Opérations sur les variables aléatoires
Lois limites

A. Indépendance de deux variables aléatoires.

Exemple 1.

Pour améliorer le stockage d'un produit un supermarché fait une étude sur la vente de packs de 6 bouteilles d'eau de marques *A* et *B*.

On note :

X, la variable aléatoire mesurant le nombre de packs d'eau de marque *A* achetés ;
Y, la variable aléatoire mesurant le nombre de packs d'eau de marque *B* achetés ;

La probabilité $P(X = x_i \text{ et } Y = y_i)$ est donnée dans le tableau suivant.

X \ Y	1	2	3	Totaux
1	0,1	0,2	0,2	
2	0,1	0,3	0,1	
Totaux				

Définition. Soit *X* et *Y* deux variables aléatoires discrètes

On pose : $X(\Omega) = \{x_1, x_2, \dots, x_i, \dots, x_n\}$
$Y(\Omega) = \{y_1, y_2, \dots, y_j, \dots, x_m\}$

X et *Y* sont deux variables aléatoires indépendantes si et seulement si :

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

Ce qui s'écrit encore : $P((X = x_i) \cap (Y = y_j)) = P(X = x_i) \times P(Y = y_j)$

Exemple 2.

Reprenons l'exemple 1.

Les variables *X* et *Y* sont-elles indépendantes ?

Calculons par exemple $P(X = 1) \times P(Y = 2)$ et comparons avec $P(X = 1 \text{ et } Y = 2)$

Conclusion : Les deux résultats sont , les variables aléatoires *X* et *Y*

Exemple 3.

On donne la loi de probabilité du couple de variables aléatoires *X* et *Y*.

Y / X	10	20	Totaux
10	0,08	0,12	
20	0,20	0,30	
30	0,12	0,18	
Totaux			

1° Compléter le tableau.

2° Pour tout $i \in \{1, 2, 3\}$ et $j \in \{1, 2\}$ comparer $P(X = x_i \text{ et } Y = y_j)$ $P(X = x_i) \times P(Y = y_j)$

3° Conclusion :

B. Opérations sur les variables aléatoires.

B.1. Somme de deux variables aléatoires.

Exemple 4.

Reprenons l'exemple1.

On s'intéresse au nombre de packs d'eau achetés par le client
On obtient une nouvelle variable aléatoire *S* égale à la somme des variables *X* et *Y*.

• **Définissons cette variable aléatoire *S***.

Dans le tableau insérons dans chaque case la somme :

Y / *X*	1	2	3
1	s = 2 0,1	s = 3 0,2	s = 4 0,2
2	s = 3 0,1	s = 4 0,3	s = 5 0,1

Les valeurs de *S* sont 2, 3, 4 et 5.

On associe à chacune de ces valeurs la somme des probabilités qui lui correspondent.

On défini ainsi la loi de probabilité de S présentée dans un tableau

Valeurs de s_i	2	3	4	5
$P(S = s_i)$				

B1.1. Définition. Soit X et Y deux variables aléatoires.

- La somme $X + Y$ est une variable aléatoire S : $S = X + Y$.
- La loi de probabilité de S est obtenue en associant à chaque valeur s de S, la somme des probabilités correspondants à tous les couples dont la somme des termes est égale à s.

B.1.2. Espérance mathématique de la somme de deux variables aléatoires.

Propriété. Soit X et Y deux variables aléatoires.

L'espérance mathématique de la somme $X + Y$ est égale à la somme des espérances mathématiques de X et de Y.

$$E(X + Y) = E(X) + E(Y)$$

Exemple 5.

Reprenons l'exemple1 et calculons $E(X)$, $E(Y)$ et $E(X + Y)$.

$E(X) =$

$E(Y) =$

$E(X + Y) =$

Conclusion :

Exercice 6. Reprenons l'exercice 3 ;

X \ Y	10	20
10	s = 0,08	s = 0,12
20	s = 0,20	s = 0,30
30	s = 0,12	s = 0,18

1° Définir la loi de probabilité de $S = X + Y$.

2° Calculer $E(X)$, $E(Y)$ et $E(X + Y)$, et vérifier l'égalité $E(X + Y) = E(X) + E(Y)$

1° On précise d'abord la loi de probabilité de la somme $S = X + Y$ par le tableau suivant :

Valeurs de s_i	20	30	40	50
$P(S = s_i)$				

2° $E(X) =$

$E(Y) =$

$E(X + Y) =$

Conclusion :

B.1.3. Variance de la somme de deux variables aléatoires.

