + 9 ]^{½} \}$$

Pour que f(x) ne change pas de signe à l'intérieur de l'intervalle, ces deux racines doivent être toutes deux à l'extérieur de l'intervalle [ 0 ; 1 ]. Pour que f(x) soit positif dans l'intervalle [ 0 ; 1 ], il faut que l'on ait :

- ∗ ou bien : a > 0 et les racines sont toutes deux du même côté de l'intervalle [ 0 ; 1 ],
- ∗ ou bien : a < 0 et les racines sont de part et d'autre de l'intervalle [ 0 ; 1 ].

Ces conditions peuvent s'écrire aussi :

- ∗ ou bien : a > 0 et ( $x_2 < 0$ ou $x_1 > 1$ )
- ∗ ou bien : a < 0 et ( $x_1 < 0$ et $x_2 > 1$ ).

Les conditions $f(0) \geq 0$ et $f(1) \geq 0$ limitent l'étude au secteur de plan défini par $c \geq 0$ et $a \geq 3 ( c - 2 )$. La droite c = 0 du plan ( c , a ) est tangente à l'ellipse. La droite a = 3 ( c - 2 ) est tangente, elle aussi, à l'ellipse. En effet :

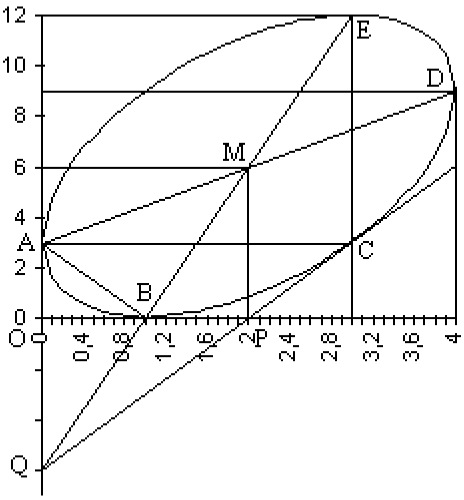
- Pour c = 3, le point de la droite a pour ordonnée a = 3, et le point ( 3 , 3 ) est sur l'ellipse puisque l'on a :
  $$a^2 - 3 a c + 9 c^2 - 6 a - 18 c + 9 = 9 - 27 + 81 - 18 - 54 + 9 = 99 - 99 = 0.$$
- En ce point de l'ellipse, la tangente a pour pente la valeur de la dérivée de $\frac{3}{2}\{c + 2 - [ 3 c ( 4 - c) ]^{½} \}$, soit :
  $$\frac{3}{2} \{ 1 - \frac{1}{2} ( - 6 c + 12 ) [ 3 c ( 4 - c) ]^{-½} \} = 3$$
- La tangente est donc la droite de pente 3 qui passe par le point ( 3 , 3 ), c'est la droite a = 3 ( c - 2 ).

Dans la région définie, entre les deux droites c = 0 et a = 3 ( c - 2 ), par $c \geq 0$ et $a \geq 3 ( c - 2 )$, on a $f(0) \geq 0$ et $f(1) \geq 0$. Quand a-t'on $f(x) \geq 0$ pour tout x entre 0 et 1 ? ***Réponse***: pour $c \leq 3$ et $a \leq \frac{3}{2}\{c + 2 - [ 3 c ( 4 - c) ]^{½} \}$, pas ailleurs. Ainsi, la région où f(x) définit une densité de probabilité est la région hachurée sur la figure. Dans cette région, les deux racines de f(x) sont à l'extérieur de l'intervalle [ 0 , 1 ] : du même côté pour a > 0, de part et d'autre pour a < 0. Au-delà de l'ellipse, les deux racines sont à l'intérieur de l'intervalle [ 0 , 1 ].

***En résumé :***

> Pour que $f(x) = a x^2 + bx + c$ définisse une densité de probabilité, il faut et il suffit que :
> 1. 2 a + 3 b + 6 c = 6 (condition de ***normalisation***)
> 2. Le point ( c , a ) appartienne à la région définie (par ***positivité*** de f(x)) par :
>    - l'adhérence de l'ellipse (E) d'équation $a^2 - 3 a c + 9 c^2 - 6 a - 18 c + 9 = 0$
>    - le triangle curviligne limité par les droites c = 0 et 2 a - 3 c = 6 et la branche inférieure de l'ellipse (E).

On peut maintenant préciser un peu les choses sur le graphique de l'ellipse :




**Points particuliers** :

- Sur l'ellipse :
  - ∗ A : c = 0 ; a = 3 ( b = 0 )
  - ∗ B : c = 1 ; a = 0 ( b = 0 )
  - ∗ C : c = 3 ; a = 3 ( b = – 6 )
  - ∗ D : c = 4 ; a = 9 ( b = – 12 )
  - ∗ E : c = 3 ; a = 12 ( b = – 12 )
- Dans l'ellipse :
  - ∗ M : c = 2 ; a = 6 ( b = – 6 )
- Hors de l'ellipse :
  - ∗ O : c = 0 ; a = 0 ( b = 2 )
  - ∗ P : c = 2 ; a = 0 ( b = – 2 )
  - ∗ Q : c = 0 ; a = – 6 ( b = 6 )

**Droites particulières :**

- Segment QOA : c = f ( 0 ) = 0. A droite de ce segment : f ( 0 ) > 0.
- Segment QBME : a = 6 ( c – 1 ).
- Segment QPC : a = 3 ( c – 2 ). A gauche de ce segment : f ( 1 ) > 0.
- Segment OBP : a = 0 . Sur cette droite, on a b = 2 ( 1 - c ) et ***f(x) est du premier degré*** .
  Pour c = 1 : pas de racine, f(x) = 1, ∀ x ∈ [ 0 ; 1 ].
  Pour c ≠ 1, f(x) a pour racine : x = $\dfrac{c}{2\ (c-1)}$
  
  Pour 0 < c < 1 : une racine < 0. Droite ascendante ( b > 0 )
  Pour 1 < c < 2 : une racine > 1. Droite descendante ( b < 0 )
- Segment AMD : 2 a = 3 ( c + 2 ). Axe de symétrie verticale de l'ellipse.
- Droite AB : a = 3 ( 1 – c ). Intersection des plans 2 a + 3 b + 6 c = 6 et b = 0

***Régions particulières et résultats s'y rattachant :***

- **Intérieur de l'ellipse** : f(x) est toujours positif et f(x) définit une densité de probabilité.
  0 < c < 4
  $\frac{3}{2}$ {c + 2 - [ 3 c ( 4 – c) ]$^{½}$ } ≤ a ≤ $\frac{3}{2}$ { c + 2 + [ 3 c ( 4 – c) ]$^{½}$ }
  b > 0 à gauche de la droite AB ; b < 0 à droite de la droite AB ; b = 0 sur le segment de droite AB.
  - ∗ Point M : c = 2 ; a = 6 ; ( b = – 6 ) ; f(x) = 2 ( 3 x² - 3 x + 1 ) est symétrique par rapport à x = $\frac{1}{2}$ ; f(0) = f(1) = 2 ; f ( $\frac{1}{2}$ ) = $\frac{1}{2}$ ; f(x) décroît de 0 à $\frac{1}{2}$, puis croît de $\frac{1}{2}$ à 1.
  - ∗ Segment BE : 1 < c < 3 ; a = 6 ( c – 1 ) ; b < 0 ; f(x) est symétrique par rapport à x = $\frac{1}{2}$ ; f(x) décroît de 0 à $\frac{1}{2}$, puis croît de $\frac{1}{2}$ à 1.
  - ∗ Triangle curviligne AEB : 0 < c < 3 ; Sup ( 3 ( 1 – c ) ; 6 ( c – 1 ) ) < a < $\frac{3}{2}$ { c + 2 + [ 3 c ( 4 – c) ]$^{½}$ } ; b < 0 ; f(x) n'a pas de racine réelle ; son axe de symétrie est à gauche de x = $\frac{1}{2}$ ; f(x) est croissante au voisinage de 1, et décroissante au voisinage de 0.
  - ∗ Secteur d'ellipse sous AB : 0 < c < 1 ; $\frac{3}{2}$ { c + 2 – [ 3 c ( 4 – c) ]$^{½}$ } < a < 3 ( 1 – c ) ; b > 0 ; f(x) est croissante au voisinage de 1 et croissante au voisinage de 0.
  - ∗ Segment AB : 0 < c < 1 ; a = 3 ( 1 – c ) ; b = 0 ; f(x) = 3 ( 1 – c ) x² + c est minimum pour x = 0 et croissante sur [ 0 ; 1 ].
  - ∗ Intérieur de la demi-ellipse BCDEM : 1 < c < 4 ; $\frac{3}{2}$ { c + 2 – [ 3 c ( 4 – c) ]$^{½}$ } < a < Inf ( 6 ( 1 – c ) ; $\frac{3}{2}$ { c + 2 + [ 3 c ( 4 – c) ]$^{½}$ } ) ; b < 0 ; l'axe de symétrie de f(x) est à droite de la droite x = $\frac{1}{2}$ ; f(x) est décroissante au voisinage de 0 et croissante au voisinage de 1.

- **Frontière** de l'ellipse : f(x) définit toujours une densité de probabilité.
  - Point A : c = 0 , a = 3 , ( b = 0 ) : $f(x) = 3x^2$ a une racine double en x = 0.
  - Arc AB : f(x) a une racine double négative $-\sqrt{\frac{c}{a}}$, 0 < c < 1 , a = $\frac{3}{2}$ {c + 2 - [ 3 c ( 4 – c) ]$^{½}$ } , b = 2 $\sqrt{a\,c}$ .
  - Point B : c = 1 , a = 0 , b = 0 : f(x) = 1. Densité de probabilité constante.
  - Arc BC : f(x) a une racine double > 1. 1 < c < 3 , a = $\frac{3}{2}$ {c + 2 - [ 3 c ( 4 – c) ]$^{½}$ } , b = – 2 $\sqrt{a\,c}$ .
  - Point C : c = 3 , a = 3 ( b = – 6 ) : $f(x) = 3 ( x - 1 )^2$, 1 est racine double.
  - Arc CD : f(x) a une racine double positive $\sqrt{\frac{c}{a}}$ comprise entre 1 et $\frac{2}{3}$ ;

    3 < c < 4 , a = $\frac{3}{2}$ {c + 2 - [ 3 c ( 4 – c) ]$^{½}$ } , b = – 2 $\sqrt{a\,c}$ .
  - Point D : c = 4 ; a = 9 ( b = - 12 ) : $f ( x ) = ( 3 x - 2 )^2$ ; $\frac{2}{3}$ est racine double. c atteint là sa valeur maximum.
  - Arc DE : f(x) a une racine double positive $\sqrt{\frac{c}{a}}$ comprise entre $\frac{2}{3}$ et $\frac{1}{2}$;

    3 < c < 4 , a = $\frac{3}{2}$ {c + 2 + [ 3 c ( 4 – c) ]$^{½}$ } , b = – 2 $\sqrt{a\,c}$ .
  - Point E : c = 3 ; a = 12 ( b = – 12 ) : $f(x) = 3 ( 2 x - 1 )^2$ ; $\frac{1}{2}$ est racine double a atteint là sa valeur maximum.
  - Arc EA : f(x) a une racine double positive $\sqrt{\frac{c}{a}}$ comprise entre $\frac{1}{2}$ et 0 ;

    0 < c < 3 , a = $\frac{3}{2}$ {c + 2 + [ 3 c ( 4 – c) ]$^{½}$ } , b = – 2 $\sqrt{a\,c}$ .
- **Intérieur du triangle curviligne AOB** : f(x) a deux racines réelles simples < 0 ; f(x) est croissant sur [ 0 ; 1 ].
- **Intérieur du triangle curviligne BPC** : f(x) a deux racines réelles simples > 1 ; f(x) est décroissant sur [ 0 ; 1 ].
- **Point O** : c = 0 ; a = 0 ( b = 2 ) : f(x) = 2 x ; f(x) a une racine réelle simple x = 0 ; f(x) est croissant sur [ 0 ; 1 ].
- **Segment OB** : 0 < c < 1 ; a = 0 ( b = 2 ( 1 - c ) ) ; f(x) = 2 ( 1 - c ) x + c a une racine réelle simple < 0 ; f(x) est croissant sur [ 0 ; 1 ].
- **Segment BP** : 1 < c < 2 ; a = 0 ( b = 2 ( 1 - c ) ) ; f(x) = 2 ( 1 - c ) x + c a une racine réelle simple > 1 ; f(x) est décroissant sur [ 0 ; 1 ].
- **Point P** : c = 2 ; a = 0 ( b = – 2 ) ; f(x) = – 2 ( x – 1 ) a une racine réelle simple x = 1 ; f(x) est décroissant sur [ 0 ; 1 ].
- **Segment OA** : c = 0 ; 0 < a < 2 ( b = 2 ( 1 – $\frac{a}{3}$ ) ) ; f(x) = x [ a x + 2 ( 1 – $\frac{a}{3}$ ) ] a deux racines réelles distinctes : 0 et 2 ( $\frac{1}{3} - \frac{1}{a}$ ) qui est inférieure à – $\frac{1}{3}$ ; f(x) est croissante sur [ 0 ; 1 ].

- **Segment QO** : c = 0 ; – 6 < a < 0 ( b = 2 ( 1 – $\frac{a}{3}$ ) ) ; f(x) = x [ a x + 2 ( 1 – $\frac{a}{3}$ ) ] a deux racines réelles distinctes : 0 et 2 ( $\frac{1}{3} - \frac{1}{a}$ ) qui est supérieure à 1 ; f(x) est croissante au voisinage de 0.
- **Point Q** : c = 0 ; a = – 6 ( b = 6 ) : f(x) = – 6 x ( x - 1 ) a deux racines réelles distinctes 0 et 1 ; f(x), croissante au voisinage de 0, admet pour axe de symétrie la droite x = $\frac{1}{2}$.

