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 $p(X \geq a) \leq \frac{E(X)}{a}$. 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 : $p(X \geq a) \leq \frac{E(X)}{a}$.

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 ?

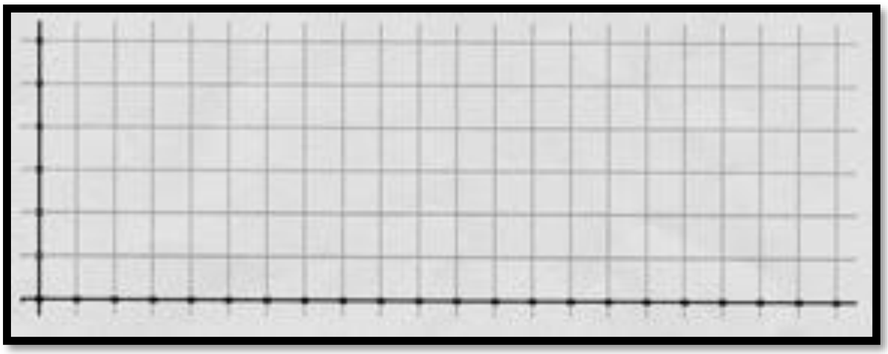
De l'espérance à la variance

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

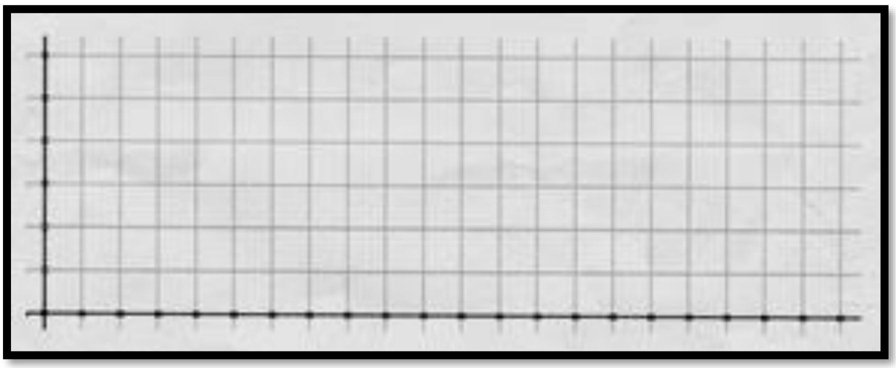
On indique que la variance (indicateur de dispersion) répond à la définition suivante $V(X)=E\left(\left[X-E(X)\right]^2\right)$ et se calcule ainsi $V(X)==\sum_i \left(x_i-E(X)\right)^2 \times p(X=x_i)$.

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	$\frac{1}{17}$	$\frac{2}{17}$	$\frac{3}{17}$	$\frac{5}{17}$	$\frac{3}{17}$	$\frac{2}{17}$	$\frac{1}{17}$



y_i	3	5	6	7	8	9	10	11	12	13	15	17	20
p_i	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{2}{17}$	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{1}{17}$	$\frac{2}{17}$	$\frac{3}{17}$	$\frac{1}{17}$	$\frac{1}{17}$



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

Démonstration

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 qui se réécrit ainsi $p\left(\left|X-E(X)\right|\geq a\right)\leq\dfrac{V(X)}{a^2}$.

Remarque

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.

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 $a>0$, on a $p\left(\left|X-E(X)\right|<a\right)\geq 1-\dfrac{V(X)}{a^2}$. C'est une reformulation de l'inégalité de Bienaymé Tchebychev.

Majorer ou minorer la probabilité de certains événements

Situation 1

On considère la variable aléatoire D donnant le débit de la Loire en mètres cube par seconde. Une étude statistique permet d'affirmer que $E(D)=350$ et que $V(D)=28000$.

La navigation est interrompue lorsque le débit est trop faible (inférieur à 150 mètres cube par seconde) ou trop fort (supérieur à 550 mètres cube par seconde). A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que la navigation soit interrompue ?

Un batelier indique que les conditions de navigation sur la Loire sont idéales lorsque le débit est strictement compris entre 300 et 400 mètres cube par seconde. A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que la navigation soit idéale ?

Situation 2

Une compagnie de transport met en vente des billets pour effectuer un trajet en bus. Des études statistiques lui indiquent que le nombre de passagers ayant acheté leur biller et qui vont réellement se présenter à l'embarquement est une variable aléatoire X d'espérance 32 et de variance 9.

