

Chapitre 6- Variables aléatoires et répétitions d'une expérience aléatoire

Table des matières

Chapitre 6 - Variables aléatoires et répétitions d'une expérience aléatoire	1
1. Les variables aléatoires (discrètes)	2
1.1. Fonction de distribution ou loi de probabilité	2
1.2. Espérance mathématique, notée $E(X)$	3
1.3. Variance, notée $V(X)$ et écart-type, noté $\sigma(X)$ (pour aller plus loin)	3
2. La loi de Bernoulli	5
3. Simulation et fluctuation d'échantillonnage	6
3.1. Création de la simulation	6
3.2. Distance à p : fréquent ou rare ?	8

Dans le domaine de la gestion, de nombreux phénomènes sont soumis à des aléas : qualité des produits, comportement des consommateurs, concurrence, environnement, etc. Bien souvent, certains **sont soumis à des lois statistiques**, et à ce titre, nécessitent la mise en œuvre d'un certain nombre d'outils.

1. Les variables aléatoires (discrètes)

Une variable aléatoire est une application qui permet d'associer à **chaque résultat d'une épreuve (éventualité) une probabilité**.

Exemple : on prélève au hasard une marchandise dans un lot et on relève son poids. Le poids relevé est aléatoire et s'exprime par un nombre : c'est donc une variable aléatoire.

Dans le cas d'une variable aléatoire discrète, il s'agit d'une variable dont l'ensemble des **valeurs possibles est dénombrable ou fini** (ex. nombre d'incidents d'impressions, nombre de gagnants à un jeu, nombre quotidien de commandes).

1.1. Fonction de distribution ou loi de probabilité

La fonction de distribution (ou Loi de probabilité) d'une variable aléatoire X est l'application, qui à **chaque valeur possible de x_i** , associe une probabilité « p_i [$P(x_i)$ ou $P(X=x_i)$] ».

La somme des probabilités associées à une **variable aléatoire X est égale à 1** (probabilité d'un événement certain).

$$\sum_{i=1}^n P(x_i) = 1$$

Application pratique :

Un attaché commercial a relevé le nombre de commandes obtenues sur 300 clients.

Nombre quotidien de commandes	0	1	2	3	4	Total
Nombre de clients	30	60	120	75	15	300

Détermination de la loi de probabilité du nombre quotidien de commandes

Soit la variable aléatoire étudiée X : « nombre quotidien de commandes ». Il s'agit bien d'un ensemble fini de valeurs.

Les probabilités associées à chaque valeur de x_i sont assimilables aux fréquences de la série statistique.

Ainsi, nous obtenons respectivement : $(0,10 = 30/300)$

x_i	0	1	2	3	4	
$P(X = x_i)$ ou (P_{x_i})	0,10	0,20	0,40	0,25	0,05	$\sum P(X) = 1$

1.2. Espérance mathématique, notée $E(X)$

Exprime la tendance centrale de la variable aléatoire X . Elle est la **moyenne arithmétique** des valeurs

$$E(X) = \sum_{i=1}^n P(x_i) * x_i$$

prises par X pondérées par les probabilités associées à chacune de ces valeurs.

Application pratique (suite) :

Nbre de commande (X)	0	1	2	3	4	Total
Nombre de clients	30	60	120	75	15	300
$P(X = x_i)$ ou (P_{x_i})	10%	20%	40%	25%	5%	100%
Moyenne du nbre de commande	0	0,2	0,8	0,75	0,2	1,95

Moyenne = $0*10\% + 1*20\% + \dots + 4*5\% = 1.95$ commande en moyenne.

1.3. Variance, notée $V(X)$ et écart-type, noté $\alpha(X)$ (pour aller plus loin)

La **variance** exprime la **dispersion** de la variable aléatoire X par rapport à la tendance centrale (moyenne).

