Variables aléatoires, concentration et loi des grands nombres

I Variables aléatoires $X+Y$ et aX

X et Y sont deux variables aléatoires définies sur l'univers d'une expérience aléatoire.

X prend les valeurs a_1, a_2, ..., a_n et Y prend les valeurs b_1, b_2, ..., b_n.

Variable aléatoire $X+Y$

- La variable aléatoire $X+Y$ prend toutes les valeurs possibles a_i+b_j avec $1 \leq i \leq n$ et $1 \leq j \leq m$.
- Loi de probabilité de $X+Y$: pour toute valeur w prise par $X+Y$, $P(X+Y=w)$ est la somme de toutes les probabilités $P\left(X=a_i \cap Y=b_j\right)$ où $a_i+b_j=w$.

Exemple : En période de réglage de ses machines, le contrôleur qualité d'une usine effectue de nombreux pélévement dans une production et relève, pour chaque pièce, si elle a un défaut A, un defaut B. Compte tenu du grand nombre de pièces prélévées, les fréquences indiquées dans ce tableau peuvent être assimmilées à des probabilités.

X (resp Y) est la variable éléatoire qui compte le nombre de défauts A (resp. B)

X et Y	0	1	2	Loi de X
0	$0,07$	$0,18$	$0,17$	$0,42$
1	$0,15$	$0,20$	$0,23$	$0,58$
Loi de Y	$0,22$	$0,38$	$0,40$	1

Compléter le tableau suivant avec les valeurs prises par la variable aléatoire somme $S=X+Y$

s	0	1	2	3
$P(S=s)$				

Variable aléatoire aX

a désigne un réel non nul.

- La variable aléatoire aX prend toutes les valeurs possibles $a \times a_i$ avec $1 \leq i \leq n$.
- Loi de probabilité de aX : pour toute valeur w prise par aX, $P(aX=w)$ est la somme de toutes les probabilités $P(X=a_i)$ où $a \times ai=w$.

Exemple : On lance un dé équilibré à six faces numérotés de 1 à 6.

X est la variable aléatoire qui donne le numéro obtenu et Y celle qui donne le double de ce numéro, alors $Y=$ ....

Linéarité de l'espérance

- $E(X+Y) = E(X)+E(Y)$
- $E(aX) = aE(X)$

Exemple : On reprend l'exemple du contrôleur de qualité

x	0	1
$P(X=x)$	0,42	0,58

y	0	1	2
$P(Y=y)$	0,22	0,38	0,40

s	0	1	2	3
$P(S=s)$	0,07	0,33	0,37	0,23

$E(X) =$ ....

$E(Y) =$ ....

$E(S) =$ ....

On constate qu'on a bien $E(X)+E(Y) =$ ....

Propriétés de la variance

- $V(X+Y) = V(X)+V(Y)$ si X et Y sont deux variables aléatoires **indépendantes**
- $V(aX) = a^2V(X)$

Exemple : On reprend l'exemple du contrôleur de qualité

- $V(S) =$ ....
- $\sigma(S) =$ ....

II Variables aléatoires indépendantes

Succession d'épreuves aléatoires indépendantes

On effectue successivement deux épreuves aléatoires indépendantes l'une de l'autre. Ainsi, l'issue de la première épreuve n'influence pas l'issue de la seconde.
On note X (resp Y) la variable aléatoire qui donne le résultat de la première (resp. seconde) épreuve.
On dit que les variables aléatoires sont **indépendantes**.

Conséquence

Les événements $X = a_i$ et $Y = b_j$ sont indépendants, donc $P\left(X = a_i \cap Y = b_j\right) = P\left(X = a_i\right) \times P\left(Y = b_j\right)$

Exemple : On lance successivement deux dés équilibrés, l'un a quatre faces numéroté de 1 à 4 et l'autre à 6 faces numérotées 0, 3, 3, 6, 6 et 6.
X (resp. Y) est la variable aléatoire qui donne le numéro obtenu avec la premier (resp. le second) dé.
Les variables aléatoires sont indépendantes car les deux lancers de dés le sont.

Ainsi, par exemple,$P(X = 1 \cap Y = 3) =$ ....

III Somme de variables aléatoires identiques et indépendantes

Somme de variables indépendantes suivant une même loi de Bernoulli

Si X_1, X_2, ..., X_n sont des variables aléatoires indépendantes suivant toutes une même loi de Bernoulli de paramètre p, alors $X_1 + X_2 + ... + X_n$ suit une loi binomiale de paramètres n et p.

Exemples : Si X_i suit une loi binomiale de paramètre $p = 0,17$ pour tout $1 \leq i \leq 12$, alors $X_1 + X_2 + ... + X_{12}$ suit une loi binomiale de paramètres $n = 12$ et $p = 0,17$.

