13

Chapitre

Statistiques et probabilités

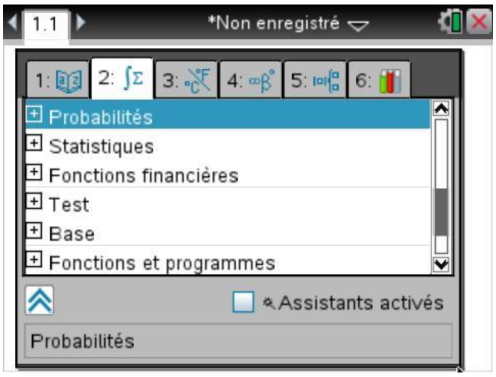
Les statistiques et les probabilités occupent une place importante dans l'enseignement de certaines classes préparatoires. Les principales fonctions nécessaires pour travailler dans ce domaine se trouvent dans les applications Calculs et Tableur & listes. L'application Données & statistiques permet d'effectuer des représentations graphiques de données statistiques, l'application Graphiques ne permettant que la représentation de nuages de points. Des fonctions définies dans ce chapitre vous permettront d'étendre les possibilités de votre unité nomade.

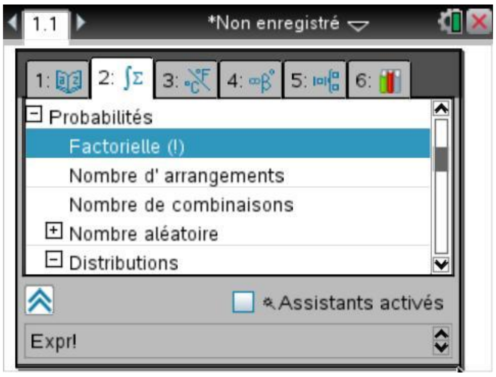
Sommaire

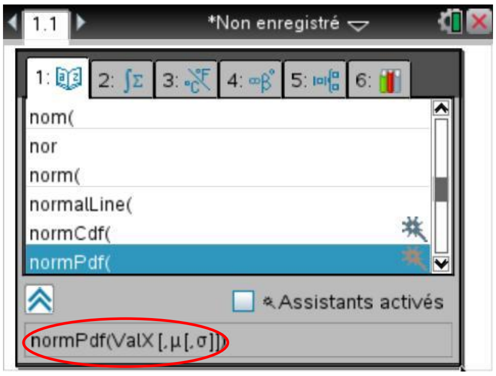
1. Les fonctions disponibles

1.1 *Où trouver ces fonctions*

Vous aurez accès aux fonctions utilisables dans le catalogue (📖) page 2, dans les rubriques : **Probabilité**, **Statistiques** mais aussi dans **Listes**. On peut y accéder aussi directement page 1 (taper la première lettre de la fonction cherchée).

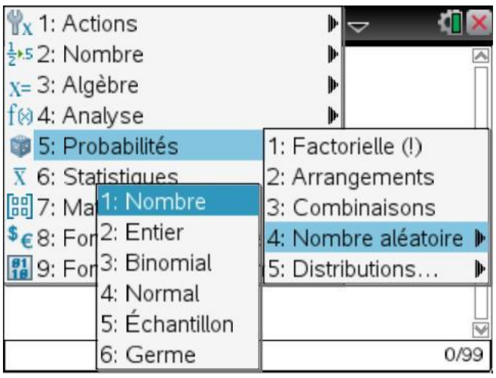


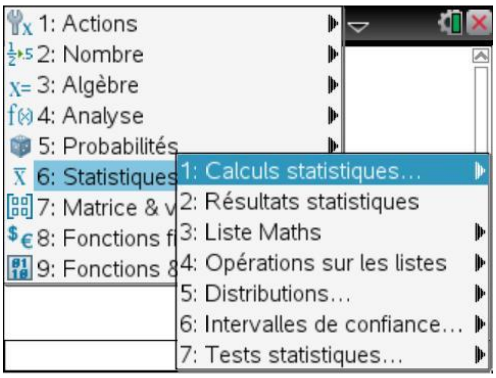




☞ *La syntaxe de la fonction sélectionnée se trouve affichée au bas gauche de l'écran.*

Dans l'application Calculs, on accède à ces fonctions dans les menus **Probabilité** et **Statistiques** (touches [menu] [5] ou [6]) :

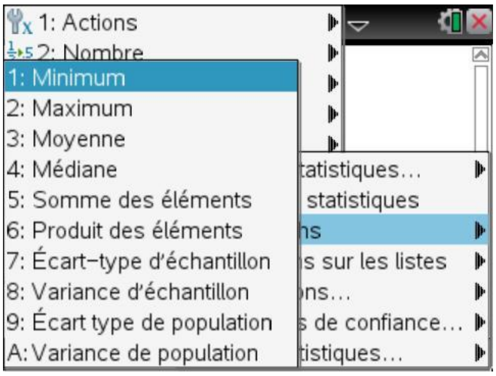




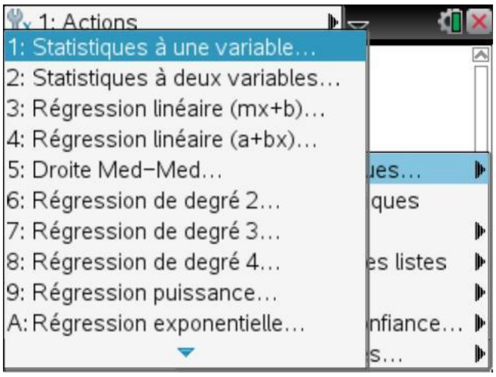
☞ *Le symbole ! permettant le calcul des factorielles est disponible à l'aide de la touche* [?!▸] *et également dans la palette de symboles* ([ctrl] [∞β°]) *ligne 4.*

Le nombre C_n^p, que l'on note $\begin{pmatrix} n \\ p \end{pmatrix}$, se calcule par **nCr**$(n, p)$, A_n^p par **nPr**(n, p) (resp. [menu] [5] [3] et [menu] [5] [2]).