Propriété. Si X et Y sont deux variables aléatoires **indépendantes**,
alors la variance de la somme $X + Y$ est égale à la somme des variances de X et de Y.

$$V(X + Y) = V(X) + V(Y)$$

{La réciproque est fausse !}

Exercice 7.

1° Reprenons l'exemple1 et calculons $V(X)$, $V(Y)$ et $V(X + Y)$.

2° L'égalité $V(X + Y) = V(X) + V(Y)$ est- elle vérifiée. Que peut-on conclure ?

1°

$V(X) = E(X^2) - [E(X)]^2 = \Sigma\, p_i x_i^{\,2} - [E(X)]^2 =$

$V(Y) = E(Y^2) - [E(Y)]^2 = \Sigma\, q_i y_i^{\,2} - [E(Y)]^2 =$

$V(X + Y) = \Sigma\, P(S = s_i) \times s_\mathrm{i}^{\,2} - [E(X + Y)]^2$

2°

Exercice 8.

Reprenons l'exemple 2.

Calculer $V(X)$, $V(Y)$ et $V(X + Y)$

$V(X) = E(X^2) - [E(X)]^2 = \Sigma\, p_i x_i^{\,2} - [E(X)]^2 =$

$V(Y) = E(Y^2) - [E(Y)]^2 = \Sigma\, q_i y_i^{\,2} - [E(Y)]^2 =$

$V(X + Y) = \Sigma\, P(S = s_i) \times s_\mathrm{i}^{\,2} - [E(X + Y)]^2 =$

B.2. Différence de deux variables aléatoires.

Définition. Soit X et Y deux variables aléatoires.

- La différence $X - Y$ est une variable aléatoire D.
- La loi de probabilité de D est obtenue en associant à chaque valeur d de D, la somme des probabilités correspondants à tous les couples dont la différence des termes est égale à d.

Propriété. Soit X et Y deux variables aléatoires.

- $E(X - Y) = E(X) - E(Y)$
- Si X et Y sont **indépendantes**, alors $V(X - Y) = V(X) + V(Y)$

Exercice 9. La loi de probabilité du couple (X, Y) est donnée par le tableau

X \ Y	– 2	– 1	0	Totaux
0	0,12	0,24	0,24	
1	0,08	0,16	0,16	
Totaux				

1° Déterminer la loi de probabilité de $X - Y$.

2° Calculer les nombres $E(X)$, $E(Y)$ et $E(X - Y)$ et les comparer.

3° Calculer les nombres $V(X)$, $V(Y)$ et $V(X - Y)$ et les comparer. Justifier Lerésultat.

C. Lois limites.

C.1.Théorèmes admis..

Théorème de Bernoulli (Loi faible des grands nombres).

Soit X une variable aléatoire et n variables $X_1, X_2, \ldots, X_n$, de même loi de probabilité que celle de X.

Si on pose : $S_n = X_1 + X_2 + \ldots + X_n$ et $Y_n = \dfrac{S_n}{n}$.

alors pour tout $\varepsilon > 0$, $\lim\limits_{n \to +\infty} P(|Y_n - E(X)| < \varepsilon) = 1$.

Signification de ce théorème.

• Soit une expérience aléatoire qui conduit à deux résultats :

- succès
- échec

On lui associe la variable aléatoire suivante

$X : \Omega \to [0 ; 1]$
$\omega \to 1$ {succès}
$\omega \to 0$ {échec}

$E(X) = p$ nombre fixé en théorie

• On répète n fois cette même expérience.

Les n variables aléatoires $X_1, X_2, \ldots, X_n$, ont la même loi de probabilité.

$$E(X_1) = E(X_2) = \ldots = E(X_n) = E(X) = p.$$

• Pour connaître le nombre de succès, on étudie la variable aléatoire : Y_n : « Fréquence des succès »

$$Y_n = \frac{\text{Nombre de succès}}{\text{Nombre d'éxpériences aléatoires}}$$

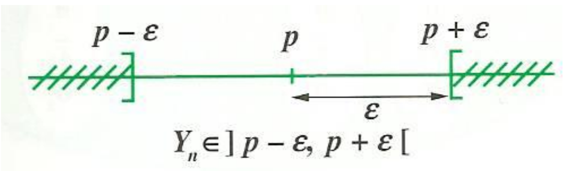
$$Y_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

$$E(Y_n) = \frac{E(X_1) + E(X_2) + \ldots + E(X_n)}{n} = \frac{np}{p} = p$$

• Est-on éloigné du résultat théorique ?

