

# **PROBABILITÉS ET STATISTIQUE**

---

AGRÉGATION – PROGRAMME COMMUN ET OPTION A

Si vous rencontrez des coquilles, n'hésitez pas à les signaler à  
[julien.chevallier1@univ-grenoble-alpes.fr](mailto:julien.chevallier1@univ-grenoble-alpes.fr), [raphael.rossignol@univ-grenoble-alpes.fr](mailto:raphael.rossignol@univ-grenoble-alpes.fr) ou  
[Loren.Coquille@univ-grenoble-alpes.fr](mailto:Loren.Coquille@univ-grenoble-alpes.fr).  
Merci.



# SOMMAIRE

---

<b>1</b>	<b>INTRODUCTION .....</b>	<b>7</b>
	<i>Objectifs - Différences entre probabilités et statistique - Bibliographie</i>	

I

## Programme commun

<b>2</b>	<b>PROBABILITÉS ET VARIABLES ALÉATOIRES .....</b>	<b>11</b>
<b>2.1</b>	<b>Espace probabilisé</b>	<b>11</b>
	<i>Cas discret - Cas continu - Indépendance et conditionnement - Propriétés asymptotiques</i>	
<b>2.2</b>	<b>Variables aléatoires</b>	<b>13</b>
	<i>Définitions et exemples - Fonction de répartition - Vecteur aléatoire - Tribu engendrée et indépendance - Espérance, variance, moments - Covariance et indépendance - Fonction caractéristique et génératrice</i>	
<b>3</b>	<b>CONVERGENCES DE VARIABLES ALÉATOIRES .....</b>	<b>17</b>
<b>3.1</b>	<b>Types de convergence</b>	<b>17</b>
	<i>Inégalités - Plus de détails sur la convergence en loi</i>	
<b>3.2</b>	<b>Théorèmes fondamentaux</b>	<b>18</b>

II

## Programme spécifique

<b>4</b>	<b>LOIS USUELLES .....</b>	<b>21</b>
<b>4.1</b>	<b>Modélisation</b>	<b>21</b>
	<i>Lois discrètes - Lois continues</i>	
<b>4.2</b>	<b>Méthodes de simulation</b>	<b>26</b>
	<i>Inversion - Rejet - Loi conjointe</i>	
<b>5</b>	<b>CHAÎNES DE MARKOV .....</b>	<b>27</b>
<b>5.1</b>	<b>Définitions</b>	<b>27</b>
	<i>Matrice de transition - Principaux exemples simples - Itérations de la chaîne</i>	
<b>5.2</b>	<b>Classification</b>	<b>29</b>
	<i>Communication entre états - Récurrence/Transience</i>	
<b>5.3</b>	<b>Mesure stationnaire</b>	<b>30</b>
	<i>Récurrence positive ou nulle - Exemple principal</i>	
<b>5.4</b>	<b>Théorèmes de convergence</b>	<b>31</b>
	<i>Loi des grands nombres - Convergence en loi</i>	
<b>5.5</b>	<b>Espace d'états fini</b>	<b>32</b>

<b>6</b>	<b>PROCESSUS DE POISSON .....</b>	<b>33</b>
6.1	<b>Définitions</b>	33
	<i>Deux construction possibles - Loi des temps de saut</i>	
6.2	<b>Propriétés</b>	34
	<i>Superposition et amincissement - Théorème de convergence</i>	
6.3	<b>Compléments classiques (pour votre culture)</b>	35
	<i>Pourquoi la loi de Poisson - Le paradoxe de l'autobus</i>	
<b>7</b>	<b>ESPÉRANCE CONDITIONNELLE ET MARTINGALES .....</b>	<b>37</b>
7.1	<b>Espérance conditionnelle</b>	37
	<i>Cas discret - Cas général</i>	
7.2	<b>Loi conditionnelle</b>	38
7.3	<b>Processus stochastique</b>	38
	<i>Définitions - Temps d'arrêt</i>	
7.4	<b>Martingales</b>	39
	<i>Martingales arrêtées - Théorèmes de convergence</i>	
7.5	<b>Complément classique (pour votre culture)</b>	40
<b>8</b>	<b>ESTIMATION .....</b>	<b>41</b>
8.1	<b>Introduction</b>	41
8.2	<b>Estimation ponctuelle</b>	41
	<i>Définitions - Méthodes - Principaux exemples et propriétés</i>	
8.3	<b>Estimation par région de confiance</b>	43
	<i>Région de confiance et quantile - Exemples</i>	
<b>9</b>	<b>VECTEURS GAUSSIENS .....</b>	<b>45</b>
9.1	<b>Définition</b>	45
9.2	<b>Applications importantes</b>	45
<b>10</b>	<b>MODÈLE LINÉAIRE GAUSSIEN .....</b>	<b>47</b>
10.1	<b>Cas général</b>	47
	<i>Définition - Estimation par moindres carrés - Intervalles de confiance et tests</i>	
10.2	<b>Régression linéaire</b>	48
	<i>Définition - Exemple</i>	
10.3	<b>Compléments classiques (pour votre culture)</b>	49
<b>11</b>	<b>TESTS PARAMÉTRIQUES .....</b>	<b>51</b>
11.1	<b>Généralités sur les tests</b>	51
11.2	<b>Test paramétrique</b>	51
	<i>Définitions - Exemples</i>	
11.3	<b>Compléments classiques (pour votre culture)</b>	53

<b>12 TESTS D'AJUSTEMENT .....</b>	<b>55</b>
<b>12.1 Test de Kolmogorov-Smirnov</b>	<b>55</b>
<i>Fonction de répartition empirique - Théorèmes de convergence - Ajustement à une loi</i>	
<b>12.2 Tests du khi-deux</b>	<b>56</b>
<i>Ajustement à une loi - Ajustement à une famille de lois</i>	



# 1 INTRODUCTION

## Objectifs

Les deux objectifs principaux de ces notes sont :

1. donner la liste exhaustive des définitions, méthodes et théorèmes qui apparaissent au programme commun et optionnel de l'agrégation,
2. obtenir un manuscrit le plus court possible en omettant les preuves des résultats énoncés ainsi que les résultats hors-programme.

Néanmoins, des résultats complémentaires qui nous paraissent importants sont souvent évoqués en fin de chapitre bien qu'ils n'apparaissent pas explicitement dans le programme.

Ces notes peuvent donc servir de base pour vous construire :

- une liste de références bibliographique,
- un schéma des ramifications entre les différents chapitres et résultats. Il est notamment recommandé d'avoir une idée de la démonstration de la plupart des résultats, notamment pour ceux qui sont des corollaires plus ou moins directs.

## Différences entre probabilités et statistique

Les probabilités et la statistique sont deux branches des mathématiques très inter-connectées. Cependant, leurs objectifs sont très différents.

D'un côté, la théorie des probabilités s'intéresse aux **événements futurs** et cherche, dans une certaine mesure, à les **prédirer** en utilisant une loi de probabilité connue (les paramètres du modèle sont **connus** et souvent considérés comme **fixés**).

De l'autre, la théorie de la statistique s'intéresse aux **événements passés** et cherche à **estimer** ou **ajuster** les données observées avec une loi de probabilité ou, plus généralement, une famille de lois de probabilité (les paramètres du modèle sont **inconnus** et souvent considérés comme **variables**).

En pratique, on utilise souvent une étude statistique préliminaire effectuée sur des données observées pour choisir une "bonne" loi de probabilité, et ainsi faire des prédictions fiables.

## Bibliographie

Dans l'esprit des objectifs principaux de ces notes, voici une liste succincte de livres de référence où le lecteur intéressé pourra trouver les démonstrations, ou de la bibliographie complémentaire, pour la quasi-totalité des énoncés mathématiques de ces notes.

- **M.-L. Chabanol, J.-J. Ruch.** *Probabilités et statistiques pour l'épreuve de modélisation à l'agrégation de mathématiques*. Ellipses, 2016.

Cet ouvrage fait le tour de toutes les notions de probabilités et statistique au programme commun et optionnel de l'agrégation. De plus, il contient des exercices de TP sur machine corrigés dans la syntaxe du logiciel Scilab. En revanche, certains résultats importants à nos yeux manquent et c'est pour cela que nous rajoutons les deux références suivantes.

- **P. Barbe, M. Ledoux.** *Probabilité (L3MI)*. EDP Sciences, 2012.

Cet ouvrage apporte de nombreux compléments de probabilités et peut servir à tout étudiant qui recherche des compléments plus élaborés.

- **V. Rivoirard, G. Stoltz.** *Statistique mathématique en action*. Vuibert, 2012.

C'est l'ouvrage de référence pour la statistique à l'usage du concours de l'agrégation. Toutes les notions sont abordées de manière rigoureuse et de très nombreux résultats

complémentaires sont énoncés. Pour les quelques résultats énoncés sans démonstration, un ouvrage de référence est renseigné pour le lecteur intéressé.



# *Programme commun*

<b>2</b>	<b>PROBABILITÉS ET VARIABLES ALÉATOIRES .....</b>	<b>11</b>
<b>3</b>	<b>CONVERGENCES DE VARIABLES ALÉATOIRES .....</b>	<b>17</b>



## 2 PROBABILITÉS ET VARIABLES ALÉATOIRES

### 2.1 Espace probabilisé

#### ■ Vocabulaire.

- |  |  |
|--|--|
| $\diamond \Omega$ : univers                                | $\diamond A \in \mathcal{F}$ : évènement       |
| $\diamond \omega \in \Omega$ : réalisation                 | $\diamond A \cap B$ : évènement “ $A$ et $B$ ” |
| $\diamond \mathcal{F} \subset \mathcal{P}(\Omega)$ : tribu | $\diamond A \cup B$ : évènement “ $A$ ou $B$ ” |

#### 2.1.1 Cas discret

Dans le cas discret, on prendra toujours  $\mathcal{F} = \mathcal{P}(\Omega)$  l'ensemble des parties de  $\Omega$ .

**Définition.** Soit  $f : \Omega \rightarrow [0, 1]$  telle que  $\sum_{\omega \in \Omega} f(\omega) = 1$ .

On appelle  $f$  une **fonction de masse** et la **mesure de probabilités associée** est  $\mathbb{P} : \mathcal{P}(\Omega) \rightarrow [0, 1]$  définie par

$$\mathbb{P}(A) := \sum_{\omega \in A} f(\omega).$$

**Proposition.** Soit  $(A_n)_{n \in \mathbb{N}}$  une suite d'évènements telle que  $A_i \cap A_j = \emptyset$  pour tout  $i \neq j$ . Alors

$$\mathbb{P}(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbb{P}(A_n).$$

■ **Exemple.** — Probabilité uniforme :  $\Omega = \{1, \dots, n\}$  et  $f \equiv 1/n$ .

— Lancer de pièce (biaisé) :  $\Omega = \{0, 1\}$  et  $f(0) = 1 - p$ ,  $f(1) = p$  avec  $p \in [0, 1]$ .

#### 2.1.2 Cas continu



Pourquoi ne pas se limiter à la théorie des probabilités discrètes ?

- Etudier les suites infinies de lancers de pièce.
- Donner un sens à un point aléatoire uniforme dans  $]0, 1]$ .

En général, on ne peut pas définir une probabilité sur  $\mathcal{P}(\Omega)$ . On doit se restreindre à un ensemble d'évènements plus petit (qui satisfait de bonnes propriétés).

**Définition.** Soit  $\mathcal{F}$  un sous-ensemble de  $\mathcal{P}(\Omega)$ . On dit que  $\mathcal{F}$  est une tribu si

- (i)  $\Omega \in \mathcal{F}$ ,
- (ii)  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ ,
- (iii)  $(A_n)_n \in \mathcal{F} \Rightarrow \bigcup_n A_n \in \mathcal{F}$ .

■ **Exemple.** — Tribu minimale :  $\mathcal{F}_{\min} = \{\emptyset, \Omega\}$ .

— Tribu des dénombrables :  $\mathcal{F}_{\text{dén}}$  = {ensembles dénombrables et codénombrables}.

**Proposition.** Soit  $(\mathcal{F}_t)_t$  une collection de tribus. Alors  $\mathcal{F} = \bigcap_t \mathcal{F}_t$  est une tribu.

**Définition.** Soit  $\mathcal{A} \subset \mathcal{P}(\Omega)$ . La **tribu engendrée** par  $\mathcal{A}$  est  $\sigma(\mathcal{A}) := \bigcap \mathcal{F}$  où l'intersection porte sur toutes les tribus  $\mathcal{F}$  qui contiennent  $\mathcal{A}$ . C'est la plus petite tribu qui contient  $\mathcal{A}$ .

- **Exemple.** — Si  $\mathcal{F}$  est une tribu, alors  $\sigma(\mathcal{F}) = \mathcal{F}$ .
- **Tribu borélienne :**  $\mathcal{B}(\Omega) := \sigma(\{U, U \text{ ouvert de } \Omega\})$  tribu sur un espace topologique  $\Omega$ .
- **Tribu produit :** Soit  $\Omega = \Omega_1 \times \Omega_2$  le produit de deux espaces munis des tribus  $\mathcal{F}_1$  et  $\mathcal{F}_2$ . La tribu produit est

$$\mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\{A_1 \times \Omega_2, A_1 \in \mathcal{F}_1\} \cup \{\Omega_1 \times A_2, A_2 \in \mathcal{F}_2\}).$$

**Définition.** On appelle **mesure de probabilité** toute application  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  telle que

- (i)  $\mathbb{P}(\Omega) = 1$ .
- (ii) Si  $A_1, A_2, \dots$ , sont deux à deux disjoints alors  $\mathbb{P}(\sqcup_n A_n) = \sum_n \mathbb{P}(A_n)$ .

**Proposition.**  $\diamond \quad \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$   $\diamond \quad A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$

$$\diamond \quad A_n \nearrow_{n \rightarrow +\infty} A \Rightarrow \mathbb{P}(A_n) \nearrow_{n \rightarrow +\infty} \mathbb{P}(A) \quad \diamond \quad \mathbb{P}(\cup_n A_n) \leq \sum_n \mathbb{P}(A_n)$$

$\diamond$  Formule du crible ou de Poincaré :

$$\mathbb{P}(\cup_{i=1}^n A_i) = \sum_{i=1}^n \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i \cap A_j) + \dots + (-1)^{n+1} \mathbb{P}(\cap_{i=1}^n A_i).$$

### 2.1.3 Indépendance et conditionnement

La connaissance de la réalisation d'un évènement  $A \in \mathcal{F}$  (i.e. on sait que  $\omega \in A$ ) change la probabilité de tous les évènements : en particulier, les chances d'être dans  $A^c$  deviennent nulles.

**Définition.** Soit  $A \in \mathcal{F}$  tel que  $\mathbb{P}(A) > 0$  et  $B \in \mathcal{F}$ . La **probabilité de  $B$  sachant  $A$**  est

$$\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}.$$

L'application  $\mathbb{P}(\cdot|A) : \mathcal{F} \rightarrow [0, 1]$  est une mesure de probabilité appelée **probabilité conditionnelle sachant  $A$** .

On peut noter que si la probabilité de  $B$  ne dépend pas de  $A$ , dans le sens où  $\mathbb{P}(B|A) = \mathbb{P}(B)$ , alors  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . D'où la notion d'indépendance.

**Définition.** On dit que  $A_1, A_2, \dots$  sont **indépendants** (sous  $\mathbb{P}$ ) si pour tout ensemble fini d'entiers  $I$ ,

$$\mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

On dit que  $\mathcal{F}_1, \mathcal{F}_2, \dots$  sont **indépendantes** (sous  $\mathbb{P}$ ) si pour tout ensemble fini d'entiers  $I$ ,

$$\forall (A_i)_{i \in I} \in \prod_{i \in I} \mathcal{F}_i, \quad \mathbb{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbb{P}(A_i).$$

En partitionnant l'univers  $\Omega$  et en conditionnant par des évènements bien choisis, il est possible de simplifier le calcul de  $\mathbb{P}(B)$ .

**Proposition.** Soit  $(A_n)_n$  une partition \* de  $\Omega$  telle que  $\mathbb{P}(A_n) > 0$ . Pour tout  $B \in \mathcal{F}$ ,

**Formule des probabilités totales :**

$$\mathbb{P}(B) = \sum_n \mathbb{P}(B|A_n) \mathbb{P}(A_n).$$

**Formule de Bayes :**

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k \cap B)}{\sum_n \mathbb{P}(B|A_n) \mathbb{P}(A_n)}.$$

---

\*. pour tout  $i \neq j$ ,  $A_i \cap A_j = \emptyset$  et  $\sqcup_n A_n = \Omega$ .

## 2.1.4 Propriétés asymptotiques

**Proposition — Lemmes de Borel Cantelli.** Soit  $(A_n)_n$  une suite d'évènements

- (i) Si  $\sum_n \mathbb{P}(A_n) < +\infty$ , alors  $\mathbb{P}(\limsup A_n) = 0$  (i.e.  $\{n, \omega \in A_n\}$  est un ensemble fini p.s.).
- (ii) Si  $\sum_n \mathbb{P}(A_n) = +\infty$  et que les évènements  $A_n$  sont indépendants, alors  $\mathbb{P}(\limsup A_n) = 1$  (i.e.  $\{n, \omega \in A_n\}$  est infini p.s.).

Soit  $(\mathcal{F}_n)_n$  une suite de tribus. On note  $\mathcal{G}_n = \sigma(\mathcal{F}_n, \mathcal{F}_{n+1}, \dots)$  les tribus de queue et  $\mathcal{G}_\infty = \cap_n \mathcal{G}_n$  la **tribu terminale**.

**Théorème — Loi du 0-1 de Kolmogorov.** Si les tribus  $\mathcal{F}_n$  sont indépendantes, alors la tribu terminale  $\mathcal{G}_\infty$  vérifie la loi du 0-1, i.e.

$$\forall A \in \mathcal{G}_\infty, \mathbb{P}(A) = 0 \text{ ou } \mathbb{P}(A) = 1.$$

## 2.2 Variables aléatoires

### 2.2.1 Définitions et exemples

La réalisation  $\omega$  est généralement inconnue. En revanche, on a souvent accès à des observables de cette réalisation : on les appelle **variables aléatoires**.

**Définition.** Soit  $(E, \mathcal{E})$  un espace mesurable. Une **variable aléatoire (v.a.)** est une application mesurable  $X : \Omega \rightarrow E$ .

La plupart du temps,  $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$  et on parle de v.a. réelle.



Dans la définition d'une v.a.  $X$ , l'espace  $\Omega$  peut sembler important. En général, **seule la loi de  $X$  nous intéresse** : elle ne dépend pas de l'ensemble de définition  $\Omega$ .

**Définition.** La **loi de la v.a.**  $X$  est la mesure  $\mathbb{P}_X$  sur  $(E, \mathcal{E})$  définie par

$$\forall C \in \mathcal{E}, \quad \mathbb{P}_X(C) := \mathbb{P}(X^{-1}(C)) \quad (\text{également noté } \mathbb{P}(X \in C)).$$

**Définition.** On dit que la loi de  $X$  est **discrète** si l'ensemble des valeurs possibles  $X(\Omega)$  est au plus dénombrable ou, plus précisément, si  $\mathbb{P}_X$  est une mesure discrète.

Cette loi est déterminée par la **fonction de masse**  $f_X : k \in E \mapsto \mathbb{P}(X = k)$  et peut être notée  $\mathbb{P}_X = \sum_{k \in E} f_X(k) \delta_k$  (cette somme est dénombrable même si  $E$  ne l'est pas car  $f_X$  est nulle sur  $E \setminus X(\Omega)$ ).

■ **Exemple.** — Loi uniforme  $\mathcal{U}(\{1, \dots, n\}) : f_X(k) = 1/n$ .

