Échantillonnage

1. Représentation graphique de la loi binomiale

1.1. Rappel

Si la variable aléatoire X suit la loi binomiale de paramètres n et p alors pour tout entier k, tel que $0 \leqslant k \leqslant n$: $p(X=k)=\binom{n}{k} p^k (1-p)^{n-k}$.

1.2. Calculs de p(X=k) et P(X≤k)

Exemple :

La variable aléatoire X suit la loi binomiale de paramètres $n=100$ et $p=0{,}52$.
Déterminer $p(X=2)$; $p(X=50)$.
Déterminer $p(X \leq 10)$; $p(X \leq 30)$.

i. Avec le tableur OpenOffice.

	A	B	C	D
1	k	P(X=k)	P(X≤k)	
2	2	7,73E-029		
3	50	7,35E-002		
4	10		5,70E-019	
5	30		6,95E-006	
6				
7				

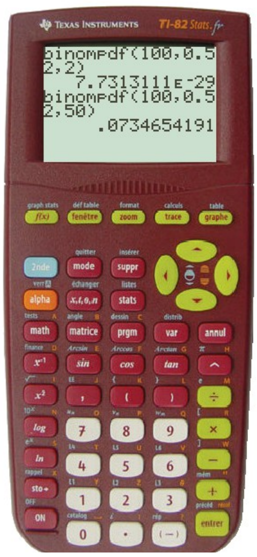
Faire un clic droit sur la cellule B2, entrer dans Formater les cellules, puis dans Nombres et dans Catégories choisir Scientifique. Faire de même dans la cellule C4.

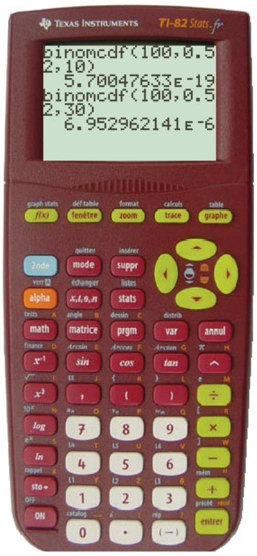
En cellule B2, pour calculer $p(X=2)$, il faut taper: « =LOI.BINOMIALE(A2;100;0,52;0).
Étendre cette formule dans la cellule B3 pour calculer $p(X=50)$.

En cellule C4, pour calculer $p(X \leq 10)$, il faut taper: « =LOI.BINOMIALE(A4;100;0,52;1).
Étendre cette formule dans la cellule C5 pour calculer $p(X \leqslant 30)$.

ii. Avec une calculatrice TI.

Sélectionner Distrib (touches 2nde Var).
Sélectionner binompdf ou binomFdp (selon le modèle de la TI).
Pour calculer $p(X=2)$, écrire binompdf(100,0.52,2).
Pour calculer $p(X=50)$, écrire binompdf(100,0.52,50).





Sélectionner Distrib (touches 2nde Var).
Sélectionner binomcdf ou binomFrép (selon le modèle de la TI).
Pour calculer $p(X\leq 10)$, écrire binomcdf (100,0.52,10).
Pour calculer $p(X\leq 30)$, écrire binomcdf (100,0.52,30).

iii. Avec une calculatrice Casio.

Sélectionner Dist (touches optn Stat)
Sélectionner Binm
Sélectionner Bpd
Pour calculer $p(X=2)$, écrire BinomialPD(2,100,0.52).
Pour calculer $p(X=50)$, écrire BinomialPD(50,100,0.52).

Sélectionner Dist (touches optn Stat)
Sélectionner Binm
Sélectionner Bcd
Pour calculer $p(X\leq 10)$, écrire BinomialCD(10,100,0.52).
Pour calculer $p(X\leq 30)$, écrire BinomialCD(30,100,0.52).

