Démonstrations des théorèmes folkloriques sur la dérivée de Radon-Nikodym

Yaiza BERMUDEZ[1] Gaetan BISSON[2] Iñaki ESNAOLA[3 4] Samir PERLAZA[1 2 4]

[1]INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France.

[2]Laboratoire GAATI, Université de la Polynésie française, Fa'a'ā, Polynésie française.

[3]School of Electrical and Electronic Engineering, University of Sheffield, Sheffield, Royaume-Uni.

[4]Department of Electrical and Computer Engineering, Princeton University, Princeton, 08544 NJ, USA.

Résumé – Des énoncés rigoureux et des démonstrations formelles sont présentés pour les théorèmes folkloriques, à la fois fondamentaux et avancés, concernant la dérivée de Radon-Nikodym, qui est un outil mathématique fondamental pour la formulation des mesures d'information.

Abstract – Rigorous statements and formal proofs are presented for both foundational and advanced folklore theorems on the Radon-Nikodym derivative, which is a fundamental mathematical tool for the formulation of information measures.

1 Introduction

En mathématiques, les « théorèmes folkloriques » désignent des résultats couramment admis et utilisés par les spécialistes, mais rarement formalisés ou attribués. En théorie des jeux, par exemple, le théorème folklorique originel, connu depuis les années 1950, reste non publié et sans auteur identifié (voir [7], [6]). En théorie de l'information, nombre d'entre eux portent sur la dérivée de Radon–Nikodym (DRN), introduite par Radon [13] puis généralisée par Nikodym [10]. Toutes les mesures d'information de Shannon s'expriment via la DRN, et ses propriétés sont souvent énoncées sans preuve dans les manuels classiques (cf. [2, 3, 8]), à l'exception de quelques démonstrations dans [5]. Shannon lui-même n'a pas fait appel à la DRN dans ses travaux fondateurs [**?**, 15], préférant traiter séparément fonctions de masse et densités de probabilité. Ce choix a influencé la présentation des manuels ultérieurs [1, 4, 9], avant que la DRN ne s'impose dans les ouvrages modernes [12] et dans la définition de nouvelles mesures (lautum information [11]), assurant une présentation unifiée quelle que soit la mesure de référence (comptage, Lebesgue, etc.) [14]

2 Préliminaires

Cette section introduit les conventions de notation, ainsi que le théorème de Radon-Nikodym. En particulier, certaines égalités présentées dans cet article sont valables presque sûrement par rapport à une mesure donnée. Pour plus de clarté, étant donné un espace mesuré $(\Omega, \mathscr{F}, P)$, la notation $\overset{p.s.}{\underset{P}{=}}$ est introduite et se lit « égal pour tout $x \in \Omega$ sauf sur un ensemble négligeable au sens de P », ou de manière équivalente « égal presque sûrement par rapport à P ». De plus, étant données deux mesures P et Q définies sur le même espace mesurable, la notation $P \ll Q$ signifie que « la mesure P est absolument continue par rapport à Q ». À l'aide de cette notation, la dérivée de Radon-Nikodym est introduite par le théorème suivant.

Theorem 1 (Théorème de Radon-Nikodym, [2, Théorème 2.2.1])**.** *Soient P et Q deux mesures définies sur un espace mesurable $(\Omega, \mathscr{F})$, telles que Q soit σ-finie et $P \ll Q$. Alors, il existe une fonction mesurable borélienne, positive, $g : \Omega \to \mathbb{R}$ telle que, pour tout $\mathcal{A} \in \mathscr{F}$,*

$$P(\mathcal{A}) = \int_{\mathcal{A}} g(x) \mathrm{d}Q(x). \tag{1}$$

De plus, si une autre fonction h satisfait, pour tout $\mathcal{A} \in \mathscr{F}$, que $P(\mathcal{A}) = \int_{\mathcal{A}} h(x) \mathrm{d}Q(x)$, alors $g(x) \overset{p.s.}{\underset{Q}{=}} h(x)$.

La fonction g dans (1) est souvent appelée la dérivée de Radon-Nikodym de P par rapport à Q, et s'écrit également $\frac{\mathrm{d}P}{\mathrm{d}Q}$, de sorte que $g(x) = \frac{\mathrm{d}P}{\mathrm{d}Q}(x)$.

Le théorème de Radon–Nikodym constitue l'outil fondamental à partir duquel prennent source nombre de résultats classiques en théorie de l'information [15]. Les sections suivantes en examinent plusieurs ; les résultats folkloriques reçoivent une démonstration rigoureuse, tandis que les énoncés déjà établis sont rappelés, parfois avec des preuves plus concises.

