Parcours probabilités au lycée

Gwénaëlle Castellan - Anne Keller

Jeudi 30 janvier 2020

Au menu de l'atelier

1. Parcours probabilités : de l'approche fréquentiste (cycle 4) à la loi faible des grands nombres (terminale)
2. Inégalité de Bienaymé-Tchebychev, motivation historique, démonstrations et applications : inégalité de concentration
3. Quelques exemples d'exercices d'application

Survol des principales modifications des programmes

Des déplacements

- Les probabilités conditionnelles et l'indépendance passent de la terminale à la première.
- La loi binomiale passe de la première à la terminale.

Des nouveautés

- Somme de variables aléatoires.
- Inégalité de Bienaymé-Tchebychev, inégalité de concentration, loi des grands nombres.

Des disparitions

- Intervalle de fluctuation, intervalle de confiance.
  L'étude des fluctuations prend une autre forme.
- Lois à densité.

Approche fréquentiste et simulations au cycle 4

Programmes de cycle 4 - Compétence

Faire le lien entre fréquence et probabilité.

Repères de progression - En 3e

Le constat de la stabilisation des fréquences s'appuie sur la simulation d'expériences aléatoires à une épreuve à l'aide d'un tableur ou d'un logiciel de programmation [...]

Attendus de fin de 3e - Ce que sait faire l'élève

À partir de dénombrements, il calcule des probabilités pour des expériences aléatoires simples à une ou deux épreuves. Il fait le lien entre stabilisation des fréquences et probabilités.

Exemples de réussite en fin de 3^{e}

- On suppose que, pour un couple, la probabilité d'avoir une fille ou un garçon est la même. Un couple souhaite avoir deux enfants.

  Calcule, en explicitant les issues possibles, la probabilité d'avoir deux garçons.

  Calcule la probabilité que le couple ait au moins une fille. Il peut utiliser le fait que c'est l'événement contraire d'avoir deux garçons.
- On tire, deux fois de suite et avec remise, une boule dans une urne contenant une boule bleue et deux boules violettes.

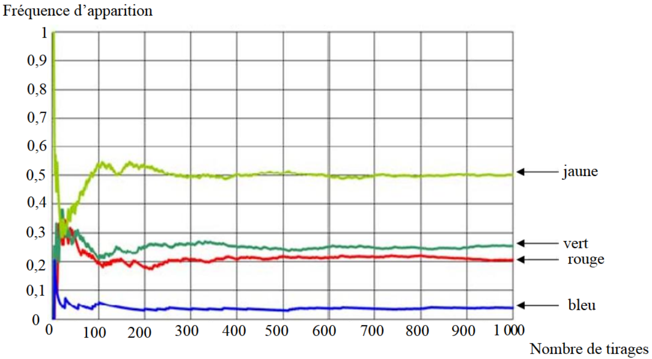
  Détermine la probabilité de tirer successivement deux boules violettes, en utilisant une méthode de dénombrement prenant appui sur un tableau à double entrée.
- On donne les fréquences d'apparition de chaque face d'un dé pour 10000 lancers.

  L'élève interprète les résultats en les comparant aux probabilités théoriques.
- *L'élève interprète des simulations effectuées sur tableur ou logiciel de programmation en fonction d'un nombre de lancers.*

Exemple : Exercice 4 du DNB 2014

Un sac contient 20 jetons qui sont soit jaunes, soit verts, soit rouges, soit bleus. On considère l'expérience suivante : tirer au hasard un jeton, noter sa couleur et remettre le jeton dans le sac. Chaque jeton a la même probabilité d'être tiré.

1) Le professeur, qui connaît la composition du sac, a simulé un grand nombre de fois l'expérience avec un tableur. Il a représenté ci-dessous la fréquence d'apparition des différentes couleurs en fonction du nombre de tirages.



a) Quelle couleur est la plus présente dans le sac ? Aucune justification n'est attendue.

En seconde : Modéliser le hasard, calculer des probabilités

L'ensemble des issues est fini.

Peu de changements par rapport aux anciens programmes.

- Notion de loi (distribution) de probabilité.
- Distinction entre modèle probabiliste et situation réelle.

En seconde : Échantillonnage (contenus)

L'objectif est de faire percevoir, sous une forme expérimentale, la loi des grands nombres, la fluctuation d'échantillonnage et le principe de l'estimation d'une probabilité par une fréquence observée sur un échantillon.

- Échantillon aléatoire de taille n pour une expérience à deux issues.
- Version vulgarisée de la loi des grands nombres : « Lorsque n est grand, sauf exception, la fréquence observée est proche de la probabilité. »
- Principe de l'estimation d'une probabilité, ou d'une proportion dans une population, par une fréquence observée sur un échantillon.

En seconde : Échantillonnage (capacités attendues)

- Lire et comprendre une fonction Python renvoyant le nombre ou la fréquence de succès dans un échantillon de taille n pour une expérience aléatoire à deux issues.
- Observer la loi des grands nombres à l'aide d'une simulation sur Python ou tableur.
- Simuler N échantillons de taille n d'une expérience aléatoire à deux issues. Si p est la probabilité d'une issue et f sa fréquence observée dans un échantillon, calculer la proportion des cas où l'écart entre p et f est inférieur ou égal à $\frac{1}{\sqrt{n}}$.

En première : Probabilités conditionnelles et indépendances

Dans cette partie, on retrouve des notions qui relevaient auparavant essentiellement du programme de terminale :

- Probabilité conditionnelle, indépendance.
- Arbres pondérés : règle du produit et de la somme.
- Formule des probabilités totales.
- Succession de deux épreuves indépendantes. Représentation par un arbre ou un tableau.

En première : Variables aléatoires réelles

Le programme ne considère que des univers finis.

- Variable aléatoire réelle : formalisation comme fonction définie sur l'univers.
- Loi d'une variable aléatoire.
- Espérance, variance et écart type d'une variable aléatoire.

La loi de Bernoulli et la loi binomiale ne sont plus vues en première.

En première : Expérimentations

- Simuler une variable aléatoire avec Python.

k	1000	1050	1100
$P(X=k)$	0,6	0,2	0,2

$$\mu = E(X) = 1030 \quad \text{et} \quad \sigma = \sigma(X) = 40$$

- Fonction Python renvoyant la moyenne d'un échantillon de taille n d'une variable aléatoire.

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- Avec Python ou un tableur, simuler N échantillons de taille n d'une variable aléatoire d'espérance μ et d'écart type σ.
  Si m désigne la moyenne d'un échantillon, proportion des échantillons où l'écart entre m et μ est inférieur ou égal à $\frac{2\sigma}{\sqrt{n}}$.