Vous trouverez d'autres fonctions utiles : calcul de la moyenne (**mean**), du maximum, du minimum, de la variance, ou encore de l'écart type des éléments d'une liste dans le menu **Liste Maths** accessible à partir du menu **Statistiques** (menu 6 3).



Certaines fonctions statistiques sont également utilisables à partir de l'application Tableur & listes, on accède à ces fonctions dans le menu **Statistiques/Calculs statistiques** (touches menu 4 1).



1.2 *Remarque importante concernant le calcul de la variance et de l'écart type*

Attention, il existe deux fonctions pour le calcul de l'écart type : **stdDevPop** et **stdDevSamp** et deux fonctions pour le calcul de la variance **varPop** et **varSamp**. L'une des deux ne donnant pas le résultat classiquement attendu en classes préparatoires.

La formule "usuelle" de calcul de la variance de la liste $\{x_1, x_2, \dots, x_n\}$ est :

$$v = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}$$

Elle est donnée par la fonction **varPop**.

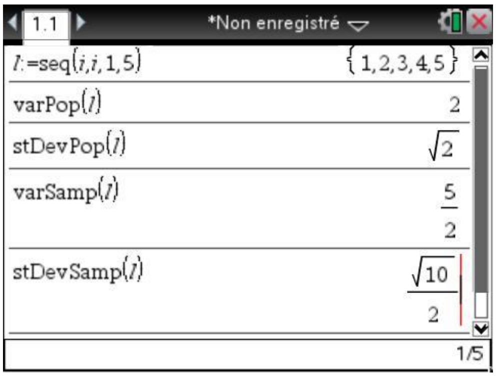
La fonction **varSamp** effectue le calcul à partir de la relation :

$$v = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n-1}$$

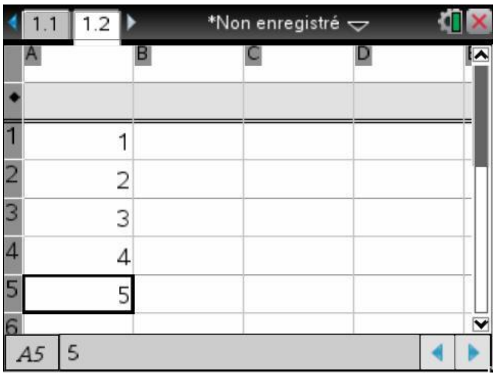
qui donne un estimateur sans biais de σ^2, variance de la population d'où est tiré l'échantillon $\{x_1, x_2, \dots, x_n\}$ (voir page 20).

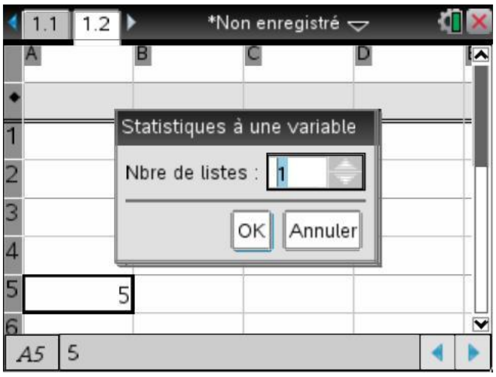
On utilise donc **varPop** et **stdDevPop** lorsque $\{x_1, x_2, \ldots, x_n\}$ représente la population entière, et **varSamp** et **stdDevSamp** lorsque $\{x_1, x_2, \ldots, x_n\}$ est un échantillon tiré d'une population dont on veut estimer les paramètres.

Prenons par exemple la liste $\{1,2,3,4,5\}$. La moyenne est 3. La variance est la moyenne des valeurs $\{(-2)^2, (-1)^2, 0, 1^2, 2^2\}$, c'est à dire $10/5 = 2$. L'écart type est égal à $\sqrt{2}$.