3 Théorèmes classiques fondamentaux

Cette section se concentre sur des théorèmes classiques fondamentaux, où « fondamentaux » désigne leur nature bien établie. L'un des théorèmes les plus courants est souvent appelé le théorème du « changement de mesure ».

Theorem 2 (Changement de mesure)**.** *Soient P et Q deux mesures sur l'espace mesurable $(\Omega, \mathscr{F})$ avec $P \ll Q$; et Q une mesure σ-finie. Soit $f : \Omega \to \mathbb{R}$ une fonction mesurable borélienne telle que l'intégrale $\int_\Omega f(x) \mathrm{d}P(x)$ existe. Alors, pour tout $\mathcal{A} \in \mathscr{F}$,*

$$\int_{\mathcal{A}} f(x) \mathrm{d}P(x) = \int_{\mathcal{A}} f(x) \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \mathrm{d}Q(x). \tag{2}$$

Une démonstration de ce théorème est donnée dans [5, Proposition 3.9]. Par souci de complétude, une preuve alternative est proposée ci-dessous, en utilisant la notation précédemment introduite.

Proof: La première partie de la preuve est développée sous l'hypothèse que la fonction f est simple. Une fonction f est dite simple lorsqu'elle est de la forme $f(x) = \sum_{i=1}^{m} a_i \mathbb{1}_{\mathcal{A}_i}(x)$, pour un $m \in \mathbb{N}$ fini, des ensembles disjoints $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_m$ dans $\mathscr{F}$ et des réels $a_1, a_2, \ldots, a_m$, pour tout x.

Pour tout $\mathcal{A} \in \mathcal{F}$, et pour tout $i \in \{1, 2, \ldots, m\}$, soit $\mathcal{B}_i = \mathcal{A} \cap \mathcal{A}_i$, alors,

$$\int_{\mathcal{A}} f(x) \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \mathrm{d}Q(x) = \int_{\mathcal{A}} \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \sum_{i=1}^{n} a_i \mathbb{1}_{\mathcal{A}_i}(x) \mathrm{d}Q(x) \quad (3)$$

$$= \sum_{i=1}^{n} a_i \int_{\mathcal{B}_i} \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \mathrm{d}Q(x) = \sum_{i=1}^{n} a_i P(\mathcal{B}_i), \quad (4)$$

où l'égalité dans (4) provient de la linéarité de l'intégrale [2, Théorème 1.6.3] et du Théorème 1.

Par ailleurs, pour tout $\mathcal{A} \in \mathcal{F}$

$$\int_{\mathcal{A}} f(x) \, \mathrm{d}P(x) = \sum_{i=1}^{m} a_i \int_{\mathcal{B}_i} \mathrm{d}P(x) \; = \; \sum_{i=1}^{m} a_i P(\mathcal{B}_i), \quad (5)$$

où l'égalité dans (5) provient de la linéarité de l'intégrale [2, Théorème 1.6.3]. Ainsi, à partir du Théorème 1, il s'ensuit que lorsque f est une fonction simple, l'égalité (2) est vérifiée.

La preuve procède en considérant les observations suivantes : (a) les fonctions simples forment un sous-ensemble dense de l'espace des fonctions boréliennes mesurables [2, Théorème 1.5.5(b)] ; et (b) l'intégrale est une application continue sur cet espace [2, Théorème 1.6.2]. Par conséquent, de (a) et (b), il s'ensuit que (2) est aussi vérifiée pour toute fonction borélienne mesurable f. Cela conclut la preuve. ■

Le théorème folklorique des « mesures proportionnelles » explicite la dérivée de Radon–Nikodym entre deux mesures proportionnelles.

Theorem 3 (Mesures proportionnelles)**.** *Soient P et Q deux mesures σ-finies sur l'espace mesurable $(\Omega, \mathcal{F})$, telles que pour tout $\mathcal{A} \in \mathcal{F}$*

$$Q(\mathcal{A}) = cP(\mathcal{A}), \quad (6)$$

avec $c > 0$. Alors,

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(x) \overset{p.s.}{\underset{Q}{=}} \frac{1}{c}, \quad et \quad \frac{\mathrm{d}Q}{\mathrm{d}P}(x) \overset{p.s.}{\underset{P}{=}} c. \quad (7)$$

