

VII. Annexes

A. Couple d'indicateurs et problèmes de minimisation

Le but de cette annexe est de présenter le lien entre un indicateur de position et un indicateur de dispersion qui lui est associé.

1. Position du problème

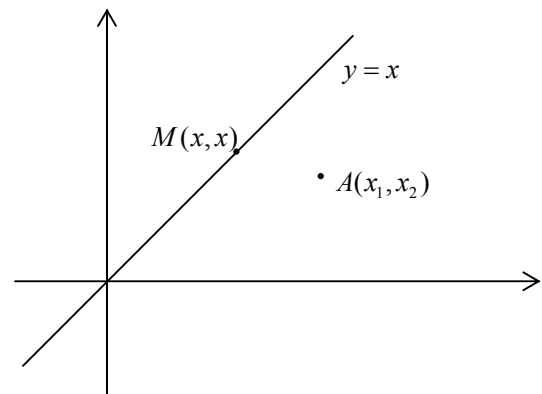
On se donne une série statistique quantitative x_1, x_2, \dots, x_n , que l'on veut « résumer » par un couple d'indicateurs donnant un renseignement sur la position et sur la dispersion de la série.

Supposons d'abord que $n = 2$ pour dégager l'idée.

La série constituée de deux valeurs est identifiée au point $A(x_1, x_2)$ du plan muni d'un repère.

On s'intéresse au point M de la droite d'équation $y = x$ qui est le plus proche de A , si toutefois ce point existe.

Si n est quelconque, on identifie la série au point $A(x_1, \dots, x_n)$ de \mathbb{R}^n et on s'intéresse, de la même manière, au point $M(x, \dots, x)$ qui est le plus proche de A . Ce point M , s'il existe, réalise la plus courte distance de A à la droite $(O, \mathbb{R}\bar{u})$, avec $\bar{u}(1, \dots, 1)$.



Si le point $M(x, \dots, x)$ précédemment décrit existe, nous convenons de nommer indicateur de position de la série le nombre x , et indicateur de dispersion associé la distance $d(A, M)$.

Il existe plusieurs distances dans \mathbb{R}^n . Recherchons les couples d'indicateurs correspondant à trois distances classiques :

$$d_1(A, M) = \frac{1}{n} \sum_{i=1}^n |x_i - x|, \quad d_2(A, M) = \frac{1}{n} \sum_{i=1}^n (x_i - x)^2, \quad d_\infty(A, M) = \max_{i=1, \dots, n} |x_i - x|.$$

Ces trois expressions dépendent uniquement de la variable réelle x . Dans la suite, nous les notons plus simplement $d_1(x)$, $d_2(x)$ et $d_\infty(x)$.

2. Étude des trois fonctions d_1 , d_2 et d_∞

Commençons par étudier la fonction d_2 . Si \bar{x} désigne la moyenne arithmétique des valeurs x_1, x_2, \dots, x_n , alors on a :

$$d_2(x) - d_2(\bar{x}) = \frac{1}{n} \sum_{i=1}^n ((x_i - x)^2 - (x_i - \bar{x})^2) = \frac{1}{n} \sum_{i=1}^n (\bar{x} - x)(2x_i - x - \bar{x}) = (\bar{x} - x)^2 \geq 0.$$

Donc $d_2(x) - d_2(\bar{x}) \geq 0$.

Ainsi pour tout x , $d_2(x) \geq d_2(\bar{x})$. La valeur minimale de la fonction d_2 est atteinte en $x = \bar{x}$ et est égale à la variance σ^2 de la série statistique. La moyenne \bar{x} est donc associée assez naturellement à l'écart-type via cette propriété. On aurait pu aussi étudier les variations de la fonction d_2 et montrer qu'elle admet un unique minimum au point $x = \bar{x}$.

Conclusion : avec la distance d_2 , le couple d'indicateurs associé à la série est le couple (moyenne, écart-type).

Le cas de la fonction d_1 est moins courant dans la littérature. On commence par ordonner par ordre croissant les observations, et on suppose donc désormais que $x_1 \leq x_2 \leq \dots \leq x_n$. Un calcul un peu plus fastidieux amène à distinguer deux cas :

- si n est impair (égal à $2p + 1$), d_1 a un unique minimum atteint en $x = x_{p+1}$;
- si n est pair (égal à $2p$), d_1 admet tout point de l'intervalle $[x_p; x_{p+1}]$ comme minimum.

Dans les deux cas, une valeur qui minimise d_1 est une médiane Me de la série statistique. Comme on l'avait déjà noté dans les classes du collège, dans le cas d'une série comportant un nombre pair d'observations, une médiane n'est pas définie de manière univoque et il appartient donc de choisir une convention si on veut définir "la" médiane. Néanmoins, on peut remarquer que la valeur minimale de d_1 obtenue est $d_1(Me) = \frac{1}{n} \sum_{i=1}^n |x_i - Me|$ qui est l'écart absolu moyen à la médiane. Ainsi une médiane est associée naturellement à cet écart moyen. Bien entendu, cet indicateur de dispersion est bien moins utilisé que l'écart interquartile $Q_3 - Q_1$.

Conclusion : avec la distance d_1 , le couple d'indicateurs associé à la série est le couple (médiane, écart moyen à la médiane).

La fonction d_∞ admet un unique minimum en $x^* = \frac{1}{2}(x_1 + x_n)$, milieu des deux valeurs extrêmes. C'est un indicateur de position qui n'est pas répandu, mais il est associé à un paramètre de dispersion qui, lui, est plus connu : la valeur minimale de d_∞ obtenue en x^* est égale à la moitié de l'étendue $x_n - x_1$.

Conclusion : avec la distance d_∞ , le couple d'indicateurs associé à la série est le couple (milieu des extrêmes, demi-étendue).

B. Loi faible des grands nombres

Pour une suite $(X_k)_{k \geq 1}$ de variables aléatoires indépendantes et de même loi admettant comme espérance commune m et comme variance σ^2 , la loi faible des grands nombres établit que pour tout $\varepsilon > 0$, la probabilité que la moyenne empirique $\frac{1}{n} \sum_{k=1}^n X_k$ s'écarte de m d'au moins ε tend vers 0 quand n tend vers l'infini, ce qui s'écrit formellement :

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \varepsilon\right) = 0.$$

On propose de montrer ce résultat dans le cas particulier d'un schéma de Bernoulli et on note S_n la variable aléatoire comptant le nombre de succès. La variable aléatoire S_n suit la loi binomiale $\mathcal{B}(n, p)$ où p est la probabilité d'obtenir un succès.

On se donne $\varepsilon > 0$ et on majore :

$$P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \varepsilon\right) = P\left(\left|\frac{1}{n} S_n - p\right| \geq \varepsilon\right) = P(|S_n - np| \geq n\varepsilon) = P(|S_n - E(S_n)| \geq n\varepsilon) \text{ où } S_n = \sum_{k=1}^n X_k$$
$$P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - m\right| \geq \varepsilon\right) \leq \frac{\text{Var}(S_n)}{n^2 \varepsilon^2} = \frac{E[(S_n - E(S_n))^2]}{n^2 \varepsilon^2}$$

où on a utilisé l'inégalité de Tchebychev qui stipule que pour une variable aléatoire Z admettant une variance, on a pour tout nombre réel $a > 0$: $P(|Z - E(Z)| \geq a) \leq \frac{\text{Var}(Z)}{a^2}$.

Comme la variance d'une loi binomiale de paramètres n et p est égale à $np(1-p)$ et que son espérance est np , on en déduit que :

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{k=1}^n X_k - p\right| \geq \varepsilon\right) \leq \lim_{n \rightarrow \infty} \frac{np(1-p)}{n^2 \varepsilon^2} = 0.$$

On dit aussi que $\frac{1}{n} \sum_{k=1}^n X_k$ converge en probabilité⁷ vers p quand n tend vers l'infini. Ceci prouve la loi faible des grands nombres dans le schéma de Bernoulli.

La preuve du cas général pour une suite quelconque de variables aléatoires indépendantes et de même loi se fait de manière analogue en utilisant le fait que la variance d'une somme de variables aléatoires indépendantes est égale à la somme des variances.

Le *théorème central-limit (terminologie anglo-saxonne)* ou *théorème de la limite centrée* donne des précisions sur la convergence de la moyenne empirique $\frac{1}{n} \sum_{k=1}^n X_k$ vers la moyenne commune m . Ce théorème indique comment se comporte, lorsque n tend vers l'infini, la probabilité que l'erreur $\frac{1}{n} \sum_{k=1}^n X_k - m$ appartienne à un intervalle $[a, b]$ quelconque.

