Chapitre 14 Somme de variables aléatoires , Concentration , loi des grands nombres

1ère partie Somme de variables aléatoires

Contenus

- 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, exemples de variables indépendantes X, Y et relation d'additivité $V(X+Y) = V(X) + V(Y)$. Relation $V(aX) = a^2V(X)$.
- Application à l'espérance, la variance et l'écart type de la loi binomiale.
- Échantillon de taille n d'une loi de probabilité : liste $(X_1,\ldots,X_n)$ de variables indépendantes identiques suivant cette loi. Espérance, variance, écart type de la somme $S_n = X_1 + \ldots + X_n$ et de la moyenne $M_n = S_n/n$.

Capacités attendues

- Représenter une variable comme somme de variables aléatoires plus simples.
- Calculer l'espérance d'une variable aléatoire, notamment en utilisant la propriété de linéarité.
- Calculer la variance d'une variable aléatoire, notamment en l'exprimant comme somme de variables aléatoires indépendantes.

I- Somme de variables aléatoires

I Définition

Définition Soit X et Y deux variables aléatoires associées à une même expérience d'univers fini Ω et a un réel

- X+Y est la variable aléatoire définie sur Ω qui prend comme valeurs la somme des valeurs de X et de Y .
- aX est la variable aléatoire définie sur Ω qui prend comme valeurs le produit de a par X
- Linéarité de l'espérance $E(X+Y) = E(X) + E(Y)$ et $E(aX) = aE(X)$
- Si, de plus, les variables X et Y sont **indépendantes** :

  Additivité de la variance $V(X+Y)=V(X)+V(Y)$ et $V(aX)=a^2V(X)$

Remarque L'indépendance est prise ici au sens intuitif du terme c'est à dire que la réalisation de X n'influe pas sur la réalisation de Y .

Exemple On lance deux dés, l'un tétraédrique numéroté de 1 à 4 et l'autre cubique numéroté de 1 à 6 .
On appelle X et Y les variables aléatoires associées respectivement à ces deux expériences.

- X + Y est la variable aléatoire qui prend les valeurs de 2 à 10
- 2X est la variable aléatoire qui prend les valeurs 2 , 4 , 6 , 8
- X prend comme valeur 1 , 2 , 3 , 4 de probabilité $\frac{1}{4}$ donc $E(X)=\frac{1}{4}(1+2+3+4)$ = 2,5

  De même, on a : $E(Y)=\frac{1}{6}(1+2+3+4+5+6)$ = 3,5

  On a donc E(X+Y) = 2,5+3,5 = 6 et $E(2X)=2\times E(X)=5$
- $V(X)=\frac{1}{4}(1+4+9+16)$ – $2{,}5^2$ = 1,25 et $V(Y)=\frac{1}{6}(1+4+9+16+25+36)$ – $3{,}5^2$ = $\frac{35}{12}$ ≈2,92

  (On rappelle que V(X) = $\sum p_i x_i^2 - E(X)^2$)

  $V(X+Y)=V(X)+V(Y)$ ≈ 4,17 et $V(2X)=2^2V(X)=4\times 1{,}25=5$

II- Sommes de variables aléatoires identiques et indépendantes

II-1 Décomposition d'une variable aléatoire suivant une loi binomiale

Théorème **Soit n variables aléatoires indépendantes X_1 , X_2 , ... , X_n suivant la même loi de Bernoulli B(p).**
La variable aléatoire $S_n = X_1 + X_2 + \dots + X_n$ suit alors la loi binomiale $B(n\,;p)$

Explication Les variables aléatoires X_i suivant la même loi de Bernoulli, elles prennent comme valeur 0 (échec) ou 1 (succès) . Comme ces variables aléatoires X_i sont indépendantes , leur somme $X_1 + X_2 + \dots + X_n$ prendra comme valeur le nombre de succès pour n expériences de Bernoulli donc S_n suit la loi binomiale B(n;p)

Théorème Toute variable aléatoire suivant la loi binomiale B(n;p) peut se décomposer en une somme de n variables indépendantes
$S_n = X_1 + X_2 + \dots + X_n$ où X_i suit une même loi de Bernoulli B(p)

II-2 Echantillon d'une variable aléatoire

Définition Soit X une variable aléatoire suivant une loi de probabilité
Une liste de variables indépendantes (X_1, X_2 , ... , X_n) suivant cette même loi est appelée échantillon de taille n associé à X .
On pose $S_n = X_1 + X_2 + \dots + X_n$ et $M_n = \frac{S_n}{n}$, on a alors

$E(S_n) = nE(X)$, $E(M_n) = E(X)$ et $V(S_n) = nV(X)$, $V(M_n) = \frac{V(X)}{n}$

A noter M_n est la moyenne empirique des n variables aléatoires

Exemple :

Soit X une variable aléatoire dont la loi de probabilité est donnée ci-contre :

-10	5	20
$\frac{1}{4}$	$\frac{11}{20}$	$\frac{1}{5}$

On considère un échantillon (X_1, X_2 ... , X_n) de la loi suivie par X et la variable aléatoire moyenne M_n

1) Calculer $E(X_{11})$ et $E(S_{11})$

2) Calculer $V(M_{11})$

3) Déterminer la taille de l'échantillon n à partir de laquelle la variance M_n devient inférieure à 0,05

2ème partie : Concentration, loi des grands nombres

Contenus

– Inégalité de Bienaymé-Tchebychev. Pour une variable aléatoire X d'espérance μ et de variance V, et quel que soit le réel strictement positif δ : $P(|X-\mu| \geqslant \delta) \leqslant \frac{V(X)}{\delta^2}$.