- Loi de Bernoulli  $\mathcal{B}(p) : f_X(0) = 1 - p$  et  $f_X(1) = p$  donc  $\mathbb{P}_X = (1 - p)\delta_0 + p\delta_1$ .
- Loi binomiale  $\mathcal{B}(n, p) : E = \{1, \dots, n\}$  et  $f_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ .
- Loi géométrique  $\mathcal{G}(p) : E = \mathbb{N}^*$  et  $f_X(k) = p(1 - p)^{k-1}$ .
- Loi de Poisson  $\mathcal{P}(\lambda) : E = \mathbb{N}$  et  $f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$ .

**Définition.** Soit  $X$  une v.a. à valeurs dans  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ . On dit que la loi de  $X$  est **absolument continue (ou à densité)** si la mesure  $\mathbb{P}_X$  est absolument continue par rapport à Lebesgue. Dans ce cas, la loi est déterminée par la densité  $f_X$  (unique à un ensemble de mesure nulle près) qui vérifie

$$\forall C \in \mathcal{B}(\mathbb{R}^d), \quad \mathbb{P}_X(C) = \int_C f_X(x) dx.$$

■ **Exemple.** — Loi uniforme  $\mathcal{U}([a, b]) : f_X(x) = 1/(b - a) \mathbf{1}_{[a, b]}(x)$ .

- Loi exponentielle  $\mathcal{E}(\lambda) : f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{[0, +\infty[}(x)$ .
- Loi de Gauss  $\mathcal{N}(\mu, \sigma^2) : f_X(x) = (\sqrt{2\pi}\sigma)^{-1} \exp(-(x - \mu)^2/2\sigma^2)$ .

## 2.2.2 Fonction de répartition

**Définition.** Soit  $X$  une v.a. réelle. On appelle **fonction de répartition** la fonction  $F_X : \mathbb{R} \rightarrow [0, 1]$  définie par, pour tout  $t \in \mathbb{R}$ ,

$$F_X(t) := \mathbb{P}(X \leq t).$$

**Proposition.** Toute fonction de répartition  $F_X$  est croissante, càdlàg<sup>†</sup>,  $F_X(-\infty) = 0$  et  $F_X(\infty) = 1$ .

**Proposition.** La fonction de répartition  $F_X$  caractérise la loi de  $X$  (i.e.  $F_X = F_Y$  implique  $\mathbb{P}_X = \mathbb{P}_Y$ ). De plus,

- (i) si la loi de  $X$  est discrète, alors  $F_X$  est une fonction en escalier;
- (ii) si la loi de  $X$  est absolument continue, alors  $F_X$  est absolument continue.



Les discontinuités de la fonction de répartition correspondent aux atomes de la loi de  $X$ . **Attention :** il existe des fonctions de répartitions continues qui correspondent à des mesures continues singulières (e.g. l'escalier de Cantor).

La croissance de  $F_X$  permet de définir son **inverse généralisé**<sup>‡</sup> : on utilisera la notation  $F_X^{-1}$ .

**Définition.** Soit  $X$  une v.a. réelle. On appelle **fonction quantile** la fonction inverse généralisée de  $F_X$  définie par, pour tout  $p \in ]0, 1[$ ,

$$F_X^{-1}(p) := \inf\{t \in \mathbb{R}, F_X(t) \geq p\}.$$

**Proposition.** Soit  $U \sim \mathcal{U}([0, 1])$  et  $F$  une fonction de répartition, alors la v.a.  $Y = F^{-1}(U)$  admet  $F$  comme fonction de répartition.

## 2.2.3 Vecteur aléatoire

**Définition.** Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $E_1$  et  $E_2$ . Le couple  $(X, Y)$  est une v.a. à valeurs dans  $E_1 \times E_2$ . Sa loi est la **loi conjointe** de  $X$  et  $Y$  définie par

$$\forall C \in \mathcal{E}_1 \otimes \mathcal{E}_2, \quad \mathbb{P}_{(X,Y)}(C) := \mathbb{P}((X, Y) \in C).$$

La loi de  $X$  est appelée **loi marginale** et se retrouve par

$$\forall C \in \mathcal{E}_1, \quad \mathbb{P}_X(C) = \mathbb{P}_{(X,Y)}(C \times E_2).$$



Les lois marginales, et plus généralement la loi de  $Z = \varphi(X, Y)$ , sont déterminées par la loi jointe. En revanche, les **lois marginales ne suffisent pas à déterminer la loi jointe** : il faudrait connaître la loi de  $Z$  pour toute fonction  $\varphi$  mesurable bornée.

## 2.2.4 Tribu engendrée et indépendance

**Définition.** La tribu engendrée par une v.a.  $X$  est définie par  $\sigma(X) := \{X^{-1}(C), C \in \mathcal{E}\}$ .

**Proposition.** Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $E_1$  et  $E_2$ . Sont équivalents :

- (i) La v.a.  $Y$  est  $\sigma(X)$ -mesurable.
- (ii) Il existe  $h : E_1 \rightarrow E_2$  mesurable telle que  $Y = h(X)$ .

La notion d'indépendance entre v.a. est un cas particulier de la notion la plus générale : indépendance entre tribus (voir page 12).

**Définition.** On dit que les v.a.  $X_1, X_2, \dots$  sont **indépendantes** si les tribus  $\sigma(X_1), \sigma(X_2), \dots$  le sont.

---

<sup>†</sup>. continue à droite, limite à gauche.

<sup>‡</sup>. Si  $F_X$  est bijective alors son inverse généralisé est son inverse classique.

**Proposition.** Soient  $X_1, \dots, X_n$  des v.a. et  $X = (X_1, \dots, X_n)$ . Sont équivalents :

- (i) Les v.a.  $X_1, \dots, X_n$  sont indépendantes.
- (ii) La loi de  $X$  est le produit des lois marginales, i.e.  $\mathbb{P}_X = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ .

■ **Exemple.** — Soient  $X$  et  $Y$  deux v.a. discrètes. Alors  $Z = (X, Y)$  est une v.a. discrète et

$$X \text{ et } Y \text{ sont indépendantesssi } f_{(X,Y)}(k_1, k_2) = f_X(k_1)f_Y(k_2).$$

— Soient  $X$  et  $Y$  deux v.a. à densité. Alors  $Z = (X, Y)$  n'est pas nécessairement une v.a. à densité, mais

$$X \text{ et } Y \text{ sont indépendantesssi } Z \text{ admet la densité } f_{(X,Y)}(x, y) = f_X(x)f_Y(y).$$

## 2.2.5 Espérance, variance, moments

**Définition.** Soit  $X$  une v.a. réelle. Si  $\int_{\Omega} |X(\omega)| d\mathbb{P}(\omega) < +\infty$ , on dit que  $X$  est intégrable et on définit son espérance par

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) \mathbb{P}(d\omega).$$



Encore une fois, l'espace  $\Omega$  semble important dans cette définition. Il n'en est rien : l'espérance ne dépend que de la loi de  $X$  comme le montre le Théorème de transfert ci-dessous. De plus, c'est souvent cette formule qui sert en pratique.

**Théorème de Transfert.** Soit  $X$  un vecteur aléatoire dans  $\mathbb{R}^d$  et  $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$  mesurable positive. Alors

$$\mathbb{E}[\varphi(X)] = \int_{\Omega} \varphi(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}^d} \varphi(x) \mathbb{P}_X(dx).$$

Ce résultat s'étend naturellement aux fonctions de signe quelconque : soit  $\varphi$  une fonction mesurable telle que

$$\mathbb{E}[|\varphi(X)|] := \int_{\Omega} |\varphi(X(\omega))| \mathbb{P}(d\omega) = \int_{\mathbb{R}^d} |\varphi(x)| \mathbb{P}_X(dx) < +\infty,$$

alors on dit que  $\varphi(X) \in L^1$  et  $\mathbb{E}[\varphi(X)] := \int_{\Omega} \varphi(X(\omega)) \mathbb{P}(d\omega) = \int_{\mathbb{R}^d} \varphi(x) \mathbb{P}_X(dx)$  est bien définie. Inversement, la connaissance des quantités  $\mathbb{E}[\varphi(X)]$  pour une classe de fonctions  $\varphi$  suffisamment générale détermine complètement la loi de  $X$ .

**Proposition.** Soit  $X$  une v.a. réelle et  $\mathbb{Q}$  une mesure sur  $\mathbb{R}$ . Si  $\mathbb{Q}$  vérifie

$$\forall \varphi \text{ mesurable bornée}, \quad \mathbb{E}[\varphi(X)] = \int \varphi(x) \mathbb{Q}(dx),$$

alors  $\mathbb{Q}$  est la loi de  $X$ .

■ **Exemple.** — Le moment d'ordre  $k \in \mathbb{N}$  d'une v.a. est donné par  $\mathbb{E}[X^k]$  et bien défini dès que  $\mathbb{E}[|X|^k] < +\infty$ . Le moment d'ordre  $k$  existessi  $X \in L^k$ .

— La variance d'une v.a. est donnée par  $\text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$  qui peut être infinie.

**Proposition.** Soient  $a$  et  $b$  deux réels. On a :

- $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$  et  $X \geq 0 \Rightarrow \mathbb{E}[X] \geq 0$ ;
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$ ,  $\text{Var}(X) = 0 \Rightarrow X$  est constante p.s.,  $\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ ;
- $\mathbb{E}[X]$  est la projection orthogonale (dans  $L^2$ ) de  $X$  sur l'espace des v.a. constantes car

$$\mathbb{E}[(X - a)^2] = \text{Var}(X) + (\mathbb{E}[X] - a)^2.$$

## 2.2.6 Covariance et indépendance

**Définition.** La covariance de deux v.a.  $X$  et  $Y$  est donnée par

$$\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y].$$

En particulier, elle est bien définie dès que  $X$  et  $Y$  sont dans  $L^2$ .

**Proposition.** La covariance est un opérateur bilinéaire et  $\text{Cov}(X, X) = \text{Var}(X)$ .

**Théorème.** Sont équivalents :

- (i) Les v.a.  $X$  et  $Y$  sont indépendantes.
- (ii) Pour toutes fonctions  $f$  et  $g$  mesurables bornées,  $\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$ .

Si elle existe, la covariance de deux v.a. indépendantes est nulle. **La réciproque est fausse.**

**Définition.** Le coefficient de corrélation de  $X$  et  $Y$  est donné par

$$\rho(X, Y) := \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

**Proposition.** Le coefficient de corrélation vérifie  $|\rho(X, Y)| \leq 1$  et sont équivalents :

- (i)  $|\rho(X, Y)| = 1$ .
- (ii) Il existe  $a \neq 0$  et  $b \in \mathbb{R}$  tels que  $Y = aX + b$ .

## 2.2.7 Fonction caractéristique et génératrice

**Définition.** La fonction caractéristique d'une v.a.  $X$  à valeurs dans  $\mathbb{R}^d$  est définie par

$$\forall t \in \mathbb{R}^d, \quad \Phi_X(t) := \mathbb{E}[e^{i\langle t, X \rangle}] = \int_{\mathbb{R}^d} e^{i\langle t, x \rangle} \mathbb{P}_X(dx).$$

C'est la transformée de Fourier de la loi  $\mathbb{P}_X$ . Elle détermine donc cette dernière.

**Proposition.** Toute fonction caractéristique  $\Phi_X$  est continue bornée sur  $\mathbb{R}^d$  et  $\Phi_X(0) = 1$ . Si  $X$  est réelle (dimension 1) et dans  $L^k$ , alors  $\Phi_X$  est  $\mathcal{C}^k$  et

$$\Phi_X^{(k)}(0) = i^k \mathbb{E}[X^k].$$

**Définition.** La fonction génératrice d'une v.a.  $X$  à valeur dans  $\mathbb{N}$  est donnée par

$$\forall z \in \mathbb{C}, \quad G_X(z) := \mathbb{E}[z^X] = \sum_{k=0}^{+\infty} z^k \mathbb{P}(X = k).$$

Le rayon de convergence de cette série est au moins égal à 1. Elle détermine la loi  $\mathbb{P}_X$ .

**Proposition — Génération des moments.** Toute fonction génératrice  $G_X$  vérifie  $G_X(1) = 1$  et, si ces quantités sont bien définies,  $G'_X(1) = \mathbb{E}[X]$ ,  $G''_X(1) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$  etc...

**Théorème — Somme et indépendance.** Soient  $X$  et  $Y$  deux v.a. indépendantes. On a :

$$\Phi_{X+Y} = \Phi_X \Phi_Y \quad \text{et} \quad G_{X+Y} = G_X G_Y.$$

# 3 CONVERGENCES DE VARIABLES ALÉATOIRES

## 3.1 Types de convergence

**Définition.** Soient  $(X_n)_n$  une suite de v.a. et  $X$  une v.a. On dit que  $X_n$  converge vers  $X$  :

- **presque-sûrement**, noté  $X_n \xrightarrow{p.s.} X$ , si  $\mathbb{P}(\{\omega, X_n(\omega) \rightarrow X(\omega)\}) = 1$ ;
- **en probabilité**, noté  $X_n \xrightarrow{\mathbb{P}} X$ , si  $\forall \varepsilon > 0, \mathbb{P}(d(X_n, X) > \varepsilon) \rightarrow 0$ ;
- **dans  $L^p$** , noté  $X_n \xrightarrow{L^p} X$ , si  $X \in L^p$  et  $\mathbb{E}[d(X_n, X)^p] \rightarrow 0$ ;
- **en loi**, noté  $X_n \xrightarrow{\mathcal{L}} X$ , si  $\forall \varphi$  continue bornée,  $\mathbb{E}[\varphi(X_n)] \rightarrow \mathbb{E}[\varphi(X)]$ .

Ces quatre notions dépendent de la distance que l'on met sur l'espace  $E$  des valeurs de  $X$ .

\*. L'espace probabilisé de  $X_n$  peut être différent de celui de  $X$ .

### Théorème — Lien entre les notions de convergence.

Les implications suivantes sont vraies en général, et celles-ci sont les seules.

$$\begin{array}{ccccccc} \xrightarrow{L^r} & \Rightarrow & \xrightarrow{L^p} & & & & \\ & r > p \geq 1 & & & & & \\ & & & \downarrow & & & \\ \xrightarrow{p.s.} & \Rightarrow & \xrightarrow{\mathbb{P}} & \Rightarrow & \xrightarrow{\mathcal{L}} & & \end{array}$$

### Proposition — Implications réciproques.

- (i) Si  $X_n \xrightarrow{\mathbb{P}} X$  et de plus  $\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < +\infty$  pour tout  $\varepsilon > 0$ , alors  $X_n \xrightarrow{p.s.} X$ .
- (ii) Si  $X_n \xrightarrow{\mathbb{P}} X$ , alors il existe une sous-suite  $n_k$  telle que  $X_{n_k} \xrightarrow{p.s.} X$ .
- (iii) Si  $X_n \xrightarrow{\mathbb{P}} X$  et qu'il existe  $r$  tel que  $\sup_n \mathbb{E}[|X_n|^r] < +\infty$ , alors  $X_n \xrightarrow{L^p} X$  pour tout  $p < r$ .
- (iv) Si  $X_n \xrightarrow{\mathcal{L}} a$  constante, alors  $X_n \xrightarrow{\mathbb{P}} a$ .

### Proposition — Opérations sur les limites.

- Soit  $(X_n)_n$  une suite de v.a. à valeurs dans  $\mathbb{R}^d$  qui converge vers  $X$ , p.s. ou en probabilité ou en loi, et  $f$  une fonction continue de  $\mathbb{R}^d$  dans  $\mathbb{R}^m$ . Alors la suite  $f(X_n)$  converge vers  $f(X)$  dans le même sens.
- Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de v.a. réelles qui convergent vers  $X$  et  $Y$ , p.s. ou en probabilité ou dans  $L^p$ . Alors  $(X_n + Y_n)_n$  converge vers  $X + Y$  dans le même sens.



Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $E$  et  $F$ . La convergence des deux v.a. implique souvent la convergence du couple. Néanmoins, il y a de nombreux contre-exemples : convergence en loi, distance exotique sur  $E \times F$ , etc. Il est conseillé de faire la démonstration si vous souhaitez utiliser un résultat de ce type.

### 3.1.1 Inégalités

**Proposition.** Soit  $X$  une v.a. et  $f$  une fonction mesurable positive réelle. Alors

$$\forall t > 0, \quad \mathbb{P}(f(X) \geq t) \leq \frac{\mathbb{E}[f(X)]}{t}.$$

### Corollaire.

**Markov :**

$$\mathbb{P}(|X| \geq t) \leq \frac{\mathbb{E}[|X|]}{t}.$$

**Bienaymé-Tchebychev :**

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}(X)}{t^2}.$$

### 3.1.2 Plus de détails sur la convergence en loi



La convergence en loi est assez déroutante car elle concerne des objets qui ne sont pas nécessairement définis sur le même espace. Il serait en fait plus juste de parler de convergence en loi uniquement pour les lois des v.a. et non pour les v.a. elles-mêmes : c'est en particulier pour cela que l'on parle de la convergence en loi de  $(X_n)_n$  vers la loi normale  $\mathcal{N}(0, 1)$  par exemple. **Beaucoup de résultats vérifiés par les notions de convergence classiques ne le sont pas pour la convergence en loi.**

**Théorème de Lévy.** Soit  $(X_n)_n$  une suite de v.a. de fonction caractéristique  $\Phi_{X_n}$ .

Si  $\Phi_{X_n}(t) \rightarrow \Phi(t)$  pour tout  $t \in \mathbb{R}^d$  et que  $\Phi$  est continue en 0, alors  $\Phi$  est une fonction caractéristique et  $X_n \xrightarrow{\mathcal{L}} X$  (où  $X$  est une v.a. de fonction caractéristique égale à  $\Phi$ ).

Il existe l'équivalent avec les fonctions génératrices :  $X_n \xrightarrow{\mathcal{L}} X$ ssi  $G_{X_n}(z) \rightarrow G_X(z)$  pour tout  $|z| < 1$ .

**Proposition.** Soient  $(X_n)_n$  une suite de v.a. de fonction de répartition  $F_{X_n}$  et  $X$  une v.a. de fonction de répartition  $F_X$ . Sont équivalents :

1.  $X_n \xrightarrow{\mathcal{L}} X$ ,
2.  $F_{X_n}(t) \rightarrow F_X(t)$  pour tout  $t$  point de continuité de  $F_X$ .

**Lemme de Slutsky.** Soient  $(X_n)_n$  et  $(Y_n)_n$  deux suites de v.a. réelles définies sur le même espace telles que  $X_n \xrightarrow{\mathcal{L}} X$  et  $Y_n \xrightarrow{\mathcal{L}} a$  constante. Alors  $(X_n, Y_n) \xrightarrow{\mathcal{L}} (X, a)$ .

## 3.2 Théorèmes fondamentaux

**Théorème — Loi faible des grands nombres.** Soit  $(X_n)_n$  une suite de v.a. indépendantes et identiquement distribuées (i.i.d.) dans  $L^1$  d'espérance  $\mu$ . Alors

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\mathbb{P}} \mu.$$

**Théorème — Loi forte des grands nombres.** Soit  $(X_n)_n$  une suite de v.a. i.i.d. dans  $L^4$  d'espérance  $\mu$ . Alors

$$\bar{X}_n \xrightarrow{p.s.} \mu.$$

— L'hypothèse  $L^4$  n'est pas nécessaire ( $L^1$  suffit) mais simplifie énormément la preuve.