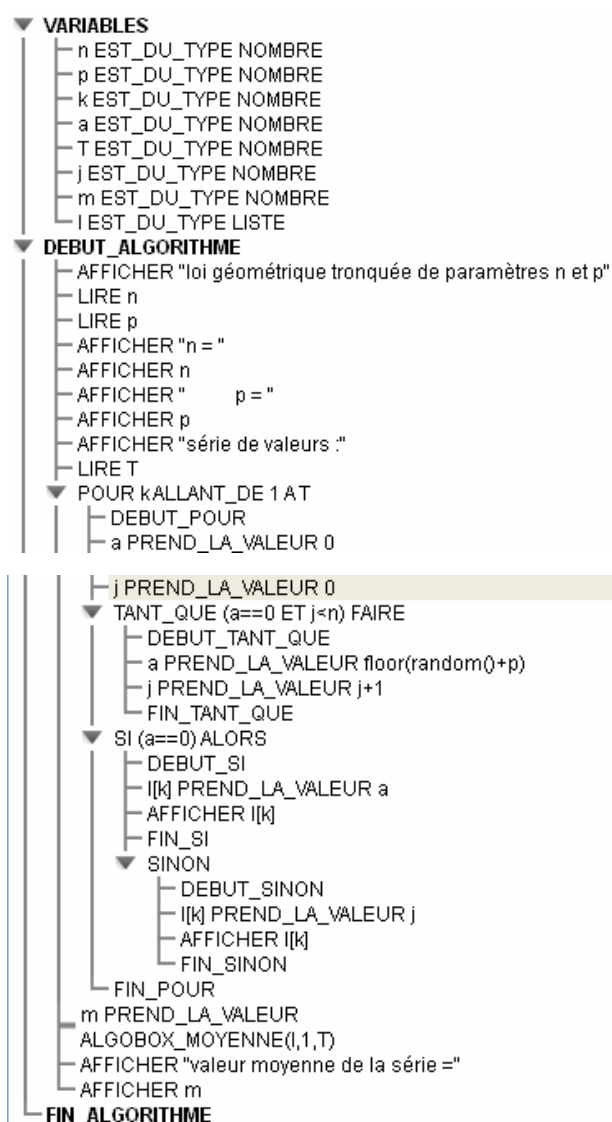
⁷ Il existe une loi forte qui correspond à un autre type de convergence, la convergence presque sûre.

C. Espérance de la loi géométrique tronquée : approches expérimentales

1. Calculatrice

<pre>PROGRAM: TPSATTEN : suite(0, 1, 1, 200) → L1 : For(N, 1, 200) : 0 → K : 0 → A : While A=0 et K < 100</pre>	<pre>PROGRAM: TPSATTEN : ent(NbrAléat+0. 07) → A : K+1 → K : End : If A=0 : Then : A → L1(N)</pre>	<pre>PROGRAM: TPSATTEN : A → L1(N) : Else : K → L1(N) : End : End : End : Disp moyenne(L1)</pre>
---	--	---

2. Algorithme modifié sur Algobox



On peut faire établir l'égalité $E(X) = \frac{1}{p} [1 - (1 + np)(1 - p)^n]$, puis utiliser un outil numérique ou graphique pour émettre une conjecture sur la limite de $E(X)$ lorsque n tend vers l'infini.

3. Point de vue numérique

On construit une feuille de calcul donnant les valeurs de $E(X)$ pour une valeur de p , en fonction de n .

	A	B	C	D	E	F	G	H	I	J	K
1	valeur de p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
2											
3	valeur de n	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$	$E(X)$
4	1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
5	5	1.14265	1.7232	1.93275	1.9168	1.78125	1.5984	1.41295	1.248	1.11105	1
6	10	3.0264312	3.38938726	2.95669967	2.42441728	1.98828125	1.66544333	1.42850394	1.24999885	1.11111111	1
7	15	4.8527217	4.29631256	3.24629471	2.49177176	1.9994812	1.66664877	1.42857119	1.25	1.11111111	1
8	20	6.35270036	4.71176962	3.31471514	2.49917736	1.99997902	1.66666643	1.42857143	1.25	1.11111111	1
9	25	7.48735704	4.8866632	3.32953364	2.49992182	1.9999992	1.66666666	1.42857143	1.25	1.11111111	1
10	30	8.30435367	4.9566721	3.33258202	2.49999282	1.99999997	1.66666667	1.42857143	1.25	1.11111111	1
11	35	8.87358002	4.98377407	3.33318812	2.49999936	2	1.66666667	1.42857143	1.25	1.11111111	1
12	40	9.26095585	4.99401847	3.33330574	2.49999994	2	1.66666667	1.42857143	1.25	1.11111111	1
13	45	9.519962	4.99782219	3.33332816	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
14	50	9.69077349	4.99921501	3.33333237	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
15	55	9.80218857	4.99971939	3.33333316	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
16	60	9.87420928	4.99990039	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
17	65	9.92041625	4.99996485	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
18	70	9.9498737	4.99998766	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
19	75	9.96855098	4.99999569	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
20	80	9.98033729	4.9999985	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
21	85	9.98774433	4.99999948	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
22	90	9.99238227	4.99999982	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
23	95	9.99527689	4.99999994	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
24	100	9.99707825	4.99999998	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
25											
26	valeur de $1/p$	10	5	3.33333333	2.5	2	1.66666667	1.42857143	1.25	1.11111111	1
27											

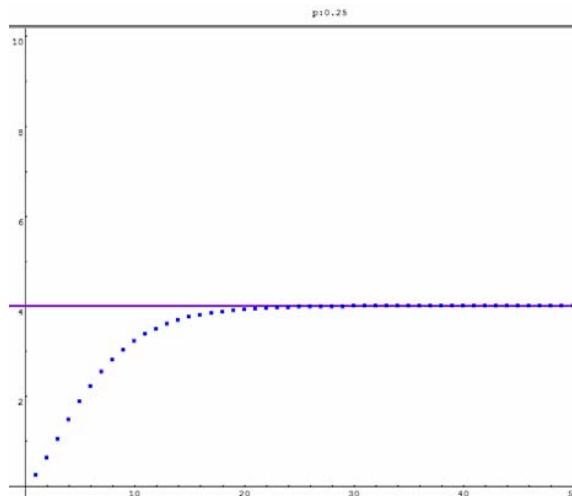
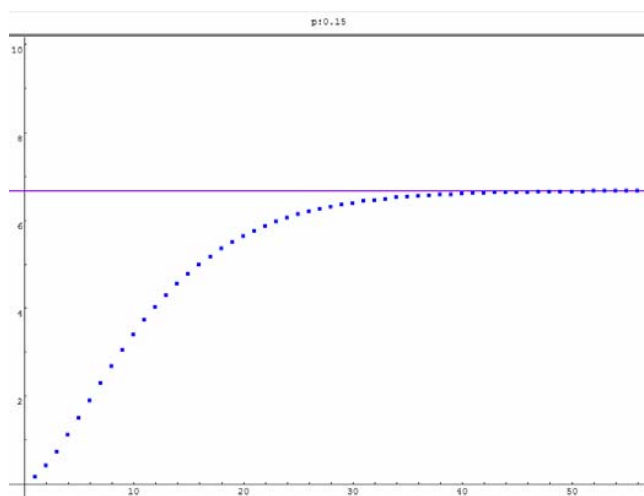
4. Point de vue graphique

On construit (ici sur géoplan) la représentation graphique de la suite définie par $u_n = \frac{1}{p} [1 - (1 + np)(1 - p)^n]$ et on observe une stabilisation pour de grandes valeurs de n .

Pour $p = 0,2$ par exemple, il semble que les valeurs se stabilisent autour de 5, d'où l'idée de tracer la droite d'équation $y = \frac{1}{p}$.

$$p = 0,15$$

$$p = 0,25$$



D. Loi géométrique

On répète dans des conditions identiques une épreuve de Bernoulli de paramètre p et on arrête le processus au premier succès obtenu. La *loi géométrique de paramètre p* est par définition la loi de la variable aléatoire X , rang du premier succès.

1. Quelques propriétés de la loi géométrique

La variable X prend ses valeurs dans \mathbf{N}^* et pour tout entier naturel k non nul : $P(X = k) = (1 - p)^{k-1} p$.

On vérifie facilement que : $\sum_{k \geq 1} P(X = k) = 1$ (somme d'une série géométrique)

On montre que : $E(X) = \frac{1}{p}$ (en utilisant la dérivée d'une série géométrique) et que $V(X) = \frac{1-p}{p^2}$ (en utilisant entre autre la dérivée seconde d'une série géométrique).

Il est à noter que l'espérance de la loi géométrique de paramètre p est la limite de l'espérance de la loi géométrique tronquée de paramètres n et p .

2. Précautions en classe de Première

La variable X prend toutes les valeurs entières sauf 0. L'univers associé n'est donc pas fini et ne figure pas aux programmes du lycée.

3. Une approche à l'aide de l'algorithmique

Le processus au cours duquel on répète dans des conditions identiques une épreuve de Bernoulli de paramètre p et que l'on arrête au premier succès obtenu, est très facile à mettre en œuvre avec un algorithme. L'instruction **ent(NbrAléat + p)** génère un nombre aléatoire entier qui vaut 1 avec une fréquence de p et 0 avec une fréquence de $(1 - p)$. On notera que, dans la pratique, le programme correspondant s'arrête toujours.