Proof: Premièrement, on note que (6) implique que les mesures P et Q sont mutuellement absolument continues. Par conséquent, d'après le Théorème 1, il s'ensuit que pour tout $\mathcal{A} \in \mathcal{F}$,

$$P(\mathcal{A}) = \int_{\mathcal{A}} \mathrm{d}P(x) = \int_{\mathcal{A}} \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \; \mathrm{d}Q(x). \quad (8)$$

et, d'après (6), il suit que $P(\mathcal{A}) = \frac{1}{c} Q(\mathcal{A}) = \int_{\mathcal{A}} \frac{1}{c} \mathrm{d}Q(x)$. Ainsi, $\frac{\mathrm{d}P}{\mathrm{d}Q}(x) \overset{p.s.}{\underset{Q}{=}} \frac{1}{c}$ découle directement du Théorème 1. Un raisonnement similaire montre que $\frac{\mathrm{d}Q}{\mathrm{d}P}(x) \overset{p.s.}{\underset{P}{=}} c$. ■

Dans le cas particulier où $c = 1$ dans (7), P et Q coïncident, ce qui conduit à $\frac{\mathrm{d}P}{\mathrm{d}Q}(x) \overset{p.s.}{\underset{Q}{=}}$ et $\frac{\mathrm{d}Q}{\mathrm{d}P}(x) \overset{p.s.}{\underset{P}{=}} 1$. Le théorème folklorique suivant, souvent désigné sous le nom de « règle de la chaîne », fait l'objet d'une démonstration dans [5, Proposition 3.9].

Theorem 4 (Règle de chaîne)**.** *Soient P, Q et R trois mesures sur l'espace mesurable $(\Omega, \mathcal{F})$ telles que $P \ll Q$; $Q \ll R$; et Q et R sont des mesures σ-finies. Alors,*

$$\frac{\mathrm{d}P}{\mathrm{d}R}(x) \overset{p.s.}{\underset{R}{=}} \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \frac{\mathrm{d}Q}{\mathrm{d}R}(x). \quad (9)$$

Le théorème folklorique suivant découle du Théorème 4 et met en évidence le lien entre la dérivée de Radon–Nikodym et son inverse multiplicatif. On retrouve également ce résultat dans [5, Corollary 3.10].

Theorem 5 (Inverse multiplicatif)**.** *Soient P et Q deux mesures mutuellement absolument continues sur l'espace mesurable $(\Omega, \mathcal{F})$; et supposons que pour tout $x \in \Omega$, on ait $\frac{\mathrm{d}Q}{\mathrm{d}P}(x) > 0$. Alors,*

$$\frac{\mathrm{d}P}{\mathrm{d}Q}(x) \overset{p.s.}{\underset{Q}{=}} \left(\frac{\mathrm{d}Q}{\mathrm{d}P}(x) \right)^{-1}. \quad (10)$$

Le théorème folklorique suivant établit la linéarité de la dérivée de Radon–Nikodym. Un énoncé similaire figure dans [5, Proposition 3.11], mais sans démonstration.

Theorem 6 (Linéarité)**.** *Soit P une mesure σ-finie sur $(\Omega, \mathcal{F})$ et soient aussi $Q_1, Q_2, \ldots, Q_n$ des mesures finies sur $(\Omega, \mathcal{F})$ absolument continues par rapport à P. Soient $c_1, c_2, \ldots, c_n$ des réels positifs ; et soit S une mesure finie sur $(\Omega, \mathcal{F})$ telle que pour tout $\mathcal{A} \in \mathcal{F}$, $S(\mathcal{A}) = \sum_{t=1}^{n} c_t Q_t(\mathcal{A})$. Alors,*

$$\frac{\mathrm{d}S}{\mathrm{d}P}(x) \overset{p.s.}{\underset{P}{=}} \sum_{t=1}^{n} c_t \frac{\mathrm{d}Q_t}{\mathrm{d}P}(x). \quad (11)$$

Proof: D'après les hypothèses du théorème, on a $S \ll P$; ainsi, pour tout $\mathcal{A} \in \mathcal{F}$,

$$\int_{\mathcal{A}} \frac{\mathrm{d}S}{\mathrm{d}P}(x) \mathrm{d}P(x) = \int_{\mathcal{A}} \mathrm{d}S(x) = \sum_{t=1}^{n} \int_{\mathcal{A}} c_t \mathrm{d}Q_t(x) \quad (12)$$

$$= \sum_{t=1}^{n} \int_{\mathcal{A}} c_t \frac{\mathrm{d}Q_t}{\mathrm{d}P}(x) \mathrm{d}P(x) = \int_{\mathcal{A}} \sum_{t=1}^{n} c_t \frac{\mathrm{d}Q_t}{\mathrm{d}P}(x) \mathrm{d}P(x), \quad (13)$$

où la première égalité de (12) et la première égalité de (13) découlent du Théorème 2, tandis que la dernière égalité de (13) résulte de l'additivité de l'intégrale [2, Corollaire 1.6.4]. La preuve se conclut en utilisant le Théorème 1, qui implique l'égalité dans (11). ■

Le théorème folklorique suivant établit la continuité de la dérivée de Radon–Nikodym.

