EXERCICES : LOI BINOMIALE

Exercice 1 : **Hasard et QCM**

Objectif : Etudier, pour une situation modélisable par un schéma de Bernoulli, des variables aléatoires qui suivent ou non une loi binomiale.

Un QCM (Questionnaire à Choix Multiples) est composé de 10 questions numérotées de 1 à 10. Pour chacune d'elles, quatre réponses possibles sont proposées, dont une seule est exacte.
La difficulté réside dans le fait que ce QCM Syldave est en chinois, et que notre candidat Fabrice ne lit pas le chinois (bien qu'il le parle couramment, évidemment). Il se voit donc obligé de répondre à chaque question au hasard, de façon indépendante (Fabrice déteste ne pas répondre du tout, il veut tenter sa chance coûte que coûte).

Partie A :

1. **a.** Justifier que la méthode de Farice pour répondre à une question est une épreuve de Bernoulli. Préciser le succès et la probabilité qu'il se réalise.
   b. Que peut-on dire de l'expérience de Fabrice sur le QCM entier ?
2. **Temps d'attente de la première bonne réponse**
   On désigne par X la variable aléatoire donnant le numéro de la première question à laquelle Fabrice répond juste. On convient que X prend la valeur 11 si toutes les réponses sont fausses.
   a. Préciser quelles valeurs peut prendre X.
   b. Calculer $P(X = 11)$ et $P(X = k)$ pour $1 \leq k \leq 10$
   c. X suit-elle une loi binomiale ?
3. **Attribution d'une note**
   On décide de donner à Fabrice un point par réponse exacte. Soit Y la variable aléatoire associant aux réponses de Fabrice sa note obtenue sur 10.
   a. Justifier que Y suit une loi binomiale et en préciser les paramètres.
   b. Sur la calculatrice ou le tableur, obtenir les valeurs arrondies à 10^{-6} de $P(Y = k)$ pour $0 \leq k \leq 10$. Les présenter dans un tableau, puis sous la forme d'un histogramme.
   c. Quelle est la probabilité que Fabrice obtienne la note maximale ?
   d. Quelle est la probabilité qu'il obtienne au moins la moyenne ?
   e. Quelle est la note la plus probabilité de Fabrice ?
   f. Quelle note Fabrice peut-il espérer obtenir (ie quelle note moyenne obtiendrait-il s'il remplissait au hasard un très grand nombre de QCM de ce type)
4. Pour pénaliser les candidats qui ne comptent que sur le hasard comme Fabrice, le gouvernement Syldave décide de toujours accorder 1 point par réponse exacte, mais cette fois d'enlever 0.2 point par réponse fausse.
   a. Prouver qu'avec cette nouvelle règle, la variable aléatoire Z donnant la note obtenue par Fabrice s'exprime par $Z = 1,2Y - 2$
   b. En déduire la probablité que Fabrice obtienne une note négative, puis une note supérieur à 5.
   c. Quelle note Fabrice peut-il espérer obtenir ? L'objectif vous paraît-il atteint ?

Partie B : On suppose que n candidats ($n \in \mathbb{N}^*$) répondent à ce QCM et qu'aucun d'entre eux ne lisant le chinois, ils suivent tous la méthode de Fabrice et sans copier.

1. Quelle est la probabilité P_n qu'au moins un candidat obtienne la note 10 ?
2. Pour quelles valeurs de n cet événement se produira-t-il avec une probabilité supérieur à 0.99 ?

Loi Binomiale à la calculatrice

Pour calculer $P(X = k)$:

- **TI 82-83-84** : Taper $\boxed{\text{2nde}}$ + $\boxed{\text{var}}$ puis descendre avec $\boxed{\blacktriangledown}$ et choisir `0:binomFdp`
  Il s'affiche `binomFdp(` que vous devez compléter par les valeurs `n,p,k)`

- **TI 89** : Taper $\boxed{\text{CATALOG}}$ + $\boxed{\text{F3}}$ puis se déplacer avec $\boxed{\blacktriangledown}$ et $\boxed{\blacktriangle}$ et choisir `binomDdP`
  Il s'affiche `TIStat.binomDdP(` que vous devez compléter par les valeurs `n,p,k)`

- **TI Nspire CX CAS** : Taper $\boxed{\text{CATALOG}}$ + $\boxed{2}$ puis se déplacer avec $\boxed{\blacktriangledown}$ et $\boxed{\blacktriangle}$ et sélectionner `Probabilités` que l'on ouvre avec $\boxed{\text{enter}}$
  Ensuite, sélectionner `Distributions`, que l'on ouvre avec $\boxed{\text{enter}}$, et enfin choisir `Binomiale DdP`
  Il s'affiche `binomPdf()` que vous devez compléter par les valeurs `n,p,k`

- **Casio** : Dans $\boxed{\text{MENU}}$, choisir l'icône `STAT`, puis `DIST>BINM BPD`
  `Numtrial` correspond au paramètre n.

