Travail de l'élève 1 : En Syldavie, le sport national des éléphants est le saut en parachute. Malheureusement pour les crocodiles, il arrive qu'ils meurent écrasés par un éléphant à son atterrissage, pas toujours contrôlé...
Les crocodiles savent que les éléphants gauchers, tout comme les éléphants daltoniens présentent plus de risque pour leur survie.
Or 46% des éléphants pratiquant le saut en parachute sont des gauchers et 18% sont daltoniens.
Aussi, lorsqu'un club vient à ouvrir près de chez eux, ils se demandent toujours si la fréquence de gauchers et celle de daltoniens dans ce club est anormale (pour éventuellement demander sa fermeture).

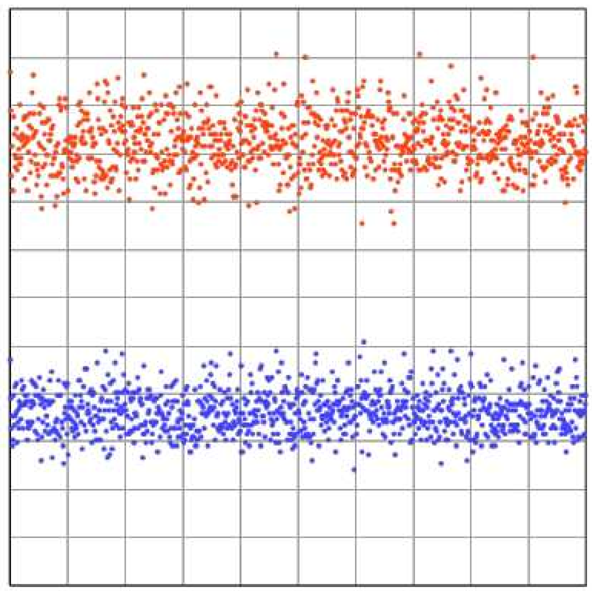
PARTIE A. **Fluctuation d'échantillonnage**

A Gataca, un club de saut en parachute pour éléphants de 324 membres vient d'ouvrir.
L'échantillon de 324 éléphants peut être assimilé à un tirage aléatoire avec remise, au vu du grand nombre d'éléphants pratiquant le saut en parachute en Syldavie.

1. Intuitivement, quelles fréquences de gauchers considéreriez-vous « normale » ? de daltoniens ?

2. On donne le programme suivant et un exemple d'affichage en sortie.

```
VARIABLES
  a EST_DU_TYPE NOMBRE
  compteur EST_DU_TYPE NOMBRE
  compteur2 EST_DU_TYPE NOMBRE
  Daltoniens EST_DU_TYPE NOMBRE
  Gauchers EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  POUR compteur ALLANT_DE 1 A 1000
    DEBUT_POUR
    Daltoniens PREND_LA_VALEUR 0
    Gauchers PREND_LA_VALEUR 0
    POUR compteur2 ALLANT_DE 1 A 324
      DEBUT_POUR
      a PREND_LA_VALEUR random()
      SI (a<=0.46) ALORS
        DEBUT_SI
        Gauchers PREND_LA_VALEUR Gauchers+1
        FIN_SI
      SI (a<=0.18) ALORS
        DEBUT_SI
        Daltoniens PREND_LA_VALEUR Daltoniens+1
        FIN_SI
      FIN_POUR
    TRACER_POINT (compteur,Gauchers/324)
    TRACER_POINT (compteur,Daltoniens/324)
    FIN_POUR
FIN_ALGORITHME
```



Xmin: 0 ; Xmax: 1000 ; Ymin: 0 ; Ymax: 0.6 ; GradX: 100 ; GradY: 0.05

a. Le programmer sur Algobox (régler la même échelle).

b. Que fait-il ?

c. Qu'observe-t-on ?

d. Placer les informations relatives à l'échelle sur les axes du graphique donné.

e. A votre avis, pour quelles fréquences de gauchers peut-on considérer « normale » ? de daltoniens ?

f. Votre critère vous semble-t-il assez convainquant pour demander ou non la fermeture du club ?

