L'inégalité de Markov

Une variable aléatoire X est dite réelle positive ou nulle lorsque toutes les valeurs prises par cette variable sont des réels positifs ou nuls.

Soit X une variable aléatoire réelle positive ou nulle d'espérance notée $E(X)$.

Pour tout $a>0$, on a $p(X \geq a) \leq \dfrac{E(X)}{a}$. Cette inégalité est l'inégalité de Markov.

Démonstration

Notons x_i pour $1 \leq i \leq n$ les n valeurs prises par la variable aléatoire X.

Nous pouvons donc écrire $E(X)=\sum_{i=1}^{n} x_i \times p(X=x_i)$ et scinder cette somme en deux parties.

$$E(X)=\sum_{x_i<a} x_i \times p(X=x_i)+\sum_{x_i \geq a} x_i \times p(X=x_i)$$

La première partie concerne les valeurs de la variable strictement inférieures à a tandis que la deuxième partie concerne les valeurs de la variable supérieures ou égales à a.

Etant donné que les valeurs de X sont positives et que les probabilités associées le sont aussi :

$$E(X)=\sum_{x_i<a} x_i \times p(X=x_i)+\sum_{x_i \geq a} x_i \times p(X=x_i) \geq \sum_{x_i \geq a} x_i \times p(X=x_i)$$

Et par conséquent on a l'inégalité : $E(X) \geq \sum_{x_i \geq a} x_i \times p(X=x_i)$.

Ensuite, étant donné que l'on travaille avec des valeurs supérieures ou égales à a, on peut écrire :

$$E(X) \geq \sum_{x_i \geq a} x_i \times p(X=x_i) \geq a \times \sum_{x_i \geq a} p(X=x_i)$$

Qui peut se réécrire de la manière suivante :

$$E(X) \geq \sum_{x_i \geq a} x_i \times p(X=x_i) \geq a \times p(X \geq a)$$

Et par conséquent on a l'égalité : $E(X) \geq a \times p(X \geq a)$.

Que l'on peut réécrire et retenir ainsi : $p(X \geq a) \leq \dfrac{E(X)}{a}$.

Remarque : Cette première inégalité de concentration permet de majorer la probabilité d'un événement (sans forcément obtenir la meilleure majoration possible, c'est-à-dire sans être sur qu'il s'agisse du plus petit des majorants)…

Majorer la probabilité de certains événements

En 2015, le salaire brut mensuel moyen en France était de 2442 euros. On choisit un salarié au hasard et on note X la variable aléatoire donnant son salaire. Les salaires étant positifs ou nuls on sait que X est une variable aléatoire positive ou nulle. Majorer la probabilité qu'un salarié français gagne plus de 7326 euros.

Dans un immeuble, l'ascenseur reste en moyenne deux minutes au rez-de-chaussée avant d'être sollicité à nouveau. Majorer la probabilité que l'ascenseur reste au rez-de-chaussée plus de cinq minutes.

On suppose que la température moyenne aux Maldives est de 28°C et que la température n'est jamais négative. Majorer la probabilité que la température dépasse 34°C un jour donné.

Majorer la probabilité demandée dans les cas suivants où X est une variable aléatoire positive ou nulle : $p(X \geq 1)$ dans le cas où $E(X) = 0,5$; $p(X \geq 24)$ dans le cas où $E(X) = 6$; $p(X \geq 4)$ dans le cas où $E(X) = 4/3$.

Majorer mais aussi minorer la probabilité de certains événements

Sur une autoroute, la vitesse moyenne des véhicules est de 120 km/h.

- Majorer la probabilité qu'un automobiliste roule à une vitesse supérieure à 150 km/h.
- Minorer la probabilité qu'un automobiliste roule à une vitesse inférieure à 140 km/h.

Léo termine le mois avec en moyenne 40 euros sur son compte. Sa banque ne lui autorise aucun découvert (on travaille donc avec des valeurs positives ou nulles).

- Majorer la probabilité que Léo finisse le mois avec plus de 70 euros sur son compte.
- Minorer la probabilité que Léo finisse le mois avec moins de 50 euros sur son compte.

