Variables aléatoires réelles

1) Définition, Espérance, variance, écart-type :

Considérons une population de pièces cylindriques de longueur variable. Notons μ la valeur moyenne des longueurs des pièces dans la population, V, sa variance et σ son écart-type. Tirons n fois de suite avec remise une pièce au hasard dans la population. La fréquence avec laquelle nous observons une valeur x parmi les valeurs possibles de longueurs de pièces tend vers la proportion de pièces dans la population ayant cette longueur x quand n devient de plus en plus grand (de manière pratique à partir de $n = 10000$ ces fréquences observées diffèrent des proportions de moins de 1%).

La variable aléatoire X associée à la longueur d'une pièce tirée au hasard est alors définie par la donnée des valeurs de longueurs pouvant être prises et de leur fréquence d'observation, ce qui se résume par un tableau définissant la **loi de probabilité de X** (il s'agit ici d'une variable aléatoire dite discrète)

$\boldsymbol{k}$	$\boldsymbol{x_1}$	$\boldsymbol{x_2}$	…	$\boldsymbol{x_m}$
$\boldsymbol{P(X = k)}$	$\boldsymbol{p_1}$	$\boldsymbol{p_2}$	…	$\boldsymbol{p_m}$

L'espérance de cette variable correspond à la valeur vers laquelle tend la moyenne des longueurs observées sur n tirages successifs avec remise quand n tend vers l'infini. Elle se définit donc mathématiquement par :

$$\boxed{\boldsymbol{E(X) = x_1\, p_1 + x_2\, p_2 + \cdots + x_m\, p_m}}$$

La variance de cette variable correspond à la valeur vers laquelle tend la variance des longueurs observées sur n tirages successifs avec remise quand n tend vers l'infini. Elle se définit donc mathématiquement par :

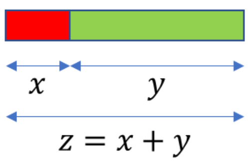
$$\boxed{\boldsymbol{V(X) = {x_1}^2\, p_1 + {x_2}^2\, p_2 + \cdots + {x_m}^2\, p_m - \big(E(X)\big)^2 = E\left(X^2\right) - \big(E(X)\big)^2}}$$

Et l'écart-type est la racine carrée de la variance :

$$\boxed{\boldsymbol{\sigma(X) = \sqrt{V(X)}}}$$

2) Somme de variables aléatoires :

On considère une population de pièces cylindriques formées de deux parties collées l'une à l'autre, une rouge, une verte. On note x la longueur de la partie rouge, y la longueur de la partie verte et z la longueur de l'ensemble.



On définit alors trois variables aléatoires X, Y, Z qui prennent respectivement les valeurs x, y, z d'une pièce tirée au hasard dans la population.

On suppose ces trois variables discrètes, pouvant prendre chacune un nombre fini de valeurs (même si celui-ci peut être très grand).

L'espérance de X est la valeur moyenne des valeurs x des pièces sur l'ensemble de la population, l'espérance de Y celle des valeurs y et l'espérance de Z, celle des valeurs z.

On a alors la propriété :

$$\boxed{\boldsymbol{E(Z) = E(X+Y) = E(X) + E(Y)}}$$

Preuve :

Afin de simplifier la rédaction de la preuve, nous allons supposer que X ne peut prendre que 3 valeurs x_1, x_2, x_3 et Y deux valeurs y_1, y_2.