4. Langage naturel

Entrée : valeur de p
Initialisation : X prend la valeur 0
 k prend la valeur 0
Traitement : tant que $k = 0$
 k prend la valeur ent(NbrAléat p)
 X prend la valeur $X + 1$
Fin du "while"
Sortie : valeur de X

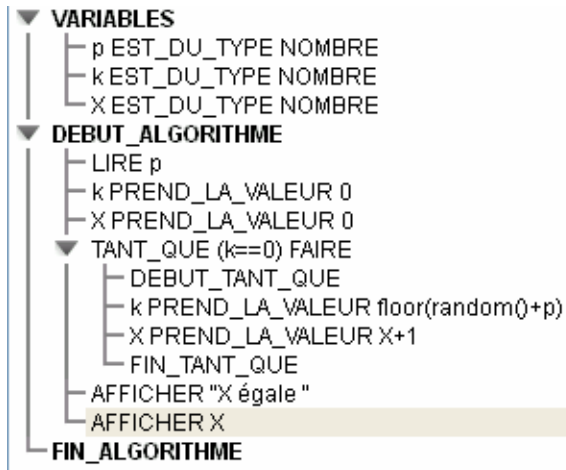
5. Sur calculatrice

```
PROGRAM: LOIGEOM
: Prompt P
: 0 → X
: 0 → K
: While K=0
: ent(NbrAléat+P)
: → K
: X+1 → X
```

```
PROGRAM: LOIGEOM
: While K=0
: ent(NbrAléat+P)
: → K
: X+1 → X
: End
: Disp X
:
```

modèle TI 84+

6. Sous Algobox



7. Sous Scilab

```
1 //loi geometrique
2 //X=rang du premier succes
3 X=0;
4 a=0;
5 p=input("donner la probabilite d'un succes : ");
6 while (a==0)
7     a=floor(rand()*p);
8     X=X+1;
9     afficher(["a="+string(a), "X="+string(X)])
10 end
11 afficher("X= "+string(X))
```

E. Quelques outils de calcul avec la loi binomiale

1. Tableur

La syntaxe $\text{LOI.BINOMIALE}(k; n; p; \text{FAUX})$ ou $\text{LOI.BINOMIALE}(k; n; p; 0)$ renvoie la probabilité $P(X = k)$, pour une variable aléatoire X de loi binomiale de paramètres n et p .

La syntaxe $\text{LOI.BINOMIALE}(k; n; p; \text{VRAI})$ ou $\text{LOI.BINOMIALE}(k; n; p; 1)$ renvoie la probabilité cumulée $P(X \leq k)$.

La syntaxe $\text{COMBIN}(n; k)$ donne la valeur du coefficient binomial $\binom{n}{k}$.

2. Un logiciel : Scilab

Calcul des coefficients binomiaux :

coef binomiaux.sce

```
1 n=input("Entrer n :")
2 coef=[0]
3 for i=1:n+1
4     for j=1:n+1
5         if j==1 then
6             coef(i,j)=1;
7         else coef(i,j)=0
8         end;
9     end;
10 end;
11 for i=2:n+1
12     for j=2:i
13         coef(i,j)=coef(i-1,j-1)+coef(i-1,j);
14     end;
15 end;
16 afficher ("Coefficients binomiaux pour n="+string(n))
17 afficher (coef(n+1,:))
```

L'algorithme ci-dessus affiche les coefficients $\binom{n}{k}$ pour k compris entre 0 et n , la valeur de n étant celle introduite au départ. (Les colonnes d'une matrice sont repérées à partir de 1, ce qui explique la présence du $n+1$).

L'algorithme ci-après affiche le triangle de Pascal.

coef binomiaux pascal.sce

```
1 n=input("Entrer n :")
2 coef=[0]
3 for i=1:n+1
4     for j=1:n+1
5         if j==1 then
6             coef(i,j)=1;
7         else coef(i,j)=0
8         end;
9     end;
10 end;
11 for i=2:n+1
12     for j=2:i
13         coef(i,j)=coef(i-1,j-1)+coef(i-1,j);
14     end;
15 end;
16 for j=1:n+1
17     afficher (coef(n+1,j))
18 end;
```


3. Deux modèles de calculatrice

a) Modèles TI (84, mais aussi 83 et 82 avec des modifications mineures)

- Calcul de probabilités avec une loi binomiale

- Probabilité de l'événement $\{X = k\}$

Instruction **DISTR** (touches **2ND** **VAR**) puis sélectionner **binomFdp**(.

Syntaxe : (nombre d'essais, probabilité de succès, valeur désirée pour la probabilité).

- Probabilité de l'événement $\{X \leq k\}$

Instruction **DISTR** (touches **2ND** **VAR**) puis sélectionner **binomFRép**(.

Syntaxe : (nombre d'essais, probabilité de succès, valeur désirée pour la probabilité).

- Valeur des coefficients binomiaux

Touche **MATH** puis **PRB** et instruction **Combinaison**. Syntaxe « $n, combinaison, k$ ».

b) Modèle Casio (graph 35+)

- Calcul de probabilités dans le cadre d'une loi binomiale

- Probabilité de l'événement $\{X = k\}$

Icône **STAT**, choisir **DIST** (touche **F5**) et **BINM** (touche **F5**). Enfin, **Bpd** (touche **F1**) et **Var** (touche **F2**).

Renseigner la boîte de dialogue :

Data : variable ; x : valeur désirée pour la probabilité ; Numtrial : nombre d'essais ; p : probabilité de succès

- Probabilité de l'événement $\{X \leq k\}$

Icône **STAT** puis saisir dans la liste 1 les valeurs prises par $k : 0, 1, \dots, n$.

Choisir **DIST** (touche **F5**) et **BINM** (touche **F5**). Enfin, **Bcd** (touche **F2**).

Renseigner la boîte de dialogue :

Data : List ; x : List1 ; Numtrial : nombre d'essais ; p : probabilité de succès

Pour chaque valeur de k , la valeur de la probabilité de l'événement $\{X \leq k\}$ est affichée dans une liste.

- Valeur des coefficients binomiaux

Touche **OPTN** puis **PRB** et instruction **nCr**. Syntaxe : « $n nCr k$ ».

F. Coefficients binomiaux et quadrillage

L'objectif est de donner une représentation des coefficients binomiaux à partir des trajets sur un quadrillage. Ce travail peut être un sujet d'étude pour la série S.

1. De l'arbre au quadrillage

On lance trois fois une pièce de monnaie équilibrée.

Il est d'usage de schématiser les huit issues de cette expérience aléatoire par les huit trajets dans un arbre tel que celui représenté ci-contre.

Sur le schéma, et pour chaque lancer, on a codé « P » pour « pile » et « F » pour « face ».

On peut convenir d'orienter la branche de l'arbre vers le nord-est pour chaque résultat « pile », vers le sud-est pour chaque résultat « face ».

On peut envisager une simplification, toujours avec la même convention : chaque déplacement vers le nord-est schématise un résultat « pile », chaque déplacement vers le sud-est schématise un résultat « face ».

L'arbre obtenu pour la même expérience aléatoire n'a plus que 4 terminaisons au lieu de 8. Toutes les issues donnant le même nombre de « pile » et de « face » correspondent en effet à plusieurs trajets qui aboutissent à une unique terminaison.

On retrouve ainsi, par exemple, qu'il y a trois trajets donnant 1 « face » et 2 « pile », les trajets étant limités aux deux directions précédentes.

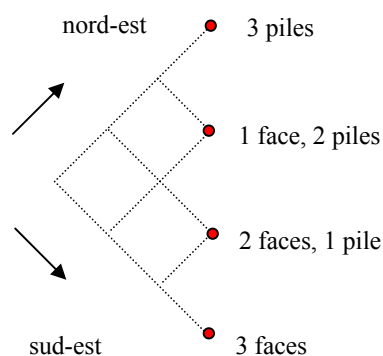
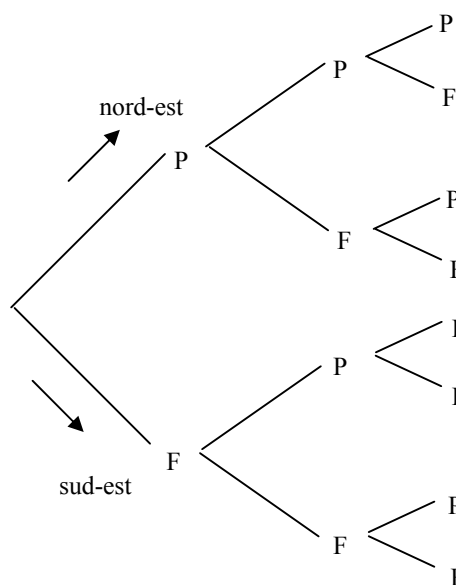
Généralisons cela à n lancers ($n \geq 1$).

