3- LES TIRAGES PROBABILISTES D'ECHANTILLONS

Dans de nombreuses applications pratiques du calcul des probabilités, on retrouve un ou plusieurs des schémas de tirages probabilistes d'échantillons que nous allons exposer.
Le cadre général peut être formalisé ainsi :

Soit une population E de N éléments numérotés de 1 à N (par exemple: N boules d'une urne, N individus d'une population statistique,..) et l'expérience aléatoire consiste à choisir au hasard n éléments dans cette population. On dit qu'on tire un n-échantillon aléatoire (ou un échantillon aléatoire de taille n). La loi de probabilité qu'on peut associer aux différents n-échantillons possibles dépend du mode de tirage.

Définition :

On dit que le tirage de l'échantillon est **avec remise** (ou **bernoullien** ou **non exhaustif**) lorsque les n individus sont tirés un à un et remis dans la population avant tirage de l'individu suivant.

Le tirage de l'échantillon est **sans remise** (ou **exhaustif** lorsque les n individus sont tirés:
- **un à un** , mais ne sont pas remis dans la population avant tirage de l'individu suivant;
- ou encore que les n sont tirés **simultanément** (ce qui est équivalent lorsqu'on ne tient pas compte de l'ordre)

Pour donner un exemple type ; on considère une urne contenant N boules et on y tire n boules. Dans de nombreux cas les N individus sont répartis en k catégories de tailles respectives N_1, N_2 ...., N_k avec

$$\sum_{i=1}^{k} N_i = N$$

On notera p_i la probabilité qu'un individu de la catégorie i soit tiré. C'est la proportion des individus de la catégorie i dans la population :

$$p_i = \frac{N_i}{N} \quad i = 1,2,...,k \ \text{ et } \ \sum_{i=1}^{k} p_i = 1$$

La probabilité qu'un n-échantillon soit formé de $n_1, n_2, ..., n_k$ individus de chaque catégorie respective 1,2,...,k dépend du mode de tirage. On a :

$$\sum_{i=1}^{k} n_i = n$$

3-1- LE TIRAGE AVEC REMISE (BERNOULLIEN)

L'espace fondamental est alors $\Omega = E^n$ et card $\Omega = N^n$. Il contient donc N^n échantillons possibles qui sont les N^n événements élémentaires de . (N^n arrangements avec répétitions de n éléments parmi N).
La probabilité d'obtenir $n_1, n_2, ..., n_k$ individus de chaque catégorie sera notée :

- Dans les cas ordonnés : $P(n_1^* n_2^* ... n_k^*)$ (avec des étoiles)

Elle se lira : probabilité dans un ordre spécifié.

- Dans les cas où l'ordre n'a pas d'importance, on notera

$P(n_1, n_2, ... n_k)$(avec des virgules) et on lira probabilité dans un ordre quelconque.

3-1-1 On tient compte de l'ordre d'apparition des individus.

On cherche p = $P(n_1^* n_2^* ... n_k^*)$ (Càd la probabilité d'un échantillon composé de cette structure mais en plus dans un certain ordre spécifié.)

$$P_a = P(n_1 {}^* n_2 {}^* ... {}^* n_k) = \frac{N_1^{n_1} N_2^{n_2} ... N_k^{n_k}}{N^n} = p_1^{n_1} p_2^{n_2} ... p_k^{n_k}$$

Démonstration :

$$Card\Omega = N^n$$

Les événements élémentaires sont ici équiprobables. Rappelons que la probabilité P(A) d'un événement A se calcule en cas d'équiprobabilité à l'aide de la formule :

$$\boxed{P(A) = \frac{Card\ A}{Card\ \Omega}}$$

cardA $= N_1^{n_1} N_2^{n_2} ... N_k^{n_k}$ car il y a $N_1^{n_1}$ façons de choisir les n_1 parmi N_1, à chacune d'elles, on peut associer $N_2^{n_2}$ façons de choisir n_2 parmi N_2, ..., à chacune des $N_1^{n_1} N_2^{n_2} ... N_{k-1}^{n_{k-1}}$ façons on peut associer $N_k^{n_k}$ façons de choisir n_k parmi les N_k

or $n = n_1 + n_2 + ... + n_k$ d' où $N^n = N^{(n_1 + n_2 + ... + n_k)}$

$$\boxed{\text{et} \quad P(n_1 {}^* n_2 {}^* ... {}^* n_k) = \left(\frac{N_1}{N}\right)^{n_1} \left(\frac{N_2}{N}\right)^{n_2} ... \left(\frac{N_k}{N}\right)^{n_k} = p_1^{n_1} p_2^{n_2} ... p_k^{n_k}}$$