Pour calculer $P(X \leq k)$: Même méthode, il suffit de choisir à la fin :

- **TI 82-83-84** : `A:binomFRép`

- **TI 89** : `binomFdR`

- **TI Nspire CX CAS** : `Binomiale FdR`

- **Casio** : `DIST>BINM ?`
  `Numtrial` correspond au paramètre n.

Exercice 2 : **Méthode du poolage**

Objectif : Etudier une méthode utilisée par exemple pour les tests sanguins

La méthode de poolage est utilisée dans la détection des porteurs d'un parasite au sein d'un ensemble donnée de N individus tirés au sort de façon indépendante, dans une population très vaste par rapport à N (ce qui permet de considérer que le tirage est équivalent à un tirage avec remise).
La proportion de porteurs du parasite dans la population est p $(0 < p < 1)$.
On dispose d'un test permettant de savoir de façon certaine qu'un échantillon de sang contient ou non le parasite, le résultat du test étant dit positif dans le premier cas, négatif dans le second.
Pour chacun des N individus, on possède un prélèvement sanguin.
La méthode de poolage consiste à répartir les N prélèvements par groupes de n prélèvements $(n < \mathrm{N})$.
On mélange les prélèvements des n individus et on teste ces mélanges.
Si le mélange est positif dans un des groupes, on teste le prélèvement de chaque individu du groupe concerné.
La question est de savoir dans quelles conditions le poolage permet d'économiser des tests par rapport au fait de tester chacun des N prélèvements.

Partie A : Etude d'un cas particulier.

On prend $\mathrm{N} = 60$ et on fait 20 groupes de 3, que l'on numérote de 1 à 20.
Pour chaque groupe i (i entier de 1 à 20), on note H_i le mélange des prélèvements des trois individus du groupe.
Soit X la variable aléatoire qui compte le nombre de groupes pour lesquels le test de H_i est négatif.
Soit T la variable aléatoire qui correspond au nombre total de tests effectués.

1. Justifier que la probabilité que le test de H_i soit négatif est $(1-p)^3$.
2. Quelle est la loi de la variable X ? En déduire son espérance mathématique.
3. Prouver que $\mathrm{T} = 80 - 3\mathrm{X}$ et en déduire l'espérance mathématique de T en fonction de p.
4. Le nombre de tests à effectuer étant de 60 lorsque l'on teste les prélèvements de chaque individu, on consièdre que le poolage est rentable si E(T) est inférieur ou égal à 60.
   a. Justifier que $\mathrm{E(T)} \leq 60 \iff \dfrac{1}{3} - (1-p)^3 \leq 0$
   b. Etudier le sens de variation de la fonction g définie sur $[0;1]$ par $g(x) = \dfrac{1}{3} - (1-x)^3$.
   c. En déduire que l'équation $g(x) = 0$ a une unique solution, puis en déterminer une valeur approchée à 0,1 près.
   On pourra utiliser un tableau de valeurs
   d. Conclure en précisant pour quelles valeurs de p la méthode du poolage est rentable.

Partie B : Autre cas particulier

On a toujours $\mathrm{N} = 60$, mais cette fois, on fait 15 groupes de 4, que l'on numérote de 1 à 15.
En s'inspirant de la méthode précédente, préciser pour quelles valeurs de p ce poolage est rentable.

Partie C : Un problème d'optimisation

Si N est le nombre d'individus, on fait $\dfrac{\mathrm{N}}{n}$ groupes de n individus (on suppose que N est assez grand devant n pour négliger le fait qu'un groupe pourrait ne pas être complet).
La démarche et les notations restent celles de la partie A.

1. Déterminer la loi suivie par X.
2. Exprimer T en fonction de X, puis déduire que $\mathrm{E(T)} = \mathrm{N}\left(1 + \dfrac{1}{n} - (1-p)^n\right)$
3. Montrer que E(T) est minimal si et seulement si $1 + \dfrac{1}{n} - (1-p)^n$ est minimale.
4. Pour chacune des valeurs suivantes de p : 0,1 ; 0.01 et 0.001 déterminer à l'aide d'une calculatrice ou d'un tableur la valeur de n telle que E(T) soit minimale.

5. Vérifier dans chacun de ces cas que le poolage est rentable.

Exercice 3 : **Affaire Castaneda contre Partida**

Des arguments de type probabiliste peuvent être avancés et pris en compte dans les cours de justice. En novembre 1976, Rodrigo Partida, d'originie mexicaine, était condamné à huit ans de prison pour vol et tentative de viol dans un comté du sud du Texas. Il attaqua le jugement sous motif que la désignation des jurés dans l'Etat du Texas était discriminatoire pour les Américains d'origine mexicaine. Son argument était que ceux-ci n'étaient pas suffisamment représentés dans les jurys populaires.