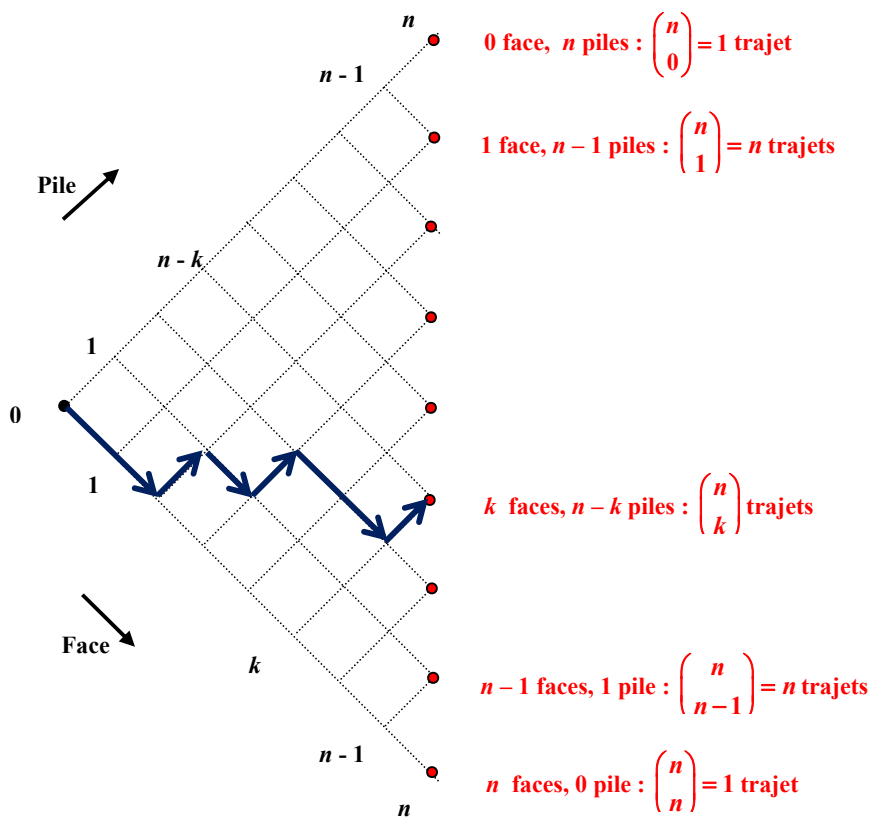
La variable X qui comptabilise le nombre de « face » suit une loi binomiale de paramètres n et $0,5$.

L'arbre correspondant aura $n+1$ terminaisons correspondant au résultat « k faces, $n-k$ piles », c'est-à-dire à l'événement $\{X = k\}$, pour tout k tel que $0 \leq k \leq n$.

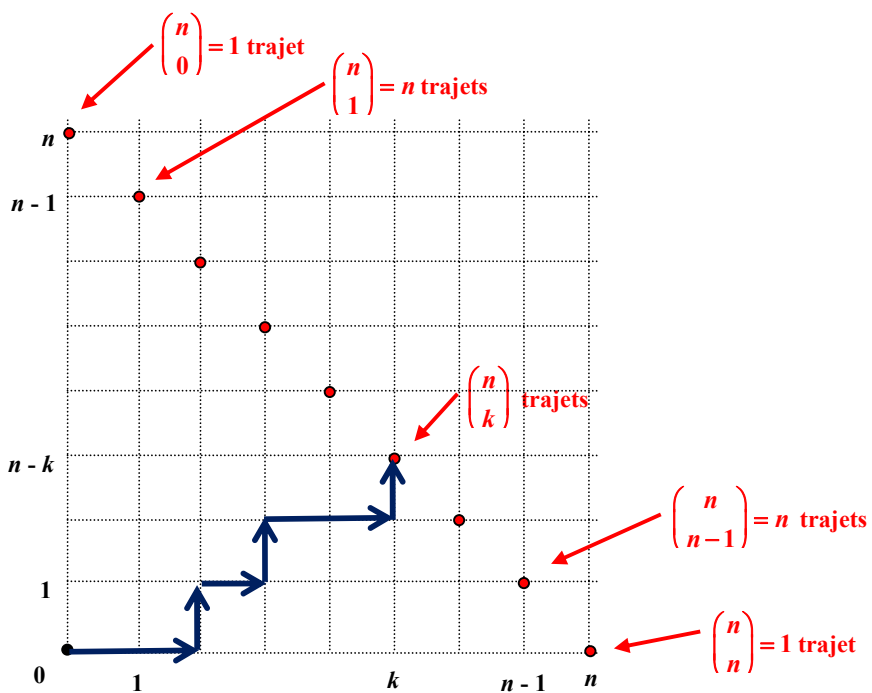
On sait que $P(X = k) = \binom{n}{k} \times \left(\frac{1}{2}\right)^n$ et que chaque trajet a pour probabilité $\left(\frac{1}{2}\right)^n$. Il en résulte que,

pour $0 \leq k \leq n$, le nombre de trajets aboutissant à la terminaison $\{X = k\}$ est égal à $\binom{n}{k}$.





Faisons tourner la figure de 45° dans le sens trigonométrique : les nœuds du schéma précédent deviennent des points à coordonnées entières dans un repère orthogonal « naturel ».



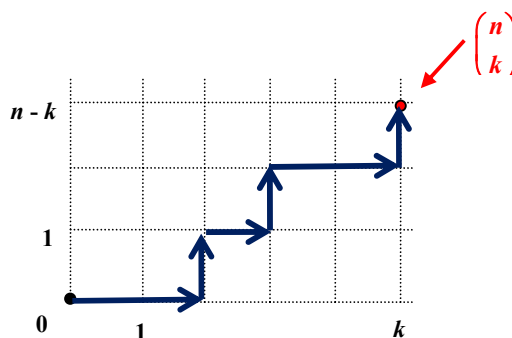
Les trajets considérés sont ceux partant de l'origine et aboutissant aux points de coordonnées $(k, n-k)$, en suivant toujours les directions vers la droite ou vers le haut (l'un d'entre eux est représenté sur la figure).

2. Interprétation des coefficients binomiaux

Retenons le résultat suivant :

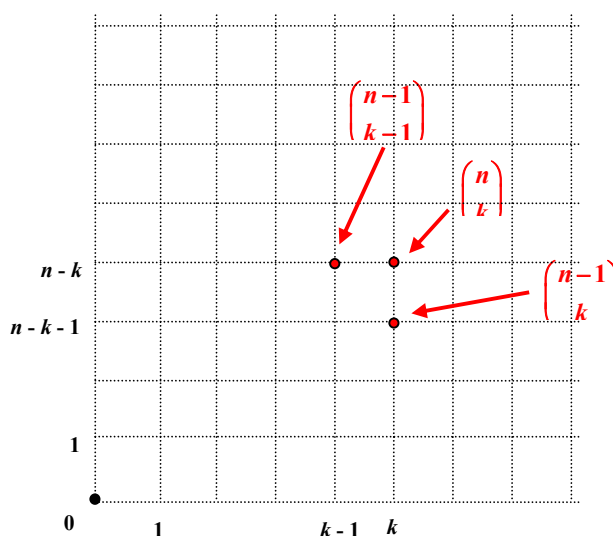
Le nombre $\binom{n}{k}$ représente le nombre de trajets reliant l'origine au point de coordonnées $(k, n-k)$, en suivant toujours les directions vers la droite ou vers le haut.

L'un de ces trajets est représenté ci-contre.



3. Formule de Pascal

Cette interprétation fournit une démonstration géométrique simple de la formule de Pascal.



Supposons que $1 \leq k \leq n-1$.

- Les trajets joignant l'origine au point de coordonnées $(k, n-k)$ se subdivisent en deux catégories :
- ceux passant par le point de coordonnées $(k, n-k-1)$, qui sont au nombre de $\binom{n-1}{k}$;
- ceux passant par le point de coordonnées $(k-1, n-k)$, qui sont au nombre de $\binom{n-1}{k-1}$.

On en déduit que, pour $1 \leq k \leq n-1$: $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$ (formule de Pascal).

Remarque

Cette interprétation des coefficients binomiaux permet l'obtention de plusieurs formules sommatoires classiques. Le développement de ce point de vue n'est pas un objectif du programme.

4. Application : le problème des pilules

Argan se croit malade, il doit prendre $2n$ pilules dans la journée. Il dispose de deux boîtes identiques A et B et, chaque matin, il place n pilules dans chacune de ses deux boîtes. À chaque prise, il choisit

une des deux boîtes de façon équiprobable, puis prend tant que c'est possible une pilule dans la boîte choisie. Au bout d'un certain temps l'une des boîtes est vide.

Combien l'autre boîte contient-elle de pilules en moyenne à ce moment-là ?

