Article

Not peer-reviewed version

Variations on the Expectation Due to Changes in the Probability Measure

Samir M. Perlaza * and Gaetan Bisson

Posted Date: 25 July 2025

doi: 10.20944/preprints202507.2126.v1

Keywords: gibbs probability measures; sensitivity; distribution shifts; variations of the expectation

 doi:10.20944/preprints202507.2126.v1

Article

Variations on the Expectation Due to Changes in the Probability Measure

Samir M. Perlaza [1,2,3,*] **and Gaetan Bisson** [3]

[1] INRIA, Centre Inria d'Université Côte d'Azur, 06902 Sophia Antipolis, France.
[2] Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544 USA.
[3] GAATI Mathematics Laboratory, University of French Polynesia, 98702 Faaa, French Polynesia.
[*] Correspondence: samir.perlaza@inria.fr

Abstract

In this paper, closed-form expressions for the variation of the expectation of a given function due to changes in the probability measure (probability distribution drifts) are presented. They unveil interesting connections with Gibbs probability measures, mutual information, and lautum information.

Keywords: gibbs probability measures; sensitivity; distribution shifts; variations of the expectation

1. Introduction

Let m be a positive integer and denote by $\triangle(\mathbb{R}^m)$ the set of all probability measures on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$, with $\mathscr{B}(\mathbb{R}^m)$ being the Borel σ-algebra on $\mathbb{R}^m$. Given a Borel measurable function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, consider the functional

$$\begin{cases} \mathsf{G}_h : \mathbb{R}^n \times \triangle(\mathbb{R}^m) \times \triangle(\mathbb{R}^m) \longrightarrow \mathbb{R} \\ \qquad\qquad (x, P_1, P_2) \longmapsto \int h(x,y) \mathrm{d}P_1(y) - \int h(x,y) \mathrm{d}P_2(y), \end{cases} \tag{1}$$

which quantifies the variation of the expectation of the measurable function h due to changing the probability measure from P_2 to P_1. These variations are often referred to as *probability distribution drifts*, in some application areas, see for instance [1,2] and [3]. The functional G_h is defined when both integrals exist and are finite.

In order to define the expectation of $\mathsf{G}_h(x, P_1, P_2)$ when x is obtained by sampling a probability measure in $\triangle(\mathbb{R}^n)$, the structure formalized below is required.

Definition 1. *A family $P_{Y|X} \triangleq (P_{Y|X=x})_{x \in \mathbb{R}^n}$ of elements of $\triangle(\mathbb{R}^m)$ indexed by $\mathbb{R}^n$ is said to be a conditional probability measure if, for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$, the map*

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

is Borel measurable. The set of all such conditional probability measures is denoted by $\triangle(\mathbb{R}^m|\mathbb{R}^n)$.

In this setting, consider the functional

$$\bar{\mathsf{G}}_h : \begin{cases} \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^n) \longrightarrow \mathbb{R} \\ \qquad\qquad \left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) \longmapsto \int \mathsf{G}_h\left(x, P^{(1)}_{Y|X=x}, P^{(2)}_{Y|X=x}\right) \mathrm{d}P_X(x). \end{cases} \tag{2}$$

 doi:10.20944/preprints202507.2126.v1

This quantity can be interpreted as the variation of the integral (expectation) of the function h when the probability measure changes from the joint probability measure $P^{(1)}_{Y|X}P_X$ to another joint probability measure $P^{(2)}_{Y|X}P_X$, both in $\triangle(\mathbb{R}^m \times \mathbb{R}^n)$. This follows from (2) by observing that

$$\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) = \int h(x,y)\mathrm{d}P^{(1)}_{Y|X}P_X(y,x) - \int h(x,y)\mathrm{d}P^{(2)}_{Y|X}P_X(y,x). \tag{3}$$

Special attention is given to the quantity $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$, for some $P_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$, with P_Y being the marginal of the joint probability measure $P_{Y|X} \cdot P_X$. That is, for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$,

$$P_Y(\mathcal{A}) = \int P_{Y|X=x}(\mathcal{A})\mathrm{d}P_X(x). \tag{4}$$

Its relevance stems from the fact that it captures the variation of the expectation of the function h when the probability measure changes from the joint probability measure $P_{Y|X}P_X$ to the product of its marginals P_YP_X. That is,

$$\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right) = \int\left(\int h(x,y)\mathrm{d}P_Y(y) - \int h(x,y)\mathrm{d}P_{Y|X=x}(y)\right)\mathrm{d}P_X(x) \tag{5}$$

$$= \int h(x,y)\mathrm{d}P_YP_X(y,x) - \int h(x,y)\mathrm{d}P_{Y|X}P_X(y,x). \tag{6}$$

1.1. Novelty and Contributions

This work makes two key contributions: First, it provides a closed-form expression for the variation $\mathsf{G}_h(x, P_1, P_2)$ in (1) for a fixed $x \in \mathbb{R}^n$ and two arbitrary probability measures P_1 and P_2, formulated explicitly in terms of relative entropies. Second, it derives a closed-form expression for the expected variation $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ in (2), again in terms of information measures, for arbitrary conditional probability measures $P^{(1)}_{Y|X}$, $P^{(2)}_{Y|X}$, and an arbitrary probability measure P_X.

A further contribution of this work is the derivation of specific closed-form expressions for $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6), which reveal deep connections with both mutual information [4,5] and lautum information [6]. Notably, when $P_{Y|X}$ is a Gibbs conditional probability measure, this variation simplifies (up to a constant factor) to the sum of the mutual and lautum information induced by the joint distribution $P_{Y|X}P_X$.

Although these results were originally discovered in the analysis of generalization error of machine learning algorithms, see for instance [7–11], where the function h in (1) was assumed to represent an empirical risk, this paper presents such results in a comprehensive and general setting that is no longer tied to such assumptions. Also, strong connections with *information projections* and Pythagorean identities [12,13] are discussed. This new general presentation not only unifies previously scattered insights but also makes the results applicable across a broad range of domains in which changes in the expectation due to variations of the underlying probability measures are relevant.

1.2. Applications

The study of the variation of the integral (expectation) of h (for some fixed $x \in \mathbb{R}^n$) due to a measure change from P_2 to P_1, i.e., the value $\mathsf{G}_h(x, P_1, P_2)$ in (1), plays a central role in the definition of *integral probability metrics* (IPMs)[14,15]. Using the notation in (1), an IPM results from the optimization problem

$$\sup_{h\in\mathcal{H}} |\mathsf{G}_h(x, P_1, P_2)|, \tag{7}$$

for some fixed $x \in \mathbb{R}^n$ and a particular class of functions $\mathcal{H}$. Note for instance that the maximum mean discrepancy is an IPM [16], as well as the Wasserstein distance of order one [17–20].

 doi:10.20944/preprints202507.2126.v1

Other areas of mathematics in which the variation $\mathsf{G}_h(x, P_1, P_2)$ in (1) plays a key role is distributionally robust optimization (DRO) [21,22] and optimization with relative entropy regularization [8,9]. In these areas, the variation $\mathsf{G}_h(x, P_1, P_2)$ is a central tool. See for instance, [7,23]. Variations of the form $\mathsf{G}_h(x, P_1, P_2)$ in (1) have also been studied in [10] and [11] in the particular case of statistical machine learning for the analysis of generalization error. The central observation is that the generalization error of machine learning algorithms can be written in the form $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6). This observation is the main building block of the *method of gaps* introduced in [11], which leads to a number of closed-form expressions for the generalization error involving mutual information, lautum information, among other information measures.