Pour le savoir, on étudie la différence $Y_n - p$

Soit ε un réel positif tel que $|Y_n - p| < \varepsilon$



Le théorème affirme que :

$$\lim_{n \to +\infty} P(|Y_n - E(X)| < \varepsilon) = 1$$

• On lance un dé. Si on obtient 6, c'est gagné et on marque 1 point, sinon c'est perdu, on marque 0 points.

$X(6) = 1$
$X(1) = X(2) = X(3) = X(4) = X(5) = 0.$

$$E(X) = \frac{1}{6}$$

• On lance un dé 30 fois

$$Y_n = \frac{\text{Nombre de 6}}{\text{Nombre de lancers}}$$

Il y a peu de chance que l'on trouve le résultat théorique $\frac{1}{6}$.

• On lance le dé 1000 fois :

$$Y_{1000} = \frac{\text{Nombre de 6}}{1000}$$

Le théorème dit que plus n est grand, plus Y_n se rapproche de la valeur théorique p. Intuitivement, ce résultat semble simple.

Théorème de la limite centrée.

Si $X_1, X_2, \dots, X_n$, sont des variables aléatoires indépendantes de même loi de probabilité, de même espérance mathématique m et de même variance σ^2,
alors lorsque n est « suffisamment grand » on admet :

- La loi de probabilité de la variable aléatoire $S_n = X_1 + X_2 + \dots + X_n$, qui mesure la somme des n variables aléatoires suit approximativement « la loi normale de moyenne nm et d'écart type $\sigma\sqrt{n}$, notée $\mathcal{N}(nm\,,\,\sigma\sqrt{n})$
- La loi de probabilité de la variable aléatoire $Y_n = \dfrac{S_n}{n}$ qui mesure la moyenne des n variables aléatoires suit approximativement « la loi normale » $\mathcal{N}\left(m\,,\,\dfrac{\sigma}{\sqrt{n}}\right)$.

C.2. Application : lois d'échantillonnage.

En statistique, il est en général impossible d'étudier un caractère sur toute une population de taille N élevée.

Avant d'aborder le problème de **l'estimation de valeurs caractéristiques inconnues** dans la population (ce problème sera traité dans le chapitre qui suit : **statistique inférentielle**), il est indispensable de commencer par l'étude de la **théorie de l'échantillonnage**.

Dans ce cas **les paramètres du caractère étudié dans la population sont connus** et on en déduit les propriétés sur l'ensemble des échantillons prélevés dans la population.

Nous n'envisagerons dans cette partie que des **échantillons aléatoires**, c'est-à-dire que tout élément de **l'échantillon est choisi au hasard**, et de plus, **les choix sont indépendants** car supposés avec remise.

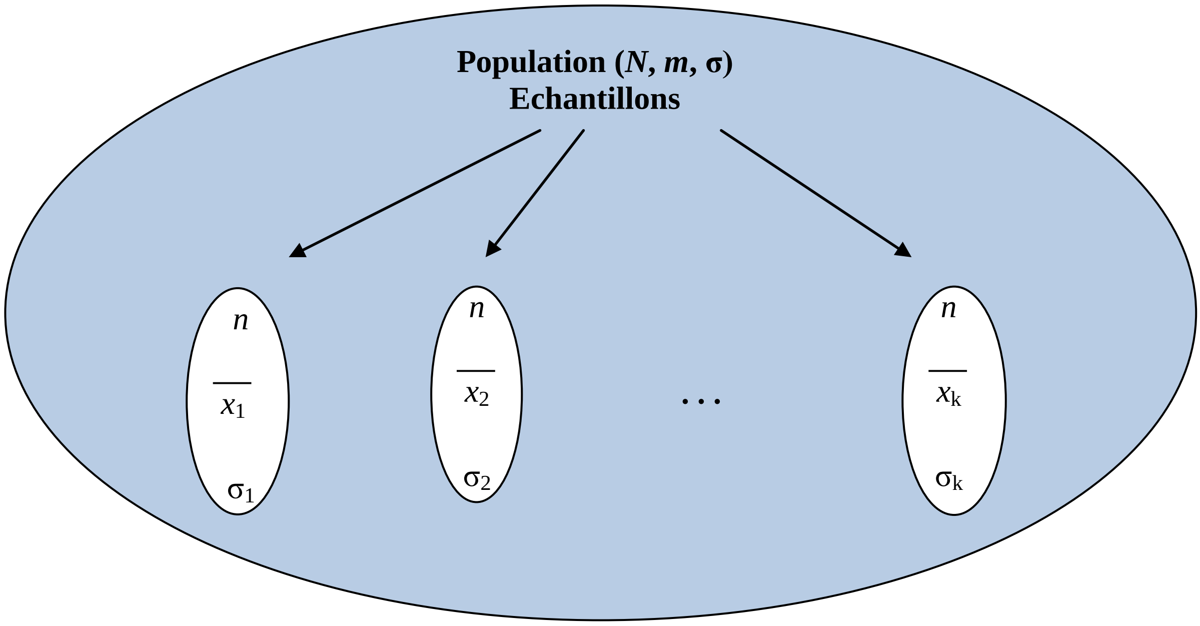
L'ensemble des échantillons de taille n est appelé échantillonnage de taille n.