En terminale : idée générale

On étudie une variable aléatoire d'espérance μ et de variance σ^2.

- On s'intéresse à la somme et à la moyenne d'un échantillon $(X_1, \ldots, X_n)$ de même loi que X :

$$S_n = X_1 + X_2 + \cdots + X_n \quad \text{et} \quad M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- On cherche à majorer la probabilité $P\left(|M_n - \mu| \geqslant \frac{k\sigma}{\sqrt{n}}\right)$.
- Objectifs :
  - Approfondir le sens de l'écart type comme mesure de la dispersion.
  - Couronner la partie *Probabilités* par la loi des grands nombres.

Comparaison ancien et nouveau programme

Ancien programme	**Nouveau programme**
X suit une loi de Bernoulli	X est quelconque
S_n suit une loi binomiale	S_n est quelconque
Approximation asymptotique par la loi normale (théorème de Moivre-Laplace)	Majoration par l'inég. de concentration via l'inég. de Bienaymé-Tchebychev

Simulations : Progression au lycée

On simule N échantillons de taille n.

2de	Exp. aléatoire à deux issues : proportion des cas où $\lvert f_n - p\rvert \leqslant \frac{1}{\sqrt{n}}$ *La variable aléatoire suit une loi de Bernoulli de paramètre* p, *on observe des proportions de l'ordre de* $0{,}95$ *ou plus* ($\sigma \leqslant \frac{1}{2}$).
1re	V. a. quelconque : proportion des cas où $\lvert m - \mu\rvert \leqslant \frac{2\sigma}{\sqrt{n}}$ *On observe des proportions de l'ordre de* $0{,}95$.
Tle	V.a. quelconque : proportion des cas où $\lvert m - \mu\rvert \leqslant \frac{k\sigma}{\sqrt{n}}$, $k = 1, 2, 3$ *L'inég. de concentration permet de majorer* $P\left(\lvert M_n - \mu\rvert \geqslant \frac{k\sigma}{\sqrt{n}}\right)$ *par* $\frac{1}{k^2}$ *Dans le cas de la loi binomiale, on calcule numériquement* $P\left(\lvert S_n - pn\rvert \geqslant \sqrt{n}\right)$.

Simulation en seconde

On simule un lancer de dé.
On compte le nombre de fois où on ne tire pas un 6 ($p = \frac{5}{6} \approx 0{,}83$).

```python
def valeur():
# 1 si on ne tire pas un 6,
# 0 sinon.
    if randint(1,6)!=6:
        return 1
    else:
        return 0
```

```python
def echantillon(n):
# renvoie un échantillon
# de taille n
    L = []
    for i in range(n):
        L = L + [valeur()]
    return L
```

```python
def moyenne(L):
# renvoie la moyenne
# de l'échantillon
    S = 0
    n = 0
    for X in L:
        S = S + X
        n = n + 1
    return S / n
```

```python
def proportion(n,N):
# simule N échantillons
# et renvoie la proportion d'échantillon
# pour lesquels l'écart entre
# la fréquence et la probabilité
# est inférieur à 1/sqrt(n)
    Nb = 0
    for i in range(N):
        m = moyenne(echantillon(n))
        if 5/6-1/sqrt(n) <= m <= 5/6+1/sqrt(n):
            Nb = Nb + 1
    return Nb/N
```

On peut utiliser les fonctions

Python `len`, `sum`, `mean`...

Simulation en seconde : résultats

Moyenne de chaque échantillon

```
>>> moyenne(echantillon(100))
0.81
>>> moyenne(echantillon(100))
0.76
>>> moyenne(echantillon(100))
0.82
>>> moyenne(echantillon(100))
0.82
>>> moyenne(echantillon(100))
0.9
>>> moyenne(echantillon(100))
0.86
```

Pour 200 échantillons, proportion de ceux pour lesquels l'écart entre la moyenne et $\frac{5}{6}$ est inférieur à $\frac{1}{\sqrt{n}} = 0{,}1$.

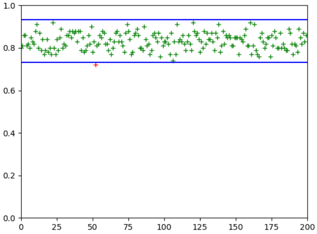
```
>>> proportion(100,200)
1.0
>>> proportion(100,200)
0.98
>>> proportion(100,200)
0.995
>>> proportion(100,200)
0.99
>>> proportion(100,200)
0.995
>>> proportion(100,200)
1.0
```

Simulation en seconde : graphique 1

```
def graphe(n,N): #  calcul de la proportion avec graphique
    Nb = 0
    for i in range(N):
        m = moyenne(echantillon(n))
        if 5/6-1/sqrt(n) <= m <= 5/6+1/sqrt(n):
            plot(i,m, '+',color='green')
            Nb = Nb + 1
        else:
            plot(i,m, '+',color='red')
    axis([0,N,0,1])
    plot([0,N],[5/6+1/sqrt(n),5/6+1/sqrt(n)],color='blue')
    plot([0,N],[5/6 -1/sqrt(n),5/6-1/sqrt(n)],color='blue')
    show()
    return Nb/N
```

Pour 200 échantillons de taille 100 :

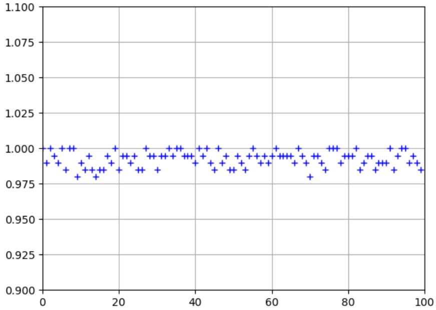
```
>>> graphe(100,200)
0.995
```



Simulation en seconde : graphique 2

```
def graphe_proportion(n,N,E): # répétitions sur plusieurs séries de N
     échantillons
    for i in range(E):
        p = proportion(n,N)
        plot(i,p,'+',color='blue')
    grid(True)
    axis([0,E,0.9,1.1])
    show()
```

Pour 100 séries de 200 échantillons de taille 100 :