**Théorème Central Limite.** Soit  $(X_n)_n$  une suite de v.a. i.i.d. dans  $L^2$  d'espérance  $\mu$  et de variance  $\sigma^2$ . Alors

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

# *Programme spécifique*

<b>4</b>	<b>LOIS USUELLES</b>	<b>21</b>
<b>5</b>	<b>CHAÎNES DE MARKOV</b>	<b>27</b>
<b>6</b>	<b>PROCESSUS DE POISSON</b>	<b>33</b>
<b>7</b>	<b>ESPÉRANCE CONDITIONNELLE ET MARTINGALES</b>	<b>37</b>
<b>8</b>	<b>ESTIMATION</b>	<b>41</b>
<b>9</b>	<b>VECTEURS GAUSSIENS</b>	<b>45</b>
<b>10</b>	<b>MODÈLE LINÉAIRE GAUSSIEN</b>	<b>47</b>
<b>11</b>	<b>TESTS PARAMÉTRIQUES</b>	<b>51</b>
<b>12</b>	<b>TESTS D'AJUSTEMENT</b>	<b>55</b>



# 4 LOIS USUELLES

## 4.1 Modélisation

### 4.1.1 Lois discrètes

Pour chacune des lois, nous listons : sa notation et ses paramètres usuels, son support, sa fonction de masse, son espérance, sa variance, l'expression de sa fonction caractéristique et la commande Python utilisée pour la simulation à partir du module `numpy.random` importé en tant que `npr`.

#### ► Loi uniforme

$\diamond X \sim \mathcal{U}(E);$		<code>npr.randint(a, b+1)</code>
$\diamond E$ est un ensemble fini (e.g. $E = [[a, b]]$ );		lancer de pièces ou de dés équilibrés
$\diamond f_X(k) = \frac{1}{\#E}$ (e.g. $\#E = n = b - a + 1$ );		jeu de la roulette
$\diamond \mathbb{E}[X] = (a + b)/2$ , $\text{Var}(X) = \frac{n^2 - 1}{12}$ ;		jeux de cartes
$\diamond \Phi_X(t) = \frac{e^{iat}}{n} \sum_{k=0}^{n-1} e^{ikt}.$		statistique d'ordre (permutation) date d'anniversaire

Le paramètre  $E$  représente le support. La loi uniforme modélise une expérience aléatoire dont les issues sont **équiprobables**. La loi uniforme est le premier exemple de loi de probabilité étudié au lycée. Les calculs se ramènent à du **dénombrement** et des arguments combinatoires.

**Proposition.** Soient  $X \sim \mathcal{U}(E)$  et  $f : E \rightarrow F$ . La v.a.  $f(X)$  est uniforme sur  $F$ ssi le cardinal de  $f^{-1}(\{y\})$  ne dépend pas de  $y \in F$ .

#### ► Loi binomiale

$\diamond X \sim \mathcal{B}(n, p)$ , $n \in \mathbb{N}$ , $p \in [0, 1]$ ;		<code>npr.binomial(n, p)</code>
$\diamond E = \{0, \dots, n\}$ ;		lancer de pièces
$\diamond f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$ ;		sélection naturelle
$\diamond \mathbb{E}[X] = np$ , $\text{Var}(X) = np(1-p)$ ;		tirage de boules avec remise
$\diamond \Phi_X(t) = (1 - p + pe^{it})^n$ .		sondage loi hypergéométrique

Le paramètre  $p$  représente la probabilité de succès d'une expérience aléatoire à deux issues (pile/face, succès/échec, etc.). Le paramètre  $n$  représente le nombre de répétitions de cette expérience. En particulier, lorsque  $n = 1$ , on parle de **loi de Bernoulli**, notée  $\mathcal{B}(p)$ . La loi binomiale modélise le nombre de succès lors de la répétition de  $n$  expériences identiques et indépendantes (**tirages avec remise**).

**Proposition — Stabilité.** Soient  $X_1$  et  $X_2$  deux v.a. indépendantes de lois respectives  $\mathcal{B}(n_1, p)$  et  $\mathcal{B}(n_2, p)$ . Alors  $X_1 + X_2$  suit la loi  $\mathcal{B}(n_1 + n_2, p)$ .

D'un point de vue historique, la loi binomiale est la première loi pour laquelle a été démontré le Théorème Central Limite (par de Moivre en 1733 pour  $p = 1/2$  et Laplace en 1812 dans le

cas général). Cette convergence est très bien illustrée par l'expérience de la **planche de Galton**.

### ► Loi multinomiale

La loi multinomiale est la loi de probabilité d'un **vecteur aléatoire**. Les lettres en gras ci-dessous désignent des vecteurs. En particulier,  $\text{Var}(\mathbf{N})$  désigne la matrice de covariance du vecteur aléatoire  $\mathbf{N}$  et  $\delta_{ij}$  est le symbole de Kronecker.

<ul style="list-style-type: none"> <li>◊ <math>\mathbf{N} \sim \mathcal{M}(n, \mathbf{p})</math>, <math>n \in \mathbb{N}</math>, <math>\mathbf{p} \in [0, 1]^m</math>, <math>m \in \mathbb{N}^*</math>,  <math>\sum p_i = 1</math>;</li> <li>◊ <math>E = \{\mathbf{N} \in \{0, \dots, n\}^m, \sum N_i = n\}</math>;</li> <li>◊ <math>f_{\mathbf{N}}(\mathbf{n}) = \frac{n!}{n_1! \dots n_m!} p_1^{n_1} \dots p_m^{n_m}</math>;</li> <li>◊ <math>\mathbb{E}[\mathbf{N}] = n\mathbf{p}</math>, <math>\text{Var}(\mathbf{N})_{ij} = np_i\delta_{ij} - np_i p_j</math>;</li> <li>◊ <math>\Phi_{\mathbf{N}}(\mathbf{t}) = (\sum_j p_j e^{it_j})^n</math>.</li> </ul>	<ul style="list-style-type: none"> <li>▀ <code>npr.multinomial(n, p)</code></li> <li>— lancer de dés pipés</li> <li>— sélection naturelle</li> <li>— tirage de boules avec remise</li> <li>— sondage</li> <li>💡 test du khi-deux</li> </ul>
---	---

Le paramètre  $m$  représente le nombre d'issues possibles lors d'une expérience aléatoire. Le paramètre  $\mathbf{p} = (p_1, \dots, p_m)$  donne la probabilité  $p_i$  d'obtenir la  $i^e$  issue. Enfin, le paramètre  $n$  représente le nombre de répétitions de l'expérience. La loi multinomiale modélise le nombre de fois que l'on obtient chacune des  $m$  issues.

En particulier, lorsque  $m = 2$ , on retrouve la loi binomiale.

**Proposition.**  $X \sim \mathcal{B}(n, p)$ ssi  $(X, n - X) \sim \mathcal{M}(n, (p, 1 - p))$ .

### ► Loi géométrique

<ul style="list-style-type: none"> <li>◊ <math>X \sim \mathcal{G}(p)</math>, <math>p \in ]0, 1]</math>;</li> <li>◊ <math>E = \mathbb{N}^*</math>;</li> <li>◊ <math>f_X(k) = (1 - p)^{k-1} p</math>;</li> <li>◊ <math>\mathbb{E}[X] = 1/p</math>, <math>\text{Var}(X) = (1 - p)/p^2</math>;</li> <li>◊ <math>\Phi_X(t) = \frac{pe^{it}}{1 - (1-p)e^{it}}</math>.</li> </ul>	<ul style="list-style-type: none"> <li>▀ <code>npr.geometric(p)</code></li> <li>— lancer de pièces</li> <li>— collectionneur de coupons</li> <li>— durée de vie (voir la loi exponentielle)</li> <li>💡 loi exponentielle</li> </ul>
--	---

Le paramètre  $p$  représente la probabilité de succès d'une expérience aléatoire. La loi géométrique modélise le nombre de répétitions de l'expérience jusqu'à l'obtention du **premier succès**.

**Proposition — Sans mémoire.** La v.a.  $X$  suit une loi géométriquessi pour tout  $n, k \in \mathbb{N}$ ,  $\mathbb{P}(X = n + k \mid X > n) = \mathbb{P}(X = k)$ .

### ► Loi hypergéométrique

<ul style="list-style-type: none"> <li>◊ <math>X \sim \mathcal{HG}(R, B, n)^*</math>, <math>R, B, n \in \mathbb{N}</math>, <math>n \leq R + B</math>;</li> <li>◊ <math>E = \{0 \vee (n - B), \dots, n \wedge R\}</math>;</li> <li>◊ <math>f_X(k) = \frac{\binom{R}{k} \binom{B}{n-k}}{\binom{R+B}{n}}</math>;</li> <li>◊ <math>\mathbb{E}[X] = n \frac{R}{R+B}</math>, <math>\text{Var}(X) = n \frac{RB}{(R+B)^2} \frac{R+B-n}{R+B-1}</math>;</li> <li>◊ <math>\Phi_X</math> est appelée série hypergéométrique.</li> </ul>	<ul style="list-style-type: none"> <li>▀ <code>npr.hypergeometric(R, B, n)</code></li> <li>— sélection naturelle</li> <li>— tirage de boules sans remise</li> <li>— sondage</li> <li>💡 loi binomiale</li> </ul>
---	---

Le paramètre  $n$  représente le nombre de **tirages sans remise** dans une urne contenant  $R$  boules rouges et  $B$  boules bleues. La loi hypergéométrique modélise le nombre de boules rouges obtenues lors de ces  $n$  tirages.

En pratique, un sondage s'effectue souvent sous la forme d'un tirage sans remise. Mais, il est beaucoup plus pratique (mathématiquement et informatiquement) d'approcher la loi hypergéométrique par la loi binomiale.

**Proposition.** *Lorsque  $N, R \rightarrow +\infty$  avec  $R/N \rightarrow p$ , la loi  $\mathcal{HG}(N, R, n)$  converge vers la loi  $\mathcal{B}(n, p)$ .*

### ► Loi de Poisson

<ul style="list-style-type: none"> <li>◊ <math>X \sim \mathcal{P}(\lambda)</math>, <math>\lambda \in \mathbb{R}_+^*</math> ;</li> <li>◊ <math>E = \mathbb{N}</math> ;</li> <li>◊ <math>f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}</math> ;</li> <li>◊ <math>\mathbb{E}[X] = \lambda</math>, <math>\text{Var}(X) = \lambda</math> ;</li> <li>◊ <math>\Phi_X(t) = \exp(\lambda(e^{it} - 1))</math>.</li> </ul>	<ul style="list-style-type: none"> <li>▀ npr.poisson(<math>\lambda</math>)</li> <li>— événements rares</li> <li>— nombre de clients servis</li> <li>— nombre de pannes</li> <li>— nombre de voitures à un péage</li> <li>💡 loi exponentielle</li> </ul>
--	---

Le paramètre  $\lambda$  représente le nombre moyen d'événements qui se produisent pendant un intervalle de temps  $[0, T]$ . La loi de Poisson modélise le nombre d'événements qui se produisent pendant cet intervalle de temps. On peut également voir le paramètre  $\lambda$  comme l'inverse du temps moyen entre deux événements (voir le chapitre **Processus de Poisson**).

Cette loi a été introduite par Poisson en 1837 comme la limite de la loi binomiale dans un certain régime asymptotique.

**Proposition — Loi des événements rares.** *Lorsque  $n \rightarrow +\infty$ ,  $p \rightarrow 0$  et  $np \rightarrow \lambda \in \mathbb{R}_+^*$ , la loi  $\mathcal{B}(n, p)$  converge vers la loi  $\mathcal{P}(\lambda)$ .*

### 4.1.2 Lois continues

Pour chacune des lois, nous listons : sa notation et ses paramètres usuels, son support, sa densité, son espérance, sa variance, l'expression de sa fonction caractéristique et la commande Python utilisée pour la simulation à partir du module `numpy.random` importé en tant que `npr`.

### ► Loi uniforme

<ul style="list-style-type: none"> <li>◊ <math>X \sim \mathcal{U}(E)</math> ;</li> <li>◊ <math>E</math> est un ensemble de mesure de Lebesgue <math>\lambda(E) \in \mathbb{R}_+^*</math> (e.g. <math>E = [a, b]</math>) ;</li> <li>◊ <math>f_X(x) = \frac{1}{\lambda(E)} \mathbf{1}_E(x)</math> (e.g. <math>\lambda(E) = b - a</math>) ;</li> <li>◊ <math>\mathbb{E}[X] = (a + b)/2</math>, <math>\text{Var}(X) = \frac{(b-a)^2}{12}</math> ;</li> <li>◊ <math>\Phi_X(t) = \frac{e^{ibt} - e^{iat}}{it(b-a)}</math>.</li> </ul>	<ul style="list-style-type: none"> <li>▀ npr.uniform(<math>a, b</math>)</li> <li>— aiguille de Buffon</li> <li>💡 estimation de <math>\pi</math></li> <li>💡 méthode d'inversion</li> </ul>
---	---

Le paramètre  $E$  représente le support. La loi uniforme modélise une expérience dont l'issue est uniforme sur  $E$  : par exemple, la probabilité que  $X$  appartienne à un intervalle de  $E$  ne dépend que de la longueur de cet intervalle.

\*. **Attention :** Les paramètres de la loi hypergéométrique ne sont pas les mêmes d'un livre à l'autre.

**Proposition.** Soient  $X \sim \mathcal{U}(E)$  et  $f : E \rightarrow F$ . Si  $f$  est bijectif, différentiable et que la valeur absolue de son jacobien est constante, alors  $f(X) \sim \mathcal{U}(F)$ .

## ► Loi exponentielle

- ◊  $X \sim \mathcal{E}(\lambda), \lambda \in \mathbb{R}_+^*$ ;
- ◊  $E = \mathbb{R}_+$ ;
- ◊  $f_X(x) = \lambda e^{-\lambda x} \mathbf{1}_{\mathbb{R}_+}(x)$ ;
- ◊  $\mathbb{E}[X] = 1/\lambda, \text{Var}(X) = 1/\lambda^2$ ;
- ◊  $\Phi_X(t) = (1 - \frac{it}{\lambda})^{-1}$ .

- ▀ npr.exponential(1/ $\lambda$ )
- durée de vie
- temps d'attente
- 💡 loi géométrique
- 💡 processus de Poisson

Le paramètre  $\lambda$  représente le taux instantané d'apparition d'un certain évènement. La loi exponentielle modélise le temps d'attente avant l'apparition de cet évènement. On peut également voir le paramètre  $\lambda$  comme l'inverse du temps d'attente moyen.

La loi exponentielle est l'analogie continu de la loi géométrique.

**Proposition — Liens exponentielle/géométrique.**

- (i) Pour tout  $n \in \mathbb{N}$ , on note  $Y_n \sim \mathcal{G}(p_n)$ . Si  $p_n \rightarrow 0$ , alors  $p_n Y_n$  converge en loi vers  $\mathcal{E}(1)$ .
- (ii) Si  $X \sim \mathcal{E}(1)$  et  $Y = \lceil \theta X \rceil$  avec  $\theta > 0$ , alors  $Y \sim \mathcal{G}(1 - e^{-\frac{1}{\theta}})$ .

**Proposition — Sans mémoire.** La v.a.  $X$  suit une loi exponentielle ssi pour tout  $t, s \in \mathbb{R}_+$ ,  $\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t)$ .

**Proposition — Stabilité par minimum.** Soient  $X \sim \mathcal{E}(\lambda)$  et  $Y \sim \mathcal{E}(\mu)$  deux v.a. indépendantes. La v.a.  $Z = \min(X, Y)$  suit la loi  $\mathcal{E}(\lambda + \mu)$ .

## ► Loi normale

- ◊  $X \sim \mathcal{N}(\mu, \sigma^2), \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+^*$ ;
- ◊  $E = \mathbb{R}$ ;
- ◊  $f_X(x) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ ;
- ◊  $\mathbb{E}[X] = \mu, \text{Var}(X) = \sigma^2$ ;
- ◊  $\Phi_X(t) = \exp(\mu it - \frac{\sigma^2 t^2}{2})$ .

- ▀ npr.normal( $\mu, \sigma$ )
- erreur de mesure
- balistique
- taille ou poids d'un individu
- 💡 Théorème Central Limite

Les paramètres  $\mu$  et  $\sigma^2$  représentent respectivement la moyenne et la variance de la loi. La loi normale modélise généralement une grandeur qui dépend d'un grand nombre de paramètres inconnus et aléatoires. Son utilisation dans ce cadre est justifiée par le **Théorème central limite**.

Cette loi apparaît dans des travaux de Moivre (1733), Laplace (1781 et 1812) et Gauss (1809) qui relie cette loi avec des tables d'erreurs d'observation en astronomie.

**Proposition — Stabilité.** Soient  $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$  et  $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$  deux v.a. indépendantes et  $a, b \in \mathbb{R}^*$ . Alors  $aX_1 + bX_2 \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$ .

## ► Loi du Khi-deux

- ◊  $X \sim \chi^2(v)$ ,  $v \in \mathbb{N}^*$ ;
- ◊  $E = \mathbb{R}_+$ ;
- ◊  $f_X(x) = \frac{1}{2^{v/2}\Gamma(v/2)}x^{\frac{v}{2}-1}e^{-x/2}\mathbf{1}_{\mathbb{R}_+}(x)$ ;
- ◊  $\mathbb{E}[X] = v$ ,  $\text{Var}(X) = 2v$ ;
- ◊  $\Phi_X(t) = (1 - 2it)^{-v/2}$ .

▀ npr.chisquare(v)

💡 carré de loi normale

💡 variance empirique

💡 test du khi-deux

Le paramètre  $v$  est généralement appelé “**degré de liberté**” de la loi du  $\chi^2$ . La loi du khi-deux est obtenue comme la loi de la somme des carrés de  $v$  v.a. indépendantes de loi  $\mathcal{N}(0, 1)$ .

Cette loi apparaît dans l’étude de l’estimateur de la variance dans un modèle gaussien grâce à la propriété suivante.

**Proposition.** Soient  $X_1, \dots, X_v$  des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ . Alors la v.a.  $Y = \sum_{i=1}^v X_i^2$  suit la loi  $\chi^2(v)$ .

## ► Loi Gamma

- ◊  $X \sim \Gamma(k, \theta)$ ,  $k, \theta \in \mathbb{R}_+^*$ ;
- ◊  $E = \mathbb{R}_+$ ;
- ◊  $f_X(k) = \frac{1}{\Gamma(k)\theta^k}x^{k-1}e^{-x/\theta}\mathbf{1}_{\mathbb{R}_+}(x)$ ;
- ◊  $\mathbb{E}[X] = k\theta$ ,  $\text{Var}(X) = k\theta^2$ ;
- ◊  $\Phi_X(t) = (1 - \theta it)^{-k}$ .

▀ npr.gamma(k, theta)

— durée de vie

— temps d’attente

💡 loi exponentielle

Les paramètres  $k$  et  $\theta$  influencent respectivement la forme et l’échelle du graphe de la densité de la loi Gamma. La loi Gamma modélise des temps aléatoires en généralisant la loi exponentielle : on ne suppose plus la perte de mémoire.

Les lois exponentielle et du khi-deux font parties de la famille des lois Gamma :  $\mathcal{E}(\lambda) = \Gamma(1, 1/\lambda)$  et  $\chi^2(v) = \Gamma(v/2, 2)$ .

**Proposition — Stabilité.** Soient  $X_1 \sim \Gamma(k_1, \theta)$  et  $X_2 \sim \Gamma(k_2, \theta)$  deux v.a. indépendantes et  $a \in \mathbb{R}^*$ . Alors  $X_1 + X_2 \sim \Gamma(k_1 + k_2, \theta)$  et  $aX_1 \sim \Gamma(k_1, a\theta)$ .

## ► Loi de Student

- ◊  $X \sim \mathcal{T}(k)$ ,  $k \in \mathbb{N}^*$ ;
- ◊  $E = \mathbb{R}$ ;
- ◊  $f_X(x) = \frac{1}{\sqrt{k\pi}} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}$ ;
- ◊  $\mathbb{E}[X] = 0$  ( $k > 1$ ),  $\text{Var}(X) = \frac{k}{k-2}$  ( $k > 2$ ).

▀ npr.standard\_t(k)

💡 loi normale modifiée

💡 variance empirique

💡 test de Student

Le paramètre  $k$  est généralement appelé “**degré de liberté**” de la loi de Student. La loi de Student est spécifiquement définie pour satisfaire la proposition suivante.