Vérifier une affirmation

Un économiste affirme la chose suivante : « moins de 6,2% de la population mondiale adulte est millionnaire ». En se basant les données suivantes, peut-on penser que l'économiste a raison ?

- La richesse mondiale est de 278,1 billions de dollars (1 billion représente 1000 milliards),
- La population adulte mondiale s'élève à 4,5 milliards d'individus.

Retrouver la moyenne

On considère la variable aléatoire donnant le nombre de buts marqués par une équipe de football au cours d'un match. Un analyste déclare : « grâce à l'inégalité de Markov, je peux dire que la probabilité que l'équipe marque au moins deux buts est inférieure à 70% ». Trouver alors le nombre moyen de buts marqués par match par cette équipe.

L'inégalité de Markov est-elle toujours utile ?

Une personne affirme que l'inégalité de Markov est inutile lorsque $a \leq E(X)$. Qu'en penser ?

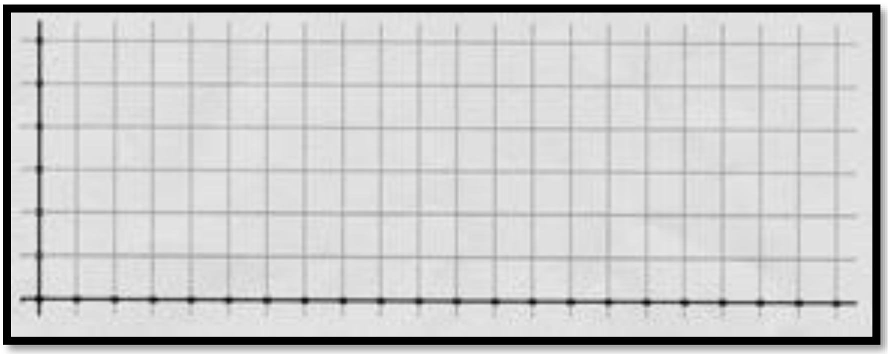
De l'espérance à la variance

On rappelle que l'espérance (indicateur de position) se calcule ainsi $E(X)=\sum_i x_i \times p(X=x_i)$.

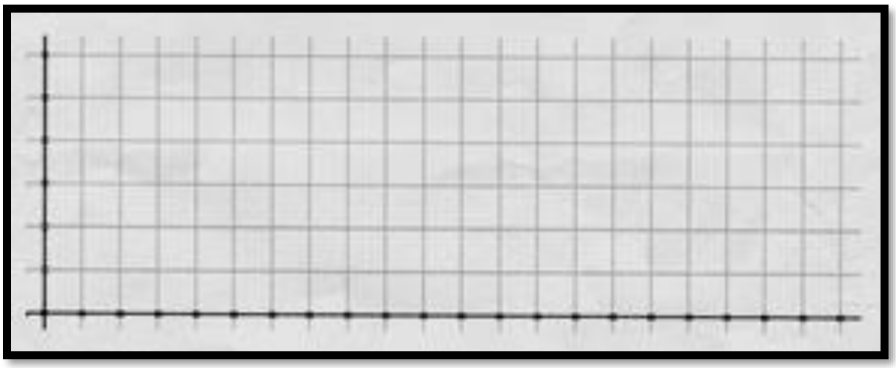
On indique que la variance (indicateur de dispersion) répond à la définition suivante $V(X)=E\left(\left[X-E(X)\right]^2\right)$ et se calcule ainsi $V(X)==\sum_i \left(x_i-E(X)\right)^2 \times p(X=x_i)$.

