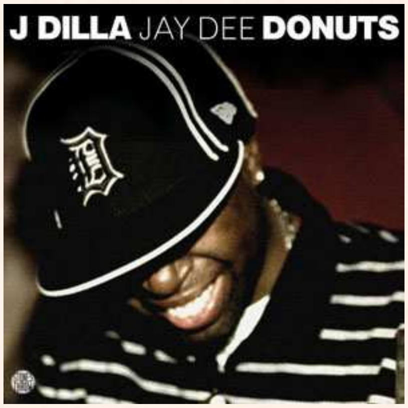
Chapitre 7

Loi binomiale

Hors Sujet



Titre : « J Dilla »
Auteur : DONUTS
Présentation succinte de l'auteur : Evoquer J Dilla aujourd'hui n'est pas une tâche facile. Parti rejoindre l'au-delà, physiquement cette fois, James Yancey aura indéniablement marqué les dix dernières années. Et au-delà de toutes les interrogations sur le bien-fondé des (nombreuses ?) sorties post-mortuaires à venir, ouvrons grandes les mirettes et reconnaissons qu'en quittant ces lieux il laisse derrière lui un héritage musical certain et une discographie pour le moins vaste, méritant une (nouvelle) exploration minutieuse.
Entièrement instrumental, et à ce titre plus ou moins suite de "Welcome to Detroit" et des "Instrumental Series", mené du début à la fin par l'ex-Jay Dee, "Donuts" désarçonne d'emblée par son format. Fort d'une trentaine de beats enchaînés parfois brutalement - mais dépassant péniblement le cap de la minute trente -, "Donuts" prend des allures de mosaïque. Une mosaïque où les boucles s'enchevêtrent et se succèdent régulièrement dans un medley détonant.

Document réalisé à l'aide de LaTeX
Auteur : D. Zancanaro
Site : wicky-math.fr.nf
Lycée : Jean Durand (Castelnaudary)

Table des matières

L'essentiel :

- ⇝ Découvrir les coefficients binomiaux
- ⇝ Reconnaître une situation modéliser par la loi binomiale.
- ⇝ Visualiser sur un tableur l'allure de l'histogramme de la loi binomiale

« *Si les gens ne croient pas que les mathématiques sont simples, c'est seulement parce qu'ils ne réalisent pas combien la vie est compliquée !* »
John Louis von Neumann

Leçon 7

Loi binomiale

Au fil du temps

Alors que les êtres humains se sont intéressés à la géométrie depuis la nuit des temps, et qu'une première présentation rigoureuse, ie **axiomatique** en a été proposée trois siécles avant JC par le grec Euclide, il a fallu attendre le XVI^{eme} siècle pour qu'on s'intéresse enfin aux probabilités, et encore était-ce pour aider les princes à améliorer leurs gains au jeu ...
Ainsi, la théorie des probabilités est jeune par rapport aux autres grandes disciplines mathématiques : elle prend réellement forme au XVII^{e} siècle, dans la correspondance entre Blaise Pascal, philosophe, théologien et mathématicien, et Pierre de Fermat, avocat et mathématicien « amateur »comme il se qualifiait lui-même, alors qu'il est l'une des grandes figures mathmétiques modernes (deux théorèmes d'arithmétiques portent son nom, dont l'un ne fut démontré qu'en 1994).
Si les probabilités ont attendu si longtemps avant de voir le jour, c'est peut-être à cause de l'étrangeté philosophique qu'elles véhiculent, à savoir l'idée qu'un événement du monde physique soit pensé non pas dans les termes « il se produit ou il ne se produit pas », mais « il peut se produire avec x% de chances »...
Si on dit qu'il y a $35{,}4\%$ de chances qu'il pleuve demain à Carcassonne à 11h, cela signifie-t-il qu'il pleuvra réellement à Carcassonne mais seulement à $35{,}4\%$ et qu'il fera en même temps soleil à $64{,}6\%$? Cela semble un pur non sens ... Ou cela sigifie-t-il que la dynamique atmosphérique porte en elle une indétermination physique qui interdit de prévoir « à 100% » ? Ou enfin qu'il n'y a aucune indétermination physique mais que nous, êtres pensants, n'avons pas assez d'informations et de moyens de calculs pour prédire parfaitement son évolution ? En fait, ces trois façons d'interpréter un résultat de probabilité sont valables, chacune dans un certain domaine.
Ainsi, en mécanique quantique, une particule peut se trouver « ici à $35{,}4\%$ et là-bas à $64{,}6\%$ »tant qu'on n'a pas mesuré concrètement sa position (et non pas « ici ou là-bas » !). Les physiciens admettent qu'avant une mesure, une particule unique peut être en plusieurs lieux simultanément...
En ce qui concerne les deux autres interprétations, indétermination physique objective ou manque d'informations du physicien, le débat est ouvert depuis 1958, date à laquelle James Maxwell introduit les probabilités en physique. Pour lui, la température d'un gaz est liée aux probabilités de mouvement des milliards de particules qui le composent ... Très choquant pour les physiciens de l'époque : comment un phénomène aussi réel et objectif que la température d'un gaz peut-il dépendre d'un état de connaissance, c'est-à-dire d'une donnée subjective ? Pourtant, cette théorie est l'une des grandes réussites de la physique moderne ...
Ainsi, si la théorie mathématique des probabilités est aujourd'hui bien comprise et acceptée, dès qu'on cherche à en comprendre le sens physique, l'étonnement reprend le dessus et motive des générations de futurs chercheurs.