C.2.1. Loi d'échantillonnage des moyennes.

Étant donné une population de taille N et X une variable aléatoire définissant le caractère étudié telle que :

$$E(X) = m \quad \text{et} \quad \sigma(X) = \sigma.$$

Pour prélever les échantillons de taille n, on a procédé à n épreuves indépendantes auxquelles correspondent n variables aléatoires $X_1, X_2, \dots, X_n$ de même loi que X.
La variable aléatoire $Y_n = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ associe à tout échantillon de taille n la moyenne de cette échantillon.



Quand n est suffisamment grand, la loi d'échantillonnage de la variable aléatoire $Y_n = \dfrac{X_1 + X_2 + \dots + X_n}{n}$ qui mesure la moyenne des échantillons de taille n suit d'après le théorème de la limite centrée, approximativement une loi normale.

On retrouve les valeurs caractéristiques de cette loi en calculant $E(Y_n)$ et $\sigma(Y_n)$.

$$E(Y_n) = E\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} E(X_1 + X_2 + \dots + X_n) = \frac{1}{n}\left[E(X_1) + E(X_2) + \dots + E(X_n)\right] = \frac{1}{n} \times nE(X)$$

$$\boldsymbol{E(Y_n) = E(X) = m}$$

$$V(Y_n) = V\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n^2} V(X_1 + X_2 + \dots + X_n) = \frac{1}{n^2}\left[V(X_1) + V(X_2) + \dots + V(X_n)\right]$$

$$\boldsymbol{V(Y_n) = \frac{1}{n^2} \times n \times V(X) = = \frac{1}{n} \times V(X) = \frac{1}{n}\sigma^2 \text{ d'où } \sigma(Yn) = \frac{\sigma}{\sqrt{n}}}$$

Théorème. La loi d'échantillonnage de taille n de la moyenne Y_n, quand n devient grand ($n \geq 30$), peut être approchée par la loi normale $\mathcal{N}\left(m, \dfrac{\sigma}{\sqrt{n}}\right)$

C.2.2. Loi d'échantillonnage des fréquences.

On étudie sur une population de taille N Soit un caractère à deux résultats :

On considère la variable aléatoire associée

$X : \Omega \to [0 ; 1]$
$\omega \to 1$ si ω possède le caractère {succès}
$\omega \to 0$ sinon

Dans cette situation : $E(X) = p$ est la probabilité de succès et $V(X) = p(1 - p)$

Pour constituer chacun des échantillons de taille n, on répète n fois la même épreuve de manière indépendante.

On obtient n variables aléatoires $X_1, X_2, \dots, X_n$ de même loi que X.

Considérons la variable aléatoire $S_n = X_1 + X_2 + \dots + X_n$ qui mesure le nombre succès

On a : $E(S_n) = nE(X) = np$ et $V(S_n) = nV(X) = np(1 - p)$

Quand n devient suffisamment grand, la loi d'échantillonnage de $f_n = \dfrac{S_n}{n}$ qui mesure la fréquence des succès dans les échantillons de taille n peut être approchée par une loi normale d'après le théorème de la limite centrée.

On retrouve les valeurs caractéristiques de cette loi en calculant :

$$E(f_n) = E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(S_n) = \frac{1}{n} \times np = p$$

$$V(f_n) = V\left(\frac{S_n}{n}\right) = \frac{1}{n^2} V(S_n) = \frac{1}{n^2} \times np(1 - p) = \frac{p(1-p)}{n} = \frac{pq}{n} \text{ avec } q = (1 - p) \text{ et donc } \sigma(f_n) = \sqrt{\frac{pq}{n}}$$

Théorème. La loi d'échantillonnage de taille n de la fréquence f_n, quand n devient grand ($n \geq 30$), peut être approchée par la loi normale $\mathcal{N}\left(p, \sqrt{\dfrac{pq}{n}}\right)$