On considère ci-dessous deux variables aléatoires X et Y. Calculer $E(X)$, $E(Y)$, $V(X)$, $V(Y)$.

x_i	8	9	10	11	12	13	14
p_i	$1/17$	$2/17$	$3/17$	$5/17$	$3/17$	$2/17$	$1/17$



y_i	3	5	6	7	8	9	10	11	12	13	15	17	20
p_i	$1/17$	$1/17$	$1/17$	$1/17$	$2/17$	$1/17$	$1/17$	$1/17$	$1/17$	$2/17$	$3/17$	$1/17$	$1/17$



Exercice d'application directe

Un trader a analysé plusieurs scénarios quant à l'évolution de deux actions notées A et B. On note X la variable aléatoire donnant l'évolution en euros de l'action A et Y celle donnant l'évolution en euros de l'action B. Voici les lois de probabilités de X et de Y. Le trader ne souhaite pas prendre trop de risques et décide d'investir sur l'action la moins volatile. Quelle action lui conseillez-vous ?

Valeur de X	-50	0	10	40
Probabilité	0,1	0,3	0,5	0,1

Valeur de Y	-30	10	30
Probabilité	0,3	0,4	0,3

L'inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

Pour tout $a>0$, on a $p\left(\left|X-E(X)\right|\geq a\right)\leq\dfrac{V(X)}{a^2}$.

Cette inégalité est l'inégalité de Bienaymé Tchebychev

<u>Démonstration</u>

Comme $a>0$, les inégalités $\left|X-E(X)\right|\geq a$ et $\left[X-E(X)\right]^2\geq a^2$ sont équivalentes.

De plus la variable aléatoire $Y=\left[X-E(X)\right]^2$ est positive ou nulle.

On peut donc appliquer l'inégalité de Markov à la variable aléatoire Y et au réel a^2.

Ainsi on a : $p\left(Y\geq a^2\right)\leq\dfrac{E(Y)}{a^2}$ et donc : $p\left(\left[X-E(X)\right]^2\geq a^2\right)\leq\dfrac{E\left(\left[X-E(X)\right]^2\right)}{a^2}$.

Nous savons par ailleurs que $V(X)=E\left(\left[X-E(X)\right]^2\right)$.

Par conséquent nous obtenons : $p\left(\left[X-E(X)\right]^2\geq a^2\right)\leq\dfrac{V(X)}{a^2}$.

Inégalité que nous pouvons réécrire et retenir ainsi : $p\left(\left|X-E(X)\right|\geq a\right)\leq\dfrac{V(X)}{a^2}$.

<u>Remarque :</u> Cette deuxième inégalité de concentration permet de majorer la probabilité que l'écart entre une variable aléatoire et son espérance soit supérieur ou égal à une quantité donnée. Là encore, la majoration n'est pas la meilleure possible (ce n'est pas le plus petit des majorants). Dans la réalité il est possible que la probabilité soit inférieure au majorant obtenu.

Exercice d'application directe

Le taux moyen de glycémie est de 1 g/L avec une variance de 0,1. Une personne présente un taux X critique si son taux ne se situe pas dans l'intervalle [0,5 ; 1,5].

Majorer la probabilité qu'une personne présente un taux critique.

Et si on souhaite minorer la probabilité d'un événement ?

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$.

Pour tout $a>0$, l'inégalité de on a $p\left(\left|X-E(X)\right|<a\right)\geq 1-\dfrac{V(X)}{a^2}$.

Cette inégalité est une reformulation de l'inégalité de Bienaymé Tchebychev.

Majorer ou minorer la probabilité de certains événements

La taille moyenne des hommes est de 170 cm et la variance est de 7,2. On prend un homme au hasard et on note X la variable aléatoire donnant sa taille en centimètres. Jean mesure 165 cm. Paul mesure 175 cm. Que peut-on dire de la probabilité qu'un homme soit plus petit que Jean ou plus grand que Paul.

La moyenne des notes des élèves d'un Lycée est de 11,4 et la variance est de 1,2. On prend un élève au hasard et on note X la variable aléatoire correspondant à sa note. La note de Jean est 9. La note de Paul est 13,8. Que peut-on dire de la probabilité qu'élève ait une note strictement comprise entre celle de Jean et celle de Paul.

Sur les 20 matchs précédents, une équipe de rugby a marqué 60 essais. On note X la variable aléatoire donnant le nombre d'essais marqués au cours d'un match. Que vaut l'espérance de X ? On suppose que la variance est égale à 0,67. Majorer la probabilité qu'au cours du prochain match l'écart entre le nombre d'essais marqués et la moyenne soit supérieur ou égal à 1. Minorer la probabilité que l'écart entre le nombre d'essais marqués et la moyenne soit strictement inférieur à 2.