Pour déterminer la loi de probabilité de Z, il nous faut connaitre la loi conjointe du couple variable (X, Y), c'est-à-dire, pour tous les couples (x_i, y_j) possibles, la probabilité $P(X = x_i \cap Y = y_j)$ autrement dit la proportion de pièces dans la population ayant une partie rouge de longueur x_i et une partie verte de longueur y_j, grandeur que nous noterons p_{ij}. On résume cela sous forme d'un tableau :

	y_1	y_2	Total
x_1	p_{11}	p_{12}	$P(X = x_1)$
x_2	p_{21}	p_{22}	$P(X = x_2)$
x_3	p_{31}	p_{32}	$P(X = x_3)$
Total	$P(Y = y_1)$	$P(Y = y_2)$	1

L'espérance de Z s'en déduit :

$$\begin{aligned} E(Z) &= (x_1 + y_1)\, p_{11} + (x_1 + y_2)\, p_{12} \\ &+ (x_2 + y_1)\, p_{21} + (x_2 + y_2)\, p_{22} \\ &+ (x_3 + y_1)\, p_{31} + (x_3 + y_2)\, p_{32} \end{aligned}$$

En développant et en refactorisant, on obtient :

$$E(Z) = x_1\,(p_{11} + p_{12}) + x_2\,(p_{21} + p_{22}) + x_3\,(p_{31} + p_{32})$$
$$+ y_1\,(p_{11} + p_{21} + p_{31}) + y_2\,(p_{12} + p_{22} + p_{32})$$

Soit :

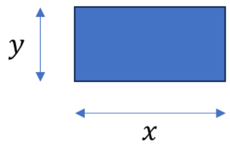
$$E(Z) = x_1\,P(X = x_1) + x_2\,P(X = x_2) + x_3\,P(X = x_3)$$
$$+ y_1\,P(Y = y_1) + y_2\,P(Y = y_2)$$

Finalement :

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

3) <u>Produit de variables aléatoires</u> :

On considère une population de pièces rectangulaires. On note x la longueur de la pièce en cm, y sa largeur en cm et z son aire en cm^2. On a donc : $z = x\,y$



On définit alors trois variables aléatoires X, Y, Z qui prennent respectivement les valeurs x, y, z d'une pièce tirée au hasard dans la population.

On suppose ces trois variables discrètes, c'est-à-dire pouvant prendre chacune un nombre fini de valeurs (même si celui-ci peut être très grand).

On dit que les variables aléatoires X et Y sont indépendantes si pour tout couple de valeurs possibles (x_i, y_j), on a $P(X = x_i \cap Y = y_j) = P(X = x_i) \times P(Y = y_j)$ ce qui revient à dire que les évènements $(X = x_i)$ et $(Y = y_j)$ sont indépendants.

On a alors, **si les variables $\boldsymbol{X}$ et $\boldsymbol{Y}$ sont indépendantes**, la propriété :

$$\boxed{\boldsymbol{E(Z) = E(X\,Y) = E(X) \times E(Y)}}$$

<u>Preuve</u> :

Nous reprenons les hypothèses faites dans la preuve précédente.

L'espérance de Z s'écrit alors :

$$\begin{aligned} E(Z) &= (x_1\,y_1)\,p_{11} + (x_1\,y_2)\,p_{12} \\ &+ (x_2\,y_1)\,p_{21} + (x_2\,y_2)\,p_{22} \\ &+ (x_3\,y_1)\,p_{31} + (x_3\,y_2)\,p_{32} \end{aligned}$$

En appliquant la propriété liée à l'indépendance de X et Y :

$$\begin{aligned} E(Z) &= (x_1\,P(X = x_1))\,(y_1\,P(Y = y_1)) + (x_1\,P(X = x_1))\,(y_2\,P(Y = y_2)) \\ &+ (x_2\,P(X = x_2))\,(y_1\,P(Y = y_1)) + (x_2\,P(X = x_2))\,(y_2\,P(Y = y_2)) \\ &+ (x_3\,P(X = x_3))\,(y_1\,P(Y = y_1)) + (x_3\,P(X = x_3))\,(y_2\,P(Y = y_2)) \end{aligned}$$

Et en constatant qu'il s'agit d'un développement :

$$E(Z) = \big(x_1\,P(X = x_1) + x_2\,P(X = x_2) + x_3\,P(X = x_3)\big) \times \big(y_1\,P(Y = y_1) + y_2\,P(Y = y_2)\big)$$