2. Preliminaries

The main results presented in this work involve Gibbs conditional probability measures. Such measures are parametrized by a Borel measurable function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$; a σ-finite measure Q on $\mathbb{R}^m$; and a vector $x \in \mathbb{R}^n$. Note that the variable x will remain inactive until Section 4. Although it is introduced now for consistency, it could be removed altogether from all results presented in this section and Section 3.

Consider the following function:

$$\mathsf{K}_{h,Q,x} : \begin{cases} \mathbb{R} \longrightarrow \mathbb{R} \\ t \longmapsto \log(\int \exp(t\, h(x,y)) \mathrm{d}Q(y)). \end{cases} \tag{8}$$

Under the assumption that Q is a probability measure, the function $\mathsf{K}_{h,Q,x}$ in (8) is the cumulant generating function of the random variable $h(x, Y)$, for some fixed $x \in \mathbb{R}^n$ and $Y \sim Q$. Using this notation, the definition of the Gibbs conditional probability measure is presented hereunder.

Definition 2 (Gibbs Conditional Probability Measure). *Given a Borel measurable function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$; a σ-finite measure Q on $\mathbb{R}^m$; and a $\lambda \in \mathbb{R}$, the probability measure $P_{Y|X}^{(h,Q,\lambda)} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$ is said to be an (h, Q, λ)-Gibbs conditional probability measure if*

$$\forall x \in \mathcal{X}, \mathsf{K}_{h,Q,x}(-\lambda) < +\infty; \tag{9}$$

for some set $\mathcal{X} \subseteq \mathbb{R}^n$; and for all $(x, y) \in \mathcal{X} \times \operatorname{supp} Q$,

$$\frac{\mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}}{\mathrm{d}Q}(y) = \exp\big(-\lambda h(x,y) - \mathsf{K}_{h,Q,x}(-\lambda)\big), \tag{10}$$

where the function $\mathsf{K}_{h,Q,x}$ is defined in (8); $\operatorname{supp} Q$ stands for the support of the σ-finite measure Q; and the function $\frac{\mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}}{\mathrm{d}Q}$ is the Radon-Nikodym derivative [24,25] of the probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ with respect to Q.

Note that, while $P_{Y|X}^{(h,Q,\lambda)}$ is an (h, Q, λ)-Gibbs conditional probability measure, the measure $P_{Y|X=x}^{(h,Q,\lambda)}$, obtained by conditioning it upon a given vector $x \in \mathcal{X}$, is referred to as an (h, Q, λ)-Gibbs probability measure.

The condition in (9) is easily met under certain assumptions. For instance, if h is a non-negative function and Q is a finite measure, then it holds for all $\lambda \in (0, +\infty)$. Let $\triangle_Q(\mathbb{R}^m) \triangleq \{P \in \triangle(\mathbb{R}^m) : P \ll Q\}$, with $P \ll Q$ standing for "P absolutely continuous with respect to Q". The

 doi:10.20944/preprints202507.2126.v1

relevance of (h, Q, λ)-Gibbs probability measures relies on the fact that under some conditions, they are the unique solutions to problems of the form,

$$\min_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x,y) \mathrm{d}P(y) + \frac{1}{\lambda} D(P\|Q), \text{ and} \tag{11}$$

$$\max_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x,y) \mathrm{d}P(y) + \frac{1}{\lambda} D(P\|Q), \tag{12}$$

where $\lambda \in \mathbb{R} \setminus \{0\}$, $x \in \mathbb{R}$, and $D(P\|Q)$ denotes the relative entropy (or KL divergence) of P with respect to Q.

Definition 3 (Relative Entropy). *Given two σ-finite measures P and Q on the same measurable space, such that P is absolutely continuous with respect to Q, the relative entropy of P with respect to Q is*

$$D(P\|Q) = \int \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(x)\right) \mathrm{d}Q(x), \tag{13}$$

where the function $\frac{\mathrm{d}P}{\mathrm{d}Q}$ is the Radon-Nikodym derivative of P with respect to Q.

The connection between the optimization problems (11) and (12) and the Gibbs probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ in (10) has been pointed out by several authors. See for instance, [8, Theorem 3] and [26–35] for the former; and [10, Theorem 1] together with [36–38] for the latter. In these references a variety of assumptions and proof techniques have been used to prove such connections. A general and unified statement of these observations is presented hereunder.

Lemma 1. *Assume that the optimization problem in (11) (respectively, in (12)) admits a solution. Then, if $\lambda > 0$ (respectively, if $\lambda < 0$), the probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ in (10) is the unique solution.*

Proof: For the case in which $\lambda > 0$, the proof follows the same approach as the proof of [8, Theorem 3]. Alternatively, for the case in which $\lambda < 0$, the proof follows along the lines of the proof of [10, Theorem 1]. ■

The following lemma highlights a key property of (h, Q, λ)-Gibbs probability measures.

Lemma 2. *Given an (h, Q, λ)-Gibbs probability measure, denoted by $P_{Y|X=x}^{(h,Q,\lambda)}$, with $x \in \mathbb{R}^n$,*

$$-\frac{1}{\lambda} \mathsf{K}_{h,Q,x}(-\lambda) = \int h(x,y) \mathrm{d}Q(y) - \frac{1}{\lambda} D\left(Q \| P_{Y|X=x}^{(h,Q,\lambda)}\right) \tag{14}$$

$$= \int h(x,y) \mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}(y) + \frac{1}{\lambda} D\left(P_{Y|X=x}^{(h,Q,\lambda)} \| Q\right); \tag{15}$$

moreover, if $\lambda > 0$,

$$-\frac{1}{\lambda} \mathsf{K}_{h,Q,x}(-\lambda) = \min_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x,y) \mathrm{d}P(y) + \frac{1}{\lambda} D(P\|Q); \tag{16}$$

alternatively, if $\lambda < 0$,

$$-\frac{1}{\lambda} \mathsf{K}_{h,Q,x}(-\lambda) = \max_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x,y) \mathrm{d}P(y) + \frac{1}{\lambda} D(P\|Q), \tag{17}$$

where the function $\mathsf{K}_{h,Q,x}$ is defined in (8).