Pour ce trajet, la compagnie est en difficulté financière si le nombre de passagers qui se présente est inférieur à 18 (le trajet n'est pas rentable) ou si le nombre de passagers qui se présente est supérieur à 46 (pas assez de place dans le bus qui compte 45 places). A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que cette compagnie se retrouve en difficulté ?

Le gérant explique que les conditions idéales pour le bon déroulement de ce trajet est d'avoir un nombre de passagers strictement compris entre 24 et 40 passagers (le trajet est rentable et le bus n'est pas bondé). A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que cette compagnie se retrouve dans des conditions idéales ?

Situation 3

On considère la variable aléatoire X donnant la vitesse du vent exprimée en km/h dans une région où on souhaite implanter des éoliennes pour produire de l'électricité. Une étude statistique permet d'affirmer que $E(X)=75$ et que $V(X)=120$. La variable aléatoire X est dite continue.

La production électrique de ces éoliennes est interrompue lorsque le vent est trop faible (inférieur ou égal à 35 km/h) ou trop fort (supérieur ou égal à 115 km/h). A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que la production électrique soit interrompue dans cette région ?

Le fabriquant des éoliennes indique que les conditions de production sont idéales pour le modèle proposé lorsque la vitesse du vent est strictement comprise entre 55 et 95 km/h. A l'aide de l'inégalité de Bienaymé-Tchebychev, que peut-on dire de la probabilité que les conditions soient idéales pour la production électrique de ses éoliennes ?

Situation 3

Une élève entrée en classe préparatoire aux grandes écoles se renseigne : les étudiants de deuxième année lui disent que le nombre de feuilles utilisées pour écrire les cours de maths dans l'année suit une loi F d'espérance 1250 et de variance 6000. A la rentrée, cette élève a acheté deux paquets de feuilles : un paquet de 1000 feuilles qu'elle ouvrira en premier et un paquet de 500 feuilles qu'elle ouvrira en deuxième. Quelqu'un lui dit : « tu as plus de 90% de chances d'utiliser au moins toutes les feuilles du premier paquet sans être obligé d'en acheter un troisième ». Que penser de cette affirmation ? La variable aléatoire F est discrète, elle prend des valeurs entières.

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.

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

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 $0 \leq k \leq n$ on a : $\boxed{p(X=k)=\binom{n}{k}p^kq^{n-k}}$

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 $i\in\{1;\ldots;n\}$ 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 $X=X_1+\ldots+X_n$ 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 $X=X_1+\ldots+X_n$ on peut en déduire :

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

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

Inégalité de Bienaymé-Tchebychev et loi binomiale

Situation 1

Si X suit une loi binomiale de paramètres n (répétitions indépendantes) et p (probabilité de succès), comment calcule-t-on l'espérance de X en fonction des paramètres ? Comment calcule-t-on la variance en fonction des paramètres ?

Situation 2

Un couple veut avoir 10 enfants. On suppose qu'ils ont, à chaque naissance, la même probabilité d'obtenir une fille ou un garçon. Minorer la probabilité que ce couple a d'avoir entre 4 et 6 filles.

Situation 3

Je réponds au hasard aux 100 questions d'un QCM ayant deux réponses pour chaque question : l'une juste, l'autre fausse. Minorer la probabilité d'obtenir entre 40 et 60 bonnes réponses.

Situation 4

On lance 1000 fois une pièce que l'on suppose bien équilibrée tombant soit sur « Pile » soit sur « Face ». Minorer la probabilité d'obtenir entre 400 et 600 fois « Pile » au cours des mille lancers.

Situation 5

Des études statistiques montrent qu'il y a seulement 75% de chance qu'une personne inscrite se présente réellement à l'examen d'entrée d'une université. On considère la session du mois de mars à laquelle 800 jeunes se sont inscrits. Minorer la probabilité que le nombre de présents soit compris entre 551 et 649.

Situation 6

Des études statistiques montrent qu'il y a une chance sur dix qu'une personne soit gauchère. On considère les 600 élèves de notre établissement. Donner une minoration de la probabilité que le nombre de gauchers dans notre établissement soit compris entre 51 et 69.

Situation 7

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.

Situation 8

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.

<u>Démonstration</u>

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

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

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 $p\left(\left|M_n - E(X)\right| \geq a\right) \leq \frac{V(X)}{na^2}$.

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.