5. Simulation de l'expérience sur AlgoBox

Voici un algorithme simulant 1000 expériences, avec $n = 10$.

```

1  VARIABLES
2  a EST_DU_TYPE NOMBRE
3  b EST_DU_TYPE NOMBRE
4  c EST_DU_TYPE NOMBRE
5  d EST_DU_TYPE NOMBRE
6  k EST_DU_TYPE NOMBRE
7  DEBUT_ALGORITHME
8  c PREND_LA_VALEUR 0
9  POUR k ALLANT_DE 1 A 1000
10  DEBUT POUR
11  a PREND_LA_VALEUR 10
12  b PREND_LA_VALEUR 10
13  TANT_QUE (a!=0 ET b!=0) FAIRE
14  DEBUT_TANT_QUE
15  SI (random() < 0.5) ALORS
16  DEBUT_SI
17  a PREND_LA_VALEUR a-1
18  FIN_SI
19  SINON
20  DEBUT_SINON
21  b PREND_LA_VALEUR b-1
22  FIN_SINON
23  FIN_TANT_QUE
24  c PREND_LA_VALEUR c+MAX(a,b)
25  FIN POUR
26  d PREND_LA_VALEUR c/1000
27  AFFICHER "Nombre moyen de pilules = "
28  AFFICHER d
29  FIN_ALGORITHME
  
```

RÉSULTATS :

```

***Algorithme lancé***
Nombre moyen de pilules = 3.551
***Algorithme terminé***
  
```

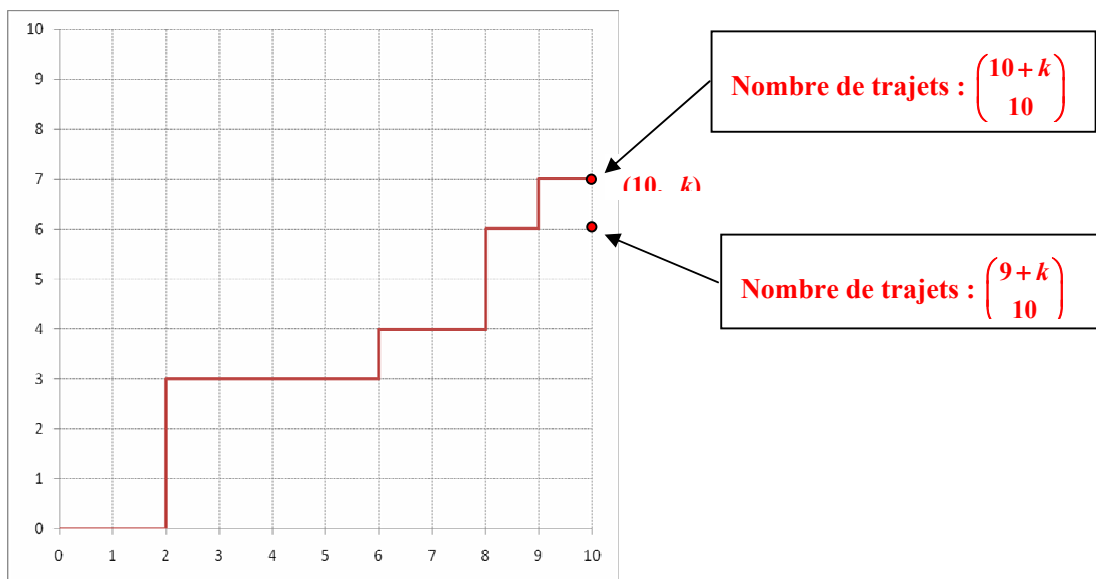
6. Traitement mathématique

On peut d'abord envisager un traitement exhaustif pour une petite valeur de n ($n = 3$ par exemple).

Pour une valeur plus élevée ($n = 10$), il est intéressant de visualiser chaque expérience comme une marche aléatoire sur un carré. Un bord est atteint (à droite ou en haut) lorsqu'une boîte est vide.

Ainsi une expérience pour laquelle il ne reste plus aucune pilule dans la boîte A et 10 pilules dans la boîte B correspond à l'unique marche aléatoire aboutissant au point de coordonnées $(10, 0)$.

Si $1 \leq k \leq 9$, une expérience pour laquelle il ne reste plus aucune pilule dans la boîte A et $10 - k$ pilules dans la boîte B correspond à une marche aboutissant au point de coordonnées $(10, k)$, sans passer par le point de coordonnées $(10, k - 1)$.



C'est l'occasion de réinvestir l'interprétation géométrique des coefficients binomiaux.

Pour $1 \leq k \leq 9$, le nombre de chemins allant de l'origine au point de coordonnées $(10, k)$ sans passer par le point de coordonnées $(10, k-1)$ est égal à $\binom{10+k}{10} - \binom{9+k}{10}$, soit encore $\binom{9+k}{9}$ d'après la formule de Pascal.

Puisqu'une telle expérience correspond à la consommation de $10+k$ pilules, la probabilité de l'événement correspondant est donc égale à $\frac{\binom{9+k}{9}}{2^{10+k}}$.

Notons X la variable aléatoire correspondant au nombre de pilules restant dans une boîte dès que l'autre est vide. Les deux boîtes jouent un rôle symétrique ; on en déduit que :

$$P(X=10) = 2 \times \frac{1}{2^{10}} = \frac{1}{2^9} ;$$

$$\text{pour } 1 \leq k \leq 9, P(X=10-k) = 2 \times \frac{\binom{9+k}{9}}{2^{10+k}} = \frac{\binom{9+k}{9}}{2^{9+k}}.$$

Remarquons que la dernière égalité englobe la première pour $k=0$.

On peut utiliser un logiciel de calcul formel pour vérifier que : $\sum_{k=0}^9 P(X=10-k) = \sum_{k=0}^9 \frac{\binom{9+k}{9}}{2^{9+k}} = 1$.

Il en est de même pour le calcul de l'espérance $E(X) = \sum_{k=0}^9 (10-k) \times \frac{\binom{9+k}{9}}{2^{9+k}} \approx 3,524$.

La simulation précédemment effectuée est en accord avec le résultat.

G. Compléments sur la prise de décision

Les compléments portent sur deux points :

- la notion d'intervalle de fluctuation unilatéral ;
- la notion d'erreur attachée à la prise de décision.

Ces compléments ne sont pas des attendus du programme.

1. L'affaire Woburn

[D'après DUCÉL Y., SAUSSEREAU B. : « La prise de décision de la Seconde à la Première », *Repères IREM*, 85, octobre 2011, Topiques éditions, Nancy (à paraître)]

Le document ressource des programmes de mathématiques de lycée professionnel propose la situation suivante :

Une petite ville des États-Unis, Woburn, a connu 9 cas de leucémie parmi les 5969 garçons de moins de 15 ans sur la période 1969-1979. La fréquence des leucémies pour cette tranche d'âge aux États-Unis est égale à 0,00052. (Source : *Massachusetts Department of Public Health*).

Les autorités concluent qu'il n'y a rien d'étrange dans cette ville. Qu'en pensez-vous ?

De façon plus précise, on peut reformuler la question sous la forme suivante : le nombre de cas observés est-il **significatif** d'une situation anormale pour cette ville ou bien peut-on considérer qu'il est simplement le fruit du hasard ?

Pour mieux faire comprendre le contexte d'expérience aléatoire sous-jacent à cette situation, on la transpose en termes de schéma d'urne : la population des garçons de moins de 15 ans sur la période 1969-1979 des États-Unis sera assimilée à une urne contenant 100 000 boules rouges ou vertes où

- les boules rouges, au nombre de 52, représentent les personnes atteintes de leucémie,
- les boules vertes représentent les personnes non atteintes.

On peut considérer, en première approximation, que la population des 5969 garçons de moins de 15 ans sur la période 1969-1979 à Woburn est assimilable à l'observation d'un échantillon (au sens de la définition donnée en Seconde) de 5969 boules, prélevées de façon équiprobable et avec remise dans l'urne.

La question posée relève alors d'un problème de **prise de décision**. Nous ne pouvons pas utiliser la démarche préconisée dans le programme de Seconde car les conditions de sa validité ne sont pas satisfaites. En effet ici n vaut 5969, $n > 25$, mais p vaut 0,00052, valeur très inférieure à 0,2.

D'après le schéma d'urne adopté, l'expérience aléatoire de départ E : « *Extraire une boule de l'urne et noter sa couleur* », est une expérience de Bernoulli de paramètre $p = 0,00052$. L'expérience aléatoire attachée à la situation est l'expérience aléatoire obtenue par 5969 « répétitions » à l'identique de E . On associe à cette nouvelle expérience aléatoire un modèle probabiliste (Ω, P) et la variable aléatoire X qui, à toute issue ω de l'expérience fait correspondre le nombre (entier), noté $X(\omega)$, de boules rouges obtenues dans l'issue ω . La variable aléatoire X suit la loi binomiale $\mathcal{B}(n; p)$ où n vaut 5969 et où p est inconnu.

a) La situation est-elle normale à Woburn ?

Se poser la question de savoir si la situation est normale à Woburn, revient à se demander si l'échantillon observé peut être considéré comme issu d'une population pour laquelle la proportion de

cas de leucémie est $p=0,00052$ comme dans tout le pays. Dans le cas contraire on sera amené à considérer que $p \neq 0,00052$.

La figure 1 représente le diagramme⁸ en bâtons de la loi binomiale $\mathcal{B}(n; p)$ avec $n=5969$ et $p=0,00052$, car on raisonne sous l'hypothèse de travail que la situation est normale à Woburn, c'est-à-dire que $p=0,00052$.

