

# 9 | Suites de variables aléatoires

## I – Suites de variables aléatoires

### 1 – Indépendance d'une famille de variables aléatoires

#### Définition 9.1 –

- Soient  $n \geq 2$  un entier et  $X_1, X_2, \dots, X_n$  des variables aléatoires. On dit que  $X_1, X_2, \dots, X_n$  sont **mutuellement indépendantes** lorsque pour tout  $n$ -uplet d'intervalles  $(I_1, I_2, \dots, I_n)$ ,

$$P(X_1 \in I_1, X_2 \in I_2, \dots, X_n \in I_n) = P(X_1 \in I_1) \times P(X_2 \in I_2) \times \dots \times P(X_n \in I_n).$$

- Soit  $(X_n)_{n \geq 1}$  une suite de variables aléatoires. On dit que la suite  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires mutuellement indépendantes si et seulement si pour tout entier  $m \in \mathbb{N}^*$ , les variables aléatoires  $X_1, X_2, \dots, X_m$  sont mutuellement indépendantes.

**Exemple 9.2 –** On lance une pièce équilibrée jusqu'à obtenir PILE. Pour tout  $n \in \mathbb{N}^*$ , on note  $X_n$  la variable aléatoire égale à 1 si on obtient PILE au  $n$ -ième lancer et 0 sinon. Alors la suite  $(X_n)_{n \geq 1}$  est une suite de variables aléatoires mutuellement indépendantes.

### 2 – Espérance et variance d'une famille de variables aléatoires

#### Théorème 9.3 – Linéarité de l'espérance

Soient  $n \geq 2$  un entier et  $X_1, X_2, \dots, X_n$  des variables aléatoires. Alors

$$E\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n E(X_k).$$

Autrement dit,

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n).$$

**Exemple 9.4 –** On considère une suite  $(X_k)_{k \in \mathbb{N}^*}$  de variables aléatoires mutuellement indépendantes, dont la loi commune est donnée pour tout  $k \in \mathbb{N}^*$  par

$$P(X_k = 1) = \frac{1}{3} \quad \text{et} \quad P(X_k = 2) = \frac{2}{3}.$$

Pour tout  $n \in \mathbb{N}^*$ , on pose  $S_n = \frac{1}{n} \sum_{k=1}^n X_k$ . Calculer  $E(S_n)$ .

**Théorème 9.5**

Soient  $n \geq 2$  un entier et  $X_1, X_2, \dots, X_n$  des variables aléatoires **mutuellement indépendantes**. Alors

$$V\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n V(X_k).$$

Autrement dit, sous réserve d'indépendance,

$$V(X_1 + X_2 + \dots + X_n) = V(X_1) + V(X_2) + \dots + V(X_n).$$

**Exemple 9.6** – On reprend l'exemple d'une suite  $(X_k)_{k \in \mathbb{N}^*}$  de variables aléatoires mutuellement indépendantes, dont la loi commune est donnée pour tout  $k \in \mathbb{N}^*$  par

$$P(X_k = 1) = \frac{1}{3} \quad \text{et} \quad P(X_k = 2) = \frac{2}{3}.$$

Pour tout  $n \in \mathbb{N}^*$ , on pose  $S_n = \frac{1}{n} \sum_{k=1}^n X_k$ . Calculer  $V(S_n)$ .

## II – Inégalités classiques en théorie des probabilités

### 1 – Inégalité de Markov

**Proposition 9.7 – Inégalité de Markov**

Soit  $X$  une variable aléatoire **positive** admettant une espérance. Alors

$$\forall a > 0, \quad P(X \geq a) \leq \frac{E(X)}{a}.$$

**Remarque 9.8 –**

- Cette inégalité se vérifie directement tant pour une variable aléatoire discrète que à densité.
- Si une variable aléatoire  $X^2$  admet une espérance, alors en particulier

$$\forall a > 0, \quad P(|X| \geq a) = P(X^2 \geq a^2) \leq \frac{E(X^2)}{a^2}.$$

## 2 – Inégalité de Bienaymé-Tchebychev

### Proposition 9.9 – Inégalité de Bienaymé-Tchebychev

Soit  $X$  une variable aléatoire, discrète ou à densité. On suppose que  $X$  admet une variance (donc que  $X^2$  admet une espérance). Alors

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

**Remarque 9.10** – Il s'agit de l'inégalité de Markov appliquée à la variable aléatoire  $(X - E(X))^2$ .

## 3 – Loi faible des grands nombres

### Théorème 9.11 – Loi faible des grands nombres

Soit  $(X_n)_{n \in \mathbb{N}^*}$  une suite de variables aléatoires mutuellement indépendantes, ayant chacune la même espérance  $m$  et la même variance. On pose  $\bar{X}_n = \frac{1}{n} \sum_{k=1}^n X_k$ .

