Statistiques - Probabilités TI graphiques (82, 83, 84)	# De la loi binomiale à la loi normale

Ce texte, rédigé par Rémy Coste il y a plusieurs années sur une idée d'André Guillemot, revient sur le devant de la scène, avec la parution des nouveaux programmes de Premières et de Terminales. Nous avons choisi d'en conserver pour l'essentiel la trame et le déroulement, en le remaniant sur certains points, avec l'accord de son auteur, pour tenir compte des nouvelles exigences du lycée.

Introduction

La nouvelle réforme occasionne de profonds changements dans les contenus, en accordant une place très importante à l'enseignement des probabilités.
La loi binomiale est désormais introduite dès la classe de première, tandis que l'étude des probabilités continues, déjà amorcée lors des précédents programmes, se trouve bien plus fortement développée en terminale.

Dans ce contexte, la calculatrice se révèle un instrument particulièrement pertinent, voire indispensable.

De la loi binomiale à la loi de Gauss

C'est un exemple important de passage en probabilité du discret au continu.

Énoncé de la situation :

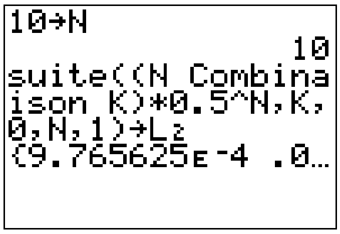
> *Un nouveau jeu électronique a fait son apparition. Dans une association,* 100 *personnes en ont acheté,* 45 *filles et* 55 *garçons. Le directeur des ventes pense que les garçons sont plus attirés par le produit que les filles. Est-ce fondé* ?

Une simulation du type de celles qui ont été utilisées en classe de seconde (répétition de 100 lancers de pièces) suffit à observer, que malgré l'hypothèse d'équiprobabilité, il n'est pas rare d'observer une répartition inégale du type 45 (filles) - 55 (garçons).
En première, on peut aller plus loin puisque l'on dispose maintenant de la loi binomiale. Considérons ici la loi binomiale X qui donne le nombre de filles dans un échantillon de 100 clients, **en supposant que les filles sont aussi intéressées que les garçons par le jeu, autrement dit que la probabilité que le client soit une fille est égale à 0,5.**
En une seule commande, on peut demander à la calculatrice de calculer, pour tout entier naturel k de 0 à n les $p(X = k)$:

$$p(X=k)=\binom{n}{k}\left(\frac{1}{2}\right)^k\left(1-\frac{1}{2}\right)^{n-k}=\binom{n}{k}\left(\frac{1}{2}\right)^n.$$

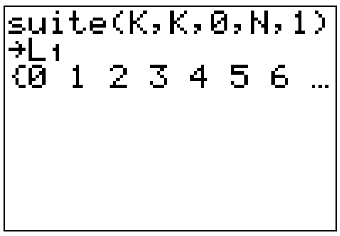
Examinons ce qui se passe avec $n = 10$, $n = 50$ puis $n = 100$.

Exemple avec n = 10



On mémorise la valeur 10 dans la variable n avec **`1Ø→N`**.

`suite((N Combinaison K)*Ø.5^N,K,Ø,N,1)→L2` génère alors la suite des valeurs de $p(X = k)$ pour la valeur choisie de n et la stocke dans la liste L_2.

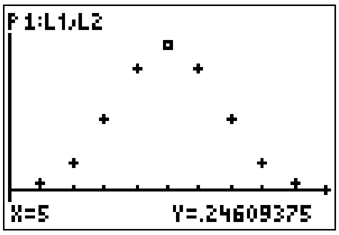


`suite(K,K,Ø,N,1)→L1` génère la suite des entiers de 0 à 10 et la stocke dans la liste L_1.

Remarque : pour l'instruction **`suite`** la valeur par défaut du pas est 1. On peut donc se dispenser d'écrire, en fin de définition, « **,1** ».

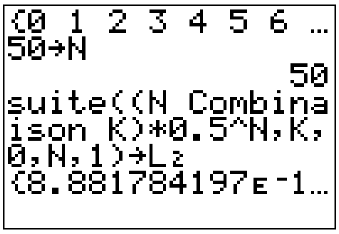
Puis on affiche le nuage de points défini par les listes 1 et 2, sur l'intervalle [0 ; n] (il suffit de faire un **Zoom Stats**).