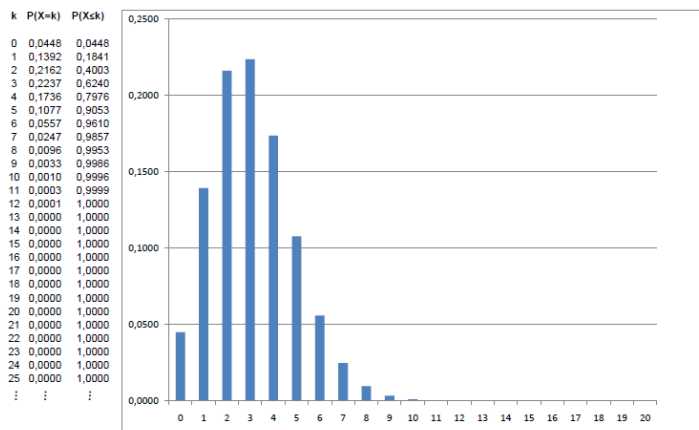


Fig. 1 : Diagramme en bâtons et probabilités cumulées de la loi binomiale pour $n=5969$ et $p=0,00052$

Un raisonnement analogue à celui fait pour les exemples précédents de prise de décision conduit à un intervalle de fluctuation, que nous qualifierons par la suite de **bilatéral**, au seuil de 95% égal à $[0 ; 7]$ pour la variable de décision X . On pourrait de même définir en adaptant les raisonnements précédents un intervalle de fluctuation bilatéral **au seuil de 90%** : c'est l'intervalle $[1 ; 6]$. Les fréquences correspondantes sont données dans le tableau de la figure 2 :

Seuil de 95 %

IF bilatéral	Effectif	Fréquence
Borne inf	0	0,0000
Borne sup	7	0,0012
p-1/racine n		-0,0124
p+1/racine n		0,0135

Seuil de 90 %

IF bilatéral	Effectif	Fréquence
Borne inf	1	0,0002
Borne sup	6	0,0010

Fig. 2 : Intervalles de fluctuations bilatéraux aux seuils de 95% et de 90%

On remarque que l'intervalle de fluctuation bilatéral $[0 ; 0,0012]$ au seuil de 95% trouvé ici est extrêmement différent de l'intervalle de fluctuation donné par la formule de la classe de seconde :

$$\left[p_0 - \frac{1}{\sqrt{n}}, p_0 + \frac{1}{\sqrt{n}} \right] = \left[0,00052 - \frac{1}{\sqrt{5969}}, 0,00052 + \frac{1}{\sqrt{5969}} \right] = [-0,0124; 0,0135]$$

⁸ Ce diagramme (Fig. 1) n'est pas obtenu par simulation mais par calcul en utilisant la fonctionnalité LOI.BINOMIALE du tableur.

Il est clair, comme on l'a déjà noté, que la formulation donnée en seconde n'est pas du tout adaptée au contexte de cette situation, ni aux conditions de l'observation, pour prendre la décision.

Les règles de décision correspondant à chacun des intervalles de fluctuation aux seuils respectifs de 95% et de 90% sont représentées dans les deux figures 3 et 4 :

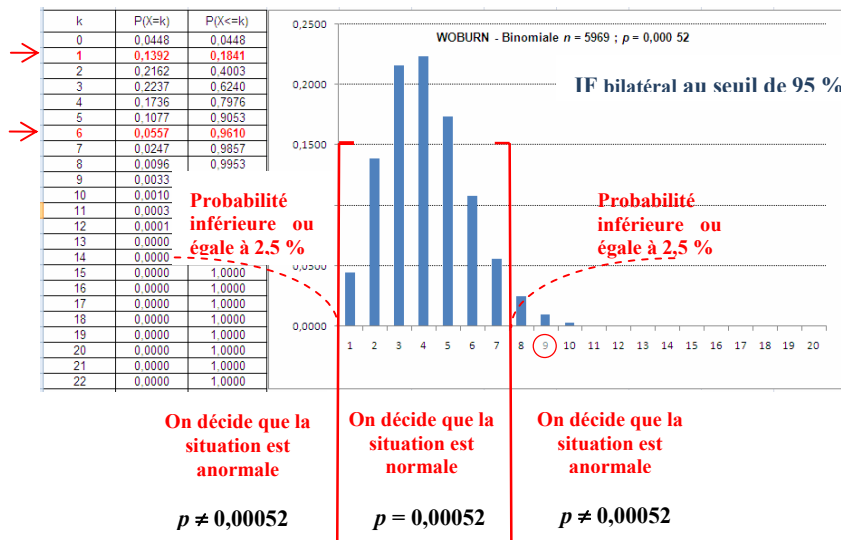


Fig. 3 : Règle de décision avec l'intervalle de fluctuation bilatéral au seuil de 95 %

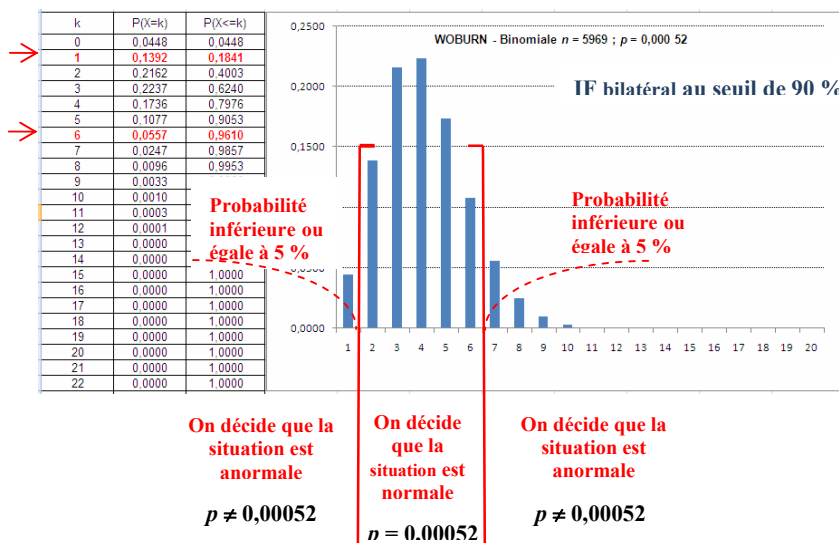


Fig. 4 : Règle de décision avec l'intervalle de fluctuation bilatéral au seuil de 90 %

Dans les deux cas, comme l'effectif observé est $x = 9$, on est conduit à décider que le nombre de cas de leucémie observé est anormal pour une proportion de référence $p = 0,00052$ valable pour tout le pays.

Le traitement mathématique est analogue aux exemples traités précédemment comme nous l'avons remarqué. Cependant si on prend en compte la signification réelle du contexte de l'affaire Woburn, nous sommes amenés à modifier le regard que nous avons sur cette situation. Dans les exemples traités jusqu'à présent la logique conduit à mettre sur le même plan, lorsque la proportion p n'est pas p_0 , le

cas où la proportion est supérieure strictement à p_0 (valeur théorique retenue) et celui où la proportion est inférieure strictement à p_0 .

Il en est tout autrement dans la situation de Woburn où il s'agit d'un problème qui touche à la santé publique. En cas de rejet de l'hypothèse $p = p_0$, le cas $p < p_0$ signifie, certes que la situation est anormale, mais concrètement qu'il y a, toute proportion gardée, moins de cas de leucémie que dans le reste du pays. Ce qui est une bonne chose en soi et peut rendre la ville de Woburn agréable à vivre et attractive. En revanche le cas $p > p_0$ signifie également que la situation est anormale, mais concrètement qu'il y a, toute proportion gardée, plus de cas de leucémie que dans le reste du pays. Ce qui rend la ville de Woburn plus dangereuse à habiter qu'ailleurs.

On voit ainsi que les enjeux ne sont pas les mêmes des deux côtés de l'intervalle de fluctuation. Aussi, plutôt que de se demander si la situation est normale, ce qui formellement revient à trancher entre les hypothèses $p = p_0$ et $p \neq p_0$, il vaut mieux se demander si la situation à Woburn ne serait pas, au vu du nombre de cas observé, plutôt dangereuse pour la santé publique.

b) La situation est-elle dangereuse pour la santé à Woburn ?

Compte tenu des enjeux, c'est plutôt la question sur la dangerosité de la situation qui est pertinente dans le cas de Woburn. Il s'agit de trancher entre l'hypothèse « *La situation n'est pas dangereuse* » ce qui formellement s'écrit $p \leq p_0$ et l'hypothèse « *La situation est dangereuse* » ce qui formellement s'écrit $p > p_0$, avec $p_0 = 0,00052$.

Pour statuer, on trace le diagramme en bâtons de la variable aléatoire X qui suit la loi binomiale pour $n = 5969$ et $p = 0,00052$. Si le nombre de cas observé x est faible, il n'y aura pas lieu de penser à un danger. En revanche, si le nombre de cas observé x est très élevé, il y aura lieu de penser à un danger. Mais on sait que si $p = 0,00052$, on peut quand même avoir des échantillons pour lesquels la valeur observée x de X est relativement élevée. **La question est de déterminer une valeur b au-delà de laquelle on estimera que la valeur élevée de x n'est plus le fruit de la fluctuation due au hasard, mais est plutôt révélatrice d'une situation dangereuse.**

L'idée est de partager l'axe de valeurs de la variable de décision X seulement en deux intervalles, $[0, b]$ et $]b, n]$, au lieu de trois comme dans les prises de décision précédentes. Tant que la valeur x observée sera proche de 0 (i.e. dans $[0, b]$), il n'y aura pas lieu de déclarer la situation dangereuse. Au-delà de b , c'est-à-dire si x est dans $]b, n]$, on déclarera la situation dangereuse. Dans cette approche, $[0, b]$ sera l'intervalle de fluctuation. On parlera alors d'intervalle de fluctuation **unilatéral** pour le distinguer de l'intervalle de fluctuation utilisé en classe qu'on a qualifié de bilatéral. Si on fixe le seuil à 95 %, b sera choisi pour que $P(X > b) \leq 0,05$.