**Proposition.** Soient  $X_1, \dots, X_k$  des v.a. i.i.d. de loi  $\mathcal{N}(0, 1)$ . On note  $\bar{X}_k$  leur moyenne empirique

et  $\hat{S}_n^2$  leur variance empirique corrigée (voir page 43). La v.a.

$$T := \sqrt{k} \frac{\bar{X}_k}{\hat{S}_k^2},$$

suit la loi de Student  $\mathcal{T}_{k-1}$ .

## 4.2 Méthodes de simulation

Pour simuler un échantillon de **loi usuelle**, on utilisera les fonctions déjà codées dans Python. Pour simuler un échantillon de **loi originale**, il faudra s'appuyer sur l'une des trois méthodes ci-dessous.

**Méthode — Inversion.** Pour simuler une v.a. réelle de fonction de répartition  $F$ , il suffit de simuler une v.a. loi uniforme.

En effet, si  $U \sim \mathcal{U}([0, 1])$ , alors  $F^{-1}(U)$  est une v.a. de fonction de répartition  $F$ . On rappelle que  $F^{-1}$  est l'inverse généralisé défini à la page 14.

On utilise la méthode d'inversion pour les lois continues dont la fonction de répartition est inversible ou dont l'inverse généralisé est simple à calculer (e.g. loi exponentielle, de Cauchy, etc). Pour les lois discrètes, c'est la méthode la plus naturelle (cf. construction d'une chaîne de Markov à partir d'une suite d'innovations de loi uniforme).

**Méthode — Rejet.** Pour simuler une v.a. de densité (ou fonction de masse)  $f$ , il suffit de savoir simuler (indépendamment) une v.a. uniforme et une v.a. de densité (ou fonction de masse)  $g$  telle que :

- (i)  $\text{supp } f \subset \text{supp } g$ ,
- (ii) il existe une constante  $C \in \mathbb{R}_+^*$  telle que  $f \leq Cg$ .

En effet, si  $(U_i, X_i)_{i \in \mathbb{N}}$  est une suite de v.a. i.i.d. de loi  $\mathcal{U}([0, 1]) \otimes g$ , alors  $Y = X_{i_0}$  est une v.a. de densité  $f$  où  $i_0$  est le plus petit entier tel que

$$U_i \leq \frac{f(X_i)}{Cg(X_i)}. \quad (4.1)$$

La méthode du rejet s'implémente sous la forme d'une boucle `while`. Tant que la condition (4.1) n'est pas vérifiée, on simule un nouveau couple  $(U, X)$ . On s'arrête la première fois <sup>†</sup> où la condition (4.1) est vérifiée et on conserve la valeur de  $X$ .

On utilise la méthode du rejet lorsque la méthode d'inversion n'est pas réalisable. Un exemple important est la restriction d'une loi simple à un sous-ensemble d'intérêt (loi uniforme sur le disque unité à partir de la loi uniforme sur  $[-1, 1]^2$ , loi uniforme sur les permutations de degré  $n$  sans point fixe à partir de la loi uniforme sur les permutations de degré  $n$ ).

**Méthode — Loi d'un couple.** Pour simuler un couple de v.a.  $(X, Y) \in E \times F$  de loi jointe  $\mathbb{P}_{(X,Y)}$ , il suffit de savoir simuler (indépendamment) une v.a. de loi  $\mathbb{P}_X$  et, pour tout  $x \in E$ , une v.a. de loi  $\mathbb{P}_{Y|x}$  (où  $\mathbb{P}_{Y|x}$  est la loi conditionnelle de  $Y$  sachant  $X = x$ ).

En effet, si  $X \sim \mathbb{P}_X$  et, qu'indépendamment,  $Y \sim \mathbb{P}_{Y|x}$ , alors  $(X, Y) \sim \mathbb{P}_{(X,Y)}$ .

Cette méthode est très utile dans le cas d'un **mélange de lois** (problème de classification). Par exemple, pour simuler  $X$  de densité donnée par  $f_X(x) = p\mathbf{1}_{[1,2]}(x) + (1-p)\mathbf{1}_{[0,1]}(x)$  où  $p \in ]0, 1[$  est connu, on peut se ramener à simuler le couple  $(\varepsilon, X)$  où  $\varepsilon \sim \mathcal{B}(p)$  et  $\mathbb{P}_{X;i} = \mathcal{U}([i, i+1])$  pour  $i = 1, 2$ .

<sup>†</sup>. le nombre d'itérations de la boucle `while`, noté  $i_0$  ci-dessus, est aléatoire (il dépend des réalisations de la simulation).

# 5 CHAÎNES DE MARKOV

Les chaînes de Markov sont des processus stochastiques. Les notions de processus stochastiques et de temps d'arrêt sont définies dans la section 7.3.1.

Dans la suite,  $E$  désigne un espace d'état **fini ou dénombrable**.

## 5.1 Définitions

**Définition.** La suite  $(X_n)_{n \geq 0}$  est une **chaîne de Markov** (homogène)ssi pour tout  $k \in \mathbb{N}$  et  $(x_0, \dots, x_{k+1}) \in E^{k+2}$  tels que  $\mathbb{P}(X_k = x_k, \dots, X_0 = x_0) > 0$ ,

$$\mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k, \dots, X_0 = x_0) = \mathbb{P}(X_{k+1} = x_{k+1} | X_k = x_k), \quad (5.1)$$

et pour tout  $(x, y) \in E^2$  tels que  $\mathbb{P}(X_0 = x) > 0$  et  $\mathbb{P}(X_k = x) > 0$ ,

$$\mathbb{P}(X_{k+1} = y | X_k = x) = \mathbb{P}(X_1 = y | X_0 = x). \quad (5.2)$$

**Définition.** On appelle **probabilité de transition** pour aller de l'état  $i$  à l'état  $j$  la probabilité

$$p_{ij} := \mathbb{P}(X_1 = j | X_0 = i). \quad (5.3)$$

On appelle **matrice de transition** la matrice  ${}^* P := (p_{ij})_{i,j \in E}$ .

\*(éventuellement ayant une infinité dénombrable de lignes et colonnes)

**Notation** (matricielle).

- On identifie <sup>†</sup>toute probabilité  $v$  sur  $E$  à un **vecteur ligne** et on note  $X_0 \sim v^0$ .  
En particulier,  $v^1 = v^0 P$  est la loi de  $X_1$ .
- On identifie toute fonction  $f : E \rightarrow \mathbb{R}$  à un **vecteur colonne**.  
En particulier, on a  $v^0 f = \mathbb{E}[f(X_0)]$  et  $v^0 P f = \mathbb{E}[f(X_1)]$ . De plus, le vecteur  $(1, 1, \dots)$  est un **vecteur propre** (à droite) associé à la valeur propre 1.

**Proposition.** Toute matrice de transition est une **matrice stochastique** :

- tous ses coefficients appartiennent à  $[0, 1]$ ,
- sur chaque ligne la somme des coefficients vaut 1.

**Lemme — Probabilité d'une trajectoire.** Pour tout  $(x_0, \dots, x_n) \in E^n$ ,

$$\mathbb{P}(X_n = x_n, \dots, X_0 = x_0) = v^0(x_0) \prod_{k=0}^{n-1} p_{x_k, x_{k+1}},$$

avec  $v^0$  la loi de  $X_0$ .

Le résultat suivant propose une construction algorithmique des chaînes de Markov.

<sup>†</sup>. (en choisissant une énumération de  $E$ )

### Théorème fondamental.

**Existence** Soit  $v^0$  une loi sur  $E$ ,  $X_0 \sim v^0$  et  $P$  une matrice stochastique. Il existe une fonction  $\Phi : E \times \mathcal{U}$  et une suite de v.a. i.i.d.  $(U_n)_{n \geq 0}$  à valeurs dans  $\mathcal{U}$  indépendante de  $X_0$ , telle que

$$X_{n+1} = \Phi(X_n, U_n), \quad (5.4)$$

définisse une chaîne de Markov de transition  $P$ .

Réciproquement, une suite construite selon (5.4) et les conditions au-dessus est une chaîne de Markov homogène.

**Unicité** La loi d'une chaîne de Markov est caractérisée par sa distribution initiale  $v^0$  et sa matrice de transition  $P$ .

#### 5.1.1 Principaux exemples simples

- ▶ **Deux états.** Les deux états peuvent représenter la santé d'un patient au jour  $n$  : Sain (0) ou Malade (1). La matrice de transition de cette chaîne a uniquement deux paramètres :

$$P = \begin{pmatrix} 1 - p_{0,1} & p_{0,1} \\ p_{1,0} & 1 - p_{1,0} \end{pmatrix}.$$

C'est l'exemple le plus simple possible. Il faut savoir sa dynamique en fonction des valeurs de  $p_{0,1}$  et  $p_{1,0}$  (par exemple si  $p_{0,1} = 0$  ou 1). C'est l'exemple “fil rouge” de cette feuille.

- ▶ **Ruine du joueur.** Deux joueurs (A et B) possèdent  $n_a$  et  $n_b$  euros. A chaque tour de jeu, chaque joueur mise 1 euro à pile ou face et la probabilité que A gagne est  $0 < p < 1$  (les parties sont indépendantes). Le jeu se termine lorsqu'un des deux joueurs est ruiné. La fortune du joueur A après la  $n$ -ème partie, notée  $X_n$ , définit une chaîne de Markov d'espace d'états  $E = \{0, 1, \dots, n_a + n_b\}$ . Sa matrice de transition est une matrice tri-diagonale :
  - diagonale :  $(1, 0, \dots, 0, 1)$ ,
  - sur-diagonale :  $(0, p, \dots, p)$ ,
  - sous-diagonale :  $(q, \dots, q, 0)$ .

- ▶ **Contre-exemples.**

- Processus qui vaut 1 à partir de la deuxième fois où on tombe malade et 0 sinon.
- Marche aléatoire auto-évitante.

#### 5.1.2 Itérations de la chaîne

**Notation.** Pour tout  $n \geq 0$  et  $i, j \in E$ , on note la probabilité de transition en  $n$  étapes,

$$p_{ij}^{(n)} := \mathbb{P}(X_n = j | X_0 = i).$$

**Proposition.** La matrice de transition en  $n$  étapes  $(p_{ij}^{(n)})_{i,j \in E}$  est égale à la matrice  $P^n$ . En particulier,  $P^n$  est une matrice stochastique et l'associativité  $(P^{m+n} = P^m P^n)$  se réécrit :

$$p_{ij}^{(n+m)} = \sum_{k \in E} p_{ik}^{(n)} p_{kj}^{(m)}.$$

**Théorème — Propriété de Markov (faible).** Soient  $n$  un entier et  $A$  un événement de  $\mathcal{F}_n^X$  tel que  $\mathbb{P}(A \text{ et } X_n = i) > 0$ . Alors

$$\mathcal{L}((X_{n+k})_{k \geq 0} | A \text{ et } X_n = i) = \mathcal{L}((X_k)_{k \geq 0} | X_0 = i).$$

**Théorème — Propriété de Markov (forte).** Soient  $T$  un temps d'arrêt et  $A$  un évènement de  $\mathcal{F}_T^X$  tels que  $\mathbb{P}(T < +\infty, A, X_T = i) > 0$ . Alors

$$\mathcal{L}((X_{T+k})_{k \geq 0} | T < \infty, A \text{ et } X_T = i) = \mathcal{L}((X_k)_{k \geq 0} | X_0 = i).$$

On se sert souvent de ce dernier résultat avec  $T = \inf\{n \geq 1, X_n = i\}$ . Cela prouve que, conditionnellement à  $\{T < +\infty\}$ , le processus  $(X_{T+n})_{n \geq 0}$  est une **chaîne de Markov d'état initial  $i$** .

## 5.2 Classification

L'étude de la matrice de transition (et de ses puissances) permet de classifier les états dans des classes d'équivalence.

### 5.2.1 Communication entre états

**Définition.** On dit que l'état  $j$  est **accessible** depuis l'état  $i$ , noté  $i \rightsquigarrow j$ ,ssi il existe  $n \geq 0$  tel que  $p_{i,j}^{(n)} > 0$  (en particulier  $i \rightsquigarrow i$ ).

On dit que  $i$  et  $j$  **communiquent**, noté  $i \rightsquigarrow\rightsquigarrow j$ ,ssi  $i \rightsquigarrow j$  et  $j \rightsquigarrow i$ .

**Proposition.** La relation de communication est une relation d'équivalence. Ses classes d'équivalence sont appelées **classes de communication**.

**Définition.** Si la chaîne ne possède qu'une seule classe de communication, on dit qu'elle est **irréductible**.

■ **Exemple** (fil rouge). La chaîne est irréductiblessi  $p_{0,1}$  et  $p_{1,0}$  sont non nuls.

### 5.2.2 Récurrence/Transience

Pour tout  $i \in E$ , on note  $T_i := \inf\{n \geq 1, X_n = i\}$  le temps d'atteinte de l'état  $i$ , et  $N_i = \sum_{n \geq 0} \mathbf{1}_{X_n=i}$  le nombre de passages en  $i$ .

**Définition.** On dit que  $i$  est **récurrent** si

$$\mathbb{P}(T_i < +\infty | X_0 = i) = 1.$$

Sinon, on dit que  $i$  est **transitoire**.

■ **Exemple** (fil rouge). Si  $p_{0,1} > 0$  et  $p_{1,0} = 0$  alors l'état 0 est transitoire. Sinon, il est récurrent.

**Proposition.**

- (i) Si  $i$  est **récurrent**, alors  $N_i = +\infty$  p.s. sous la condition initiale  $v^0 = \delta_i$ .
- (ii) Si  $i$  est **transitoire**, alors  $N_i$  suit une loi géométrique sous la condition initiale  $v^0 = \delta_i$ .  
Quelque soit  $v^0$ , le nombre moyen de passages en  $i$  est fini, i.e.  $\mathbb{E}[N_i] < +\infty$ .
- (iii) Si  $i$  est récurrent et  $i \rightsquigarrow j$ , alors  $j$  est récurrent : on peut aller d'un état transitoire vers un état récurrent, mais l'inverse est impossible.

**Corollaire.** L'état  $i$  est transitoiressi  $\sum_{n \geq 0} p_{ii}^{(n)} < +\infty$ .

**Théorème.** La récurrence/transience est une propriété de classe de communication.

L'espace d'états est naturellement partitionné en classes de communication. Voici une condition nécessaire et une condition suffisante de récurrence faisant intervenir la notion de fermeture :

une classe  $C$  est **fermée**ssi pour tout  $i \in C$  et  $j \notin C$ ,  $p_{ij} = 0$ . Autrement dit, on ne peut pas sortir d'une classe fermée.

**Proposition.** *Une classe récurrente est fermée. Toute classe fermée et finie est récurrente.*



Si la chaîne part d'un état récurrent, elle reste indéfiniment dans sa classe.

## 5.3 Mesure stationnaire

Il se peut que certaines conditions initiales ne soient pas modifiées par l'évolution de la CM.

**Définition.** *On dit que  $\mu$  est une **mesure stationnaire** si  $\mu P = \mu$ . Si  $\mu$  est une probabilité, on parlera plutôt de **probabilité stationnaire**.*

*On dit que  $\mu$  est une **mesure réversible** si pour tous  $i$  et  $j$ ,*

$$\mu(i)p_{i,j} = \mu(j)p_{j,i}$$

**Proposition.** *Si  $\mu$  est réversible, alors  $\mu$  est stationnaire.*

■ **Exemple (fil rouge).** On note  $q = p_{0,1} + p_{1,0}$ . Si  $q > 0$  alors  $\mu = \begin{pmatrix} p_{0,1}/q \\ p_{1,0}/q \end{pmatrix}$  est l'unique probabilité stationnaire (et est réversible). Si  $q = 0$  alors toutes les mesures sont stationnaires.

**Proposition.** *Si  $\mu$  est une probabilité stationnaire et que  $i$  est transitoire, alors  $\mu(i) = 0$ .*

**Théorème.** *Si  $i$  est un état récurrent, alors la mesure  $v_i$  définie par  $v_i(j) := \sum_{k=0}^{T_i-1} \mathbf{1}_{X_k=j}$  est une mesure invariante.*

### 5.3.1 Récurrence positive ou nulle

Pour un état récurrent  $i$ , on définit le **temps de retour moyen** par

$$m_{ii} := \mathbb{E}[T_i | X_0 = i].$$

C'est la masse totale de la mesure  $v_i$ .

**Définition.** *On dira que  $i$  est récurrent positif si  $m_{ii} < +\infty$ . Sinon, on dira que  $i$  est récurrent nul.*

**Théorème.** *Un état  $i$  est nulssi  $p_{ii}^{(n)} \rightarrow 0$ . Par conséquent, la positivité/nullité est une propriété de classe.*



Lorsque  $E$  est fini, tous les états récurrents sont récurrents positifs. Les états récurrents nuls posent problème lorsque  $E$  est dénombrable. Ils ont un comportement trop "proche" de celui des états transitoires.

**Théorème.** *Soit  $P$  une chaîne irréductible. Alors,*

- (i) *La chaîne est récurrente positive si et seulement si elle possède une mesure de probabilité invariante. Dans ce cas, cette mesure, notée  $\mu$ , est unique et  $\mu(i) = 1/\mathbb{E}[T_i | X_0 = i]$ .*
- (ii) *Si la chaîne est récurrente, toutes les mesures invariantes sont proportionnelles.*

En particulier, lorsqu'une chaîne irréductible est **récurrente nulle ou transiente** alors **il n'existe pas de probabilité stationnaire**.



Toute chaîne de Markov finie possède **au moins une probabilité stationnaire**  $\mu$ .

On illustre le Théorème précédent sur l'exemple de la marche aléatoire sur  $\mathbb{Z}$ .

### 5.3.2 Exemple principal

Pour la marche aléatoire, on considère l'espace d'états  $E = \mathbb{Z}$  et un paramètre  $p \in ]0, 1[$ . Sa matrice de transition  $P$  est définie par

$$p_{ij} = \begin{cases} p & \text{si } j = i + 1, \\ 1 - p & \text{si } j = i - 1, \\ 0 & \text{sinon.} \end{cases}$$

Si  $p = 1/2$ , on dit que la marche est symétrique. Cette chaîne de Markov est irréductible. En voici quelques propriétés supplémentaires :

- L'unique classe est fermée et de période égale à 2.
- L'unique classe est transitoire si  $p \neq 1/2$  : **montre qu'il n'y a pas équivalence dans la Proposition 5.2.2.**
- L'unique classe est récurrente si  $p = 1/2$ .
- La mesure de comptage est une mesure stationnaire.
- La chaîne n'admet pas de probabilité stationnaire : **exemple pour le Théorème 5.3.1.**
- La mesure  $\mu(i) = \left(\frac{1-p}{p}\right)^i$  est stationnaire, réversible. Lorsque  $p \neq 1/2$ , cela donne un exemple (transitoire) où il n'y a pas unicité à coefficient multiplicatif près des mesures stationnaires.

## 5.4 Théorèmes de convergence

### 5.4.1 Loi des grands nombres

**Théorème ergodique.** Si la chaîne est irréductible de probabilité stationnaire  $\mu$ , alors pour toute fonction  $f : E \rightarrow \mathbb{R}$ ,

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{p.s.} \int f d\mu = \sum_{i \in E} f(i) \mu(i).$$

### 5.4.2 Convergence en loi

Bien que les trajectoires soient aléatoires, il se peut qu'elles satisfassent un phénomène plus ou moins périodique, ce qui pose problème pour la convergence en loi de la chaîne. La notion suivante précise cette idée.

**Définition.** Soit  $i \in E$ . Le PGCD des possibles temps de retour en  $i$ , c'est-à-dire

$$d(i) := \text{PGCD} \left( n \geq 1, p_{ii}^{(n)} > 0 \right),$$

est appelé la **période** de  $i$ . Si  $d(i) = 1$ , on dit que  $i$  est **apériodique**. Par extension, la chaîne est apériodique si tous ses états le sont.