- **Segment QB** : $0 < c < 1$ ; $a = 6 ( c - 1 )$ ; $b = - 6 ( c - 1 )$ ; $f(x) = 6 ( c - 1 ) x^2 - 6 ( c - 1 ) + c$ admet la droite $x = \frac{1}{2}$ pour axe de symétrie. Ses racines sont $\frac{1}{2}\left(1 \pm \frac{\sqrt{1+\frac{2}{1-c}}}{\sqrt{3}}\right)$ : elles sont situées de part et d'autre de l'intervalle [ 0 ; 1 ], et symétriques par rapport au milieu du segment [ 0 ; 1 ]. f(x) est croissante de 0 à $\frac{1}{2}$ et décroissante de $\frac{1}{2}$ à 1.

- **Triangle QOB** : $0 < c < 1$ ; $6 ( c - 1 ) < a < 0$ ; f(x) a deux racines réelles distinctes de part et d'autre de l'intervalle [ 0 ; 1 ], son axe de symétrie est à droite de $x = \frac{1}{2}$ ; f(x) est croissante au voisinage de 0.
- **Triangle QBP** : $0 < c < 2$ ; $3 ( c - 2 ) < a < \text{Inf} ( 6 ( c - 1 ) ; 0 )$ ; f(x) a deux racines réelles distinctes de part et d'autre de l'intervalle [ 0 ; 1 ] , son axe de symétrie est à gauche de $x = \frac{1}{2}$ ; f(x) est décroissante au voisinage de 1.
- **Segment QP** : $0 < c < 2$ ; $a = 3 ( c - 2 )$ ; ( $b = 2 ( 3 - 2 c )$ ) ; $f(x) = ( x - 1 ) ( 3 ( c - 2 ) x - c )$ a deux racines réelles distinctes : 1 et $\frac{1}{3}\left(1 - \frac{1}{1-\frac{c}{2}}\right)$ qui est négative ; f(x) est décroissante au voisinage de 1.
- **En dehors du secteur de plan limité par QOAEDCPQ**, la fonction f(x) ne définit pas une densité de probabilité parce qu'il y a toujours au moins un morceau de l'intervalle [ 0 ; 1 ] dans lequel f(x) est négative.

| |
|---|
| Dans le plan défini par la condition de normalisation : $2 a + 3 b + 6 c = 6$, la seule région de l'espace (a , b , c ) qui donne pour $f(x) = a x^2 + b x + c$ une densité de probabilité est l'adhérence de la région qui se projette, parallèlement à l'axe b sur le secteur de plan QOAEDCPQ |

.

## 2°/ Densité de probabilité f(x) qui s'annule pour x = 0 et pour x = 1.

### a. Densité de probabilité f(x).

Elle est définie par les trois conditions :

$f(0) = c = 0$
$f(1) = a + b + c = 0$
Condition de normalisation : $2 a + 3 b + 6 c = 6$.

Comme on l'a vu dans l'étude précédente, elle correspond au point Q , projection dans le plan ( c , a ) du point :

$a = - 6$
$b = 6$
$c = 0$

| | |
|---|---|
| $f ( x ) = - 6 x^2 + 6 x = 6 x ( 1 - x )$ | pour $0 < x < 1$ |
| $f ( x ) = 0$ | ailleurs . |

### b. Fonction de répartition F ( x ).

La fonction de répartition F ( x ) correspondant à f(x) est définie par $F(x) = \int_{-\infty}^{x} f(t)\,dt$.

Pour $x \leq 0$, la fonction f(t) est nulle sur tout l'intervalle d'intégration, donc F(x) est nulle.
Pour $0 < x < 1$, on peut décomposer l'intervalle d'intégration en deux morceaux :

- Un morceau de $- \infty$ à 0 sur lequel f(t) est nulle, donc sur lequel l'intégrale est nulle ;

- Un morceau de 0 à x sur lequel f ( t ) = – 6 t² + 6 t, donc sur lequel l'intégrale vaut : $[-2t^3+3t^2]_0^x = -2x^3 + 3x^2$.

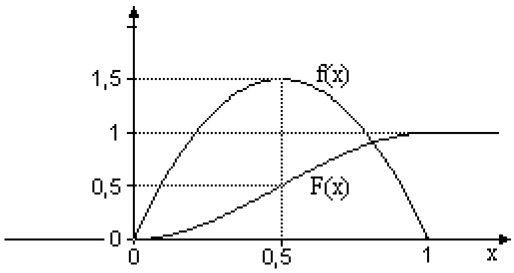
Pour x ≥ 1, on peut décomposer l'intervalle d'intégration en trois morceaux :

- Un morceau de – ∞ à 0 sur lequel f(t) est nulle, donc sur lequel l'intégrale est nulle ;
- Un morceau de 0 à 1 sur lequel f ( t ) = – 6 t² + 6 t, donc sur lequel l'intégrale vaut : $[-2t^3+3t^2]_0^1 = 1$ ;
- Un morceau de 1 à x sur lequel f(t) est nulle, donc sur lequel l'intégrale est nulle.

Au total, pour x ≥ 1, on obtient F ( x ) = 1.

$$F(x) = \begin{cases} 0 & \text{pour } x \le 0 \\ -2x^3+3x^2 & \text{pour } 0 < x < 1 \\ 1 & \text{pour } x \ge 1 \end{cases}$$

### c. Courbes représentatives de f(x) et de F(x).




## 3°/ Moyenne, variance et écart-type de X.

### a) Moyenne.

Comme la distribution est symétrique autour de la valeur 0,5, on voit tout de suite, sur le graphique précédent que la moyenne, égale à la médiane ( F(x) = 0,5), est 0,5. On peut aussi le montrer par le calcul :

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_0^1 x\,(-6x^2+6x)\, dx = \left[-\frac{3}{2}x^4 + 2x^3\right]_0^1 = \frac{1}{2}.$$

### b. Variance.

1. La variance peut se calculer à partir de la ***définition*** :

$$\mathrm{Var}(X) = E((X - E(X))^2) = \int_{-\infty}^{+\infty} \left(x - \frac{1}{2}\right)^2 f(x)\, dx = \int_0^1 6x\left(x-\frac{1}{2}\right)^2 (1-x)\, dx$$

Changement de variable : $x - \frac{1}{2} = u$ ; $dx = du$ ; $x = 0 \Rightarrow u = -\frac{1}{2}$ ; $x = 1 \Rightarrow u = \frac{1}{2}$.

$$\mathrm{Var}(X) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} 6\left(u+\frac{1}{2}\right)u^2\left(\frac{1}{2}-u\right) du = 6\int_{-\frac{1}{2}}^{+\frac{1}{2}} u^2\left(\frac{1}{4}-u^2\right) du = 2\times\left[\frac{1}{2}u^3 - \frac{6}{5}u^5\right]_0^{\frac{1}{2}} = \frac{1}{8} - \frac{3}{40} = \frac{1}{20}.$$

2. La variance peut se calculer à partir de la ***formule de la variance*** : Var ( X ) = E ( X² ) – [ E ( X ) ]².

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx = \int_0^1 x^2(-6x^2+6x)\, dx = -6\int_0^1 x^4\, dx + 6\int_0^1 x^3\, dx = -\frac{6}{5}[x^5]_0^1 + \frac{3}{2}[x^4]_0^1 = \frac{3}{10}$$

$$\mathrm{Var}(X) = \frac{3}{10} - \left(\frac{1}{2}\right)^2 = \frac{3}{10} - \frac{1}{4} = \frac{1}{20}.$$

L'écart-type est la racine carrée de la variance : $\sigma_X = \frac{\sqrt{5}}{10}$.

## 4°/ Médiane et mode.

### a) Médiane.

La médiane $M_e$ est définie par : $\Pr(X \leq M_e) = \frac{1}{2}$. On a donc $F(M_e) = \frac{1}{2}$ ; comme $F(M_e)$ est différent de 0 et de 1, $M_e$ est compris entre 0 et 1, et $M_e$ est solution de l'équation : $-2M_e^3 + 3M_e^2 = \frac{1}{2}$. Il y a une solution évidente que l'on avait déjà trouvée plus haut : $M_e = \frac{1}{2}$. Divisons le polynôme $-2M_e^3 + 3M_e^2 - \frac{1}{2}$ par $(M_e - \frac{1}{2})$ :

$$-2M_e^3 + 3M_e^2 - \frac{1}{2} = (M_e - \frac{1}{2}) \times (-2M_e^2 + 2M_e + 1) = -2\,(M_e - \frac{1}{2})(M_e - \frac{1+\sqrt{3}}{2})(M_e - \frac{1-\sqrt{3}}{2}).$$

La seule racine du polynôme qui soit comprise entre 0 et 1 est $M_e = \frac{1}{2}$, les autres sont ou négative ( $M_e = \frac{1-\sqrt{3}}{2}$ ), ou plus grande que 1 ($M_e = \frac{1+\sqrt{3}}{2}$ ). La médiane vaut donc :

$$M_e = \frac{1}{2}.$$

### b) Mode.

Par définition, le mode de X est la valeur de X pour laquelle f(x) atteint un maximum. Dans la construction de la courbe représentative de la densité de probabilité f(x), nous avons vu que f(x) a un maximum, et un seul pour x = 0,5. Le mode de X est donc :

$$M_o = \frac{1}{2}.$$

## 5°/ Calcul de diverses probabilités.

a) $\Pr(X = \frac{1}{3})$.

C'est la probabilité d'une valeur particulière : pour une distribution continue, la probabilité d'une valeur particulière est toujours nulle.

$$\Pr(X = \frac{1}{3}) = 0$$

b) $\Pr(X < \frac{1}{3})$.

$$\Pr(X < \frac{1}{3}) = F(\frac{1}{3}) = -2\left(\frac{1}{3}\right)^3 + 3\left(\frac{1}{3}\right)^2 = \frac{1}{3} - \frac{2}{27} = \frac{7}{27}$$

c) $\Pr(X > \frac{2}{3})$.

Comme la distribution est symétrique par rapport à $\frac{1}{2} = \frac{1}{2}\left(\frac{1}{3}+\frac{2}{3}\right)$, on a $\Pr(X > \frac{2}{3}) = \Pr(X < \frac{1}{3}) = \frac{7}{27}$.

On peut aussi montrer ce résultat par le calcul :

$$\Pr(X > \frac{2}{3}) = 1 - F(\frac{2}{3}) = 1 - [-2\left(\frac{2}{3}\right)^3 + 3\left(\frac{2}{3}\right)^2] = 1 + \frac{16}{27} - \frac{4}{3} = \frac{7}{27}.$$

6°/ Probabilité des intervalles de fluctuation $| X - \mu | < \sigma$ et $| X - \mu | < 2\sigma$ .

**a) Valeur des probabilités.**

| x | $\mu - 2\sigma = \frac{5 - 2\sqrt{5}}{10}$ | $\mu - \sigma = \frac{5 - \sqrt{5}}{10}$ | $\mu + \sigma = \frac{5 + \sqrt{5}}{10}$ | $\mu + 2\sigma = \frac{5 + 2\sqrt{5}}{10}$ |
|---|---|---|---|---|
| F ( x ) | $\frac{25 - 11\sqrt{5}}{50}$ | $\frac{25 - 7\sqrt{5}}{50}$ | $\frac{25 + 7\sqrt{5}}{50}$ | $\frac{25 + 11\sqrt{5}}{50}$ |

$$\Pr ( | X - \mu | < \sigma ) = F ( \mu + \sigma ) - F ( \mu - \sigma ) = \frac{7\sqrt{5}}{25} = 0{,}6261$$

$$\Pr ( | X - \mu | < 2\sigma ) = F ( \mu + 2\sigma ) - F ( \mu - 2\sigma ) = \frac{11\sqrt{5}}{25} = 0{,}9839$$

**b) Majorations données par l'inégalité de Bienaymé - Tchebitchev.**

L'inégalité de Bienaymé - Tchebitchev, valable pour une variable aléatoire discontinue, est encore valable pour une variable continue, quelle que soit sa loi de probabilité :

$$\boxed{\Pr \{ | X - \mu | \geq k\sigma \} \leq \frac{1}{k^2}.}$$

Pour k = 1 et k = 2, cette inégalité donne :

$\Pr \{ | X - \mu | \geq \sigma \} \leq 1$ et $\Pr \{ | X - \mu | \geq 2\sigma \} \leq \frac{1}{4}$

On en tire seulement : $\Pr \{ | X - \mu | < \sigma \} \geq 0$ et $\Pr \{ | X - \mu | < 2\sigma \} \geq \frac{3}{4}$, ce qui n'est pas très fin comme minoration.

# Exercice 33.

On suppose que la glycémie est distribuée normalement dans la population, avec une moyenne de 1 g/l et un écart-type de 0,03 g/l. On mesure la glycémie chez un individu.

1- Calculer la probabilité pour que sa glycémie soit :
   a) inférieure à 1,06 ;
   b) supérieure à 0,9985 ;
   c) comprise entre 0,94 et 1,08 ;
2- On mesure la glycémie chez 1 000 individus. Donner le nombre moyen d'individus dont la glycémie est supérieure à 0,99.

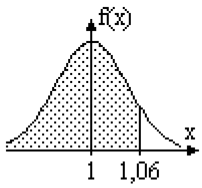
# SOLUTION.

## 1°/ Probabilité de divers intervalles de valeurs de la glycémie.

Notons X la glycémie mesurée sur un individu de la population. X suit une loi de Gauss G ( 1,00 ; 0,03 ). La variable aléatoire centrée réduite correspondante $U = \frac{X - \mu}{\sigma}$ suit une loi de Gauss G ( 0 ; 1 ).