1.3. Construire une table de valeurs p(X=k) et P(X≤k)

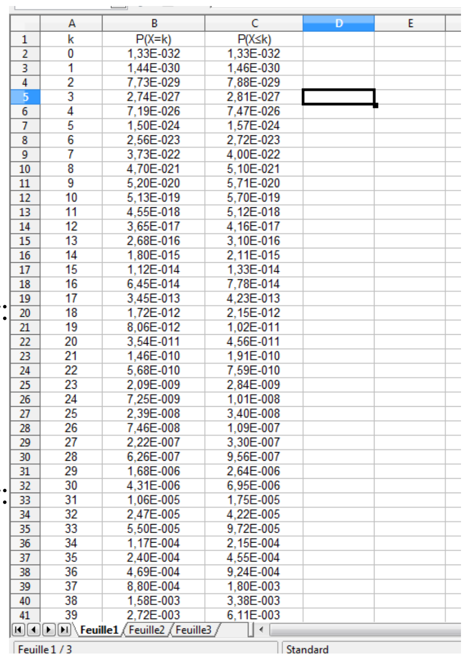
Construire une table de valeurs de $p(X=k)$ et de $p(X\leqslant k)$ pour k compris entre 0 et 100 de la variable aléatoire X qui suit la loi binomiale de paramètres $n=100$ et $p=0{,}52$.

i. Avec le tableur OpenOffice.

Dans la colonne A, on entre les entiers de 0 à 100.

Dans la colonne B, on calcule les $p(X=k)$.
En cellule B2, pour calculer $p(X=0)$, il faut taper: « =LOI.BINOMIALE(A2;100;0,52;0).
Étendre cette formule dans la colonne B pour calculer les $p(X=k)$.

Dans la colonne C, on calcule les $p(X\leq k)$.
En cellule C2, pour calculer $p(X\leq 0)$, il faut taper: « =LOI.BINOMIALE(A2;100;0,52;1).
Étendre cette formule dans la colonne C pour calculer les $p(X\leqslant k)$.



	A	B	C	D	E
1	k	P(X=k)	P(X≤k)		
2	0	1,33E-032	1,33E-032		
3	1	1,44E-030	1,46E-030		
4	2	7,73E-029	7,88E-029		
5	3	2,74E-027	2,81E-027		
6	4	7,19E-026	7,47E-026		
7	5	1,50E-024	1,57E-024		
8	6	2,56E-023	2,72E-023		
9	7	3,73E-022	4,00E-022		
10	8	4,70E-021	5,10E-021		
11	9	5,20E-020	5,71E-020		
12	10	5,13E-019	5,70E-019		
13	11	4,55E-018	5,12E-018		
14	12	3,65E-017	4,16E-017		
15	13	2,68E-016	3,10E-016		
16	14	1,80E-015	2,11E-015		
17	15	1,12E-014	1,33E-014		
18	16	6,45E-014	7,78E-014		
19	17	3,45E-013	4,23E-013		
20	18	1,72E-012	2,15E-012		
21	19	8,06E-012	1,02E-011		
22	20	3,54E-011	4,56E-011		
23	21	1,46E-010	1,91E-010		
24	22	5,68E-010	7,59E-010		
25	23	2,09E-009	2,84E-009		
26	24	7,25E-009	1,01E-008		
27	25	2,39E-008	3,40E-008		
28	26	7,46E-008	1,09E-007		
29	27	2,22E-007	3,30E-007		
30	28	6,26E-007	9,56E-007		
31	29	1,68E-006	2,64E-006		
32	30	4,31E-006	6,95E-006		
33	31	1,06E-005	1,75E-005		
34	32	2,47E-005	4,22E-005		
35	33	5,50E-005	9,72E-005		
36	34	1,17E-004	2,15E-004		
37	35	2,40E-004	4,55E-004		
38	36	4,69E-004	9,24E-004		
39	37	8,80E-004	1,80E-003		
40	38	1,58E-003	3,38E-003		
41	39	2,72E-003	6,11E-003		

Feuille1 Feuille2 Feuille3
Feuille 1 / 3 Standard

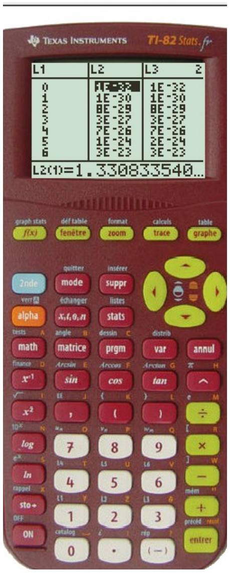
ii. Avec une calculatrice TI.