PARTIE B. **Intervalle de fluctuation de seconde et critère de décision**

Norbert, fervent militant de la défense des crocodiles, cherche à savoir s'il peut considérer les fréquences d'apparition des deux caractères dans le club comme exceptionnelles ou non.
Il émet donc l'hypothèse que la probabilité qu'un éléphant du club soit gaucher est $p_1 = 0.46$ et celle qu'il soit daltonien est $p_2 = 0.18$ (comme dans la population totale d'éléphants pratiquant le saut en parachute).
Puis il dénombre dans le club les gauchers et les daltoniens : il y a 165 gauchers et 252 non daltoniens.

Rappel de seconde

Pour un caractère donné, on note p sa probabilité d'apparition et f sa fréquence d'apparition dans un échantillon.
La fréquence f se situe dans au moins 95% des cas dans l'intervalle

$$\left[\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\right]$$

appelé **intervalle de fluctuation au seuil de** 95%.

- **Avantages** :

- **Inconvénients** :

Sans parler du fait que cet intervalle n'a jamais été justifié par votre enseignant ...

Remarques :

– Il s'agit d'un intervalle dans lequel se situent au moins 95% ...

Cet intervalle dépend évidemment de
 –
 –

– Dans la pratique, on connait …… et on peut faire en sorte que notre échantillon soit établi ...
Par contre, pour ……, à moins d'avoir regarder la population entière (et dans ce cas, ce chapitre n'a plus d'intérêt) il s'agit d'une hypothèse.
Cet intervalle de fluctuation servira donc surtout à vérifier ...

1. Déterminer l'intervalle de fluctuation au seuil de 95% du critère « gaucher » dans la population des éléphants pratiquant le saut en parachute.

2. Peut-on ou non réfuter l'hypothèse faite sur les gauchers ? Argumenter.

3. Et pour le critère « daltonien » ?

Critère de décision

Pour une étude statistique d'un caractère connu de la population (âge, sexe, taille, etc), on considère qu'une hypothèse sur une proportion p d'apparition de ce caractère dans la population est crédible (ou qu'un échantillon est **représentatif** lorsqu'on connaît p), si

Remarques :

- Si la fréquence observée n'appartient pas à l'intervalle de fluctuation au seuil de 95%, on considérera que l'hypothèse de départ sur p est ...

  Dans ce cas, on a un risque de se tromper dans 5%, puisque 5% des fréquences établies dans ces conditions ne sont pas dans cet intervalle. On peut donc refuser l'hypothèse pour 5% d'échantillons « en trop ».
  Par contre, si f est dans l'intervalle de fluctuation au seuil de 95%, on considérera l'hypothèse sur p est ...

  Dans ce cas, on ne connait pas le risque d'erreur, ie le risque d'accepter une hypothèse « en trop » (à partir d'échantillons établis dans d'autres conditions).
  Dans tous les cas, **on est sûr de rien** ! Donc on évitera le vocabulaire « vrai » ou « faux ».
- Tout est une question d'équilibre :
  Si l'on veut diminuer l'erreur de rejet, par exemple en prenant un intervalle de fluctuation au seuil de 100%, on ne rejettera donc aucune hypothèse « en trop », par contre, on les acceptera tous, donc évidemment, beaucoup trop. Toute hypothèse sur p semblera crédible, ce qui n'a aucun intérêt.
  Ainsi, dans la pratique, on utilise surtout les seuils de 95% et de 99%.

PARTIE C. **Un nouvel intervalle de fluctuation**

1. Si l'on choisit au hasard un échantillon de 324 éléphants pratiquant le saut en parachute, quelle loi suit la variable aléatoire X qui dénombre le nombre d'éléphants daltoniens ? *Justifier.*

2. Sur tableur, construire :
   a. Dans la colonne A la listes des valeurs k possibles pour X.

   b. Dans la colonne B la liste des probabilités P(X = k) correspondantes.
   Utiliser `=LOI.BINOMIALE( k ; n ; p ; 0 )` *où k est la valeur d'une cellule.*

   c. Le diagramme en bâton correspondant aux colonnes A et B.
   Dans « Plage de données », cocher les deux cases.

   d. Dans la colonne C les probabilités cumulées correspondantes. **Appeler le professeur pour valider !**
3. En observant le diagramme, proposer une **méthode** pour établir un nouvel intervalle de fluctuation à 95% des fréquences $\dfrac{X}{n}$.

4. Grâce à la colonne C, déterminer précisément cet intervalle.

5. Interpréter cet intervalle et conclure.

6. A partir de combien de daltoniens peut-on considérer que la situation dans le club est anormale avec cette méthode de prise de décision ?