**a) P ( X < 1,06 )**

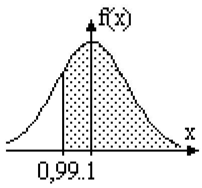
C'est la surface hachurée suivante :




$$P ( X < 1{,}06 ) = P\left( U < \frac{1{,}06 - 1}{0{,}03} \right) = P ( U < 2 ) = F ( 2 ) = 0{,}977249868052 \approx 0{,}9772$$

b) P ( X > 0,9985 )

C'est la surface hachurée suivante :



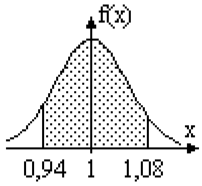

$$P ( X > 0{,}9985 ) = P\left( U > \frac{0{,}9985 - 1}{0{,}03} \right) = P ( U > -0{,}05 ) = P ( U < 0{,}05 )$$

$$P ( X > 0{,}9985 ) = F ( 0{,}05 ) = 0{,}519938805838 \approx 0{,}5199$$

$$F ( 0{,}05 ) \approx 0{,}5 + \frac{1}{\sqrt{2\pi}} \left( 0{,}05 - \frac{0{,}05^3}{6} + \ldots \right) \approx 0{,}5200$$

**c) P ( 0,94 < X < 1,08 )**

C'est la surface hachurée suivante :




P ( 0,94 < X < 1,08 ) $= P\left(\frac{0{,}94-1}{0{,}03} < U < \frac{1{,}08-1}{0{,}03}\right) = F\left(\frac{8}{3}\right) - F(-2) = F\left(\frac{8}{3}\right) - 1 + F(2)$

$= 0{,}9962 + 0{,}9772 - 1 = 0{,}9734$

$$P(0{,}94 < X < 1{,}08) = 0{,}9734$$

## 2°/ Nombre moyen d'individus, dans un échantillon de 1000 personnes, dont la glycémie est supérieure à 0,99.

$P(X > 0{,}99) = 1 - F\left(\frac{1}{3}\right) = 1 - 0{,}369441340182 = 0{,}630558659818 \approx 0{,}6306.$

En moyenne, la proportion d'individus ayant une glycémie supérieure à 0,99, tend à se rapprocher de la valeur théorique précédente donnée par la loi de Gauss, à mesure que le nombre d'individus pris en considération augmente. Pour 1 000 individus, on peut donc considérer qu'il y a en moyenne 630,6 individus qui présentent une glycémie supérieure à 0, 99 g/l.

# Exercice 34.

A l'entrée d'une station de métro, un marchand de journaux remarque qu'en moyenne, entre 8h et 9h, une personne sur 10 achète un journal.

1. Sachant qu'il passe 400 personnes entre 8h et 9h, indiquer la loi de probabilité de X, nombre de journaux vendus pendant cette période (préciser les hypothèses).
2. Donner l'espérance mathématique et la variance de X.
3. Par quelle loi de probabilité peut-on approcher la loi de X ? Utiliser cette approximation pour calculer les probabilités des évènements :
   $X = 42$ ; $X \geq 45$ ; $35 \leq X < 45$ ; $29 \leq X < 52$.

# SOLUTION.

## 1°/ Loi de probabilité du nombre X de journaux vendus entre 8h et 9h.

### a) Epreuve de Bernoulli.

Pour une personne qui prend le métro entre 8h et 9h, de deux choses l'une : ou bien elle achète un journal (succès, probabilité $p = 0{,}1$ ), ou bien elle n'en achète pas ( échec, probabilité $q = 0{,}9$ ). Regarder si la personne achète un journal constitue donc une épreuve de Bernoulli.

***Hypothèses :***

- On suppose que les 400 personnes qui prennent le métro entre 8h et 9h décident indépendamment l'une de l'autre de l'achat ou non d'un journal. (épreuves indépendantes).
- On suppose que la probabilité d'acheter un journal est la même pour toules les personnes qui prennent le métro entre 8h et 9h. (la probabilité du succès est la même pour chaque épreuve).

### b) Loi de probabilité du nombre X de succès.

Lorsqu'on répète l'épreuve de Bernoulli pour les 400 personnes qui prennent le métro entre 8h et 9h, le nombre X de journaux vendus est le nombre de succès dans cette répétition de l'épreuve de Bernoulli. Le nombre de succès X suit une loi binomiale de paramètres : $n = 400$ (nombre de répétittions) et $p = 0{,}10$ (probabilité du succès dans l'épreuve).

$$P ( X = k ) = C_{400}^{k} \; ( 0{,}1 )^{k} \, ( 0{,}9 )^{400-k}.$$

## 2°/ Espérance mathématique et variance de X.

$$E ( X ) = n\,p = 400 \times 0{,}1 = 40$$
$$Var ( X ) = n\,p\,q = 40 \times 0{,}9 = 36$$

## 3°/ Approximation de la loi binomiale par la loi normale.

Comme $n\,p = 40$ et $n\,q = 360$ sont tous deux supérieurs à 10, on peut espérer avoir une bonne approximation de la loi binomiale par une loi normale de moyenne $\mu = n\,p = 40$ et de variance $\sigma^2 = n\,p\,q = 36$.
Appelons X' la variable aléatoire normale permettant d'approcher la loi de probabilité de X. A chaque évènement concernant X correspond un évènement concernant X'. Le tableau suivant permet de comparer les résultats donnés par l'approximation normale et l'approximation par une loi de Poisson de paramètre $\mu = 40$.

| Evènement sur X | $X = 42$ | $X \geq 45$ | $35 \leq X < 45$ | $29 \leq X < 52$ |
|---|---|---|---|---|
| Evènement sur X' | $41{,}5 < X' < 42{,}5$ | $X' > 44{,}5$ | $34{,}5 < X' < 44{,}5$ | $28{,}5 < X' < 51{,}5$ |
| Loi binomiale B (400 ; 0,1) | $6{,}148 \times 10^{-2}$ | $22{,}367 \times 10^{-2}$ | $59{,}585 \times 10^{-2}$ | $94{,}542 \times 10^{-2}$ |
| Loi normale G ( 40 ; 36 ) | $6{,}283 \times 10^{-2}$ | $22{,}663 \times 10^{-2}$ | $59{,}371 \times 10^{-2}$ | $94{,}472 \times 10^{-2}$ |
| Loi de Poisson P ( 40 ) | $5{,}849 \times 10^{-2}$ | $23{,}432 \times 10^{-2}$ | $57{,}181 \times 10^{-2}$ | $93{,}188 \times 10^{-2}$ |

L'approximation par la loi normale de moyenne 40 et de variance 36 apparaît comme la meilleure des deux approximations envisageables. Cependant, il faut remarquer que le perfectionnement des calculettes actuelles permet, dans un cas comme celui-ci, de se passer d'approximation et de calculer la vraie valeur de probabilité donnée par la loi binomiale, et ceci, assez rapidement.

***Détail des calculs dans l'approximation normale :***

## a) P ( X = 42 )

La correction de continuité donne P ( X = 42 ) = P ( 41,5 < X' < 42,5 ) = $F\left(\dfrac{42{,}5-40}{6}\right) - F\left(\dfrac{41{,}5-40}{6}\right)$

$$P(X = 42) = F\left(\frac{5}{12}\right) - F\left(\frac{1}{4}\right) = 0,661\,5 - 0,598\,7 = 6,28 \times 10^{-2}.$$

## b) P ( X ≥ 45 )

$$P(X \geq 45) = P(X' > 44{,}5) = 1 - P(X' < 44{,}5) = 1 - F\left(\frac{44{,}5-40}{6}\right) = 1 - F\left(\frac{3}{4}\right) = 1 - F(0{,}75)$$

$$P(X \geq 45) = 1 - 0,773\,4 = 22,66 \times 10^{-2}.$$

## c) P ( 35 ≤ X < 45 )

$$\begin{aligned} P(35 \leq X < 45) &= P(34{,}5 < X' < 44{,}5) \\ &= F\left(\frac{44{,}5-40}{6}\right) - F\left(\frac{34{,}5-40}{6}\right) \\ &= F\left(\frac{3}{4}\right) - F\left(\frac{-11}{12}\right) = F\left(\frac{3}{4}\right) + F\left(\frac{11}{12}\right) - 1 \\ &= F(0{,}75) + F(0{,}9167) = 0,773\,4 + 0,802\,3 - 1 = 59,37 \times 10^{-2}. \end{aligned}$$

# Exercice 35.

On désire étudier la loi de probabilité du « temps d'attente » d'un véhicule devant un signal de circulation type « rouge et vert ». On suppose que la périodicité du signal est régulière et que l'arrivée du véhicule est complètement aléatoire. Le temps d'attente du véhicule est noté $X$ ; il peut varier de $a = 0$ à $b = 60$ (en secondes). On définit sur l'intervalle de définition de $X$, la fonction $f(x) = \dfrac{k}{b-a} = \dfrac{k}{60}$ ($k > 0$).

1°/ Calculer la valeur de $k$ pour que $f(x)$ puisse être considérée comme la fonction densité de probabilité de la variable aléatoire continue $X$ qui prend ses valeurs dans l'intervalle [ $a$ ; $b$ ] = [ 0 ; 60 ].

$k =$ **Réponse 1**

2°/ Calculer le temps d'attente moyen (en secondes) $E(X) =$ **Réponse 2**

3°/ Calculer l'écart-type (on arrondira le résultat à deux chiffres) : $\sigma =$ **Réponse 3**

4°/ Calculer la probabilité pour que le véhicule attende moins de 20 secondes **Réponse 4**

5°/ Calculer la probabilité pour que le véhicule attende entre 15 et 30 secondes **Réponse 5**

6°/ Calculer la probabilité suivante : $P(\ |X - E(X)| \leq \sigma\ )$ N.B. dans ce calcul, on prendra pour σ la valeur arrondie à deux chiffres significatifs de la réponse 3. **Réponse 6**

*Tableau de réponses de l'exercice 1 (Réponses 1 à 6)*

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0,05 | 0,11 | 0,19 | 0,29 | 0,41 | 0,52 | 0,65 | 0,81 | 5 | 9 | 17 | 30 | 37 | 45 | 50 |
| 0 | 0,09 | 0,15 | 0,25 | 0,33 | 0,49 | 0,57 | 0.,71 | 1 | 7 | 12 | 26 | 33 | 41 | 47 | 51 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | B | D | AB | AD | BC | BE | CE | ABC | ABE | ACE | BCD | BDE | ABCD | ABDE | BCDE |
| A | C | E | AC | AE | BD | CD | DE | ABD | ACD | ADE | BCE | CDE | ABCE | ACDE | ABCDE |

# SOLUTION.

## 1°/ Condition de normalisation.

Pour que la fonction $f(x) = \dfrac{k}{b-a} = \dfrac{k}{60}$ ($k > 0$) soit une densité de probabilité, il faut et il suffit que la condition de normalisation $\int_{-\infty}^{+\infty} f(x)\,dx = 1$ soit remplie. Cette condition s'écrit $\dfrac{k}{b-a}\int_a^b dx = 1$, soit $k = 1$.

$$\boxed{k = 1}$$

**Réponse 1 : ABD**

## 2°/ Espérance mathématique.

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x)\, dx = \int_0^{60} \frac{x\,dx}{60} = \frac{1}{60}\left[\frac{x^2}{2}\right]_{x=0}^{x=60} = \frac{60^2}{60\times 2} = 30 \text{ secondes}$$

$$\boxed{E(X) = 30 \text{ secondes}}$$

**Réponse 2 : BDE**

## 3°/ Ecart-type.

$$\sigma = \sqrt{Var(X)} = \sqrt{\int_0^{60} (x - E(X))^2 f(x)\,dx} = \sqrt{\int_0^{60} \frac{(x-30)^2}{60}\,dx} = \sqrt{\left[\frac{(x-30)^3}{180}\right]_{x=0}^{x=60}} = \sqrt{2\times\frac{30^3}{180}}$$

$$= \sqrt{300} = 10\sqrt{3} = 17{,}32050807569$$

$$\boxed{\sigma = 17 \text{ s}}$$

**Réponse 3 : BCD**

## 4°/ Probabilité pour que le véhicule attende moins de 20 secondes.

La fonction de répartition de la variable aléatoire $X$ est :

- Pour $x < 0$ : $F(x) = P(X \leq x) = 0$
- Pour $x > 60$ : $F(x) = 1$
- Pour $0 \leq x \leq 60$ : $F(x) = P(X \leq x) = \int_0^x f(x)\, dx = \frac{x}{60}$

D'où :

$$P(X < 20) = P(X \leq 20) = F(20) = \frac{20}{60} = \frac{1}{3} = 0{,}33$$

$$\boxed{P(X < 20) = 0{,}33}$$

**Réponse 4 : AE**

## 5°/ Probabilité pour que le véhicule attende entre 15 et 30 secondes.

$$P(15 < X < 30) = F(30) - F(15) = \frac{30-15}{60} = \frac{1}{4} = 0{,}25$$

$$\boxed{P(15 < X < 30) = 0{,}25}$$

**Réponse 5 : AC**

## 6°/ Probabilité pour que | $X - E(X)$ | soit inférieur ou égal à σ.

$P(\,|X - E(X)| \leq \sigma\,)$ $= P(E(X) - \sigma \leq X \leq E(X) + \sigma) = P(30 - 17 \leq X \leq 30 + 17) = P(13 \leq X \leq 47)$

$= F(47) - F(13) = \frac{47-13}{60} = \frac{34}{60} = 0{,}57$

$$\boxed{P(\,|X - E(X)| \leq \sigma\,) = 0{,}57}$$

**Réponse 6 : CD**

L'inégalité de Bienaymé-Tchebitcheff donne :

$$P(\,|X - E(X)| \geq k\,\sigma\,) \leq \frac{1}{k^2}$$

# Exercice 36.

Calculer l'espérance mathématique et la variance d'une variable aléatoire gaussienne *X* sachant que :

$$P(X \leq 2) = 0,5793 \quad \text{et} \quad P(X > 5) = 0,2119$$

# SOLUTION.

La ***table de la fonction de répartition*** de la variable normale centrée réduite (*Cours polycopié*, page 61) donne :

$$0,5793 = F(0,20)$$

$$1 - 0,2119 = 0,7881 = F(0,80)$$

On en déduit :

$$\frac{2 - E(X)}{\sigma} = 0,20$$

$$\frac{5 - E(X)}{\sigma} = 0,80$$

d'où :

$$\frac{5 - E(X)}{2 - E(X)} = \frac{0,80}{0,20} = 4$$

$$3\,E(X) = 3$$

$$\boxed{E(X) = 1}$$

$$2 - E(X) = 0,20\,\sigma$$

$$\sigma = 10 - 5\,E(X) = 10 - 5 = 5$$

$$\boxed{\sigma = 5}$$

La variance $\sigma^2$ est donc égale à $5^2 = 25$.

$$\boxed{\sigma^2 = 25}$$

# Exercice 37.

On suppose que la taille de 615 étudiants est distribuée normalement avec une moyenne de 1,75 m et un écart-type de 20 cm. Calculer le nombre d'étudiants ayant des tailles :

- inférieures ou égales à 1,50 m
- comprises entre 1,50 m et 1,65 m
- supérieures ou égales à 2 m.