La variance de X est la somme des écarts au carré entre les valeurs prises par X et son espérance mathématique, pondérés par des probabilités associées à chacune de ces valeurs.

$$V(X) = \sum [x_i - E(X)]^2 \cdot P(x_i)$$

Application pratique (suite) :

Nbre de commande (X)	P(X = x _i)	(X _i - E(x)) ²	(X _i - E(x)) ² * P(X)
0	10%	3,80	0,380
1	20%	0,90	0,181
2	40%	0,00	0,001
3	25%	1,10	0,276
4	5%	4,20	0,210

Moyenne du nbre de commande	1,95
Variance	1,05

L'écart-type est la racine carrée de la variance. Il est utilisé en gestion pour apprécier le risque. De façon plus générale, plus il est élevé, plus le risque est grand.

$$\sigma_X = \sqrt{\text{Var}(X)}.$$

Application pratique (suite) :

Variance	1,05
Écart type	1,02

2. La loi de Bernoulli

Beaucoup de situations de gestion se ramènent à une question dont la réponse est **oui ou non** : le client achète-t-il ou non ? La pièce est-elle défectueuse ou non ? Un salarié est-il absent ou non ? La loi de Bernoulli est le modèle mathématique qui correspond exactement à ce type de situation « à deux issues ».

La loi de Bernoulli est une expérience aléatoire qui ne comporte que deux issues possibles :

- Le **succès (S)**, avec une probabilité « p ».
- L'**Échec ($\bar{S}$)**, avec une probabilité « $q = 1 - p$ ».

Le Schéma de Bernoulli : On parle de schéma de Bernoulli lorsqu'on répète « n » fois la même épreuve de Bernoulli de manière **identique et indépendante**.

Application pratique :

Dans un atelier, on sait que 12 % des pièces produites sont défectueuses. On prélève une pièce au hasard et on définit $X=1$ si la pièce est défectueuse, $X=0$ sinon.

X suit une loi de Bernoulli de paramètre $p=0,12$. On a donc $E(X)=0,12$: en moyenne, sur un grand nombre de pièces prélevées une par une, la proportion de pièces défectueuses est de 12 %, ce qui est cohérent avec la donnée de départ.

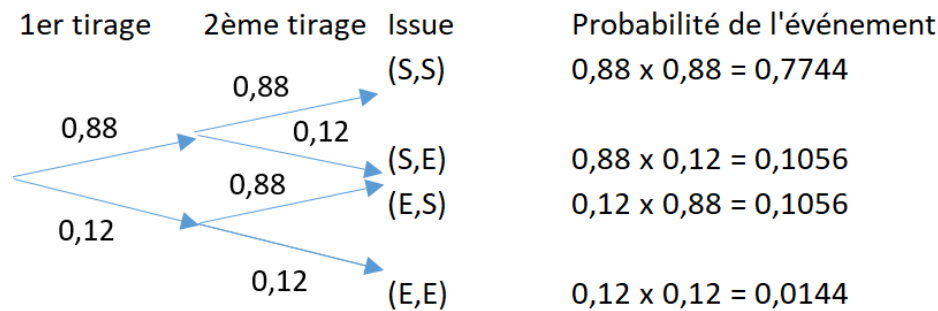
Il est possible de réaliser un arbre de probabilité pour schématiser le tout.

Exemple :

	Atelier	Probabilité
Pièce conforme	880	0,88
Pièce défectueuse	120	0,12
Total	1000	1

Succès ($X=0$) : La pièce est conforme.

Échec ($X=1$) : La pièce est défectueuse.



3. Simulation et fluctuation d'échantillonnage

3.1. Création de la simulation

On avait vu avant, que si 12 % des pièces produites sont défectueuses, on devrait trouver une moyenne de 12 % de pièces défectueuses à chaque contrôle.

Cependant, on prélève un échantillon de n pièces dans un lot où la vraie proportion de défauts est p (12%), on ne va presque jamais retomber exactement sur ces 12% (p) dans l'échantillon : **la fréquence** observée des défauts va varier légèrement d'un échantillon à l'autre, autour de p . C'est ce qu'on appelle la fluctuation d'échantillonnage. En simulant de nombreux échantillons sur ordinateur, on peut observer et mesurer cette fluctuation.

Concrètement, on simule N échantillons (idéalement plusieurs centaines) de taille n d'une loi de Bernoulli de paramètre p , puis on représente les fréquences observées des « 1 » à l'aide d'un histogramme ou d'un nuage de points.