Proof: The proof of (15) follows from taking the logarithm of both sides of (10) and integrating with respect to $P_{Y|X=x}^{(h,Q,\lambda)}$. As for the proof of (14), it follows by noticing that for all $(x, y) \in \mathbb{R}^n \times \operatorname{supp} Q$, the

 doi:10.20944/preprints202507.2126.v1

Radon-Nikodym derivative $\frac{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}{\mathrm{d}Q}(y)$ in (10) is strictly positive. Thus, from [39, Theorem 5], it holds that $\frac{\mathrm{d}Q}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y) = \left(\frac{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}{\mathrm{d}Q}(y)\right)^{-1}$. Hence, taking the negative logarithm of both sides of (10) and integrating with respect to Q leads to (14). Finally, the equalities in (16) and (17) follow from Lemma 1 and (15). ■

Lemma 2, at least equalities (15), (16), and (17), can be seen as an immediate restatement of Donsker-Varadhan variational representation of the relative entropy [40]. Alternative interesting proofs for (14) have been presented by several authors including [10, Lemma 3] and [35, Lemma 2.2]. A proof for (15) appears in [29, Lemma 3] in the specific case of $\lambda > 0$.

The following lemma introduces the main building block of this work, which is a characterization of the variation from the probability measure $P^{(h,Q,\lambda)}_{Y|X=x}$ in (10) to an arbitrary measure $P \in \triangle_Q(\mathbb{R}^m)$, i.e., $\mathsf{G}_h\left(x, P, P^{(h,Q,\lambda)}_{Y|X=x}\right)$. Such a result appeared for the first time in [7, Theorem 1] for the case in which $\lambda > 0$; and in [10, Theorem 6] for the case in which $\lambda < 0$, in different contexts of statistical machine learning. A general and unified statement of such results is presented hereunder.

Lemma 3. *Consider an (h, Q, λ)-Gibbs probability measure, denoted by $P^{(h,Q,\lambda)}_{Y|X=x} \in \triangle(\mathbb{R}^m)$, with $\lambda \neq 0$ and $x \in \mathbb{R}$. For all $P \in \triangle_Q(\mathbb{R}^m)$,*

$$\mathsf{G}_h\left(x, P, P^{(h,Q,\lambda)}_{Y|X=x}\right) = \frac{1}{\lambda}\left(D\left(P\|P^{(h,Q,\lambda)}_{Y|X=x}\right) + D\left(P^{(h,Q,\lambda)}_{Y|X=x}\|Q\right) - D(P\|Q)\right). \tag{18}$$

Proof: The proof follows along the lines of the proofs of [7, Theorem 1] for the case in which $\lambda > 0$; and in [10, Theorem 6] for the case in which $\lambda < 0$. A unified proof is presented hereunder by noticing that for all $P \in \triangle_Q(\mathbb{R}^m)$,

$$\begin{aligned}
D\left(P\|P^{(h,Q,\lambda)}_{Y|X=x}\right) &= \int \log\left(\frac{\mathrm{d}P}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right)\mathrm{d}P(y) && (19)\\
&= \int \log\left(\frac{\mathrm{d}Q}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\frac{\mathrm{d}P}{\mathrm{d}Q}(y)\right)\mathrm{d}P(y) && (20)\\
&= \int \log\left(\frac{\mathrm{d}Q}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right)\mathrm{d}P(y) + D(P\|Q) && (21)\\
&= \lambda \int h(x,y)\mathrm{d}P(y) + \mathsf{K}_{h,Q,x}(-\lambda) + D(P\|Q) && (22)\\
&= \lambda \mathsf{G}_h\left(x, P, P^{(h,Q,\lambda)}_{Y|X=x}\right) - D\left(P^{(h,Q,\lambda)}_{Y|X=x}\|Q\right) + D(P\|Q), && (23)
\end{aligned}$$

where (20) follows from [39, Theorem 4]; (22) follows from [39, Theorem 5] and (10); and (23) follows from (15). ■

It is interesting to highlight that $\mathsf{G}_h\left(x, P, P^{(h,Q,\lambda)}_{Y|X=x}\right)$ in (18) characterizes the variation of the expectation of the function $h(x,\cdot) : \mathbb{R}^m \to \mathbb{R}$, when $\lambda > 0$ (resp. $\lambda < 0$) and the probability measure changes from the solution to the optimization problem (11) (resp. (12)) to an alternative measure P. This result takes another perspective if it is seen in the context of information projections [13]. Let Q be a probability measure and $\mathcal{S} \subseteq \triangle_Q(\mathbb{R}^m)$ be a convex set. From [13, Theorem 1], it holds that for all measures $P \in \mathcal{S}$,

$$D(P\|Q) \geqslant D(P\|P^\star) + D(P^\star\|Q), \tag{24}$$

where, $P^\star$ satifies

$$P^\star \in \arg\min_{P\in\mathcal{S}} D(P\|Q). \tag{25}$$

 doi:10.20944/preprints202507.2126.v1

In the particular case in which the set $\mathcal{S}$ in (24) satisfies

$$\mathcal{S} \triangleq \left\{ P \in \triangle_Q(\mathbb{R}^m) : \int h(x,y) \mathrm{d}P(y) = c \right\}, \tag{26}$$

for some real c, with the vector x and the function h defined in Lemma 3, the optimal measure $P^\star$ in (25) is the Gibbs probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ in (10), with $\lambda > 0$ chosen to satisfy

$$\int h(x,y) \mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}(y) = c. \tag{27}$$

The case in which the measure Q in (25) is a σ-finite measure, for instance, either the Lebesgue measure or the counting measure, respectively leads to the classical framework of differential entropy maximization or discrete entropy maximization, which have been studied under particular assumptions on the set $\mathcal{S}$ in [36,37] and [38].

When the reference measure Q is a probability measure, under the assumption that (27) holds, it follows from [13, Theorem 3] that for all $P \in \mathcal{S}$, with $\mathcal{S}$ in (26),

$$D(P\|Q) = D\left(P\|P_{Y|X=x}^{(h,Q,\lambda)}\right) + D\left(P_{Y|X=x}^{(h,Q,\lambda)}\|Q\right), \tag{28}$$

which is known as the *Pythagorean theorem for relative entropy*. Such a geometric interpretation follows from admitting relative entropy as an analog of squared Euclidean distance. The first appearance of such a "Pythagorean theorem" was in [12] and was later revisited in [13]. Interestingly, the same result can be obtained from Lemma 3 by noticing that for all $P \in \mathcal{S}$, with $\mathcal{S}$ in (26),

$$\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right) = 0. \tag{29}$$

The converse of the Pythagorean theorem [41, Book I, Proposition 48] together with Lemma 3, lead to the geometric construction shown in Figure 1. A similar interpretation was also presented in [11, Figure 6] and [11, Figure 7] in the context of the generalization error of machine learning algorithms. The former considers $\lambda > 0$, while the latter considers $\lambda < 0$. Nonetheless, the interpretation in Figure 1 is general and independent of such an application.

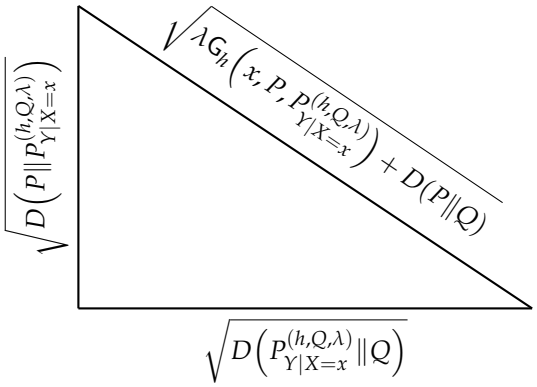

Figure 1. Geometric interpretation of Lemma 3, with Q a probability measure.