Exemple :

Soit une urne avec N=5 boules dont 3 sont blanches et 2 sont noires. On tire : n=3 boules et soient $n_1=2$ et $n_2=1$:

$$P(BNB) = \frac{3^1 2^1 3^1}{5^3} = \left(\frac{3}{5}\right)^2 \left(\frac{2}{5}\right)^1 = 0{,}6^2 0{,}4^1 = 0{,}144$$

BNB : désigne obtenir une boule blanche au premier tirage, une boule noire au 2ème tirage et une boule blanche au 3ème tirage.

Remarquons que P(BBN)=P(NBB)=0,144 . La probabilité est la même pour: BNB ; BBN ou NNB. Quel que soit l'ordre qu'on retient la probabilité est la même.

3-1-2- On ne tient pas compte de l'ordre :

b1- On considère k catégories:

on cherche $P_b = P(n_1, n_2, ..., n_k)$ (on ne s'intéresse ici qu'à la structure de l'échantillon peu importe l'ordre d'obtention des individus) :

$$\boxed{P_b = P(n_1, n_2, ..., n_k) = \frac{n!}{n_1! n_2! ... n_k!} p_1^{n_1} p_2^{n_2} ... p_k^{n_k}}$$

Démonstration**:** Comme la probabilité d'obtenir un ordre spécifié est :

$$P_a = P(n_1 * n_2 * ... * n_k) = \frac{N_1^{n_1} N_2^{n_2} ... N_k^{n_k}}{N^n} = p_1^{n_1} p_2^{n_2} ... p_k^{n_k},$$

et comme il y a autant d'ordres possibles -chacun d'eux correspondant à l'échantillon avec la même composition $(n_1, n_2, ..., n_k)$- que de permutations avec répétition de n éléments dont n_1 sont identiques entre eux, n_2 identiques entre eux, ..., n_k identiques entre eux, c'est-à-dire :

$$\frac{n!}{n_1! n_2! ... n_k!},$$

il suffit de multiplier P_a par ce nombre.

Définition : la loi de probabilité qui donne la répartition $(n_1, n_2, \ldots, n_k)$,quel que soit l'ordre, s'appelle **la loi multinomiale**

b2 -Cas particulier important : k=2

Exemple :

Soit N=5 boules d'une urne avec $n_1=3$ blanches , $n_2=2$ noires et on y tire,au hasard , n=3 boules (avec remise).L'urne est :

$$E = \{ B_1, B_2, B_3, N_1, N_2 \}.$$

Soit à calculer la probabilité d'avoir un échantillon composé de 2 boules blanches et une noire noté P(2B,1N)

$$P(2B,1N) = 3!/(2!\,.1!) \times (3^2 \cdot 2^1)/5^3 = 3 \times 0{,}144 = 0{,}432$$

Remarquons qu'on a ici un résultat attendu : on a plus de chances d'avoir un échantillon dans un ordre quelconque qu'un échantillon avec un ordre spécifié (0,432 > 0,144).