Simulation en première

k	1000	1050	1100
$P(X=k)$	0,6	0,2	0,2

$$\mu = E(X) = 1030 \quad \text{et} \quad \sigma = \sigma(X) = 40$$

```python
def valeur():
# Renvoie une valeur
      aléatoire.
    P = random()
    if P <= 0.6:
        return 1000
    elif P <= 0.8:
        return 1050
    else:
        return 1100
```

```python
def proportion(n,N):
# simule N échantillons
# et renvoie la proportion d'échantillon
# pour lesquels l'écart entre
# la moyenne et l'espérance
# est inférieur à 2*sigma/sqrt(n)
    Nb = 0
    for i in range(N):
        m = moyenne(echantillon(n))
        if 1030-2*40/sqrt(n) <= m <= 1030+2*40/sqrt(n):
            Nb = Nb + 1
    return Nb/N
```

Simulation en première : résultats

Moyenne de chaque échantillon

```
>>> moyenne(echantillon(100))
1024.5
>>> moyenne(echantillon(100))
1036.0
>>> moyenne(echantillon(100))
1030.0
>>> moyenne(echantillon(100))
1035.0
>>> moyenne(echantillon(100))
1029.0
>>> moyenne(echantillon(100))
1029.0
```

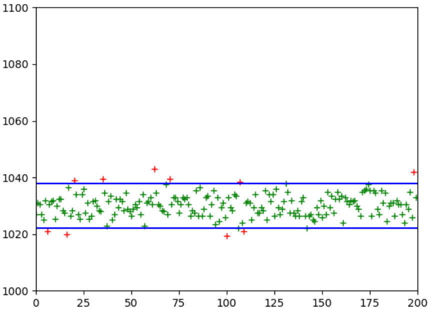
Pour 200 échantillons, proportion de ceux pour lesquels l'écart entre la moyenne et μ est inférieur à $\frac{2\sigma}{\sqrt{n}} = 0{,}8$.

```
>>> proportion(100,200)
0.955
>>> proportion(100,200)
0.98
>>> proportion(100,200)
0.97
>>> proportion(100,200)
0.96
>>> proportion(100,200)
0.95
>>> proportion(100,200)
0.935
```

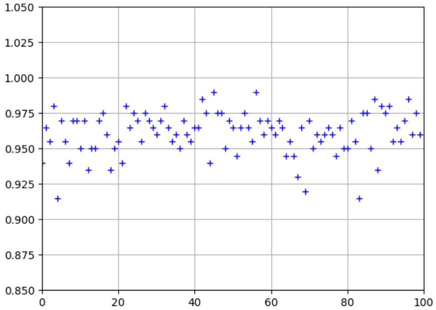
Simulation en première : graphiques

Pour 200 échantillons de taille 100 :

```
>>> graphe(100,200)
0.95
```



Pour 100 séries de 200 échantillons de taille 100 :

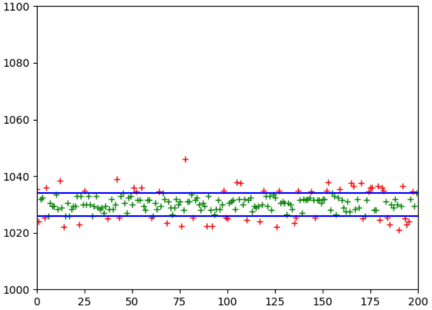


Simulation en terminale : graphiques

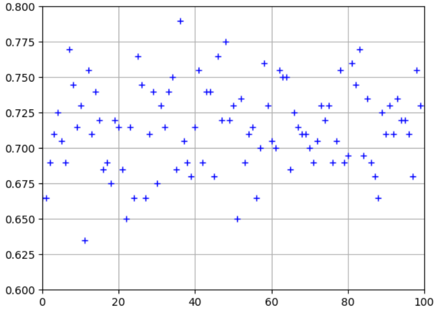
Proportion des échantillons pour lesquels l'écart entre la moyenne et μ est inférieur à $\frac{k\sigma}{\sqrt{n}}$, avec $k = 1$.

Pour 200 échantillons de taille 100 :

```
>>> graphe(100,200)
0.705
```



Pour 100 séries de 200 échantillons de taille 100 :



Dénombrement et loi binomiale

Combinatoire et denombrement

- Fait l'objet d'une sous-section dans la partie Algèbre.
- Définition des combinaisons.

Loi binomiale

- Schéma de Bernoulli.
- Loi binomiale, expression à l'aide des coefficients binomiaux.

Somme de variables aléatoires

On se place dans le cas de variables aléatoires finies.

- Somme de deux variables aléatoires. Linéarité de l'espérance : $E(X+Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$.
- Dans le cadre de la succession d'épreuves indépendantes, variables indépendantes et relation $V(X+Y) = V(X) + V(Y)$.
  Relation $V(aX) = a^2 V(X)$.
- Application : espérance, variance et écart type de la loi binomiale.
- Pour un échantillon de taille n d'une loi de probabilité :
  liste $(X_1, \ldots, X_n)$, espérance, variance et écart type de la somme $S_n = X_1 + \cdots + X_n$ et de la moyenne $M_n = \frac{S_n}{n}$.

Somme de variables aléatoires : un exemple

Exercice

On lance un dé à six faces. X est la variable aléatoire égale au nombre obtenu, Y est la variable aléatoire égale à 3 si le résultat du dé est impair et 0 sinon.
Quelle est la loi de $X+Y$? Quelle est l'espérance de $X+Y$?

Issue	1	2	3	4	5	6
Valeur de X	1	2	3	4	5	6
Valeur de Y	3	0	3	0	3	0
Valeur $X+Y$	4	2	6	4	8	6

Valeur z_k de $X+Y$	2	4	6	8
Probabilité $P(X+Y=z_k)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

$$E(X+Y)=\frac{1}{6}\times 2+\frac{1}{3}\times 4+\frac{1}{3}\times 6+\frac{1}{6}\times 8=5$$

Somme de variables aléatoires : justification de la linéarité

La démonstration de la linéarité de l'espérance nécessite de formaliser les variables aléatoires comme des fonctions sur l'univers et d'utiliser l'expression de l'espérance comme moyenne pondérée sur l'ensemble des issues. Le professeur peut choisir de l'admettre, ou de la justifier sur un exemple.

Issue	1	2	3	4	5	6
Valeur de X	1	2	3	4	5	6
Valeur de Y	3	0	3	0	3	0
Valeur $X+Y$	4	2	6	4	8	6

$$\begin{aligned}
E(X+Y) &= \tfrac{1}{6}\times 4 + \tfrac{1}{6}\times 2 + \cdots + \tfrac{1}{6}\times 6 \\
&= \tfrac{1}{6}\times (1+3) + \tfrac{1}{6}\times (2+0) + \cdots + \tfrac{1}{6}\times (6+0) \\
&= \tfrac{1}{6}\times 1 + \tfrac{1}{6}\times 2 + \cdots + \tfrac{1}{6}\times 6 + \tfrac{1}{6}\times 3 + \tfrac{1}{6}\times 0 + \cdots + \tfrac{1}{6}\times 0 \\
&= E(X) + E(Y)
\end{aligned}$$

1. Parcours probabilités : de l'approche fréquentiste (cycle 4) à la loi faible des grands nombres (terminale)
2. Inégalité de Bienaymé-Tchebychev, motivation historique, démonstrations et applications : inégalité de concentration
3. Quelques exemples d'exercices d'application

D'après un travail de Charles Suquet.