<u>Démonstration</u>

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

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

La formule de König Huygens

Soit X une variable aléatoire dont la loi de probabilité est donnée par le tableau suivant :

x_i	x_1	x_2	...	x_n
$p(X=x_i)$	p_1	p_2	...	p_n

On rappelle que l'espérance se calcule de la manière suivante :

- $E(X)=m=\sum_{i=1}^{n} p_i x_i = p_1x_1+p_2x_2+\ldots+p_nx_n$

On rappelle également que la variance se calcule de la manière suivante :

- $V(X)=\sum_{i=1}^{n} p_i(x_i-m)^2 = p_1(x_1-m)^2+p_2(x_2-m)^2+\ldots+p_n(x_n-m)^2$

En développant les termes $(x_i-m)^2$, démontrer l'égalité suivante :

- $V(X)=\sum_{i=1}^{n} p_i x_i{}^2-m^2 = p_1x_1{}^2+p_2x_2{}^2+\ldots+p_nx_n{}^2-m^2$

Cette égalité (appelée formule de König Huygens) permet d'affirmer que la variance (qui est par définition, la moyenne des carrés des écarts à la moyenne) peut également se calculer comme la différence entre la moyenne des carrés des valeurs et le carré de la moyenne.

Application directe de la formule

Devoir 1 : 8, 9, 9, 10, 10, 10, 11, 11, 11, 11, 11, 12, 12, 12, 13, 13, 14.
Devoir 2 : 3, 5, 6, 7, 8, 8, 9, 10, 11, 12, 13, 13, 15, 15, 15, 17, 20.

On note X la variable aléatoire correspondant aux notes obtenues sur le devoir 1.
On note Y la variable aléatoire correspondant aux notes obtenues sur le devoir 2.
Calculer $E(X)$ et $E(Y)$. Calculer ensuite $V(X)$ et $V(Y)$ à l'aide de la formule de König Huygens présentée et démontrée ci-dessus. Comparer avec les résultats obtenus en page 1.

Application directe de la formule, encore…

Xavier et Yves s'affrontent en vue d'une sélection lors d'une épreuve comportant 20 tirs sur cible.

Xavier : **50 – 20 – 20 – 30 – 10 – 20 – 30 – 10 – 50 – 30 – 0 – 20 – 30 – 50 – 10 – 50 – 20 – 30 – 30 – 10**
Yves : **50 – 20 – 20 – 50 – 10 – 20 – 30 – 10 – 50 – 30 – 0 – 20 – 0 – 50 – 10 – 50 – 20 – 50 – 30 – 0**

On note X la variable aléatoire correspondant aux points obtenus par Xavier.
On note Y la variable aléatoire correspondant aux points obtenus par Yves.
Calculer $E(X)$ et $E(Y)$. Calculer ensuite $V(X)$ et $V(Y)$ à l'aide de la formule de König Huygens présentée et démontrée ci-dessus. Comparer avec les résultats obtenus en page 1.

Propriétés de l'espérance

Une situation pour comprendre

Une entreprise compte 100 employés. Le tableau ci-dessous donne la répartition des salaires.

Salaire en euros	1 550	1 750	2 200	3 000
Nombre de personnes	40	35	24	1

Le directeur prend au hasard le bulletin de salaire d'un de ses employés. On note X la variable aléatoire correspondant au salaire de cet employé. Calculer l'espérance $E(X)$.

1. Le directeur décide d'augmenter tous les salaires de 10 euros. Que devient l'espérance ?
   -
2. Le directeur décide d'augmenter tous les salaires de 2%. Que devient l'espérance ?

Vos résultats seront validés par la détermination de la loi de probabilité puis le calcul de l'espérance de deux nouvelles variables aléatoires Y et Z associées à la variable aléatoire X par deux relations simples que vous déterminerez.

Une autre situation pour comprendre

On donne la loi de probabilité de la variable aléatoire C correspondant aux températures d'une région exprimées en degrés Celsius.

x_i	-1	3	5	10
$p(X = x_i)$	0,5	0,32	0,16	0,02

On rappelle que la relation permettant de transformer des degrés Celsius en degrés Fahrenheit est : $F = 1{,}8 \times C + 32$. Déterminer la loi de probabilité de la variable aléatoire F correspondant à la conversion des températures en Fahrenheit. Comparer les espérances de ces deux variables.

Une généralisation pour démontrer

Soit X une variable aléatoire dont la loi de probabilité est donnée par le tableau suivant :

Valeurs	x_1	x_2	$\dots$	x_n
Probabilités	p_1	p_2	$\dots$	p_n

On considère la variable aléatoire Y reliée à la variable aléatoire X par la relation $Y = a \times X + b$ et on donne ci-dessous la loi de probabilité de cette nouvelle variable aléatoire.

Valeurs	$a \times x_1 + b$	$a \times x_2 + b$	$\dots$	$a \times x_n + b$
Probabilités	p_1	p_2	$\dots$	p_n

Démontrer, en détaillant tous les calculs, la relation suivante : $E(a \times X + b) = a \times E(X) + b$.