C'est en XVIII^{e} siècle que les statistiques endossent quant à elles leur premier grand rôle dans la vie moderne : celui d'un outil de prévision. Le mathématicien Antoine Deparcieu établit dans un livre le « profil »de la mortalité de populations à partir de données statistiques. En se servant des méthodes d'échantillonnage, de calculs de moyennes et d'écart-type, il crée les premières « tables de mortalité »permettant d'évaluer le risque moyen de mort d'un individu en fonction de son profil. Ce risque est alors directement transformé en pécule, rente versée à quelqu'un durant toute sa vie en contrepartie de l'acquisition de son bien à sa mort. Avec Deparcieu, la statistique fait son entrée dans le monde de l'économie.
Mais ce n'est qu'au XIX^{e} siècle que la statistique finit de prendre la place qu'est la sienne aujourd'hui : celle d'une science mathématique mais aussi humaine, omniprésente dans le débat publique. C'est Adolphe Quételet, astronome belge, qui intègre dans un ouvrage toutes les lois de probabilités développées depuis Pascal et Fermat : mesure des erreurs, loi binomiale que vous allez découvrir ici, etc.
Il tente d'établir des lois statistiques des suicides et des crimes en fonction de paramètres comme l'origine sociale, l'âge, le sexe, le climat, le niveau d'études, le revenu, etc. Mais Quételet se voit reprocher de faire de l'homme un être dont le comportement est prédéterminé par des lois mathématiques...
De plus, avec les statistiques sociales, une question se pose : sont-ce les statistiques qui déterminent nos comportements ou l'inverse ? Par exemple, si le taux de meurtriers dans la population est de 5% par an, cela signifie-t-il qu'il existe une sorte de loi qui nous dépasse et qui « oblige »5% de personnes à se transformer en meutriers ?
Ce type de questionnement hantera tout le XX^{e} siècle et conduira à de tragiques dérapages où l'on voudra « neutraliser »dès le berceau tout homme né avec les « paramètres statistiques du crime »...

A-t-on besoin de savoir tout ça pour réussir au Bac ? Par exemple, depuis votre tendre enfance, vous calculez avec les nombres entiers sans connaître les axiomes de Peano, vous travaillez en géométrie euclidienne même si elle ne correspond pas à la réalité : avez-vous déjà rencontré un véritable triangle rectangle ? Et pourtant vous arrivez quand même à démontrer le théorème de Pythagore.
Mais le débat est plus passionné au sujet des probabilités car il a fallu attendre 1933 et le Russe Kolmogorov pour enfin les axiomatiser, alors qu'Euclide avait fait cela pour la géométrie 2300 ans plus tôt...
C'est ce sujet encore brûlant que nous allons explorer cette année à travers quelques chapitres qui sauront, je n'en doute pas, vous passionner !

I. Loi de Bernoulli

Exercice 1. Un vieux professeur d'histoire-géo, se décrivant lui-même « entre deux âges et très séduisant », décide de séquestrer des gens pour le fun. (« Lol ! »dixit le vieux prof d'histoire-géo) Il en a kidnappé cinq : deux filles qu'il n'arrive pas à distinguer et qu'il appelle toutes les deux « Lolo », son fils Hugo et ses collègues Dédé et Gougou.
Le vieux professeur d'histoire-géo choisit au hasard l'un d'entre eux et le séquestre en le pinçant.
Il considère que son expérience est un Succès lorsqu'il séquestre l'une des Lolos, un Echec sinon.
On appelle X la variable aléatoire égale à 1 en cas de Succès et 0 sinon.

1. Déterminer la loi de probabilité de X.
2. Calculer l'espérance de X et sa variance.

Définition 1.

Une **épreuve de Bernoulli** est une expérience aléatoire ne comportant que 2 issues (Succès et Echec). On note p la probabilité de Succès.
Soit X la variable aléatoire qui est égale à 1 en cas de succès et 0 sinon.
Alors, on dit que X suit une **loi de Bernoulli de paramètre** p. On note $X \hookrightarrow B(1; p)$.

Propriété 1.

Si $X \hookrightarrow B(1; p)$ alors :

$$E(X) = p \qquad \text{et} \qquad V(X) = p(1-p)$$

Exemple :

- Pile ou Face.
- Aimer ou non Mireille Matthieu.
- Etre ou ne pas être.
- Séquestrer ou non l'une des Lolos : $X \hookrightarrow B\left(1; \frac{2}{5}\right)$ et $E(X) = \frac{2}{5}$, $V(X) = \frac{2}{5} \times \frac{3}{5} = \frac{6}{25}$
- Lancer un dé et gagner si l'on obtient 5 : $X \hookrightarrow B\left(1; \frac{1}{6}\right)$ et $E(X) = \frac{1}{6}$, $V(X) = \frac{1}{6} \times \frac{5}{6} = \frac{5}{36}$

Preuve

$$E(X) = P(X=0) \times 0 + P(X=1) \times 1 = 0 + p = p$$

$$V(X) = E(X^2) - E(X)^2 = p - p^2 = p(1-p)$$

Remarque : Interprétation des paramètres : Si l'on répète un grand nombre de fois une même expérience de Bernoulli de paramètre p, et de manière indépendante, la fréquence d'un succès sera proche de p, avec un écart-type (ou risque) de $\sqrt{p(1-p)}$.