Aperçu historique

I. J. Bienaymé, Considérations A l'appui de la découverte de Laplace sur la loi de probabilité dans la méthode des moindres carrés, *Comptes rendus des séances de l'Académie des Sciences* t. XXXVII, séance du 29 août 1853.

« Dans la séance du 8 août 1853, [...]. M. Cauchy avait nié l'exactitude du résultat si remarquable, découvert et démontré par Laplace, et qui consiste en ce que la méthode des moindres carrés s'applique aux données des observations, quelle que soit la loi de probabilité des erreurs. ».

« Deux choses en effet ont toujours surpris, et surprendront toujours, quand, pour la première fois, on en vient à considérer le résultat final qu'il a livré aux observateurs. C'est, d'une part, que l'intégrale $\frac{2}{\sqrt{2\pi}}\int_0^a \mathrm{d}t e^{-t^2}$ se reproduit sans cesse comme expression approchée de la probabilité ; et de l'autre, que les écarts ou les erreurs sont proportionnels à la limite de cette intégrale, et à une fonction de la moyenne des carrés des différences entre chaque erreur possible et sa moyenne. C'est là tout ce qu'il reste, dans l'approximation de Laplace, du caractère primitif de la fonction de probabilité qui régnait pendant les observations ».

Théorème (Tchébychef, 1867)

Si les espérances mathématiques des quantités $x, y, z, \dots$ *et* $x^2, y^2, z^2, \dots$ *sont respectivement* $a, b, c, \dots, a_1, b_1, c_1, \dots,$ *la probabilité que la différence entre la moyenne arithmétique des* N *quantités* $x, y, z, \dots,$ *et la moyenne arithmétique des espérances mathématiques de ces quantités ne surpassera pas* $\dfrac{1}{t}\sqrt{\dfrac{a_1+b_1+c_1+\dots}{N} - \dfrac{a^2+b^2+c^2+\dots}{N}}$ *sera toujours plus grande que* $1 - \dfrac{t^2}{N}$ *quel que soit* t.

Théorème (Loi faible des grands nombres, Tchébychef, 1867)

Si les espérances mathématiques des quantités $U_1, U_2, U_3, \dots$ *et de leurs carrés* $U_1^2, U_2^2, U_3^2, \dots,$ *ne dépassent pas une limite finie quelconque, la probabilité que la différence entre la moyenne arithmétique d'un nombre* N *de ces quantités et la moyenne arithmétique de leurs espérances mathématiques, sera moindre qu'une quantité donnée, se réduit à l'unité, quand* N *tend vers l'infini.*

On propose deux démonstrations de l'inégalité de Bienaymé-Tchebychev.

1. Une preuve niveau Terminale qui s'inspire du texte initial de Bienaymé (1853), avec des notations plus lisibles de nos jours. Voir la version « lissée » pour une exploitation en classe ;
2. Une preuve graphique presque muette.

Dans chaque preuve, on utilise comme une simplification d'écritures l'inégalité de Markov ($P(Y > t) \leq \frac{\mathbf{E}\,Y}{t}$) qui est de toutes façons sous-jacente à l'inégalité de Bienaymé-Tchebychev. Ceci renvoie à la définition :

$$\mathbf{V}(X) = \mathbf{E}\left((X - \mathbf{E}\,X)^2\right). \tag{1}$$

La formule suivante

$$\mathbf{V}(X) = \sum_{j=1}^{k} (x_j - \mathbf{E}\,X)^2 P(X = x_j) = \sum_{x \in X(\Omega)} (x - \mathbf{E}\,X)^2 P(X = x), \tag{2}$$

où $X(\Omega) = \{x_1, \ldots, x_k\}$ désigne l'ensemble des valeurs prises par X, n'est qu'une formule de calcul obtenue à partir de la définition (1).

Pour ce faire on est amené à poser $Y = (X - \mathbf{E}\,X)^2$, et à remarquer que

$$\mathbf{V}\,X = \mathbf{E}\,Y = \sum_{y \in Y(\Omega)} yP(Y = y).$$

Ensuite on note que pour chaque $y \in Y(\Omega)$ il y a un ou deux $x \in X(\Omega)$ tels que $(x - \mathbf{E}\,X)^2 = y$.
S'il y en a un seul, notons le x, alors $P(Y = y) = P(X = x)$ et $yP(Y = y) = (x - \mathbf{E}\,X)^2 P(X = x)$.
S'il y en a deux, notés x' et x'', $P(Y = y) = P(X = x') + P(X = x'')$ et $yP(Y = y) = (x' - \mathbf{E}\,X)^2 P(X = x') + (x'' - \mathbf{E}\,X)^2 P(X = x'')$.
En recollant tous les morceaux on obtient la formule (2).
Notons aussi pour tout réel $t \geq 0$ l'équivalence

$$|X - \mathbf{E}\,X| \geq t \quad \text{si et seulement si} \quad Y \geq t^2,$$

d'où

$$P(|X - \mathbf{E}\,X| \geq t) = P(Y \geq t^2). \tag{3}$$

Preuve basée sur l'idée de Bienaymé

On travaille directement sur les sommes pondérées permettant de calculer la variance. Pour éviter le recours à la notation $\sum$ et aux indices, on peut les présenter sous forme de tableau. Si on choisit cette démonstration, il pourra être utile de commencer par traiter « à la main » un exemple où Y ne prend que 3 ou 4 valeurs.

		Somme des termes de la forme	pour y vérifiant
$\mathbf{V}(X)$	$=$	$yP(Y=y)$	$y \in Y(\Omega)$
	$\geq$	$yP(Y=y)$	$y \geq t^2$
	$\geq$	$t^2P(Y=y)$	$y \geq t^2$
	$=$	$t^2P(Y \geq t^2) = t^2P(\lvert X - \mathbf{E}\,X\rvert \geq t)$	

On a ainsi prouvé que $\mathbf{V}(X) \geq t^2P(|X - \mathbf{E}\,X| \geq t)$.
En lisant cette formule de droite à gauche et en divisant par $t^2 > 0$, on obtient l'inégalité de Bienaymé-Tchebychev.