The relevance of Lemma 3, with respect to this body of literature on information projections, follows from the fact that Q might be a σ-finite measure, which is a class of measures that includes the class of probability measures, and thus, unifies the results separately obtained in the realm of maximum entropy methods and information-projection methods. More importantly, when $P \notin \mathcal{S}$, with $\mathcal{S}$ in (26), it might hold that $\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right) < 0$ or $\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right) > 0$, with $\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right)$

 doi:10.20944/preprints202507.2126.v1

in (18), which resonates with the fact that (h, Q, λ)-Gibbs conditional probability measures are also related to another class of optimization problems, as shown by the following lemma.

Lemma 4. *Assume that the following optimization problems possess at least one solution for some $x \in \mathbb{R}^n$,*

$$\min_{P \in \triangle_Q(\mathbb{R}^m)} \quad \int h(x,y) \mathrm{d}P(y) \tag{30a}$$
$$\text{s.t.} \quad D(P\|Q) \leqslant \rho. \tag{30b}$$

and

$$\max_{P \in \triangle_Q(\mathbb{R}^m)} \quad \int h(x,y) \mathrm{d}P(y) \tag{31a}$$
$$\text{s.t.} \quad D(P\|Q) \leqslant \rho. \tag{31b}$$

Consider the (h, Q, λ)-Gibbs probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ in (10), with $\lambda \in \mathbb{R} \setminus \{0\}$ such that $\rho = D\left(P_{Y|X=x}^{(h,Q,\lambda)} \| Q\right)$. Then, the (h, Q, λ)-Gibbs probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$ is a solution to (30) if $\lambda > 0$; or to (31) if $\lambda < 0$.

Proof: Note that if $\lambda > 0$, then, $\frac{1}{\lambda} D\left(P \| P_{Y|X=x}^{(h,Q,\lambda)}\right) \geqslant 0$. Hence, from Lemma 3, it holds that for all probability mesures P such that $D(P\|Q) \leqslant \rho$,

$$\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right) \geqslant \frac{1}{\lambda}\left(D\left(P_{Y|X=x}^{(h,Q,\lambda)} \| Q\right) - D(P\|Q)\right) \tag{32}$$
$$= \frac{1}{\lambda}(\rho - D(P\|Q)) \tag{33}$$
$$\geqslant 0, \tag{34}$$

with equality if $D\left(P \| P_{Y|X=x}^{(h,Q,\lambda)}\right) = 0$. This implies that $P_{Y|X=x}^{(h,Q,\lambda)}$ is a solution to (30). Note also that if $\lambda < 0$, from Lemma 3, it holds that for all probability mesures P such that $D(P\|Q) \leqslant \rho$,

$$\mathsf{G}_h\left(x, P, P_{Y|X=x}^{(h,Q,\lambda)}\right) \leqslant \frac{1}{\lambda}\left(D\left(P_{Y|X=x}^{(h,Q,\lambda)} \| Q\right) - D(P\|Q)\right) \tag{35}$$
$$= \frac{1}{\lambda}(\rho - D(P\|Q)) \tag{36}$$
$$\leqslant 0, \tag{37}$$

with equality if $D\left(P \| P_{Y|X=x}^{(h,Q,\lambda)}\right) = 0$. This implies that $P_{Y|X=x}^{(h,Q,\lambda)}$ is a solution to (31). ■

3. Characterization of $\mathsf{G}_h(x, P_1, P_2)$ in (1)

The main result of this section is the following theorem.

Theorem 1. *For all probability measures P_1 and P_2, both absolutely continuous with respect to a given σ-finite measure Q on $\mathbb{R}^m$, the variation $\mathsf{G}_h(x, P_1, P_2)$ in (1) satisfies,*

$$\mathsf{G}_h(x, P_1, P_2) = \frac{1}{\lambda}\left(D\left(P_1 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - D\left(P_2 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) + D(P_2\|Q) - D(P_1\|Q)\right), \tag{38}$$

where the probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$, with $\lambda \neq 0$, is an (h, Q, λ)-Gibbs probability measure.

Proof: The proof follows from Lemma 3 and by observing that

$$\mathsf{G}_h(x, P_1, P_2) = \mathsf{G}_h\left(x, P_1, P_{Y|X=x}^{(h,Q,\lambda)}\right) - \mathsf{G}_h\left(x, P_2, P_{Y|X=x}^{(h,Q,\lambda)}\right).$$

■

 doi:10.20944/preprints202507.2126.v1

Theorem 1 might be particularly simplified in the case in which the reference measure Q is a probability measure. Consider for instance the case in which $P_1 \ll P_2$ (or $P_2 \ll P_1$). In such a case, the reference measure might be chosen as P_2 (or P_1), as shown hereunder.

Corollary 1. *Consider the variation* $\mathsf{G}_h(x, P_1, P_2)$ *in* (1)*. If the probability measure* P_1 *is absolutely continuous with respect to* P_2*, then,*

$$\mathsf{G}_h(x, P_1, P_2) = \frac{1}{\lambda}\left(D\left(P_1 \| P_{Y|X=x}^{(h,P_2,\lambda)}\right) - D\left(P_2 \| P_{Y|X=x}^{(h,P_2,\lambda)}\right) - D(P_1 \| P_2)\right). \tag{39}$$

Alternatively, if the probability measure P_2 *is absolutely continuous with respect to* P_1*, then,*

$$\mathsf{G}_h(x, P_1, P_2) = \frac{1}{\lambda}\left(D\left(P_1 \| P_{Y|X=x}^{(h,P_1,\lambda)}\right) - D\left(P_2 \| P_{Y|X=x}^{(h,P_1,\lambda)}\right) + D(P_2 \| P_1)\right), \tag{40}$$

where the probability measures $P_{Y|X=x}^{(h,P_1,\lambda)}$ *and* $P_{Y|X=x}^{(h,P_2,\lambda)}$ *are respectively* (h, P_1, λ)*- and* (h, P_2, λ)*-Gibbs probability measures, with* $\lambda \neq 0$*.*

In the case in which neither P_1 is absolutely continuous with respect to P_2; nor P_2 is absolutely continuous with respect to P_1, the reference measure Q in Theorem 1 can always be chosen as a convex combination of P_1 and P_2. That is, for all Borel sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$, $Q(\mathcal{A}) = \alpha P_1(\mathcal{A}) + (1-\lambda)P_2(\mathcal{A})$, with $\alpha \in (0,1)$.

Theorem 1 can be specialized to the specific cases in which Q is the Lebesgue or the counting measure.