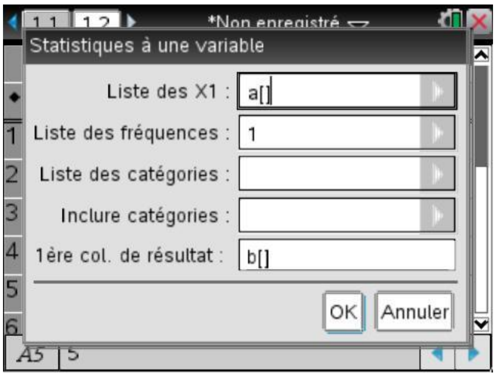


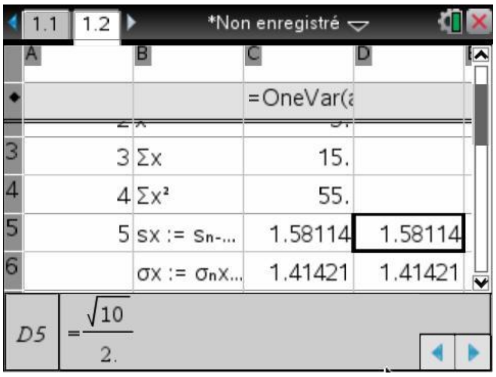
La situation est identique dans l'application Tableur & listes. La valeur affichée **Sx** correspond à la valeur $\sqrt{10}/2$, et la valeur de **σx** correspond à la valeur $\sqrt{2}$. Ajoutez une page (ctrl I), choisissez l'application Tableur & listes (menu 4). Entrez ensuite la liste dans la première colonne, puis ouvrez le menu **Statistiques à une variable** (menu 4 1 1)





Validez en cliquant sur OK.





Validez en cliquant sur OK, les résultats s'affichent.

Il est possible de redimensionner les colonnes afin de les rendre plus lisibles. Sélectionnez une cellule de la colonne à élargir, puis choisissez **Redimensionner, Largeur des colonnes** dans le menu contextuel accessible par `ctrl` `menu`. Déplacez ensuite la limite droite à l'aide du curseur validez par `enter`.
Les fonctions statistiques travaillent ici en mode approché, alors que les fonctions **varPop**, **stdDevPop**, **varSamp**, **stdDevSamp** font des calculs "exacts".

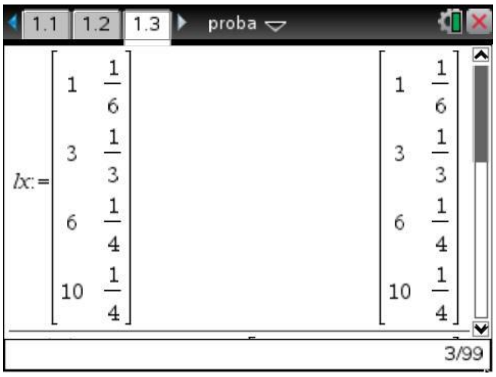
2. L'écriture de quelques fonctions utiles

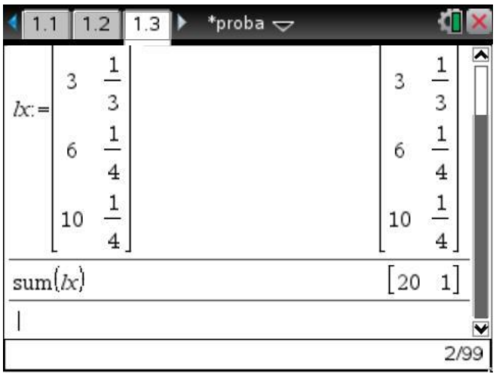
Une bibliothèque de programmes **proba.tns** facilite certains calculs classiques, elle est jointe à ce document.

2.1 *Tableau de calcul, espérance, variance, écart type d'une variable aléatoire discrète*

Considérons par exemple la variable aléatoire définie par $X(\Omega)=\{1,3,6,10\}$, $p(X=1)=\frac{1}{6}$, $p(X=3)=\frac{1}{3}$, $p(X=6)=\frac{1}{4}$, $p(X=10)=\frac{1}{4}$.

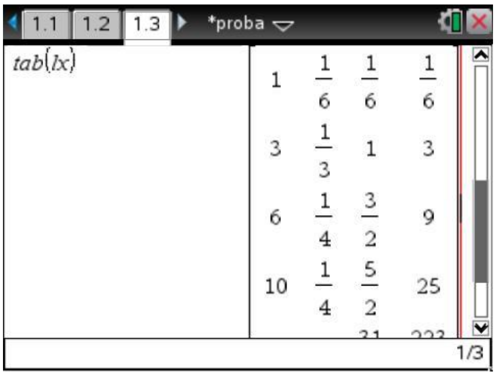
Nous allons placer les éléments définissant cette variable aléatoire dans une matrice **lx**. La première colonne contiendra les valeurs x_i, la seconde les probabilités p_i.

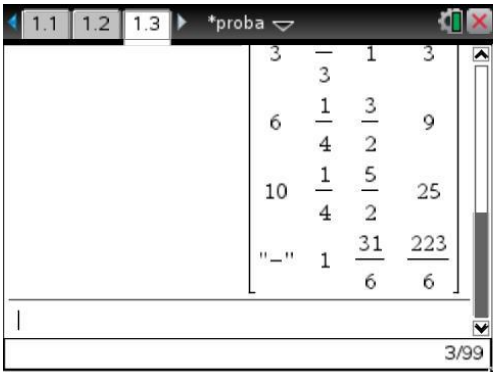




L'utilisation de **sum** (`menu` `6` `3` `5`) faite dans le deuxième écran montre que la somme des probabilités est égale à 1. Jusque-là, tout va bien. Construisons à présent la matrice contenant également la colonne formée par les $p_i x_i$ et celle formée par les $p_i x_i^2$.