Compagnie aérienne

Sur un vol de 325 voyageurs, une compagnie aérienne décide d'attribuer 340 billets (elle pratique la surréservation). Elle perd de l'argent pour un nombre strictement inférieur à 317 passagers et ne peut pas trouver un autre vol aux personnes en surréservation pour un nombre strictement supérieur à 329 passagers. On considère la variable aléatoire X qui compte le nombre de passagers effectivement présents.

Des études statistiques montrent que $E(X)=323$ et que $V(X)=16{,}15$. Que peut-on dire de la probabilité que la compagnie ne perde pas d'argent et trouve un vol à tout le monde ?

Retrouver la variance d'une variable aléatoire

Une personne souffre d'hyperglycémie si son taux de glycémie à jeun est supérieur à 1,8 g/L et souffre d'hypoglycémie si ce taux est inférieur à 0,4 g/L. Sachant que 15% de la population présente un problème de glycémie et que le taux de glycémie moyen est de 1,1 g/L, un médecin se demande s'il est possible de déterminer la variance de la variable aléatoire donnant le taux de glycémie d'une personne. Pourriez-vous l'aider ?

Des clients estiment que leur temps d'attente en magasin varie trop fortement selon les périodes. Le gérant décide donc de vérifier cette impression. Il mesure que le temps d'attente moyen de ses clients est de 12 minutes et que la probabilité qu'un client attende strictement entre 9 et 15 minutes est de 0,55. Pourriez-vous l'aider à retrouver la variance du temps d'attente ?

Bienaymé-Tchebychev versus Markov

Le nombre de pièces fabriquées dans une usine en une journée suit une variable aléatoire d'espérance 50 et de variance 25. Donner deux majorations de la probabilité que la production dépasse sur une journée les 75 pièces. La première sera donnée avec l'inégalité de Markov et la deuxième sera établie avec l'inégalité de Bienaymé-Tchebychev. Un commentaire pourra être envisagé à la lecture des deux inégalités obtenues.

Notion d'échantillons de variables aléatoires

On considère un entier naturel $n \geq 1$ et $X_1;...;X_n$ des variables aléatoires indépendantes et identiquement distribuées. On note $S_n = X_1 + ... + X_n$ la somme de ces variables aléatoires et $M_n = \frac{S_n}{n} = \frac{X_1 + ... + X_n}{n}$ la moyenne de ces variables aléatoires.

- $E(S_n) = n \times E(X_i)$ et $V(S_n) = n \times V(X_i)$ quel que soit $i \in \{1;...;n\}$
- $E(M_n) = E(X_i)$ et $V(M_n) = \frac{1}{n} \times V(X_i)$ quel que soit $i \in \{1;...;n\}$

Démonstration

$$E(S_n) = E(X_1 + ... + X_n) = \underbrace{E(X_1) + ... + E(X_n)}_{\text{meme espérance}} = n \times E(X_i)$$

$$V(S_n) = V\left(\underbrace{X_1 + ... + X_n}_{\text{indépendantes}}\right) = \underbrace{V(X_1) + ... + V(X_n)}_{\text{meme variance}} = n \times V(X_i)$$

$$E(M_n) = E\left(\frac{1}{n} \times S_n\right) = \frac{1}{n} E(S_n) = \frac{1}{n} \times n \times E(X_i) = E(X_i)$$

$$V(M_n) = V\left(\frac{1}{n} \times S_n\right) = \frac{1}{n^2} V(S_n) = \frac{1}{n^2} \times n \times V(X_i) = \frac{V(X_i)}{n}$$

Exercices d'application directe

On considère 10 variables aléatoires indépendantes et identiquement distribuées. On donne ci-dessous la loi de probabilité de chacune d'entre elles :

x_i	-5	0	1	3
p_i	0,4	0,3	0,2	0,1

On note $S_{10} = X_1 + ... + X_{10}$ et $M_{10} = \frac{X_1 + ... X_{10}}{10}$.

