Loi Binomiale

I. Loi de Bernoulli

1) Epreuve et Loi de Bernoulli

Une épreuve de Bernoulli est une expérience admettant exactement 2 issues l'une appelée succès S et l'autre échec $\bar{S}$ leur probabilités respectives seront notées p et (1-p).

Etant donné une épreuve de Bernoulli, considérons la variable aléatoire *X* prenant pour valeur 1 quand le succès est réalisé, ou prenant pour valeur 0 dans le cas de la réalisation de l'échec.
Une variable aléatoire ainsi définie s'appelle variable de Bernoulli, on dit aussi que *X* suit une loi de Bernoulli de paramètre *p*, où *p* est la probabilité du succès.

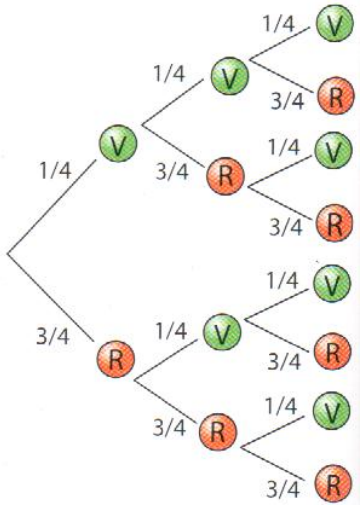
Exemple : on tire une boule dans une urne contenant 3 rouges et une verte
Si X est la variable aléatoire associant 1 quand la boule verte est tirée et 0 quand c'est la rouge, X suit une loi de Bernoulli de paramètre $\frac{1}{4}$.

2) Espérance, écart type d'une loi de Bernoulli

Considérons une variable aléatoire suivant une loi de Bernoulli de paramètre *p* .
Alors :$E(X) = 1P(X = 1) + 0P(X = 0) = 1p + 0(1 - p) = p$.
Et $V(X) = (1 - p)^2P(X = 1) + (0 - p)^2P(X = 0) = (1 - p)^2p + p^2(1 - p) = p(1 - p)[1 - p + p] = p(1 - p)$
L'espérance d'une loi de Bernoulli est égale à *p* et sa variance à *p(1-p)*.

3) Schéma de Bernoulli

Considérons une épreuve de Bernoulli où *p* désigne la probabilité du "succès" et *n* un entier naturel non nul.
On appelle schéma de Bernoulli à *n* épreuves de paramètre *p*, la répétition *n* fois de suite avec indépendance de la même épreuve de Bernoulli de paramètre *p*.



Exemple :
En revenant à l'exemple précédent : on tire trois boules avec remise (ce qui garantit l'indépendance). Cela constitue une schéma de Bernoulli de paramètres *n* = 3 et *p* = $\frac{1}{4}$. : voir schéma ci-contre (on remarquera que les événements contraires sont V et R , la forme S , $\bar{S}$ n'est pas obligatoire)

II. Loi binomiale

1) Définition

Considérons un schéma de Bernoulli de paramètres *n* et *p*.
Les *n* épreuves de Bernoulli peuvent se modéliser par un arbre qui aurait 2 branches initiales (succès-échec) puis chacune de ces branches donnant naissance à 2 nouvelles branches (succès-échec) etc.. A chaque épreuve, le nombre de branches est doublé.
Considérons une variable aléatoire *X* qui compte le nombre de réalisations du succès au cours des *n* épreuves. On dit que *X* suit une loi binomiale de paramètre *n* et *p*.
X peut donc prendre toutes les valeurs entières de 0 à *n*.
Prenons *k* une de ces valeurs entières.
Sur l'arbre il existe un certain nombre de branches (ou chemins) qui comporte *k* succès et *n-k* échecs. Ce nombre de chemins se note $\binom{n}{k}$. $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{1\times2\times3\times\ldots\times(n-1)\times n}{(1\times2\times\ldots\times(k-1)\times k)(1\times2\times\ldots\times(n-k-1)\times(n-k))}$.
On peut trouver aussi $\binom{n}{k}$ à l'aide du triangle de pascal : on regarde l'intersection de la ligne n et de la colonne k (attention dans le comptage des lignes et des colonnes : il y a une colonne 0 et une ligne 0)
Le coefficient binomial est donné par la calculatrice.
- Pour la Texas instrument **TI 82** : Dans le menu Maths, choisir l'option PRB et sélectionner **nCr**.
- Pour la Casio **Graph 25** : Dans le menu OPTION, choisir PROB, puis **nCr**.

		colonne					
		0	1	2	3	4	5
ligne	0	1					
	1	1	1				
	2	1	2	1			
	3	1	3	3	1		
	4	1	4	6	4	1	
	5	1	5	10	10	5	1

Chacun de ces chemins comporte *k* succès et *n-k* échecs. La probabilité qu'un de ces chemins se réalise est égale à $p^k(1-p)^{n-k}$. (C'est le principe du produit des probabilités sur les branches...)