# SOLUTION.

Soit $X$ la taille d'un étudiant. Le nombre d'étudiants ayant une taille comprise entre deux limites $a$ et $b$ est :

$$N = 615 \times P(a < X < b) = 615 \times \left[ F\left(\frac{b-1,75}{0,20}\right) - F\left(\frac{a-1,75}{0,20}\right) \right] = 615 \times [F(5b - 8,75) - F(5a - 8,75)].$$

On peut donc dresser le tableau suivant des résultats :

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $X$ | $-\infty$ | | $1,50$ | | $1,65$ | | $2$ | | $+\infty$ |
| $5(X-1,75)$ | $-\infty$ | | $-1,25$ | | $-0,50$ | | $1,25$ | | $+\infty$ |
| $F(5X-8,75)$ | $0$ | | $0,1056$ | | $0,3085$ | | $0,8944$ | | $1$ |
| $P(X < 1,5)$ | $\uparrow$ | $0,1056$ | $\uparrow$ | | | | | | |
| $P(1,5 < X < 1,65)$ | | | $\uparrow$ | $0,2029$ | $\uparrow$ | | | | |
| $P(X \geq 2)$ | | | | | | | $\uparrow$ | $0,1056$ | $\uparrow$ |
| $N$ | $\uparrow$ | $65$ | $\uparrow$ | $125$ | $\uparrow$ | | $\uparrow$ | $65$ | $\uparrow$ |

# Exercice 38.

On jette un dé 6 000 fois. On appelle $X$ la variable aléatoire correspondant au nombre de fois où on observe la face « 1 » après les 6 000 jets.

1. Montrer que $X$ est une variable binomiale.
2. Justifier l'approximation de la loi binomiale par la loi de Gauss.
3. En utilisant l'approximation gaussienne, calculer les probabilités suivantes :
   $P\ (\ 980 \leq X \leq 1\ 030\ )$
   $P\ (\ 980 < X < 1\ 030\ )$

# SOLUTION.

## 1°/ Loi de probabilité de la variable aléatoire $X$.

L'épreuve consistant à jeter le dé et regarder si la face supérieure porte le chiffre « 1 » est une épreuve de Bernoulli avec deux résultats possibles :

— le succès, la face supérieure porte le chiffre « 1 », de probabilité $p = \frac{1}{6}$

— l'échec, la face supérieure ne porte pas le chiffre « 1 », de probabilité $q = 1 - p = \frac{5}{6}$.

Lorsqu'on répète cette épreuve 6 000 fois, le nombre $X$ de succès, c'est-à-dire le nombre de « 1 » qui sont sortis, suit une loi binomiale de paramètres $n$ = 6 000 et $p = \frac{1}{6}$.

$$P(X = k) = C_{6000}^{k}\,\frac{5^{6000-k}}{6^{6000}}$$

La probabilité pour que $X$ soit compris entre 980 et 1 030 est donc :

$$P(980 \leq X \leq 1\ 030) = \sum_{k=980}^{k=1030} C_{6000}^{k}\,\frac{5^{6000-k}}{6^{6000}} = \left(\frac{5}{6}\right)^{6000} \times \sum_{k=980}^{k=1030} \frac{C_{6000}^{k}}{5^{k}}$$

Les calculettes donnent un résultat aberrant.

## 2°/ Approximation de la loi binomiale par la loi de Gauss.

Comme le calcul de la loi binomiale à la machine dure un certain temps (5 minutes) et donne finalement un résultat aberrant 0, on peut envisager de faire le calcul de la loi binomiale par approximation par une loi de Gauss. En effet, la variance $n\,p\,q$ est égale à 833 : elle est donc supérieure à 20, ce qui justifie l'approximation de la loi binomiale de paramètres $n$ et $p$ par une loi de Gauss de paramètres :

$\mu = n\,p = 1\ 000$

$\sigma^2 = n\,p\,q = \frac{5000}{6}$

Dans cette approximation :

- la probabilité binomiale d'une valeur $k$ est approchée par la valeur de la densité de probabilité normale :

$$P\ (\ X = k\ ) \approx \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{(k-\mu)^2}{2\sigma^2}} = \sqrt{\frac{3}{10000\ \pi}} \times e^{-\frac{6\times(k-1000)^2}{10000}}$$

**Exemple** : $P\ (\ X = 1\ 030\ ) \approx \sqrt{\frac{3}{10000\ \pi}} \times e^{-0{,}54} = 8{,}05 \times 10^{-3}$.

On peut prendre aussi la surface sous la courbe de Gauss, dans un intervalle $[\ k - 0{,}5\ ;\ k + 0{,}5\ ]$ de longueur 1 entourant $k$ :

$$P\ (\ X = k\ ) \approx F\left(\frac{k + 0{,}5 - \mu}{\sigma}\right) - F\left(\frac{k - 0{,}5 - \mu}{\sigma}\right)$$

**Exemple** : $P\ (X = 1\ 030) \approx F\left(30{,}5 \times \sqrt{\dfrac{6}{5000}}\right) - F\left(29{,}5 \times \sqrt{\dfrac{6}{5000}}\right) = F\ (1{,}057) - F\ (1{,}022)$

$= 0{,}85464 - 0{,}84659 = 0{,}00805$

- Pour un intervalle [ $a$ ; $b$ ] de valeurs de $X$, la probabilité binomiale s'obtiendra en faisant la somme des probabilités des valeurs entières contenues dans l'intervalle. La somme des surfaces sous la courbe de Gauss donne l'approximation normale avec correction de continuité (*Cours polycopié*, page50) :

$$
\begin{aligned}
P(980 \le X \le 1\ 030) &= F\left(\frac{1030{,}5 - 1000}{28{,}86751345948}\right) - F\left(\frac{979{,}5 - 1000}{28{,}86751345948}\right) \\
&= F(1{,}056550992617) - F(-0{,}7101408311033) \\
&= F(1{,}056550992617) + F(0{,}7101408311033) - 1 \\
&= 0{,}854641721579 + 0{,}761191595859 - 1 \\
&= 0{,}6158331744
\end{aligned}
$$

**Remarque** : Le calcul sans correction de continuité donne : $P(980 \le X \le 1\ 030) = 0{,}606440063691$. Nous retiendrons donc la valeur avec correction de continuité :

$$\boxed{P(980 \le X \le 1\ 030) = 0{,}6158}$$

Le calcul avec la ***table de la fonction de répartition*** de la variable normale centrée réduite donne, par interpolation linéaire :

$$P(980 \le X \le 1\ 030) = F(1{,}057) + F(0{,}710) - 1 = 0{,}8547 + 0{,}7611 - 1 = 0{,}6158$$

Sa précision est donc bien suffisante, par rapport à celle de la machine à calculer.

En enlevant les valeurs entières bornes de l'intervalle, il reste :

$$
\begin{aligned}
P(980 < X < 1\ 030) &= F\left(29{,}5 \times \sqrt{\frac{6}{5000}}\right) - F\left(-19{,}5 \times \sqrt{\frac{6}{5000}}\right) \\
&= F\ (1{,}02190997647) - F\ (-0{,}675499814952) \\
&= 0{,}846588244907 - 0{,}249679134450 \\
&= 0{,}596909110457 \\
&\approx 0{,}5969
\end{aligned}
$$

# Exercice 39.

On lance un dé 450 fois. On appelle « *succès* » le fait d'obtenir un 5 ou un 6, et « *échec* » le fait d'obtenir un autre chiffre.

1°/ Soit $X_i$ la variable aléatoire « *Nombre de succès observés au* $i^{ème}$ *lancer* ». Ecrire la distribution de probabilité de $X_i$ ; calculer son espérance mathématique et sa variance.

2°/ Soit $Y$ la variable aléatoire « *Nombre total de succès observés après les 450 lancers* ». En utilisant le théorème central limite, déduire la loi de probabilité suivie par $Y$ ; puis calculer $E(Y)$ et $Var(Y)$.

# SOLUTION.

## 1°/ Loi de probabilité du nombre de succès lors du i^ème^ lancer.

Les valeurs possibles de la variable aléatoire $X_i$ sont 0 et 1. La valeur 0 correspond à zéro succès, c'est-à-dire à l'échec. La valeur 1 correspond au succès. Lors du $i^{ème}$ lancer, les évènements élémentaires constituant le « succès » sont les évènements « 5 » et « 6 ». La probabilité du succès est donc :

$$P(X_i = 1) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}} = \frac{2}{6} = \frac{1}{3}$$

et la probabilité de l'échec est :

$$P(X_i = 0) = \frac{2}{3}$$

La loi de probabilité de la variable aléatoire $X_i$ est donc une *loi de Bernoulli*, donnée par :

| Valeurs $k$ de $X_i$ | 0 | 1 |
|---|---|---|
| $P(X_i = k)$ | $\frac{2}{3}$ | $\frac{1}{3}$ |

L'*espérance mathématique* et la *variance* de la variable aléatoire $X_i$ se déduisent de cette loi de probablité :

$$E(X_i) = \frac{2}{3} \times 0 + \frac{1}{3} \times 1 = \frac{1}{3}$$

$$Var(X_i) = \frac{2}{3} \times \left(0 - \frac{1}{3}\right)^2 + \frac{1}{3} \times \left(1 - \frac{1}{3}\right)^2 = \frac{2}{27} + \frac{4}{27} = \frac{6}{27} = \frac{2}{9}$$

$$\boxed{E(X_i) = \frac{1}{3} \qquad Var(X_i) = \frac{2}{9}}$$

## 2°/ Loi de probabilité du nombre de succès après 450 lancers.

Le théorème central limite indique que lorsque le nombre $n$ de lancers tend vers l'infini, la loi de probabilité de la moyenne $\overline{X} = \frac{1}{n}\sum_{i=1}^{i=n} X_i$ des variables aléatoires $X_i$ correspondant aux $n$ lancers, tend vers une *loi de Gauss* de moyenne $E(X_i)$ et de variance $\frac{1}{n} Var(X_i)$. Pour $n = 450$, on peut donc affirmer, par multiplication par $n$, que la loi de probabilité de la variable aléatoire $Y = 450 \times \overline{X} = 450 \times \left( \frac{1}{450} \times \sum_{i=1}^{i=450} X_i \right)$ est, à peu près, une loi de Gauss de moyenne :

$$\boxed{E(Y) = 450 \times E(X_i) = 150}$$

et de variance :

$$\boxed{Var(Y) = 450 \times Var(X_i) = 100}$$

# Exercice 40.

Soit U $\rightarrow$ N ( 0 , 1 ), une variable aléatoire de Gauss centrée réduite
Soit T $\rightarrow$ loi de Student à $\nu$ = 5 ddl, une variable aléatoire de Student à 5 degrés de liberté,

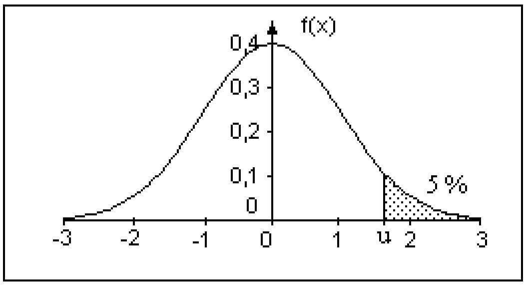
Calculer pour chacune des deux variables aléatoires les intervalles de pari au risque "$\alpha$" pour $\alpha$ = 0,05 et $\alpha$ = 0,01. On calculera les intervalles unilatéraux et les intervalles bilatéraux (en répartissant alors "$\alpha$" de façon équivalente à droite et à gauche).

# SOLUTION.

## 1°) Loi de Gauss.

- **Risque unilatéral.**
  - **$\alpha$ = 0,05**

On cherche la valeur u de la variable normale U centrée réduite telle que la surface hachurée soit égale à 5 %.




La table de la fonction de répartition de la variable de Gauss centrée réduite donne, par interpolation linéaire :

F( 1, 645 ) = 0, 9500

On a donc : Prob ( U $\in$ ] – $\infty$ ; 1,645 ] ) = 0,95 = 1 – 0,05, et ***l'intervalle de pari au risque 5 % à droite*** est

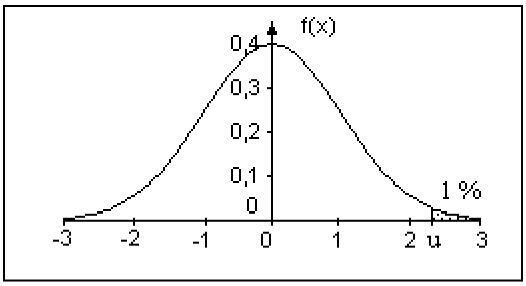
L'intervalle ] – $\infty$ ; 1,645 ]

Par symétrie, on obtient F ( – 1,645 ) = 1 – F ( 1,645 ) = 0,05. On a donc Prob ( U $\in$ [ – 1,645 ; + $\infty$ [ ) = 0,95 = 1 - 0,05 et ***l'intervalle de pari au risque 5 % à gauche*** est :

L'intervalle [ – 1,645 ; + $\infty$ [

  - **$\alpha$ = 0,01**

On cherche la valeur u de la variable normale U centrée réduite telle que la surface hachurée soit égale à 1 %.