Dans la liste 1, on entre les entiers de 0 à 100.
Sélectionner 2nde stats OPS seq (ou suite selon le modèle de la TI)
Écrire (k,k,0,100,1)
Sélectionner sto 2nde 1.

Dans la liste 2, on calcule les $p(X=k)$.
Sélectionner Distrib (touches 2nde Var).
Sélectionner binompdf ou binomFdp (selon le modèle de la TI).
Écrire binompdf(100,0.52)
Sélectionner sto 2nde 2.

Dans la liste 3, on calcule les $p(X\leq k)$.
Sélectionner Distrib (touches 2nde Var).
Sélectionner binomcdf ou binomFrép (selon le modèle de la TI).
Ecrire binomcdf (100,0.52)
Sélectionner sto 2nde 3.

On obtient:



iii. Avec une calculatrice Casio.

Dans la liste 1, on entre les entiers de 0 à 100.
Sélectionner OPTN LIST seq
Écrire (k,k,0,100,1)
Sélectionner $\rightarrow$ F1 F1   1

Dans la liste 2, on calcule les $p(X=k)$.
Sélectionner Dist (touches optn Stat)
Sélectionner Binm
Sélectionner Bpd
Écrire BinomialPD(100,0.52)
et $\rightarrow$SHIFT 1 2

Dans la liste 3, on calcule les $p(X \leq k)$.
Sélectionner Dist (touches optn Stat)
Sélectionner Binm
Sélectionner Bcd
Écrire BinomialCD(100,0.52)
et $\rightarrow$SHIFT 1 3

1.4. Représentation graphique de la loi binomiale

Grâce aux listes 1 et 2 crées précédemment, on obtient un diagramme en bâtons.

2. Intervalle de fluctuation à 95% d'une fréquence correspondant à la loi binomiale

2.1. Exemple

La proportion de personnes portant des lunettes dans une population est $p=0{,}52$. On veut déterminer un intervalle de fluctuation de la fréquence des personnes portant des lunettes dans des échantillons de taille 100.

La variable aléatoire X compte le nombre de personnes portant des lunettes dans un tel échantillon.
On peut considérer que la variable aléatoire X suit la loi binomiale de paramètres 100 et 0,52.

On recherche un intervalle [a;b] (avec a et b entiers) qui contient au moins 95% des valeurs de X.

Pour cela, on partage l'intervalle [0;100] en 3 intervalles: [0;a-1], [a;b], [b+1;100]. On détermine a et b de façon que la probabilité pour que X appartienne aux intervalles [0;a-1] et [b+1;100] soit inférieure à 0,025.

On choisit donc a tel que a-1 soit le plus grand entier tel que $p(X \leqslant a-1) \leqslant 0{,}025$, c'est à dire on choisit a tel que a soit le plus entier tel que $p(X \leqslant a) > 0{,}025$

On choisit b tel que b soit le plus petit entier tel que $p(X \geqslant b+1) \leqslant 0{,}025$.
Or, $p(X \geqslant b+1) = 1 - p(X \leqslant b)$
Donc, $1 - p(X \leqslant b) \leqslant 0{,}025$
C'est à dire $p(X \leqslant 0{,}025) \geqslant 0{,}975$
<u>Conclusion:</u>
a est <u>le plus petit entier</u> tel que $p(X \leqslant a) > 0{,}025$
b est <u>le plus petit entier</u> tel que $p(X \leqslant b) \geqslant 0{,}975$

On regarde dans la table des probabilités cumulées de X calculées précédemment, ou on fait chercher la calculatrice dans la table.