Pour finir, puisqu'il y a $\binom{n}{k}$chemins ayant chacun la probabilité $p^k(1-p)^{n-k}$de se produire, on peut en déduire que $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

2) Espérance et écart type d'une loi binomiale
On admet les résultats suivants sur l'espérance et l'écart type d'une loi binomiale de paramètres *n* et *p* .
$E(X)=np$ et $V(X)=np(1-p)$

3) Exemple
Revenons au schéma de Bernoulli précédent où on tire une boule 3 fois de suite.
La variable aléatoire X qui compte le nombre « Vert » obtenus sur 3 tirages suit une loi binomiale de paramètres *n*=3 et *p*=$\frac{1}{4}$.
Par exemple: $P(X=2)=\binom{3}{2}\left(\frac{1}{4}\right)^k\left(1-\frac{1}{4}\right)^{3-2}=3$. On a aussi $(X)=np=3\times\frac{1}{4}=\frac{3}{4}$.

Reprenons cette loi avec *n* =4, l'espérance devient égale à $E(X)=4\frac{1}{4}=1$ ce qui est logique, sur quatre tirage on en aura en moyenne un qui sera vert.

III. Echantillonnage
L'année dernière on avait une population présentant une caractéristique dans une proportion p, et on prenait un échantillon de taille n de cette population et présentant la caractéristique avec une fréquence f.
Connaitre f mais pas p, nous permettait d'estimer ce dernier à l'aide de l'intervalle de confiance au niveau de confiance de 95% , on a $p\in\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$
Quand on connait p, si $f\in\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$ *(intervalle de fluctuation de la fréquence f au seuil de 95%) l'échantillon est dit représentatif de la population, dans le cas contraire on dira que l'échantillon est biaisé*

Cette année l'approche est différente, on conjecture que la population présente le caractère dans une proportion p. On va mettre notre conjecture à l'épreuve d'un échantillon représentatif. Plus f est éloigné de p, plus il y a des chances que notre conjecture soit fausse. On ne sera jamais sûr à 100%, donc il faut bien choisir un seuil à partir duquel on rejettera notre conjecture. Notre priorité est de ne pas rejeter à tort la conjecture.

Intervalle de fluctuation à 95%
Choisir un échantillon de taille n dans une population présente la caractéristique avec une proportion p, revient à faire n tirages successifs avec la probabilité p d'avoir la caractéristique à chaque tirage (le fait qu'il n'y ait pas de remise, n'empêche pas l'indépendance car la population est trop importante pour qu'un tirage influence le suivant), ainsi X le nombre de personnes parmi les n de l'échantillon ayant la caractéristique suit une loi Binomiale de paramètres n et p.
En lisant la ligne des probabilités cumulées croissantes je peux identifier :
a le plus petit entier tel que $P(X\leq a)>2{,}5\%$ et *b* le plus petit entier tel que $P(X\leq b)\geq 97{,}5\%$
ainsi $P(a\leq X\leq b)$ sera à peine supérieur à 95%
ainsi le nombre de personnes présentant la caractéristique aura 95 pourcent de chance d'être entre a et b.

Définition / règle
L'intervalle de fluctuation au seuil de 95% de la fréquence f est l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$
avec *a* le plus petit entier tel que $P(X\leq a)>2{,}5\%$ et *b* le plus petit entier tel que $P(X\leq b)\geq 97{,}5\%$
Si f n'est pas dans $\left[\frac{a}{n};\frac{b}{n}\right]$, on peut rejeter notre conjecture, il n'y a que 5% de chance que l'on se trompe en le faisant, par contre si f est dans l'intervalle rejeter la conjecture serait bien trop risqué, on l'acceptera donc.

Remarque :
L'approche de 1ère est plus lourde que celle de seconde car on doit chercher à la main a et b, mais on comprends mieux ce que l'on fait (vu qu'on est des pros de la loi binomiale) , l'approche de seconde s'appuie sur la loi Normale que vous verrez plus tard.