Preuve presque muette 1/4

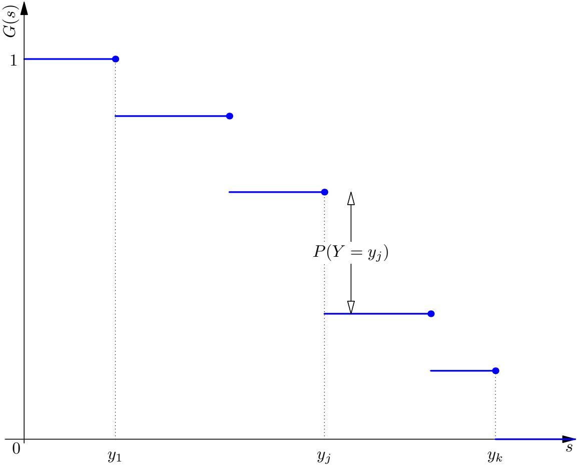


Figure – Fonction $G : s \mapsto P(Y \geq s)$, Y est la v.a. positive $(X - \mathbf{E}\,X)^2$

Preuve presque muette 2/4

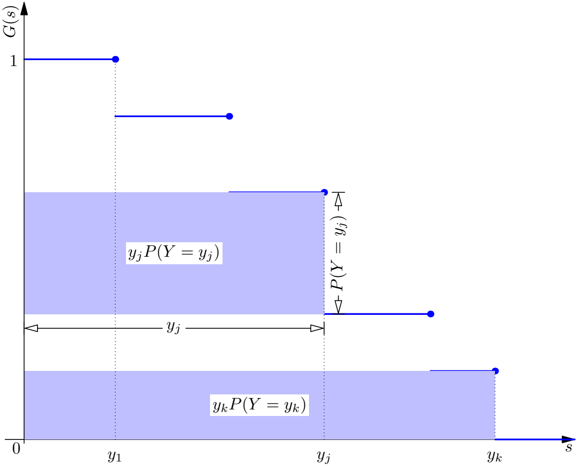


Figure – Découpage en tranches de l'aire sous l'escalier G

Preuve presque muette 3/4

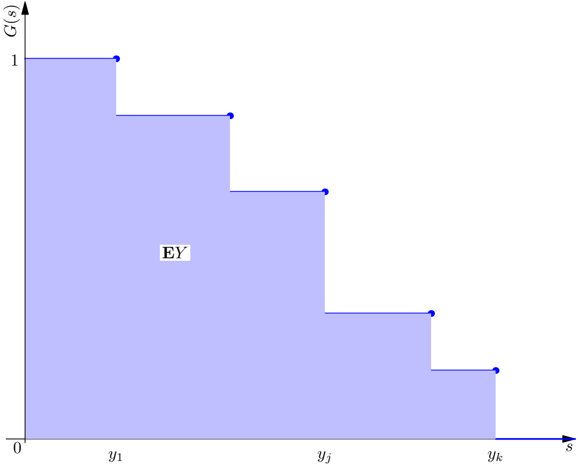


Figure – L'espérance de Y est l'aire sous l'escalier G

Preuve presque muette 4/4

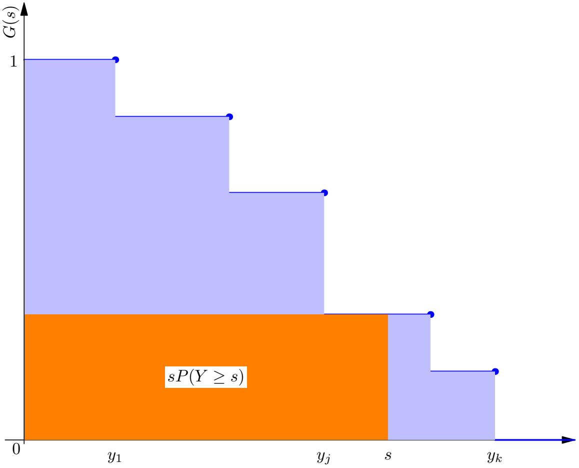


Figure – $sP(Y \geq s) \leq \mathbf{E}\,Y$, d'où avec $s = t^2$, $P(|X - \mathbf{E}\,X| \geq t) \leq \frac{\mathbf{V}(X)}{t^2}$

Pertinence de l'utilisation de l'inégalité de B.-T.

L'inégalité $P(|X - \mathbf{E}\, X| \geq t) \leq \frac{\mathbf{V}(X)}{t^2}$ est vraie pour tout réel $t > 0$, MAIS si l'on sait calculer facilement le premier membre, ce qui nécessite en général la connaissance exacte de la loi de X, elle est inutile. Elle peut être utile si

1. on connaît avec une bonne précision les valeurs de $\mathbf{E}\, X$ et de $\mathbf{V}(X)$ sans forcément connaître la loi de X (exemple : X somme ou produit de v.a. indépendantes de lois connues).
2. on connaît $\mathbf{E}\, X$ (par exemple loi symétrique autour d'une constante connue c, alors $\mathbf{E}\, X = c$) et si on sait majorer $\mathbf{V}(X)$ (exemple bien connu : X v.a. de Bernoulli de paramètre p inconnu, $\mathbf{V}(X) \leq \frac{1}{4}$, dans ce cas $\mathbf{E}\, X = p$ est inconnu mais on peut utiliser B.-T. pour un intervalle de confiance).
3. X peut s'écrire comme somme de v.a. indépendantes[1] d'espérances connues et de variances majorables facilement.

1. Seul cas étudié par Bienaymé et Tchébychef.

Petite digression sur la variance 1/2

Une identité utile

$$\forall t \in \mathbb{R}, \qquad \mathbf{E}\left((X-t)^2\right) = \mathbf{V}(X) + (t - \mathbf{E}\,X)^2. \qquad (\star)$$

Preuve. Ingrédients : linéarité de l'espérance et $\mathbf{E}\,c = c$. Posons $m = \mathbf{E}\,X$.

$$\begin{aligned}
\mathbf{E}\left((X-t)^2\right) &= \mathbf{E}\left(\left((X-m)-(t-m)\right)^2\right) \\
&= \mathbf{E}\left((X-m)^2 - 2(t-m)(X-m) + (t-m)^2\right) \\
&= \mathbf{V}(X) - 2(t-m)\underbrace{\mathbf{E}\left(X-m\right)}_{=\mathbf{E}\,X - \mathbf{E}\,m = 0} + \underbrace{\mathbf{E}\left((t-m)^2\right)}_{=(t-m)^2}.
\end{aligned}$$

Conséquences

1. En prenant $t = 0$ on obtient la formule (hors programme) $\mathbf{V}(X) = \mathbf{E}\left(X^2\right) - (\mathbf{E}\,X)^2$.
2. Pour tout réel t, $\mathbf{V}(X) \leq \mathbf{E}\left((X-t)^2\right)$.