II. Schéma de Bernoulli - Loi Binomiale

Exercice 2. On reprend le contexte et les notations de l'activité précédente. Le vieux professeur d'histoire-géo considère toujours que son expérience est un Succès lorsqu'il séquestre l'une des Lolos, un Echec sinon et X désigne toujours la variable aléatoire qui compte le nombre de Succès.

1. **Le vieux professeur d'histoire-géo choisit au hasard, successivement et avec remise deux personnes parmi les cinq.**
   (a) Quelles valeurs k peut prendre X ?
   (b) On note $\binom{2}{k}$ (se lit « k parmi 2 ») le nombre de chemins dans l'arbre à deux étapes décrivant la situation, qui mènent à l'événement $(X = k)$, pour $k \in \{0; 1; 2\}$.
   Ainsi $\binom{2}{0}$ désigne le nombre de chemins menant à l'événement $(X = 0)$ parmi les deux expériences.
   Que valent $\binom{2}{0}$? $\binom{2}{1}$ et $\binom{2}{2}$?
   (c) Déterminer la loi de X, son espérance et sa variance.
2. **Le vieux professeur d'histoire-géo choisit au hasard, successivement et avec remise trois personnes parmi les cinq.**
   (a) A votre avis, comment note-t-on le nombre de chemins dans l'arbre décrivant la situation menant à 0 Succès ? Combient vaut-il ?
   (b) Même question pour 1 Succès, puis 2, puis 3.
   (c) Déterminer alors la loi de X, son espérance et sa variance.
3. **Le vieux professeur d'histoire-géo choisit au hasard, successivement et avec remise n personnes parmi les cinq.**
   (a) Conjecturer une formule déterminant $P(X = k)$ en fonction du nombre n d'expériences, du nombre k de Succès et de la probabilité p d'un Succès *(on utilisera notamment la notation $\binom{n}{k}$).*
   (b) Conjecturer une formule pour déterminer l'espérance de X en fonction du nombre n d'expériences et de la probabilité p d'un Succès.
   (c) Ces formules sont-elles valables si le vieux professeur d'histoire-géo choisit au hasard n personnes succesivement et sans remise ? Expliquer.

Définition 2.

On dit que deux expériences aléatoires sont **indépendantes** lorsque les résultats de l'une n'influencent pas les probabilités des issues de l'autre.

Remarque : Pour résoudre des problèmes étudiant une répétition de n expériences aléatoires indépendantes, on utilise très souvent un arbre. Cependant, lorsqu'il s'agit d'une épreuve de Bernoulli répétée indépendamment n fois, on pourra désormais utiliser la propriété définition suivante.

Définition 3.

Soit $n \in \mathbb{N}^*$. La répétition de n épreuves de Bernoulli, identiques et indépendantes, de paramètre p, s'appelle un **schéma de Bernoulli de paramètres n et p**.
On considère l'arbre de probabilité représentant une telle situation.
On note $\binom{n}{k}$ le nombre de chemins réalisant k Succès parmi les n épreuves de Bernoulli, avec $k \in \{0, 1, \dots, n\}$.
Ce nombre (entier) s'appelle un **coefficient binomial** et se lit « k parmi n »

Exercice 3. Combien de branches contient un arbre représentant un schéma de Bernoulli à 4 étapes ?
Déterminer dans cet ordre à l'aide de la définition $\binom{4}{0}$, $\binom{4}{4}$, $\binom{4}{1}$ et $\binom{4}{3}$.
En déduire $\binom{4}{2}$.

Définition 4.

On considère un schéma de Bernoulli de paramètre n et p.
Soit X la variable aléatoire comptant le nombre de succès (X est à valeurs dans $\{0; 1; \dots; n\}$).
Alors, on dit que X suit une **loi binomiale de paramètres n et p**. On note $X \hookrightarrow B(n; p)$.

Théorème 1.

Si $X \hookrightarrow B(n; p)$ alors pour tout $k \in \{0; 1; \dots; n\}$ on a :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

De plus :

$$E(X) = np \qquad \text{et} \qquad V(X) = np(1-p)$$

Preuve

La probabilité d'avoir k succès suivis de $n-k$ échecs est :

$$p^k(1-p)^{n-k}$$

Mais les Succès et les Echecs n'apparaissent pas nécessairement dans cet ordre.... Cependant, on sait par définition qu'il y a $\binom{n}{k}$ chemins de l'arbre qui contiennent k succès parmi les n épreuves (donc $n-k$ échecs) et ces chemins ont tous la même probabilité $p^k(1-p)^{n-k}$ d'être réalisés, car les expériences sont identiques et indépendantes.
On en déduit :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Le reste est admis.

Remarques :

⇝ Si on note q la probabilité d'échec alors $P(X = k) = \binom{n}{k} p^k q^{n-k}$

- ⇝ La probabilité d'avoir n succès est : $P(X = n) = p^n$
- ⇝ La probabilité de n'avoir aucun succès est : $P(X = 0) = q^n$
  Par conséquent, on retrouve des résultats déjà « connus », comme la probabilité d'avoir au moins un succès est :

$$P(X \geq 1) = 1 - P(X = 0) = 1 - q^n$$

Exemple :

Dans lequel des cas suivants X suit-elle une loi binomiale ? Si oui, donner les paramètres de la loi et calculer $P(X = 3)$ si c'est possible, puis l'espérance et la variance de X.

