Chapitre 5B : loi binomiale

Valentin Melot — Terminale spé maths A

4 janvier 2021

1 Schéma de Bernoulli

1.1 Un premier exemple : la planche de Galton

Remarque 1 Il est recommandé de visionner la vidéo à l'adresse suivante :

`https://youtu.be/SorBOCTWvws`.

Une planche de Galton consiste en le dispositif illustré à la figure 1. Des billes, insérées dans un dispositif, heurtent une série de picots. Chaque picot a une « probabilité » 1/2 de faire tomber la bille à gauche, et la même de la faire tomber à droite.

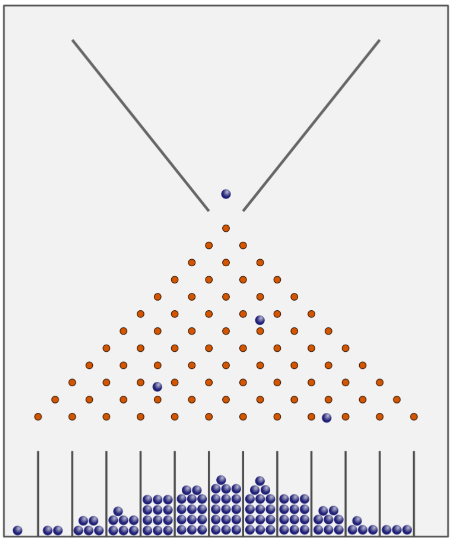

Figure 1 – Description schématique d'une planche de Galton.
Marcin Floryan, Wikimedia Commons, CC-BY-SA 3.0

Notons n le nombre de lignes de picots ($n = 12$ sur la figure). Pour chaque bille lancée, une issue élémentaire correspond à un chemin de la bille. Ce chemin se code par la donnée de n basculements à gauche ou à droite. Autrement dit, notre univers est :

$$\Omega = \{g; d\}^n.$$

Ici, chaque bille subit une série de n épreuves : à chacun des n picots, la bille tombera-t-elle à gauche ou à droite ?

Ces épreuves sont identiques et indépendantes. Si la planche est bien conçue, l'issue de la k-ième épreuve ne dépend pas de celui des précédentes. Et la probabilité [1] de tomber à gauche ou à droite est de 0.5 à chaque fois.

Après avoir passé la série de picots, chaque bille atterrit dans un réservoir. Si les réservoirs sont numérotés de 0 à n, alors le numéro du réservoir où tombe une bille est une variable aléatoire réelle discrète X à valeurs dans $\{0; \dots; n\}$. Nous allons essayer de déterminer la loi de X.

On constate que la valeur de X ne dépend *que* du nombre de fois où la bille est tombée à droite (et donc du nombre de fois où elle est tombée à gauche). Si la bille tombe trois fois à droite, et neuf fois à gauche, alors elle sera dans le réservoir numéro 3, etc.

On remarque que chaque événement élémentaire (chaque trajectoire de bille) a une probabilité de $\frac{1}{2^n}$. Mais combien de trajectoires différentes peuvent-elles amener à avoir $X = k$? Une trajectoire vérifie $X = k$ si, parmi les n épreuves, k d'entre elles exactement ont eu pour issue « droite » – et $n-k$ ont eu pour issue « gauche ». Choisir une trajectoire vérifiant $X = k$ revient donc à choisir k issues « droite » parmi n issues au total. Il y a $\binom{n}{k}$ trajectoires possibles.

En conséquence, on a pour tout $k \in \{0; \dots; n\}$:

$$\mathbf{P}(X = k) = \binom{n}{k} \times \frac{1}{2^n}.$$

On dit que X suit la loi binomiale de paramètres n et 0.5, ce que l'on note :

$$X \sim \mathscr{B}(n; 0.5).$$

Remarque 2 On peut facilement simuler en Python m tirages successifs d'une planche de Galton à n lignes de picots. Vers quel motif tend-on si l'on représente un histogramme statistique des valeurs prises par X lorsque n et m deviennent grands ?