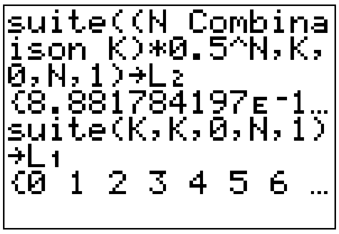
Au passage, on peut observer que la répartition 5-5, si elle est la plus probable, n'a qu'une chance sur 4 environ de se produire.

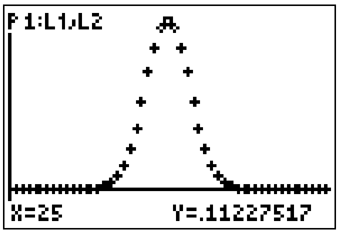


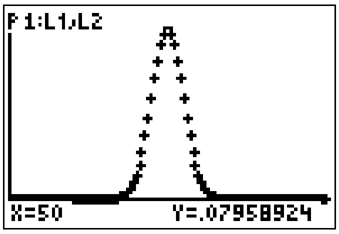
Recommençons avec n = 50 puis avec n = 100

En changeant la valeur de la variable n et en allant rechercher avec ψ ⊆ les instructions précédentes…







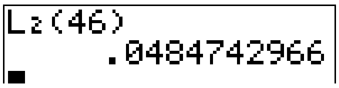


… on obtient les deux écrans ci-contre.

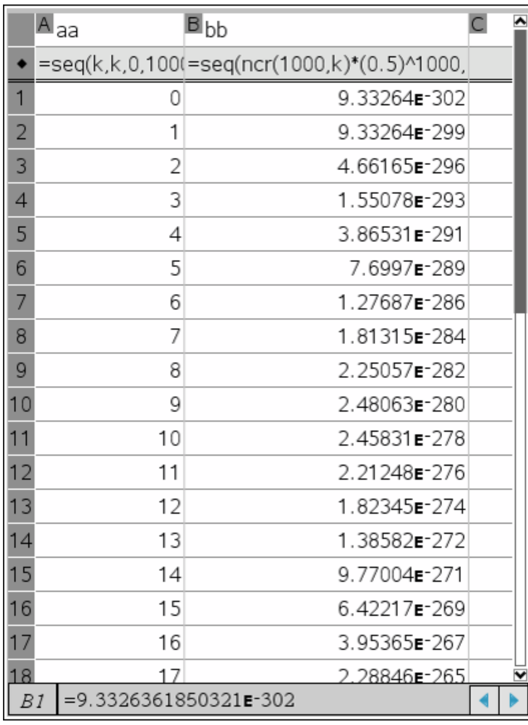
On observe que la répartition égale (25-25 ou 50-50) est toujours la plus probable, mais que sa probabilité est de plus en plus faible : ainsi, avec n = 100, $p(X = 50) \approx 0{,}080$.

Retour au problème posé

Quelle conclusion peut-on tirer de ces valeurs ?
Tout d'abord, **L2 (46)** nous renvoie une approximation de $p(X = 45)$, soit environ 0,048.

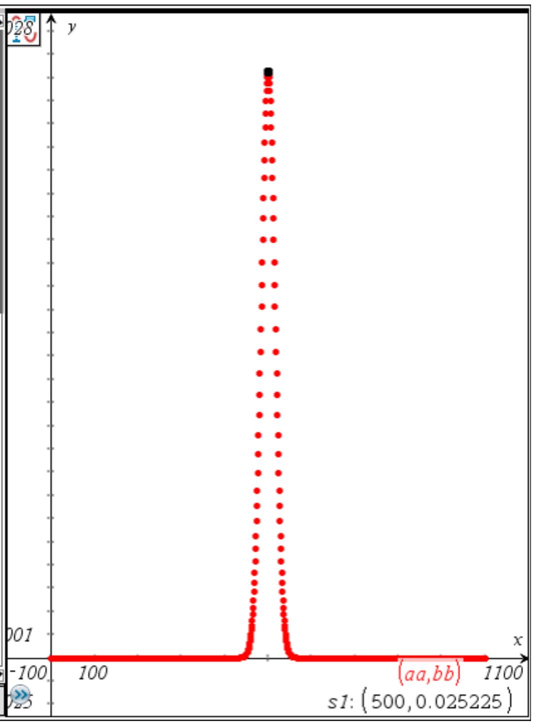


Que penser de cette probabilité ?
Elle est certes faible, mais finalement la probabilité de chacune des éventualités l'est aussi, comme le montrent les graphiques.
Les valeurs de $p(X = k)$ avec n = 1 000 (obtenues sur ordinateur avec le logiciel TI-Nspire) renforcent encore ces observations :



