

Loi binomiale



1. Épreuve et loi de Bernoulli

Définition On appelle **épreuve de Bernoulli** une expérience aléatoire présentant deux issues : l'une, notée S , appelée **succès** et l'autre $\bar{S}$ appelée **échec**. On note p la probabilité du succès, et parfois $q = 1 - p$ la probabilité de l'échec.

La variable aléatoire X qui prend la valeur 1 en cas de succès et 0 en cas d'échec est appelée variable aléatoire de Bernoulli.

La loi de probabilité, appelée loi de Bernoulli de paramètre p , est alors donnée par :

x_i	0	1
$P(X = x_i)$	$1 - p$	p

Propriété Soit X une variable aléatoire suivant une loi de Bernoulli de paramètre p . Alors :

$$E(X) = p \quad V(X) = p(1 - p) \quad \text{et} \quad \sigma(X) = \sqrt{p(1 - p)}$$

2. Schéma de Bernoulli et loi binomiale

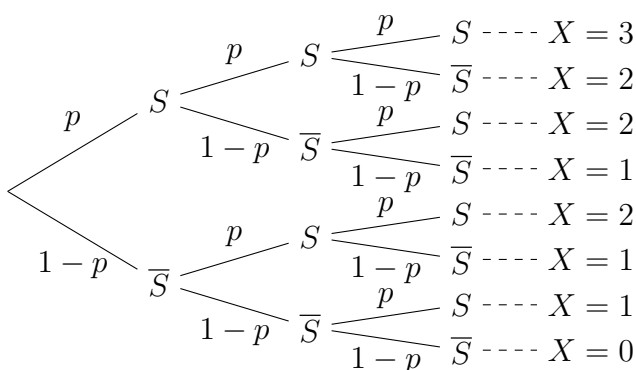
Définition L'expérience aléatoire consistant à répéter n fois ($n \in \mathbb{N}^*$) de manière indépendante une épreuve de Bernoulli de paramètre p s'appelle un **schéma de Bernoulli de paramètres n et p** .

On considère la variable aléatoire X égale au nombre de succès obtenus au cours des n épreuves.

On appelle alors **loi binomiale de paramètres n et p** la loi de probabilité de X .

On note cette loi $\mathcal{B}(n, p)$ et on écrit que $X \sim \mathcal{B}(n, p)$ (le symbole $\sim$ se lit « suit la loi »).

Exemple Voici une représentation avec un arbre pondéré pour $n = 3$:



Il y a une rédaction à faire pour justifier qu'une variable aléatoire suit une loi binomiale.

Exemple Un QCM comporte cinq questions et, pour chaque question, quatre réponses sont proposées dont une seule est exacte. Un élève répond au hasard à ce QCM, ses réponses étant alors indépendantes. En particulier, pour chaque question, la probabilité que la réponse de l'élève soit exacte est $\frac{1}{4}$. On nomme X la variable aléatoire comptant le nombre de réponses exactes obtenues par cet élève. Quelle est la loi suivie par X ?

On considère l'**épreuve de Bernoulli** consistant pour l'élève à répondre à **une** question du QCM. L'**événement succès** est le fait que la réponse soit exacte. La **probabilité de succès** est alors $p = \frac{1}{4}$. L'**expérience est répétée $n = 5$ fois de manière indépendante** et on s'intéresse au **nombre X de succès**, on obtient donc un schéma de Bernoulli et X suit la loi binomiale de paramètres $n = 5$ et $p = \frac{1}{4} = 0,25$: On le note $X \sim \mathcal{B}(5; 0,25)$.

3. Coefficients binomiaux

Remarque À un nombre de succès k donné correspond généralement plusieurs branches de l'arbre (voir l'arbre plus haut).

Ses branches sont de même probabilité (car il y a toujours k succès et $n - k$ échecs, mais dans un ordre différent).

Il suffit donc de compter le nombre de branches pour donner une formule de la loi de probabilité.

Définition On considère un schéma de Bernoulli correspondant à la répétition de n épreuves de Bernoulli. Le nombre de façons d'obtenir k succès parmi les n épreuves se note $\binom{n}{k}$ et se lit « k parmi n ».

Les nombres $\binom{n}{k}$ s'appellent les coefficients binomiaux.