La table de la fonction de répartition de la variable de Gauss donne, par interpolation linéaire :

F( 2,327 ) = 0, 9900

On a donc : Prob ( U ∈ ] – ∞ ; 2,327 ] ) = 0,99 = 1 – 0,01, et ***l'intervalle de pari au risque 1 % à droite*** est

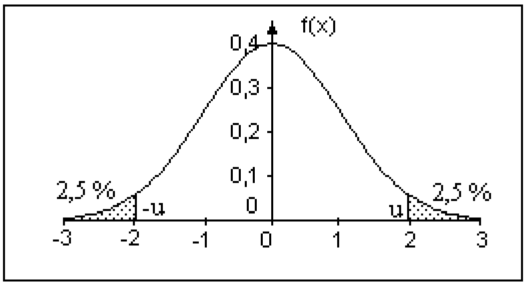
L'intervalle ] – ∞ ; 2,327 ]

Par symétrie, on obtient F ( – 2,327 ) = 1 – F ( 2,327 ) = 0,01. On a donc Prob ( U ∈ [ – 2,327 ; + ∞ [ ) = 0,99 = 1 - 0,01 et ***l'intervalle de pari au risque 1 % à gauche*** est :

L'intervalle [ – 2,327 ; + ∞ [

- **Risque bilatéral**
  - **$\alpha$ = 0,05**

    La densité de probabilité de la variable de Gauss est une fonction symétrique. Si l'on répartit le risque également à droite et à gauche, on cherche la valeur u de la variable normale U centrée réduite telle que la surface hachurée soit égale à 5 %.




La table de la fonction de répartition de la variable aléatoire normale centrée réduite donne :

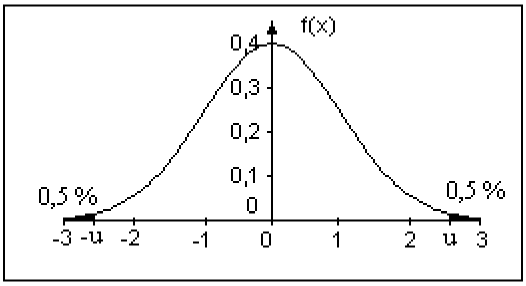
F ( 1,96 ) = 0, 975 0 = 1 – 0, 025 0.

On a donc Prob ( U ∈ [ – 1, 96 ; + 1, 96 ] ) = 0, 95 = 1 – 0, 05 et l'intervalle de pari au risque bilatéral de 5 % est

L'intervalle [ – 1, 96 ; + 1, 96 ]

  - **$\alpha$ = 1 %**

    La densité de probabilité de la variable de Gauss est une fonction symétrique. Si l'on répartit le risque également à droite et à gauche, on cherche la valeur u de la variable normale U centrée réduite telle que la surface hachurée soit égale à 1 %.




La table de la fonction de répartition de la variable aléatoire normale centrée réduite donne, par interpolation linéaire :

F ( 2, 575 ) = 0, 995 0 = 1 – 0, 005 0

On a donc : Prob ( U ∈ [ – 2, 575 ; + 2, 575 ] ) = 0,99 et l'intervalle de pari au risque bilatéral de 1 % est :

L'intervalle [ – 2, 575 ; + 2, 575 ]

# 2°) Loi de Student.

- **Fonction de répartition de la variable de Student à 5 degrés de liberté.**

La fonction de répartition de la loi de Student à 5 degrés de liberté est la fonction :

$$F_5(t) = \frac{1}{\sqrt{5\pi}} \frac{\Gamma\left(\frac{5+1}{2}\right)}{\Gamma\left(\frac{5}{2}\right)} \int_{-\infty}^{t} \left(1+\frac{u^2}{5}\right)^{-\frac{5+1}{2}} du = \frac{1}{\sqrt{5\pi}} \frac{2}{\frac{3}{4}\sqrt{\pi}} \int_{-\infty}^{t} \frac{du}{\left(1+\frac{u^2}{5}\right)^3} = \frac{8}{3\pi\sqrt{5}} \int_{-\infty}^{t} \frac{du}{\left(1+\frac{u^2}{5}\right)^3}$$

Pour calculer l'intégrale, on pose $\frac{u}{\sqrt{5}}$ = tg x, d'où du = $\sqrt{5}$ ( 1 + tg² x ) dx.

Pour u tendant vers – ∞, x tend vers $-\frac{\pi}{2}$. Pour u = t, x = Arc tg $\frac{t}{\sqrt{5}}$. On obtient donc :

$$F_5(t) = \frac{8}{3\pi} \int_{-\frac{\pi}{2}}^{\mathrm{Arctg}\frac{t}{\sqrt{5}}} \frac{dx}{(1+\mathrm{tg}^2 x)^2} = \frac{8}{3\pi} \int_{-\frac{\pi}{2}}^{\mathrm{Arctg}\frac{t}{\sqrt{5}}} \cos^4 x \, dx .$$

$$\cos^4 x = \frac{1}{8} ( 3 + 4 \cos 2x + \cos 4x )$$

$$F_5(t) = \frac{1}{\pi}\mathrm{Arc\,tg}\frac{t}{\sqrt{5}} + \frac{1}{2} + \frac{2}{3\pi}\sin\left(2\,\mathrm{Arc\,tg}\frac{t}{\sqrt{5}}\right) + \frac{1}{12\pi}\sin\left(4\,\mathrm{Arc\,tg}\frac{t}{\sqrt{5}}\right)$$

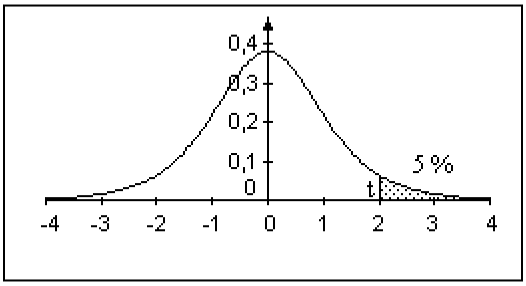
$$F_5(t) = \frac{1}{\pi}\mathrm{Arc\,tg}\frac{t}{\sqrt{5}} + \frac{1}{2} + \frac{4\sqrt{5}}{3\pi}\frac{t}{t^2+5} + \frac{\sqrt{5}}{3\pi}\frac{t(5-t^2)}{(5+t^2)^2}$$

$$\boxed{F_5(t) = \frac{1}{2} + \frac{1}{\pi}\mathrm{Arc\,tg}\frac{t}{\sqrt{5}} + \frac{\sqrt{5}}{3\pi}\frac{t(3t^2+25)}{(t^2+5)^2}}$$

Bien que cette fonction soit facile à calculer à l'aide d'une calculette, il en existe des tables.

- **Risque unilatéral**

  ∗ **α = 5 %**
  On cherche la valeur t de la variable aléatoire de Student à 5 degrés de liberté, telle que la surface hachurée soit égale à 5 %.




La table de la fonction de répartition de la variable de Student à 5 degrés de liberté fournit :

$$F_5 ( 2{,}015) = 0{,}95 = 1 - 0{,}05$$

tandis que la résolution à la machine de l'équation $F_5$ ( t ) = 0,95 fournit t = 2, 015 048 373 34. On a donc :

$$\mathrm{Prob}\,( T \in \,]-\infty \,;\, 2{,}015\,]\,) = 0{,}95$$

et l'***intervalle de pari au risque de 5 % à droite*** est :

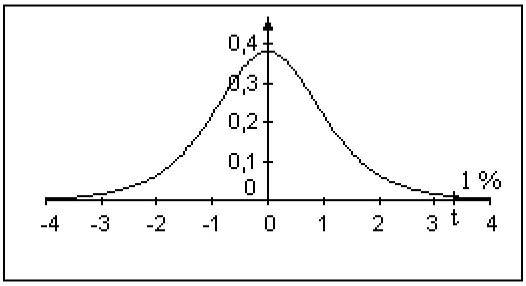
$$\boxed{\text{L'intervalle } ]-\infty \,;\, 2{,}015\,]}$$

Comme la densité de probabilité $f_5(t) = \frac{8}{3\pi\sqrt{5}} \frac{1}{\left(1+\frac{t^2}{5}\right)^3}$ est une fonction paire, donc symétrique par rapport à t=0, on obtient, par symétrie : $F_5$ ( – 2,015) = 1 – $F_5$ ( 2,015) = 0,05

On a donc : Prob ( T ∈ ] – 2, 015 ; + ∞ ] ) = 0,95 et l'***intervalle de pari au risque de 5 % à gauche*** est :

| L'intervalle ] – 2, 015 ; + ∞ ] |
|---|

∗ **α = 1 %**
On cherche la valeur t de la variable aléatoire de Student à 5 degrés de liberté, telle que la surface hachurée soit égale à 1 %.




La table de la fonction de répartition de la variable de Student à 5 degrés de liberté fournit :

$$F_5 ( 3, 365) = 0,99 = 1 – 0,01$$

tandis que la résolution à la machine de l'équation $F_5$ ( t ) = 0,95 fournit t = 3, 364 929 998 92. On a donc : Prob ( T ∈ ] – ∞ ; 3, 365 ] ) = 0,99 et l'***intervalle de pari au risque de 1 % à droite*** est :

| L'intervalle ] – ∞ ; 3, 365 ] |
|---|

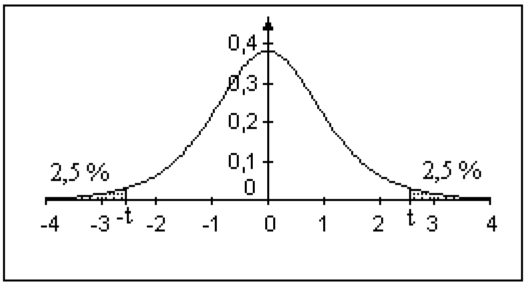
Comme la densité de probabilité $f_5 ( t ) = \frac{8}{3\pi\sqrt{5}} \frac{1}{\left(1+\frac{t^2}{5}\right)^3}$ est une fonction paire, donc symétrique par rapport à t=0, on obtient, par symétrie : $F_5$ ( – 3, 365) = 1 – $F_5$ ( 3, 365) = 0,01

On a donc : Prob ( T ∈ ] – 3, 365 ; + ∞ ] ) = 0,99 et l'***intervalle de pari au risque de 1 % à gauche*** est :

| L'intervalle ] – 3, 365 ; + ∞ ] |
|---|

- **Risque bilatéral.**

∗ **α = 5 %**
La densité de probabilité de la variable de Student est une fonction symétrique. Si l'on répartit le risque également à droite et à gauche, on cherche la valeur u de la variable normale U centrée réduite telle que la surface hachurée soit égale à 5 %.




La table de la fonction de répartition de la variable de Student à 5 degrés de liberté fournit :

$$F_5 ( 2, 571) = 0,975 = 1 – 0,025$$

tandis que la résolution à la machine de l'équation $F_5$ ( t ) = 0,975 fournit t = 2, 570 581 835 66. On a donc :

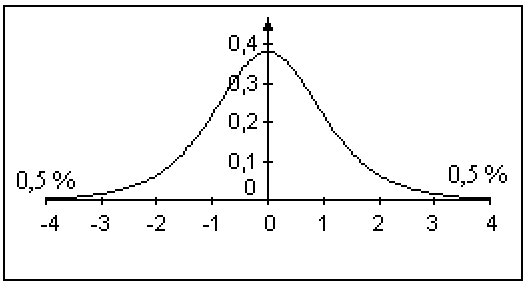
$$\text{Prob} ( T \in ] – 2, 571 ; 2, 571 ] ) = 0,95$$

et l'***intervalle de pari au risque bilatéral de 5 %*** est :

| L'intervalle ] – 2, 571 ; 2, 571 ] |
|---|

* **$\alpha$ = 1 %**

La densité de probabilité de la variable de Student est une fonction symétrique. Si l'on répartit le risque également à droite et à gauche, on cherche la valeur t de la variable de Student T à 5 degrés de liberté telle que la surface hachurée soit égale à 1 %.




La table de la fonction de répartition de la variable de Student à 5 degrés de liberté fournit :

$$F_5 ( 4, 032) = 0,995 = 1 - 0,005$$

tandis que la résolution à la machine de l'équation $F_5$ ( t ) = 0,995 fournit t = 4, 032 142 983 76. On a donc :

$$\text{Prob} ( T \in \, ] - 4, 032 ; 4, 032 ] ) = 0,99$$

et l'***intervalle de pari au risque bilatéral de 1 %*** est :

| |
|---|
| L'intervalle ] – 4, 032 ; 4, 032 ] |

# REVISIONS

# Exercice 41.

Un joueur lance une fléchette au hasard dans une cible de rayon 1 et découpée en couronnes concentriques par des cercles de rayons $\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}$ (« au hasard » signifiant que la fléchette touche la cible et que la probabilité pour que la fléchette pointe dans un domaine donné est proportionnelle à l'aire de ce domaine). Si la fléchette touche la cible dans la couronne limitée par les cercles de rayon $\frac{i}{n}$ et $\frac{i-1}{n}$, pour $1 \leq i \leq n$, le joueur gagne $n - i$ francs. On appelle $X$ la variable aléatoire représentant le gain du joueur.

1. Soit $D$ un domaine de la cible, d'aire $a(D)$. Quelle est la probabilité pour que la fléchette pointe dans le domaine $D$ ?
2. Déterminer la loi de la variable $X$ et calculer son espérance (on rappelle que $\sum_{k=1}^{k=n} k^2 = \frac{n(n+1)(2n+1)}{6}$).