7. Avec cette méthode, donner l'intervalle de fluctuation au seuil de 95% de la fréquence des gauchers. *Préciser la loi de la variable aléatoire* Y *qui compte le nombre de gauchers.*

8. Commenter.

Pour résumer

On appelle X la variable aléatoire qui compte le nombre de personnes d'un échantillon de taille n, possèdant un certain caratère, de probabilité d'apparition p. On a donc $\mathrm{X} \hookrightarrow \mathrm{B}(n, p)$

On note $f = \dfrac{\mathrm{X}}{n}$ la fréquence d'apparition de caractère.

La fréquence f se situe dans au moins 95% des cas dans l'intervalle $\left[\ldots\ldots\ldots;\ldots\ldots\ldots\right]$, où :

- a désigne le plus petit entier k à partir duquel $\mathrm{P}(\mathrm{X} \leq k)$ dépasse strictement ....
- b désigne le plus petit entier k à partir duquel $\mathrm{P}(\mathrm{X} \leq k)$ dépasse ....

- **Avantages** :

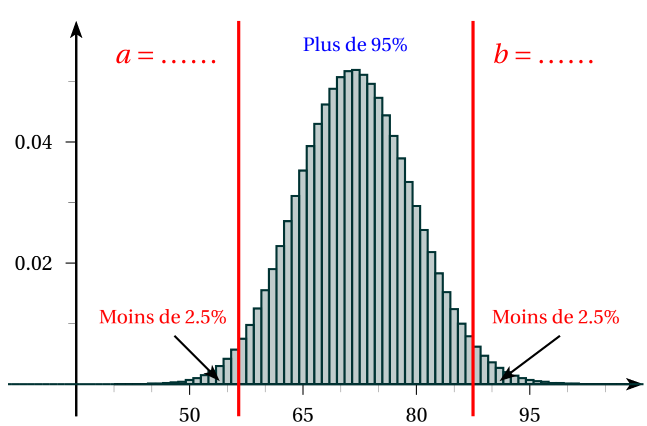
  Sans parler du fait que cet intervalle est justifié par vos connaissances sur la loi binomiale ...

- **Inconvénients** :

Ici $\mathrm{X} \hookrightarrow \mathrm{B}(400; 0.18)$.

k	$\mathrm{P}(\mathrm{X} \leq k)$
...	...
54	0.00953
55	0.01375
56	0.01944
57	0.02699
58	0.03678
59	0.04924
...	...

k	$\mathrm{P}(\mathrm{X} \leq k)$
...	...
83	0.94587
84	0.94587
85	0.95825
86	0.96821
87	0.97609
88	0.98225
...	...



Exemple :

On admet que dans la population d'enfants de 11 à 14 ans d'un département français le pourcentage d'enfants ayant déjà eu une crise d'asthme dans leur vie est de 13%.
Un médecin d'une ville de ce département est surpris du nombre important d'enfants le consultant ayant des crises d'asthme et en informe les services sanitaires. Ceux–ci décident d'entreprendre une étude et d'évaluer la proportion d'enfants de 11 à 14 ans ayant déjà eu des crises d'asthme.
Ils sélectionnent de manière aléatoire 100 jeunes de 11 à 14 ans de la ville.
La règle de décision prise est la suivante : si la proportion observée est supérieure à la borne supérieure de l'intervalle de fluctuation asymptotique au seuil de 95% alors une investigation plus complète sera mise en place afin de rechercher les facteurs de risque pouvant expliquer cette proportion élevée.

1. Déterminer l'intervalle de fluctuation asymptotique au seuil de 95% de la proportion de jeunes de 11 à 14 ans ayant eu une crise d'asthme dans un échantillon de taille 100.
2. L'étude réalisée auprès des 100 personnes a dénombré 19 jeunes ayant déjà eu des crises d'asthme. Que pouvez-vous conclure ?
3. Le médecin n'est pas convaincu par cette conclusion et déclare que le nombre de personnes interrogées était insuffisant pour mettre en évidence qu'il y avait plus de jeunes ayant eu des crises d'asthme que dans le reste du département.

   Combien faudrait-il prendre de sujets pour qu'une proportion observée de 19% soit en dehors de l'intervalle de fluctuation asymptotique ?