Pour cela on peut utiliser une fonction **tab** dont vous trouverez la définition page 6 :



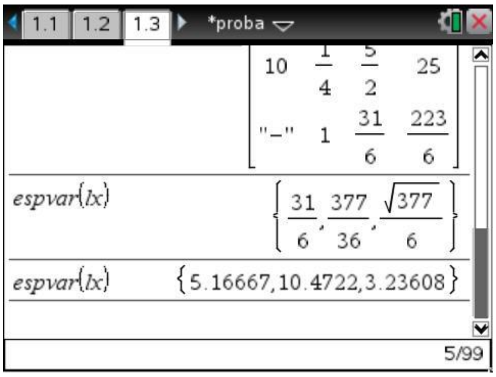


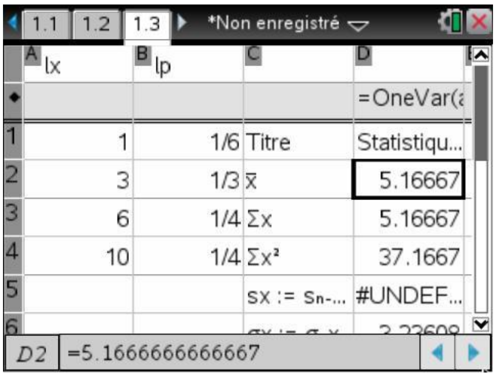
Une ligne contenant la somme des termes de chaque colonne a été ajoutée à la matrice. Cela peut faciliter la construction du tableau de calcul de l'espérance et de la variance.

On peut, par exemple, y lire que $E(X)=\dfrac{31}{6}$ et $E(X^2)=\dfrac{223}{6}$.

Pour calculer ces deux dernières, il suffit d'utiliser la fonction **espvar**, que vous trouverez également au paragraphe suivant, qui utilise la fonction précédente pour retourner une liste formée par l'espérance, la variance, et l'écart type. Vous pouvez retrouver ces résultats à l'aide de l'application Tableur & listes, entrez dans la colonne A la liste $\{1,3,6,10\}$, dans la colonne B, la liste $\left\{\dfrac{1}{6},\dfrac{1}{3},\dfrac{1}{4},\dfrac{1}{4}\right\}$.

[menu][4][1][1], nombre de listes : 1, Liste des x1 : a[ ], Liste des fréquences : b[ ], on valide (voir écran ci-dessous à gauche).





2.2 *Texte des fonctions tab et espvar*

L'écriture de ces deux fonctions utilise des fonctions de calcul matriciel, ce qui en fait l'intérêt. Si vous n'êtes pas familiarisé avec ces dernières, cela risque de vous paraître un peu mystérieux. Voici quelques explications permettant d'en suivre le fonctionnement. Ces explications sont beaucoup plus longues que la fonction **tab** !

☞ *Ces calculs peuvent être également faits de façon très simple dans l'application Tableur & listes.*

1. **xᵀ[i]** forme le vecteur ligne obtenu à partir de la *i*-ième colonne de la matrice **x** passée en argument. Cette matrice comporte les valeurs de x_i sur la première colonne, et celle de p_i dans la seconde.
2. La fonction **mat▸list** permet de convertir ce vecteur ligne en liste. On obtient les listes **lx** et **lp** des valeurs et probabilités à utiliser pour les calculs suivants.
3. À partir des quatre listes, **lx**, **lp**, **lx*lp**, **lx^2*lp**, il est possible de construire une matrice de quatre lignes contenant les valeurs de x_i, de p_i, de $p_i x_i$ et de $p_i x_i^2$.
4. En la transposant, on obtient la matrice **m** représentant le tableau dans sa présentation classique avec ses quatre colonnes.
5. La fonction **sum** permet de faire la somme de chacune de ces colonnes et d'obtenir la matrice $\begin{bmatrix}\sum_{i=1}^{n} x_i & \sum_{i=1}^{n} p_i & \sum_{i=1}^{n} p_i x_i & \sum_{i=1}^{n} p_i x_i^2\end{bmatrix}$.
6. On "empile" ensuite cette matrice et la matrice **m** pour former le tableau dont la dernière ligne comporte les sommes de chaque colonne. Cela se fait en utilisant la fonction **colAugment**.
7. Enfin, on modifie le terme situé sur la première colonne de la dernière ligne. Il correspond au cumul des valeurs de x_i, et n'est pas utile pour la suite. On utilise la fonction **rowDim** pour connaître le nombre de lignes de la matrice.

☞ *L'opérateur de transposition* ᵀ *s'obtient dans le menu* [📖][2] **Matrice.**

```
Define LibPub tab(mat)=Func
©mat:tableau de calcul proba/stat
Local  lx,lp,u,v,m
lx:=mat▸list(matᵀ[1])
lp:=mat▸list(matᵀ [2])
m:=({lx,lp,lx*lp,lx^2*lp})ᵀ
m:=colAugment(m,sum(m))
m[rowDim(m),1]:="−"
m
EndFunc
```

☞ *L'utilisation de la syntaxe*

```
Define LibPub… =
```

permet de faire apparaître cette fonction dans le catalogue, voir **chapitre 15**.

La fonction **espvar** est beaucoup plus simple à comprendre.

