# Variables aléatoires discrètes

## 1 Variable aléatoire & Loi de probabilité

**Définition 1**

On considère $\Omega$ un ensemble discret (fini la plupart du temps).
**Une variable aléatoire** X est une fonction définie sur $\Omega$ à valeurs dans $\mathbb{R}$.
X prend alors un nombre discret de valeurs $x_1;\dots;x_r;\dots$ on note alors « $X = x_i$ » l'événement « X prend la valeur $x_i$ ».

**Remarque :** Un ensemble est **discret** si on peut *énumérer* ses éléments : ainsi tout ensemble fini ainsi que $\mathbb{N}$ sont discrets... mais $\mathbb{R}$ ne l'est pas.

**Exemple :**
On considère un jeu de 32 cartes et on y tire une carte.
L'univers de cette expérience aléatoire est $\Omega = \{$ 7 de trèfle ;...;as de pique$\}$.
On est en situation d'équiprobabilité, la probabilité de tirer une carte est $\frac{1}{32}$.
On convient de jouer au jeu suivant : on mise 5€, on gagne 10€ si c'est un as et 5€ si c'est une figure.
On note X le gain algébrique ; X peut prendre les valeurs $-5+10 = 5$€ si on tire un as, $5-5 = 0$€ si c'est une figure et $0-5 = -5$€ dans les autres cas.
X est donc une variable aléatoire défini sur $\Omega$ dans $\{-5;0;5\}$.

On peut aussi définir la loi de probabilité associé :

| Valeurs $x_i$ | 5 | 0 | $-5$ |
|---|---|---|---|
| $\mathbf{P}(X = x_i)$ | $\frac{4}{32} = \frac{1}{8}$ | $\frac{12}{32} = \frac{3}{8}$ | $\frac{16}{32} = \frac{1}{2}$ |

.

**Définition 2**

On considère X une variable aléatoire qui prend des valeurs $x_1;\dots;x_r$ .
**Une loi de probabilité associée à** X est la donnée de nombres $p_i \in [0;1]$ associés à $x_i$ tels que
$$\sum_{i=1}^{r} p_i = p_1 + \cdots + p_r = 1.$$

On la présente généralement sous la forme d'un tableau :

| Valeurs $x_i$ | $x_1$ | ... | $x_r$ |
|---|---|---|---|
| $\mathbf{P}(X = x_i)$ | $p_1$ | ... | $p_r$ |

## 2 Outils d'étude des variables aléatoires

**Définition 3**

X est une une variable aléatoire de loi de probabilité :

| Valeurs $x_i$ | $x_1$ | ... | $x_r$ |
|---|---|---|---|
| $\mathbf{P}(X = x_i)$ | $p_1$ | ... | $p_r$ |

.

**L'espérance de** X est le nombre noté $\mathbf{E}(X)$ défini par :

$$\mathbf{E}(X)=\sum_{i=1}^{r}p_ix_i=p_1x_1+\cdots+p_rx_r$$

**Propriété 1**

X est une variable aléatoire.
L'espérance de X s'interprète comme la valeur moyenne prise par X au bout « d'un très grand nombre » d'expériences.

**Remarque :** On appelle cette propriété « **loi des grands nombres** ».

**Définition 4**

X est une variable aléatoire de loi de probabilité :

| Valeurs $x_i$ | $x_1$ | ... | $x_r$ |
|---|---|---|---|
| $\mathbf{P}(X=x_i)$ | $p_1$ | ... | $p_r$ |

.

- **La variance de** X est le nombre noté $\mathbf{V}(X)$ défini par :

$$\mathbf{V}(X)=\mathbf{E}\left(\left(X-\mathbf{E}(X)\right)^2\right)$$

$$\mathbf{V}(X)=\sum_{i=1}^{r}p_i\left(x_i-\mathbf{E}(X)\right)^2=p_1\left(x_1-\mathbf{E}(X)\right)^2+\cdots+p_r\left(x_r-\mathbf{E}(X)\right)^2$$

- **L'écart-type de** X est le nombre noté $\sigma(X)$ défini par :

$$\sigma(X)=\sqrt{\mathbf{V}(X)}$$

**Propriété 2**

X est une variable aléatoire.
L'écart-type de X s'interprète comme l'écart-type de X déterminé à partir « d'un très grand nombre » d'expériences.