Theorem 7 (Continuité)**.** *Soit P une mesure σ-finie sur $(\Omega, \mathcal{F})$, et soit $Q_1, Q_2, \cdots$ une suite infinie de mesures σ-finies sur $(\Omega, \mathcal{F})$, convergeant vers une mesure Q. Supposons que pour tout $n \in \mathbb{N}$, $Q_n \ll P$. Alors, $Q \ll P$ et*

$$\lim_{n \to \infty} \frac{\mathrm{d}Q_n}{\mathrm{d}P}(x) \overset{p.s.}{\underset{P}{=}} \frac{\mathrm{d}Q}{\mathrm{d}P}(x). \quad (14)$$

Proof: D'après les hypothèses du théorème et le Théorème 2, pour tout $\mathcal{A} \in \mathcal{F}$, on obtient

$$Q(\mathcal{A}) = \lim_{n \to \infty} Q_n(\mathcal{A}) \; = \; \int_{\mathcal{A}} \lim_{n \to \infty} \frac{\mathrm{d}Q_n}{\mathrm{d}P}(x) \, \mathrm{d}P(x), \quad (15)$$

où la dernière égalité de (15) découle du Théorème 2 et de [2, Theorem1.6.2]. On en déduit que $Q \ll P$ et, pour tout $\mathcal{A} \in \mathcal{F}$,

$$Q(\mathcal{A}) \; = \; \int_{\mathcal{A}} \frac{\mathrm{d}Q}{\mathrm{d}P}(x) \, \mathrm{d}P(x). \quad (16)$$

L'égalité (16), combinée au Théorème1, conduit alors à (14). ■

Le théorème folklorique suivant relie la dérivée de Radon–Nikodym d'une mesure produit à celles de ses mesures composantes. Ce résultat figure également dans [5, Exercise 3.12].

Theorem 8 (Produit de mesures)**.** *Pour tout $i \in \{1, 2\}$, soient P_i et Q_i une mesure finie et une mesure σ-finie respectivement sur $(\Omega_i, \mathcal{F}_i)$; avec $P_i \ll Q_i$. Soient aussi $P_1 P_2$ et $Q_1 Q_2$ les mesures produits sur $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ formées respectivement par P_1 et P_2 ; et Q_1 et Q_2. Alors,*

$$\frac{\mathrm{d}P_1 P_2}{\mathrm{d}Q_1 Q_2}(x_1, x_2) \overset{p.s.}{\underset{Q_1 Q_2}{=}} \frac{\mathrm{d}P_1}{\mathrm{d}Q_1}(x_1) \frac{\mathrm{d}P_2}{\mathrm{d}Q_2}(x_2). \quad (17)$$

Proof: D'après les hypothèses du théorème, pour tout $\mathcal{A} \in (\Omega_1 \times \Omega_2)$,

$$P_1P_2(\mathcal{A}) = \int_{\mathcal{A}} \mathrm{d}P_1P_2(x_1, x_2) = \iint_{\mathcal{A}_{x_2}} \mathrm{d}P_1(x_1)\,\mathrm{d}P_2(x_2) \quad (18)$$

$$= \iint_{\mathcal{A}_{x_2}} \frac{\mathrm{d}P_1(x_1)}{\mathrm{d}Q_1} \mathrm{d}Q_1(x_1)\,\mathrm{d}P_2(x_2) \quad (19)$$

$$= \iint_{\mathcal{A}_{x_2}} \frac{\mathrm{d}P_1(x_1)}{\mathrm{d}Q_1} \frac{\mathrm{d}P_2(x_2)}{\mathrm{d}Q_2} \mathrm{d}Q_1(x_1)\,\mathrm{d}Q_2(x_2) \quad (20)$$

$$= \int_{\mathcal{A}} \frac{\mathrm{d}P_1}{\mathrm{d}Q_1}(x_1) \frac{\mathrm{d}P_2}{\mathrm{d}Q_2}(x_2)\,\mathrm{d}Q_1Q_2(x_1, x_2), \quad (21)$$

où $\mathcal{A}x_2$ désigne la section de $\mathcal{A}$ déterminée par x_2, c'est-à-dire $\mathcal{A}x_2 = x_1 : (x_1, x_2) \in \mathcal{A}$; l'égalité de (18) découle de la définition de P_1P_2 comme produit de P_1 et P_2 ; les égalités de (19) et (20) résultent directement du Théorème 2 ; enfin, l'égalité de (21) provient de la construction de Q_1Q_2 en tant que mesure produit de Q_1 et Q_2. D'après (21), on a $P_1P_2 \ll Q_1Q_2$. Par conséquent, pour tout $\mathcal{A} \in \mathcal{F}1 \times \mathcal{F}2$,

$$P1P2(\mathcal{A}) = \int_{\mathcal{A}} \frac{\mathrm{d}P_1P_2}{\mathrm{d}Q_1Q_2}(x_1, x_2)\mathrm{d}Q_1Q_2(x_1, x_2). \quad (22)$$