Propriété

1. Pour tout entier k compris entre 0 et n , $\binom{n}{k} = \binom{n}{n-k}$.
2. (Formule de Pascal) pour tout entier n non nul et tout entier k compris entre 0 et $n - 1$,

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

Pour calculer les coefficients binomiaux :

- On peut utiliser la calculatrice :
 - * Casio : OPTN → PROB → nCr ;
En écriture linéaire, taper d'abord n , puis la commande, puis k .
 - * TI : MATH → PRB → Combinaison
En écriture linéaire, taper d'abord n , puis la commande, puis k .
 - * Numworks : boîte à outils, menu Dénombrement.
- On peut aussi utiliser le **Triangle de Pascal** pour les petites valeurs :

$k \rightarrow$	0	1	2	3	4	...
$n = 0$	1					
$n = 1$	1	1				
$n = 2$	1	2	1			
$n = 3$	1	3	3	1		
$n = 4$	1	4	6	4	1	
...	...	...	...	...	...	...

Propriété Soit X une variable aléatoire qui suit la loi binomiale $\mathcal{B}(n; p)$. Alors, pour tout k compris entre 0 et n ,

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

Propriété Soit X une variable aléatoire qui suit la loi $\mathcal{B}(n; p)$. Alors :

$$E(X) = np, \quad V(X) = np(p - 1), \quad \text{et } \sigma(X) = \sqrt{np(p - 1)}$$

Utilisation de la calculatrice :

- $\mathbb{P}(X = k)$ En TI : binomFdp(n,p,k)
En Casio : BinominalPD(k,n,p)
En Numworks : menu probabilités
- $\mathbb{P}(X \leq k)$ En TI : binomFRép(n,p,k)
En Casio : BinominalCD(k,n,p)
En Numworks : menu probabilités

Échantillonnage et estimation



1. Intervalle de fluctuation et prise de décision

Dans une population, on souhaite vérifier si la proportion d'apparition d'un caractère donné est une valeur p annoncée (ou souhaitée).

Pour décider si cette affirmation est acceptable ou non, on utilise un échantillon de la population, sur lequel on calcule la fréquence d'apparition du caractère étudié.

Bien sûr, d'un échantillon à l'autre, la fréquence obtenue varie, car il s'agit d'une expérience aléatoire, et elle n'est presque jamais égale à p ; elle peut lui être supérieure ou inférieure.

Comment décider alors que la valeur de la fréquence « correspond » à la valeur p ?

Pour cela on utilise un protocole rigoureux basé sur la loi binomiale.

En effet, si on prélève un échantillon aléatoire de taille n et si on considère la variable aléatoire X , nombre d'individus de cet échantillon ayant ce caractère, cette variable aléatoire X suit la loi binomiale $\mathcal{B}(n; p)$ (à l'approximation près liée au fait que la taille de l'échantillon est très petite devant celle de la population et que l'on peut considérer l'expérience comme un tirage avec remise).

La valeur X étant le nombre d'individus ayant le caractère donné, on note $F = \frac{X}{n}$ la variable aléatoire donnant la fréquence du caractère dans l'échantillon.

Définition On appelle intervalle de fluctuation au seuil (de confiance) $0,95 = 95\%$ de la fréquence l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ où :

- a est le plus petit entier tel que $\mathbb{P}(X \leq a) > 0,025$;
- b est le plus petit entier tel que $\mathbb{P}(X \leq b) \geq 0,975$.

Il faut retenir que la fréquence pour un échantillon, autrement dit $F = \frac{X}{n}$, a alors une probabilité au moins égale à $0,95$ d'appartenir à cet intervalle.

De manière plus générale, on peut fixer un seuil de confiance $1 - \alpha$ (avec $0 < \alpha < 1$ généralement très proche de 0) et définir alors un intervalle de fluctuation au seuil (de confiance) $1 - \alpha$:

Définition On appelle intervalle de fluctuation au seuil $1 - \alpha$ de F l'intervalle $\left[\frac{a}{n}; \frac{b}{n}\right]$ où :

- a est le plus petit entier tel que $\mathbb{P}(X \leq a) > \frac{\alpha}{2}$;
- b est le plus petit entier tel que $\mathbb{P}(X \leq b) \geq 1 - \frac{\alpha}{2}$.

Remarque L'intervalle de fluctuation au seuil 95% vu plus haut, assez courant, correspond à celui obtenu pour $\alpha = 0,05$.

On obtient les valeurs de a et de b à l'aide de la calculatrice :

Casio : InvBinomCD(k, n, p) (disponible avec les touches de fonction dans $\text{dist} > \text{binomial}$) où l'on remplace k par $\frac{\alpha}{2}$ puis par $1 - \frac{\alpha}{2}$;

TI : Pour les modèles récents, il existe la fonction InvBinom (disponible dans distrib).

On donne pour l'aire les valeurs de $\frac{\alpha}{2}$ puis $1 - \frac{\alpha}{2}$.