Finalement :

$$E(Z) = E(X) \times E(Y)$$

4) Variance d'une somme de variables aléatoires indépendantes

La variance d'une somme de deux variables aléatoires n'est pas nécessairement égale à la somme des variances de ces deux variables. Mais si X et Y sont deux variables indépendantes alors :

$$\boldsymbol{V(X+Y)=V(X)+V(Y)}$$

Cette propriété se généralise à la somme de n variables $X_1, X_2, \dots, X_n$ **indépendantes deux à deux** :

$$\boldsymbol{V(X_1+X_2+\cdots+X_n)=V(X_1)+V(X_2)+\cdots+V(X_n)}$$

Preuve :

$$V(X+Y)=E((X+Y)^2)-\big(E(X+Y)\big)^2=E(X^2+2\,X\,Y+Y^2)-\big(E(X)+E(Y)\big)^2$$

$$=E(X^2)+2\,E(X\,Y)+E(Y^2)-\Big(\big(E(X)\big)^2+2\,E(X)\,E(Y)+\big(E(Y)\big)^2\Big)=$$

$$=E(X^2)-\big(E(X)\big)^2+E(Y^2)-\big(E(Y)\big)^2+2\,\big(E(X\,Y)-E(X)\,E(Y)\big)$$

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

5) Linéarité de l'espérance, variance d'une fonction affine d'une variable aléatoire :

Reprenons l'exemple d'une pièce formée de deux parties cylindriques, la première de couleur rouge pouvant avoir une longueur variable x mesurée en cm et la seconde de couleur verte, une longueur fixe b mesurée en mm. Si on note z la longueur totale de la pièce en mm, alors on a : $z=a\,x+b$ avec $a=10$.

En faisant intervenir les variables aléatoires X, Z associées respectivement aux valeurs x, z d'une pièce tirée au hasard dans la population , on a :

$$\boldsymbol{E(Z)=E(a\,X+b)=a\,E(X)+b}$$

$$\boldsymbol{V(Z)=V(a\,X+b)=a^2\,V(X)}$$

Preuve :

$$E(Z)=(a\,x_1+b)\,p_1+(a\,x_2+b)\,p_2+\cdots+(a\,x_m+b)\,p_m$$

$$=a\,(x_1\,p_1+x_2\,p_2+\cdots+x_m\,p_m)+b\,(\,p_1+\ p_2+\cdots+\ p_m)$$

$$=a\,E(X)+b$$

$$V(Z)=E((a\,X+b)^2)-\big(E(a\,X+b)\big)^2$$

$$=E(a^2\,X^2+2\,a\,b\,X+b^2)-(a\,E(X)+b)^2$$

$$=a^2\,E(X^2)+2\,a\,b\,E(X)+b^2-\Big(a^2\,\big(E(X)\big)^2+2\,a\,b\,E(X)+b^2\Big)$$

$$=a^2\,\Big(E(X^2)-\ \big(E(X)\big)^2\Big)=a^2\,V(X)$$

6) Variable de Bernoulli

Considérons une population où il y a une proportion p de boules de couleur rouge et une proportion $1-p$ de boules de couleur verte. Définissons alors une variable aléatoire X qui, dans l'expérience d'un tirage aléatoire dans la population, prend la valeur 1 si la boule tirée est rouge et 0 sinon.

Une telle variable est qualifiée **de variable de Bernoulli**. Sa loi de probabilité est donnée par le tableau suivant :

k	0	1
$P(X=k)$	$1-p$	p

L'espérance d'une telle variable est alors :

$$\boldsymbol{E(X) = 0 \times (1-p) + 1 \times p = p}$$

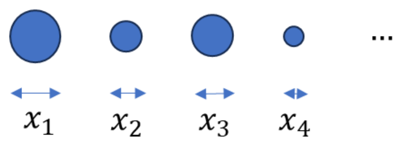
Sa variance :

$$\boldsymbol{V(X) = 0^2 \times (1-p) + 1^2 \times p - p^2 = p - p^2 = p\,(1-p)}$$

5) Variable moyenne empirique

Considérons une population de perles de tailles différentes et désignons par X la variable aléatoire qui prend, lors d'un tirage aléatoire unique la valeur x du diamètre de la perle tirée. Notons μ la valeur moyenne des diamètres des perles dans la population. On a alors : $E(X) = \mu$.