Petite digression sur la variance 2/2

3. $\mathbf{V}(X)$ est le minimum de $t \mapsto \mathbf{E}\left((X-t)^2\right)$, atteint pour $t = m = \mathbf{E}\,X$.
4. Si $[a, b]$ est le plus petit intervalle contenant $X(\Omega)$, autrement dit, $P(a \leq X \leq b) = 1$ et $P(X = a) \neq 0$, $P(X = b) \neq 0$,

$$\mathbf{V}(X) \leq \frac{(b-a)^2}{4} \qquad \text{(inégalité optimale)}$$

Vérification du point 4. On note d'abord que si Y vérifie $0 \leq Y \leq c$ constante, alors $0 \leq \mathbf{E}\,Y \leq c$, (facile en revenant à la déf. de $\mathbf{E}\,Y$). Ensuite distance d'un réel x de $[a, b]$ au milieu du segment est au plus $(b-a)/2$. En appliquant ceci à $x = X(\omega)$ pour tout $\omega \in \Omega$, on vérifie que

$$0 \leq \left|X - \frac{a+b}{2}\right| \leq \frac{b-a}{2} \quad \text{d'où} \quad 0 \leq \left(X - \frac{a+b}{2}\right)^2 \leq \frac{(b-a)^2}{4}.$$

En appliquant le point 2 avec $t = \frac{a+b}{2}$ et en posant $Y = \left(X - \frac{a+b}{2}\right)^2$, on en déduit :

$$\mathbf{V}(X) \leq \mathbf{E}\left(\left(X - \frac{a+b}{2}\right)^2\right) \leq \frac{(b-a)^2}{4}.$$

Retour à la pertinence de l'inégalité B.-T.

On note encore $[a, b]$ le plus petit intervalle contenant $X(\Omega)$.

$$\text{L'inégalité} \quad P(|X - \mathbf{E}\, X| \geq t) \leq \frac{\mathbf{V}(X)}{t^2}$$

est vraie pour tout réel $t > 0$, MAIS est

- sans intérêt pour $0 < t \leq \sigma$ où $\sigma = \sqrt{\mathbf{V}(X)}$;
- sans intérêt pour $t > d = \max(\mathbf{E}\, X - a; b - \mathbf{E}\, X)$, plus grande des distances de $\mathbf{E}\, X$ aux bornes a et b (premier membre nul).

Comparaison avec la loi Normale : Si $X \sim \mathfrak{N}(0, 1)$ et $t = 2$ alors

- l'inégalité de B.-T. donne

$$P(X \notin [-2, 2]) = P(|X - \mathbf{E}\, X| \geq 2) \leq 0.25,$$

- alors que la valeur est $P(X \notin [-2, 2]) = 0.0456$.

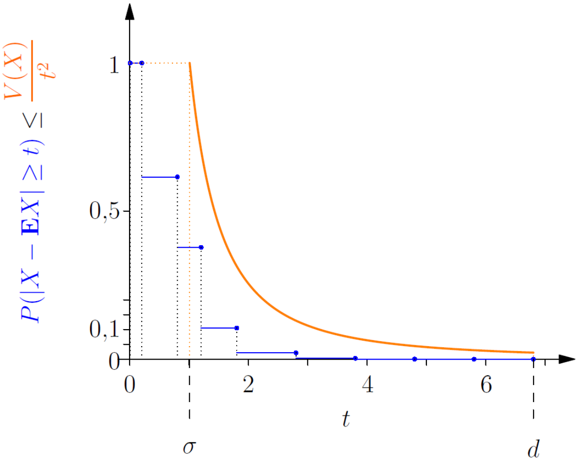


Figure – Comparaison de $P(|X - \mathbf{E}\,X| \geq t)$ et $\frac{\mathbf{V}(X)}{t^2}$ pour $X \sim \mathrm{Bin}(8; 0{,}15)$

Utilisation unilatérale de l'inégalité de B.-T.

$$\forall t > 0, \quad P(X \geq \mathbf{E}\,X + t) \leq \frac{\mathbf{V}(X)}{t^2}, \quad P(X \leq \mathbf{E}\,X - t) \leq \frac{\mathbf{V}(X)}{t^2}.$$

On a l'impression de prendre ainsi un majorant deux fois trop grand. C'est vrai si la loi de X est symétrique autour de $\mathbf{E}\,X$, par exemple :

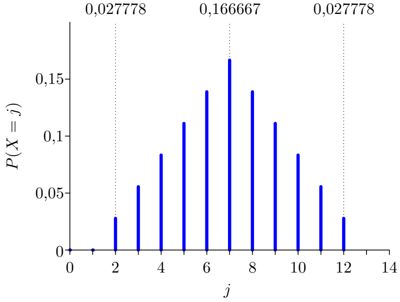


Figure – Loi de X somme des points de deux dés équilibrés

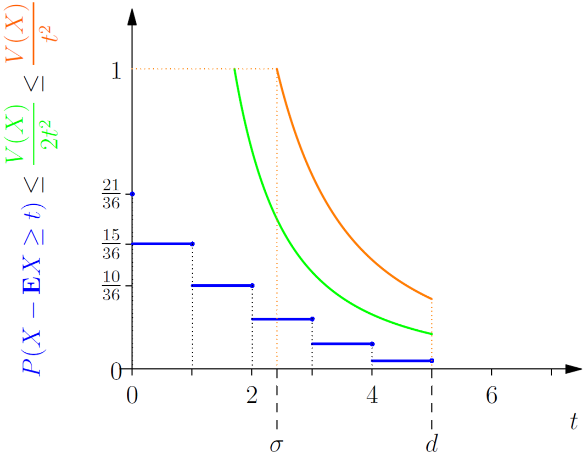


Figure – Inégalité B.T. unilatérale pour X somme des points de deux dés

Mais sinon on n'en sait rien !
Exemple : $P(X = 0) = 0{,}9$, $P(X = 10) = 0{,}1$, $t = 9$. Dans ce cas, $\mathbf{E}\,X = 1$, $\mathbf{V}(X) = 9$ et

$$P(X - \mathbf{E}\,X \geq 9) = P(|X - \mathbf{E}\,X| \geq 9) = P(X = 10) = 0{,}1$$

tandis que

$$\frac{\mathbf{V}(X)}{9^2} = \frac{1}{9} \simeq 0{,}111, \quad \text{d'où} \quad \frac{\mathbf{V}(X)}{2 \times 9^2} < P(X - \mathbf{E}\,X \geq 9).$$

Forme universelle de l'inégalité de Bienaymé-Tchébychef

$$\forall u > 0, \quad P(|X - \mathbf{E}\, X| \geq u\sigma) \leq \frac{1}{u^2} \qquad (\sigma \text{ écart-type de } X).$$

Sous cette forme, le majorant obtenu ne dépend plus de la loi de X. Ceci permet de comprendre pourquoi $\sigma = \sqrt{\mathbf{V}(X)}$ s'appelle écart-type ou unité d'écart de X.