Application pratique :

Pour cette dernière partie, on va travailler sur une probabilité de succès de 10% (10% des pièces sont défectueuses). Avec un échantillon (n) de 50 produits. Enfin pour la simulation on travaille sur 100 échantillons simulés.

Tirage i	Résultat	13	0	26	1
1	0	14	0	27	0
2	0	15	0	28	0
3	1	16	0	29	0
4	0	17	1	30	0
5	1	18	0	31	0
6	0	19	0	32	0
7	0	20	0	33	0
8	0	21	0	34	0
9	0	22	0	35	0
10	1	23	0	36	0
11	0	24	1	37	0
12	0	25	1	38	0

39	0
40	0
41	0
42	0

43	0
44	0
45	0
46	0

47	0
48	0
49	1
50	0

Fréquence observée des 1 (f = moyenne des 0/1)	0,16
Probabilité théorique p	0,100
Écart entre f et p	0,060

Fréquence : 8 produits défectueux / 50 produits testés = 0,16. Normalement, on devrait à une probabilité de 10% (0,1), d'où un écart de 0,06 (6%).

NB : formule à utiliser sur Excel pour créer une simulation : chaque tirage vaut 1 avec probabilité p, et 0 sinon : =SI(ALEA()<p ; 1 ; 0)

Lorsqu'on lance la simulation avec 100 échantillons simulés on arrive aux résultats suivants :

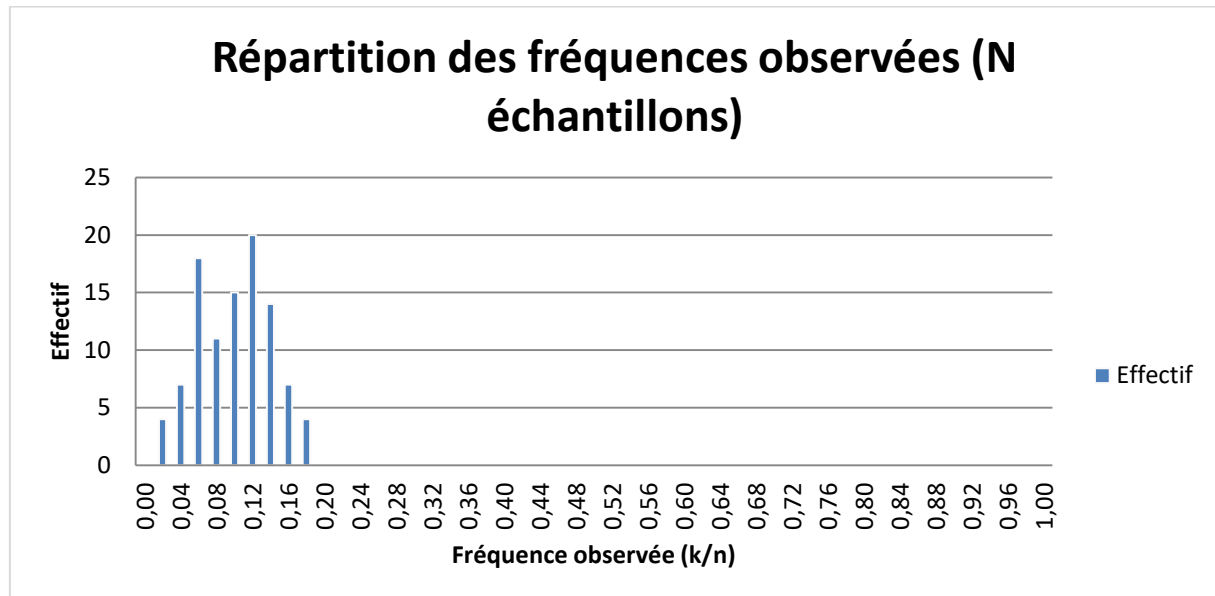
k (succès)	Fréquence k/n	Effectif
0	0,00	1
1	0,02	2
2	0,04	7
3	0,06	14
4	0,08	19
5	0,10	10
6	0,12	17
7	0,14	16
8	0,16	7
9	0,18	3
10	0,20	2
11	0,22	0
12	0,24	2
13	0,26	0
14	0,28	0
15	0,30	0
16	0,32	0
17	0,34	0
18	0,36	0
19	0,38	0
20	0,40	0
21	0,42	0
22	0,44	0
23	0,46	0
24	0,48	0