Dans la liste 3, on a calculé les $p(X \leq k)$.
<u>Avec une calculatrice TI</u>
Sélectionner 2nde stats MATH 5
Écrire sum($L_3 \leq 0{,}025$)
Écrire sum($L_3 \leq 0{,}975$)

Avec une calculatrice casio
Sélectionner OPTN LIST F6 F6
Écrire Sum(List 3≤0,025)
Écrire Sum(List 3≤0,975)

La calculatrice trouve a=42 et b=62
Ainsi, l'intervalle [42;62] contient au moins 95% des valeurs de la variables X.
La taille de l'échantillon est $n=100$.
L'intervalle de fluctuation à 95% de la fréquence correspondant à X est $\left[\frac{42}{100};\frac{62}{100}\right]=[0{,}42;0{,}62]$

2.2. Propriété

L'intervalle de fluctuation à 95% de la fréquence correspondant à la réalisation, sur un échantillon aléatoire de taille n, d'une variable aléatoire X suivant une loi binomiale, est $\left[\frac{a}{n};\frac{b}{n}\right]$, où a est le plus petit entier tel que $p(X\leqslant a)>0{,}025$ et b le plus petit entier tel que $p(X\leqslant b)\geqslant 0{,}975$.

2.3. Prise de décision

On émet l'hypothèse que la proportion d'un caractère étudié dans une population est p. Pour valider cette hypothèse, on prélève au hasard et avec remise un échantillon de taille n. On observe une fréquence f du caractère.
Soit X la variable aléatoire suivant une loi binomiale $\mathscr{B}$(n;p). L'intervalle de fluctuation à 95% de la fréquence correspondant à la réalisation, de X est $\left[\frac{a}{n};\frac{b}{n}\right]$.
Cet intervalle permet de fixer **le seuil de décision** de sorte que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, est inférieure à 5%.

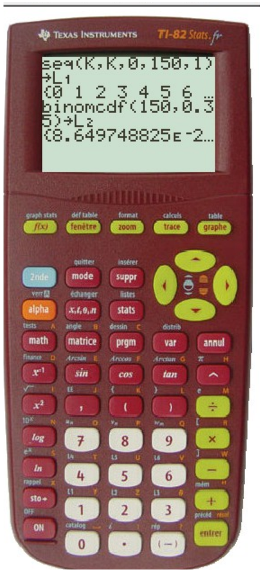
Si f n'appartient pas à l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$, alors **on rejette l'hypothèse** selon laquelle la proportion est p avec un risque d'erreur de 5%.
Si f appartient à l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$, l'hypothèse selon laquelle la proportion est p **n'est pas remise en question.**

3. Exercice

Au journal télévisé, un journaliste affirme que 35% des collégiens lisent au moins 5 livres par an.
On interroge 150 collégiens au hasard et parmi ses collégiens 33 ont lu au moins 5 livres dans l'année.
Peut-on mettre en doute, au seuil de 5%, l'affirmation du journaliste?

On considère X la variable aléatoire qui suit la loi binomiale $\mathscr{B}(150;0,35)$.

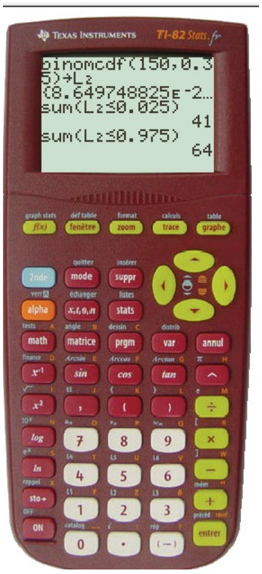
<u>Première étape:</u> on construit la table des des probabilités cumulées de X.



<u>Deuxième étape:</u> Recherche de a et b

a est le plus petit entier tel que $p(X \leqslant a) > 0{,}025$.
b est le plus petit entier tel que $p(X \leqslant b) \geqslant 0{,}975$.

a=41
b=64



<u>Troisième étape:</u> Recherche de l'intervalle $\left[\frac{a}{n};\frac{b}{n}\right]$.

$$\left[\frac{a}{n};\frac{b}{n}\right]=\left[\frac{41}{150};\frac{64}{150}\right]\approx[0{,}27;0{,}43]$$

<u>Quatrième étape:</u> On regarde si la fréquence observée fait partie ou non de l'intervalle.

$$f=\frac{33}{150}=0{,}22$$

f n'appartient pas à l'intervalle.
Au seuil de risque de 5%, on rejette l'hypothèse selon laquelle 35% des collégiens lisent au moins 5 livres par an.