Attendu de la Cour Suprême des Etats-Unis (affaire Castaneda contre Partida)

« Si les jurés étaient tirés au hasard dans l'ensemble de la population, le nombre d'américains mexicains dans l'échantillon pourrait alors être modélisé par une distribution binomiale ...

Etant donnée que 79.1% de la population est mexico-américaine, le nombre attendu d'américains mexicains parmi les 870 personnes convoquées en tant que grands jurés pendant la période de 11 ans est approximativement de 688. Le nombre observé est 339.

Bien sûr, dans n'importe quel tirage considéré, une certaine fluctuation par rapport au nombre attendu est prévisible. Le point essentiel, cependant, est que le modèle statistique montre que les résultats d'un tirage au sort tombent vraisemblablement dans le voisinage de la valeur attendue ...

La mesure des fluctuations prévues par rapport à la valeur attendue est l'écart-type, défini pour la distribution binomiale comme la racine carré de la taille de l'échantillon (ici 870) multiplié par la probabilité de sélectionner un américain mexicain (ici 0.791) et par la probabilité de sélectionner un non américain mexicain (ici 0.209) ... Ainsi, dans ce cas, l'écart-type est approximativement de 12.

En règle générale, pour de si grands échantillons, si la différence entre la valeur attendue et le nombre observé est plus grande que deux ou trois écarts-types, alors l'hypothèse que le tirage du jury était au hasard serait suspecte à un spécialiste des sciences humaines.

Les données sur 11 années reflètent ici une différence d'environ 29 écarts-types. Un calcul détaillé revèle qu'un éloignement aussi important de la valeur attendue se produirait avec moins d'une chance sur 10^{140}. »

Source : « *Prove it with Figures (Statistics for Social Science and Behaviour Sciences)* », Hans Zeisel et David Kaye ; Springer (2006)

1. Définir la variable aléatoire X qui, dans cette situation, suit une loi binomiale.
   Quels sont les paramètres de cette loi binomiale ?
2. A quel calcul correspond la valeur 688 ?
3. Effectuer le calcul de l'écart-type de X. A quoi correspond la « différence de 29 écarts-types » ?
4. A quel événement correspond la probabilité de 10^{-140} ? Faites le calcul à l'aide d'un tableur. Etes-vous d'accord ?
5. Peut-on considérer que la consitution des jurys résultait du hasard ?
6. La cour d'appel a donc finalement donné raison à la défense Partida. Cependant, elle exclut une démonstration mathématique de discrimination raciale. N'allez pas croire qu'il y ait eu un complot !
   Quel critique pouvez-vous faire sur la modélisation ?
   A votre avis, qu'est-ce qui peut justifier une telle composition de jury ?

Exercice 4 : Le « président » de Syldavie pense que les filles de son pays sont trop intelligentes et risquent de renverser son pouvoir, pourtant bien établi depuis 30 ans.
Pour limiter le nombre de filles en Syldavie il décide que chaque famille aura au moins un enfant et arrêtera de procréer après la naissance d'un garçon, dans un maximum de 4 enfants par famille.
On considère que chaque enfant autant de chances d'être un garçon qu'une fille, indépendamment du sexe d'éventuel(s) enfant(s) précédent(s).
On se demande si ce choix a la conséquence attendue, à savoir de diminuer le nombre de filles dans la population.

1. On présente l'algorithme suivant :

Algorithme 1 :

```
Variables
    X, NbeEnfants, NbeFilles sont des nombres entiers

Début
    NbeEnfants= 0, X = 0
    Tant que (X == 0 et NbeEnfants< 4) Faire
        Affecter à X un entier aléatoire entre 0 et 1
        NbeEnfants=NbeEnfants+1
    Fin Tant que
    Si (X == 1) Alors
        NbeFilles=NbeEnfants−1
        Afficher « La famille a eu », NbeFilles , « fille(s) et 1 garçon »
    Sinon
        La famille a eu 4 filles et 0 garçon
    Fin Si

Fin
```

   a. Que fait-il ?
   b. Modifier cet algorithme pour qu'il simule les naissances dans N familles quelconque en Syldavie et qu'il renvoit le nombre de garçons et le nombre de filles.
   c. Programmer votre algorithme sur Scilab.
   d. Conjecturer une réponse au problème posé.

2. Le « président » de Syldavie appelle S (pour Succès) et E (pour Echec) les événements suivants :
   S : « La famille a un garçon » et E : « La famille n'a pas de garçon »
   a. Schématiser la situation par un arbre.
   b. On appelle X la variable aléatoire égale à k si le premier Succès est rencontré au $k^{ième}$ enfant, et à 0 si aucun Succès n'a été obtenu.
   Déterminer la loi de X.
   c. Répondre au problème posé.