Pour les modèles plus anciens, on peut afficher le tableau des valeurs des $\mathbb{P}(X \leq k)$ en définissant la fonction : $Y1 = \text{binomFRep}(n, p, X)$

Puis l'on cherche dans la table de valeurs celles de X qui font dépasser $\frac{\alpha}{2}$ puis $1 - \frac{\alpha}{2}$.

Numworks : Dans le menu, choisir Probabilité puis loi binomiale. Donner les valeurs des paramètres n et p .

Choisir l'affichage sous forme $P(X \leq \dots) = \dots$. Choisir la valeur de probabilité, la calculatrice cherche alors la valeur de X , à laquelle il faut parfois ajouter un (car elle cherche la plus proche de la probabilité donnée, mais qui peut être inférieure).

Exemple Pour $n = 100$ et $p = 0,5$, avec $\alpha = 0,05$, on trouve $a = 40$ et $b = 60$.

L'intervalle de fluctuation au seuil 0,95 est donc $\left[\frac{40}{100}; \frac{60}{100} \right]$ soit $[0,4; 0,6]$.

Remarque Il s'agit d'un intervalle de fréquences, les valeurs sont donc nécessairement comprises entre 0 et 1.

Remarque plus α est petit, autrement dit plus le seuil augmente, plus l'intervalle est grand (si la probabilité que F appartienne à l'intervalle doit augmenter, il est nécessaire que la longueur de l'intervalle augmente).

Remarque Pour une valeur de α fixé, plus n est grand, plus l'intervalle est petit (les fluctuations diminuent).

Une fois l'intervalle obtenu, **on fait l'hypothèse** que « la proportion du caractère étudié dans la population est p ».

On utilise alors la **règle de décision** suivante :

Soit f_{obs} la fréquence observée dans l'échantillon (la réalisation de la variable $F = \frac{X}{n}$).

- Si f_{obs} appartient à l'intervalle de fluctuation, on valide l'hypothèse selon laquelle la proportion dans la population est p .
- Si f_{obs} n'appartient pas à l'intervalle de fluctuation, l'hypothèse est rejetée.
À savoir : il y a une probabilité α de se tromper la rejeter à tort.

Exemple Comme dans l'exemple précédent on suppose que $p = 0,5$ et $n = 100$. Supposons maintenant que dans un échantillon de 100 individus, 36 individus aient le caractère étudié.

Alors $f = \frac{36}{100} = 0,36$.

On rappelle que l'intervalle de fluctuation au seuil de confiance 95% est $I = [0,4; 0,6]$.

On observe que $f \notin I$, donc on rejette l'hypothèse que la proportion p soit égale à 0,5 dans la population.

2. Intervalle de confiance et estimation

Ici on s'intéresse au problème contraire : on ignore la valeur de la proportion p dans la population, et on souhaite l'estimer à partir d'un échantillon. C'est le problème des instituts de sondage.

Plutôt que donner seulement la valeur f_{obs} , dont on sait, à cause des fluctuations d'un échantillon à l'autre, qu'elle ne peut pas être égale à p mais qu'elle en est proche, on souhaite donner un intervalle, pour lequel on peut être sûr à 95% qu'il contient la valeur p .

Définition Un **intervalle de confiance** pour une proportion p au niveau de confiance 0,95 est la réalisation, à partir d'un échantillon, d'un intervalle aléatoire contenant la proportion p avec une probabilité supérieure ou égale à 0,95

Propriété Si f_{obs} est la fréquence observée sur un échantillon de taille n , on donne comme intervalle de confiance au niveau de confiance 0,95 l'intervalle suivant :

$$\left[f_{obs} - \frac{1}{\sqrt{n}}; f_{obs} + \frac{1}{\sqrt{n}} \right]$$

Exemple Un sac opaque contient un très grand nombre de boules rouges ou bleues indiscernables au toucher. Lors d'un tirage de 100 boules, on obtient 41 boules rouges et 59 boules bleues. La fréquence observée de boules rouges dans l'échantillon est donc $f_{obs} = 0,41$.

L'intervalle $\left[0,41 - \frac{1}{\sqrt{100}}; 0,41 + \frac{1}{\sqrt{100}} \right] = [0,40; 0,42]$ est un intervalle de confiance au niveau de confiance de 95% de la proportion p de boules rouges dans le sac.

En principe il y a des conditions pour assurer un niveau de confiance de 95%, qui sont les suivantes : $n \geq 30$, $N \times f \geq 5$ et $n \times (1 - f) \geq 5$.

Autrement dit, il faut n « assez grand » et f ni trop grand ni trop petit.