1.2 Généralisation : loi de Bernoulli, loi binomiale

Plus généralement, on considère un espace probabilisé quelconque $(\Omega, \mathbf{P})$. On s'intéresse dans un premier temps à des épreuves à deux issues possibles.

Définition 3 (loi de Bernoulli) *Soit* $(\Omega, \mathbf{P})$ *un espace probabilisé, et* E *une variable aléatoire réelle discrète sur* $(\Omega, \mathbf{P})$.

On dit que E *suit une* loi de Bernoulli de paramètre $p \in [0, 1]$ *si* $\mathbf{P}(X = 0) = 1 - p$ *et* $\mathbf{P}(X = 1) = p$.

On note : $E \sim \mathscr{B}(p)$.

Exemple 4 Soit $(\Omega, \mathbf{P})$ un espace probabilisé correspondant à un tirage équiprobable d'une carte parmi un jeu de cinquante-deux.

Soit A une variable aléatoire valant 1 si la carte tirée est un as et 0 sinon. Alors $A \sim \mathscr{B}\left(\frac{1}{13}\right)$.

1. Ici, la « probabilité » est en quelque sorte une mesure de notre ignorance. En réalité, chaque issue est déterministe, la trajectoire de la bille pourrait être calculée exactement. Mais elle dépend du mouvement initial, au nanomètre près, et d'imperfections minimes sur le système, de sorte que le calcul exact de la trajectoire est impossible en pratique.

On considère généralement que $E = 0$ correspond à l'échec de l'issue et que $E = 1$ correspond au succès. On s'intéresse enfin à la situation d'une répétition d'épreuves de Bernoulli indépendantes.

Définition 5 *Soit* $(\Omega, \mathbf{P})$ *un espace probabilisé. L'on considère des variables aléatoires* $E_1, \ldots, E_n$ *et* C*, telles que :*

(i) Chacune des variables aléatoires $E_1, \ldots, E_n$ *suit une même loi de Bernoulli* $\mathscr{B}(p)$*.*

(ii) Les variables aléatoires $E_1, \ldots, E_n$ *sont indépendantes. Dit autrement, ces variables aléatoires représentent l'issue de* n *expériences dont chacune est indépendante des résultats de l'ensemble des autres.*

(iii) La variable aléatoire C *compte le nombre de succès de ces expériences, c'est-à-dire que* $C = E_1 + \cdots + E_n$*.*

Alors on dit que C *suit une* loi binomiale de paramètres n et p*, et l'on note :* $C \sim \mathscr{B}(n; p)$*.*

La situation précédente est appelée un schéma de Bernoulli.

À chaque fois que la situation présentée dans un exercice correspond à un schéma de Bernoulli, on veillera à bien faire apparaître la nature des expériences successives, à indiquer qu'elles sont identiques et indépendantes, à décrire ce que sont le succès et l'échec, à identifier la variable aléatoire comptant le nombre de succès, et à utiliser les mots schéma de Bernoulli et loi binomiale.

Théorème 6 *Soit* $(\Omega, \mathbf{P})$ *un espace probabilisé, et* $X \sim \mathscr{B}(n; p)$*.*

Pour tout $k \in \{0; \ldots; n\}$*,*

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

(démonstration 1 exigible, démonstration 2 à comprendre)

Exemple 7 Je considère un jeu de 52 cartes, et je tire sept cartes avec remise en mélangeant à chaque fois. Quelle est la probabilité de tirer quatre cœurs ?

Corollaire 8 *Soit* $n \in \mathbf{N}$ *et* $p \in [0, 1]$*. On a :*

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1.$$

Ce corollaire est en fait un cas particulier d'un résultat plus général, formellement hors-programme mais qu'il est bon de savoir redémontrer et qui sera nécessaire pour la suite.