	A aa	B bb	C
◆	=seq(k,k,0,100(	=seq(ncr(1000,k)*(0.5)^1000,	
1	0	9.33264E-302	
2	1	9.33264E-299	
3	2	4.66165E-296	
4	3	1.55078E-293	
5	4	3.86531E-291	
6	5	7.6997E-289	
7	6	1.27687E-286	
8	7	1.81315E-284	
9	8	2.25057E-282	
10	9	2.48063E-280	
11	10	2.45831E-278	
12	11	2.21248E-276	
13	12	1.82345E-274	
14	13	1.38582E-272	
15	14	9.77004E-271	
16	15	6.42217E-269	
17	16	3.95365E-267	
18	17	2.28846E-265	

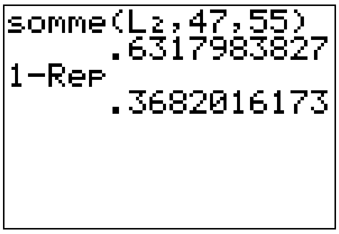
B1 =9.3326361850321E-302



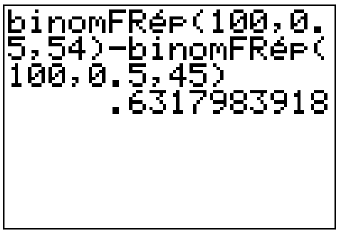
On peut conjecturer que lorsque n tend vers l'infini, $p(X = k)$ tend vers 0 pour *toutes* les valeurs de k.
Le fait que $p(X = 45)$ soit faible ne permet pas de donc pas de répondre de façon satisfaisante à la question du directeur des ventes.

Ce qui lui importe, c'est de savoir si l'hypothèse que les filles et les garçons sont également intéressés par le jeu, permet raisonnablement d'observer la distribution 45-55 rencontrée ici, avec son léger déséquilibre par rapport à 50.

On peut donc par exemple déterminer la probabilité que le nombre de filles soit compris strictement entre 45 et 55, soit $p(45 < X < 55)$, et la probabilité de son événement contraire.



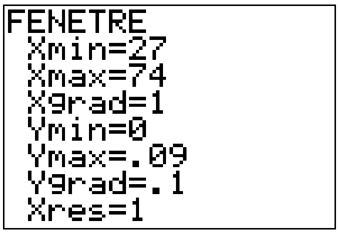
Calculons d'abord $p(45 < X < 55)$: la fonction **somme** autorise que l'on restreigne le calcul de la somme entre les termes d'indice 47 et 55, comme on le voit sur l'écran ci-contre (attention aux décalages d'indice car on commence à 0).
On peut aussi utiliser dans le menu = (ψ □) la fonction de répartition de la loi binomiale avec
binomFRép(1ØØ,Ø.5,54)-binomFRép(1ØØ,Ø.5,45).



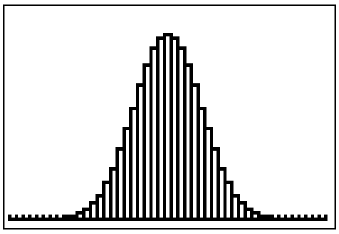
Sous l'hypothèse que les filles et les garçons sont également intéressés par le jeu, un déséquilibre *au moins* aussi important que 45-55 sur un échantillon de 100 personnes s'observe donc dans plus d'un cas sur trois : il n'est donc pas si rare.

On ne peut donc pas décemment rejeter l'hypothèse que les filles et les garçons sont également intéressés par le jeu, ce que l'on peut transmettre au directeur des ventes. Les différences observées ne sont pas significatives et ne sont probablement dues qu'aux fluctuations d'échantillonnage.

Remarque : plutôt qu'un nuage de points, affichons sur la calculatrice un diagramme en barre d'amplitudes 1 (l'intervalle choisi est [27 ; 74]).



$p(45 < X < 55)$ s'interprète alors comme **la somme des aires** de rectangles construits sur la graduation de l'axe des abscisses entre 46 et 54.



Vers la loi de Gauss

Dans les exemples que nous avons abordés, on peut constater que tous les nuages de points obtenus ont la même allure de courbe en « cloche » et c'est d'autant plus vrai que n est grand. C'est d'ailleurs le cas à chaque fois que l'on travaille avec une loi binomiale.
On peut se poser la question de savoir s'il existe une fonction dont la courbe passerait suffisamment près des points de ce nuage.

On démontre qu'il existe un modèle de fonction qui convient pour toute valeur de n suffisamment grande.