If Q is the Lebesgue measure the probability measures P_1 and P_2 in (38) admit probability density functions f_1 and f_2, respectively. Moreover, the terms $-D(P_1\|Q)$ and $-D(P_2\|Q)$ are Shannon's differential entropies [4] induced by P_1 and P_2, denoted by $\mathsf{h}(P_1)$ and $\mathsf{h}(P_2)$, respectively. That is, for all $i \in \{1,2\}$,

$$\mathsf{h}(P_i) \triangleq -\int f_i(x)\log f_i(x)\mathrm{d}x. \tag{41}$$

The probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$, with $\lambda \neq 0$, $x \in \mathbb{R}^n$, and Q the Lebesgue measure, possesses a probability density function, denoted by $f_{Y|X=x}^{(h,Q,\lambda)} : \mathbb{R}^m \to (0, +\infty)$, which satisfies

$$f_{Y|X=x}^{(h,Q,\lambda)}(y) = \frac{\exp(-\lambda h(x,y))}{\int \exp(-\lambda h(x,y))\mathrm{d}y}. \tag{42}$$

If Q is the counting measure the probability measures P_1 and P_2 in (38) admit probability mass functions $p_1 : \mathcal{Y} \to [0,1]$ and $p_2 : \mathcal{Y} \to [0,1]$, with $\mathcal{Y}$ a countable subset of $\mathbb{R}^m$. Moreover, $-D(P_1\|Q)$ and $-D(P_2\|Q)$ are respectively Shannon's discrete entropies [4] induced by P_1 and P_2, denoted by $\mathsf{H}(P_1)$ and $\mathsf{H}(P_2)$, respectively. That is, for all $i \in \{1,2\}$,

$$\mathsf{H}(P_i) \triangleq -\sum_{y\in\mathcal{Y}} p_i(y)\log p_i(y). \tag{43}$$

The probability measure $P_{Y|X=x}^{(h,Q,\lambda)}$, with $\lambda \neq 0$ and Q the counting measure, possesses a conditional probability mass function, denoted by $p_{Y|X=x}^{(h,Q,\lambda)} : \mathcal{Y} \to (0, +\infty)$, which satisfies

$$p_{Y|X=x}^{(h,Q,\lambda)}(y) = \frac{\exp(-\lambda h(x,y))}{\sum_{y\in\mathcal{Y}} \exp(-\lambda h(x,y))}. \tag{44}$$

These observations lead to the following corollary of Theorem 1.

 doi:10.20944/preprints202507.2126.v1

Corollary 2. *Given two probability measures P_1 and P_2, with probability density functions f_1 and f_2 respectively, the variation $\mathsf{G}_h(x, P_1, P_2)$ in (1) satisfies,*

$$\mathsf{G}_h(x, P_1, P_2) = \frac{1}{\lambda}\left(D\left(P_1 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - D\left(P_2 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - \mathsf{h}(P_2) + \mathsf{h}(P_1)\right), \tag{45}$$

where the probability density function of the measure $P_{Y|X=x}^{(h,Q,\lambda)}$, with $\lambda \neq 0$ and Q the Lebesgue measure, is defined in (42); and the entropy functional h is defined in (41). Alternatively, given two probability measures P_1 and P_2, with probability mass functions p_1 and p_2 respectively, the variation $\mathsf{G}_h(x, P_1, P_2)$ in (1) satisfies,

$$\mathsf{G}_h(x, P_1, P_2) = \frac{1}{\lambda}\left(D\left(P_1 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - D\left(P_2 \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - \mathsf{H}(P_2) + \mathsf{H}(P_1)\right), \tag{46}$$

where the probability mass function of the measure $P_{Y|X=x}^{(h,Q,\lambda)}$, with $\lambda \neq 0$ and Q the counting measure, is defined in (44); and the entropy functional H is defined in (43).

4. Characterizations of $\bar{\mathsf{G}}_h\left(P_{Y|X}^{(1)}, P_{Y|X}^{(2)}, P_X\right)$ in (2)

The main result of this section is a characterization of $\bar{\mathsf{G}}_h\left(P_{Y|X}^{(1)}, P_{Y|X}^{(2)}, P_X\right)$ in (2).

Theorem 2. *Consider the variation $\bar{\mathsf{G}}_h\left(P_{Y|X}^{(1)}, P_{Y|X}^{(2)}, P_X\right)$ in (2) and assume that for all $x \in \operatorname{supp} P_X$, the probability measures $P_{Y|X=x}^{(1)}$ and $P_{Y|X=x}^{(2)}$ are both absolutely continuous with respect to a σ-measure Q. Then,*

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P_{Y|X}^{(1)}, P_{Y|X}^{(2)}, P_X\right) = \frac{1}{\lambda}\int \Bigg(&D\left(P_{Y|X=x}^{(1)} \| P_{Y|X=x}^{(h,Q,\lambda)}\right) - D\left(P_{Y|X=x}^{(2)} \| P_{Y|X=x}^{(h,Q,\lambda)}\right) \\ &+ D\left(P_{Y|X=x}^{(2)} \| Q\right) - D\left(P_{Y|X=x}^{(1)} \| Q\right)\Bigg) \mathrm{d}P_X(x),\end{aligned} \tag{47}$$

where the probability measure $P_{Y|X}^{(h,Q,\lambda)}$, with $\lambda \neq 0$, is an (h, Q, λ)-Gibbs conditional probability measure.

Proof: The proof follows from (2) and Theorem 1. ∎

Two special cases are particularly noteworthy.

When the reference measure Q is the Lebesgue measure

both $-\int D\left(P_{Y|X=x}^{(1)} \| Q\right) \mathrm{d}P_X(x)$ and $-\int D\left(P_{Y|X=x}^{(2)} \| Q\right) \mathrm{d}P_X(x)$ in (47) become Shannon's differential conditional entropies, denoted by $\mathsf{h}\left(P_{Y|X}^{(1)} | P_X\right)$ and $\mathsf{h}\left(P_{Y|X}^{(2)} | P_X\right)$, respectively. That is, for all $i \in \{1, 2\}$,

$$\mathsf{h}\left(P_{Y|X}^{(i)} | P_X\right) \triangleq \int \mathsf{h}\left(P_{Y|X=x}^{(i)}\right) \mathrm{d}P_X(x), \tag{48}$$

where h is the entropy functional in (41).

When the reference measure Q is the counting measure

both $-\int D\left(P_{Y|X=x}^{(1)} \| Q\right) \mathrm{d}P_X(x)$ and $-\int D\left(P_{Y|X=x}^{(2)} \| Q\right) \mathrm{d}P_X(x)$ in (47) become Shannon's discrete conditional entropies, denoted by $\mathsf{H}\left(P_{Y|X}^{(1)} | P_X\right)$ and $\mathsf{H}\left(P_{Y|X}^{(2)} | P_X\right)$, respectively. That is, for all $i \in \{1, 2\}$,

$$\mathsf{H}\left(P_{Y|X}^{(i)} | P_X\right) \triangleq \int \mathsf{H}\left(P_{Y|X=x}^{(i)}\right) \mathrm{d}P_X(x), \tag{49}$$

where H is the entropy functional in (43).