# SOLUTION.

## 1.1. Probabilité du domaine *D*.

Comme on suppose que la fléchette pointe forcément dans la cible, la probabilité du domaine $D$ est le rapport de l'aire du domaine $D$ à l'aire $2\pi$ de la cible :

$$\boxed{P(D) = \frac{a(D)}{2\pi}}$$

## 1.2. Loi de probabilité du gain.

La probabilité pour que le gain $X$ soit égal à $n - i$ est égale à la probabilité pour que la fléchette atteigne la cible dans la couronne limitée par les cercles de rayons $\frac{i}{n}$ et $\frac{i-1}{n}$. Cette probabilité est égale au rapport de l'aire de la couronne à l'aire totale de la cible :

$$P(X = n - i) = \frac{2\pi\left(\frac{i}{n}\right)^2 - 2\pi\left(\frac{i-1}{n}\right)^2}{2\pi} = \left(\frac{i}{n}\right)^2 - \left(\frac{i-1}{n}\right)^2 = \frac{1}{n^2}(2\,i - 1)$$

$$\boxed{P(X = n - i) = \frac{1}{n^2}(2\,i - 1)}$$

L'espérance mathématique de $X$ est donnée par la formule :

$$E(X) = \sum_{i=1}^{i=n} (n - i)\,P(X = n - i) = \frac{1}{n^2}\sum_{i=1}^{i=n} (n - i)(2\,i - 1)$$

$$E(X) = \frac{2}{n}\sum_{i=1}^{i=n} i - \frac{2}{n^2}\sum_{i=1}^{i=n} i^2 - \frac{1}{n}\sum_{i=1}^{i=n} 1 + \frac{1}{n^2}\sum_{i=1}^{i=n} i = \frac{2n+1}{n^2}\sum_{i=1}^{i=n} i - 1 - \frac{2}{n^2}\sum_{i=1}^{i=n} i^2$$

$$\sum_{i=1}^{i=n} i = \frac{n(n+1)}{2} \; ; \; \sum_{i=1}^{i=n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$E(X) = \frac{(n+1)(2n+1)}{2n} - 1 + \frac{(n+1)(2n+1)}{3n} = \frac{n+1}{6n}\,(3\,(2\,n+1) - 6\,n + 2\,(2\,n+1))$$

$$E(X) = \frac{n+1}{6n}\,(10\,n+5) = \frac{5(n+1)(2n+1)}{6n}$$

$$\boxed{E(X) = \frac{5(n+1)(2n+1)}{6n}}$$

# Exercice 42.

Dans une banque, le nombre de chèques émis par les clients chaque jour est une variable aléatoire $X$ qui suit une loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$. On suppose que les chèques sont émis indépendamment les uns des autres. Pour un chèque émis, la probabilité pour que ce chèque soit sans provision est $p \in ]\, 0 \,, 1\, [$. On appelle $Y$ la variable aléatoire correspondant au nombre de chèques émis sans provision lors d'une journée.

1. Soit $n \in \mathbb{N}^*$. Montrer que la loi conditionnelle de la variable $Y$ sachant l'évènement $\{ X = n \}$ est une loi binomiale dont on déterminera les paramètres.
2. Déterminer la loi du couple $(X,Y)$.
3. En déduire la loi de la variable aléatoire $Y$ et calculer son espérance.
4. Les variables $Y$ et $X - Y$ sont-elles indépendantes ?

# SOLUTION.

## 2.1. Loi de probabilité conditionnelle de la variable *Y*.

Lorsque le nombre de chèques émis dans une journée est $X = n$, considérons un chèque émis, parmi ces $n$ chèques. De deux choses l'une, ou bien il est couvert par une provision (probabilité $1 - p$), ou bien il est sans provision (probabilité $p$). Le fait de regarder s'il a ou non une provision constitue une épreuve de Bernoulli. Lorsqu'on répète $n$ fois cette épreuve de Bernoulli, le nombre de succès dans cette répétition est le nombre $Y$ de chèques sans provision : on sait que le nombre de succès dans une répétition d'épreuves de Bernoulli, suit une loi binomiale de paramètre $n$ (nombre de répétitions) et $p$ (probabilité du succès). Lorsque le nombre de chèques émis en une journée est $X = n$, le nombre $Y$ de chèques sans provision suit une loi binomiale de paramètres $n$ et $p$ :

$$\boxed{P\,(Y = k) = \binom{n}{k} p^k\,(1 - p)^{n-k}}$$

où $\binom{n}{k}$ est le coefficient binomial d'indices $n$ et $k$, qu'on peut généraliser en prenant une valeur nulle lorsque $k$ est strictement plus grand que $n$.

## 2.2. Loi de probabilité conjointe du couple *(X,Y)*.

La loi de probabilité du couple $(X,Y)$ est donnée par :

$$P\,(X = n \text{ et } Y = k) = P\,(Y = k \mid X = n\,) \times P\,(X = n)$$

par définition de la probabilité conditionnelle $P\,(Y = k \mid X = n\,)$. Or on a :

$$P\,(Y = k \mid X = n\,) = \binom{n}{k} p^k\,(1 - p)^{n-k}$$

$$P\,(X = n) = e^{-\lambda}\,\frac{\lambda^n}{n!}$$ , par définition de la loi de Poisson de paramètre $\lambda$.

On en déduit la loi de probabilité conjointe du couple $(X,Y)$ :

$$P\,(X = n \text{ et } Y = k) = e^{-\lambda}\,\frac{\lambda^n}{n!}\binom{n}{k} p^k\,(1 - p)^{n-k} \quad \text{pour } 0 \le k \le n$$

$$P\,(X = n \text{ et } Y = k) = 0 \text{ pour } k > n.$$

$$\boxed{P\,(X = n \text{ et } Y = k) = \lambda^n\, e^{-\lambda}\,\frac{p^k}{k!}\,\frac{(1-p)^{n-k}}{(n-k)!}}$$

## 2.3. Loi de probabilité de la variable *Y*.

L'évènement $Y = k$ est la réunion des évènements incompatibles ($Y = k$ et $X = n$) pour les valeurs de $n \geq k$.

La loi de probabilité de la variable aléatoire *Y* est donnée par la somme des probabilités :

$$P\,(Y = k) = \sum_{n \geq k} P\,(Y = k \text{ et } X = n)$$

$$P\,(Y = k) = \sum_{n \geq k} P\,(Y = k \text{ et } X = n) = e^{-\lambda}\, \frac{p^k}{k\,!} \sum_{n \geq k} \frac{(1-p)^{n-k}}{(n-k)!}\, \lambda^n$$

Or le développement en série de l'exponentielle donne :

$$e^{\lambda(1-p)} = \sum_{j \geq 0} \frac{(\lambda(1-p))^j}{j\,!}$$

On a donc :

$$P\,(Y = k) = e^{-\lambda}\, \frac{(\lambda p)^k}{k\,!}\, e^{\lambda(1-p)}$$

$$\boxed{P\,(Y = k) = e^{-\lambda p}\, \frac{(\lambda p)^k}{k\,!}}$$

C'est une loi de Poisson de paramètre $\lambda\, p$.

Son espérance mathématique est :

$$\boxed{E\,(Y) = \lambda\, p}$$

Sa variance est aussi $\lambda\, p$.

## 2.4. Indépendance des variables *Y* et *X* – *Y*.

Dire que les variables *Y* et $X - Y$ sont indépendantes, c'est dire que, quelles que soient les valeurs entières *k* et *k'*, on a :

$$P\,(Y = k \text{ et } X - Y = k') = P\,(Y = k) \times P\,(X - Y = k')$$

Or on a :

$$(Y = k \text{ et } X - Y = k') \Leftrightarrow (Y = k \text{ et } X = k + k')$$

$$P\,(Y = k \text{ et } X - Y = k') = P\,(Y = k \text{ et } X = k + k') = \lambda^{k+k'}\, e^{-\lambda}\, \frac{p^k}{k\,!}\, \frac{(1-p)^{k'}}{k'\,!} = e^{-\lambda}\, \frac{(\lambda p)^k}{k\,!}\, \frac{(\lambda(1-p))^{k'}}{k'\,!}$$

$$P\,(Y = k \text{ et } X - Y = k') = e^{-(\lambda p + \lambda(1-p))}\, \frac{(\lambda p)^k}{k\,!}\, \frac{(\lambda(1-p))^{k'}}{k'\,!} = e^{-\lambda p}\, \frac{(\lambda p)^k}{k\,!} \times e^{-\lambda(1-p)}\, \frac{(\lambda(1-p))^{k'}}{k'\,!}$$

$e^{-\lambda p}\, \frac{(\lambda p)^k}{k\,!}$ est la probabilité pour que $Y = k$ ; $e^{-\lambda(1-p)}\, \frac{(\lambda(1-p))^{k'}}{k'\,!}$ est la probabilité pour le nombre de chèques avec provision soit égal à *k'*, puisque, étant donné un chèque reçu dans la journée, la probabilité pour qu'il soit couvert par une provision est $1 - p$. Or le nombre de chèques avec provision est $X - Y$, donc $P\,(X - Y = k') = e^{-\lambda(1-p)}\, \frac{(\lambda(1-p))^{k'}}{k'\,!}$. On a bien :

$$P\,(Y = k \text{ et } X - Y = k') = P\,(Y = k) \times P\,(X - Y = k')$$

$$\boxed{\text{Les variables } Y \text{ et } X - Y \text{ sont indépendantes}}$$

Comme les chèques sont émis indépendamment les uns des autres, le nombre de chèques sans provision émis dans une journée est indépendant du nombre de chèques émis avec provision suffisante.

# Exercice 43

Soit $a \in R$ et soit $X$ une variable aléatoire continue dont la densité $f$ est définie, pour tout $x \in R$, par :

$$f(x) = a\, x\, e^{-\frac{x^2}{4}}\, \mathbf{1}_{[\,0\,,\,+\infty\,[}(x).$$

où $\mathbf{1}_A$ désigne la fonction indicatrice de l'ensemble $A$.

1. Que vaut $a$ ?
2. On pose $Y = X^2$. Déterminer la fonction de répartition de la variable aléatoire $Y$ et préciser sa densité. Quelle est la loi de la variable aléatoire $Y$ ?
3. Calculer l'espérance de la variable aléatoire $Y$.

# SOLUTION

## 3.1. Valeur de la constante *a*.

La constante $a$ et déterminée par la condition de normalisation :

$$\int_R f(x)\, dx = 1$$

Cette condition s'écrit ici :

$$a \int_0^{+\infty} x\, e^{-\frac{x^2}{4}}\, dx = 1$$

$\int_0^{+\infty} x\, e^{-\frac{x^2}{4}}\, dx$ est la variation d'une primitive de $x\, e^{-\frac{x^2}{4}}$ sur l'ensemble $[\,0\,;+\infty\,[$. Le changement de variable

$$u = \frac{x^2}{4}$$

donne $du = \frac{1}{2}\, x\, dx$

$$\int_0^{+\infty} x\, e^{-\frac{x^2}{4}}\, dx = 2 \int_0^{+\infty} e^{-u}\, du = -2\,[\, e^{-u}\,]_0^{\infty} = 2$$

La condition de normalisation s'écrit : $2\, a = 1$.

$$\boxed{a = \frac{1}{2}}$$

## 3.2. Fonction de répartition de la variable aléatoire *Y*.

La fonction de répartition de la variable aléatoire $Y$ est définie par :

$$F_Y(y) = P(\,Y \leq y)$$

Pour $y \leq 0$, cette fonction de répartition est nulle puisque $Y$ est un carré et n'a que des valeurs positives.
Pour $y > 0$, l'évènement $Y \leq y$ s'écrit $X \leq \sqrt{y}$ et sa probabilité est donnée par :

$$P(Y \leq y) = P(X \leq \sqrt{y}\,) = \frac{1}{2} \int_0^{\sqrt{y}} x\, e^{-\frac{x^2}{4}}\, dx = \int_0^{\frac{y}{4}} e^{-u}\, du = 1 - e^{-\frac{y}{4}}$$

$$\boxed{F_Y(y) = P(\,Y \leq y) = 1 - e^{-\frac{y}{4}}}$$

La densité de probabilité de la variable $Y$ est la dérivée de la fonction de répartition :

— elle est nulle pour $y < 0$ ;

— pour $y > 0$, elle vaut : $\frac{1}{4}\, e^{-\frac{y}{4}}$

$$\boxed{f_Y(y) = \frac{1}{4}\, e^{-\frac{y}{4}}\ \mathbf{1}_{[\,0\,,\,+\infty\,[}(y).}$$

C'est une ***loi exponentielle***.

## 3.3. Espérance de la variable aléatoire *Y*.

L'espérance mathématique de *Y* est définie par :

$$E(Y) = \int_{R} y\, f_Y(y)\, dy$$

Ici, l'intégrale se réduit à :

$$E(Y) = \frac{1}{4} \int_{0}^{+\infty} y\, e^{-\frac{y}{4}}\, dy$$

Pour calculer l'intégrale, on intègre par parties, en posant :

— $u = y$, d'où $du = dy$

— pour avoir $dv = e^{-\frac{y}{4}}\, dy$ , on pose $v = -\,4\, e^{-\frac{y}{4}}$

La formule d'intégration par parties :

$$\int u\, dv = u\, v - \int v\, du$$

donne :

$$\int_{0}^{+\infty} y\, e^{-\frac{y}{4}}\, dy = -\,4\,\left[\, y\, e^{-\frac{y}{4}}\,\right]_{0}^{\infty} + 4 \int_{0}^{+\infty} e^{-\frac{y}{4}}\, dy = -\,16\,\left[\, e^{-\frac{y}{4}}\,\right]_{0}^{\infty} = 16$$

d'où :

$$\boxed{E(Y) = 4}$$

# Exercice 44.

Soit $E = \{a,b,c,d,e\}$. On choisit au hasard un sous-ensemble $F$ de $E$.

1. Décrire un espace probabilisé associé à cette expérience aléatoire.
2. Déterminer les probabilités des deux évènements suivants :
   $a \in F$
   $\{a,c,d\} \subset F$
3. Calculer la probabilité que $\{a,b,c\} \subset F$ sachant que $a \in F$.
4. Calculer la probabilité que $\{a,b\} \subset F$ sachant que $c \notin F$ ou $d \in F$.
5. Soit $X$ la variable aléatoire égale au nombre d'éléments de $F$.
   a) Quelle est la loi suivie par $X$ ?
   b) Donner, sans calcul la valeur de $E(X)$ et de $Var(X)$.