Décomposition d'une variable aléatoire suivant une loi binomiale

Pour X suivant une loi binomiale de paramètres n et p, on a $X = X_1 + X_2 + ... + X_n$ où les variables aléatoires X_1, X_2, ..., X_n sont des variable aléatoires indépendantes suivant la même loi de Bernoulli de paramètre p.

Exemples : Si X suit une loi binomiale de paramètres $n = 3$ et $p = 0,4$, alors $X = X_1 + X_2 + X_3$ où X_1, X_2 et X_3 sont indépendantes et suivent une même loi de Bernoulli de paramètre $p = 0,4$.

Espérance et variance

Pour X suivant une loi binomiale de paramètres n et p, on a : $E(X) = np$ et $V(X) = np(1-p)$

Exemples : Une étude dans un restauranr montre que 85% des clients consomment un dessert.
On interroge 10 clients du restaurant. On supoose qu'on peut assimiler cette expérience à un tirage avec remise.
On note X la variable aléatoire correspondant au nombre de clients commandant un dessert parmi ceux interrogés.

1. Quelle loi suit la variable aléatoire X ?

   On répete 10 fois de manière supposée .... et .... l'expérience ayant deux issues :

   - succès : « le client choisi a choisi de consommé un dessert » avec probabilité $p = 0,85$
   - échec : « le client choisi n'a pas choisi de consommé un dessert » avec $1 - p =$ ....

   X compte le nombre de succès donc X suit la loi .... de paramètres $n = 10$ et $p = 0,85$

2. Calculer et interpréter l'espérance de la variable aléatoire X

   Comme X suit la loi binomiale de paramètres $n = 10$ et $p = 0,85$, alors $E(X) =$ ....

   Cela sigifie qu'au bout d'un grand nombre de répétitions de l'expérience, en moyenne sur 10 client ..... consomment un dessert.

3. Calculer la variance et l'écart-type (arrondi au millième) de la variable aléatoire X

   Comme X suit la loi binomiale de paramètres $n = 10$ et $p = 0,85$

   Alors $V(X) =$ ....

   Et $\sigma(X) =$ ....

IV Somme et moyenne d'un échantillon

Echantillon d'une variable aléatoire

Une liste $(X_1;X_2;...;X_n)$ de variables aléatoires indépendantes suivant toutes la même loi est appelé échantillon de taille n associé à cette loi.

Exemple : Bernard prend le même train cinq jours par semaine. On admet que la variable aléatoire X qui compte le nombre de retards suit la loi binomiale de paramètres $n = 5$ et $p = 0,1$.
En répétant cette expérience pendant 8 semaines, on construit un échantillon (X1 ; ... ; X8) de taille 8 de cette loi de probabilité.

Espérance et variance

En considérant un échantillon de taille n, $(X_1;...;X_n)$ d'une variable aléatoire X, et en posant $S_n = X_1 + X_2 + ... + X_n$ (variable aléatoire somme) et $M_n = \dfrac{S_n}{n}$ (variable aléatoire moyenne), on a :

- $E(S_n) = nE(X)$; $V(S_n) = nV(X)$ et $\sigma(S_n) = \sqrt{n}\sigma(X)$
- $E(M_n) = E(X)$; $V(M_n) = \dfrac{V(X)}{n}$ et $\sigma(M_n) = \dfrac{\sigma(X)}{\sqrt{n}}$

Exemple : Pendant la fête de l'école, Bianca a prévu de jouer 10 fois au jeu de la grenouille, dans lequel elle peut gagner un certain nombre de points, noté X.
Voici la loi de probabilité de X.

a	0	5	10	20	50	100
$P(X = a)$	$0,6$	$0,2$	$0,1$	$0,06$	$0,03$	$0,01$

Déterminer l'espérance et l'écart type du gain moyen par partie pour une série de 10 parties.

Les 10 parties sont .... les unes des autres, donc elles constituent un échantillon $(X_1;...;X_{10})$ de taille 10 de la loi de probabilité de X.
La variable aléatoire M_{10} moyenne de cet échantillon modélise alors le gain ....

On obtient $E(X) =$ ....

Et $\sigma(X) =$ ....

Donc $E(M_{10}) =$ ....

Et $\sigma(M_{10}) \approx$ ....