**Propriété 3**

*Formule de König-Huygens*

X est une variable aléatoire de loi de probabilité :

| Valeurs $x_i$ | $x_1$ | ... | $x_r$ |
|---|---|---|---|
| $\mathbf{P}(X=x_i)$ | $p_1$ | ... | $p_r$ |

.

$$\mathbf{V}(X)=\sum_{i=1}^{r}p_ix_i^2-\mathbf{E}(X)^2$$

**Propriété 4**

X est une une variable aléatoire d'espérance et $a$ et $b$ deux nombres réels.
- $\mathbf{E}(aX+b)=a\,\mathbf{E}(X)+b$
- $\mathbf{V}(aX+b)=a^2\,\mathbf{V}(X)$
- $\sigma(aX+b)=|a|\sigma(X)$

## 3 Répétitions d'expériences aléatoires identiques et indépendantes

On considère une expérience aléatoire avec, par exemple, trois issues A, B et C.
On répète cette expérience *dans les mêmes conditions* deux fois de suite : les probabilités $\mathbf{P}(A)$ ; $\mathbf{P}(B)$ et $\mathbf{P}(C)$ ne varient donc pas d'une expérience à une autre.

On peut alors représenter la situation grâce à **un arbre de répétitions d'expériences aléatoires** comme ci-contre.

Dans cette situation particulière, la probabilité de réaliser successivement deux événements dans un ordre précis est égal au produit de leur probabilité. Ainsi :
- $\mathbf{P}(\text{« AA »}) = \mathbf{P}(A) \times \mathbf{P}(A)$
- $\mathbf{P}(\text{« AC »}) = \mathbf{P}(A) \times \mathbf{P}(C)$
- $\mathbf{P}(\text{« CA »}) = \mathbf{P}(C) \times \mathbf{P}(A)$

nodesep=0mm,levelsep=20mm,treesep=10mm [treemode=R] [tnpos=a]A$\mathbf{P}(A)$ [tnpos=r]A$\mathbf{P}(A)$ [tnpos=r]B$\mathbf{P}(B)$ [tnpos=r]C$\mathbf{P}(C)$ [tnpos=a]B$\mathbf{P}(B)$ [tnpos=r]A$\mathbf{P}(A)$ [tnpos=r]B$\mathbf{P}(B)$ [tnpos=r]C$\mathbf{P}(C)$ [tnpos=a]C$\mathbf{P}(C)$ [tnpos=r]A$\mathbf{P}(A)$ [tnpos=r]B$\mathbf{P}(B)$ [tnpos=r]C$\mathbf{P}(C)$

# 4 Épreuve et loi de Bernoulli

**Définition 5**

**Une épreuve de Bernoulli de paramètre** $p \in [0;1]$ est une expérience aléatoire présentant deux issues, dont l'une est « **le succès** », de probabilité $p$ et l'autre « **l'échec** », de probabilité $1-p$.

On donne ci-contre une représentation d'une telle épreuve : nodesep=0mm,levelsep=20mm,treesep=10mm [treemode=R] [tnpos=r]S$p$ [tnpos=r]$\overline{S}$$1-p$

La variable aléatoire associée à cette expérience et qui prend la valeur 0 en cas d'échec et 1 en cas de succès est appelée **variable aléatoire de Bernoulli**.
La loi de probabilité de cette variable aléatoire est appelée **loi de Bernoulli**.
On note alors $X \sim \mathcal{B}(p)$ (« X suit la loi de Bernoulli de paramètre $p$ »).