**Théorème.** La période est un **invariant de classe de communication**.

■ **Exemple (fil rouge).** Si  $p_{0,1} = p_{1,0} = 1$  alors la période est égale à 2. Sinon, la chaîne est apériodique.

**Théorème.** Si la chaîne est irréductible, apériodique et de probabilité stationnaire  $\mu$ , alors pour toute condition initiale  $v^0$ ,

$$X_n \xrightarrow{\mathcal{L}} \mu.$$

## 5.5 Espace d'états fini

Dans cette partie on ajoute quelques précisions lorsque l'espace d'état est fini.



Toute chaîne de Markov finie admet au moins un état récurrent et admet au moins une probabilité invariante.

Si la chaîne part d'un état transitoire, elle finira par sortir de sa classe en un temps fini p.s. Ensuite, elle atteindra, en un temps fini p.s., une classe récurrente dans laquelle elle restera.

**Théorème — Perron-Frobenius.** Soit  $P$  la matrice de transition d'une chaîne de Markov irréductible, apériodique à espace d'état fini. Pour toute norme matricielle  $\|\cdot\|$ , il existe deux constantes  $\lambda > 0$  et  $C > 0$  telles que, pour toute loi initiale  $v^0$  et tout  $n \geq 0$ ,

$$\|v^0 P^n - \mu\| \leq C e^{-\lambda n}.$$

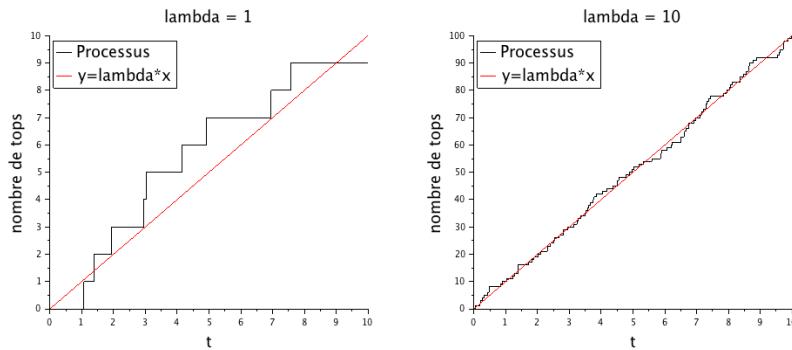
# 6 PROCESSUS DE POISSON

## 6.1 Définitions

Le processus de Poisson est le principal exemple de processus de comptage : on compte les évènements qui se sont produits entre un temps de référence 0 et un temps  $t \geq 0$  quelconque. Les instants auxquels ces évènements se produisent sont appelés **points** ou **temps de saut** du processus. Nous choisissons d'utiliser le mot **temps de saut**.

Les exemples d'application de ce modèle sont nombreux.

- L'arrivée d'un tram à une station,
- Le passage d'une voiture à un péage,
- L'arrivée d'un paquet de données sur un serveur.



**Définition.** Un processus de Poisson d'intensité  $\lambda$  est une fonction  $t \mapsto N_t$  de  $\mathbb{R}_+$  dans  $L^1(\Omega, \mathcal{F}, \mathbb{P})$  telle que

- (i) **condition initiale** :  $N_0 = 0$ ;
- (ii) **accroissements indépendants** : pour tout entier  $k$  et  $0 \leq t_1 \leq \dots \leq t_k$ ,

les v.a.  $N_{t_1}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$  sont indépendantes;

- (iii) **accroissements stationnaires** : pour tout  $t \geq s \geq 0$ ,  $N_t - N_s \sim \mathcal{P}(\lambda(t-s))$ .

*Remarque :* La condition (iii) peut être vue comme une propriété d'homogénéité en temps du processus.

La proposition suivante montre qu'il ne peut pas y avoir deux sauts arbitrairement proches l'un de l'autre.

**Proposition — Continuité locale.** Pour tout  $t \geq 0$ ,

- (i)  $\mathbb{P}(N_{t+h} - N_t = 0) \rightarrow_{h \rightarrow 0} 1$ ,
- (ii)  $\mathbb{P}(N_{t+h} - N_t = 1) \sim_{h \rightarrow 0} \lambda h$ ,
- (iii)  $\mathbb{P}(N_{t+h} - N_t \geq 2) \sim_{h \rightarrow 0} (\lambda^2/2)h^2$ .

**Notation.** Pour un processus de Poisson  $(N_t)_{t \geq 0}$ , on définit :

- les **temps de saut**  $(T_n)_{n \geq 0}$  par  $T_n = \inf\{t \geq 0, N_t \geq n\}$ ;
- les **délais inter-sauts**  $(S_n)_{n \geq 1}$  par  $S_n = T_n - T_{n-1}$ .

### 6.1.1 Deux construction possibles

On présente ici deux constructions/simulations possibles pour un processus de Poisson. Par la même occasion, cela démontre l'existence d'un processus de Poisson tel que définit plus haut.

**Proposition — A partir des délais inter-sauts.** *Le processus  $(N_t)_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda$ ssi les délais inter-sauts  $(S_n)_{n \geq 1}$  sont i.i.d. de loi  $\mathcal{E}(\lambda)$ .*

**Proposition — A partir du nombre de sauts.** *Le processus  $(N_t)_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda$ ssi pour tout  $t \geq 0$ ,  $N_t \sim \mathcal{P}(\lambda t)$  et que conditionnellement à  $N_t = n$ , le vecteur des temps de saut  $(T_1, \dots, T_n)$  a même loi que la **statistique d'ordre**\* de  $n$  v.a. i.i.d. de loi  $\mathcal{U}([0, t])$  (en particulier on retrouve l'homogénéité temporelle).*

**Corollaire** (deux méthodes de simulation).

1. *On se donne un **nombre de sauts maximal**  $n$ . On génère  $n$  v.a.  $S_1, \dots, S_n$  de loi  $\mathcal{E}(\lambda)$ . On calcule les  $n$  temps de saut  $T_1, \dots, T_n$  par  $T_k = \sum_{i=1}^k S_i$ .*
2. *On se donne un **temps maximal**  $t$ . On génère  $N_t$  de loi  $\mathcal{P}(\lambda t)$ . On génère  $N_t$  v.a.  $\tilde{T}_1, \dots, \tilde{T}_{N_t}$  de loi  $\mathcal{U}([0, t])$ . On trie le vecteur des valeurs  $\tilde{T}$  pour obtenir le vecteur des temps de saut  $(T_1, \dots, T_{N_t})$ .*

*Le processus de comptage s'obtient alors en traçant la **fonction en escalier** avec les  $T_k$  en abscisse et les entiers  $k$  en ordonnée.*

### 6.1.2 Loi des temps de saut



Le paramètre  $\lambda$  d'un processus de Poisson peut être vu de deux manières.

- Comme le **nombre moyen de sauts** dans l'intervalle unité ( $\lambda = \mathbb{E}[N_1]$ ).
- Comme l'**inverse du délai moyen inter-sauts** ( $\lambda = 1/\mathbb{E}[S_1]$ ).

**Lemme — Loi multi-dimensionnelle des temps de saut.** *Soient  $S_1, \dots, S_n$  des v.a. i.i.d. de loi  $\mathcal{E}(\lambda)$  et  $T_k = \sum_{i=1}^k S_i$ . Alors  $T_n$  suit la loi Gamma  $\Gamma(\lambda, n)$  de densité  $\frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} \mathbf{1}_{x>0}$  et la densité de  $(T_1, \dots, T_n)$  est*

$$\lambda^n e^{-\lambda t_n} \mathbf{1}_{0 < t_1 < \dots < t_n}.$$

## 6.2 Propriétés

### 6.2.1 Superposition et amincissement

Les deux résultats qui suivent montrent une certaine stabilité de la classe des processus de Poisson.

**Proposition — Amincissement.** *Soit  $(N_t)_{t \geq 0}$  un processus de Poisson d'intensité  $\lambda$  et  $(\varepsilon_n)_{n \geq 1}$  une suite indépendante de v.a. i.i.d. de loi  $\mathcal{B}(p)$ . Pour chaque temps de saut  $T_n$  de  $(N_t)_{t \geq 0}$ , on dit qu'il est de type 0 si  $\varepsilon_n = 0$  et de type 1 sinon. Alors les processus de comptage des sauts de type 0, noté  $(N_t^{(0)})_{t \geq 0}$ , et de type 1, noté  $(N_t^{(1)})_{t \geq 0}$ , sont deux processus de Poisson indépendants et d'intensité  $(1-p)\lambda$  et  $p\lambda$ .*

**Proposition — Superposition.** *Soient  $(N_t)_{t \geq 0}$  et  $(M_t)_{t \geq 0}$  deux processus de Poisson indépendants d'intensité  $\lambda$  et  $\mu$ . On pose  $L_t = N_t + M_t$ . Alors  $(L_t)_{t \geq 0}$  est un processus de Poisson d'intensité  $\lambda + \mu$ .*

---

\*. la statistique d'ordre d'un vecteur  $(x_1, \dots, x_n)$  est le vecteur  $(x_{\sigma(1)}, \dots, x_{\sigma(n)})$  où  $\sigma$  est une permutation de  $\{1, \dots, n\}$  telle que pour tout  $i < j$ ,  $x_{\sigma(i)} \leq x_{\sigma(j)}$ .

## 6.2.2 Théorème de convergence

**Théorème — Loi des grands nombres et TCL.** *On a les convergences suivantes :*

$$\frac{N_t}{t} \xrightarrow[t \rightarrow +\infty]{} \lambda \quad \text{et} \quad \sqrt{t} \left( \frac{N_t}{t} - \lambda \right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \lambda).$$

**Proposition — Intervalle de confiance.** *La v.a.  $N_t/t$  est un estimateur convergent de  $\lambda$ .*

*L'intervalle*

$$\left[ \frac{N_t}{t} - q_{1-\alpha/2} \frac{\sqrt{N_t}}{t}, \frac{N_t}{t} + q_{1-\alpha/2} \frac{\sqrt{N_t}}{t} \right],$$

*où  $q_{1-\alpha/2}$  est le quantile d'ordre  $1 - \alpha/2$  de la loi normale  $\mathcal{N}(0, 1)$  est un **intervalle de confiance asymptotique** pour  $\lambda$  de niveau  $1 - \alpha$ .*

## 6.3 Compléments classiques (pour votre culture)

### 6.3.1 Pourquoi la loi de Poisson

On peut se demander si le choix de la loi de Poisson dans la condition (iii) de la définition est arbitraire ou bien si il est imposé par les propriétés d'indépendance et de stationnarité. On peut montrer qu'un processus de comptage localement continu (Proposition 3) qui vérifie (ii) et (iii') : " $N_t - N_s \sim N_{t-s}$ " est **nécessairement un processus de Poisson**. Ce résultat rappelle la propriété d'absence de mémoire qui caractérise la loi exponentielle.

### 6.3.2 Le paradoxe de l'autobus

Un usager arrive au temps  $s \geq 0$  à un arrêt de bus. On suppose que les bus passent à l'arrêt selon un processus de Poisson  $(N_t)_{t \geq 0}$ . Le temps d'attente de l'usager est  $V_s = T_{N_s+1} - s$  et le temps écoulé depuis le dernier passage est  $W_s = s - T_{N_s}$ . On sait que le temps moyen entre deux passages du bus est  $1/\lambda$ . On pourrait donc penser que le temps moyen d'attente de l'usager est strictement inférieur à  $1/\lambda$ . **Il n'en est rien**.

En effet, on peut montrer que

- (i)  $V_s$  et  $W_s$  sont indépendants,
- (ii)  $\mathbb{E}[T_{N_s+1} - T_{N_s}] = \mathbb{E}[V_s + W_s] = \frac{1}{\lambda} + \frac{1}{\lambda}(1 - e^{-\lambda s})$ ,
- (iii)  $V_s \sim \mathcal{E}(\lambda)$  et  $W_s \sim e^{-\lambda s} \delta_s(dw) + \lambda e^{-\lambda w} \mathbf{1}_{[0,s]}(u)du$ .

En particulier, lorsque  $s \rightarrow +\infty$ , le temps moyen entre les deux bus de l'intervalle de temps qui contient  $s$ , à savoir  $\mathbb{E}[T_{N_s+1} - T_{N_s}]$ , tend vers  $2/\lambda$ . Ce paradoxe peut s'expliquer par le fait que **les délais inter-sauts exceptionnellement grands ont plus de poids** : l'usager a plus de chance d'arriver pendant un de ces intervalles.



# 7 ESPÉRANCE CONDITIONNELLE ET MARTINGALES

## 7.1 Espérance conditionnelle

### 7.1.1 Cas discret

C'est le cas le plus simple (qui est au programme commun de l'agrégation) car on conditionne par un événement de **probabilité non nulle**.

**Définition.** Soit  $B$  un événement tel que  $\mathbb{P}(B) > 0$ . On appelle **probabilité conditionnelle sachant  $B$**  la mesure définie par, pour tout événement  $A$ ,

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Pour toute v.a.  $X$  positive ou dans  $L^1$ , l'espérance conditionnelle  $\mathbb{E}[X|B]$  est l'espérance de  $X$  sous  $\mathbb{P}(\cdot|B)$ , i.e.  $\mathbb{E}[X|B] = \mathbb{E}[X\mathbf{1}_B]/\mathbb{P}(B)$ .

**Définition.** Soit  $X$  positive ou dans  $L^1$  et  $Y$  une v.a. à valeurs dans  $E$  dénombrable. L'espérance conditionnelle de  $X$  sachant  $Y$  est la v.a.

$$\mathbb{E}[X|Y] = \varphi(Y),$$

où  $\varphi : E \mapsto \mathbb{R}$  est définie par  $\varphi(y) = \mathbb{E}[X|Y=y]$  si  $\mathbb{P}(Y=y) > 0$  et  $\varphi(y) = 0$  sinon.



Comme toute v.a., l'espérance conditionnelle sachant  $Y$  est définie à un ensemble de mesure nulle près. C'est pour cela que l'on peut choisir  $\varphi(y) = 0$  ou n'importe quelle autre valeur lorsque  $\mathbb{P}(Y=y) = 0$ .

### 7.1.2 Cas général

**Théorème.** Soit  $\mathcal{B}$  une sous-tribu et  $X$  une v.a. dans  $L^1$ . Il existe une unique v.a. dans  $L^1$  et  $\mathcal{B}$ -mesurable, notée  $\mathbb{E}[X|\mathcal{B}]$  et appelée **espérance conditionnelle de  $X$  sachant  $\mathcal{B}$** , telle que

$$\forall Z \text{ v.a. } \mathcal{B}\text{-mesurable bornée}, \quad \mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}]Z].$$

Dans la suite on suppose que toutes les v.a. dont on prend l'espérance conditionnelle sont intégrables.



Si  $\mathcal{B} = \sigma(Y)$ , on dira que  $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$  est l'**espérance de  $X$  sachant  $Y$** .

Lorsque  $X$  est dans  $L^2(\Omega, \mathcal{F}, \mathbb{P})$ ,  $\mathbb{E}[X|\mathcal{B}]$  est aussi la projection orthogonale de  $X$  sur  $L^2(\Omega, \mathcal{B}, \mathbb{P})$  et une formule de Pythagore existe.

**Proposition — Analogue à l'espérance classique.** L'espérance conditionnelle est linéaire. De plus, le Lemme de Fatou, l'inégalité de Jensen ainsi que les convergences monotones et dominées sont valides.

**Proposition — Spécifique à l'espérance conditionnelle.**

- Si  $Z$  est  $\mathcal{B}$ -mesurable alors  $\mathbb{E}[XZ|\mathcal{B}] = Z\mathbb{E}[X|\mathcal{B}]$ .
- Si  $\mathcal{B}_1 \subset \mathcal{B}_2$  sont deux sous-tribus alors  $\mathbb{E}[\mathbb{E}[X|\mathcal{B}_1]|\mathcal{B}_2] = \mathbb{E}[\mathbb{E}[X|\mathcal{B}_2]|\mathcal{B}_1] = \mathbb{E}[X|\mathcal{B}_1]$ .
- Si  $\sigma(X)$  est indépendante de  $\mathcal{B}$  alors  $\mathbb{E}[X|\mathcal{B}] = \mathbb{E}[X]$ .

### ■ Exemple.

- Si  $(X, Y)$  admet une densité, on peut définir la **densité conditionnelle** par  $f_{X|Y=y}(x) = f_{(X,Y)}(x,y)/f_Y(y)$  (et de manière arbitraire pour les  $y$  tels que  $f_Y(y) = 0$ ).
- Soient  $X$  et  $Y$  deux v.a. indépendantes et  $f$  borélienne telle que  $f(X, Y)$  est  $L^1$ . Alors

$$\mathbb{E}[f(X, Y) | Y] = \varphi(Y) \text{ où } \varphi(y) = \mathbb{E}[f(X, y)] \text{ (si cela a du sens et } \varphi(y) = 0 \text{ sinon).}$$

- Soit  $(X_1, \dots, X_{n+1})$  un vecteur gaussien. Il existe un unique  $(a_0, \dots, a_n)$  tel que

$$\mathbb{E}[X_{n+1} | (X_1, \dots, X_n)] = a_0 + a_1 X_1 + \dots + a_n X_n.$$



Pour voir la différence entre une espérance par rapport à  $(X, Y)$  et une espérance par rapport à  $X$  ou  $Y$ , on peut utiliser les formulations intégrales par rapport aux lois  $\mathbb{P}_{(X,Y)}$ ,  $\mathbb{P}_X$  et  $\mathbb{P}_Y$ .

## 7.2 Loi conditionnelle

Pour un couple de v.a.  $(X, Y)$ , on veut donner une définition rigoureuse de la loi de  $Y$  sachant que  $X = x$  par exemple. Pour cela, il faut parler de noyau de transition.

**Définition.** Soient  $(E, \mathcal{E})$  et  $(F, \mathcal{F})$  des espaces mesurables. On appelle **noyau de transition** de  $E$  dans  $F$  une famille de mesures de probabilités sur  $F$ , notée  $(\mathbb{P}_x(dy), x \in E)$ , telle que, pour tout  $A \in F$ ,  $x \mapsto \mathbb{P}_x(A)$  soit  $\mathcal{E}$ -mesurable.

**Définition.** Soient  $X$  et  $Y$  deux v.a. à valeurs dans  $E$  et  $F$ . La **loi conditionnelle** de  $Y$  sachant  $X = x$  est un noyau de transition, noté  $\mathbb{P}_{Y|x}$ , qui vérifie

$$\forall f, g \text{ mesurables bornées}, \quad \mathbb{E}[f(X)g(Y)] = \int f(x) \left\{ \int g(y) \mathbb{P}_{Y|x}(dy) \right\} \mathbb{P}_X(dx)$$

## 7.3 Processus stochastique

### 7.3.1 Définitions

**Définition.** Soit  $(E, \mathcal{E})$  un espace mesurable. On appelle **processus stochastique** à valeurs dans  $E$  une famille de v.a.  $(X_n)_{n \geq 0}$  à valeurs dans  $E$ .

**Définition.** Une **filtration** de la tribu  $\mathcal{F}$  est une suite croissante  $(\mathcal{F}_n)_{n \geq 0}$  de sous-tribus de  $\mathcal{F}$ . On note généralement  $\mathcal{F}_\infty = \vee_{n \geq 0} \mathcal{F}_n = \sigma(\cup_{n \geq 0} \mathcal{F}_n)$ .

**Définition.** On dit qu'un processus  $(X_n)_{n \geq 0}$  est  $(\mathcal{F}_n)_{n \geq 0}$ -adapté si pour tout  $n \geq 0$ ,  $X_n$  est  $\mathcal{F}_n$ -mesurable.

La filtration canonique associée à  $(X_n)_{n \geq 0}$  est  $(\mathcal{F}_n^X)_{n \geq 0}$  définie par  $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$ . C'est la plus petite filtration qui rend le processus adapté.