La lecture des probabilités cumulées de la loi binomiale pour $n = 5969$ et $p = 0,00052$ de la figure 1 donne $b = 6$. Ce qui conduit, comme $x = 9$, à décider que la situation est dangereuse pour la santé à Woburn.

De plus, par construction de l'intervalle de fluctuation $[0, 6]$, on peut affirmer qu'en prenant cette décision on a moins de 5 % de chances de se tromper. Plus précisément, la probabilité de décider que la situation est dangereuse, alors qu'elle ne l'est pas, est égale à $P(X > 6) = 1 - P(X \leq 6) \approx 1 - 0,9610$, soit environ 4,9 %.

La règle de décision est représentée graphiquement par la figure 5 :

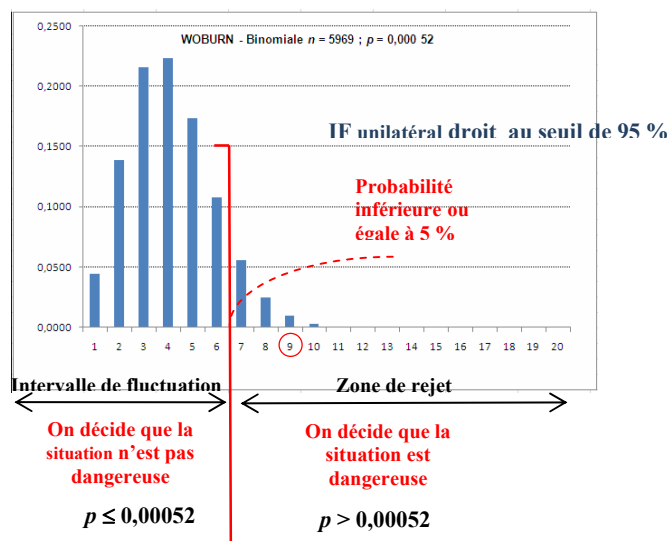


Fig. 5 : Règle de décision avec l'intervalle de fluctuation unilatéral au seuil de 95%

Les bornes des intervalles de fluctuation bilatéraux et celles de l'intervalle de fluctuation unilatéral au seuil de 95% sont rassemblées dans le tableau de la figure 6 :

Seuil de 95 %

IF bilatéral	Effectif	Fréquence
Borne inf	0	0,0000
Borne sup	7	0,0012
$p-1/\text{racine } n$		-0,0124
$p+1/\text{racine } n$		0,0135

Seuil de 90 %

IF unilatéral	Effectif	Fréquence
Borne inf	0	0,0000
Borne sup	6	0,0010

Fig. 6 : Comparaison du cas bilatéral et du cas unilatéral au seuil de 95%

c) Commentaires

Alors que les autorités locales et les experts gouvernementaux ont conclu, dans un premier temps, qu'il n'y avait rien d'étrange dans le nombre de cas de leucémie observés, à la suite d'actions et d'études entreprises par les familles avec leurs propres experts, le Département de Santé Publique du Massachusetts a officiellement confirmé en avril 1980 que le taux de leucémie constaté était anormalement élevé. La recherche des causes a conduit à soupçonner l'eau de la ville polluée par le trichloréthylène. Cette petite histoire illustre bien les enjeux de la démarche statistique.

On a vu dans cet exemple qu'une même situation peut donner lieu à plusieurs questionnements possibles et, en conséquence, à des traitements mathématiques différents :

- soit on souhaite décider entre *La situation est normale* (i.e. $p = 0,00052$) et *La situation n'est pas normale* (i.e. $p \neq 0,00052$) : le risque est à prendre en compte des deux côtés de l'intervalle de fluctuation, ce qui conduit à un intervalle de fluctuation bilatéral.
- soit on souhaite décider entre *La situation présente un danger pour la santé* (i.e. $p > 0,00052$) et *La situation ne présente pas un danger pour la santé* (i.e. $p \leq 0,00052$) : le risque est à prendre en compte d'un seul côté de l'intervalle de fluctuation, ce qui conduit à un intervalle de fluctuation unilatéral.

Dans les deux cas les calculs sont effectués avec $p = 0,00052$ et la variable de décision utilisée, $X = nF$, suit la loi binomiale $\mathcal{B}(n; p)$ avec $n = 5969$ et $p = 0,00052$.

Concernant par exemple la situation à Woburn, comme l'observation donne $x = 9$, on décidera que la situation est, avec le même seuil de 95 %, anormale en utilisant l'intervalle de fluctuation bilatéral, et dangereuse en utilisant l'intervalle de fluctuation unilatéral. Mais si l'observation avait donné $x = 0$ ou $x = 7$, on aurait été amené à déclarer la situation anormale avec l'intervalle de fluctuation bilatéral au seuil de 90 %, mais normale avec l'intervalle de fluctuation bilatéral au seuil de 95 %. De même, avec le même seuil de 95 %, si l'observation avait donné $x = 7$, on aurait été amené à déclarer la situation normale en utilisant l'intervalle de fluctuation bilatéral, mais dangereuse en utilisant l'intervalle de fluctuation unilatéral.

On voit bien que le choix de la question à poser (*La situation est-elle normale ? La situation présente-t-elle un danger pour la santé ?*), le choix du seuil (90 % ou 95 % par exemple), la forme de l'intervalle de fluctuation (bilatéral ou unilatéral) vont influencer sur la règle de décision à adopter, et par conséquent sur la décision elle-même.

Dans une démarche de prise de décision, il est donc nécessaire de clarifier en premier lieu le choix des hypothèses pour formaliser un questionnement qui, lui-même, dépend des préoccupations du décideur liées aux enjeux (économiques, sociaux, sanitaires, politiques, ...) de la situation. Le seuil devra également être défini en amont de la mise en forme mathématique. Ce choix des hypothèses déterminera ensuite la forme de l'intervalle de fluctuation à utiliser. Il est donc nécessaire de veiller à ce que la forme de l'intervalle de fluctuation utilisé soit toujours en cohérence avec le questionnement naturellement induit par la situation concrète, même si pour des raisons pédagogiques on se limite en classe aux intervalles de fluctuation bilatéraux.

Dans la réalité, toute prise de décision statistique suppose, en préliminaire à la mise en œuvre mathématique, une analyse approfondie de la signification concrète des risques encourus qui ne peut être que le fruit d'une concertation interdisciplinaire entre les divers acteurs professionnels (dont le statisticien n'est qu'un des éléments) concernés par cette prise de décision.

d) Bibliographie

Ministère Éducation nationale-DGESCO : *Ressources pour la classe en baccalauréat professionnel – extrait : Probabilités et statistiques*, Document de travail, avril 2009.

Ministère Éducation nationale-DGESCO : *Ressources pour la classe de Seconde – extrait : Probabilités et statistiques*, 2009.

2. Radioactivité ou bruit de fond ?

L'activité suivante est inspirée⁹ de présentations effectuées par Monsieur Alain VIVIER, enseignant-chercheur à l'INSTN¹⁰, institut dépendant du Commissariat à l'énergie atomique à Saclay. Convaincu de l'intérêt pédagogique des expérimentations sur tableur, Monsieur VIVIER déclare : « pour ma part je n'aurais jamais approfondi ces aspects [de statistique et probabilités], indispensables en physique,

⁹ Avec l'aimable autorisation de Monsieur Alain Vivier.

¹⁰ Institut national des sciences et techniques nucléaires.

sans le tableur. Cela m'a été utile non seulement pour des aspects d'enseignement, mais aussi de recherche, notamment dans le domaine de la problématique du seuil de décision, qui peut s'avérer parfois difficile ».

Une activité sur le thème de la radioactivité, consistant à rechercher un seuil de différence significative à avec le modèle binomial, peut se présenter comme suit.

On mesure en laboratoire, avec un compteur Geiger, un objet pouvant être « radioactif ». Le compteur est réglé selon une certaine sensibilité et on effectue une mesure à un mètre de l'objet, pendant dix secondes. L'instrument compte 37 désintégrations ou « coups ». Cependant, avec ce réglage et dans ces conditions, une mesure de « bruit de fond » (correspondant à l'environnement du laboratoire) donne en moyenne un comptage de 30 coups. La question qui se pose est de savoir si la différence observée est assez importante pour considérer l'objet comme « radioactif ».

On suppose que dans le laboratoire, durant chaque centième de seconde, le compteur est aléatoirement susceptible de compter un coup de bruit de fond avec une probabilité 0,03, ce que l'on simulera à l'aide d'un tableur avec l'instruction =ENT(ALEA()+0,03).