| $x_i$ | 0 | 1 |
|---|---|---|
| $\mathbf{P}(X = x_i)$ | $1-p$ | $p$ |

**Remarques :**

- Une épreuve de Bernoulli sert à modéliser une expérience aléatoire dont on ne considère que deux issues.
  Ainsi, un lancer d'un dé à 20 faces, où on s'intéresse à la réalisation de l'événement S : « on obtient 13 » est une épreuve de Bernoulli, dont le paramètre est $p = \frac{1}{20}$ si ce dé est équilibré.
  Sonder une personne d'une population humaine en lui demandant si elle va voter oui ou non à un référendum est aussi une épreuve de Bernoulli : on convient qu'on a un succès si cette personne sondée répond « oui » et un échec si la personne sondée répond « non ».
- La notation $X \sim \mathcal{B}(p)$ n'est pas officielle, certains préfèrent $X \hookrightarrow \mathcal{B}(p)$, $X \rightsquigarrow \mathcal{B}(p)$, *etc.* .
- De nombreux manuels et formulaires notent $q = 1-p$.

**Propriété 5**

Pour une variable aléatoire $X \sim \mathcal{B}(p)$, X possède une espérance et une variance et

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

## 5 Schéma de Bernoulli et loi binomiale

**Définition 6**

On considère une épreuve de Bernoulli de paramètre $p \in [0;1]$. En répétant cette expérience aléatoire, de manière indépendante, $n \in \mathbb{N}^{\times}$ fois, on obtient un **schéma de Bernoulli**.

On a donné ci-contre une représentation, sous forme d'un arbre de répétition, de ce schéma pour $n = 3$ où S est l'événement « obtenir un succès à l'épreuve de Bernoulli ».

La loi de probabilité d'une variable aléatoire X qui compte après la réalisation de $n$ épreuves le nombre de succès se nomme **la loi binomiale de paramètres $n$ et $p$.**

On note alors $X \sim \mathcal{B}(n;p)$ (« X suit la loi binomiale de paramètres $n$ et $p$ »).

La loi binomiale de paramètres 1 et $p$ est la loi de Bernoulli de paramètre $p$.

nodesep=0mm,levelsep=20mm,treesep=10mm [treemode=R] [tnpos=a]S$p$ [tnpos=a]S$p$ [tnpos=r]S$p$ [tnpos=r]$\overline{S}1-p$ [tnpos=a]$\overline{S}1-p$ [tnpos=r]S$p$ [tnpos=r]$\overline{S}1-p$ [tnpos=a]$\overline{S}1-p$ [tnpos=a]S$p$ [tnpos=r]S$p$ [tnpos=r]$\overline{S}1-p$ [tnpos=a]$\overline{S}1-p$ [tnpos=r]S$p$ [tnpos=r]$\overline{S}1-p$

**Remarque :**

Une loi binomiale correspond à une épreuve de Bernoulli réalisée plusieurs fois, sous les mêmes conditions.

Lancer 4 fois un dé équilibré à 20 faces et s'intéresser à l'événement S : « obtenir 13 » correspond à loi binomiale de paramètres 4 et $\frac{1}{20}$.

Pour le sondage pour un référendum, il existe un petit écueil. En effet, *a priori*, comme un sondeur ne ré-interroge pas plusieurs un même individu, on ne peut pas parler d'indépendance (il faudrait tout le temps choisir dans la même situation d'avant la première question, sans exclusion, car sinon, les probabilités changent : penser à la situation extrême d'une population de 10 individus, 5 pour, 5 contre ; si on interroge en excluant, et si par exemple on interroge les 5 pour, alors la probabilité d'interroger un pour dans ceux qui restent est de 0). Ces situations, dites « sans remise », sont modélisées par une loi plus complexe, la loi hypergéométrique.

Néanmoins, dans le cas d'un sondage, dès qu'on a une « grande » population, et que l'on ne l'interroge pas plus de 10% de celle-ci, on supposera que les tirages sont indépendants et donc qu'on est dans le cadre de la loi binomiale (et pour les arrondis faits pour un sondage, l'erreur est négligeable devant la probabilité).