On pourrait l'écrire directement, mais on peut aussi utiliser la fonction **tab** précédente. Il suffit de construire la matrice précédente et d'aller y chercher les informations utiles, c'est-à-dire les valeurs de $\sum_{i=1}^{n} p_i x_i$ et de $\sum_{i=1}^{n} p_i x_i^2$, désignées par **ex** et **ex2** dans cette fonction.

```
Define LibPub espvar(x)=Func
©mat:calcul de esp et var
Local  m,k,ex,ex2,v
m:= tab(x)
k:=rowDim(m)
ex:=m[k,3]
ex2:=m[k,4]
v:=ex2−ex^2
{ex,v, √ (v)}
EndFunc
```

3. Utilisation en Statistique : régression linéaire

La statistique suivante donne l'évolution des stocks d'une entreprise :

Année	2000	2001	2002	2003	2004	2005
Stock Q_i	6 400	7 200	8 700	10 400	12 600	15 000

Effectuer un ajustement à l'aide d'une régression linéaire, puis un ajustement à l'aide d'une fonction exponentielle. Que peut-on en conclure ? Donner une estimation du stock en 2009.

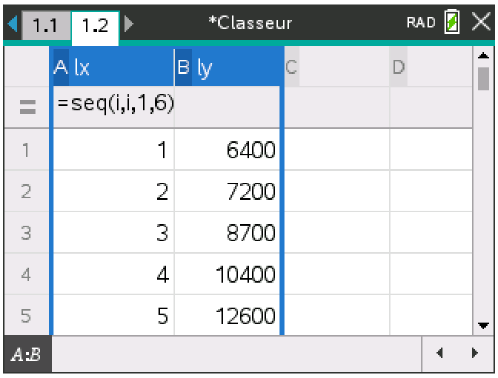
Solution

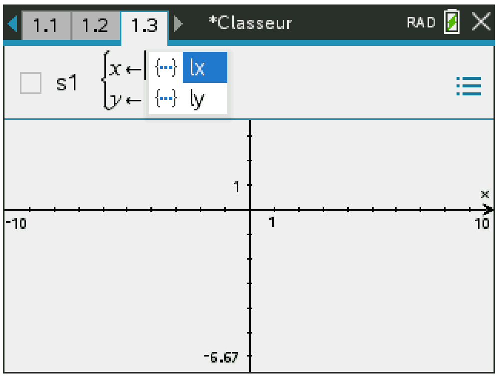
On entre dans l'éditeur la liste des numéros des années et la liste des Q_i.

$\mathrm{lx} = \{1,2,3,4,5,6\}$, $\mathrm{ly} = \{6400,7200,8700,10400,12600,15000\}$.

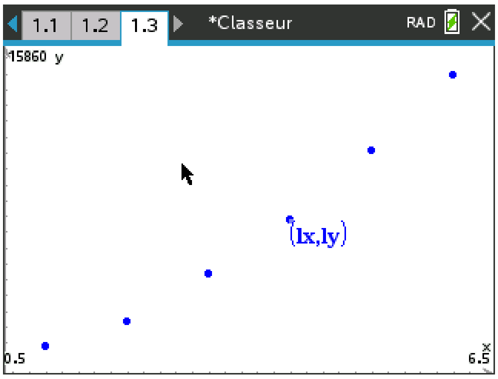
Pour représenter le nuage de points, on peut le faire sur la même page à l'aide de **Graphe rapide** que vous trouverez dans le menu contextuel accessible par [ctrl][menu], choisir [6] Données.

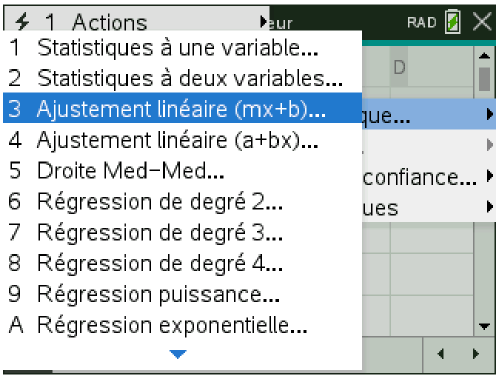
On peut aussi ajouter une page Graphiques en utilisant [ctrl][I]. Dans la ligne de saisie, sélectionner **Nuage de points** dans le menu contextuel ([ctrl][menu][1][6]). Il suffit ensuite d'entrer les deux listes : appuyer sur [ctrl][L], choisir la liste et valider, [tab] pour passer à l'autre liste et recommencer.

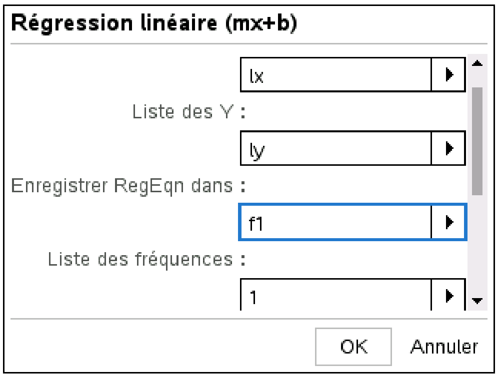


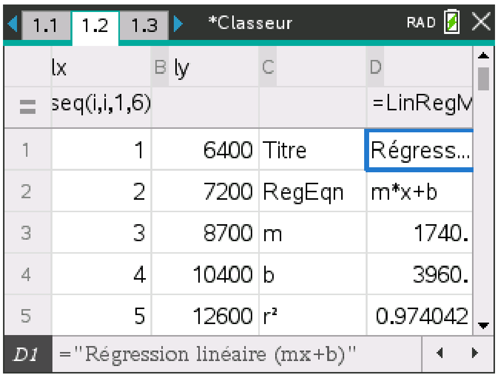


On valide, esc, pour sortir de la ligne de saisie, ctrl menu 4 9 pour choisir un zoom adapté aux données. Effectuons une régression linéaire, ctrl ◀ pour revenir à la page précédente, puis menu 4 1 3.