Alors pour tout réel  $\varepsilon$  strictement positif,

$$\lim_{n \rightarrow +\infty} P(|\bar{X}_n - m| \geq \varepsilon) = 0.$$

**Remarque 9.12** – On reconnaît qu'il faut se servir de la loi faible des grands nombres lorsqu'il s'agit de calculer la limite d'une probabilité. Le sens de l'inégalité peut être retourné en considérant l'événement contraire. La seule présence de valeurs absolues dans la probabilité conduit elle à penser à l'utilisation de l'inégalité de Bienaymé-Tchebychev.

## III – Estimation

Les statisticiens connaissent, en général, le type de loi qui décrit tel ou tel phénomène, par l'observation. Mais souvent ils ne connaissent pas tous les paramètres de la dite loi. Ils doivent donc les estimer : c'est l'objectif de ce que l'on appelle la statistique inférentielle.

On considère une variable aléatoire  $X$ , dont le type de loi est connu et dépend d'un paramètre réel  $\theta$  inconnu (ce peut être le paramètre  $\lambda$  d'une variable exponentielle, l'étendue  $b - a$  d'une variable uniforme sur  $[a, b]$ , le paramètre  $p$  d'une variable de Bernoulli, l'espérance  $m$  d'une loi normale, etc).

L'objectif est de donner une *estimation* de la vraie valeur du paramètre  $\theta$ . Il existe deux façons de répondre à un tel problème : l'estimation ponctuelle et l'estimation par intervalle de confiance.



### Méthode 9.13 – Estimer ponctuellement le paramètre d'une loi

Pour estimer ponctuellement le paramètre d'une loi lorsque l'on connaît un échantillon de réalisations de cette loi, il suffit de faire la moyenne des valeurs observées.

**Exemple 9.14** – On considère un dé non équilibré que l'on lance 500 fois et on compte le nombre de fois où l'on obtient les différentes faces du dé.

Face du dé	1	2	3	4	5	6
Nombre d'apparitions	110	87	42	69	91	101

On considère la variable aléatoire  $X$  qui vaut 1 si le dé donne 6 et 0 sinon.

1. Reconnaître la loi suivie par  $X$ .

2. Donner une estimation à 0.01 près du paramètre de cette loi.

Les estimations ponctuelles ne fournissent pas d'information sur la précision des estimations, c'est-à-dire qu'elles ne tiennent pas compte de l'erreur possible attribuable aux fluctuations d'échantillonnage. Or deux échantillons distincts donnent presque certainement des valeurs distinctes pour l'estimation. Ici, il s'agit toujours d'estimer un paramètre inconnu, mais au lieu de lui attribuer une valeur unique en faisant appel à une estimation ponctuelle, on construit un intervalle aléatoire qui permet de "recouvrir" avec une certaine fiabilité, la vraie valeur du paramètre estimé.

**Définition 9.15** – Soient  $X$  une variable aléatoire suivant une loi de Bernoulli de paramètre  $p$  inconnu et  $a \in ]0, 1[$ . On appelle **intervalle de confiance** au niveau  $1 - a$ , tout intervalle  $I$  tel que

$$P(p \in I) \geq 1 - a.$$

Il s'agit d'un intervalle qui contient le paramètre  $p$  à estimer avec une probabilité minimale donnée.



**Méthode 9.16 – Déterminer un intervalle de confiance**

Déterminer un intervalle de confiance consiste à trouver un intervalle dont on sait qu'il contient le paramètre inconnu  $p$  d'une variable aléatoire de Bernoulli, avec une probabilité minimale donnée. Les différentes étapes pour déterminer cet intervalle de confiance sont toujours détaillées par l'énoncé. L'idée est d'appliquer l'inégalité de Bienaymé-Tchebychev, puis de choisir une valeur pour  $\varepsilon$  assurant le seuil de probabilité demandé. On donne ci-dessous un enchaînement classique de questions menant à la détermination d'un intervalle de confiance.

**Exemple 9.17** – On suppose que le paramètre  $p$  d'une loi de Bernoulli est inconnu et on cherche à l'estimer. Pour ce faire, on considère une suite de variables aléatoires  $(X_k)_{k \in \mathbb{N}^*}$  indépendantes, suivant toutes la loi de Bernoulli de paramètre  $p$ .

Pour tout  $n \in \mathbb{N}^*$ , on pose  $S_n = \frac{1}{n} \sum_{k=1}^n X_k$ .

1. a) Montrer que  $E(S_n) = p$ .

b) Calculer la variance de  $S_n$ .

2. a) On admet que pour tout  $p \in ]0, 1[$ ,  $p(1-p) \leq \frac{1}{4}$ . À l'aide de l'inégalité de Bienaymé-Tchebychev, montrer que pour tout  $\varepsilon > 0$ ,

$$P(|S_n - p| \leq \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

- b) Montrer que  $|S_n - p| \leq \varepsilon \iff p \in [S_n - \varepsilon, S_n + \varepsilon]$ .

- c) En déduire que l'intervalle  $\left[ S_n - \sqrt{\frac{5}{n}}, S_n + \sqrt{\frac{5}{n}} \right]$  est un intervalle de confiance de  $p$  au niveau de confiance 0.95.