# 6 Coefficient binomiaux

## 6.1 Définition

**Exemple :** On lance trois fois de suite le même dé équilibré à 12 faces, et on s'intéresse à la sortie du 7.
On note X la variable aléatoire comptant le nombre de « 7 » après les trois lancers.
Lancer le dé et observer l'apparition du « 7 » est une épreuve de Bernoulli, de paramètre $\frac{1}{12}$ car le dé est équilibré. Cette épreuve est répétée de manière indépendante 3 fois. La variable aléatoire X suit ainsi la loi binomiale de paramètres 3 et $\frac{1}{12}$.
On veut calculer $\mathbf{P}(\mathrm{X}=1)$.
En représentant le schéma de Bernoulli associée à cette expérience aléatoire, trois chemins amènent à un succès (et donc deux échecs) : $\mathrm{S}\overline{\mathrm{S}}\overline{\mathrm{S}}$ ; $\overline{\mathrm{S}}\mathrm{S}\overline{\mathrm{S}}$ et $\overline{\mathrm{S}}\overline{\mathrm{S}}\mathrm{S}$.

En outre, $\begin{cases} \mathbf{P}(\text{« } \mathrm{S}\overline{\mathrm{S}}\overline{\mathrm{S}} \text{ »}) = p\times(1-p)\times(1-p) = p(1-p)^2 = \frac{1}{12}\times\left(\frac{11}{12}\right)^2 = \frac{121}{1\,728} \\ \mathbf{P}(\text{« } \overline{\mathrm{S}}\mathrm{S}\overline{\mathrm{S}} \text{ »}) = (1-p)\times p\times(1-p) = p(1-p)^2 = \frac{1}{12}\times\left(\frac{11}{12}\right)^2 = \frac{121}{1\,728} \\ \mathbf{P}(\text{« } \overline{\mathrm{S}}\overline{\mathrm{S}}\mathrm{S} \text{ »}) = (1-p)\times(1-p)\times p = p(1-p)^2 = \frac{1}{12}\times\left(\frac{11}{12}\right)^2 = \frac{121}{1\,728} \end{cases}$.

Les probabilités de chacun de réalisation des chemins amenant un succès sont toutes égales, donc $\mathbf{P}(\mathrm{X}=1)$ est égale au nombre de chemins multiplié par cette probabilité commune, d'où $\mathbf{P}(\mathrm{X}=1) = 3\times\frac{121}{1\,728} = \frac{121}{576}$.
Pour cette valeur faible de $n$, on a pu déterminer le nombre de chemins en réalisant l'arbre mais dès que $n \geqslant 5$, cela devient trop peu pratique. On va chercher un moyen de déterminer le nombre de chemins sans avoir à tracer l'arbre.

**Définition 7**

Pour un schéma de Bernoulli de paramètres $p \in [0;1]$ et $n \in \mathbb{N}^\times$, et pour $k \in [\![0;n]\!]$, le nombre de chemins réalisant $k$ succès, et donc $n-k$ échecs, est **le coefficient binomial**, noté $\binom{n}{k}$.

**Propriété 6**

Pour $n \in \mathbb{N}^\times$ et $k \in [\![0;n]\!]$, on a

- $\binom{n}{0} = \binom{n}{n} = 1$.
- $\binom{n}{1} = \binom{n}{n-1} = n$.
- $\binom{n}{k} = \binom{n}{n-k}$ (Symétrie des coefficients binomiaux).
- $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$ (Relation de Pascal).

**Remarque :** À l'aide de la calculatrice, $\binom{n}{k} = n$ `Combinaison` $k$, « `Combinaison` » étant dans « `math` » puis « `PROB` ».

## 6.2 Calculs à l'aide des coefficients binomiaux

‡ **Propriété 7**
Pour un schéma de Bernoulli de paramètres $p \in [0;1]$ et $n \in \mathbb{N}^{\times}$, et pour $k \in [\![0;n]\!]$, la probabilité d'une liste comportant $k$ succès, et donc $n-k$ échecs, ne dépend pas de l'ordre de la liste.