Calculer $E(S_{10})$ et $V(S_{10})$. Calculer ensuite $E(M_{10})$ et $V(M_{10})$.

On considère n variables aléatoires indépendantes et identiquement distribuées. On donne ci-dessous la loi de probabilité de chacune d'entre elles :

x_i	4	10	12
p_i	0,25	0,5	0,25

On note $S_n = X_1 + ... + X_n$ et $M_n = \frac{X_1 + ... X_n}{n}$.

Pour quelle valeur de n a-t-on $E(S_n) \geq 2453$? Pour quelle valeur de n a-t-on $V(S_n) \leq 3600$?

Rappel sur la loi binomiale

On répète n fois de manière **indépendante** une **même** expérience aléatoire présentant **deux issues** S et $\overline{S}$ de probabilités respectives p et $q = 1 - p$. On considère la variable aléatoire X égale au **nombre de succès** S obtenus au cours de ces n expériences. La loi de probabilité de la variable aléatoire X s'appelle alors **loi binomiale** de paramètres n et p, notée $\mathrm{B}(n, p)$.

k	0	1	2	$\dots$	$n-1$	n
$p(X = k)$	$\binom{n}{0} q^n$	$\binom{n}{1} pq^{n-1}$	$\binom{n}{2} p^2 q^{n-2}$	$\dots$	$\binom{n}{n-1} p^{n-1} q$	$\binom{n}{n} p^n$

On retrouve dans la loi binomiale **les coefficients binomiaux** qui correspondent au nombre de chemins de l'arbre pondéré permettant d'obtenir k « succès » au cours des n générations.

Pour tout entier k tel que $0 \le k \le n$ on a : $\boxed{p(X = k) = \binom{n}{k} p^k q^{n-k}}$

Deux indicateurs incontournables

Pour une loi binomiale, l'**espérance**, la **variance** et l'**écart type** sont donnés par les formules :

$$\boxed{E(X) = n \times p} \qquad \boxed{V(X) = n \times p \times q}$$

Démonstration

La variable X sui la loi binomiale de paramètres n et p. On considère un schéma de Bernoulli composé de n épreuves indépendantes de probabilité de succès p. La variable aléatoire X est alors égale au nombre de succès obtenus au cours de la répétition des n épreuves de Bernoulli.

Pour tout $i \in \{1; \dots; n\}$ on note X_i la variable aléatoire qui prend pour valeur 1 si on obtient un succès à la i-ième épreuve et 0 sinon. On a $X = X_1 + \dots + X_n$ avec pour chaque variable X_i :

x_i	1	0
p_i	p	$1-p$

$$E(X_i) = 1 \times p + 0 \times (1-p) = p \text{ et } V(X_i) = (1-p)^2 \times p + (0-p)^2 \times (1-p) = p(1-p)$$

Pour la somme $X = X_1 + \dots + X_n$ on peut en déduire :

$$E(X) = E(X_1 + \dots + X_n) = \underbrace{E(X_1) + \dots + E(X_n)}_{\text{meme espérance}} = n \times p$$

$$V(X) = V\left(\underbrace{X_1 + \dots + X_n}_{\text{indépendantes}}\right) = \underbrace{V(X_1) + \dots + V(X_n)}_{\text{meme variance}} = n \times p \times (1-p)$$

Inégalités et loi binomiale

Exercice 1

On lance 800 fois une pièce de monnaie non truquée, c'est-à-dire bien équilibrée. Donner une minoration de la probabilité que le nombre de « Pile » obtenus au cours des 800 lancers soit compris entre 381 et 419.

Exercice 2

Une lanceuse de fléchettes met sa fléchette « dans le mille » 60% du temps et on suppose que tous ses lancers sont indépendants. Quelle loi suit la variable aléatoire M donnant le nombre de lancers « dans le mille » sur 20 tentatives effectuées. Quand on lui demande combien elle pense mettre de fléchettes « dans le mille » elle répond ainsi : « moins de 15 mais plus de 9 ». Donner une minoration de la probabilité qu'elle ait raison.