These observations lead to the following corollary of Theorem 2.

 doi:10.20944/preprints202507.2126.v1

Corollary 3. *Consider the variation* $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ *in* (2) *and assume that for all* $x \in \operatorname{supp} P_X$*, the probability measures* $P^{(1)}_{Y|X=x}$ *and* $P^{(2)}_{Y|X=x}$ *possess probability density functions. Then,*

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) = \frac{1}{\lambda}\int\left(D\left(P^{(1)}_{Y|X=x}\|P^{(h,Q,\lambda)}_{Y|X=x}\right) - D\left(P^{(2)}_{Y|X=x}\|P^{(h,Q,\lambda)}_{Y|X=x}\right)\right)\mathrm{d}P_X(x)\\ -\frac{1}{\lambda}\mathsf{h}\left(P^{(2)}_{Y|X}|P_X\right) + \frac{1}{\lambda}\mathsf{h}\left(P^{(1)}_{Y|X}|P_X\right),\end{aligned} \tag{50}$$

where the probability density function of the measure $P^{(h,Q,\lambda)}_{Y|X=x}$*, with* $\lambda \neq 0$ *and* Q *the Lebesgue measure, is defined in* (42)*; and for all* $i \in \{1,2\}$*, the conditional entropy* $\mathsf{h}\left(P^{(i)}_{Y|X}|P_X\right)$ *is defined in* (48)*. Alternatively, assume that for all* $x \in \operatorname{supp} P_X$*, the probability measures* $P^{(1)}_{Y|X=x}$ *and* $P^{(2)}_{Y|X=x}$ *possess probability mass functions. Then,*

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) = \frac{1}{\lambda}\int\left(D\left(P^{(1)}_{Y|X=x}\|P^{(h,Q,\lambda)}_{Y|X=x}\right) - D\left(P^{(2)}_{Y|X=x}\|P^{(h,Q,\lambda)}_{Y|X=x}\right)\right)\mathrm{d}P_X(x)\\ -\frac{1}{\lambda}\mathsf{H}\left(P^{(2)}_{Y|X}|P_X\right) + \frac{1}{\lambda}\mathsf{H}\left(P^{(1)}_{Y|X}|P_X\right),\end{aligned} \tag{51}$$

where the probability mass function of the measure $P^{(h,Q,\lambda)}_{Y|X=x}$*, with* $\lambda \neq 0$ *and* Q *the counting measure, is defined in* (44)*; and for all* $i \in \{1,2\}$*, the conditional entropy* $\mathsf{H}\left(P^{(i)}_{Y|X}|P_X\right)$ *is defined in* (49)*.*

Note that, from (2), it follows that the general expression for the expected variation $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ might be simplified according to Corollary 1. For instance, if for all $x \in \operatorname{supp} P_X$, the probability measure $P^{(1)}_{Y|X=x}$ is absolutely continuous with respect to $P^{(2)}_{Y|X=x}$, the measure $P^{(2)}_{Y|X=x}$ can be chosen to be the reference measure in the calculation of $\mathsf{G}_h\left(x, P^{(1)}_{Y|X=x}, P^{(2)}_{Y|X=x}\right)$ in (2). This observation leads to the following corollary of Theorem 2.

Corollary 4. *Consider the variation* $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ *in* (2) *and assume that for all* $x \in \operatorname{supp} P_X$*,* $P^{(1)}_{Y|X=x} \ll P^{(2)}_{Y|X=x}$*. Then,*

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) = \frac{1}{\lambda}\int\left(D\left(P^{(1)}_{Y|X=x}\|P^{\left(h,P^{(2)}_{Y|X=x},\lambda\right)}_{Y|X=x}\right) - D\left(P^{(2)}_{Y|X=x}\|P^{\left(h,P^{(2)}_{Y|X=x},\lambda\right)}_{Y|X=x}\right)\right.\\ \left.-D\left(P^{(1)}_{Y|X=x}\|P^{(2)}_{Y|X=x}\right)\right)\mathrm{d}P_X(x).\end{aligned} \tag{52}$$

Alternatively, if for all $x \in \operatorname{supp} P_X$*, the probability measure* $P^{(2)}_{Y|X=x}$ *is absolutely continuous with respect to* $P^{(1)}_{Y|X=x}$*, then,*

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right) = \frac{1}{\lambda}\int\left(D\left(P^{(1)}_{Y|X=x}\|P^{\left(h,P^{(1)}_{Y|X=x},\lambda\right)}_{Y|X=x}\right) - D\left(P^{(2)}_{Y|X=x}\|P^{\left(h,P^{(1)}_{Y|X=x},\lambda\right)}_{Y|X=x}\right)\right.\\ \left.+D\left(P^{(2)}_{Y|X=x}\|P^{(1)}_{Y|X=x}\right)\right)\mathrm{d}P_X(x),\end{aligned} \tag{53}$$

where the measures $P^{\left(h,P^{(1)}_{Y|X=x},\lambda\right)}_{Y|X=x}$ *and* $P^{\left(h,P^{(2)}_{Y|X=x},\lambda\right)}_{Y|X=x}$ *are respectively* $(h, P^{(1)}_{Y|X=x}, \lambda)$*- and* $(h, P^{(2)}_{Y|X=x}, \lambda)$*-Gibbs probability measures.*

 doi:10.20944/preprints202507.2126.v1

The Gibbs probability measures $P_{Y|X=x}^{\left(h,P_{Y|X=x}^{(1)},\lambda\right)}$ and $P_{Y|X=x}^{\left(h,P_{Y|X=x}^{(2)},\lambda\right)}$ in Corollary 4 are particularly interesting as their reference measures depend on x. Gibbs measures of this form appear, for instance, in [8, Corollary 10].

5. Characterizations of $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6)

The main result of this section is a characterization of $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6), which describes the variation of the expectation of the function h when the probability measure changes from the joint probability measure $P_{Y|X}P_X$ to the product of its marginals $P_Y \cdot P_X$. This result is presented hereunder and involves the mutual information $I\left(P_{Y|X}; P_X\right)$ and lautum information $L\left(P_{Y|X}; P_X\right)$, defined as follows:

$$I\left(P_{Y|X}; P_X\right) \triangleq \int D\left(P_{Y|X=x} \| P_Y\right) \mathrm{d}P_X(x); \text{ and} \tag{54}$$

$$L\left(P_{Y|X}; P_X\right) \triangleq \int D\left(P_Y \| P_{Y|X=x}\right) \mathrm{d}P_X(x). \tag{55}$$

Theorem 3. *Consider the expected variation $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6) and assume that, for all $x \in \operatorname{supp} P_X$:*

1. *The probability measures P_Y and $P_{Y|X=x}$ are both absolutely continuous with respect to a given σ-finite measure Q; and*
2. *The probability measures P_Y and $P_{Y|X=x}$ are mutually absolutely continuous.*

Then, it follows that

$$\begin{aligned}\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right) &= \frac{1}{\lambda}\Bigg(I\left(P_{Y|X}; P_X\right) + L\left(P_{Y|X}; P_X\right) \\ &+ \iint \log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}}(y)\right) \mathrm{d}P_Y(y)\mathrm{d}P_X(x) - \iint \log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_{Y|X=x}^{(h,Q,\lambda)}}(y)\right) \mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)\Bigg),\end{aligned} \tag{56}$$

where the probability measure $P_{Y|X}^{(h,Q,\lambda)}$, with $\lambda \neq 0$, is an (h, Q, λ)-Gibbs conditional probability measure.