Théorème 9 (formule du binôme de Newton) *Soient a et b deux réels. Pour tout $n \in \mathbf{N}$, on a :*

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

(démonstration à comprendre)

2 Utiliser la loi binomiale

2.1 Indicateurs de la loi binomiale

Propriété 10 *Soit $X \sim \mathscr{B}(p)$. On a :*

1. $\mathbf{E}[X] = p$.
2. $V(X) = p(1-p)$.

Propriété 11 *Soit $X \sim \mathscr{B}(n;p)$. On a :*

1. $\mathbf{E}[X] = np$.
2. $V(X) = np(1-p)$.

On effectue préalablement les calculs suivants :

Lemme 12 *Soit $n \in \mathbf{N}^*$, $1 \leqslant k \leqslant n$ et $2 \leqslant \ell \leqslant n$. Alors :*

$$k\binom{n}{k} = n\binom{n-1}{k-1} \quad ; \quad \ell(\ell-1)\binom{n}{\ell} = n(n-1)\binom{n-2}{\ell-1}.$$

(démonstration à comprendre)

2.2 Problèmes de seuils et d'optimisation

Quelles que soient les valeurs de n et de p, la calculatrice est capable de calculer avec une grande précision, pour tout $k \in \{0, \dots, n\}$, la valeur de $\mathbf{P}(X \leqslant k)$, où $X \sim \mathscr{B}(n;p)$.

Cette valeur est égale à :

$$F_{n,p}(k) = \sum_{i=0}^{k} \mathbf{P}(X=i) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}.$$

On peut ainsi retrouver les probabilités de différents intervalles par différence : pour tous $k < \ell$,

— $\mathbf{P}(X < k) = \mathbf{P}(X \leqslant k-1)$;
— $\mathbf{P}(X > \ell) = \mathbf{P}\left(\overline{X \leqslant \ell}\right) = 1 - \mathbf{P}(X \leqslant \ell)$;
— $\mathbf{P}(X \leqslant \ell) = 1 - \mathbf{P}(X \leqslant \ell - 1)$;
— $\mathbf{P}(k < X \leqslant \ell) = \mathbf{P}\left((X \leqslant \ell) \setminus (X \leqslant k)\right) = \mathbf{P}(X \leqslant \ell) - \mathbf{P}(X \leqslant k)$;
— etc.

Remarque 13 Il est nécessaire d'être capable de réaliser rapidement ces petites transformations au brouillon pour ne pas faire d'erreur. Il est inutile de chercher à les apprendre par cœur.

Remarque 14 La calculatrice et la plupart des outils numériques (Excel, LibreOffice Calc, Python, R, Stata, etc.) disposant de bibliothèques de calcul numérique permettent de calculer facilement les $F_{n,p}(k)$. Certaines proposent également de calculer les probabilités d'intervalles d'autres formes. Il est nécessaire de :
— Savoir utiliser sa calculatrice ;
— Savoir rechercher et lire les documentations des logiciels que l'on utilise moins souvent.

Définition 15 *Soit* $p \in [0,1]$*. On appelle* intervalle de probabilité p pour la variable aléatoire X *un intervalle* I *tel que* $\mathbf{P}(X \in I) \geqslant p$*.*

Exemple 16 Soit $B \sim \mathscr{B}(20; 0,5)$. Donner un intervalle de la forme $[10-\alpha, 10+\alpha]$ de probabilité $0,95$ pour X.

Exemple 17 (un problème de surréservation) La compagnie ProBaBus vend des billets d'autocar pour effectuer le trajet de Vincennes à Bâle en Suisse — où est né, a vécu et est mort un certain Jacques Bernoulli.

Il y a soixante places dans le bus. La compagnie constate empiriquement que pour chaque billet vendu sur cette ligne, il y a 9 % de chances que le passager ne se présente pas au départ, et décide donc de faire une surréservation.

Chaque billet est vendu 100 €, mais si un passager ne peut pas embarquer, la compagnie doit lui rembourser le double du prix d'achat du billet.