Le Théorème 1 implique alors l'égalité (17), ce qui conclut la démonstration. ∎

4 Théorèmes classiques avancés

Cette section requiert une notation supplémentaire. En particulier, on note $\triangle(\mathcal{X}, \mathcal{F}_\mathcal{X})$, ou simplement $\triangle(\mathcal{X})$, l'ensemble de toutes les mesures de probabilité sur l'espace mesurable $(\mathcal{X}, \mathcal{F}_\mathcal{X})$, où $\mathcal{F}_\mathcal{X}$ est une σ-algèbre sur $\mathcal{X}$. En utilisant cette notation, les mesures de probabilité conditionnelles peuvent être définies comme suit.

Definition 1 (Probabilité conditionnelle)**.** *Une famille $P_{Y|X} \triangleq (P_{Y|X=x})_{x\in\mathcal{X}}$ d'éléments de $\triangle(\mathcal{Y}, \mathcal{F}_\mathcal{Y})$ indexée par $\mathcal{X}$ est dite mesure de probabilité conditionnelle si, pour tout ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{Y}$, l'application*

$$\begin{cases} \mathcal{X} \longrightarrow [0,1] \\ x \longmapsto P_{Y|X=x}(\mathcal{A}) \end{cases}$$

est borélienne. L'ensemble de telles mesures de probabilité conditionnelles est noté $\triangle(\mathcal{Y}|\mathcal{X})$.

Une mesure de probabilité conditionnelle $P_{Y|X} \in \triangle(\mathcal{Y}|\mathcal{X})$ et une mesure de probabilité $P_X \in \triangle(\mathcal{X})$ définissent deux mesures de probabilité uniques dans $\triangle(\mathcal{X} \times \mathcal{Y})$ et $\triangle(\mathcal{Y} \times \mathcal{X})$, respectivement. Ces mesures sont notées P_{XY} et P_{YX}, respectivement, et pour tout ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{X} \times \mathcal{F}_\mathcal{Y}$, on a

$$P_{XY}(\mathcal{A}) = \int P_{Y|X=x}(\mathcal{A}_x)\,\mathrm{d}P_X(x), \quad (23)$$

où $\mathcal{A}_x$ est la section de l'ensemble $\mathcal{A}$ déterminée par x, à savoir,

$$\mathcal{A}_x \triangleq \{y \in \mathcal{Y} : (x, y) \in \mathcal{A}\}. \quad (24)$$

De manière équivalente, pour tout ensemble $\mathcal{B} \in \mathcal{F}_\mathcal{Y} \times \mathcal{F}_\mathcal{X}$, on a

$$P_{YX}(\mathcal{B}) = \int P_{Y|X=x}(\mathcal{B}_x)\,\mathrm{d}P_X(x), \quad (25)$$

où $\mathcal{B}_x$ est la section de l'ensemble $\mathcal{B}$ déterminée par x. Notez que P_{XY} est une mesure sur $\triangle(\mathcal{X} \times \mathcal{Y})$, tandis que P_{YX} est une mesure sur $\triangle(\mathcal{Y} \times \mathcal{X})$. Par conséquent, obtenir P_{XY} à partir de P_{YX} ne consiste pas simplement à permuter X et Y ; il faut également redéfinir correctement les espaces mesurables correspondants. Plus précisément, étant donné un ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{X} \times \mathcal{F}_\mathcal{Y}$, on définit l'ensemble $\widehat{\mathcal{A}} \in \mathcal{F}_\mathcal{Y} \times \mathcal{F}_\mathcal{X}$ tel que

Pour tout ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{X} \times \mathcal{F}_\mathcal{Y}$, on note l'ensemble $\widehat{\mathcal{A}} \in \mathcal{F}_\mathcal{Y} \times \mathcal{F}_\mathcal{X}$ tel que

$$\widehat{\mathcal{A}} = \{(y, x) \in \mathcal{Y} \times \mathcal{X} : (x, y) \in \mathcal{A}\}. \quad (26)$$

Alors, d'après (23) et (25), on a

$$P_{XY}(\mathcal{A}) = P_{YX}\left(\widehat{\mathcal{A}}\right). \quad (27)$$

En utilisant cette notation, la notion de mesures marginales peut être introduite comme suit.