1. Dans une classe, on tire au sort sans remise 5 élèves, X est le nombre d'élèves abonnés à Star'Ac mag dans le lot tiré au sort.
2. Dans un sac de 20 billes contenant 7 noires et 13 blanches, on tire avec remise 3 d'entre elles, X étant le nombre de billes noires obtenues.
3. On lance 4 dés, X est le nombre de 5 obtenus.
4. Un circuit comprend 2 lampes en série. Pour chacune d'elle, la probabilité qu'elle fonctionne est de 0.03. X est le nombres de lampes qui s'allument lorsqu'on appuie sur l'interrupteur.
   Même question avec cette fois des lampes en parallèles.

Exercice 4. On lance n dés ($n \geq 1$). On note A l'événement « obtenir au moins deux 4 (sur l'ensemble des n lancers) », et X la variable aléatoire qui compte le nombre de 4 obtenus.

1. Déterminer la loi de X.
2. On considère le cas $n = 3$.
   (a) Calculer $P(X = 2)$, $P(X = 3)$.
   (b) En déduire $p(A)$.
   (c) Calculer $E(X)$ et $\sigma(X)$.
3. Dans cette question, on suppose n quelconque.
   (a) Décrire l'événement $\overline{A}$ à l'aide d'une phrase.
   (b) Calculer $P(X = 0)$ et $P(X = 1)$ en fonction de n.
   (c) En déduire $p(A)$ en fonction de n.
   (d) A l'aide d'un tableau de valeurs, déterminer le nombre de dés qu'il faut lancer pour que la probabilité d'obtenir au moins un quatre soit supérieure à $\frac{3}{4}$.

Solutions :

1. X suit la loi $B\left(n;\frac{1}{6}\right)$.
2. (a) $P(X=2)=\binom{3}{2}\times\left(\frac{1}{6}\right)^2\times\left(\frac{5}{6}\right)^1$. Il y a trois chemins dans l'arbre représentant la situation qui mène à 2 Succès (SSE, SES, ESS).
   Donc $P(X=2)=3\times\frac{1}{36}\times\frac{5}{6}=\frac{15}{216}$.
   $P(X=3)=\binom{3}{3}\times\left(\frac{1}{6}\right)^3\times\left(\frac{5}{6}\right)^0$. Il y a un seul chemin dans l'arbre représentant la situation qui mène à 3 Succès (SSS).
   Donc $P(X=3)=1\times\frac{1}{216}\times 1=\frac{1}{216}$.
   (b) $p(A)=P(X=2)+P(X=3)=\frac{15}{216}+\frac{1}{216}=\frac{16}{216}$.
   (c) $E(X)=np=3\times\frac{1}{6}=\frac{1}{2}$ $\qquad V(X)=np(1-p)=3\times\frac{1}{6}\times\frac{5}{6}=\frac{5}{12}$ $\qquad$ donc $\qquad \sigma(X)=\sqrt{\frac{5}{12}}$
3. $\overline{A}$ est l'événement « Ne pas obtenir de 4 ».
4. $P(X=0)=\binom{n}{0}\times\left(\frac{1}{6}\right)^0\times\left(\frac{5}{6}\right)^n$. Il y a un seul chemin dans l'arbre représentant la situation qui mène à 0 Succès (EEE... E).
   Donc $P(X=0)=1\times 1\times\left(\frac{5}{6}\right)^n=\left(\frac{5}{6}\right)^n$.
   $P(X=1)=\binom{n}{1}\times\left(\frac{1}{6}\right)^1\times\left(\frac{5}{6}\right)^{n-1}$. Il y a n chemins dans l'arbre représentant la situation qui mène à 1 Succès (n positions possibles pour le Succès dans un tel chemin).
   Donc $P(X=1)=n\times\frac{1}{6}\times\left(\frac{5}{6}\right)^{n-1}$.
5. Par conséquent $P(A)=1-P(\overrightarrow{A})=1-P(X=0)-P(X=1)=1-\left(\frac{5}{6}\right)^n-n\times\frac{1}{6}\times\left(\frac{5}{6}\right)^{n-1}$
6. On cherche le plus petit entier n tel que $P(A)\geq\frac{3}{4}\Longleftrightarrow 1-\left(\frac{5}{6}\right)^n-n\times\frac{1}{6}\times\left(\frac{5}{6}\right)^{n-1}\geq\frac{3}{4}$.
   Nous ne savons pas encore résoudre ce type d'inéquation, mais nous l'apprendrons l'an prochain ...
   A la calculatrice, avec un tableau de valeur, on trouve que pour $n=8$, $P(A)\simeq 0.72$ et pour $n=9$, $P(A)\simeq 0.76$.
   Donc il faut au minimum 9 dés pur que la probabilité d'obtenir au mois deux quatre soit supérieur à 0.75.