# SOLUTION.

## 1. Espace probabilisé.

Un espace probabilisé est un triplet $(\Omega, \mathsf{F}, P)$ dans lequel :

– $\Omega$ est l'ensemble des évènements élémentaires d'une épreuve,
– $\mathsf{F}$ est l'ensemble des observables de l'épreuve, c'est-à-dire l'ensemble des évènements possibles,
– $P$ une probabilité définie sur $\Omega$.

Dans l'épreuve considérée, les évènements élémentaires sont les parties de $E$ :

$$\Omega = \mathsf{P}(E)$$

Les évènements possibles sont alors les ensembles de parties de $E$ :

$$\mathsf{F} = \mathsf{P}(\mathsf{P}(E))$$

La phrase « On choisit au hasard un sous-ensemble $F$ de $E$ » indique que toutes les parties de $E$ sont équiprobables, chaque partie $F$ de $E$ a une probabilité $\frac{1}{2^5} = \frac{1}{32}$ d'être choisie au hasard :

$$\boxed{F \in P(E) \Rightarrow P(F) = \frac{1}{32}}$$

## Exemple.

On peut prendre pour $E$ l'ensemble des 5 enfants $a, b, c, d, e,$ d'une famille. On étudie la répartition des sexes. $F$ désigne l'ensemble des filles. Chacune des 32 distributions possibles des sexes est supposée équiprobable.

## 2. Probabilité d'évènements.

### 2.1. Evènement $a \in F$.

Comme toutes les parties sont équiprobables, la probabilité pour qu'une partie contienne l'élément $a$ est le rapport entre le nombre de parties contenant $a$ et le nombre total de parties. Le nombre de parties ne contenant pas $a$ est $2^4$, donc le nombre de parties contenant $a$ est $2^5 - 2^4 = 2^4$ et la probabilité de l'évènement $a \in F$ est :

$$P(a \in F) = \frac{2^4}{2^5} = \frac{1}{2}$$

Autrement dit : de deux choses l'une, ou bien $a \in F$, ou bien $a \notin F$, et les deux évènements ont la même probabilité.

Dans notre exemple, il revient au même de dire que la probabilité pour qu'un enfant donné soit une fille est $\frac{1}{2}$.

### 2.2. Evènement $\{a, c, d\} \subset F$.

Il y a $2^2 = 2$ façons de compléter l'ensemble $\{ a , c , d \}$ en une partie de $E$, donc il y a 4 parties de $E$ contenant la partie $\{ a , c , d \}$. La probabilité de l'évènement $\{ a , c , d \} \subset F$ est donc $\frac{4}{32} = \frac{1}{8}$.

$$\boxed{P ( \{ a , c , d \} \subset F) = \frac{1}{8}}$$

Dans notre exemple, cela revient à dire que la probabilité pour que les 1^er^, 3^e^ et 4^e^ enfants soient des filles est de $\frac{1}{8}$ : c'est le produit des probabilités pour que chacun des enfants considérés soient des filles. C'est ce qui se passe lorsque le sexe d'un enfant est indépendant en probabilité du sexe des autres enfants.

# 3. Probabilité conditionnelle.

$$P ( \{ a , c , d \} \subset F \mid a \in F ) = \frac{P(\{a;c;d\} \subset F )}{P(a \in F )} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4}$$

$$\boxed{P ( \{ a , c , d \} \subset F \mid a \in F ) = \frac{1}{4}}$$

Dans notre exemple, cela revient à dire que la probabilité pour que les 1^er^, 3^e^ et 4^e^ enfants soient des filles lorsqu'on sait que le 1^er^ est une fille est de $\frac{1}{4}$ : c'est le produit des probabilités pour que le 3^e^ soit une fille et pour que le 4^e^ soit une fille.

# 4. Probabilité conditionnelle.

$$P ( \{ a ; b \} \subset F \mid (c \notin F \text{ ou } d \in F )) = \frac{P(\{a;b\} \subset F )}{P( c \notin F \text{ ou } d \in F )}$$

Il y a $2^3$ parties de $E$ contenant $\{ a ; b \}$, donc $P ( \{ a ; b \} \subset F ) = \frac{8}{32} = \frac{1}{4}$.

$$P (c \notin F \text{ ou } d \in F ) = P (c \notin F ) + P (d \in F ) - P (c \notin F \text{ et } d \in F )$$

$$P (c \notin F ) = \frac{1}{2}$$

$$P ( d \in F ) = \frac{1}{2}$$

Il y a $2^3$ parties de $E$ contenant $d$ et ne contenant pas $c$, donc $P (c \notin F \text{ et } d \in F ) = \frac{8}{32} = \frac{1}{4}$, d'où :

$$P (c \notin F \text{ ou } d \in F ) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$$

On aurait pu écrire aussi :

$$P (c \notin F \text{ ou } d \in F ) = 1 - P (c \in F \text{ et } d \notin F ) = 1 - \frac{1}{4} = \frac{3}{4}.$$

$$\boxed{P ( \{ a ; b \} \subset F \mid (c \notin F \text{ ou } d \in F )) = \frac{1}{4} \times \frac{4}{3} = \frac{1}{3}}$$

On tombe là sur un paradoxe déjà souligné dans de nombreux jeux de la Revue « Pour la Science » : la probabilité pour que les deux premiers enfants soient des filles lorsqu'on sait que si le 3^e^ est une fille, alors le 4^e^ est une fille, n'est pas égale au produit des probabilités pour que le premier soit une fille et pour que le deuxième soit une fille, même si l'on suppose que les sexes des enfants successifs sont indépendants.

## 5. Loi de probabilité du nombre d'éléments de *F*.

Il y a $C_5^0$ = 1 partie de *E* à 0 élément.
Il y a $C_5^1$ = 5 parties de *E* à 1 élément.
Il y a $C_5^2$ = 10 parties de *E* à 2 éléments.
Il y a $C_5^3$ = 10 parties de *E* à 3 éléments.
Il y a $C_5^4$ = 5 parties de *E* à 4 éléments.
Il y a $C_5^5$ = 1 partie de *E* à 5 éléments.

La loi de probablité de *X* coïncide donc avec la loi binomiale de paramètres $n = 5$ et $p = \frac{1}{2}$.

$$\boxed{X \text{ suit une loi binomiale de paramètres } n = 5 \text{ et } p = \frac{1}{2}}$$

Les paramètres de la variable aléatoire *X* s'en déduisent :

$$E(X) = n\,p = \frac{5}{2}$$

$$Var(X) = n\,p\,q = n\,p\,(1 - p) = \frac{5}{4}$$

Ce résultat signifie que le nombre de filles dans une famille de 5 enfants suit une loi binomiale. On s'y attendait car, pour un enfant, de deux choses l'une : ou c'est un garçon (probabilité $\frac{1}{2}$, dans notre hypothèse), ou c'est une fille (probabilité $\frac{1}{2}$). Regarder de quel sexe est un enfant est une épreuve de Bernoulli, dont la répétition dans une famille de 5 enfants donne, pour le nombre *X* de filles, une variable binomiale de paramètres $n = 5$, et $p = \frac{1}{2}$.

# Exercice 45.

Soit une variable aléatoire $X$ à valeurs dans $N^*$ telle que :

$$P\ (X = n) = \frac{3^{n-1}}{n!}\ \alpha$$

1. Quelle valeur doit-on donner au nombre réel α pour que la loi de $X$ soit parfaitement déterminée ?
2. a) Calculer l'espérance mathématique $E\ (X)$.

b) Calculer la variance $Var\ (X)$.

3. Déterminer la fonction génératrice de $X$ en fonction de α, puis retrouver les valeurs de α, $E(X)$, $Var(X)$.

# SOLUTION.

## 1. Condition de normalisation.

La valeur de α est déterminée par la condition : « la somme des probabilités est égale à 1 ». Cette condition s'écrit :

$$\sum_{n=1}^{\infty} P\ (X = n) = 1$$

$$\sum_{n=1}^{\infty} \frac{3^{n-1}}{n!}\ \alpha = 1$$

$$\frac{\alpha}{3} \sum_{n=1}^{\infty} \frac{3^n}{n!} = 1$$

Or on a :

$$\sum_{n=1}^{\infty} \frac{3^n}{n!} = \sum_{n=0}^{\infty} \frac{3^n}{n!} - 1 = e^3 - 1$$

donc :

$$\boxed{\alpha = \frac{3}{e^3 - 1}}$$

## 2. Paramètres de la variable aléatoire *X*.

$$E\ (X) = \sum_{n=1}^{\infty} n\ P\ (X = n) = \sum_{n=1}^{\infty} n\ \frac{3^{n-1}}{n!}\ \alpha = \frac{3}{e^3 - 1} \sum_{n=1}^{\infty} \frac{3^{n-1}}{(n-1)!} = \frac{3}{e^3 - 1} \times e^3$$

$$\boxed{E\ (X) = 3\ \frac{e^3}{e^3 - 1}}$$

$Var\ (X) = E\ (X^2) - (E\ (X))^2$

$$E\ (X^2) = \sum_{n=1}^{\infty} n^2\ P\ (X = n) = \sum_{n=1}^{\infty} n^2\ \frac{3^{n-1}}{n!}\ \alpha = \frac{3}{e^3 - 1} \sum_{n=1}^{\infty} n\ \frac{3^{n-1}}{(n-1)!}$$

$\sum_{n=1}^{\infty} n \frac{3^{n-1}}{(n-1)!}$ est la valeur, pour $x = 3$, de la dérivée de $\sum_{n=1}^{\infty} \frac{x^n}{(n-1)!} = x\, e^x$. Or la dérivée de $y = x\, e^x$ est :

$$y' = (x + 1)\, e^x$$

Pour $x = 3$, elle vaut :

$$\sum_{n=1}^{\infty} n \frac{3^{n-1}}{(n-1)!} = 4\, e^3$$

$$E(X^2) = \frac{3}{e^3-1} \times 4\, e^3 = 12 \frac{e^3}{e^3-1}$$

$$Var(X) = 12 \frac{e^3}{e^3-1} - \left( 3 \frac{e^3}{e^3-1} \right)^2 = 3 \frac{e^3}{e^3-1} \left( 4 - 3 \frac{e^3}{e^3-1} \right)$$

$$\boxed{Var(X) = 3 \frac{e^3}{(e^3-1)^2} (e^3 - 4)}$$

# 3. Fonction génératrice de la variable aléatoire *X*.

La fonction génératrice de la variable aléatoire $X$ est, par définition, l'espérance mathématique de $u^X$ :

$$FG(u) = \sum_{n=1}^{\infty} u^n P(X = n) = \sum_{n=1}^{\infty} u^n \frac{3^{n-1}}{n!} \alpha = \frac{\alpha}{3} \sum_{n=1}^{\infty} \frac{(3u)^n}{n!} = \frac{\alpha}{3} (e^{3u} - 1)$$

$$\boxed{FG(u) = \frac{\alpha}{3} (e^{3u} - 1)}$$

La fonction génératrice permet de retrouver les paramètres de la distribution de probabilité :

$$FG(1) = \sum_{n=1}^{\infty} P(X = n) = 1 \Rightarrow \frac{\alpha}{3} (e^3 - 1) = 1 \Rightarrow \alpha = \frac{3}{e^3-1} \text{ et } FG(u) = \frac{e^{3u}-1}{e^3-1}$$

$$\frac{d(FG(u))}{du} = 3 \frac{e^{3u}}{e^3-1} \quad \text{et} \quad \frac{d^n(FG(u))}{du^n} = 3^n \frac{e^{3u}}{e^3-1}, \text{ pour } n \geq 1.$$

$$E(X) = \left( \frac{d(FG(u))}{du} \right)_{u=1} = 3 \frac{e^3}{e^3-1}$$

$$E(X^2) = \sum_{n=1}^{\infty} n^2 P(X = n) = \sum_{n=1}^{\infty} n(n-1) P(X = n) + \sum_{n=1}^{\infty} n P(X = n)$$

$$E(X^2) = \left( \frac{d^2(FG(u))}{du^2} \right)_{u=1} + \left( \frac{d(FG(u))}{du} \right)_{u=1} = (3^2 + 3) \frac{e^3}{e^3-1} = 12 \frac{e^3}{e^3-1}$$

$$Var(X) = E(X^2) - (E(X))^2 = 3 \frac{e^3}{e^3-1} \left( 4 - 3 \frac{e^3}{e^3-1} \right) = 3\, e^3 \frac{e^3-4}{(e^3-1)^2}$$

On retrouve bien les mêmes résultats que ceux obtenus précédemment.

# Exercice 46

On observe un phénomène aléatoire. Pendant la durée d'observation, la probabilité que ce phénomène se produise $k$ fois est égale à $(1 - \alpha)\,\alpha^k$ avec $k$ entier positif et $0 < \alpha < 1$. Lorsque ce phénomène se produit, il a une probabilité $p$ d'être enregistré. On note $X$ le nombre de phénomènes enregistrés au cours de la durée d'observation.

1. Calculer l'espérance mathématique de $X$.
2. Calculer la variance de $X$.
3. a) Déterminer la fonction génératrice de $X$.
   b) Retrouver $E(X)$ et $Var(X)$.