Cette forme est utile notamment lorsque l'on cherche un intervalle de confiance pour $\mathbf{E}\, X$ inconnue avec σ connu ou facilement majorable :

$$P(X - u\sigma \leq \mathbf{E}\, X \leq X + u\sigma) \geq 1 - \frac{1}{u^2}.$$

Inégalité de B.-T. pour une somme de v.a. indépendantes

- $X_1, \dots, X_n$, v.a. indépendantes et de même loi, $\mathbf{E}\, X_i = \mathbf{E}\, X_1 = \mu$, $\mathbf{V}(X_i) = \mathbf{V}(X_1)$.
- $M_n = \frac{S_n}{n} = \frac{X_1+\cdots+X_n}{n}$.

Alors $\mathbf{E}\, S_n = n\mu$, $\mathbf{E}\, M_n = \frac{\mathbf{E}\, S_n}{n} = \mu$ et par indépendance des X_i,

$$\mathbf{V}(S_n) = \sum_{i=1}^{n} \mathbf{V}(X_i) = n\, \mathbf{V}(X_1),\ \mathbf{V}(M_n) = \frac{\mathbf{V}(S_n)}{n^2} = \frac{\mathbf{V}(X_1)}{n}.$$

Inégalité de concentration

Sous les hypothèses précédentes, $\forall \delta > 0,\ \forall n \geq 1$,

$$P(|M_n - \mu| \geq \delta) \leq \frac{\mathbf{V}(X_1)}{n\delta^2}.$$

Application statistique : Intervalle de confiance ou de fluctuations (tests).

Théorème (Loi faible des grands nombres)

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi à valeurs dans un ensemble fini. Notons $\mu = \mathbf{E}\, X_1$ et M_n la moyenne arithmétique des n variables $X_1, \ldots, X_n$. Alors pour tout $\delta > 0$,

$$\lim_{n\to+\infty} P(|M_n - \mu| \geq \delta) = 0.$$

On dit que M_n converge vers μ en probabilité.

Preuve. Il suffit de faire $n \to +\infty$ dans l'inégalité de concentration.
La conclusion peut se reformuler ainsi :

$$\forall \delta > 0, \quad \lim_{n\to+\infty} P(\mu - \delta < M_n < \mu + \delta) = 1.$$

Corollaire (Loi des G.N. pour les fréquences de J. Bernoulli)

Soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre p. Pour $n \geq 1$, notons $S_n = X_1 + \cdots + X_n$. Alors quand n tend vers l'infini,

$$\frac{S_n}{n} \quad \textit{converge en probabilité vers} \quad \mathbf{E}\, X_1 = p.$$

Interprétation dans le cas d'un modèle d'épreuves répétées.

Pour une suite d'épreuves répétées dans des conditions identiques (par exemple lancers d'un dé bleu et d'un dé jaune), on considère une suite d'évènements E_i, dont la réalisation ne dépend que du résultat de la i-ème épreuve et ayant la même définition à part le numéro d'épreuve (par exemple le dé bleu donne plus de points que le jaune au i-ème lancer). Les E_i ont alors même probabilité p et en prenant $X_i = \mathbf{1}_{E_i}$, la fréquence S_n/n de réalisation des E_i en n épreuves converge vers p.

Une suggestion de TP : aire d'une ellipse 1/3

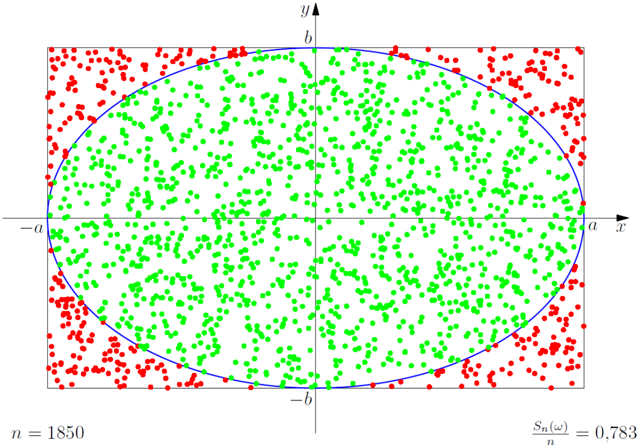


Figure – Aire d'une ellipse par la méthode de Monte-Carlo

Une suggestion de TP : aire d'une ellipse 2/3

- On génère n points aléatoires $M_1, \ldots, M_n$ dans le rectangle $R = [-a, a] \times [-b, b]$. Comment ?
- Pour chaque M_i, on teste s'il appartient au domaine B délimité par l'ellipse d'équation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. On code $Z_i = 1$ si oui, $Z_i = 0$ sinon.
- On admet que $P(M_i \in B) = \frac{\mathrm{aire}(B)}{\mathrm{aire}(R)}$. On pourrait l'expliquer en remplaçant B par une réunion A de carreaux d'une grille quadrillant R et en montrant par dénombrement que $P(M_i \in A) = \frac{\mathrm{aire}(A)}{\mathrm{aire}(R)}$.
- On note S_n la somme des Z_i. Ainsi S_n/n est la variable aléatoire fréquence des points M_i tombés dans B.

Par la loi des grands nombres, S_n/n converge en probabilité vers

$$\mathbf{E}\, Z_1 = P(Z_1 = 1) = \frac{\mathrm{aire}(B)}{\mathrm{aire}(R)} = \frac{\mathrm{aire}(B)}{4ab}$$

On en déduit que pour « n grand »,

$$\mathrm{aire}(B) \simeq 4\frac{S_n}{n}ab.$$

Une suggestion de TP : aire d'une ellipse 3/3

- En pratique, une fois le programme de calcul de S_n/n mis au point, on peut fournir à chaque élève une valeur différente du couple (a, b) et collecter les résultats obtenus.
- On constate qu'ils sont tous proches de $0{,}75$. Ceci amène à conjecturer l'existence d'une constante c approximée par $4S_n/n$ telle que $\mathrm{aire}(B) = c \times ab$.
- En comparant avec le cas « connu » du cercle ($a = b$), on peut conjecturer que $c = \pi$.