Il y a deux grands types de processus stochastiques : les chaînes de Markov et les martingales. Certains processus rentrent dans les deux catégories mais il n'y a pas de lien en général.

### 7.3.2 Temps d'arrêt

Les temps d'arrêts sont des temps aléatoires. Ils se distinguent par le fait que l'on peut décider de leur valeur (décider de s'arrêter) en se basant uniquement sur l'information passée et actuelle (et pas sur l'information future).

**Définition.** Une variable aléatoire  $T$  à valeurs dans  $\mathbb{N} \cup \{+\infty\}$  est un  $(\mathcal{F}_n)_{n \geq 0}$ -temps d'arrêt si  $\{T = n\} \in \mathcal{F}_n$  pour tout  $n$  (ou  $\{T \leq n\} \in \mathcal{F}_n$  pour tout  $n$ , ou  $\{T > n\} \in \mathcal{F}_n$  pour tout  $n$ ).



On autorise  $T = +\infty$  en général. Très souvent, il faut démontrer que  $T < +\infty$  p.s. et ce n'est pas trivial en général. En revanche, on remarquera que pour tout temps d'arrêt,  $\{T = +\infty\} \in \mathcal{F}_\infty$ .

- **Exemples :** temps constant, temps d'atteinte  $T = \inf\{n \geq 0, X_n \in A\}$  (qui peut être infini).
- **Contre-exemple :** temps de sortie  $T = \sup\{n \geq 0, X_n \in A\}$ .

**Proposition.** Soient  $S$  et  $T$  deux temps d'arrêt. Alors  $S + T$ ,  $S \vee T$  et  $S \wedge T$  sont des temps d'arrêts.

**Définition.** On appelle tribu des événements antérieurs à  $T$  la tribu

$$\mathcal{F}_T = \{A \in \mathcal{F}, \forall n \geq 0, A \cap \{T = n\} \in \mathcal{F}_n\}.$$

**Proposition.** Soient  $S$  et  $T$  deux temps d'arrêts tels que  $S \leq T$  p.s. Alors  $\mathcal{F}_S \subset \mathcal{F}_T$ .

## 7.4 Martingales

**Définition.** Un processus  $(X_n)_{n \geq 0}$  est une **martingale** par rapport à  $(\mathcal{F}_n)_{n \geq 0}$  si :

1. pour tout  $n \geq 0$ ,  $X_n$  est dans  $L^1$ ,
2.  $(X_n)_{n \geq 0}$  est  $(\mathcal{F}_n)_{n \geq 0}$ -adapté,
3.  $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$ .



Ne pas oublier les deux premiers points de la définition. Pour démontrer qu'un processus est une martingale, on vérifiera toujours le point 3 en dernier. En effet, le point 3 n'a de sens que si les deux premiers sont vérifiés.

**Définition.** Un processus  $(X_n)_{n \geq 0}$  est une **sous-martingale** (resp. sur-martingale) par rapport à  $(\mathcal{F}_n)_{n \geq 0}$  si elle vérifie les trois propriétés d'une martingale sauf que le signe d'égalité du point 3 est remplacé par le signe  $\geq$  (resp.  $\leq$ ).



L'espérance d'une martingale  $(X_n)_{n \geq 0}$  est constante, i.e.  $\mathbb{E}[X_n] = \mathbb{E}[X_0]$ . En comparaison, l'espérance d'une sous-martingale est croissante (attention : cela peut paraître contre-intuitif). Remarque :  $(X_n)_{n \geq 0}$  est une sur-martingale ssi  $(-X_n)_{n \geq 0}$  est une sous-martingale.

**Proposition.** Soit  $(X_n)_{n \geq 0}$  une martingale et  $\varphi$  une fonction convexe. Le processus  $(\varphi(X_n))_{n \geq 0}$ , s'il est intégrable, est une sous-martingale.

### 7.4.1 Martingales arrêtées

**Définition.** Soit  $(X_n)_{n \geq 0}$  un processus adapté et  $T$  un temps d'arrêt. Le processus  $(X_n^T)_{n \geq 0}$  défini par  $X_n^T = X_{n \wedge T} = X_T \mathbf{1}_{T \leq n} + X_n \mathbf{1}_{T > n}$  est appelé **processus arrêté**.

**Proposition.**

- Si  $T < +\infty$  p.s., alors  $X_{n \wedge T} \rightarrow X_T$  p.s.
- Si  $(X_n)_{n \geq 0}$  est une (sous-)martingale, alors  $(X_n^T)_{n \geq 0}$  est aussi une (sous-)martingale.



Même si on a  $\mathbb{E}[X_{n \wedge T}] = \mathbb{E}[X_0]$  pour tout  $n \geq 0$  et  $X_{n \wedge T} \rightarrow X_T$  p.s., cela n'implique pas forcément  $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ . Des conditions suffisantes sont données par les théorèmes d'arrêt ci-dessous qui reposent à chaque fois sur une interversion limite-intégrale.

**Théorème — Arrêt borné.** Soient  $(X_n)_{n \geq 0}$  une (sous-)martingale et  $T$  un temps d'arrêt borné par  $k$ . Alors  $\mathbb{E}[X_k] = \mathbb{E}[X_T] = \mathbb{E}[X_0]$  (resp.  $\geq$ ).

**Théorème — Arrêt dominé.** Soient  $(X_n)_{n \geq 0}$  une (sous-)martingale et  $T$  un temps d'arrêt p.s. fini. S'il existe  $Y$  dans  $L^1$  telle que  $|X_{n \wedge T}| \leq Y$  alors  $\mathbb{E}[X_T] = \mathbb{E}[X_0]$  (resp.  $\geq$ ).

### 7.4.2 Théorèmes de convergence

**Théorème.** Soit  $(X_n)_{n \geq 0}$  une (sous-)martingale bornée dans  $L^1$ . Alors  $(X_n)_{n \geq 0}$  converge p.s. vers  $X_\infty$ . De plus,  $X_\infty$  est dans  $L^1$ . Mais  $X_n$  ne converge pas forcément vers  $X_\infty$  dans  $L^1$ .

■ **Exemple.** Le théorème s'applique pour les martingales de signe constant (martingales bornées en particulier) ou dominées et également pour les sous-martingales négatives.

**Théorème.** Soit  $(X_n)_{n \geq 0}$  une martingale bornée dans  $L^2$ . Alors  $(X_n)_{n \geq 0}$  converge p.s. et dans  $L^2$  vers  $X_\infty$ . De plus,  $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$  et en particulier  $\mathbb{E}[X_\infty] = \mathbb{E}[X_0]$ .

■ **Contre-exemple.** On peut trouver une sur-martingale  $(X_n)_{n \geq 1}$  telle que la loi de  $X_n$  soit  $(1 - 1/n^2)\delta_0 + (1/n^2)\delta_n$ . Elle est bornée dans  $L^2$  mais ne converge pas dans  $L^2$ . ■

## 7.5 Complément classique (pour votre culture)

**Théorème de convergence  $L^1$ .** Soit  $(X_n)_{n \geq 0}$  une (sous-)martingale. Les trois conditions suivantes sont équivalentes :

- (i)  $X_n$  converge p.s. et dans  $L^1$ ,
- (ii) **martingale fermée** : il existe une v.a.  $Y$  dans  $L^1$  telle que  $X_n = \mathbb{E}[Y | \mathcal{F}_n]$ ,
- (iii) **uniforme intégrabilité** : la suite  $(X_n)_{n \geq 0}$  est uniformément intégrable.

# 8 ESTIMATION

## 8.1 Introduction

Dans la théorie des probabilités, on suppose parfaitement connu l'aléa (la loi) et on cherche à donner des informations sur la variable aléatoire. La but de la statistique est inverse : **on part de réalisations de variables aléatoires et on cherche à inférer la loi sous-jacente.**

**Définition.** Un modèle statistique est un triplet  $(\Omega, \mathcal{F}, \mathcal{P})$ .

On dit que le modèle est paramétré si  $\mathcal{P} = \{\mathbb{P}_\theta\}_{\theta \in \Theta}$  et identifiable si  $\theta \mapsto \mathbb{P}_\theta$  est injective.

On dit que le modèle est paramétrique si  $\Theta \subset \mathbb{R}^d$ .

**Notation.** L'espérance et la variance sous la loi  $\mathbb{P}_\theta$  sont notées  $\mathbb{E}_\theta$  et  $\text{Var}_\theta$ .

Dans la suite, on suppose que le modèle est paramétrique. Nous verrons deux possibilités d'estimation d'un paramètre d'une loi de probabilité :

- ponctuelle (on renvoie une valeur dont on peut mesurer la précision),
- par intervalle ou région (dont on peut mesurer le degré de confiance).

**Définition.** Un *n*-échantillon d'une variable aléatoire  $X$  est  $(X_1, \dots, X_n)$  tel que les  $X_i$  sont i.i.d. de même loi que  $X$ . Une **réalisation** de cet échantillon est le vecteur des valeurs observées, généralement noté  $(x_1, \dots, x_n)$ .



De manière générale, on utilise des lettres majuscules pour les quantités aléatoires et des lettres minuscules pour leur pendant déterministe. C'est également le cas en probabilités, mais cela est encore plus flagrant en statistique.

**Définition.** Les moments d'une v.a. aléatoire  $X$  sont les quantités  $\mathbb{E}[X^k]$  pour  $k$  entier. Les moments empiriques correspondant sont

$$\overline{X_n^k} = \frac{1}{n} \sum_{i=1}^n X_i^k.$$

L'exemple le plus important est la **moyenne empirique** généralement notée  $\bar{X}_n$ .

**Définition.** La loi empirique de l'échantillon  $(X_1, \dots, X_n)$  est  $\mu_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ . C'est une loi discrète et sa fonction de répartition (appelée **fonction de répartition empirique**) est la fonction en escalier définie par  $F_n(t) = n^{-1} \sum_{i=1}^n \mathbf{1}_{]-\infty, t]}(X_i)$ .

## 8.2 Estimation ponctuelle

### 8.2.1 Définitions

**Définition.** Un **estimateur**  $T$  de  $\theta$  est une v.a. de la forme  $\varphi(X_1, \dots, X_n)$ . L'estimation de  $\theta$  correspondante est alors  $\hat{\theta} = \varphi(x_1, \dots, x_n)$ .



Un estimateur dépend de  $n$  (notamment lorsque l'on parle de consistance, voir la définition ci-dessous). Généralement, on ne précise pas cette dépendance car un estimateur représente très souvent une méthode d'estimation plus qu'une formule à  $n$  fixé. **Attention :**  $\varphi$  ne dépend pas du paramètre inconnu  $\theta$ .

**Définition.** Le **biais** de l'estimateur  $T$  de  $\theta$  est  $b_\theta(T) = \mathbb{E}_\theta[T] - \theta$ . Si le biais est nul, on dit que l'estimateur est **sans biais**. Si le biais tend vers 0, on dit que l'estimateur est **asymptotiquement sans biais**.

**Définition.** On dit que  $T$  est **faiblement** (resp. **fortement**) **consistant** si  $T \rightarrow \theta$  en probabilité (resp. presque sûrement) sous  $\mathbb{P}_\theta$ . De plus, si  $\sqrt{n}(T - \theta)$  converge en loi vers une loi normale centrée, on dit que  $T$  est **asymptotiquement normal**.

**Définition.** Pour mesurer la performance d'un estimateur, on utilise le **risque quadratique**

$$R(T, \theta) = \mathbb{E}_\theta [(T - \theta)^2].$$

**Proposition — Décomposition biais/variance.** Le risque quadratique se décompose en

$$R(T, \theta) = b_\theta(T)^2 + \text{Var}_\theta(T).$$

### 8.2.2 Méthodes

► **Méthode des moments** On exhibe une fonction  $\Phi$  telle que  $g(\theta) = \mathbb{E}_\theta[\Phi(X)]$  est “simple”. On utilise alors la “moyenne empirique”  $\hat{g}(\theta) = n^{-1} \sum_{i=1}^n \Phi(X_i)$  pour estimer la valeur  $g(\theta)$ . Enfin, si  $g$  est inversible et  $g^{-1}$  continue alors  $g^{-1}(\hat{g}(\theta))$  est un estimateur consistant de  $\theta$ .

► **Méthode du maximum de vraisemblance**

1. **Cas discret.** On note  $f_\theta(x) = \mathbb{P}_\theta(X = x)$ .
2. **Cas continu.** On note  $f_\theta$  la densité.

On définit la **vraisemblance** de  $\theta$  au vu de l'échantillon  $(X_1, \dots, X_n)$  :

$$L_{X_1, \dots, X_n}(\theta) = \prod_{i=1}^n f_\theta(X_i).$$

Pour estimer  $\theta$  on prendra  $\hat{\theta} = \varphi(X_1, \dots, X_n) \in \arg \max_\theta L_{X_1, \dots, X_n}(\theta)$ . En général,  $L_{X_1, \dots, X_n}$  admet un unique maximum qu'il est plus simple de trouver en considérant la **log-vraisemblance**.

■ **Exemple.**

- Pour les modèles gaussien  $\mathcal{N}(m, 1)$ , exponentielle  $\mathcal{E}(\lambda)$  et Bernoulli  $\mathcal{B}(p)$  les deux méthodes donnent le même estimateur :  $\bar{X}_n$ ,  $1/\bar{X}_n$  et  $\bar{X}_n$  respectivement.
- Pour le modèle uniforme  $\mathcal{U}([0, \theta])$ , les deux méthodes donnent respectivement  $2\bar{X}_n$  et  $\max_{1 \leq i \leq n} X_i$ .

### 8.2.3 Principaux exemples et propriétés

Dans la suite, on suppose que  $X$  est  $L^1$  sous  $\mathbb{P}_\theta$  et on note  $\mu = \mathbb{E}_\theta[X]$ . Si  $X$  est  $L^2$ , on note  $\sigma^2 = \text{Var}_\theta(X)$ .

**Théorème — Moyenne empirique.** La moyenne empirique est un estimateur sans biais et consistant de la moyenne théorique  $\mu$ . Si  $X$  est  $L^2$ , il est également asymptotiquement normal (limite  $\mathcal{N}(0, \sigma^2)$ ).

**Théorème — Variance empirique à moyenne connue.** Si  $X$  est  $L^2$ , la variance empirique

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

est un estimateur sans biais et consistant de la variance théorique  $\sigma^2$ . Si  $X$  est  $L^4$ , il est également asymptotiquement normal (limite  $\mathcal{N}(0, \mathbb{E}[(X - \mu)^4] - \sigma^4)$ ).

Lorsque la moyenne est inconnue, on peut la remplacer par la moyenne empirique. Attention : on corrige la formule pour obtenir un estimateur sans biais.

**Théorème — Variance empirique à moyenne inconnue.** Si  $X$  est  $L^2$ , la variance empirique corrigée

$$\hat{S}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

est un estimateur sans biais et consistant de la variance théorique  $\sigma^2$ . Si  $X$  est  $L^4$ , il est également asymptotiquement normal (limite  $\mathcal{N}(0, \mathbb{E}[(X - \mu)^4] - \sigma^4)$ ).

## 8.3 Estimation par région de confiance

Une estimation ponctuelle ne sera jamais exactement égale au paramètre recherché. Pour attraper ce paramètre inconnu, on peut donner une région autour de notre estimateur et mesurer le degré de confiance de cette région. Dans la suite, on note  $\alpha \approx 0$  le niveau de risque et donc  $1 - \alpha \approx 1$  est le niveau de confiance.

### 8.3.1 Région de confiance et quantile

**Définition.** Une **région de confiance** pour  $\theta$  de niveau  $1 - \alpha$  est un ensemble  $C(X_1, \dots, X_n) \subset \Theta$  tel que

$$\mathbb{P}_\theta(\theta \in C(X_1, \dots, X_n)) \geq 1 - \alpha. \quad (8.1)$$

Pour préciser que c'est une inégalité, on peut dire qu'on a une **région par excès**. Si l'égalité est réalisée dans (8.1), on dit que le niveau est **exact**. Si (8.1) est vérifié à la limite  $n \rightarrow +\infty$  on dit que le niveau est **asymptotique**.

**Définition.** Soit  $\mathbb{P}$  une loi de probabilité sur  $\mathbb{R}$  de fonction de répartition  $F$  et  $\alpha \in ]0, 1[$ . On appelle **quantile** d'ordre  $\alpha$  le réel

$$z_\alpha = \inf \{x \in \mathbb{R}, F(x) \geq \alpha\}.$$

**Proposition.** La fonction quantile est parfois appelée **fonction inverse généralisée** de  $F$ . Si  $F$  est continue alors  $F(z_\alpha) = \alpha$ . Si de plus  $F$  est strictement croissante, c'est l'inverse fonctionnel standard, et dans ce cas le support de  $\mathbb{P}$  est l'adhérence de  $]z_0, z_1[$ .

### 8.3.2 Exemples

– Par l'inégalité de Bienaymé-Tchebychev.

**Proposition.** Soit  $(\mathbb{P}_\mu)_{\mu \in \Theta}$  un modèle tel que  $\mu = \mathbb{E}_\mu[X]$  et  $\text{Var}_\mu(X) \leq v$  pour tout  $\mu$ . Alors

$$\left[ \bar{X}_n - \sqrt{\frac{v}{n\alpha}}, \bar{X}_n + \sqrt{\frac{v}{n\alpha}} \right]$$

est un intervalle de confiance par excès de niveau  $1 - \alpha$ .

– **Modèle gaussien** : les lois des estimateurs  $\bar{X}_n$ ,  $S_n^2$  et  $\hat{S}_n^2$  sont connues et cela permet de donner des intervalles de confiances exacts pour la moyenne ou la variance selon que l'autre paramètre est connu ou non.

Les résultats condensés dans le tableau ci-dessous sont justifiés par le Théorème de Cochran et la propriété fondamentale de la loi de Student.

	Autre paramètre connu	Autre paramètre inconnu
IC pour $m$	$\left[ \bar{X}_n \pm u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$ quantile de $\mathcal{N}(0, 1)$	$\left[ \bar{X}_n \pm t_{1-\alpha/2} \frac{\hat{S}_n}{\sqrt{n}} \right]$ quantile de $\mathcal{T}(n-1)$
IC pour $\sigma^2$	$\left[ \frac{nS_n^2}{z_{1-\alpha/2}}, \frac{nS_n^2}{z_{\alpha/2}} \right]$ quantile de $\chi^2(n)$	$\left[ \frac{(n-1)\hat{S}_n^2}{z_{1-\alpha/2}}, \frac{(n-1)\hat{S}_n^2}{z_{\alpha/2}} \right]$ quantile de $\chi^2(n-1)$

– **Loi asymptotique connue.**

Le **Théorème Central Limite** et le **Lemme de Slutsky** permettent de construire des intervalles de confiance asymptotique en se ramenant au cas gaussien. L'exemple le plus fréquent est celui du **modèle de Bernoulli** (sondages) : l'intervalle de confiance asymptotique pour la proportion  $p$  est

$$\left[ \bar{X}_n - u_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}, \bar{X}_n + u_{1-\alpha/2} \sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}} \right].$$

# 9 VECTEURS GAUSSIENS

Dans ce chapitre, on notera  $\langle \cdot, \cdot \rangle$  le produit scalaire canonique sur  $\mathbb{R}^d$ .

## 9.1 Définition

**Définition.** On dit qu'un vecteur aléatoire  $X$  à valeurs dans  $\mathbb{R}^d$  est un **vecteur gaussien** si pour tout  $a \in \mathbb{R}^d$ , la variable aléatoire réelle  $\langle a, X \rangle$  suit une loi gaussienne.