Definition 2 (Marginales)**.** *Étant données deux mesures de probabilité conjointes $P_{XY} \in \triangle(\mathcal{X} \times \mathcal{Y})$ et $P_{YX} \in \triangle(\mathcal{Y} \times \mathcal{X})$ satisfaisant* (27)*, les mesures marginales sur $\triangle(\mathcal{X})$ et $\triangle(\mathcal{Y})$, respectivement notées P_X et P_Y, satisfont pour tout ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{X}$ et pour tout ensemble $\mathcal{B} \in \mathcal{F}_\mathcal{Y}$,*

$$P_X(\mathcal{A}) \triangleq P_{XY}(\mathcal{A} \times \mathcal{Y}) = P_{YX}(\mathcal{Y} \times \mathcal{A}); \quad (28)$$

$$P_Y(\mathcal{B}) \triangleq P_{YX}(\mathcal{B} \times \mathcal{X}) = P_{XY}(\mathcal{X} \times \mathcal{B}). \quad (29)$$

D'après le théorème de la probabilité totale [2, Theorem 4.5.2], on a, pour tout $\mathcal{B} \in \mathcal{F}_\mathcal{Y}$ et pour tout $\mathcal{A} \in \mathcal{F}_\mathcal{X}$,

$$P_Y(\mathcal{B}) = \iint_{\mathcal{B}} \mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x); \quad (30)$$

$$P_X(\mathcal{A}) = \iint_{\mathcal{A}} \mathrm{d}P_{X|Y=y}(x)\mathrm{d}P_Y(y). \quad (31)$$

Les mesures de probabilité conjointes P_{XY} et P_{YX} peuvent être décrites via une mesure de probabilité conditionnelle $P_{Y|X} \in \triangle(\mathcal{Y}|\mathcal{X})$ et la mesure de probabilité P_X comme dans (23) et (25) ; ou via une mesure de probabilité conditionnelle $P_{X|Y} \in \triangle(\mathcal{X}|\mathcal{Y})$ et la mesure marginale $P_Y \in \triangle(\mathcal{Y})$.

Plus précisément, pour tout ensemble $\mathcal{A} \in \mathcal{F}_\mathcal{X} \times \mathcal{F}_\mathcal{Y}$, on a

$$P_{XY}(\mathcal{A}) = \int P_{X|Y=y}(\mathcal{A}_y)\,\mathrm{d}P_Y(y), \quad (32)$$

où $\mathcal{A}_y$ est la section de l'ensemble $\mathcal{A}$ déterminée par y.

Dans ce contexte, le théorème suivant met en évidence une propriété des mesures conditionnelles, rappelant l'axiome de la mesure unité en théorie des probabilités.

Theorem 9 (Mesure unité)**.** *Considérons les mesures de probabilité conditionnelles $P_{Y|X} \in \triangle(\mathcal{Y}|\mathcal{X})$ et $P_{X|Y} \in \triangle(\mathcal{X}|\mathcal{Y})$; ainsi que les mesures de probabilité $P_Y \in \triangle(\mathcal{Y})$ et $P_X \in \triangle(\mathcal{X})$ qui satisfont à* (30) *et* (31)*. Supposons que pour tout $x \in \mathcal{X}$, la mesure de probabilité $P_{Y|X=x} \ll P_Y$. Alors,*

$$\int \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_X(x) \overset{p.s.}{\underset{P_Y}{=}} 1. \quad (33)$$

Proof: Pour tout $\mathcal{A} \in \mathcal{F}_\mathcal{Y}$, d'après (30), on a :

$$P_Y(\mathcal{A}) = \iint_{\mathcal{A}} \mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x) \quad (34)$$

$$= \iint_{\mathcal{A}} \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_Y(y)\mathrm{d}P_X(x) \quad (35)$$

$$= \int_{\mathcal{A}} \int \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_X(x)\mathrm{d}P_Y(y), \quad (36)$$

L'égalité (35) découle du changement de mesure (Théorème 2) ; tandis que (36) s'obtient par le théorème de Fubini [2, Theorem 2.6.6]. Puisque $P_Y(\mathcal{A}) = \int_{\mathcal{A}} \mathrm{d}P_Y(y)$, on déduit de (36) et du Théorème 1 que

$$\int \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_X(x) \overset{p.s.}{\underset{P_Y}{=}} 1. \quad (37)$$

∎

Le théorème classique suivant est une version généralisée de la règle de Bayes.