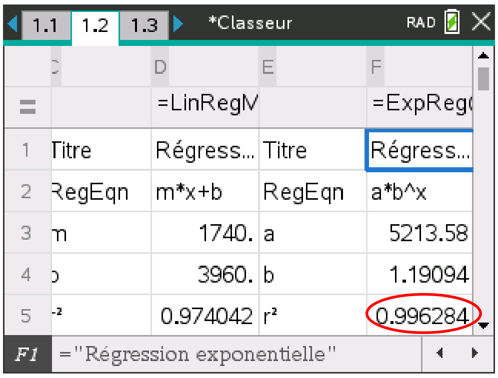


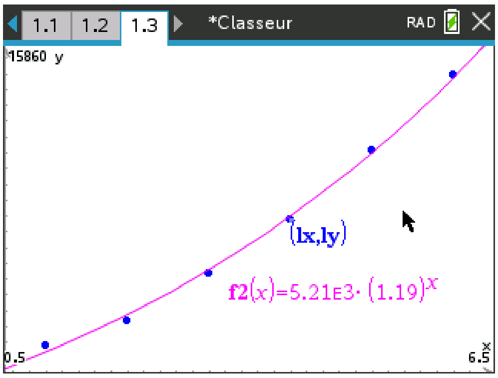


L'ajustement linéaire donne un coefficient de corrélation $r \simeq 0{,}987$.

☞ *La troisième rubrique* **Enregistrer RegEqn dans :** *permet d'indiquer le nom de la variable dans laquelle sera mémorisée l'équation de la fonction servant à effectuer l'ajustement* (f1 *pour l'ajustement linéaire et* f2 *pour l'exponentiel*).

menu 4 1 A. L'ajustement par une fonction de la forme $Q = a * b^t$ donne un meilleur résultat comme l'indique la représentation graphique et le coefficient de corrélation $r \simeq 0{,}998$.





☞ *Pour afficher la courbe, il faut utiliser de nouveau le menu contextuel (ctrl menu 1 1), après avoir ouvert la ligne de saisie (ctrl G).*

Pour avoir l'estimation du stock en 2009, il suffit de calculer f2(10), ce qui donne approximativement 29 926, alors que l'ajustement linéaire aurait donné 21 360 (f1(10)).

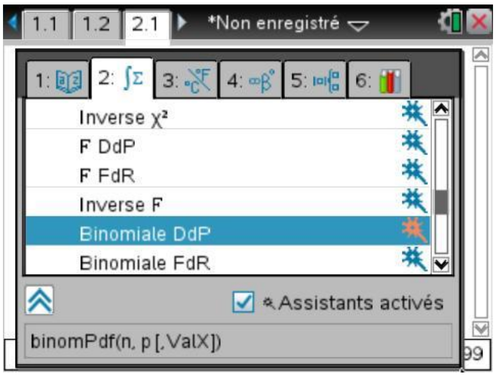
4. Lois discrètes usuelles

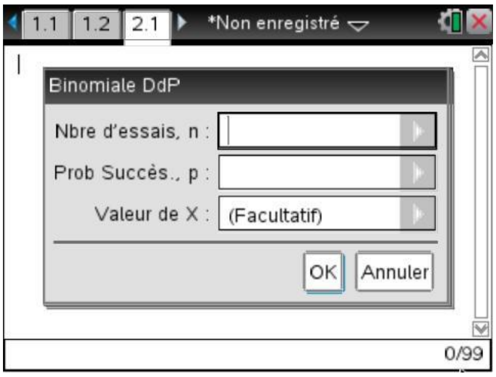
4.1 *Les différentes fonctions présentes sur la TI-Nspire*

Les lois discrètes usuelles : loi binomiale, loi géométrique, loi de Poisson sont directement intégrées. Il sera de plus très simple de définir d'autres lois lorsque vous en aurez l'utilité, voir par exemple la loi hypergéométrique.

Vous trouverez les fonctions correspondantes dans le menu **Distributions** du menu **Statistiques** (touches menu 6 5) de l'application Calculs, et dans le menu **Distributions** du menu **Statistiques** (touches menu 4 2) de l'application Tableur & listes.
Vous aurez aussi accès à ces fonctions dans le catalogue (📖) page 2, dans les rubriques : **Probabilité**, et **Statistiques** sous-menu **Distributions**.

On peut soit utiliser l'assistant (voir écrans ci-dessous), soit entrer directement les paramètres.