1. La compagnie ne souhaite pas consacrer plus de 15 % de son chiffre d'affaires à l'indemnisation des passagers en surréservation. Combien de billets doit-elle vendre au maximum pour avoir 95 % de chances de respecter cette contrainte ?
2. La compagnie vend effectivement ce nombre de billets. On appelle X le nombre de passagers présents au départ. Donner la loi de X.
3. On appelle F les frais payés par la compagnie en indemnisation du nombre de passagers en surréservation. Décrire la loi de F.
4. La stratégie de surréservation de la compagnie est-elle rentable ?

Exemple 18 (plan d'investissements pour le RER B) Le RER B est parcouru par 570 trains par jour ouvrable, soit environ 150 000 trains par an. Chaque mission a une certaine probabilité de connaître une défaillance technique.

Les infrastructures sont vétustes, et la RATP doit procéder à un vaste programme d'investissement. Elle modélise, en première approximation, que l'investissement à réaliser (en M€) pour atteindre une probabilité de panne par mission de α est de $\frac{1}{4\times\alpha^2}$.

D'autre part, elle estime que pour chaque train tombant en panne, les retards cumulés sur la ligne lui coûtent un million d'euros au total, sous la forme de pénalités d'irrégularité versées à IDFM.

1. La RATP souhaite, avec une probabilité de 95 % au moins, subir moins de 1 000 défaillances techniques par an. Quel est, à 10 M€ près, l'investissement minimal à réaliser pour atteindre cet objectif ? Quel sera alors le nombre moyen de trajets entre deux pannes ? *On exprimera dans un tableur la probabilité qu'il y ait moins de* 1 000 *défaillances en fonction de l'investissement réalisé.*
2. Quelles seront alors les pénalités moyennes versées par la RATP à IDFM chaque année ?
3. Dans une optique de performance financière, la RATP cherche désormais à minimiser les dépenses (c'est-à-dire, la somme des pénalités moyennes IDFM et de l'investissement réalisé) sur dix ans. Quel est, à 10 M€ près, l'investissement optimal à effectuer ? Quel sera alors le nombre moyen de trajets entre deux pannes ? *On pourra étudier la fonction qui au prix p investi associe la dépense totale cumulée sur dix ans, et rechercher son minimum.*

3 Approfondissement : événements rares et loi de Poisson

On admet le résultat suivant, qui pourra être démontré une fois vu le cours sur les logarithmes :

Lemme 19 (provisoirement admis) *Soit* $(p_n)_{n\in\mathbf{N}}$ *une suite de nombres appartenant à l'intervalle* $[0,1]$. *On suppose qu'il existe un réel* $\ell \in \mathbf{R}_+^*$ *tel que* $\lim\limits_{n\to+\infty} np_n = \ell$.

Alors, pour tout $k \in \mathbf{N}$,

$$\lim_{n\to+\infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\ell^k}{k!}e^{-\ell}.$$

Définition 20 *Soit* $(\Omega, \mathbf{P})$ *un espace probabilisé*[2]. *On dit qu'une variable aléatoire* X *suit une* loi de Poisson de paramètre $\ell \in \mathbf{R}_+^*$ *si* X *prend ses valeurs dans* $\mathbf{N}$ *et si pour tout* $n \in \mathbf{N}$,

$$\mathbf{P}(X = n) = \frac{\ell^k}{k!}e^{-\ell}.$$

On note :

$$X \sim \mathscr{P}(\ell).$$

On *admet* que la formule suivante définit bien une variable aléatoire, et en particulier que $\mathbf{P}(X=0) + \mathbf{P}(X=1) + \cdots + \mathbf{P}(X=n) + \dots$ a un sens et est égale à 1.