D'où la formule

$$\boxed{\mathrm{aire}(B) = \pi ab.}$$

Pour une justification géométrique, on peut ensuite approcher l'aire d'un disque de rayon 1 par quadrillage et examiner l'effet sur ce quadrillage et sur le disque de la transformation

$$M(x, y) \longmapsto M'(ax, by).$$

Interprétation graphique de la loi faible des grands nombres

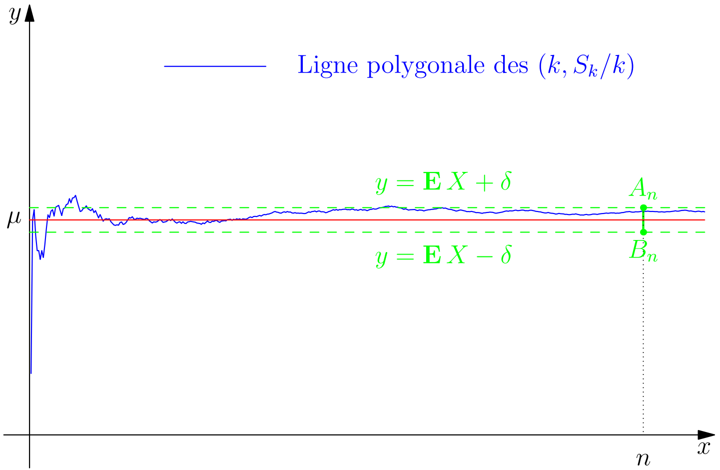


Figure – Pour tout $\delta > 0$, la probabilité que la ligne bleue traverse le segment $A_n B_n$ tend vers 1 quand n tend vers l'infini

Vitesse de convergence dans la LGN

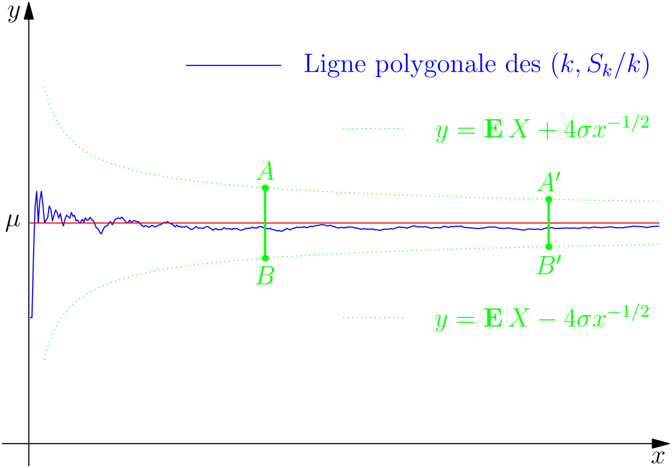


Figure – Chaque segment AB est traversé par la ligne bleue avec probabilité d'au moins $15/16$

Pour aller plus loin

Intérêt de l'inégalité de concentration :

- inégalité valable quelques soit le nombres de données n,
- se démontre « facilement »,
- peu d'hypothèses, valables sans connaître la loi (juste l'espérance et la variance),
- permet de construire des intervalles de fluctuation ou des intervalles de confiance.

Et par rapport à l'ancien programme :

- moins bon lorsque n est grand que le Théorème Limite Central,
- mais qui n'était énoncé que pour des lois binomiales,
- et dont la démonstration n'est pas accessible.

Et pour aller plus loin :

- la démonstration passe très facilement lorsque les variables aléatoires ne sont pas discrètes
- les vitesses peuvent être améliorées grandement lorsqu'on suppose un peu plus...

Exercice 1

On lance 3 600 fois un dé équilibré à six faces. On souhaite minorer la probabilité que le nombre d'apparitions du chiffre 1 soit compris entre 480 et 720.

1. Soit S la variable aléatoire comptant le nombre d'apparitions du chiffre 1 au cours des 3 600 lancers. Justifier que S suit une loi binomiale dont on précisera les paramètres.
2. Déterminer la variance et l'espérance de S.
3. Justifier l'équivalence :
$$480 < S < 720 \iff |S - 600| < 120$$
4. En utilisant l'inégalité de Bienaymé-Tchebychev, prouver que
$$P(|S - 600| \geqslant 120) \leqslant 0{,}035$$
5. En déduire que la probabilité que le numéro 1 apparaisse entre 480 et 720 fois au cours de ces 3600 lancers est supérieure à $0{,}96$.
6. En utilisant la loi de la variable aléatoire S, calculer numériquement la probabilité cherchée.

Exercice 2

On effectue une suite de lancers d'un dé à six faces. Quel est le nombre de lancers suffisant pour pouvoir affirmer avec un risque d'erreur inférieur à 5 % que la fréquence d'apparition du 6 est comprise entre $\frac{1}{6} - 0{,}01$ et $\frac{1}{6} + 0{,}01$?
Notons S_n le nombre de 6 apparus au cours des n premiers lancers.

1. Quelle est la loi de S_n ? Son espérance ? Sa variance ?
2. Montrer que la question revient à trouver n de sorte que
$$P\left(\left|\frac{S_n}{n} - \frac{1}{6}\right| \geqslant 0{,}01\right) \leqslant 0{,}05.$$
3. En déduire le nombre de lancers.

Exercice 3

Une élection oppose deux candidats A et B. On note p la proportion d'électeurs, dans la population totale, décidés à voter pour le candidat A.
On souhaite estimer cette proportion inconnue et effectuer un sondage (assimilé à un tirage avec remise) auprès de n personnes.
L'institut de sondage veut une fourchette de 1 % et un niveau de confiance de 95 %.
On note S_n la variable aléatoire égale au nombre de personnes votant pour le candidat A et on souhaite quantifier le fait que $\frac{S_n}{n}$ approche p à 10^{-2} près avec une probabilité supérieure ou égale à 95 %.

1. Quelle est la loi de S_n ? Son espérance ? Sa variance ?
2. Démontrer que
$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant 0{,}01\right) \leqslant \frac{1}{4n \times 10^{-4}}.$$
3. En déduire une condition sur n pour que $\frac{S_n}{n}$ soit une valeur approchée de p à 10^{-2} près avec une probabilité supérieure ou égale à $95\,\%$.