‡ **Propriété 8**
Pour $X \sim \mathcal{B}(n;p)$, avec $n \in \mathbb{N}^{\times}$, $p \in [0;1]$ et $k \in [\![0;n]\!]$, on a

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

$$\mathbf{P}(X=k) = \underbrace{\binom{n}{k}}_{\text{nombre de chemins amenant } k \text{ succès}} \times \underbrace{p^k}_{\text{probabilité d'un succès}^{\text{nombre de succès}}} \times \underbrace{(1-p)^{n-k}}_{\text{probabilité d'un échec}^{\text{nombre d'échecs}}}$$

‡ **Propriété 9**
Pour une variable aléatoire X suivant la loi binomiale de paramètres $n \in \mathbb{N}^{\times}$ et $p \in [0;1]$, X admet une espérance et une variance et

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

# 7 Représentations graphiques d'une loi binomiale

On donne ci-dessous des diagrammes en bâtons, avec les mêmes échelles, correspondant à des lois binomiales.

rctzbb0.10980392156862745 0.2235294117647059 0.7333333333333333 qqqqffo. 0. 1. xunit=0.6cm,yunit=20.0cm,algebraic=true,dimen=middle,dotstyle=o,dotsize=3pt 0,linewidth=0.8pt,arrowsize=3pt 2,arrowinset=0.25 (-1.,-0.04682295380956209)(9.563380281690133,0.32019350839999993) [labelFontSize=,xAxis=true,yAxis=true,Dx=2.,Dy=0.05,ticksize=-2pt0,subticks=2]->(0,0)(-1.,-0.04682295380956209)(9.563380281690133,0.32019350839999993)[linewidth=1.6pt,linecolor=rctzbb](0.,0)(0.,0.0282475249)[linewidth=1.6pt,linecolor=rctzbb](1.,0)(1.,0.121060821)[linewidth=1.6pt,linecolor=rctzbb](2.,0)(2.,0.2334744405)[linewidth=1.6pt,linecolor=rctzbb](3.,0)(3.,0.266827932)[linewidth=1.6pt,linecolor=rctzbb](4.,0)(4.,0.200120949)[linewidth=1.6pt,linecolor=rctzbb](5.,0)(5.,0.1029193452)[linewidth=1.6pt,linecolor=rctzbb](6.,0)(6.,0.036756909)[linewidth=1.6pt,linecolor=rctzbb](7.,0)(7.,0.009001692)[linewidth=1.6pt,linecolor=rctzbb](8.,0)(8.,0.0014467005)

$n = 10$ et $p = 0,3$

rctzbb0.10980392156862745 0.2235294117647059 0.7333333333333333 xunit=0.6cm,yunit=20.0cm,algebraic=true,dimen=middle,dotstyle=o,dotsize=3pt 0,linewidth=0.8pt,arrowsize=3pt 2,arrowinset=0.25 (13.,-0.03157981747823285)(30.68,0.22518242619244386) [labelFontSize=,xAxis=true,yAxis=true,Dx=5.,Dy=0.05,ticksize=-2pt0,subticks=2]->(0,0)(13.,-0.03157981747823285)(30.68,0.22518242619244386)[linewidth=1.6pt,linecolor=rctzbb](0.,0)(0.,0)[linewidth=1.6pt,linecolor=rctzbb](1.,0)(1.,0.)[linewidth=1.6pt,linecolor=rctzbb](2.,0)(2.,0.)[linewidth=1.6pt,linecolor=rctzbb](3.,0)(3.,0.)[linewidth=1.6pt,linecolor=rctzbb](4.,0)(4.,0.)[linewidth=1.6pt,linecolor= [illegible] rctzbb](17.,0)(17.,0.00220917930325584)[linewidth=1.6pt,linecolor=rctzbb](18.,0)(18.,0.0063820735427391)[linewidth=1.6pt,linecolor=rctzbb](19.,0)(19.,0.0161231331606041)[linewidth=1.6pt,linecolor=rctzbb](20.,0)(20.,0.0354708929533289)[linewidth=1.6pt,linecolor=rctzbb](21.,0)(21.,0.0675636056253884)[linewidth=1.6pt,linecolor=rctzbb](22.,0)(22.,0.110558627386999)[linewidth=1.6pt,linecolor=rctzbb](23.,0)(23.,0.153820698973216)[linewidth=1.6pt,linecolor=rctzbb](24.,0)(24.,0.179457482135419)[linewidth=1.6pt,linecolor=rctzbb](25.,0)(25.,0.172279182850003)[linewidth=1.6pt,linecolor=rctzbb](26.,0)(26.,0.132522448346156)[linewidth=1.6pt,linecolor=rctzbb](27.,0)(27.,0.0785318212421665)[linewidth=1.6pt,linecolor=rctzbb](28.,0)(28.,0.0336564948180714)[linewidth=1.6pt,linecolor=rctzbb](29.,0)(29.,0.00928455029464037)[linewidth=1.6pt,linecolor=rctzbb](30.,0)(30.,0.00123794003928538)