Exercice 3

On considère une urne contenant 800 boules blanches et 200 boules noires. On tire 400 fois une boule de l'urne en prenant soin de la replacer dans l'urne après avoir relevé sa couleur (les tirages sont donc indépendants les uns des autres). On note X le nombre de boules blanches obtenues au cours des 400 tirages successifs. On cherche à minorer la probabilité d'obtenir entre 300 et 340 boules blanches au cours des 400 tirages. Est-il possible d'effectuer cette minoration ?

L'inégalité de concentration

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$. On pose M_n la variable aléatoire moyenne d'un échantillon de taille n de X c'est-à-dire que $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ où les variables X_i sont indépendantes et de même loi de probabilité que celle de X.

Alors pour tout $a > 0$ on a $p\left(\left|M_n - E(X)\right| \geq a\right) \leq \frac{V(X)}{na^2}$.
Cette inégalité est appelée l'inégalité de concentration.

Démonstration

$$E(M_n) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n} \times nE(X) = E(X).$$

$$V(M_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} \times nV(X) = \frac{V(X)}{n}.$$

J'applique l'inégalité de Bienaymé-Tchebychev à la variable aléatoire M_n :

Pour tout $a > 0$, on a $p\left(\left|M_n - E(M_n)\right| \geq a\right) \leq \frac{V(M_n)}{a^2}$ et donc $p\left(\left|M_n - E(X)\right| \geq a\right) \leq \frac{V(X)}{na^2}$.

Application de l'inégalité de concentration

Situation 1

Une urne contient deux jetons sur lesquels figurent le numéro 3, deux jetons sur lesquels figurent le numéro 5 et un jeton sur lequel figure le numéro 10. On tire un jeton dans cette urne et on considère la variable aléatoire X donnant le nombre obtenu. Déterminer dans un premier temps l'espérance E(X) et la variance V(X) de cette variable aléatoire. Combien de tirages avec remise peut-on effectuer dans cette urne pour être sur au risque d'erreur de 5% (ou au seuil de confiance de 95%) que la moyenne des nombres obtenus soit comprise entre 5 et 5,4 ?

Situation 2

On lance un dé tétraédrique équilibré à quatre faces numérotées de 1 à 4. On appelle R la variable aléatoire donnant le résultat obtenu. Déterminer dans un premier temps l'espérance E(R) et la variance V(R) de cette variable aléatoire. Combien de lancers peut-on effectuer pour être sûr au risque d'erreur de 5% (ou au seuil de confiance de 95%) que la moyenne des résultats de ces lancers soit comprise entre 2 et 3 ?

Situation 3

Le nombre de lignes réalisées au jeu de Tetris par Alexei est donné par une variable aléatoire d'espérance 125 et de variance 100. En supposant que toutes les parties sont indépendantes, déterminer à partir de combien de parties jouées par Alexei peut être sûr, au risque de 5% (ou au seuil de confiance de 95%) que sa moyenne de lignes par parties est comprise entre 120 et 130 ?

Situation 4

Compte tenu de l'âge, de la corpulence et de l'activité physique de son patient, le médecin arrive à la conclusion qu'il aura besoin de 2900 à 3100 calories par jour en moyenne. La variable aléatoire donnant l'apport calorique de cette personne chaque jour suit une loi d'espérance 3000 et de variance 2500. Au bout de combien de jours peut-il être sûr, au risque d'erreur de 5% (ou au seuil de confiance de 95%), de respecter les préconisations de son médecin ?

Situation 5

Pour se rendre au travail, une personne prend le métro. La variable aléatoire donnant le temps exprimé en minutes du trajet suit une loi d'espérance 6 et de variance 1. Au bout de combien de trajets cette personne peut-elle être sure au risque d'erreur de 5% d'avoir mis en moyenne entre 5 minutes 45 secondes et 6 minutes 15 secondes pour se rendre au travail ?