Propriétés de la variance

Une situation pour une première conjecture

On considère ci-dessous la loi de probabilité d'une variable aléatoire X. Calculer $E(X)$, $V(X)$.

x_i	0	3	8	200
$p(X=x_i)$	0,4	0,3	0,2	0,1

On s'intéresse à variable aléatoire $Z = X+10$. Sauriez-vous déterminer la loi de probabilité de cette nouvelle variable aléatoire ? Calculer $E(Z)$ et $V(Z)$. Proposez un commentaire.

Rappeler la relation entre $E(X+b)$ et $E(X)$ pour tout nombre réel b.
Conjecturer la relation entre $V(X+b)$ et $V(X)$ pour tout nombre réel b.

Une autre situation pour une deuxième conjecture

On considère ci-dessous la loi de probabilité d'une variable aléatoire Y. Calculer $E(Y)$, $V(Y)$.

y_i	-4	5	10	100
$p(Y=y_i)$	0,25	0,2	0,4	0,15

On s'intéresse à variable aléatoire $Z = 2\times Y$. Sauriez-vous déterminer la loi de probabilité de cette nouvelle variable aléatoire ? Calculer $E(Z)$ et $V(Z)$. Proposez un commentaire.

Rappeler la relation entre $E(a\times X)$ et $E(X)$ pour tout nombre réel a.
Conjecturer la relation entre $V(a\times Y)$ et $V(Y)$ pour tout nombre réel a.

Une généralisation pour démontrer

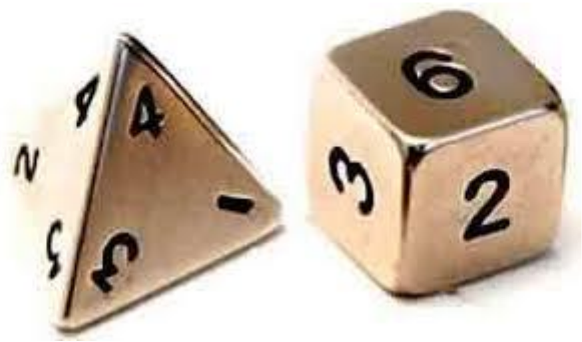
Soit X une variable aléatoire dont la loi de probabilité est donnée par :

x_i	x_1	x_2	...	x_n
$p(X=x_i)$	p_1	p_2	...	p_n

Montrer que $V(a\times X) = a^2 \times V(X)$ et que $V(X+b) = V(X)$ pour tout nombre réel a et b.

Somme de variables aléatoires

On lance deux dés équilibrés : un dé cubique et un dé tétraédrique. On s'intéresse à la somme des deux dés. On suppose que les deux lancers sont indépendants. On note X la variable aléatoire correspondant au résultat du premier dé et Y la variable aléatoire correspondant au résultat du deuxième dé. On s'intéresse donc à la variable aléatoire Z somme des variables aléatoires X et Y.



1. Dresser la loi de probabilité de X. Calculer E(X). Calculer V(X).
   Dresser la loi de probabilité de Y. Calculer E(Y). Calculer V(Y).

2. Dresser la loi de probabilité de Z. Calculer E(Z). Calculer V(Z).

3. Comparer l'espérance de Z et la somme des espérances de X et de Y.
   Comparer la variance de Z et la somme des variances de X et de Y.

4. Quelles conjectures cette étude permet-elle de formuler ?

Démonstration des conjectures dans un cas simple

On considère la variable aléatoire X prenant les valeurs x_1 ou x_2 de loi de probabilité $\begin{bmatrix} x_1 & x_2 \\ p_1 & p_2 \end{bmatrix}$.

On considère la variable aléatoire Y prenant les valeurs y_1 ou y_2 de loi de probabilité $\begin{bmatrix} y_1 & y_2 \\ q_1 & q_2 \end{bmatrix}$.

On considère la variable aléatoire Z somme des deux variables aléatoires X et Y. On suppose que ces deux variables aléatoires sont indépendantes.