Généralisons cet exemple à de nombreuses situations se retrouvant dans l'application pratique du calcul des probabilités et qui concernent une population avec 2 catégories d'individus : boules noires/blanches, vote: oui/non, sexe : masculin/féminin, pièces: valables/défectueuses.... Dans ces cas, la probabilité d'avoir un échantillon, avec la répartition (n_1, n_2) et quelque soit l'ordre peut s'exprimer d'une autre manière qu'on retrouvera fréquemment par la suite :

$$P(n_1, n_2) = \frac{n!}{n_1!\,n_2!}\, p_1^{n_1} p_2^{n_2}$$ (en vertu du résultat général précédent)

or, $\frac{n!}{n_1!\,n_2!} = C_n^{n_1}$ car $n = n_1 + n_2$ d'où $n_2 = n - n_1$ et:

$$\frac{N_1^{n_1} N_2^{n_2}}{N^n} = p_1^{n_1} p_2^{n_2}$$ D'habitude, on note

: $p_1 = p$ et $p_2 = q$ et comme $p_1 + p_2 = p + q = 1$, on a:

$$q = 1 - p$$

Finalement $P(n_1, n_2)$ peut être écrite autrement :

$$\boxed{P(n_1, n_2) = C_n^{n_1} p^{n_1} q^{n-n_1} = C_n^{n_1} p^{n_1} (1-p)^{n-n_1}}$$

Cette fonction est la loi de probabilité de la composition d'un n-échantillon (sans ordre spécifié) lorsqu'il y a 2 catégories dans la population. Elle s'appelle: **la loi binomiale.**

Remarque :

- Cette loi de probabilité ne dépend pas de la taille de la population N mais seulement de la taille de l'échantillon n et de la proportion p des individus d'une certaine catégorie. n et p sont les deux paramètres de la loi.

Exemple :

1- On considère une urne avec une proportion p de boules blanches et q de boules noires et on tire n boules avec remise. La probabilité d'avoir "i " boules blanches est alors :

$$C_n^i p^i q^{n-i} \quad \text{pour :} \qquad 0 \leq i \leq n.$$

on a pour ; n=3 et p=q = 1/2 ; l'espace fondamental suivant :

Ω'	0 blanche	1 blanche	2 blanches	3 blanches	total
$C_n^i p^i q^{n-i}$	$C_3^0 (\frac{1}{2})^3$	$C_3^1 (\frac{1}{2})^3$	$C_3^2 (\frac{1}{2})^3$	$C_3^3 (\frac{1}{2})^3$	1
probabilités	1/8	3/8	3/8	1/8	1

' compte 4 événements élémentaires non équiprobables. C'est le nombre de combinaisons avec répétitions de 3 parmi 2 éléments :

$K_2^3 = 4$

Remarque :

on est passé d'un espace fondamental où card = 8 et où les 8 événements sont équiprobables,(dans le cas où on tient compte de l'ordre), à un espace ' où les événements ne sont plus équiprobables. On s'est servi de pour probabiliser '.

2- On jette une pièce de monnaie n fois on s'intéresse au nombre de piles obtenus. Cet exemple est analogue au tirage avec remise de n boules d'une urne contenant des boules blanches et noires en proportions égales (p=q=1/2) comme dans l'exemple 1.

Ainsi si A : "avoir 2 piles en jetant la pièce 3 fois"

$$P(A)=C_3^2\left(\frac{1}{2}\right)^3=3/8$$

3- On jette un dé n fois et on s'intéresse au nombre de fois où l'on a obtenu un "as". Cet exemple est analogue au tirage de boules d'une urne où la proportion des boules blanches est 1/6 et où, on s'intéresse au nombre de boules blanches obtenues, la probabilité est donnée par la loi binomiale.

3-2- LE TIRAGE EST SANS REMISE (EXHAUSTIF)

3-2-1- La probabilité est dans un ordre spécifié

Avec les notations précédentes, la probabilité d'avoir un n-échantillon avec la répartition n_1 n_2, ... n_k (où n_i= nombre d'individus de la catégorie i(i=1, ...k) et ceci, dans un ordre spécifié est :

$$\boxed{P_a = P(n_1 * n_2 * \ldots * n_k) = \frac{A_{N_1}^{n_1} A_{N_2}^{n_2} \ldots A_{N_k}^{n_k}}{A_N^n}}$$

Démonstration : il y a autant d'échantillons possibles que d'arrangements sans répétitions de n parmi N. Donc card $=A_N^n$ événements élémentaires équiprobables. Parmi eux, il y a $A_{N_1}^{n_1} A_{N_2}^{n_2} \ldots A_{N_k}^{n_k}$ échantillons favorables à la réalisation de la répartition et de l'ordre spécifié.