25	0,50	0
26	0,52	0
27	0,54	0
28	0,56	0
29	0,58	0
30	0,60	0
31	0,62	0
32	0,64	0
33	0,66	0
34	0,68	0
35	0,70	0
36	0,72	0
37	0,74	0
38	0,76	0
39	0,78	0
40	0,80	0
41	0,82	0
42	0,84	0
43	0,86	0
44	0,88	0
45	0,90	0
46	0,92	0
47	0,94	0
48	0,96	0
49	0,98	0
50	1,00	0

Par exemple, 14 échantillons ont obtenu 3 produits défectueux sur des lots de 50 produits.

NB : Sur excel, la fréquence ne peut prendre que les valeurs k/n . On compte, pour chaque k , le nombre d'échantillons ayant exactement k succès : $=NB.SI(plage\ des\ Succès^* ; k)$.

* plage des succès : création d'une feuille en générant 100 simulations avec la formule : $SI(ALEA()<p ; 1 ; 0)$



3.2. Distance à p : fréquent ou rare ?

Certaines fréquences observées sont proches de p (ce qui arrive souvent), d'autres s'en éloignent beaucoup (ce qui arrive plus rarement). On peut mesurer cet écart à l'aide de l'écart-type s de la série des fréquences observées.

On évalue le pourcentage d'échantillons dont la fréquence observée se situe à une distance s , $2s$ ou $3s$ de p (où s est l'écart-type de la série des fréquences observées). Plus une fréquence s'éloigne de p (en termes de multiples de s), plus la situation est **rare** dans le cadre du modèle ; plus elle est proche de p , plus elle est **fréquente**.

$$\text{Ecart-type} = 1/2\sqrt{n}$$

Plus la taille n de l'échantillon est grande, moins les fréquences observées « dansent » autour de p : **les gros échantillons sont plus stables que les petits.**

En moyenne,

- 68% des fréquences appartiennent à l'intervalle $[p - s ; p + s]$.
- 95% des fréquences appartiennent à l'intervalle $[p - 2s ; p + 2s]$.
- 99% des fréquences appartiennent à l'intervalle $[p - 3s ; p + 3s]$

Application pratique (suite) :

Pour cette dernière partie, on va travailler sur une probabilité de succès de 10% (10% des pièces sont défectueuses). Avec un échantillon (n) de 50 produits. Enfin pour la simulation on travailler sur 100 échantillons simulés.

Avec $p = 0,1$, $n = 50$, $N = 100$ échantillons, l'écart-type **observé** de la série des fréquences est :

$$\text{Ecart-type} = \sqrt{1/2 \cdot 50} = 0,071$$

Intervalle	Avec $s = 0,071$
$[p-s ; p+s]$	$[0,029 ; 0,171]$
$[p-2s ; p+2s]$	$[0 ; 0,242]^*$
$[p-3s ; p+3s]$	$[0 ; 0,313]^*$

$$[p-s ; p+s] \Rightarrow [0,1 - 0,071 ; 0,1 + 0,071] \Rightarrow [0,029 ; 0,171]$$

* Les bornes inférieures sont négatives, ce qui n'a pas de sens pour une fréquence (toujours comprise entre 0 et 1). En pratique, on remplace cette borne par 0 : l'intervalle utile devient **[0 ; 0,313]**.

Intervalle	Bornes	k concernés	Effectif	% observé	% moyen
$[p-s ; p+s]$	$[0,029 ; 0,171]$	$k = 2 \text{ à } 8$	$7+14+19+10+17+16+7 = 90$	90%	68%
$[p-2s ; p+2s]$	$[0 ; 0,242]$	$k = 0 \text{ à } 12$	tous les échantillons	100%	95%
$[p-3s ; p+3s]$	$[0 ; 0,313]$	$k = 0 \text{ à } 15$	tous les échantillons	100%	99%