1. Calculer E(X). Calculer V(X). Calculer E(Y). Calculer V(Y).

2. Dresser la loi de probabilité de Z. Calculer E(Z). Calculer V(Z).

3. Comparer l'espérance de Z et la somme des espérances de X et de Y.

4. Comparer la variance de Z et la somme des variances de X et de Y.

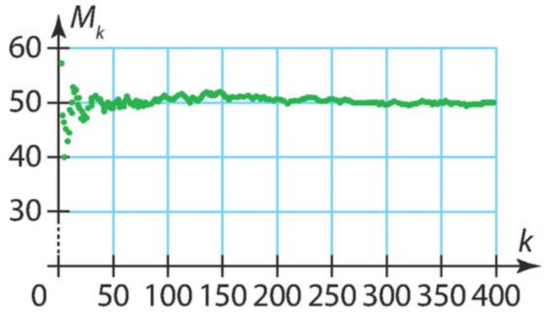
5. Enoncer clairement les deux formules que vous venez de démontrer…

Moyenne de variables aléatoires

Soient X_1, X_2, …, X_n n variables aléatoires indépendantes et $M_n = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ la variable aléatoire représentant la moyenne des n variables aléatoires précédentes. Montrer que $E(M_n) = \dfrac{E(X_1) + E(X_2) \ldots + E(X_n)}{n}$ et que $V(M_n) = \dfrac{V(X_1) + V(X_2) \ldots + V(X_n)}{n^2}$.

Visualisation de la loi des grands nombres

On considère un échantillon X_1, …, X_{400} de variables aléatoires indépendantes et identiquement distribués d'espérance notée m.
On note $M_n = \frac{X_1 + ... + X_n}{n}$ la moyenne des n premières variables aléatoires et on donne ci-contre l'allure du nuage des points $(n; M_n)$.

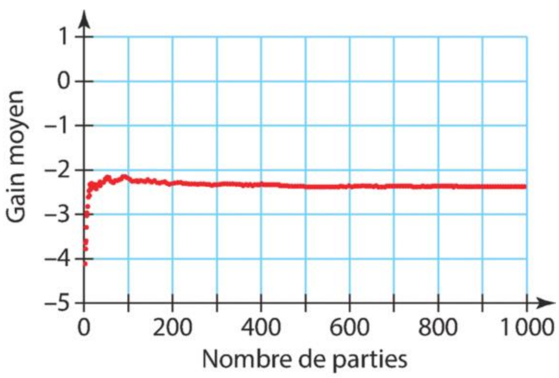


1. Est-on en mesure d'estimer la valeur de m par lecture graphique de ce nuage de points ?

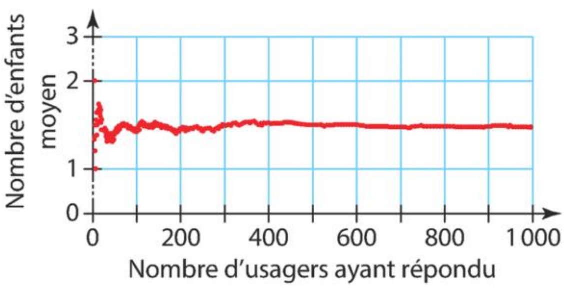
2. Préciser, pour tout réel $a > 0$, un majorant de la probabilité $p(|M_n - m| \geq a)$.
   Que vaut $\lim\limits_{n \to +\infty} p(|M_n - m| \geq a)$ quel que soit $a > 0$?

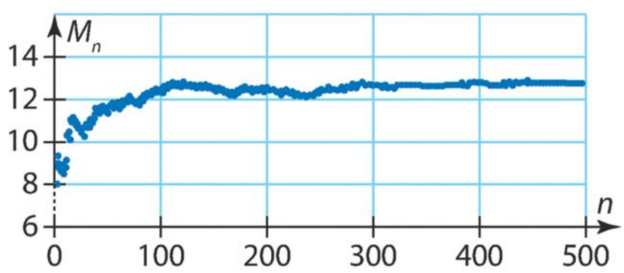
3. Sur quel théorème de calcul des limites se base le résultat de la question précédente.

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 $M_n = \frac{X_1 + ... + X_n}{n}$ 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

Application de l'inégalité de concentration

Situation 1

On estime que dans la population une personne sur dix est gauchère (c'est-à-dire « écrit avec la main gauche »). Combien de personnes doit-on réunir pour être sûr, au risque d'erreur de 5% (ou au seuil de confiance de 95%) que la moyenne de gauchers dans l'assemblée ainsi réunie soit strictement comprise entre 5 et 15% ?

Situation 2

Lorsque je joue aux fléchettes, mon taux de réussite pour atteindre le centre de la cible est de 60%. Combien de lancers dois-je envisager pour être sûr, au risque d'erreur de 5% (ou au seuil de confiance de 95%) que la moyenne des flèches ayant atteint le centre de la cible soit strictement comprise entre 55 et 65% ?

Situation 3

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. 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 4

On lance un dé équilibré. On appelle R la variable aléatoire donnant le résultat obtenu 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 3 et 4 ?

Situation 5

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 6

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 7

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 ?