Loi des grands nombres

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$. On pose M_n la variable aléatoire moyenne d'un échantillon de taille n de X c'est-à-dire que $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ où les variables X_i sont indépendantes et de même loi de probabilité que celle de X.

Alors pour tout $a > 0$ on a $\lim_{n\to+\infty} p\left(\left|M_n - E(X)\right| \geq a\right) = 0$.

Autrement dit, plus la taille de l'échantillon d'une variable aléatoire X est grande, plus l'écart entre la moyenne de cet échantillon et l'espérance de la variable X est faible. On dit que M_n converge en probabilité vers $E(X)$ lorsque n tend vers $+\infty$. Pour estimer l'espérance (la moyenne théorique), on peut répéter l'expérience de manière indépendante un grand nombre de fois et calculer la moyenne empirique.

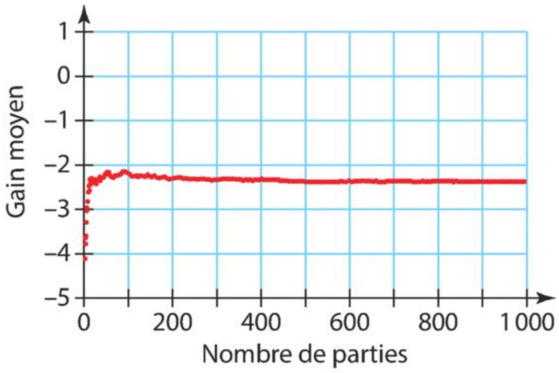
Démonstration

D'après l'inégalité de concentration pour tout $a>0$ on a $0 \leq p\left(\left|M_n - E(X)\right| \geq a\right) \leq \dfrac{V(X)}{na^2}$.

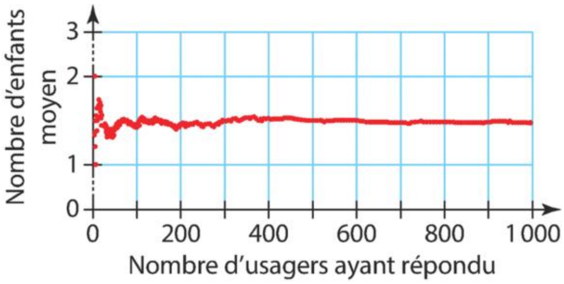
Or $\lim\limits_{n\to+\infty} \dfrac{V(X)}{na^2} = 0$ donc d'après le théorème des gendarmes $\lim\limits_{n\to+\infty} p\left(\left|M_n - E(X)\right| \geq a\right) = 0$.

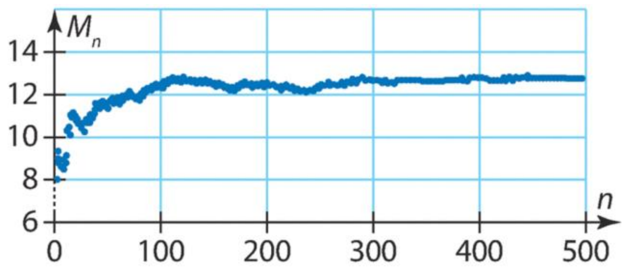
Visualisation de la loi des grands nombres

L'évolution du gain moyen à une table de jeu est donnée par le graphique ci-contre. Estimer l'espérance de la variable aléatoire donnant le gain algébrique.



Les réponses aux questionnaires d'une administration demandant d'indiquer le nombre d'enfants de moins de 18 ans dans le foyer des usagers sont représentées par le graphique ci-dessous à gauche. Estimer l'espérance de la variable aléatoire donnant ce nombre d'enfants.





On considère un échantillon de variables aléatoires X_i et $M_n = \dfrac{X_1 + \ldots + X_n}{n}$ la moyenne des n premières variables dont l'évolution est donnée par le graphique proposé ci-dessus à droite.

Parmi les deux lois de probabilité proposées ci-dessous, laquelle est celle des variables X_i ?

a_i	8	12	24
$p(A=a_i)$	0,5	0,25	0,25

b_i	5	7,5	15
$p(B=b_i)$	0,1	0,7	0,2