1.
 - a. Simuler en colonne A un comptage de bruit de fond pendant 10 secondes, puis recopier vers la droite pour obtenir la simulation de 100 comptages.
 - b. Calculer la moyenne des 100 comptages simulés. Est-elle proche de 30 coups ? (Faire F9 pour obtenir d'autres simulations.)
 - c. Un comptage supérieur ou égal à 37 coups vous semble-t-il exceptionnel ?
2.
 - a. Déterminer les paramètres n et p de la loi binomiale suivie par la variable aléatoire X modélisant un comptage de bruit de fond pendant dix secondes.
 - b. Sur une nouvelle feuille, calculer une table fournissant $P(X \leq k)$ pour k allant de 0 à 1 000.
3. Soit N est le plus petit entier tel que : $P(X \leq N) \geq 0,95$. On dira qu'il y a radioactivité significative si le nombre de coups est supérieur ou égal $N + 1$.
 - a. Déterminer la valeur de N .
 - b. On observe un comptage de 37 coups. Peut-on considérer que la radioactivité est significative ?
 - c. Quelle est la probabilité de considérer que la radioactivité est significative alors que c'est un bruit de fond ?
4. On considère un objet radioactif pour lequel, durant chaque centième de seconde, le compteur est aléatoirement susceptible de compter un coup avec la probabilité 0,05. On considère la variable aléatoire Y modélisant le comptage des désintégrations pendant dix secondes.
 - a. Donner les paramètres de la loi binomiale suivie par Y .
 - b. Déterminer la probabilité de ne pas détecter comme radioactif l'objet considéré.

Éléments de réponse

3. a. $N = 39$.

k	$P(X \leq k) \approx$
37	0,9142
38	0,9381
39	0,9563
40	0,9698

b. Un comptage de 37 coups n'est pas significatif. On considère qu'il y a radioactivité à partir d'un comptage de 40 coups.

c. La probabilité cherchée correspond à $P(X \geq 40)$ où X suit la loi binomiale de paramètres $n = 1000$ et $p = 0,03$. On a choisi un seuil de 95 % de façon à rendre cette erreur peu probable, inférieure à 5 %.

4. a. $n = 1000$ et $p = 0,05$.

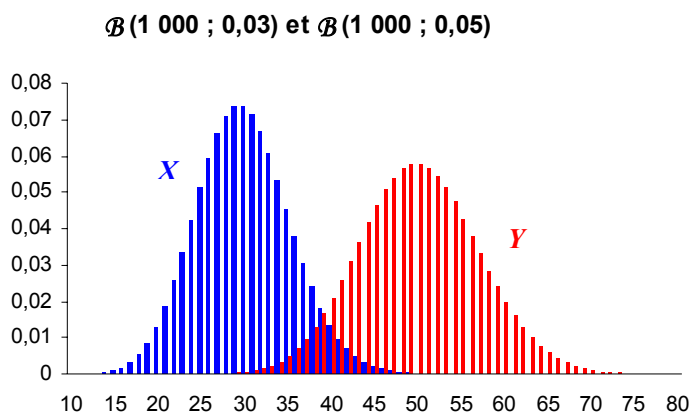
b. L'objet n'est pas détecté comme radioactif correspond à l'événement $\{Y \leq 39\}$ dont la probabilité vaut environ 0,06.

Remarques :

On a deux façons de se tromper :

- déclarer qu'un objet est radioactif alors que c'est un bruit de fond (c'est l'objet de la question 3) ;
- déclarer qu'il n'y a que du bruit de fond alors que l'objet est radioactif (c'est l'objet de la question 4).

Il est utile d'appuyer le raisonnement sur l'observation des diagrammes en bâtons des lois binomiales $\mathcal{B}(1\ 000 ; 0,03)$ et $\mathcal{B}(1\ 000 ; 0,05)$.



3. Cartes de contrôle¹¹

Dans l'industrie automobile, certains véhicules, après leur passage en peinture, présentent un défaut de type « grains ponctuels ». Ce défaut est pratiquement imperceptible, mais constitue un témoin de la qualité du processus de peinture.

On dit que le processus est « sous contrôle » lorsque 20 % des capots produits ont ce type de défaut. Des modifications apportées au processus de fabrication sont susceptibles de modifier ce pourcentage, dans un sens ou dans l'autre.

On contrôle la production en prélevant des échantillons de 50 capots. La production est suffisamment importante pour considérer qu'il s'agit de tirages au hasard avec remise.

On choisit de fixer le seuil de décision de sorte que la probabilité de rejeter l'hypothèse à tort soit inférieure à 5 %.

On accepte l'hypothèse selon laquelle la proportion dans la production est $p = 0,2$, lorsque la fréquence f observée sur l'échantillon se situe dans l'intervalle de fluctuation au seuil de 95 %.

¹¹ Mises au point en 1924 à la *Bell Telephone Company* par Walter Shewhart (1891-1967), les « cartes de contrôle » sont à la base de la « maîtrise statistique des procédés ». Toujours utilisées en raison de leur simplicité (on reporte les valeurs observées sur le graphique de la carte), elles définissent des limites de contrôle de certains paramètres de la production, comme la fréquence, la moyenne ou l'écart-type, telles que si ces limites sont dépassées, des actions de correction puissent être menées.

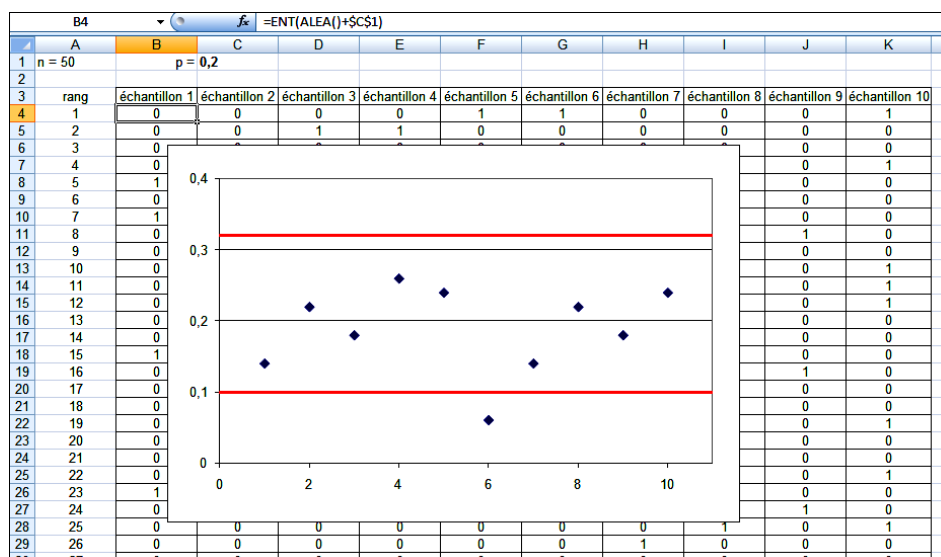
Les « limites de contrôle » sont les bornes de l'intervalle de fluctuation au seuil de 95 %, calculées en considérant la variable aléatoire X correspondant au nombre de capots présentant le défaut sur un échantillon de taille 50. Sous l'hypothèse $p = 0,2$, cette variable aléatoire suit la loi binomiale de paramètres $n = 50$ et $p = 0,2$.

1. Calculer les « limites de contrôle ».
2. Simuler le fonctionnement de cette carte de contrôle lorsque le processus est sous contrôle.
3. Le processus est sous contrôle, quelle est la probabilité de commettre une erreur de décision à partir d'un échantillon ?
- 4*¹². En réalité, la proportion de capots présentant le défaut dans la production est 0,3. On considère la variable aléatoire Y correspondant au nombre de capots présentant le défaut sur un échantillon de taille 50.
 - a. Donner les paramètres de la loi binomiale suivie par Y .
 - b. Quelle est la probabilité de décider que le processus est sous contrôle ?

Éléments de réponse

1. Les limites de contrôle sont 0,1 et 0,32 correspondant à des effectifs de 5 et 16 capots présentant le défaut.

2.



3. Si X suit la loi binomiale de paramètres $n = 50$ et $p = 0,2$, on a $P(X \leq 4) \approx 0,0185$ et $P(X \geq 17) \approx 0,0144$. La probabilité de commettre une erreur de décision lorsque $p = 0,2$ correspond à la probabilité de la zone de rejet soit environ 3,3 %.

4. a. $n = 50$ et $p = 0,3$.

b. On est dans une situation où l'hypothèse $p = 0,2$ ayant permis la construction de la carte de contrôle est fautive. En décidant que le processus est sous contrôle, on commet une erreur. La probabilité de commettre cette erreur, à partir de l'observation d'un seul échantillon, est $P(5 \leq Y \leq 16)$ où Y suit la loi binomiale de paramètres $n = 50$ et $p = 0,3$. On trouve environ 68,4 %.

¹² Le questionnement sur les erreurs n'est pas un attendu du programme. Aucune connaissance à ce propos n'est donc exigible. Cette question peut cependant être posée dans un cadre de réflexion ou de recherche.

Remarques :

- On peut, dans un premier temps, observer par simulation la fréquence des erreurs (points situés entre les lignes de contrôle alors que $p \neq 0,2$) en introduisant la valeur 0,3 en cellule C1 et en actionnant de nombreuses fois la touche F9.
- Il est difficile de distinguer $p = 0,2$ et $p = 0,3$ avec un seul échantillon de taille 50. La procédure de décision a tendance à être « conservatrice » et privilégie l'hypothèse $p = 0,2$ qui n'est rejetée que si la différence observée est réellement significative. Il est utile d'appuyer le raisonnement sur l'observation des diagrammes en bâtons des lois binomiales $\mathcal{B}(50; 0,2)$ et $\mathcal{B}(50; 0,3)$.