III. Loi binomiale et coefficient binomiaux

III.1. Coefficient binomiaux - calcul de proche en proche

Exercice 5. Le vieux professeur d'histoire-géo a décidé de piquer ses otages au cure-dent, juste pour le fun (« Lol ! »dixit le vieux prof d'histoire-géo complètement aveugle).
L'expérience consiste à choisir au hasard l'un des cinq otages, puis à tenter (car l'otage se débat) de le piquer au cure-dent. La probabilité qu'il touche sa cible est de $\frac{2}{3}$ quand il s'agit d'un garçon et de $\frac{1}{2}$ quand il s'agit d'une fille (les deux Lolos étant plus menues que Dédé et plus petites que Hugo et Gougou, il est plus difficile de les atteindre).
Hugo a eu un avertissement travaille ce trimestre-ci, ainsi, le vieux professeur d'histoire-géo considère que son expérience est un Succès lorsqu'il arrive à le piquer.

1. Quelle est la probabilité que le vieux professeur d'histoire-géo choisisse et pique Hugo avec un cure-dent ?
2. Le vieux professeur d'histoire-géo répète n fois cette expérience, de manière indépendante. On appelle X_n la variable aléatoire qui compte le nombre k de Succès S sur les n expériences.

(a) Quelle loi suit X_n ?

(b) Pour $n = 2$, lister intelligemment les chemins de l'arbre décrivant la situation pour déterminer les valeurs de $\binom{n}{k}$ possibles.

(c) Même question pour $n = 3$, puis $n = 4$.

3. (a) Grâce aux résultats des questions **2.b.** à **2.c.**, compléter les lignes correspondantes du tableau ci-contre, donnant les valeurs de $\binom{n}{k}$ en fonction de n et de k.

n \ k	0	1	2	3	4	5	6
0							
1							
2							
3							
4							
5							
6							

(b) A l'aide de la définition de $\binom{n}{k}$ et de votre tête, compléter les lignes $n = 0$ et $n = 1$.

(c) A l'aide de la calculatrice, compléter les lignes $n = 5$ et $n = 6$.

Calculatrices

Casio 35+	**TI 82 à 84**	**TI 89**	**TI Npisre CX CAS**

`6 nCr 2`	`6 Combinaison 2`	`nCr(6,2)`
		ou
		`nbrComb(6,2)`
			`nCr(6,2)`

OPTN puis choisir `PROB`	math puis choisir `PRB` avec ▶	2ND + 5 + 7	catalogue + 2 puis choisir `Probabilités` avec ▼ puis `Nombre de combinaisons`

4. (a) Quelle(s) première(s) constatation(s) pouvez-vous faire sur la valeur des coefficients binomiaux sur une même ligne ?

(b) On veut désormais établir un lien entre deux lignes consécutives. Pour cela, on a fait apparaître en couleur trois séries de 3 cellules.
Trouver une relation simple entre ces 3 cellules.

Propriété 2.

- **Cas particuliers :** Pour tout $n \in \mathbb{N}^*$ on a $\binom{n}{0} = \binom{n}{n} = 1$ et $\binom{n}{1} = \binom{n}{n-1} = n$.
- **Symétrie :** Pour tous entiers naturels n et k tels que $0 \leq k \leq n$ on a

$$\binom{n}{k} = \binom{n}{n-k}$$

- **Triangle de Pascal :** Pour tous entiers naturels n et k tels que $0 \leq k \leq n-1$ on a

$$\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$$

Preuve

On considère un schéma de Bernoulli à n épreuves représenté par un arbre, et k un entier naturel tel que $0 \leq k \leq n$.

1. Il y a un seul chemin de l'arbre conduisant à 0 Succès (c'est celui correspondant à n Echecs) et inversement. Il y a n chemins dans l'arbre conduisant à 1 Succès (celui où le Succès est placer en 1^{er}, en 2^{eme}, ..., en dernier) et inversement.
2. Compter le nombre de chemins de l'arbre contenant n succès revient par complémentarité à compter le nombre de chemins de l'arbre contenant $n-k$ échecs.
   Par symétrie des rôles des succès et des échecs on a donc :
$$\binom{n}{k} = \binom{n}{n-k}$$
3. On considère un schéma de Bernoulli à $n+1$ épreuves représenté par un arbre, et k un entier naturel tel que $0 \leq k \leq n-1$.
   $\binom{n+1}{k+1}$ est, par définition, le nombre de chemins réalisant $k+1$ succès lors des $n+1$ épreuves.
   On peut distinguer deux façons d'obtenir ce type de chemin :
   - Ceux finissent par un succès, ie qui contiennent k succès parmi les n premières épreuves : il y en a $\binom{n}{k}$
   - Ceux finissent par un échec, ie qui contiennent $k+1$ succès parmi les n premières épreuves : il y en a $\binom{n}{k+1}$

   On obtient alors que
$$\binom{n+1}{k+1} = \binom{n}{k} + \binom{n}{k+1}$$

n \ k	0	1	2	3	4	k	$k+1$	n	$n+1$
0									
1									
2									
3									
4									
k									
$k+1$									
n									
$n+1$									

III.2. Applications

III.2.a. Triangle de Pascal

La relation de Pascal[1] permet de calculer les coefficient binomiaux de la façon suivante : pour trouver un certain coefficient, on additionne dans le tableau suivant les coefficients situés "juste au dessus" et "juste au dessus à gauche" entre eux.

n \ p	0	1	2	3	4	…	p-1	p	…	n-1	n
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
…											
$n-1$	1	$n-1$					$\binom{n-1}{p-1}$	$\binom{n-1}{p}$			1
n	1	n						$\binom{n}{p}$		n	1

1. Le tableau est appelé triangle de Pascal en hommage à ce dernier qui écrivit en 1654 son "traité du triangle arithmétique" dans lequel il expose d'innombrables applications du triangle déjà connu de Tartaglia (1556), Stiefel (1543) et des Chinois (1303).