$n = 30$ et $p = 0,8$ (les valeurs $\mathbf{P}(X = 0)$ à $\mathbf{P}(X = 13)$ sont tellement faibles qu'elles sont représentées par une longueur non détectable et il a été fait le choix de tronquer la représentation)

**Remarques :**

- Pour les valeurs de $k$ proches de $\mathbf{E}(X) = np$, les probabilités cumulées $\mathbf{P}(X = k)$ sont plus élevées et correspondent aux « plus hauts bâtons ».
- Plus $n$ devient grand, plus l'allure du diagramme fait apparaître une « cloche » (appelée courbe de Gauß).

## 8 Loi binomiale et questionnement successif

On se place dans la situation suivante : on considère une population dont on étudie un caractère ; on suppose que l'on sait que le caractère est présent dans la proportion $p$.
On étudie un individu de la population : ou il présente le caractère ou il ne présente pas. Il s'agit donc d'une épreuve de Bernoulli dont la probabilité de succès est $p$.
On étudie maintenant successivement deux individus de la population. On est alors face à la situation suivante (en notant S l'événement : « l'individu présente le caractère »).

nodesep=0mm,levelsep=20mm,treesep=10mm [tree-mode=R] [tnpos=a]S$p$ [tnpos=r]S$p$ [tnpos=r]$\overline{\text{S}}1-p$ [tnpos=a]$\overline{\text{S}}1-p$ [tnpos=r]S$p$ [tnpos=r]$\overline{\text{S}}1-p$

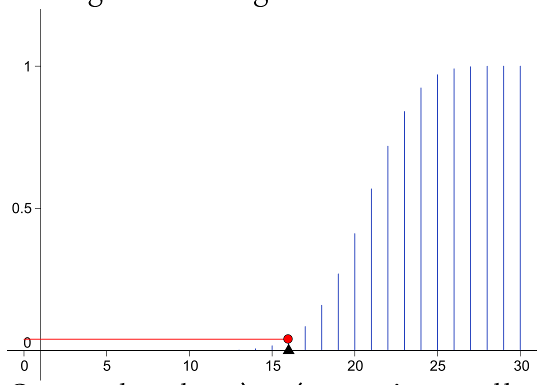
On reconnait le schéma de Bernoulli de paramètres $n = 2$ et $p$.
Ainsi, la variable aléatoire comptant le nombre d'individus présentant le caractère suit une loi binomiale de paramètres $n = 2$ et $p$ : $\mathcal{B}(2\,;p)$.

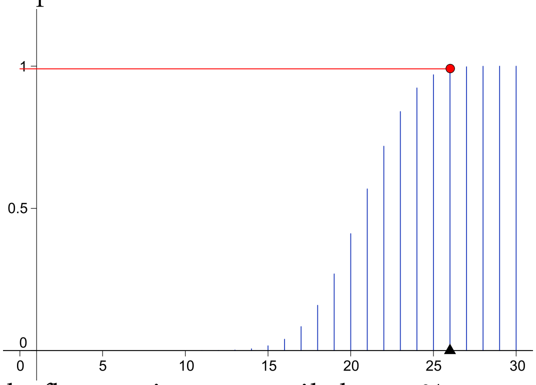
Remarque : On sous-entend que les « tirages » sont faits avec remise ; quand la population considérée est « grande », cette approximation est valable.
De manière plus précise, quand on interroge (ou qu'on prélève) moins de 10% de la population totale, l'approximation est couramment acceptée en statistique.
Maintenant, si on étudie le caractère sur $n$ individus, on aura le schéma de Bernoulli d'une loi binomiale de paramètres $n$ et $p$.