V L'inégalité de Bienaymé-Tchebychev

Inégalité de Markov

Soit X une variable aléatoire à valeurs positives et soit a un nombre réel strictement positif.
On a $p(\mathrm{X} \geq a) \leq \dfrac{\mathrm{E(X)}}{a}$

Interprétation : Cette propriété signifie que la probabilité que la variable aléatoire X prenne des valeurs plus grandes que a est d'autant plus petite que a est grande

Remarque : Si $a \leq \mathrm{E(X)}$, alors $\dfrac{\mathrm{E(X)}}{a} > 1$ et donc l'inégalité de Markov n'a pas d'intérêt. On dit alors qu'elle est triviale.
En effet la borne $\dfrac{\mathrm{E(X)}}{a}$ est alors supérieure à 1 et donc, nécessairement supérieure à $p(\mathrm{X} \geq a)$

Inégalité de Bienaymé-Tchebyche

Soit X une variable aléatoire et soit δ un nombre réel strictement positif.
On a $p(|\mathrm{X}-\mathrm{E(X)}| \geq \delta) \leq \dfrac{\mathrm{V(X)}}{\delta^2}$

Interprétation : La probabilité que les valeurs prises par la variable aléatoire X s'écartent d'au moins δ de l'espérance $\mathrm{E(X)}$ est d'autant plus petite que δ est grand

Remarque :

- $1-p(|\mathrm{X}-\mathrm{E(X)}| \geq \delta) = p(|\mathrm{X}-\mathrm{E(X)}| < \delta) = p(\mathrm{E(X)}-\delta < \mathrm{X} < \mathrm{E(X)}+\delta)$
- On dit que l'intervalle $[\mathrm{E(X)}-\delta ; \mathrm{E(X)}+\delta]$ est un intervalle de fluctuation de la variable aléatoire X

Exemple : Le nombre de pièces sortant d'une usine en une journée est une variable aléatoire d'espérance 50. On veut estimer la probabilité que la production de demain dépasse 75 pièces.

1. En utilisant l'inégalité de Markov, quelle estimation obtient-on sur cette probabilité ?

   Inégalité de Markov : $p(\mathrm{X} \geq 75) \leq$ ....

2. Que peut-on dire de plus sur cette probabilité si on sait que l'écart-type de la production quotidienne est 5 ?

   Inégalité de Tchebychev : $p(|\mathrm{X}-50| \geq 25) \leq$ ....

   Donc $p(\mathrm{X} \geq 75) \leq$ .....

VI Loi des grands nombres

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.

Autrement dit, $M_n = \frac{1}{n}\sum_{k=1}^{n} X_k$, où X_k sont indépendantes et de même loi de probabilité (celle de X).

Alors, pour tout réel $\delta > 0$, $p(|M_n - E(X)| \geq \delta) \leq \frac{V(X)}{n\delta^2}$

Cette inégalité est appelée **inégalité de concentration**

Exemple : On efectue n lancers successifs supposés indépendants d'une piece équilibrée.

- Avec 10 000 lancers, quelle la probabilité que la proportion de pile obtenue s'écarte de plus d'un centième de $\frac{1}{2}$?

  On associe à chaque tirage i la variable aléatoire x_i prenant la valeur 0 si on obtient face et 1 si on obtient pile.

  On pose $S_n = X_1 + X_2 + \ldots + X_n$ la variable aléatoire donnant le nombre de piles obtenus.

  On pose $M_n = \frac{S_n}{n}$

  On a pour $i \in \{1;\ldots;n\}$, $E(X_i) = \frac{1}{2}$ et $V(X_i) = \frac{1}{4}$

  Pour $n = 10\ 000$, l'inégalité de concertation donne $p\left(|M_n - \frac{1}{2}| \geq 0{,}01\right) \leq$ ....

  donc $p\left(|M_n - \frac{1}{2}| \geq 0{,}01\right) \leq$ ....

  Ainsi pour 10 000 lancers, la probabilité que la proportion de pile obtenue s'écarte de plus d'un centième de $\frac{1}{2}$ est inférieur à ....

- On souhaite que l'écart entre la proportion de pile obtenue et $\frac{1}{2}$ soit inférieur ou égal à $0{,}01$.

  Quelle est la valeur minimale de n pour que le risque d'erreur soit inférieur ou égale à 5% ?

  On doit résoudre l'inéquation $\frac{\frac{1}{4}}{n \times 0{,}01^2} \leq 0{,}05 \iff$ .... $\iff$ ....

  Il faut donc au moins .... lancers pour que le risque d'erreur soit inférieur ou égale à 5% .

Loi faible des grands nombres

Soit (X_n) un échantillon d'une variable aléatoire. On pose $M_n = \frac{1}{n}\sum_{k=1}^{n} X_k$

Alors, pour tout réel $\delta > 0$: $\lim_{n \to +\infty} p(|M_n - E(X)| \geq \delta) = 0$