III.2.b. Formule du binôme de Newton

Théorème 2 : *Formule du binôme*

Pour tous nombres réels a et b et tout entier naturel n non nul :

$$(a+b)^n = \sum_{p=0}^{n} \binom{n}{p} a^{n-p} b^p$$

Exemple :

A l'aide de cette formule et du triangle pascal on retrouve des résultats bien utiles :

1. pour $n=2$ $(a+b)^2 = \binom{2}{0}a^2b^0 + \binom{2}{1}a^1b^1 + \binom{2}{2}a^0b^2 = a^2+2ab+b^2$
2. pour $n=3$ $(a+b)^3 = a^3+3a^2b+3ab^2+b^2$
3. pour $n=4$ $(a+b)^4 = a^4+4a^3b^+6a^2b^2+4ab^3+b^4$

Notons qu'il n'est pas inutile de savoir substituer $(-b)$ à b dans la formule pour obtenir :

$$(a-b)^n = \sum_{p=0}^{n} \binom{n}{p} a^{n-p}(-b)^p = \sum_{p=0}^{n} (-1)^p \binom{n}{p} a^{n-p} b^p$$

En pratique, les signes obtenus en développant cette dernière formule alternent ; par exemple :

$$(a-b)^5 = a^5 - 5a^4b + 10a^3b^2 - 10a^2b^3 + 5ab^4 - b^5$$

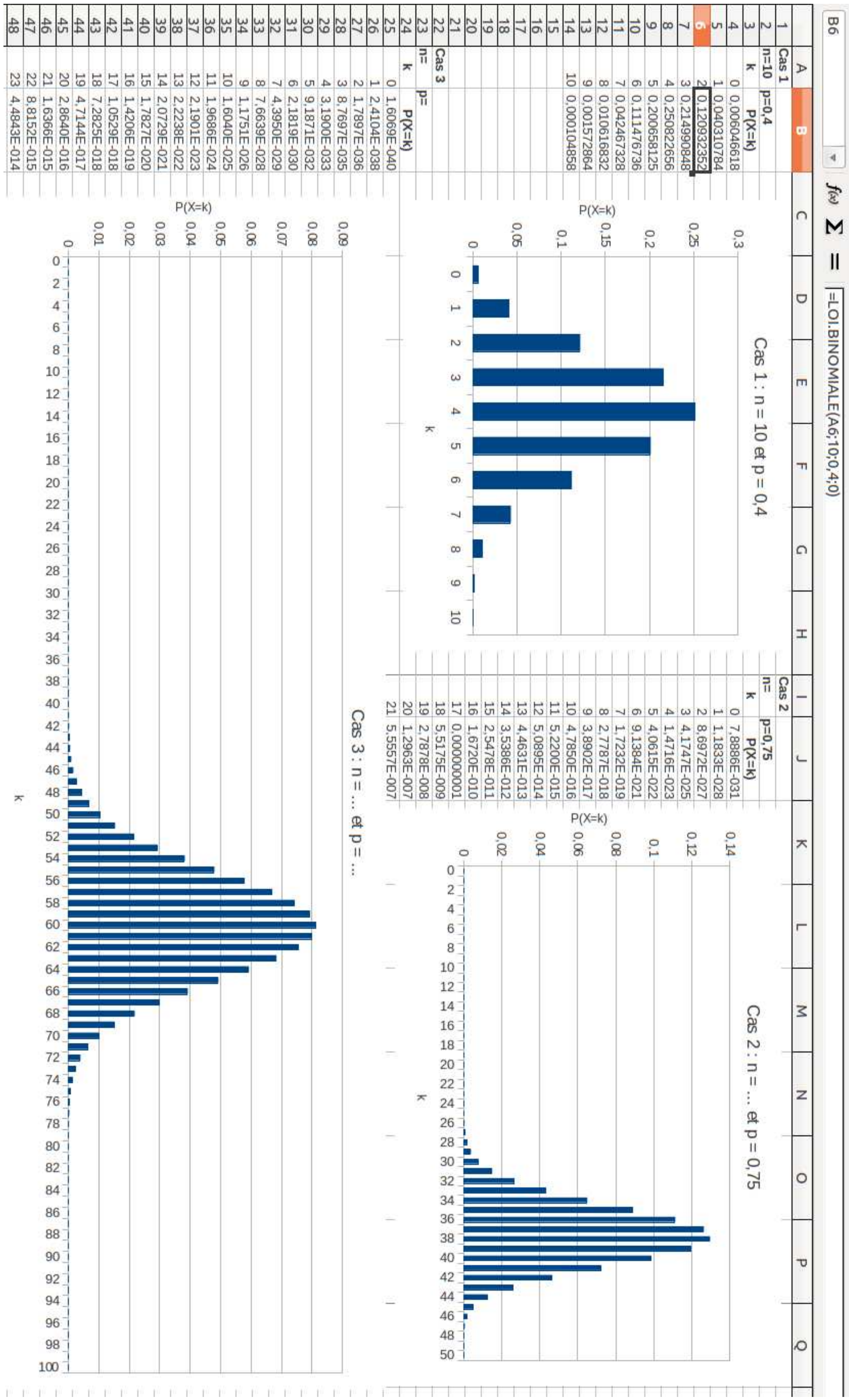
Il est aussi utile de savoir utiliser la formule avec des valeurs particulières de a et b :