– Inégalité de concentration. Si M_n est la variable aléatoire moyenne d'un échantillon de taille n d'une variable aléatoire d'espérance μ et de variance V, alors pour tout $\delta > 0$, $P(|M_n - \mu| \geqslant \delta) \leqslant \frac{V}{n\delta^2}$

– Loi des grands nombres.

Capacité attendue

– Appliquer l'inégalité de Bienaymé-Tchebychev pour définir une taille d'échantillon, en fonction de la précision et du risque choisi.

III- Concentration et loi des grands nombres

III-1 Inégalité de Bienaymé-Tchebychev

Théorème : Soit X une variable aléatoire d'espérance μ et de variance V

Pour tout $\delta \in]0;+\infty[$, $P(|X-\mu| \geq \delta) \leq \frac{V}{\delta^2}$

Autrement dit, la probabilité que X se trouve **en dehors** de l'intervalle $[X-\delta ; X+\delta]$ est inférieur à $\frac{V}{\delta^2}$

Exemple : La taille moyenne d'une femme française est de 1 m 65 et la variance est évaluée à 0,0025. Majorer la proportion des femmes françaises dont la taille est inférieure ou égale à 1,55 et supérieure ou égale à 1,75.

Soit X la variable aléatoire donnant la taille moyenne d'une femme française . On a donc $\mu = 1{,}65$ et V=0,0025.
On remarque que 1,55 et 1,75 sont distant de 0,1 de la moyenne μ donc en appliquant l'inégalité de Bienaymé-Tchebichev , on a : $P(|T-\mu| \geq 0{,}1) \leq \frac{V}{0{,}1^2}$ c'est à dire

$$P(|T-1{,}65| \geq 0{,}1) \leq \frac{0{,}0025}{0{,}1^2} = 0{,}25$$

Il y a donc au plus un quart de femmes françaises dont la taille est inférieure ou égale à 1,55 ou supérieure ou égale à 1,75

III-2 Inégalité de concentration

Théorème Soit ($X_1, X_2, \ldots, X_n$) un échantillon de variables aléatoires d'espérance μ et de variance V et M_n la variable aléatoire moyenne de cet échantillon .

Pour tout $\delta \in]0;+\infty[$, $P(|M_n-\mu| \geq \delta) \leq \frac{V}{n\delta^2}$

On rappelle que $M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$ avec $E(M_n) = \mu$ et $V(M_n) = \frac{V}{n}$

Pour démontrer ce résultat, il suffit d'appliquer l'inégalité de Bienaymé-Tchebichev à la variable aléatoire M_n .

On a donc : Pour tout $\delta \in]0;+\infty[$, $P(|M_n-\mu| \geq \delta) \leq \frac{V(M_n)}{\delta^2}$ avec $V(M_n) = \frac{V}{n}$ d'où

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

Exemple : Une variable aléatoire X suit la loi de Bernoulli de paramètres 0,13. Pour un échantillon de taille n, on note M_n la moyenne associée.

a) Justifier que $P(|M_n - 0{,}13| \geq 0{,}12) \leq \frac{277}{48n}$

E(X)=0,13 et V(X) = $0{,}13 \times (1-0{,}13) = 0{,}1131$. On applique alors l'inégalité de concentration :

$P(|M_n - 0{,}13| \geq 0{,}12) \leq \frac{0{,}1131}{n \times 0{,}12^2} = \frac{277}{48n}$

b) En déduire la valeur de n pour laquelle $P(|M_n - 0{,}13| \geq 0{,}12) \leq 0{,}05$.

Il suffit de résoudre $\frac{277}{48n} \leq 0{,}05$ ssi $\frac{48n}{277} \geq 0{,}05$ ssi $n \geq \frac{0{,}05 \times 277}{48}$ =157,08

Pour n ≥ 158, la probabilité qu'une valeur de M_n n'appartiennent pas à l'intervalle [0,01;0,25]est inférieure à 0,05

III-3 Loi des grands nombres

> Théorème : Soit ($X_1, X_2, \ldots, X_n$) un échantillon de variables aléatoires d'espérance μ et M_n la variable aléatoire moyenne de cet échantillon .
>
> Pour tout $\delta \in]0;+\infty[$, $\lim\limits_{n \to +\infty} P(|M_n - \mu| \geq \delta) = 0$

En d'autres termes, pour un δ donné aussi petit soit-il, la limite de la probabilité que M_n soit en dehors de l'intervalle $|\mu - \delta ; \mu + \delta|$ est nul. Ce qui montre de façon rigoureuse que lorsque l'on lance un grand nombre de fois une pièce de monnaie bien équilibrée, on a une chance sur deux en moyenne que la pièce tombe sur pile ou face