Theorem 10 (Règle de type Bayes)**.** *Considérons les mesures de probabilité conditionnelles $P_{Y|X}$ et $P_{X|Y}$; les mesures de probabilité P_Y et P_X satisfaisant* (30) *et* (31)*; ainsi que les mesures de probabilité conjointes P_{YX} et P_{XY} de* (25) *et* (32) *respectivement. Soient $P_X P_Y \in \triangle(\mathcal{X}\times\mathcal{Y}, \mathscr{F}_{\mathcal{X}}\times\mathscr{F}_{\mathcal{Y}})$ et $P_Y P_X \in \triangle(\mathcal{Y}\times\mathcal{X}, \mathscr{F}_{\mathcal{Y}}\times\mathscr{F}_{\mathcal{X}})$ les mesures produits des marginales P_X et P_Y.*

Supposons que : (a) Pour tout $x \in \mathcal{X}$, $P_{Y|X=x} \ll P_Y$;

(b) Pour tout $y \in \mathcal{Y}$, $P_{X|Y=y} \ll P_X$.

Alors,

$$\frac{\mathrm{d}P_{XY}}{\mathrm{d}P_X P_Y}(x,y) \overset{p.s.}{\underset{P_X P_Y}{=}} \frac{\mathrm{d}P_{X|Y=y}}{\mathrm{d}P_X}(x) \tag{38}$$

$$\overset{p.s.}{\underset{P_X P_Y}{=}} \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y) \overset{p.s.}{\underset{P_X P_Y}{=}} \frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,. \tag{39}$$

Proof: Notons que les hypothèses (a) et (b) sont suffisantes pour garantir l'existence des dérivées de Radon–Nikodym de P_{XY} par rapport à $P_X P_Y$ et de P_{YX} par rapport à $P_Y P_X$. Par conséquent, pour tout ensemble $\mathcal{A} \in \mathscr{F}_{\mathcal{X}}\times\mathscr{F}_{\mathcal{Y}}$, on a

$$P_{XY}(\mathcal{A}) = \int_{\mathcal{A}} \frac{\mathrm{d}P_{XY}}{\mathrm{d}P_X P_Y}(x,y)\,\mathrm{d}P_X P_Y(x,y), \tag{40}$$

ce qui découle du Théorème 2.

Notons également que, d'après (32), il vient

$$\begin{aligned}
P_{XY}(\mathcal{A}) &= \iint_{\mathcal{A}_y} \mathrm{d}P_{X|Y=y}(x)\,\mathrm{d}P_Y(y) &(41)\\
&= \iint_{\mathcal{A}_y} \frac{\mathrm{d}P_{X|Y=y}}{\mathrm{d}P_X}(x)\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y) &(42)\\
&= \int \mathbb{1}_{\mathcal{A}}(x,y)\frac{\mathrm{d}P_{X|Y=y}}{\mathrm{d}P_X}(x)\,\mathrm{d}P_X P_Y(x,y) &(43)\\
&= \int_{\mathcal{A}} \frac{\mathrm{d}P_{X|Y=y}}{\mathrm{d}P_X}(x)\,\mathrm{d}P_X P_Y(x,y), &(44)
\end{aligned}$$

où l'ensemble $\mathcal{A}_y$ est la section de . De plus, l'égalité (42) découle de l'hypothèse (b) et du Théorème 1.

De même, d'après (23), on obtient

$$\begin{aligned}
P_{XY}(\mathcal{A}) &= \iint_{\mathcal{A}_x} \mathrm{d}P_{Y|X=x}(y)\,\mathrm{d}P_X(x) &(45)\\
&= \iint_{\mathcal{A}_x} \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\,\mathrm{d}P_Y(y)\,\mathrm{d}P_X(x) &(46)\\
&= \iint \mathbb{1}_{\mathcal{A}_y}(x)\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y) &(47)\\
&= \int \mathbb{1}_{\mathcal{A}}(x,y)\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\,\mathrm{d}P_X P_Y(x,y) &(48)\\
&= \int_{\mathcal{A}} \frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\,\mathrm{d}P_X P_Y(x,y), &(49)
\end{aligned}$$

où $\mathcal{A}_x$ est la section de l'ensemble $\mathcal{A}$ déterminée par x. L'égalité (46) découle de l'hypothèse (a) et du Théorème 1, tandis que l'égalité (47) provient de l'échange de l'ordre d'intégration [2, Theorem 2.6.6].