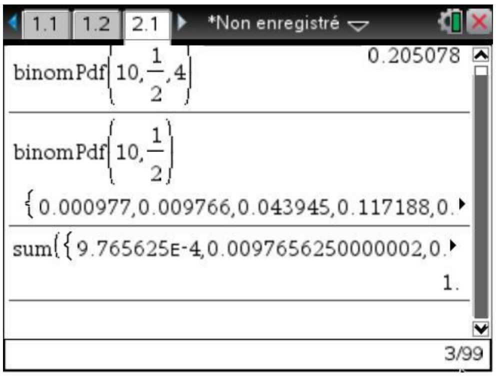
- Loi binomiale de paramètres n et p, $n \in \mathbb{N}^*$, $p \in]0,1[$

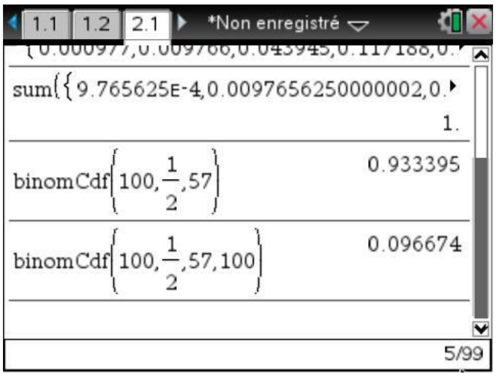
On a $p(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$ pour tout $k \in [\![0,n]\!]$.

syntaxe : **binomPdf**(n, p [, k])[1] donne $P(X=k)$

fonction de répartition : **binomCdf**(n, p [, k]) donne $P(X\leq k)$ et **binomCdf** (n, p, a, b) donne $P(a\leq X\leq b)$.

Il est possible de calculer par exemple la probabilité d'avoir 4 “pile” lorsqu'on lance 10 fois une pièce de monnaie, et de vérifier que la somme des probabilités lorsque k varie de 0 à n est bien égale à 1. Dans le second écran, on calcule la probabilité d'obtenir au plus 57 “pile” au cours de 100 lancers ($\simeq 0{,}93$) et la probabilité d'obtenir au moins 57 “pile” au cours de 100 lancers ($\simeq 0{,}097$).





- Loi géométrique de paramètre p ($p\in]0,1[$)

Soit une expérience élémentaire dont l'issue est un succès ou un échec avec des probabilités respectives p et $q=1-p$. On renouvelle cette expérience jusqu'à l'obtention d'un succès. Cette loi caractérise le nombre d'expériences nécessaires pour obtenir le premier succès. $P(X=k)=p(1-p)^{k-1}$, $\quad k\in\mathbb{N}^*$

$E(X)=\dfrac{1}{p}$ et $V(X)=\dfrac{q}{p^2}$

syntaxe : **geomPdf**(p, k)

fonction de répartition : **geomCdf**(p,[a,] b) donne $P(a\leq X\leq b)$, (par défaut a est égal à 1).

La fonction **geomPdf** permet par exemple de calculer la probabilité d'obtenir un “6” ou bout de 3 lancers d'un dé (non pipé). La fonction **geomCdf** permet dans l'exemple ci-dessous de calculer la probabilité d'obtenir un “6” en au plus 10 lancers, puis en au plus 20 lancers et au moins 10.

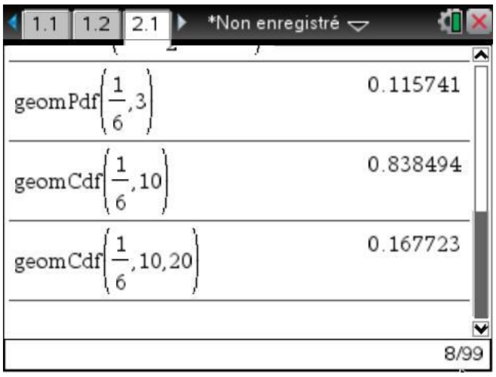
Remarque : pour calculer la probabilité $P(X>k)$, on peut utiliser la formule :

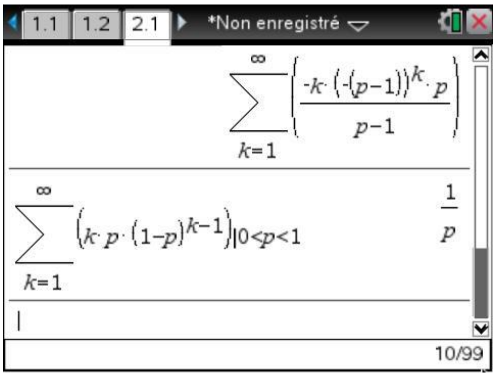
$$P(X>k)=1-P(X\leq k)=1-\textbf{geomCdf}\ \ p,k\ \ .$$

Le deuxième écran montre le calcul de l'espérance mathématique de la loi géométrique de paramètre p, l'utilisation de **geomPdf** ne permet pas ici d'obtenir le résultat, les valeurs données par cette fonction sont des valeurs approchées, il faut donc entrer la 0 ci-dessus valeur formelle de la probabilité et de plus préciser que $0<p<1$ à l'aide de l'opérateur “sachant que” |. On peut calculer de même $E(X^2)$ et en déduire la variance :

$$E(X)=\frac{1}{p}\quad\text{et}\quad V(X)=\frac{1-p}{p^2}.$$

[1] Si k est omis, **binomPdf**(n, p) (resp. **binomCdf**(n, p)) donne la liste des probabilités $P\left(X=k\right)$ (resp. $P\left(X\leq k\right)$) pour $k\in [\![0,n]\!]$.