Résumé

	Intervalle de **fluctuation** Au seuil 0.95 (p connue)	Intervalle de **confiance** Au seuil 0.95 (p inconnue)
SECONDE	$n \geq 25$, $0.2 \leq p \leq 0.8$ $\mathrm{I} = \left[p - \dfrac{1}{\sqrt{n}}, p + \dfrac{1}{\sqrt{n}}\right]$	Sensibilisation
PREMIERE	Avec la loi binomiale $\mathrm{I} = \left[\dfrac{a}{n}; \dfrac{b}{n}\right]$	

Voici un algorithme qui permet de retrouver le a et le b cherchés ci-dessus, ainsi que sa traduction pour algobox et pour chacun des modèles de vos TI. Ayez l'initiative de le recopier, cela sera bien pratique en exercice !

Algorithme 1 :

Entrée(s) :
n est un entier naturel
p est un réel de l'intervalle $[0;1]$
Variables
s est un nombre réel de l'intervalle $[0,1]$
k est un entier naturel
Début
$k := 0$
$s := \binom{n}{0} p^0 (1-p)^n$
Tant que $(s \leq 0.025)$ **Faire**
$k := k+1$
$s := s + \binom{n}{k} p^k (1-p)^{n-k}$
Fin Tant que
Afficher « $a =$ » et la valeur k
Tant que $(s < 0.975)$ **Faire**
$k := k+1$
$s := s + \binom{n}{k} p^k (1-p)^{n-k}$
Fin Tant que
Afficher « $b =$ » et la valeur k
Fin

```
VARIABLES
  n EST_DU_TYPE NOMBRE
  p EST_DU_TYPE NOMBRE
  s EST_DU_TYPE NOMBRE
  k EST_DU_TYPE NOMBRE
DEBUT_ALGORITHME
  LIRE n
  LIRE p
  k PREND_LA_VALEUR 0
  s PREND_LA_VALEUR ALGOBOX_LOI_BINOMIALE(n,p,0)
  TANT_QUE (s<=0.025) FAIRE
    DEBUT_TANT_QUE
    k PREND_LA_VALEUR k+1
    s PREND_LA_VALEUR s+ALGOBOX_LOI_BINOMIALE(n,p,k)
    FIN_TANT_QUE
  AFFICHER "a ="
  AFFICHER k
  TANT_QUE (s<0.975) FAIRE
    DEBUT_TANT_QUE
    k PREND_LA_VALEUR k+1
    s PREND_LA_VALEUR s+ALGOBOX_LOI_BINOMIALE(n,p,k)
    FIN_TANT_QUE
  AFFICHER "b ="
  AFFICHER k
FIN_ALGORITHME
```

Algorithme 2 : ***TI 82 à 84***

```
PROGRAM:INTERV
:Prompt N
:Prompt P
:0 → K
:BinomFdp(N,P,0) → S
:While S<=0.025
:K+1 → K
:S+BinomFdp(N,P,K) → S
:End
:Disp "A=",K
:While S<=0.975
:K+1 → K
:S+BinomFdp(N,P,K) → S
:End
:Disp "B=",K
```

Algorithme 3 : ***TI 89***

```
:interv(n,p)
:Prgm :0 → k
:tistat.binomddp(n,p,0) → s
:While s<=0.025
:k+1 → k
:s+tistat.binomddp(n,p,k) → s
:EndWhile
:Disp "a= ", k
:While s<=0.975
:k+1 → k
:s+tistat.binomddp(n,p,k) → s
:EndWhile
:Disp "b= ", k
:EndPrgm
```

Algorithme 4 : ***TI Nspire***

Define LibPub **intervfluc**(n,p) =
Prgm
Local s, k
$k := 0$
$s :=$ binomPdf$(n,p,0)$
While $s \leq 0.025$
$k := k+1$
$s := s+$ binomPdf(n,p,k)
EndWhile
Disp "a= ", k
While $s \leq 0.975$
$k := k+1$
$s := s+$ binomPdf(n,p,k)
EndWhile
Disp "b= ", k
EndPrgm

BinomFdp :
$\boxed{\text{2nde}}$ + $\boxed{\text{Vars}}$ + $\boxed{\text{A : binomFdp}}$

tistat.binomddp (binomDdP) :
$\boxed{\text{CATALOG}}$ + $\boxed{\text{F3}}$

binomPdf (Binomiale Ddp) :
$\boxed{\text{catalogue}}$ + $\boxed{2}$

Ouvrir Probabilités puis Distributions