Elle est définie par : $$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

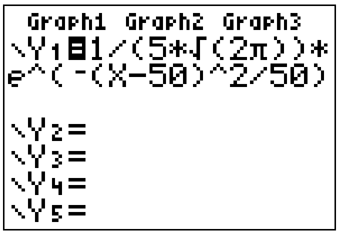
où m et σ sont respectivement l'espérance et l'écart-type de la loi binomiale considérée.

Cette fonction définit la loi de probabilité de la loi normale d'espérance m et d'écart-type σ.

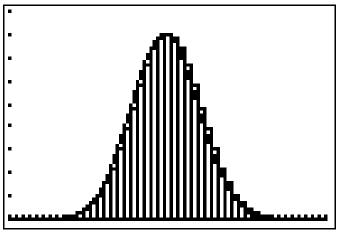
Examinons la situation pour $n = 100$.
Le calcul nous donne :

$$m = 100 \times \frac{1}{2} = 50 \text{ et } \sigma = \sqrt{100 \times \frac{1}{2} \times \left(1 - \frac{1}{2}\right)} = \sqrt{25} = 5.$$

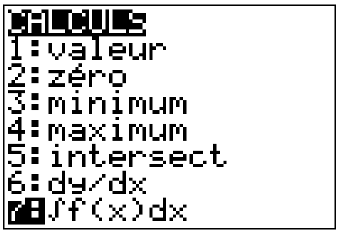
On peut écrire : $f(x) = \dfrac{1}{5\sqrt{2\pi}} e^{-\frac{(x-50)^2}{50}}$.



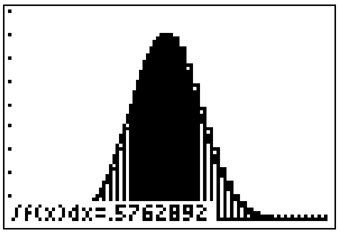
Vérifions sur la calculatrice… La courbe suit bien la forme des diagrammes en bâtons et passe assez près des points du nuage.



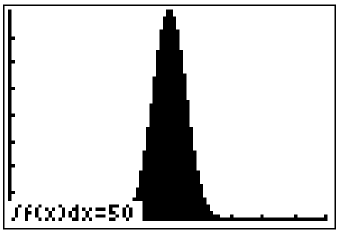
Une application importante : ce qui était la somme des aires des rectangles du diagramme en barres peut aussi s'écrire à très peu près comme l'aire sous la courbe représentative de f sur l'intervalle [46 ; 54]), soit $\int_{46}^{54} f(x)\,\mathrm{d}x$.



La calculatrice peut nous en donner une approximation. On obtient environ 0,5763 (contre 0,6318)[1].



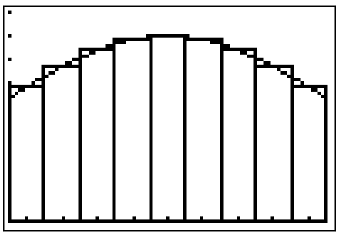
De même, on peut retrouver la moyenne de la loi binomiale : la formule $\overline{x} = \sum_{k=0}^{k=100} \left(k \times p(k)\right)$ devient $\overline{x} = \int_{0}^{100} x\, f(x).\mathrm{d}x$, proche de 50.



Correction de continuité

L'aire sous la courbe de la fonction f donne un ordre de grandeur intéressant, mais avec une précision médiocre.

Revenons sur le calcul de $p(45 < X < 55)$, que nous avons interprété comme la somme des aires de rectangles.
En leur conservant une largeur de 1, décalons ces rectangles pour que leurs sommets sur l'axe des abscisses n'aient plus des abscisses entières, mais des abscisses de type un nombre entier + 0,5.
Pour les mêmes raisons que précédemment, $p(45 < X < 55)$ s'interprète comme la somme des aires de ces rectangles, commençant à 45,5 à gauche et terminant à 54,5 à droite (c'est aussi le réglage de la fenêtre ci-contre, avec une graduation toutes les unités, la première indiquant 46).



[1] Comment améliorer la précision de l'approximation ? C'est l'objet du paragraphe suivant « Correction de continuité ».

On constate aussi que le décalage d'un demi entraîne que la somme des aires des rectangles est très proche de l'aire sous la courbe entre 45,5 et 46,5 : c'est ce que l'on appelle la correction de continuité.
Calculons cette dernière aire au moyen de la calculatrice. On obtient une nouvelle valeur approchée, 0,6319, très proche cette fois de la valeur exacte calculée plus haut.