**Proposition.** Soit  $X$  un vecteur gaussien dans  $\mathbb{R}^d$ . Alors  $X$  admet des moments d'ordre 2. De plus, la loi de  $X$  est caractérisée par son espérance, notée  $\mu \in \mathbb{R}^d$ , et sa matrice de covariance, notée  $\Gamma$ . On note  $\mathcal{N}_d(\mu, \Gamma)$  cette loi, qui est appelée la **loi gaussienne de moyenne  $\mu$  et de matrice de covariance  $\Gamma$** . Enfin, pour tout  $a \in \mathbb{R}^d$ ,

$$\mathbb{E}[e^{i\langle a, X \rangle}] = e^{i\langle a, \mu \rangle - \frac{1}{2}\langle a, \Gamma a \rangle}.$$

**Corollaire.** Si  $A$  est une matrice réelle  $n \times d$  et  $X \sim \mathcal{N}_d(\mu, \Gamma)$  alors  $AX \sim \mathcal{N}_n(A\mu, A\Gamma A^T)$ .



Pour simuler un vecteur  $X \sim \mathcal{N}_d(\mu, \Gamma)$ , il suffit de trouver une matrice  $S$  telle que  $SS^T = \Gamma$  (par exemple la racine carrée symétrique positive de  $\Gamma$ ) et de simuler  $Z = (Z_1, \dots, Z_d)$  où les  $Z_i$  sont i.i.d. de loi  $\mathcal{N}(0, 1)$ . En effet,  $X = \mu + SX$  est bien distribué selon  $\mathcal{N}_d(\mu, \Gamma)$ .

**Corollaire.** Soit  $X \sim \mathcal{N}_d(\mu, \Gamma)$  et  $I \subset \{1, \dots, d\}$ . Alors, les  $(X_i)_{i \in I}$  sont **indépendantes** si et seulement si  $\Gamma_{i,j} = 0$  pour tous  $i \neq j$  dans  $I$ .

**Proposition.** La loi  $\mathcal{N}_d(\mu, \Gamma)$  admet une densité par rapport à la mesure de Lebesgue sur  $\mathbb{R}^d$  si et seulement si  $\Gamma$  est inversible. Dans ce cas, la densité est donnée par

$$x \mapsto \frac{1}{(2\pi)^{d/2} \sqrt{\det(\Gamma)}} e^{-\frac{1}{2}\langle x - \mu, \Gamma^{-1}(x - \mu) \rangle}.$$

## 9.2 Applications importantes

**Théorème — TCL vectoriel.** Soit  $(X_i)_{i \geq 1}$  une suite de vecteurs aléatoires i.i.d à valeurs dans  $\mathbb{R}^d$ , admettant des moments d'ordre 2. On note  $\mu$  l'espérance commune et  $\Gamma$  la matrice de covariance commune. Alors

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mu) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, \Gamma).$$

**Théorème de Cochran.** Soit  $X \sim \mathcal{N}_d(0, I_d)$  et  $V$  un s.e.v de  $\mathbb{R}^d$ . On note  $\Pi_V$  la matrice  $d \times d$  de projection orthogonale sur  $V$ . Alors,

- $\Pi_V X \sim \mathcal{N}_d(0, \Pi_V)$ ,
- $\Pi_{V^\perp} X$  et  $\Pi_V X$  sont indépendants,
- $\|\Pi_V X\|^2 \sim \chi^2(\dim(V))$ .

**Proposition — Loi conditionnelle.** Soit  $X \sim \mathcal{N}_d(\mu, \Gamma)$ ,  $A$  une matrice réelle  $a \times d$  et  $B$  une matrice réelle  $b \times d$ . Alors, pour tout  $y$  appartenant au support de  $BX$ , la loi conditionnelle de  $AX$  sachant " $BX = y$ " est une loi gaussienne. En notant  $\alpha$  une matrice  $b \times a$  telle que :

$$B\Gamma B^T \alpha = B\Gamma A^T,$$

alors la loi conditionnelle de  $AX$  sachant " $BX = y$ " est

$$\mathcal{N}_b(A\mu + (y - B\mu)^T \alpha, A\Gamma A^T - \alpha^T B\Gamma A^T).$$

■ **Exemple.** Pour  $d = 2$ ,  $A = (1 \ 0)$  et  $B = (0 \ 1)$ , le résultat précédent donne la loi de  $X_1$  sachant  $X_2 = y$  pour un vecteur gaussien  $(X_1, X_2)$  tel que  $\text{Var}(X_2) > 0$ . Cette loi est

$$\mathcal{N}\left(\mathbb{E}[X_1] + \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)}(y - \mathbb{E}[X_2]), \text{Var}(X_1) - \frac{\text{Cov}(X_1, X_2)^2}{\text{Var}(X_2)}\right).$$

# 10 MODÈLE LINÉAIRE GAUSSIEN

## 10.1 Cas général

### 10.1.1 Définition

Le modèle linéaire gaussien (**MLG**) est un cadre simple dans lequel on peut mettre en œuvre de nombreux tests et intervalles de confiance. L'outil technique qui en simplifie l'étude est le **Théorème de Cochran**.

**Définition.** Soit  $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$  un vecteur aléatoire. On dit que  $Y$  suit un **MLG** si

$$Y = \mu + \varepsilon \quad \text{où } \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n).$$

Les paramètres  $\mu \in \mathbb{R}^n$  et  $\sigma^2 > 0$  sont en général inconnus, mais on suppose que  $\mu$  appartient à  $V \subset \mathbb{R}^n$ . On dit que la dimension du modèle est  $p$ .

Le vecteur  $Y$  n'est pas un  $n$ -échantillon : les coordonnées sont indépendantes, mais elles n'ont pas la même moyenne en général.



L'ensemble  $V$  est connu du statisticien. C'est généralement un sous-e.v. de  $\mathbb{R}^n$  de dimension  $p < n$ . Typiquement,  $p$  est fixé alors  $n$  tend vers l'infini. Le principal type d'espaces  $V$  étudiés sera explicité dans la section "Régression linéaire".

### 10.1.2 Estimation par moindres carrés

Dans la suite, on suppose que  $V$  est connu et on note  $\Pi_V$  la projection orthogonale sur  $V$ .

**Théorème.** Soit  $Y$  un vecteur aléatoire qui suit un **MLG**.

1. L'estimateur du maximum de vraisemblance de  $\theta = (\mu, \sigma^2)$  est  $\hat{\theta}_n = (\hat{\mu}_n, \hat{s}_n^2)$  où

$$\hat{\mu}_n := \Pi_V Y \quad \text{et} \quad \hat{s}_n^2 := \frac{1}{n} \|Y - \Pi_V Y\|_2^2.$$

2. Les estimateurs  $\hat{\mu}_n$  et  $\hat{s}_n^2$  sont indépendants et

$$\hat{\mu}_n \sim \mathcal{N}_n(\mu, \sigma^2 \Pi_V) \quad \text{et} \quad \frac{n}{\sigma^2} \hat{s}_n^2 \sim \chi^2(n-p).$$

3. L'estimateur  $\hat{\sigma}_n^2 := \frac{n}{n-p} \hat{s}_n^2$  est un estimateur sans biais de  $\sigma^2$ .

**Proposition — Estimateur des moindres carrés.** L'estimateur  $\hat{\mu}_n$  vérifie

$$\hat{\mu}_n \in \arg \min_{\mu \in V} \|Y - \mu\|_2.$$

### 10.1.3 Intervalles de confiance et tests

► **Pour la variance.** On se donne une valeur de référence  $\sigma_0 > 0$  et on souhaite tester  $\mathcal{H}_0 : \sigma \leq \sigma_0$  contre  $\mathcal{H}_1 : \sigma > \sigma_0$ . Le test de niveau  $\alpha$  est donné par :

- la statistique  $T_n = \frac{n-p}{\sigma_0^2} \hat{\sigma}_n^2$  qui suit la loi  $\chi^2(n-p)$  quand  $\sigma = \sigma_0$ ,
- et la zone de rejet  $R_n = ]z_{n-p, 1-\alpha}, +\infty[$ , où  $z_{v, \alpha}$  est un quantile de la loi  $\chi^2(v)$ .

De même, on en déduit un intervalle de confiance pour  $\sigma^2$  de niveau  $1 - \alpha$  :

$$\left[ \frac{n-p}{z_{n-p,1-\alpha}} \hat{\sigma}_n^2, +\infty \right].$$

► **Pour la moyenne.** Pour se ramener à la dimension 1, on se donne un vecteur  $c$  (par exemple un vecteur de base de l'espace  $V$  où l'on sait que la moyenne  $\mu$  se trouve). On va chercher à tester ou estimer la coordonnée  $c^T \mu$ .

On se donne une valeur de référence  $a \in \mathbb{R}$  et on souhaite tester  $\mathcal{H}_0 : c^T \mu \geq a$  contre  $\mathcal{H}_1 : c^T \mu < a$ . Le test de niveau  $\alpha$  est donné par :

- la statistique  $T_n = \frac{c^T \hat{\mu}_n - a}{\hat{\sigma}_n \sqrt{c^T \Pi_V c}}$  qui suit la loi  $\mathcal{T}(n-p)$  quand  $c^T \mu = a$ ,
- et la zone de rejet  $R_n = ]-\infty, t_{n-p,\alpha}[$ , où  $t_{v,\alpha}$  est un quantile de la loi  $\mathcal{T}(v)$ .

De même, on en déduit un intervalle de confiance pour  $c^T \mu$  de niveau  $1 - \alpha$  :

$$\left] -\infty, c^T \hat{\mu}_n - t_{n-p,\alpha} \hat{\sigma}_n \sqrt{c^T \Pi_V c} \right].$$

## 10.2 Régression linéaire

### 10.2.1 Définition

En pratique, le MLG est principalement utilisé dans le cadre de la régression linéaire :

$V = \text{Im}(\mathbf{X})$  où  $\mathbf{X}$  est une matrice réelle injective de dimension  $p \times n$  ( $p \leq n$ ).

**Définition.** Soit  $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$  un vecteur aléatoire. On dit que  $Y$  suit un **MLG de régression linéaire** si

$$Y = \mathbf{X}\beta + \varepsilon \quad \text{où } \varepsilon \sim \mathcal{N}_n(0, \sigma^2 I_n).$$

Les paramètres  $\beta \in \mathbb{R}^p$  et  $\sigma^2 > 0$  sont en général inconnus, mais on suppose que  $\mathbf{X}$  est connue. On dit que la dimension du modèle est  $p$ .

En remarquant que la projection orthogonale  $\Pi_V$  s'écrit comme  $\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ , on peut reformuler les estimateurs.

**Théorème.** Soit  $Y$  un vecteur aléatoire qui suit un MLG de régression linéaire.

1. L'estimateur du maximum de vraisemblance de  $\theta = (\beta, \sigma^2)$  est  $\hat{\theta}_n = (\hat{\beta}, \hat{s}_n^2)$  où

$$\hat{\beta} := (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y \quad \text{et} \quad \hat{s}_n^2 := \frac{1}{n} \|Y - \mathbf{X}\hat{\beta}\|_2^2.$$

2. Les estimateurs  $\hat{\beta}_n$  et  $\hat{s}_n^2$  sont indépendants et

$$\hat{\beta}_n \sim \mathcal{N}_p(\beta, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}) \quad \text{et} \quad \frac{n}{\sigma^2} \hat{s}_n^2 \sim \chi^2(n-p).$$

3. L'estimateur  $\hat{\sigma}_n^2 := \frac{n}{n-p} \hat{s}_n^2$  est un estimateur sans biais de  $\sigma^2$ .

De même, des intervalles de confiance et tests pour  $\sigma^2$  et  $\beta$  peuvent se déduire de la section précédente.

Il est bon de noter que l'on peut faire de la sélection de modèle : est-ce que l'ajout de la dernière coordonnée  $\beta_p$  est pertinent ? En effet, si le modèle de dimension  $p$  n'est pas plus efficace que celui de dimension  $p-1$ , on va préférer le second qui est plus simple. Ainsi, on souhaite tester  $\mathcal{H}_0 : \beta_p = 0$  (la dernière coordonnée n'est pas pertinente) contre  $\mathcal{H}_1 : \beta_p \neq 0$ . Le test de niveau  $\alpha$  est donné par :

- la statistique  $T_n = \frac{\hat{\beta}_p}{\hat{\sigma}_n \sqrt{(\mathbf{X}^T \mathbf{X})_{p,p}^{-1}}}$  qui suit la loi  $\mathcal{T}(n-p)$  quand  $\beta_p = 0$ ,
- et la zone de rejet  $R_n = ]-\infty, t_{n-p,\alpha/2}[ \cup ]t_{n-p,1-\alpha/2}, +\infty[$ .

### 10.2.2 Exemple

Le cas le plus simple est celui de la régression linéaire simple :  $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$  avec  $\beta_1, \beta_2 \in \mathbb{R}$ . Cela correspond au cas général avec  $p = 2$  et

$$X = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}.$$

Si l'on dispose du nuage de points  $\{(X_i, Y_i)\}$ , le but est de **trouver la droite qui représente au mieux ce nuage de points**. En notant  $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$  et  $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ , on a

$$\hat{\beta}_1 := \bar{Y}_n - \hat{\beta}_2 \bar{X}_n \quad \text{et} \quad \hat{\beta}_2 := \frac{n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)(Y_i - \bar{Y}_n)}{n^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2},$$

où le coefficient directeur de la droite de régression, i.e.  $\hat{\beta}_2$ , s'interprète comme la covariance entre  $X$  et  $Y$ , renormalisée par la variance de  $X$ .



La régression linéaire permet de modéliser des relations linéaires entre des **variables explicatives**  $X$  et des **variables à expliquer**  $Y$ . Mais, des relations non linéaires (prescrites) sont tout à fait possibles. Par exemple, l'estimation des paramètres  $\beta_1, \beta_2, \beta_3$  du modèle  $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i$  donne la parabole qui représente au mieux le nuage de points.

## 10.3 Compléments classiques (pour votre culture)

Le livre **Statistique en Action** de Rivoirard et Stoltz regorge de cas particuliers (analyse de la variance, ...), de tests et de régions de confiance qui s'appliquent dans ce modèle. On notera notamment la loi jointe de l'estimateur  $(\hat{\beta}_1, \hat{\beta}_2)$  de la régression linéaire simple, ainsi que les intervalles de confiance et de prédiction pour une nouvelle observation.

Le livre **Probabilités et Statistiques pour l'Épreuve de Modélisation à l'Agrégation de Mathématiques** de Chabanol et Ruch présente la régression linéaire dans un cadre plus général que celui du MLG. Dans ce cas, la régression consiste en une projection à l'aide de l'espérance conditionnelle, et cette régression est linéaire lorsque  $\mathbb{E}[Y | X] = aX + b$ .



# 11 TESTS PARAMÉTRIQUES

## 11.1 Généralités sur les tests

Un **test d'hypothèse** vise à vérifier la compatibilité (ou plutôt l'**absence d'incompatibilité**) entre les observations et une hypothèse de travail, généralement appelée **hypothèse nulle**  $\mathcal{H}_0$  (e.g. “l’espérance est  $\mu_0$ ” ou “les deux échantillons sont issus de la même loi”). D’un point de vue pratique, un test statistique peut répondre aux questions suivantes.

- Est-ce que les paquets produits par la machine respectent le poids indiqué ?
- Est-ce qu’un nouveau traitement est plus efficace que l’ancien ?
- Est-ce que la campagne publicitaire augmente la popularité de la marque ?
- Est-ce que le taux de réussite d’une opération chirurgicale dépend de l’âge du patient ?

**Méthode.** Le schéma de construction/application d’un test est le suivant :

1. définir l’hypothèse nulle  $\mathcal{H}_0$  et l’hypothèse alternative  $\mathcal{H}_1$  (en général  $\mathcal{H}_0^c$ ),
2. choisir une **statistique/estimateur**  $T$  pertinent pour contrôler  $\mathcal{H}_0$ ,
3. déterminer la loi (asymptotique) de la statistique  $T$  sous l’hypothèse  $\mathcal{H}_0$ ,
4. définir une **zone de rejet**  $R_\alpha$  en fonction du niveau  $\alpha \approx 0$ ,
5. calculer la réalisation  $t$  de  $T$  sur le jeu de données et prendre une décision :
  - si l’estimation  $t \in R_\alpha$ , **on rejette**  $\mathcal{H}_0$  au profit de  $\mathcal{H}_1$ ,
  - si l’estimation  $t \notin R_\alpha$ , **on ne rejette pas**  $\mathcal{H}_0$ .

**Définition.** Il existe deux types d’erreur possibles pour un test.

1. **Risque d’erreur de première espèce** : “ $\mathbb{P}_{\mathcal{H}_0}(\text{rejeter } \mathcal{H}_0)$ ”.
2. **Risque d’erreur de seconde espèce** : “ $\mathbb{P}_{\mathcal{H}_1}(\text{ne pas rejeter } \mathcal{H}_0)$ ”.



On ne peut pas contrôler les deux erreurs en même temps. C’est pour cette raison qu’un **test est asymétrique** vis à vis de  $\mathcal{H}_0$  et  $\mathcal{H}_1$ . La convention est de contrôler le risque de première espèce par  $\alpha \approx 0$ . Ainsi, si les observations nous poussent à rejeter  $\mathcal{H}_0$ , on a de fortes raisons de faire confiance à cette décision.

La notion de puissance d’un test est parfois utilisée pour désigner le complémentaire du risque de seconde espèce : **plus un test est puissant plus il sera apte à détecter**  $\mathcal{H}_1$ .

## 11.2 Test paramétrique

Dans la suite on se place dans un modèle paramétrique  $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta \in \Theta})$  et on suppose disposer d’un échantillon  $(X_1, \dots, X_n)$ . On se donne deux sous-ensembles,  $\Theta_0$  et  $\Theta_1$  (en général  $\Theta_1 = \Theta_0^c$ ), disjoints dans  $\Theta$  et on cherche à tester

$$\mathcal{H}_0 : \theta \in \Theta_0 \quad \text{contre} \quad \mathcal{H}_1 : \theta \in \Theta_1.$$

### 11.2.1 Définitions

**Définition.** Soit  $T_n$  une statistique réelle et  $R_n \subset \mathbb{R}$  une zone de rejet. La règle de décision du test défini par  $(T_n, R_n)$  est :

- si  $T_n \in R_n$ , le test rejette  $\mathcal{H}_0$  au profit de  $\mathcal{H}_1$ ,
- si  $T_n \notin R_n$ , le test ne rejette pas  $\mathcal{H}_0$ .

**Définition.** On définit les fonctions “erreur de première et seconde espèce” par

$$\tilde{\alpha}_n : \theta \in \Theta_0 \mapsto \mathbb{P}_\theta(T_n \in R_n) \quad \text{et} \quad \tilde{\beta}_n : \theta \in \Theta_1 \mapsto \mathbb{P}_\theta(T_n \notin R_n).$$

La fonction **puissance** est donc  $1 - \tilde{\beta}$  et la **taille du test** est définie par  $\bar{\alpha} = \sup \tilde{\alpha}(\theta)$ .

**Définition.** On dit que le test  $(T_n, R_n)$  est de **niveau**  $\alpha \in ]0, 1[$  si sa taille est majorée par  $\alpha$ , i.e.

$$\forall \theta \in \Theta_0, \mathbb{P}_\theta(T_n \in R_n) \leq \alpha.$$

Comme pour les intervalles de confiance, on parle également de **niveau asymptotique**.

De plus, on dit que le test est **consistant** si pour tout  $\theta \in \Theta_1$ ,  $\tilde{\beta}_n(\theta) \rightarrow 0$ .



Par contrôler le niveau d'un test, on garde la même statistique  $T$  mais on ajuste la zone de rejet  $R$ . C'est pour cela que pour un test de niveau  $\alpha$  on la note  $R_\alpha$ . En général,  $R_\alpha$  est décroissant par rapport à  $\alpha$  au sens de l'inclusion. On peut alors s'intéresser au plus grand  $\alpha$  pour lequel  $T \notin R_\alpha$ .