On peut définir, de façon rigoureuse, la notion de convergence d'une suite de variables aléatoires[3]. Au vu du lemme 19, on peut dire que si $(X_n)_{n\in\mathbf{N}}$ est une suite de variables

2. Cet espace doit être de taille infinie ; on touche là aux limites du programme.
3. Nous approfondirons cette idée dans un prochain chapitre de probabilités.

aléatoires telle que pour tout $n \in \mathbf{N}$, $X_n \sim \mathscr{B}(n; p_n)$ avec $\lim_{n \to +\infty} np_n = \ell$, alors la suite $(X_n)_{n\in\mathbf{N}}$ « converge en loi » vers une variable aléatoire X_∞, de loi $\mathscr{P}(\ell)$. Autrement dit, une loi binomiale $\mathscr{B}(n; p)$ avec n grand et p petit peut être approximée par une loi de Poisson $\mathscr{P}(np)$, qui n'a plus qu'un paramètre et est donc plus simple à étudier.

La loi de Poisson, asymptotiquement, compte donc les succès d'un nombre très grands d'expériences avec une très faible issue de succès. C'est ce qui justifie qu'on la qualifie de « loi des événements rares ».

Exemple 21 Une usine produit un grand nombre d'objets sur une même chaîne de montage, et suit des protocoles très stricts de qualité pour éviter les défaillances des objets qu'elle produit. Des défaillances exceptionnelles peuvent toutefois arrivées, de façon aléatoire, mais en temps normal, elles doivent être improbables et indépendantes.

L'ingénieur qualité de l'usine appelle X le nombre de défaillances de produits constatée une semaine donnée. X compte le nombre de succès des expériences « le produit considéré a un défaut », expérience répétée autant de fois qu'il y a de produits pendant une semaine (nombre que l'on note n) et avec une probabilité de succès faible (que l'on note $\frac{\ell}{n}$, ℓ étant donc le nombre moyen de défauts attendus).

X suit une loi binomiale $\mathscr{B}\left(n; \frac{\ell}{n}\right)$. Mais puisque n est grand et varie chaque semaine, on fait le choix d'approximer X par une loi de Poisson $\mathscr{P}(\ell)$, ce qui est possible d'après le lemme précédent.

Aussi, chaque semaine, indépendamment du nombre de produits sortis de la chaîne, l'ingénieur qualité s'attend à ce que X corresponde à un tirage d'une même loi de Poisson $\mathscr{P}(\ell)$. Dans l'hypothèse où les mesures empiriques de X s'éloigneraient trop d'une loi de Poisson, l'ingénieur en déduirait que, avec une haute probabilité, sa modélisation est fausse. L'hypothèse la plus improbable serait alors que les défaillances ne soient pas indépendantes, ce qui traduirait la présence d'un problème sur la chaîne, à résoudre au plus vite.

Exemple 22 De façon similaire, un spécialiste des mathématiques financières travaillant dans une banque pourrait modéliser par une loi de Poisson le nombre de ménages bénéficiaires d'un crédit faisant défaut sur une période de temps donnée, à supposer que la capacité de remboursement des ménages est indépendante.

S'il observait que les données réelles ne correspondent plus à une loi de Poisson, il en déduirait que les capacités de remboursement des ménages ne sont plus indépendantes. Cela traduirait probablement le début d'une crise systémique, à laquelle il faudrait réagir rapidement pour minimiser ses pertes.

Les acquis de ce chapitre

Démonstration exigible :

1. Calcul de $\mathbf{P}(X=k)$ lorsque $X \sim \mathscr{B}(n;p)$ par le dénombrement.

Savoirs et savoirs-faire indispensables.

1. Justifier qu'une situation est un schéma de Bernoulli, modéliser par une variable aléatoire suivant une loi binomiale.
2. Utiliser les indicateurs des lois binomiales.
3. Résoudre un problème de seuil faisant intervenir une loi binomiale.

Démonstrations qu'il faut avoir comprises :

1. Calcul de $\mathbf{P}(X=k)$ lorsque $X \sim \mathscr{B}(n;p)$ par récurrence.
2. Calcul de l'espérance et de la variance d'une variable aléatoire suivant une loi binomiale.
3. Formule du binôme de Newton.

Principaux approfondissements.

1. Connaître et utiliser la formule du binôme de Newton.
2. Approximer des situations faisant intervenir des lois binomiales avec des événements rares par une loi de Poisson.
3. Simuler une planche de Galton.