Proof: The proof is presented in Appendix A. ■

An alternative expression for $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6) involving only relative entropies is presented by the following theorem.

Theorem 4. *Consider the expected variation $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6) and assume that, for all $x \in \operatorname{supp} P_X$, the probability measure $P_{Y|X=x}$ is absolutely continuous with respect to a given σ-finite measure Q. Then, it follows that*

$$\begin{aligned}&\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right) \\ &= \frac{1}{\lambda}\iint\left(D\left(P_{Y|X=x_2} \| P_{Y|X=x_1}^{(h,Q,\lambda)}\right) - D\left(P_{Y|X=x_2} \| P_{Y|X=x_2}^{(h,Q,\lambda)}\right)\right) \mathrm{d}P_X(x_1)\mathrm{d}P_X(x_2),\end{aligned} \tag{57}$$

where $P_{Y|X}^{(h,Q,\lambda)}$, with $\lambda \neq 0$, is an (h, Q, λ)-Gibbs conditional probability measure.

Proof: The proof is presented in Appendix B. ■

Theorem 4 expresses the variation $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ in (6) as the expectation (w.r.t. $P_X\, P_X$) of a comparison of the conditional probability measure $P_{Y|X}$ with a Gibbs conditional probability measure $P_{Y|X}^{(h,Q,\lambda)}$ via relative entropy. More specifically, the expression consists in the expectation of the difference of two relative entropies. The former compares $P_{Y|X=x_1}$ with $P_{Y|X=x_2}^{(h,Q,\lambda)}$, where $(x_1, x_2) \in$

 doi:10.20944/preprints202507.2126.v1

$\mathcal{X} \times \mathcal{X}$ are independently sampled from the same probability measure P_X. The latter compares these two conditional measures conditioning on the same element of $\mathcal{X}$. That is, it compares $P_{Y|X=x_2}$ with $P^{(h,Q,\lambda)}_{Y|X=x_2}$.

An interesting observation from Theorem 3 and Theorem 4 is that the last two terms in the right-hand side of (56) are both zero in the case in which $P_{Y|X}$ is an (h, Q, λ)-Gibbs conditional probability measure. Similarly, in such a case, the second term in the right-hand side of (57) is also zero. This observation is highlighted by the following corollary.

Corollary 5. *Consider an (h, Q, λ)-Gibbs conditional probability measure, denoted by $P^{(h,Q,\lambda)}_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$, with $\lambda \neq 0$; and a probability measure $P_X \in \triangle(\mathbb{R}^n)$. Let the measure $P^{(h,Q,\lambda)}_{Y} \in \triangle(\mathbb{R}^m)$ be such that for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$,*

$$P^{(h,Q,\lambda)}_{Y}(\mathcal{A}) = \int P^{(h,Q,\lambda)}_{Y|X=x}(\mathcal{A}) \mathrm{d}P_X(x). \tag{58}$$

Then,

$$\bar{\mathsf{G}}_h\left(P^{(h,Q,\lambda)}_{Y}, P^{(h,Q,\lambda)}_{Y|X}, P_X\right) = \frac{1}{\lambda}\left(I\left(P^{(h,Q,\lambda)}_{Y|X}; P_X\right) + L\left(P^{(h,Q,\lambda)}_{Y|X}; P_X\right)\right) \tag{59}$$

$$= \frac{1}{\lambda} \iint D\left(P^{(h,Q,\lambda)}_{Y|X=x_2} \| P^{(h,Q,\lambda)}_{Y|X=x_1}\right) \mathrm{d}P_X(x_1) \mathrm{d}P_X(x_2). \tag{60}$$

Note that mutual information and lautum information are both nonnegative information measures, which from Corollary 5, implies that $\bar{\mathsf{G}}_h\left(P^{(h,Q,\lambda)}_{Y}, P^{(h,Q,\lambda)}_{Y|X}, P_X\right)$ in (60) might be either positive or negative depending exclusively on the sign of the regularization factor λ. The following corollary exploits such an observation to present a property of Gibbs conditional probability measures and their corresponding marginal probability measures.

Corollary 6. *Given a probability measure $P_X \in \triangle(\mathbb{R}^n)$, the (h, Q, λ)-Gibbs conditional probability measure $P^{(h,Q,\lambda)}_{Y|X}$ in (10) and the probability measure $P^{(h,Q,\lambda)}_{Y}$ in (58) satisfy*

$$\iint h(x,y) \mathrm{d}P^{(h,Q,\lambda)}_{Y}(y) \mathrm{d}P_X(x) \geqslant \iint h(x,y) \mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}(y) \mathrm{d}P_X(x) \text{ if } \lambda > 0; \tag{61}$$

or

$$\iint h(x,y) \mathrm{d}P^{(h,Q,\lambda)}_{Y}(y) \mathrm{d}P_X(x) \leqslant \iint h(x,y) \mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}(y) \mathrm{d}P_X(x) \text{ if } \lambda < 0. \tag{62}$$

Corollary 6 highlights the fact that a deviation from the joint probability measure $P^{(h,Q,\lambda)}_{Y|X} P_X \in \triangle(\mathcal{Y} \times \mathcal{X})$ to the product of its marignals $P^{(h,Q,\lambda)}_{Y} P_X \in \triangle(\mathcal{Y} \times \mathcal{X})$ might increase or decrease the expectation of the function h depending on the sign of λ.

6. Final Remarks

A simple re-formulation of Varandan's variational representation of relative entropy (Lemma 2) has been harnessed to provide an explicit expression of the variation of the expectation of a multi-dimensional real function when the probability measure changes from a Gibbs probability measure to an arbitrary measure (Lemma 3). This result reveals strong connections with information projection methods, Pythagorean identities involving relative entropy, and mean optimization problems constrained to a neighborhood around a reference measure, where the neighborhood is defined via an upper bound on relative entropy with respect to the reference measure (Lemma 4). An algebraic manipulation on Lemma 3 leads to an explicit expression for the variation of the expectation under study when the probability measure changes from an arbitrary measure to another arbitrary measure (Theorem 1). The astonishing simplicity in the proof, which is straight forward from Lemma 3,

 doi:10.20944/preprints202507.2126.v1

contrasts with the generality of the result. In particular, the only assumption is that both measures, before and after the variation, are absolutely continuous with a reference measure. The underlying observation is the central role played by Gibbs probability distributions in this analysis. In particular, such a variation is expressed in terms of comparisons, via relative entropy, of the initial and final probability measures with respect to the Gibbs probability measure specifically built for the function under study. Interestingly, the reference measure of such Gibbs probability measures can be freely chosen beyond probability measures. When such a reference is a σ-finite measure, e.g., Lebesgue measure or a counting measure, the expressions mentioned above include Shannon's fundamental information measures, e.g., entropy and conditional entropy (Corollary 2).