**Définition.** La **p-valeur** d'une famille de test  $(T, (R_\alpha)_{\alpha \in [0,1]})$  est définie par

$$\alpha^*(T) = \sup\{\alpha \in [0, 1], T \notin R_\alpha\}.$$

La p-valeur est un seuil de décision :

- si  $\alpha > \alpha^*$  alors le test rejette  $\mathcal{H}_0$ ,
- si  $\alpha < \alpha^*$  alors le test ne rejette pas  $\mathcal{H}_0$ .



La **p-valeur** est une variable aléatoire : elle dépend des observations. Une petite p-valeur exprime le fait que les observations rejettent significativement  $\mathcal{H}_0$ . Une grande p-valeur ne justifie pas  $\mathcal{H}_0$  (cela dépend de la puissance et n'est pas contrôlé par le niveau).

## 11.2.2 Exemples

► **Test de conformité.** On cherche à comparer le paramètre  $\theta$  à une valeur de référence  $\theta_0$ .

– **Contrôle de qualité, sondage** : on suppose que les observations sont distribuées selon la loi  $\mathcal{B}(p)$ . On souhaite tester le paramètre inconnu  $p$  par rapport à une valeur de référence  $p_0$ . On obtient des tests de **taille asymptotique**  $\alpha$  et **consistants** avec la statistique

$$T = \sqrt{n} \frac{\bar{X}_n - p_0}{\sqrt{p_0(1-p_0)}},$$

et les zones de rejets suivantes :

$\mathcal{H}_0 : p = p_0$	$\mathcal{H}_0 : p \geq p_0$	$\mathcal{H}_0 : p \leq p_0$
$R_\alpha = [-u_{1-\alpha/2}, u_{1-\alpha/2}]^c$	$R_\alpha = ]-\infty, -u_{1-\alpha}[$	$R_\alpha = ]u_{1-\alpha}, +\infty[$

– **Modèle gaussien** : on suppose que les observations sont distribuées selon la loi  $\mathcal{N}(\mu, \sigma^2)$ . Le tableau ci-dessous donne la statistique de test pour tester la moyenne inconnue  $\mu$  par rapport à une moyenne de référence  $\mu_0$  dans les deux cas où la variance est connue ou inconnue. Dans chaque cas, trois zones de rejet  $R_\alpha$  (qui rendent le test de **taille**  $\alpha$ ) sont données : une bilatérale et deux unilatérales. A chaque fois on suppose que  $\mathcal{H}_1 = \mathcal{H}_0^c$ .

	Variance connue	Variance inconnue
Statistique $T$	$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma}$	$\sqrt{n} \frac{\bar{X}_n - \mu_0}{\hat{S}_n}$
$\mathcal{H}_0: \mu = \mu_0$	$R_\alpha = [-u_{1-\alpha/2}, u_{1-\alpha/2}]^c$ quantile de $\mathcal{N}(0, 1)$	$R_\alpha = [-t_{1-\alpha/2}, t_{1-\alpha/2}]^c$ quantile de $\mathcal{T}(n-1)$
$\mathcal{H}_0: \mu \geq \mu_0$	$R_\alpha = ]-\infty, -u_{1-\alpha}[$	$R_\alpha = ]-\infty, -t_{1-\alpha}[$
$\mathcal{H}_0: \mu \leq \mu_0$	$R_\alpha = ]u_{1-\alpha}, +\infty[$	$R_\alpha = ]t_{1-\alpha}, +\infty[$

► **Test de comparaison.** On cherche à comparer les paramètres  $\theta_1$  et  $\theta_2$  des lois de deux échantillons.

– **Moyennes :** si les deux échantillons ont la même taille, on se ramène à tester si la moyenne des différences est nulle.

Sinon, supposons que  $(X_1, \dots, X_{n_1})$  est distribué selon  $\mathcal{N}(\mu_1, \sigma_1^2)$  et que  $(Y_1, \dots, Y_{n_2})$  est distribué selon  $\mathcal{N}(\mu_2, \sigma_2^2)$ . Si les variances sont connues, on considère la statistique normale

$$T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}.$$

Sinon, en supposant que les variances sont égales, on peut considérer la statistique de Student

$$T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{\hat{S}_{n_1, n_2}^2(1/n_1 + 1/n_2)}},$$

où

$$\hat{S}_{n_1, n_2}^2 = \frac{1}{n_1 + n_2 - 2} \left( \sum_{i=1}^{n_1} (X_i - \bar{X}_{n_1})^2 + \sum_{i=1}^{n_2} (Y_i - \bar{Y}_{n_2})^2 \right).$$

– **Proportions :** on suppose  $(X_1, \dots, X_{n_1})$  distribué selon  $\mathcal{B}(p_1)$  et  $(Y_1, \dots, Y_{n_2})$  distribué selon  $\mathcal{B}(p_2)$ . On considère alors la statistique asymptotiquement normale

$$T = \frac{\bar{X}_{n_1} - \bar{Y}_{n_2}}{\sqrt{F(1-F)(1/n_1 + 1/n_2)}} \quad \text{où} \quad F = \frac{n_1 \bar{X}_{n_1} + n_2 \bar{Y}_{n_2}}{n_1 + n_2}.$$

### 11.3 Compléments classiques (pour votre culture)

Il existe de nombreux autres tests paramétriques. Le Chapitre XIII du livre de **Chabanol et Ruch** en est bien fourni.



# 12 TESTS D'AJUSTEMENT

Les tests d'ajustement, également appelés tests d'adéquation, permettent de mesurer l'ajustement entre les données observées et une loi de probabilité, ou un modèle statistique.

## 12.1 Test de Kolmogorov-Smirnov

### 12.1.1 Fonction de répartition empirique

Dans toute la suite,  $n$  désigne la taille de l'**échantillon** i.i.d.  $(X_1, \dots, X_n)$  de loi  $\mathbb{P}_X$  sur  $\mathbb{R}$ . On note  $F$  la fonction de répartition associée à  $\mathbb{P}_X$ . A partir de l'échantillon, on peut construire des objets **empiriques** à ne pas confondre avec les objets **théoriques** analogues.

**Définition.** On appelle **loi empirique** la probabilité discrète  $P_n := n^{-1} \sum_{i=1}^n \delta_{X_i}$ .

Par extension, on appelle **fonction de répartition (f.d.r.) empirique** la f.d.r.  $F_n$  associée à  $P_n$ .

Plus précisément, pour tout  $t \in \mathbb{R}$ ,

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{]-\infty, t]}(X_i).$$



Une simple application de la loi des grands nombres nous dit que pour tout  $t \in \mathbb{R}$ ,  $F_n(t) \rightarrow F(t)$  p.s. Les résultats qui suivent montrent que la convergence est **uniforme** en  $t$  et que l'on connaît la vitesse de convergence.

**Définition.** Pour une loi  $P^0$  de f.d.r. continue  $F^0$ , on définit la distance de Kolmogorov-Smirnov par

$$d_{KS}(F_n, F^0) := \|F_n - F^0\|_\infty.$$

Si on note  $(X_{(1)}, \dots, X_{(n)})$  la statistique d'ordre de l'échantillon, alors



$$d_{KS}(F_n, F^0) = \max_{i=1, \dots, n} \max \left\{ F^0(X_{(i)}) - \frac{i-1}{n}, \frac{i}{n} - F^0(X_{(i)}) \right\}.$$

### 12.1.2 Théorèmes de convergence

**Théorème de Glivenko-Cantelli.** Si  $F$  est continue, alors

$$d_{KS}(F_n, F) \xrightarrow{p.s.} 0.$$

**Proposition.** Soit  $F^0$  une f.d.r. telle que  $F^0 \neq F$ . Alors

$$\liminf d_{KS}(F_n, F^0) > 0 \quad p.s.$$

**Théorème de Kolmogorov-Smirnov.** Si  $F$  est continue, alors

$$D_n := \sqrt{n} d_{KS}(F_n, F) \xrightarrow{\mathcal{L}} P_{KS}.$$

La loi  $P_{KS}$  est appelée loi de K.-S. et sa f.d.r. est  $F_{KS}(t) := 1 + 2 \sum_{k=1}^{+\infty} (-1)^k e^{-2k^2 t^2}$  pour  $t > 0$ .

**Proposition.** La loi de  $D_n$  ne dépend pas de la loi sous-jacente  $\mathbb{P}_X$  (absolument continue). Cette loi est tabulée pour des petites valeurs de  $n$ .

### 12.1.3 Ajustement à une loi

Soit  $P^0$  une loi de f.d.r.  $F^0$  **continue**. La loi  $\mathbb{P}_X$  est supposée **inconnue** et on cherche à tester

$$\mathcal{H}_0 : \mathbb{P}_X = P^0 \quad \text{contre} \quad \mathcal{H}_1 : \mathbb{P}_X \neq P^0.$$

**Définition.** Le test de Kolmogorov-Smirnov de paramètre  $\alpha \in ]0, 1[$  est défini par :

- la statistique  $D_n^0 := \sqrt{n} d_{KS}(F_n, F^0)$ ,
- la zone de rejet  $R_\alpha := [d_{1-\alpha}, +\infty[$  où  $d_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  de la loi  $P_{KS}$ . Plus précisément,  $d_{1-\alpha}$  est la solution de  $F_{KS}(d_{1-\alpha}) = 1 - \alpha$ .

**Corollaire.** Le test de K.-S. de paramètre  $\alpha$  est un test **consistant** et de **taille asymptotique**  $\alpha$ .



Le test de K.-S. n'est utilisable que pour les lois **absolument continues** par rapport à Lebesgue. Pour les lois discrètes, on utilise souvent le **test du Khi-Deux** présenté ci-dessous. En pratique, le test de K.-S. est en général **plus puissant** que le test du Khi-Deux.



Si  $n$  est petit, on préfère prendre le quantile de la loi tabulée de  $D_n$ . Le test de K.-S. peut être adapté pour tester l'adéquation à une famille de loi. De plus, il peut être adapté pour tester l'homogénéité entre deux échantillons pour répondre à la question : est-ce que les deux échantillons proviennent de la même loi sous-jacente.

## 12.2 Tests du khi-deux

### 12.2.1 Ajustement à une loi

**Proposition.** Soit  $\pi = (\pi(1), \dots, \pi(k))$  une loi de probabilité sur  $\{1, \dots, k\}$  qui charge tous les points. Soit  $(X_i)_{i \geq 1}$  i.i.d de loi  $\pi$ . On note  $\hat{\pi}(j) = \frac{\sum_{i=1}^n \mathbf{1}_{X_i=j}}{n}$  l'estimateur empirique (et du max de vraisemblance) de  $\pi(j)$ . Alors,

- $\hat{\pi}$  est un **estimateur consistant** de  $\pi$ ,
- $\sqrt{n}(\hat{\pi} - \pi) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, \Gamma)$  avec  $\Gamma = \text{Diag}(\pi) - \pi^T \pi$ ,
- $D_n^2(\pi) := n \sum_{j=1}^k \frac{(\hat{\pi}(j) - \pi(j))^2}{\pi(j)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k-1)$ .

$\hat{\pi} - \pi$  appartient à un s.e.v.  $V$  de dimension  $(k-1)$  de  $\mathbb{R}^k$  : l'orthogonal du vecteur  $(1, 1, \dots, 1)$ . En prenant dans  $V$  le produit scalaire suivant :



$$\langle u, v \rangle_\pi := \sum_{j=1}^k \frac{u_j v_j}{\pi(j)}$$

on peut voir que  $\sqrt{n}(\hat{\pi} - \pi)$  converge en loi vers un vecteur gaussien standard dans  $(V, \langle \cdot, \cdot \rangle_\pi)$ , ce qui signifie, en fixant une b.o.n. de  $(V, \langle \cdot, \cdot \rangle_\pi)$ , la convergence du vecteur des coordonnées de  $\sqrt{n}(\hat{\pi} - \pi)$  vers un vecteur gaussien standard de  $\mathbb{R}^{k-1}$ .

Soit  $\pi_0$  une loi sur  $\{1, \dots, k\}$  qui charge tous les points. La loi  $\pi$  est supposée **inconnue** et on cherche à tester

$$\mathcal{H}_0 : \pi = \pi_0 \quad \text{contre} \quad \mathcal{H}_1 : \pi \neq \pi_0.$$

**Définition.** Le test du **khi-deux** (de la loi  $\pi_0$ ) de paramètre  $\alpha \in ]0, 1[$  est défini par :

- la statistique  $D_n^2(\pi_0)$  définie dans la proposition ci-dessus,
- la zone de rejet  $R_\alpha := [z_{1-\alpha}, +\infty[$  où  $z_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  de la loi  $\chi^2(k-1)$ .

**Corollaire.** Le test du khi-deux de paramètre  $\alpha$  est un test **consistant** et de **taille asymptotique**  $\alpha$ .

### 12.2.2 Ajustement à une famille de lois

On se donne un sous-ensemble strict  $\mathcal{F}$  de lois sur  $\{1, \dots, k\}$  et on cherche à tester

$$\mathcal{H}_0 : \pi \in \mathcal{F} \quad \text{contre} \quad \mathcal{H}_1 : \pi \notin \mathcal{F}.$$

On note  $\hat{\pi}_{\mathcal{F}}$  l'estimateur du maximum de vraisemblance de  $\pi$  dans le modèle  $\mathcal{F}$ . Le théorème suivant est vrai sous des hypothèses de régularité sur  $\mathcal{F}$  du type "variété lisse de dimension  $r$ ", qui sont détaillées dans le chapitre 6 du livre de Rivoirard et Stoltz.

**Théorème.** Sous certaines hypothèses de régularité, sous  $H_0$ ,

$$D_n^2(\hat{\pi}_{\mathcal{F}}) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k - 1 - r)$$

où  $r$  est la dimension de  $\mathcal{F}$ .

De plus, sous  $H_1$ ,  $D_n^2(\hat{\pi}_{\mathcal{F}})$  tend vers l'infini presque sûrement.

**Définition.** Le test du **khi-deux** (de la famille  $\mathcal{F}$ ) de paramètre  $\alpha \in ]0, 1[$  est défini par :

- la statistique  $D_n^2(\hat{\pi}_{\mathcal{F}})$  définie ci-dessus,
- la **zone de rejet**  $R_{\alpha} := [z_{1-\alpha}, +\infty[$  où  $z_{1-\alpha}$  est le quantile d'ordre  $1 - \alpha$  de la loi  $\chi^2(k - 1 - r)$ .



Par rapport au test d'ajustement à une loi donnée  $\pi_0$ , le test d'ajustement à une famille revient à remplacer, dans la statistique de test, la cible  $\pi_0$  par l'estimateur  $\hat{\pi}_{\mathcal{F}}$ . Ainsi, cette statistique fait intervenir deux estimateurs issus du même échantillon :  $\hat{\pi}$  et  $\hat{\pi}_{\mathcal{F}}$ . C'est pour compenser cela que le degré de liberté qui intervient dans le choix du quantile est modifié.

**Corollaire.** Le test du khi-deux de paramètre  $\alpha$  est un test **consistant** et de **taille asymptotique**  $\alpha$ .



Une application classique est le test du khi-deux d'indépendance : les données se présentent sous la forme de couples  $X = (Y, Z)$  à valeurs dans  $\{1, \dots, s\} \times \{1, \dots, t\}$  et on souhaite tester l'indépendance de  $Y$  et  $Z$ . Cela correspond à la famille  $\mathcal{F}$  des lois produits sur  $\{1, \dots, s\} \times \{1, \dots, t\}$ . La dimension de  $\mathcal{F}$  est  $r = (s - 1)(t - 1)$  et on a :

- $\hat{\pi}(j, k) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(Y_i, Z_i)=(j, k)}$ ,
- $\hat{\pi}_{\mathcal{F}}(j, k) = \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i=j} \right) \left( \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Z_i=k} \right)$ .



# Index

- biais, 42
- chaîne de Markov, 27
  - construction, 28
  - irréductible, 29
- chaîne de Markov
  - loi des grands nombres, 31
- classe
  - de communication, 29
  - fermée, 30
- consistance, 42, 52
- convergence
  - $L^p$ , 17
  - en loi, 17, 18
  - presque-sûre, 17
  - probabilité, 17
- corrélation, 16
- covariance, 16
- densité, 13
  - conditionnelle, 38
- échantillon, 41
- espace d'états, 28
- espérance, 15
  - conditionnelle, 37
- estimateur, 41, 51
- estimation, 41
- état
  - accessible, 29
  - apériodique, 31
  - communiquant, 29
  - période, 31
  - récurrent, 29
    - nul, 30
    - positif, 30
  - transitoire, 29
- filtration, 38
- fonction caractéristique, 16
- fonction de masse, 11, 13
- fonction de répartition, 14
  - empirique, 41, 55
- fonction génératrice, 16
- formule de Bayes, 12
- formule des probabilités totales, 12
- formule du crible (Poincaré), 12
- hypothèse
  - alternative, 51
  - nulle, 51
- indépendance, 12
  - tribus, 12
  - variables aléatoires, 14
- inégalité
  - Bienaymé-Tchebychev, 18
  - Markov, 18
- intervalle de confiance, 43
- inverse généralisé, 14, 43
- lemme
  - de Borel Cantelli, 13
  - de Slutsky, 18
- loi
  - binomiale, 21
  - conditionnelle, 38, 46
  - conjointe, 14
  - d'une variable aléatoire, 13
  - de Bernoulli, 21, 52
  - de Kolmogorov-Smirnov, 55
  - de Poisson, 23, 33
  - de Student, 25, 43, 52
  - du khi-deux, 25, 43, 45, 47, 56
  - empirique, 41, 55
  - exponentielle, 24, 34
  - gama, 25
  - gamma, 34
  - géométrique, 22
  - hypergéométrique, 22
  - marginale, 14
  - multinomiale, 22
  - normale, 18, 24, 43, 45, 52
  - sans mémoire, 22, 24
  - uniforme, 21, 23, 34
- loi des événements rares, 23
- loi des grands nombres, 18, 35
- loi du 0-1 de Kolmogorov, 13

marche aléatoire simple, 31  
 martingale, 39  
     arrêtée, 39  
     sous-, 39  
 matrice de transition, 27  
 matrice stochastique, 27  
 maximum de vraisemblance, 42  
 mesure  
     de probabilité, 12  
     probabilité, 11  
     réversible, 30  
     stationnaire, 30  
 modèle  
     paramétrique, 41  
     paramétré, 41  
     statistique, 41  
 modèle linéaire gaussien, 47  
 modélisation  
     durée de vie, 24, 25  
     erreur de mesure, 24  
     sondage, 21, 22  
     temps d'attente, 24, 25  
 moindres carrés, 47  
 moment, 15  
     empirique, 41  
 noyau de transition, 38  
 p-valeur, 52  
 probabilité conditionnelle, 12, 37  
 probabilité de transition, 27  
 processus de Poisson, 33  
     simulation, 34  
 processus stochastique, 38  
     adapté, 38  
 propriété de Markov, 28  
 quantile, 14, 43  
 risque d'un test, 51  
 risque quadratique, 42  
 règle de décision, 51  
 région de confiance, 43  
 régression linéaire, 48  
 simulation  
     couple de v.a., 26  
     méthode d'inversion, 26  
     méthode du rejet, 26  
     vecteur gaussien, 45  
 sondage, 52  
 statistique d'ordre, 34, 55  
 statistique de test, 51  
 temps d'arrêt, 39  
 test statistique  
     définition, 51  
     khi-deux, 56  
     Kolmogorov-Smirnov, 56  
     niveau, 52  
     puissance, 52  
     taille, 52  
 théorème  
     central limite, 18, 35  
     vectoriel, 45  
     d'arrêt (martingales), 39  
     de Cochran, 45  
     de Glivenko-Cantelli, 55  
     de Kolmogorov-Smirnov, 55  
     de Lévy, 18  
     de transfert, 15  
     ergodique, 31  
 transformée de Fourier, 16  
 tribu, 11  
     (et temps d'arrêt), 39  
     borélienne, 12  
     engendrée, 11, 14  
     produit, 12  
     terminale, 13  
 variable aléatoire, 13  
 variance, 15  
 vecteur gaussien, 45  
 zone de rejet, 51