Remarque :

Cette probabilité est la même quelque soit l'ordre spécifié, c'est à dire, pour chaque ordre.

3.2.2. Dans un ordre quelconque.

i/ On distinguera ici deux cas :

b1 : on tire les n individus un à un.

soit : $P_{b_1} = P(n_1, n_2, ..., n_k) = \dfrac{n!}{n_1!n_2!...n_k!} \dfrac{A_{N_1}^{n_1} A_{N_2}^{n_2} ... A_{N_k}^{n_k}}{A_N^n}$

En effet ; la probabilité pour un ordre spécifié est P_a et celle-ci étant la même pour chacun des $\dfrac{n!}{n_1!n_2!...n_k!}$ ordres. Donc il suffit de les multiplier : $P_{b_1} = \dfrac{n!}{n_1!n_2!...n_k!} . Pa$

b2: on tire les n individus simultanément (d'un seul coup)

$$\boxed{P_{b_2} = P(n_1, n_2, ..., n_k) = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2} ... C_{N_k}^{n_k}}{C_N^n}}$$

Il y a C_N^n façons d'avoir l'échantillon de taille n et card ' $= C_N^n$ événements élémentaires équiprobables. Parmi eux $C_{N_1}^{n_1} C_{N_2}^{n_2} ... C_{N_k}^{n_k}$ échantillons favorables à la réalisation de l'échantillon désiré (caractérisé par la composition mais pas par l'ordre).

Remarques :

1- $P_{b_1} = P_{b_2}$ car $\dfrac{A_{N_i}^{n_i}}{n_i!} = C_{N_i}^{n_i}$ i=1,2,....,k

Cela revient au même de dire : "qu'on tire les n individus un à un et de ne pas s'intéresser à l'ordre d'apparition des individus dans l'échantillon que de dire : "que l'échantillon a été tiré d'un seul coup"

2- Dans P_a ; card $= A_N^n$. il y a : A_N^n événements élémentaires équiprobables et ils s'agit du cas où on tient compte de l'ordre.

- dans P_{b1} ou P_{b2} card ' $= C_N^n$. Il y a C_N^n événements élémentaires équiprobables également et il s'agit du cas où l'ordre n'est pas important.

Par contre, dans le cas du tirage avec remise : card $= N^n$; il y a N^n événements élémentaires équiprobables dans le cas où on tient compte de l'ordre et card ' $= K_N^n$ événements élémentaires non équiprobables dans le cas où l'ordre n'est pas important.

La raison pour laquelle les C_N^n échantillons du tirage sans remise sont équiprobables et les K_N^n échantillons du tirage avec remise ne le sont pas est que ; chaque combinaison sans répétitions parmi C_N^n renferme n! arrangements sans répétitions. Par contre chacune des combinaisons avec répétitions parmi les K_N^n ne renferme pas le même nombre d'arrangements avec répétitions (cf: introduction de l'analyse combinatoire)

ii/ Cas particulier important $k = 2$

La loi de probabilité de la composition d'un n-échantillon (sans ordre spécifié) et où on considère 2 catégories dans la population est donnée par

$$P(n_1, n_2) = \frac{C_{N_1}^{n_1} C_{N_2}^{n_2}}{C_N^n}$$

Il s'agit de **la loi hypergéométrique**.

Remarque : pour que l'expérience soit possible il faut que

$n_1 \leq N_1$, $n_2 \leq N_2$ et $n \leq N$.

Exemple :

On tire 3 boules d'une urne contenant 5 boules dont 3 blanches et 2 noires. La probabilité d'avoir n_1 boules blanches est donnée par le tableau suivant pour : n_1=1,2,3.

'	1	2	3	Somme
$\frac{C_{N_1}^{n_1} C_{N_2}^{n_2}}{C_N^n}$	$\frac{C_3^1 C_3^2}{C_5^3}$	$\frac{C_3^2 C_3^1}{C_5^3}$	$\frac{C_3^3 C_3^0}{C_5^3}$	1
probabilités	3/10	6/10	1/10	1