‡ **Propriété** 10
La variable aléatoire X qui compte le nombre d'individus de l'échantillon présentant le caractère suit une loi binomiale de paramètres $n$ et $p$ : $\mathcal{B}(n\,;p)$.

# 9 Intervalle de fluctuation de la loi binomiale

Exemple :
Selon les chiffres du ministère de la Santé, 70 % des français ne fument pas.
Dans un lycée, on interroge 30 lycéens et 19 déclarent ne pas fumer. On voudrait savoir maintenant si on peut considérer que ces lycéens fument dans la même proportion que la population française.
On suppose donc que la proportion vaut $p = 70\% = 0,7$.
On note X la variable aléatoire qui compte le nombre de lycéens ne fumant pas, alors X suit une loi $\mathcal{B}(30\,;0,7)$.
On va s'intéresser à la variable aléatoire $\mathrm{F} = \dfrac{\mathrm{X}}{n}$, qui correspond elle à la fréquence observée dans l'échantillon.
On regarde le diagramme en bâtons des probabilités cumulées :




On va chercher à créer un intervalle de fluctuation au seuil de 95 % : pour cela, il faut donc « enlever » 5% des valeurs : on enlève donc les $2,5\% = 0,025$ du « bas » et de même en « haut ».
Ainsi, graphiquement, on trace les droites correspondant à $0,025$ et $1 - 0,025 = 0,975$ et entre ces deux traits, on trouve l'intervalle de fluctuation au seuil de confiance 0,95 recherché : on trouve ici apparemment $\left[\frac{16}{30}\,;\frac{26}{30}\right]$.
Pour avoir un résultat plus précis, on utilise un tableur pour obtenir les probabilités cumulées $\mathbf{P}(\mathrm{X} \leqslant k)$ de la loi $\mathcal{B}(30\,;0,7)$ :

| Nombre de succès $k$ | $\mathbf{P}(\mathrm{X} \leqslant k)$ |
|---|---|
| 0 | $2,058\,91 \times 10^{-16}$ |
| 1 | $1,461\,83 \times 10^{-14}$ |
| $\vdots$ | $\vdots$ |
| 14 | $0,006\,370\,346$ |
| 15 | $0,016\,937\,311$ |
| 16 | $0,040\,052\,548$ |
| 17 | $0,084\,470\,060$ |
| $\vdots$ | $\vdots$ |
| 24 | $0,923\,405\,248$ |
| 25 | $0,969\,845\,057$ |
| 26 | $0,990\,683\,433$ |
| 27 | $0,997\,886\,822$ |
| $\vdots$ | $\vdots$ |
| 30 | 1 |

On cherche alors la première valeur de $k$ (sur la colonne de gauche) telle $\mathbf{P}(\mathrm{X} \leqslant k)$ (colonne de droite) soit strictement supérieure à $0,025$ ; on trouve ici 16.
Puis, on recherche la première valeur de $k$ telle $\mathbf{P}(\mathrm{X} \leqslant k)$ (colonne de droite) soit supérieure ou égale à $0,975$ ; on trouve ici 26.
On en déduit donc l'intervalle de fluctuation :
$\left[\frac{16}{30}\,;\frac{26}{30}\right]$.

On peut retrouver ce tableau de valeurs grâce à la calculatrice en entrant « `binomFRép(N,P,X)` » dans la partie fonction (`f(x)=`) puis en faisant défiler les valeurs de `X` dans la partie `TABLE` (en veillant à bien la configurer (menu `DéfTable`), par exemple : `DébTable=0 ; Pas=1 ; Auto ; Auto`).