1. Lorsque $a=b=1$ on a alors :
$$2^n = \sum_{p=0}^{n} \binom{n}{p}$$
2. Lorsque $a=1$ et $b=-1$ on a alors :
$$0 = \sum_{p=0}^{n} (-1)^p \binom{n}{p}$$

III.3. Table des valeurs et représentation graphique de la loi binomiale

Exercice 6. Sur un tableur, on a calculé la table de valeurs et représenté l'histogramme d'une loi binomiale dans trois cas.

1. Comment décririez-vous l'allure générale de ces histogrammes ?
2. A quoi correspondent les trois premiers paramètres de la ligne de commande entrée en **B6** ?
   Le dernier paramètre égal à `0` *indique au tableur de calculer* $P(X = k)$.
   Il peut également prendre la valeur `1`*, et dans ce cas, il indiquerait au tableur de calculer* $P(X \leq k)$. Déterminer alors la commande écrite en `B10`.
3. Regardons le cas 1.
   (a) Autour de quelle valeur sont situées les valeurs de X les plus probables ?
   (b) Cela vous semble-t-il logique ?
   (c) A quel paramètre de la loi cela correspond-il ?
   (d) Quelles valeurs de X vous semblent obsolètes ? Pourquoi ?
4. Regardons le cas 2.
   (a) Déterminer la valeur de n sur l'histogramme.
   (b) Déterminer alors la commande écrite en `J6`.
   (c) Autour de quelle valeur sont situées les valeurs de X les plus probables ?
   (d) Cela vous semble-t-il logique ?
   (e) Pourquoi cette valeur ne correspond-elle pas à l'espérance de X ?
   (f) Quelles valeurs de X vous semblent obsolètes ? Pourquoi ?
5. Regardons le cas 3.
   (a) Déterminer la valeur de n sur l'histogramme.
   (b) Autour de quelle valeur sont situées les valeurs de X les plus probables ?
   (c) En supposant qu'il s'agit exactement de l'espérance de X, déterminer p.
   (d) Déterminer alors les commandes écrites en `B14` et `100`.
   (e) Quelles valeurs de X vous semblent obsolètes ? Pourquoi ? Votre critère vous semble-t-il rigoureux ?



Cas 1

n=10	p=0,4
k	P(X=k)
0	0,006046618
1	0,040310784
2	0,120932352
3	0,214990848
4	0,250822656
5	0,200658125
6	0,111476736
7	0,042467328
8	0,010616832
9	0,001572864
10	0,000104858

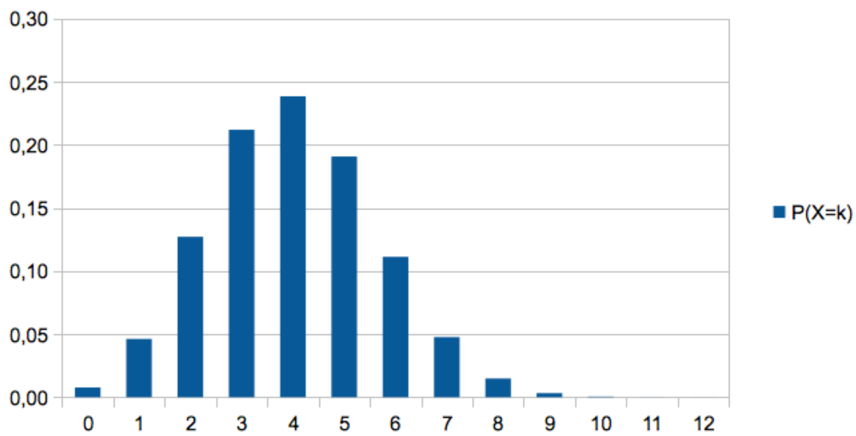
Cas 2

n=	p=0,75
k	P(X=k)
0	7,8886E-031
1	1,1833E-028
2	8,6972E-027
3	4,1747E-025
4	1,4716E-023
5	4,0615E-022
6	9,1384E-021
7	1,7232E-019
8	2,7787E-018
9	3,8902E-017
10	4,7850E-016
11	5,2200E-015
12	5,0895E-014
13	4,4631E-013
14	3,5386E-012
15	2,5478E-011
16	1,6720E-010
17	0,000000001
18	5,5175E-009
19	2,7878E-008
20	1,2963E-007
21	5,5557E-007