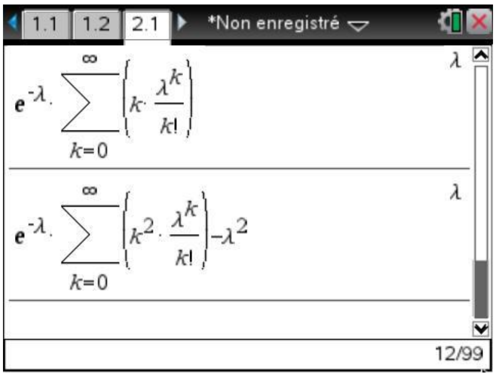


- Loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$

  syntaxe : **poissPdf**(λ, k) donne $P(X=k)=\mathrm{e}^{-\lambda}\dfrac{\lambda^k}{k!}$, pour $k\in\mathbb{N}$

  fonction de répartition : **poissCdf**(λ,[a,] b) donne $P(a \leq X \leq b)$. Par défaut a est égal à 0.

On peut vérifier que l'espérance mathématique et la variance sont égales au paramètre λ.



4.2 *Ajout de fonctions*

Comme on l'a vu dans les exemples ci-dessus les fonctions considérées donnent des résultats en valeur approchée. Si vous avez besoin de résultats formels sur une TI-Nspire CAS, comme on le voit dans les deux derniers exemples, vous pouvez réécrire ces fonctions comme ci-dessous pour la loi hypergéométrique. Vous retrouverez ces fonctions dans la bibliothèque **proba** (fichier proba.tns téléchargeable avec ce chapitre).

- Loi hypergéométrique N, n et p :

```
Define LibPub lhyp(m,n,p,k)=nCr(m*p,k)*nCr(m*(1−p),n−k)/nCr(m,n)
Define LibPub fhyp(m,n,p,k)=∑(nCr(m*p,i)*nCr(m*(1−p),n−i),i,0,k)/nCr(m,n)
```

☞ *On a utilisé* **m** *car* **N** *et* **n** *sont interprétés de la même façon par l'unité nomade TI-Nspire. De plus, dans la définition des fonctions de répartition, on a "manuellement" fait la mise en facteurs nécessaire pour diminuer le temps nécessaire au calcul.*

Voici les définitions des fonctions donnant la loi de probabilité (l) et la fonction de répartition (f) des trois lois usuelles précédentes : binomiale, géométrique et Poisson :

```
Define LibPub lbinom(n,p,k)=nCr(n,k)*p^k*(1−p)^(n−k)
Define LibPub fbinom(n,p,k)=∑(lbinom(n,p,i),i,0,k)
Define LibPub lgeom(p,k)=p*(1−p)^(k−1)
```

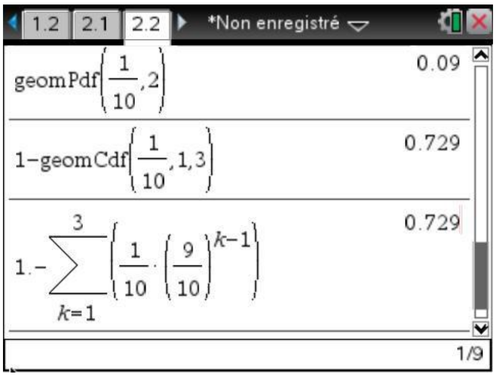
```
Define LibPub fgeom(p,k)=∑(lgeom(p,i),i,1,k)
Define LibPub lPoisson(λ,k)=e^(-λ)*λ^k/(k!)
Define LibPub fPoisson(λ,k)= ∑( λ^i/(i!),i,0,k)*e^(-λ)
```

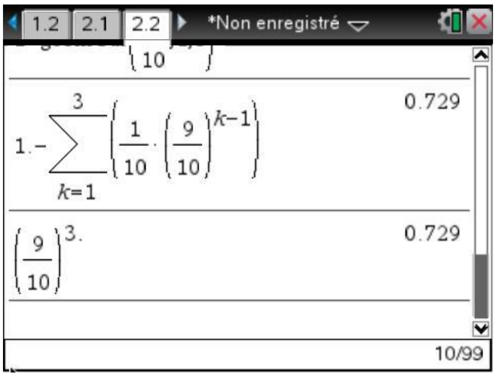
4.3 *Exemple de calcul utilisant les lois géométriques*

On considère une variable aléatoire X suivant une loi géométrique de paramètre $p=1/10$.

Nous allons calculer les probabilités $P(X=2)$ et $P(X>3)$, puis calculer l'espérance et la variance. Cela ne pose aucun problème en utilisant les fonctions décrites dans ce chapitre.

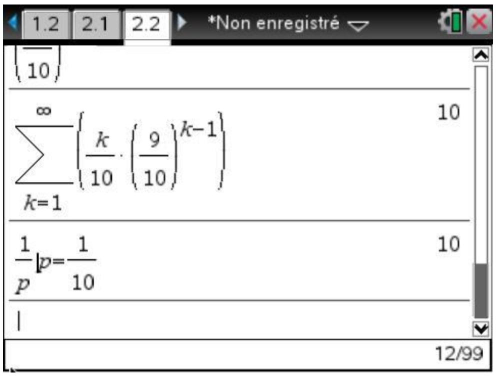
1. Calcul de $P(X=2)$ et de $P(X>3)=1-P(X\leq 3)$ (avec vérification).





☞ *(On a également fait un calcul direct de cette probabilité : $X>3$, ou encore $X\geq 4$ signifie que les trois premiers essais ont été des échecs.)*

3. Calcul de l'espérance.



4. Calcul de la variance et vérification.