Enfin, d'après (27), on a

$$\begin{aligned}
&P_{XY}(\mathcal{A}) = \int_{\widehat{\mathcal{A}}} \mathrm{d}P_{YX}(y,x)\\
&= \int_{\widehat{\mathcal{A}}} \frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,\mathrm{d}P_Y P_X(y,x) &(50)\\
&= \int \mathbb{1}_{\widehat{\mathcal{A}}}(y,x)\frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,\mathrm{d}P_Y P_X(y,x) &(51)\\
&= \iint \mathbb{1}_{\widehat{\mathcal{A}}_y}(x)\frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,\mathrm{d}P_X(x)\,\mathrm{d}P_Y(y) &(52)\\
&= \int \mathbb{1}_{\widehat{\mathcal{A}}}(x,y)\frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,\mathrm{d}P_X P_Y(x,y) &(53)\\
&= \int_{\mathcal{A}} \frac{\mathrm{d}P_{YX}}{\mathrm{d}P_Y P_X}(y,x)\,\mathrm{d}P_X P_Y(x,y), &(54)
\end{aligned}$$

où l'ensemble $\widehat{\mathcal{A}}$ est défini en (26). L'égalité (50) est obtenue par un changement de mesure via le Théorème 2 sous l'hypothèse (a), et l'égalité (52) résulte de l'échange de l'ordre d'intégration [2, Theorem 2.6.6].

La démonstration s'achève grâce au Théorème 1 et en combinant les équations (40), (44), (49) et (54), lesquelles établissent (38) et (39). ■

Le théorème suivant est une version inverse de la règle de type Bayes.

Theorem 11 (Règle de type Bayes inverse)**.** *Considérons les mesures de probabilité conditionnelles $P_{Y|X}$ et $P_{X|Y}$; ainsi que les mesures de probabilité P_Y et P_X satisfaisant* (30) *et* (31)*; et les mesures de probabilité conjointes P_{YX} et P_{XY} en* (25) *et* (32)*, respectivement.*

Supposons que : (a)Pour tout $x \in \mathcal{X}$, $P_Y \ll P_{Y|X=x}$;

(b) Pour tout $y \in \mathcal{Y}$, $P_X \ll P_{X|Y=y}$.

Alors,

$$\begin{aligned}
&\frac{\mathrm{d}P_X P_Y}{\mathrm{d}P_{XY}}(x,y) \overset{p.s.}{\underset{P_{XY}}{=}} \frac{\mathrm{d}P_X}{\mathrm{d}P_{X|Y=y}}(x)\\
&\overset{p.s.}{\underset{P_{XY}}{=}} \frac{\mathrm{d}P_Y}{\mathrm{d}P_{Y|X=x}}(y) \overset{p.s.}{\underset{P_{YX}}{=}} \frac{\mathrm{d}P_Y P_X}{\mathrm{d}P_{YX}}(y,x)\,. &(55)
\end{aligned}$$

Proof: La démonstration suit les mêmes lignes que celle du Théorème 10. ■

Références

[1] Robert B ASH : *Information Theory*. Dover Publications, Mineola, NY, USA, first édition, 1990.

[2] Robert B. ASH et Catherine A. DOLEANS-DADE : *Probability and Measure Theory*. Academic Press, Burlington, MA, USA, second édition, 2000.

[3] Patrick BILLINGSLEY : *Probability and Measure*. John Wiley & Sons, Hoboken, NJ, USA, third édition, 2012.

[4] Thomas M. COVER et Joy A. THOMAS : *Elements of Information Theory*. Wiley-Interscience, Hoboken, NJ, second édition, 2006.

[5] Gerald B. FOLLAND : *Real Analysis : Modern Techniques and Their Applications*. John Wiley & Sons, New York, NY, USA, 2 édition, 1999.

[6] Drew FUDENBERG : *Game Theory*. MIT Press, Cambridge, MA, USA, first édition, 1991.

[7] Drew FUDENBERG et Eric MASKIN : The Folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, 54(3):533–554, 1986.

[8] Paul R HALMOS : *Measure Theory*. Van Nostrand, Princeton, NJ, USA, first édition, 1950.

[9] Solomon KULLBACK : *Information Theory and Statistics*. John Wiley & Sons, Inc., New York, NY, USA, first édition, 1959.

[10] Otton NIKODYM : Sur une généralisation des intégrales de mj radon. *Fundamenta Mathematicae*, 15(1):131–179, 1930.

[11] Daniel P PALOMAR et Sergio VERDÚ : Lautum information. *IEEE Transactions on Information Theory*, 54(3):964–975, Mar. 2008.

[12] Yury POLYANSKIY et Yihong WU : *Information Theory : From Coding to Learning*. Cambridge University Press, Cambridge, UK, first édition, 2024.

[13] Johann RADON : *Theorie und Anwendungen der absolut additiven Mengenfunktionen*. Hölder, Vienna, Austria, first édition, 1913.

[14] Igal SASON et Sergio VERDÚ : f -divergence inequalities. *IEEE Transactions on Information Theory*, 62(11):5973–6006, Jun. 2016.

[15] C. E. SHANNON : A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, Jul. 1948.

[16] C. E. SHANNON : A mathematical theory of communication. *The Bell System Technical Journal*, 27:623–656, Oct. 1948.