Cas 3

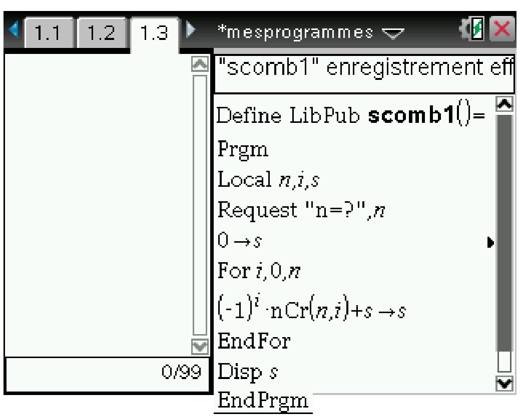
n=	p=
k	P(X=k)
0	1,6069E-040
1	2,4104E-038
2	1,7897E-036
3	8,7697E-035
4	3,1900E-033
5	9,1871E-032
6	2,1819E-030
7	4,3950E-029
8	7,6639E-028
9	1,1751E-026
10	1,6040E-025
11	1,9686E-024
12	2,1901E-023
13	2,2238E-022
14	2,0729E-021
15	1,7827E-020
16	1,4206E-019
17	1,0529E-018
18	7,2825E-018
19	4,7144E-017
20	2,8640E-016
21	1,6366E-015
22	8,8152E-015
23	4,4843E-014

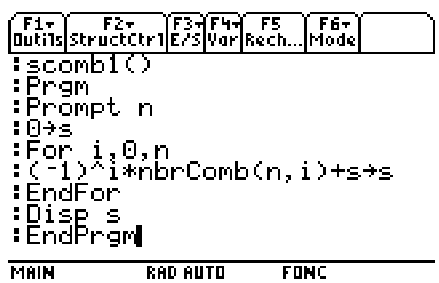
Exercice 7. Soit X une variable aléatoire qui suit une loi binomiale de paramètres n et p.
On a représenté graphiquement les différentes valeurs de $p(X = k)$ à l'aide d'un diagramme en bâtons pour toutes les valeurs de k possibles.



1. Quelle est la valeur de n ?
2. Quelle semble-être la valeur moyenne de X ?
3. En déduire la valeur de p.

Exercice 8. On a créé le programme suivant :





1. Que calcule ce programme ?
2. Entrer ce programme dans la calculatrice et le lancer pour plusieurs valeurs de N.
3. Que peut-on conjecturer ?

IV. TP

IV.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 Fabrice 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 $\mathrm{P}(\mathrm{X}=k)$:

⇝ **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 $\mathrm{P}(\mathrm{X}\leq k)$: 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.

Loi Binomiale à la calculatrice

Pour calculer $\mathrm{P}(\mathrm{X}=k)$:

⇝ **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 $\mathrm{P}(\mathrm{X}\leq k)$: 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.

IV.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 $(n < N)$.
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 C. **Etude d'un cas particulier**

On prend $N = 60$ 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 $T = 80 - 3X$ 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 $E(T) \leq 60 \iff \dfrac{1}{3} - (1-p)^3 \leq 0$
   (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 D. **Autre cas particulier**

On a toujours $N = 60$, 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 E. **Un problème d'optimisation**

Si N est le nombre d'individus, on fait $\dfrac{n}{n}$ 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 $E(T) = N\left(1 + \dfrac{1}{n} - (1-p)^n\right)$
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.

IV.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. Donner ces paramètres.
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 ? Qu'est-ce qui peut justifier une telle composition de jury ?

Il est alors aussi de notre responsabilité d'être critique à l'égard des chiffres sur lesquels toute la simulation repose.

« Alors que 79,1% de la population de comté était d'origine mexicaine, sur les 870 personnes convoqués pour être jurés lors d'une certaine période de référence, il n'y eût que 339 personnes d'origine mexicaine. »

Que signifie : « lors d'une certaine période de référence » ? En réalité, cette période de référence correspond à l'étude des listes des jurés lors des 11 années qui ont précédé le jugement de Partida. Sur ces listes, 339 citoyens d'origine mexicaine ont été comptabilisés sur 870. Or, non seulement il n'est pas raisonnable de penser que durant cette longue période la proportion de citoyens d'origine mexicaine est resté constante et égale à 79,1% mais en plus sur la période de deux ans et demi précédent le jugement de l'affaire opposant le shérif Castaneda au prévenu Partida, la proportion de citoyens d'origine mexicaine ayant été membre d'un jury dans ce même comté était d'environ 56%. Cet argument a été soutenu par l'accusation sans succès car la sous représentation des citoyens d'origine mexicaine restait trop importante.

Une deuxième piste pour contredire la thèse du complot repose sur l'étude des modalités de sélection des jurés. D'abord des listes de citoyens sont éditées de manière aléatoire mais ensuite, conformément à la loi, le juge du comté vérifie que le citoyen est alphabétisé et de « bonnes mœurs ». Ainsi, la loi elle-même impose une sous représentation d'une partie de la population qui est moins alphabétisée.

Enfin, il faut remarquer que la cour suprême ne conclut pas à la démonstration formelle de discrimination raciale. Elle précise, en effet : « Etant donné les nombreuses facettes de la motivation humaine, il serait peu approprié de prendre comme loi établie que des humains appartenant à un groupe ne pratiqueront pas de discrimination à l'égard des membres d'un autre groupe ».

IV.4. Loi géométrique tronquée

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 :

Données: X, NbeEnfants, NbeFilles sont des nombres entiers
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

   (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 $k^{\text{ième}}$ enfant, et à 0 si aucun Succès n'a été obtenu.
   Déterminer la loi de X.
   (c) Répondre au problème posé.