Voici le résultat général :

‡ **Propriété 11**

On étudie un caractère au sein d'une population ; on suppose qu'il est présent en une proportion $p$.
Sur un échantillon de $n$ individus, le caractère est présent en une fréquence observée $\mathbf{f}$.
La variable aléatoire X qui compte le nombre d'individus présentant le caractère suit une loi binomiale $\mathcal{B}(n\,;p)$.
L'intervalle de fluctuation au seuil de confiance $0,95$ de la variable aléatoire $\mathrm{F} = \frac{\mathrm{X}}{n}$ est $\left[\frac{a}{n}\,;\frac{b}{n}\right]$ où $a$ et $b$ sont déterminés par :

- $a$ est le plus petit entier tel que $\mathbf{P}(\mathrm{X} \leqslant a) > 0,025$.
- $b$ est le plus petit entier tel que $\mathbf{P}(\mathrm{X} \leqslant b) \geqslant 0,975$.

Ainsi, la fréquence observée $\mathbf{f}$ appartient à l'intervalle $\left[\frac{a}{n}\,;\frac{b}{n}\right]$ avec une probabilité de $0,95$.

Remarque : Sous les conditions suivantes : $\begin{cases} n \geqslant 30 \\ np \geqslant 5 \\ n(1-p) \geqslant 5 \end{cases}$,
l'intervalle de fluctuation au seuil de confiance $0,95$ est quasiment le même que celui de seconde : $\left[p - \frac{1}{\sqrt{n}}\,;p + \frac{1}{\sqrt{n}}\right]$.

Remarque : Quand « $n$ est grand » ($n > 100$), les approximations :

- $a \approx np - 2\sqrt{np(1-p)} = \mathbf{E}(\mathrm{X}) - 2\sigma(\mathrm{X})$
- $b \approx np + 2\sqrt{np(1-p)} = \mathbf{E}(\mathrm{X}) + 2\sigma(\mathrm{X})$

  permettent de trouver plus rapidement les valeurs recherchées.

## 10 Règle de décision

‡ **Propriété 12**

On étudie un caractère au sein d'une population. On fait l'hypothèse : « la proportion théorique du caractère est $p$ »(on ne connait donc pas la valeur de $p$, on fait juste une hypothèse).
Sur un échantillon de $n$ individus, le caractère est présent en une fréquence observée $f$.
Alors, au seuil de confiance $0,95$ :
• Si $\mathbf{f}$ appartient à l'intervalle de fluctuation $\left[\frac{a}{n}\,;\frac{b}{n}\right]$ : $\mathbf{f} \in \left[\frac{a}{n}\,;\frac{b}{n}\right]$, alors les données ne nous permettent pas de rejeter l'hypothèse.
• Si $\mathbf{f}$ n'appartient pas à l'intervalle de fluctuation $\left[\frac{a}{n}\,;\frac{b}{n}\right]$ : $\mathbf{f} \notin \left[\frac{a}{n}\,;\frac{b}{n}\right]$, alors on rejette l'hypothèse « la proportion théorique du caractère est $p$ » au risque d'erreur de 0,05.

**Remarque :** Il est important de noter que l'on ne connaît la probabilité de se tromper (le risque d'erreur) uniquement dans le cas où $\mathbf{f} \notin \left[\frac{a}{n}\,;\frac{b}{n}\right]$, *i.e.* la cas où on rejette l'hypothèse.

On reprend l'exemple du lycée où on s'intéressait à la proportion de non-fumeurs.
On a fait l'hypothèse « la proportion théorique de non-fumeurs est $p = 0,7$ ». On travaille au niveau de confiance 0,95.
La fréquence observée est $\mathbf{f} = \frac{19}{30} \approx 0,63$.
Or on a établi que l'intervalle de fluctuation au seuil de 0,95 est : $\left[\frac{16}{30}\,;\frac{26}{30}\right]$.

On en déduit que $\mathbf{f} = \frac{19}{30} \in \left[\frac{16}{30}\,;\frac{26}{30}\right]$ et donc on est dans le cas où on ne peut pas rejeter l'hypothèse.
Si dans un autre lycée, sur 30 élèves, 11 déclarent ne pas fumer, on a cette fois une fréquence observée de $\mathbf{f} = \frac{11}{30} \notin \left[\frac{16}{30}\,;\frac{26}{30}\right]$ et ici on peut dire avec 5% de chance de se tromper que dans ce lycée, le taux de fumeur est significativement différent que dans le reste de la population française.