Considérons alors l'expérience aléatoire qui consiste à tirer n fois avec remise une perle dans la population et définissons n variables aléatoire $X_1, X_2, \dots, X_n$ associées à l'échantillon $(x_1, x_2, \dots, x_n)$ des n diamètres obtenus. Ces variables ont alors la même loi que celle de la variable X.



On définit alors la variable **moyenne empirique**, notée $\boldsymbol{\overline{X}_n}$ comme étant la variable qui prend pour valeur la moyenne des diamètres de l'échantillon, c'est-à-dire :

$$\boldsymbol{\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$

L'espérance de cette variable est :

$$\boldsymbol{E(\overline{X}_n) = \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \frac{\mu + \mu + \cdots + \mu}{n} = \frac{n \times \mu}{n} = \mu}$$

Cela s'interprète par le fait qu'en réalisant un nombre toujours plus grand d'échantillon, la valeur moyenne des valeurs moyennes de ces échantillons tendra vers la valeur moyenne dans la population.

Les variables X_i étant indépendantes deux à deux, la variance de cette variable est :

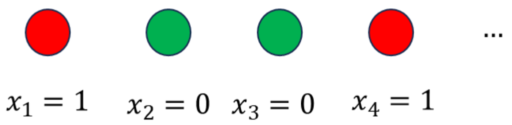
$$\boldsymbol{V(\overline{X}_n) = \frac{V(X_1) + V(X_2) + \cdots + V(X_n)}{n^2} = \frac{\sigma + \sigma + \cdots + \sigma}{n^2} = \frac{n \times \sigma}{n^2} = \frac{\sigma}{n}}$$

A noter que cette variance tend vers 0 quand n tend vers l'infini, propriété qui sera exploité pour estimer une grandeur relative à la population comme une valeur moyenne.

6) Variable fréquence empirique

Considérons une population de perles de couleur rouge ou verte et désignons par X la variable aléatoire de Bernoulli qui prend, lors d'un tirage aléatoire unique la valeur 1 si la perle tirée est rouge et 0 sinon. Notons p la proportion de perles rouges dans la population. On a alors : $E(X) = p$.

Considérons alors l'expérience aléatoire qui consiste à tirer n fois avec remise une perle dans la population et définissons n variables aléatoires $X_1, X_2, \dots, X_n$ de Bernoulli, où la variable X_i prend la valeur 1 si la perle tirée au $i-$ème tirage est rouge, 0 sinon . Ces variables ont alors la même loi que celle de la variable X.



On définit alors la variable **fréquence empirique**, notée $\boldsymbol{F_n}$ comme étant la variable qui prend pour valeur la proportion de perles rouges dans l'échantillon, c'est-à-dire :

$$\boldsymbol{F_n = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$

L'espérance de cette variable est :

$$\boldsymbol{E(F_n) = \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \frac{p + p + \cdots + p}{n} = \frac{n \times p}{n} = p}$$

Cela s'interprète par le fait qu'en réalisant un nombre toujours plus grand d'échantillon, la valeur moyenne des proportion de boules rouges dans ces échantillons tendra vers la proportion de boules rouges dans la population.

La variance de cette variable est :

$$\boldsymbol{V(F_n) = \frac{V(X_1) + V(X_2) + \cdots + V(X_n)}{n^2} = \frac{p\,(1-p) + p(1-p) + \cdots + p(1-p)}{n^2} = \frac{n \times p\,(1-p)}{n^2} = \frac{p\,(1-p)}{n}}$$

A noter que cette variance tend vers 0 quand n tend vers l'infini, propriété qui sera exploitée pour estimer une grandeur relative à la population comme une proportion d'individus de caractère donné.