# SOLUTION

## 1. Loi de probabilité de *X*.

Soit $Y$ le nombre de fois que le phénomène se produit pendant la durée d'observation. Pour un phénomène qui se produit pendant la durée d'observation, de deux choses l'une : ou bien il est observé (probabilité $p$), ou bien il n'est pas observé (probabilité $1 - p$). Le nombre de phénomènes observés sur les $k$ qui se sont produits suit donc une loi binomiale de paramètres $k$ et $p$ :

$$P\,(X = h \mid Y = k) = \binom{k}{h} p^h\,(1-p)^{k-h}$$

Par hypothèse, on a :

$$P\,(Y = k) = (1-\alpha)\,\alpha^k$$

On en déduit :

$$P\,(X = h \text{ et } Y = k) = P\,(X = h \mid Y = k) \times P\,(Y = k) = (1-\alpha)\,\alpha^k \binom{k}{h} p^h\,(1-p)^{k-h}$$

$$P(X = h) = \sum_{k \geq h} P\,(X = h \text{ et } Y = k) = \sum_{k \geq h} (1-\alpha)\,\alpha^k \binom{k}{h} p^h\,(1-p)^{k-h}$$

$$\boxed{P(X = h) = (1-\alpha) \left(\frac{p}{1-p}\right)^h \sum_{k \geq h} \binom{k}{h} (\,\alpha\,(1-p)\,)^k}$$

## 2. Paramètres de *X*.

$$E\,(X) = \sum_{h \geq 0} h\,P\,(X = h) = (1-\alpha) \sum_{h \geq 0} \sum_{k \geq h} h \left(\frac{p}{1-p}\right)^h \binom{k}{h} (\,\alpha\,(1-p)\,)^k$$

$$E\,(X) = (1-\alpha) \sum_{k \geq 0} (\,\alpha\,(1-p)\,)^k \sum_{h \leq k} h \binom{k}{h} \left(\frac{p}{1-p}\right)^h$$

$$\sum_{h \leq k} h \binom{k}{h} x^h = x \sum_{h=1}^{h=k} \binom{k}{h} h\,x^{h-1} = x \sum_{h=1}^{h=k} \binom{k}{h} \frac{d(x^h)}{dx} = x\,\frac{d}{dx}\left(\sum_{h=1}^{h=k} \binom{k}{h} x^h\right)$$

$$\sum_{h \leq k} h \binom{k}{h} x^h = x\,\frac{d((1+x)^k - 1)}{dx} = k\,x\,(1+x)^{k-1}$$

Pour $x = \dfrac{p}{1-p}$, il vient :

$$\sum_{h \leq k} h \binom{k}{h} \left(\frac{p}{1-p}\right)^h = k\,\frac{p}{1-p} \times \frac{1}{(1-p)^{k-1}} = \frac{k\,p}{(1-p)^k}$$

D'où :

$$E\,(X) = (1-\alpha)\sum_{k\geq 0}\ (\,\alpha\,(1-p)\,)^k\ \frac{k\,p}{(1-p)^k} = p\,\alpha\,(1-\alpha)\sum_{k\geq 0}\ k\,\alpha^{k-1}$$

$$\sum_{k\geq 0}\ k\,x^{k-1} = \frac{d}{dx}\left(\ \sum_{k\geq 0}\ x^k\ \right) = \frac{d}{dx}\left(\ \frac{1}{1-x}\ \right) = \frac{1}{(1-x)^2}$$

$$\sum_{k\geq 0}\ k\,\alpha^{k-1} = \frac{1}{(1-\alpha)^2}$$

$$\boxed{E\,(X) = p\,\frac{\alpha}{1-\alpha}}$$

On a :

$$E\,(Y) = \sum_{k\geq 0}\ k\,P\,(Y=k) = \sum_{k\geq 0}\ k\,(1-\alpha)\,\alpha^k = \alpha\,(1-\alpha)\,\frac{d}{d\alpha}\left(\ \frac{1}{1-\alpha}\ \right) = \frac{\alpha}{1-\alpha}$$

$$E\,(X) = p\,E\,(Y) = E\,(p\,Y)$$

$pY$ est la moyenne de $X$ pour $Y$ fixé : c'est la « moyenne conditionnelle » de $X$, on la note $E_Y\,(X)$, et l'on a :

$$\boxed{E\,(X) = E\,(E_Y\,(X))}$$

La moyenne de $X$ est la moyenne de sa moyenne conditionnelle.

# Variance.

Pour calculer $Var\,(Y)$, calculons $E\,(Y^2)$.

$$E\,(Y^2) = \sum_{k\geq 0}\ k^2\,P\,(Y=k) = \sum_{k\geq 0}\ k^2\,(1-\alpha)\,\alpha^k\ = \alpha\,(1-\alpha) + (1-\alpha)\sum_{k\geq 2}\ k^2\,\alpha^k$$

$$E\,(Y^2) = \alpha\,(1-\alpha) + (1-\alpha)\left(\ \sum_{k\geq 2}\ k\,(k-1)\,\alpha^k + \sum_{k\geq 2}\ k\,\alpha^k\ \right)$$

$$E\,(Y^2) = \alpha\,(1-\alpha) + (1-\alpha)\left(\ \alpha^2\,\frac{d^2}{d\alpha^2}\left(\ \frac{1}{1-\alpha}\ \right) + \alpha\,\frac{d}{d\alpha}\left(\ \frac{1}{1-\alpha}\ \right) - \alpha\ \right)$$

$$E\,(Y^2) = (1-\alpha)\left(\ \alpha^2 \times \frac{2}{(1-\alpha)^3} + \alpha \times \frac{1}{(1-\alpha)^2}\ \right) = \frac{\alpha\,(1+\alpha)}{(1-\alpha)^2}$$

$$Var\,(Y) = E\,(Y^2) - (E\,(Y))^2 = \frac{\alpha\,(1+\alpha)}{(1-\alpha)^2} - \frac{\alpha^2}{(1-\alpha)^2} = \frac{\alpha}{(1-\alpha)^2}$$

$$Var\,(p\,Y) = \frac{p^2\,\alpha}{(1-\alpha)^2}$$

Calculons la variance de $X$ à partir de la loi de probabilité de $X$ :

$$E\,(X^2\,) = \sum_{h\geq 0}\ h^2\,P\,(X=h) = (1-\alpha)\sum_{h\geq 0}\ \sum_{k\geq h}\ h^2\left(\ \frac{p}{1-p}\ \right)^h\binom{k}{h}\,(\,\alpha\,(1-p)\,)^k$$

$$E\,(X^2\,) = (1-\alpha)\sum_{k\geq 0}\ \sum_{h\leq k}\ h^2\left(\ \frac{p}{1-p}\ \right)^h\binom{k}{h}\,(\,\alpha\,(1-p)\,)^k$$

$$E\,(X^2\,) = (1-\alpha)\sum_{k\geq 0}\ (\,\alpha\,(1-p)\,)^k\sum_{h\leq k}\ h^2\left(\ \frac{p}{1-p}\ \right)^h\binom{k}{h}$$

$$\sum_{h\leq k}\binom{k}{h}\,h^2\,x^h = x^2\sum_{h\leq k}\binom{k}{h}\,h\,(h-1)\,x^{h-2} + x\sum_{h\leq k}\binom{k}{h}\,h\,x^{h-1}$$

$$x^2 \sum_{h\le k} \binom{k}{h} h\,(h-1)\,x^{h-2} = x^2 \frac{d^2(1+x)^k}{dx^2} = x^2\, k\,(k-1)\,(1+x)^{k-2}$$

$$x \sum_{h\le k} \binom{k}{h} h\, x^{h-1} = x \frac{d(1+x)^k}{dx} = x\, k\,(1+x)^{k-1}$$

Pour $x = \frac{p}{1-p}$, on a $1 + x = \frac{1}{1-p}$, et :

$$\sum_{h\le k} h^2 \left(\frac{p}{1-p}\right)^h \binom{k}{h} = \left(\frac{1}{1-p}\right)^k (k\,(k-1)\,p^2 + k\,p)$$

$$E\,(X^2) = (1-\alpha)\sum_{k\ge 0} (\,\alpha\,(1-p)\,)^k \times \frac{k(k-1)p^2+kp}{(1-p)^k} = (1-\alpha)\left( p^2 \sum_{k\ge 0} k\,(k-1)\,\alpha^k + p \sum_{k\ge 0} k\,\alpha^k \right)$$

$$\sum_{k\ge 0} k\,\alpha^k = \alpha \frac{d}{d\alpha}\left(\frac{1}{1-\alpha}\right) = \frac{\alpha}{(1-\alpha)^2}$$

$$\sum_{k\ge 0} k\,(k-1)\,\alpha^k = \alpha^2 \frac{d^2}{d\alpha^2}\left(\frac{1}{1-\alpha}\right) = \frac{2\alpha^2}{(1-\alpha)^3}$$

$$E\,(X^2) = p^2 \frac{2\alpha^2}{(1-\alpha)^2} + p \frac{\alpha}{1-\alpha}$$

$$Var\,(X) = E\,(X^2) - (E\,(X))^2 = p^2 \frac{2\alpha^2}{(1-\alpha)^2} + p \frac{\alpha}{1-\alpha} - p^2 \frac{\alpha^2}{(1-\alpha)^2}$$

$$\boxed{Var\,(X) = p \frac{\alpha}{1-\alpha} + p^2 \frac{\alpha^2}{(1-\alpha)^2}}$$

$$Var\,(X) = p^2 \frac{\alpha}{(1-\alpha)^2} \times \frac{1-(1-p)\,\alpha}{p} = p^2 \frac{\alpha}{(1-\alpha)^2}\left(1 + \frac{(1-p)\,(1-\alpha)}{p}\right)$$

$$Var\,(X) = p^2 \frac{\alpha}{(1-\alpha)^2} + p\,(1-p)\,\frac{\alpha}{1-\alpha}$$

$$\boxed{Var\,(X) = Var\,(pY) + E\,(p(1-p)Y)}$$

$pY$ est la moyenne de $X$ lorsque $Y$ est fixé, c'est la « *moyenne conditionnelle* » de $X$ :

$$pY = E_Y\,(X)$$

$p\,(1-p)\,Y$ est la variance de $X$ lorsque $Y$ est fixé, c'est la « *variance conditionnelle* » de $X$ :

$$p\,(1-p)\,Y = \sigma_Y^{\,2}\,(X)$$

La variance totale de $X$ est la somme de la variance de la moyenne conditionnelle $p\,Y$ et de la moyenne de la variance conditionnelle $p\,(1-p)\,Y$ : on dit que « *la variance de X est la somme de la moyenne de la variance et de la variance de la moyenne* ».

# 3. Fonction génératrice de *X*.

La fonction génératrice est l'espérance mathématique de $u^X$ :

$$FG\,(u) = \sum_{h\ge 0} u^h\,P\,(X=h) = (1-\alpha) \sum_{h\ge 0} u^h \sum_{k\ge h} \left(\frac{p}{1-p}\right)^h \binom{k}{h} (\,\alpha\,(1-p)\,)^k$$

$$FG\,(u) = (1-\alpha) \sum_{k\ge 0} \sum_{h\le k} u^h \left(\frac{p}{1-p}\right)^h \binom{k}{h} (\,\alpha\,(1-p)\,)^k$$

$$FG(u) = (1-\alpha) \sum_{k\geq 0} (\alpha(1-p))^k \sum_{h\leq k} u^h \left(\frac{p}{1-p}\right)^h \binom{k}{h}$$

$$\sum_{h\leq k} u^h \left(\frac{p}{1-p}\right)^h \binom{k}{h} = \left(1 + \frac{p\,u}{1-p}\right)^k$$

$$FG(u) = (1-\alpha) \sum_{k\geq 0} (\alpha(1-p))^k \left(1 + \frac{p\,u}{1-p}\right)^k$$

$$FG(u) = (1-\alpha) \sum_{k\geq 0} \alpha^k (1-p(1-u))^k = \frac{1-\alpha}{1-\alpha(1-p(1-u))}$$

$$\boxed{FG(u) = \frac{1-\alpha}{1-\alpha((1-p)+p\,u)}}$$

Sur cette fonction génératrice, on retrouve :

$$FG(1) = 1$$

$$E(X) = \sum_{h\geq 0} h\,P(X=h) = \left(\frac{d(FG(u))}{du}\right)_{u=1}$$

$$\frac{d(FG(u))}{du} = \frac{\alpha p(1-\alpha)}{(1-\alpha((1-p)+pu))^2} \Rightarrow \left(\frac{d(FG(u))}{du}\right)_{u=1} = \frac{\alpha p(1-\alpha)}{(1-\alpha)^2} = p\,\frac{\alpha}{1-\alpha}$$

$$\boxed{E(X) = p\,\frac{\alpha}{1-\alpha}}$$

$$E(X^2) = \sum_{h\geq 0} h^2\,P(X=h) = \sum_{h\geq 0} h(h-1)\,P(X=h) + \sum_{h\geq 0} h\,P(X=h) = \left(\frac{d^2(FG(u))}{du^2}\right)_{u=1} + E(X)$$

$$\frac{d^2(FG(u))}{du^2} = \frac{2\alpha^2 p^2(1-\alpha)}{(1-\alpha((1-p)+pu))^3} \Rightarrow \left(\frac{d^2(FG(u))}{du^2}\right)_{u=1} = \frac{2\alpha^2 p^2(1-\alpha)}{(1-\alpha)^3} = 2\,p^2\,\frac{\alpha^2}{(1-\alpha)^2}$$

$$E(X^2) = 2\,p^2\,\frac{\alpha^2}{(1-\alpha)^2} + p\,\frac{\alpha}{1-\alpha}$$

$$Var(X) = E(X^2) - (E(X))^2 = 2\,p^2\,\frac{\alpha^2}{(1-\alpha)^2} + p\,\frac{\alpha}{1-\alpha} - \left(p\,\frac{\alpha}{1-\alpha}\right)^2 = p^2\,\frac{\alpha^2}{(1-\alpha)^2} + p\,\frac{\alpha}{1-\alpha}$$

$$\boxed{Var(X) = p^2\,\frac{\alpha^2}{(1-\alpha)^2} + p\,\frac{\alpha}{1-\alpha}}$$

Ces résultats sont bien les résultats obtenus soit directement à partir de la loi de probabilité de $X$, soit à partir des formules :

$$\boxed{\begin{gathered} E(X) = E(E_Y(X)) \\ Var(X) = Var(E_Y(X)) + E(Var_Y(X)) \end{gathered}}$$