Using these initial results, the variations of expectations of multi-dimensional functions due to variations of joint probability measures has been studied. In this case, the focus has been on two particular measure changes, which unveil connections with both mutual and lautum information. First, one of the marginal probability measures remains the same after the change (Theorem 2); and second, the joint probability measure changes to the product of its marginals (Theorem 3 and Theorem 4). In the case of Gibbs joint probability measures, these expressions involve exclusively well known information measures: mutual information; lautum information; and relative entropy. These expressions reveal general connections between the variation in the expectation of arbitrary functions, induced by changes in the underlying probability measure, and both mutual and lautum information. These connections extend beyond those previously established in the analysis of generalization error in machine learning algorithms; see, for instance, [26, Theorem 1] and [8, Theorem 14].

Author Contributions: All authors contributed equally to this research.

Funding: This research was supported in part by the European Commission through the H2020-MSCA-RISE-2019 project 872172; the French National Agency for Research (ANR) through the Project ANR-21-CE25-0013 and the project ANR-22-PEFT-0010 of the France 2030 program PEPR Réseaux du Futur; and in part with the Agence de l'innovation de défense (AID) through the project UK-FR 2024352

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Proof of Theorem 3

The proof follows from Theorem 2, which holds under assumption (a) and leads to

$$\begin{aligned}&\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)\\ =&\frac{1}{\lambda}\int\left(D\left(P_Y\|P_{Y|X=x}^{(h,Q,\lambda)}\right)-D\left(P_{Y|X=x}\|P_{Y|X=x}^{(h,Q,\lambda)}\right)+D\left(P_{Y|X=x}\|Q\right)-D(P_Y\|Q)\right)\mathrm{d}P_X(x).\end{aligned}\tag{A1}$$

 doi:10.20944/preprints202507.2126.v1

The proof continues by noticing that

$$\int D\Big(P_{Y|X=x}\|Q\Big)\mathrm{d}P_X(x)=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}Q}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x) \tag{A2}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\frac{\mathrm{d}P_Y}{\mathrm{d}Q}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x) \tag{A3}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)$$
$$+\int\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}Q}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x) \tag{A4}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)$$
$$+\int\int\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}Q}(y)\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x) \tag{A5}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)$$
$$+\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}Q}(y)\right)\left(\int\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_X(x)\right)\mathrm{d}P_Y(y) \tag{A6}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\right)\mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)$$
$$+\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}Q}(y)\right)\mathrm{d}P_Y(y) \tag{A7}$$

$$=I\Big(P_{Y|X};P_X\Big)+D(P_Y\|Q), \tag{A8}$$

where (A3) follows from [39, Theorem 4]; (A5) follows from [39, Theorem 2]; and (A7) follows from [39, Theorem 10], which implies that for all $y\in\mathbb{R}^m$, $\int\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P_Y}(y)\mathrm{d}P_X(x)=1$.

Note also that

$$\int D\Big(P_Y\|P^{(h,Q,\lambda)}_{Y|X=x}\Big)\mathrm{d}P_X(x)=\int\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x) \tag{A9}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}P_{Y|X=x}}(y)\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x) \tag{A10}$$

$$=\int\int\log\left(\frac{\mathrm{d}P_Y}{\mathrm{d}P_{Y|X=x}}(y)\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x)$$
$$+\int\int\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x) \tag{A11}$$

$$=L\Big(P_{Y|X=x};P_X\Big)+\iint\log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right)\mathrm{d}P_Y(y)\mathrm{d}P_X(x), \tag{A12}$$

where (A10) follows from [39, Theorem 4]. Finally, using (A8) and (A12) in (A1) yields (56), which completes the proof.

Appendix B. Proof of Theorem 4

The proof follows by observing that the functional $\bar{\mathsf{G}}_h$ in (6) satisfies

$$\bar{\mathsf{G}}_h\Big(P_Y,P_{Y|X},P_X\Big)$$
$$=\iint\left(\int h(x_2,y)\mathrm{d}P_{Y|X=x_1}(y)-\int h(x_1,y)\mathrm{d}P_{Y|X=x_1}(y)\right)\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1). \tag{A13}$$

 doi:10.20944/preprints202507.2126.v1

Using the functional G_h in (1), the terms above can be written as follows

$$\int h(x_1,y)\mathrm{d}P_{Y|X=x_1}(y)=\mathsf{G}_h\left(x_1,P_{Y|X=x_1},P_{Y|X=x_1}^{(h,Q,\lambda)}\right)+\int h(x_1,y)\mathrm{d}P_{Y|X=x_1}^{(h,Q,\lambda)}(y), \tag{A14}$$

and

$$\int h(x_2,y)\mathrm{d}P_{Y|X=x_1}(y)=\mathsf{G}_h\left(x_2,P_{Y|X=x_1},P_{Y|X=x_2}^{(h,Q,\lambda)}\right)+\int h(x_2,y)\mathrm{d}P_{Y|X=x_2}^{(h,Q,\lambda)}(y). \tag{A15}$$

Using (A14) and (A15) in (A13) yields

$$\begin{aligned}
&\bar{\mathsf{G}}_h\left(P_Y,P_{Y|X},P_X\right)\\
=&\iint\left(\mathsf{G}_h\left(x_2,P_{Y|X=x_1},P_{Y|X=x_2}^{(h,Q,\lambda)}\right)+\int h(x_2,y)\mathrm{d}P_{Y|X=x_2}^{(h,Q,\lambda)}(y)\right. &\text{(A16)}\\
&\left.-\mathsf{G}_h\left(x_1,P_{Y|X=x_1},P_{Y|X=x_1}^{(h,Q,\lambda)}\right)-\int h(x_1,y)\mathrm{d}P_{Y|X=x_1}^{(h,Q,\lambda)}(y)\right)\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1),\\
=&\iint\left(\mathsf{G}_h\left(x_2,P_{Y|X=x_1},P_{Y|X=x_2}^{(h,Q,\lambda)}\right)-\int\mathsf{G}_h\left(x_1,P_{Y|X=x_1},P_{Y|X=x_1}^{(h,Q,\lambda)}\right)\right)\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1) &\text{(A17)}\\
=&\frac{1}{\lambda}\iint\left(D\left(P_{Y|X=x_2}\|P_{Y|X=x_1}^{(h,Q,\lambda)}\right)-D\left(P_{Y|X=x_2}\|P_{Y|X=x_2}^{(h,Q,\lambda)}\right)\right)\mathrm{d}P_X(x_1)\mathrm{d}P_X(x_2), &\text{(A18)}
\end{aligned}$$

where the last equality holds from Lemma 3, which implies

$$\begin{aligned}
&\iint\mathsf{G}_h\left(x_2,P_{Y|X=x_1},P_{Y|X=x_2}^{(h,Q,\lambda)}\right)\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1)\\
=&\frac{1}{\lambda}\iint D\left(P_{Y|X=x_1}\|P_{Y|X=x_2}^{(h,Q,\lambda)}\right)\frac{1}{\lambda}\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1)+\frac{1}{\lambda}\int D\left(P_{Y|X=x_2}^{(h,Q,\lambda)}\|Q\right)\mathrm{d}P_X(x_2)\\
&-\frac{1}{\lambda}\int D\left(P_{Y|X=x_1}\|Q\right)\mathrm{d}P_X(x_1),
\end{aligned} \tag{A19}$$

and

$$\begin{aligned}
&\iint\mathsf{G}_h\left(x_1,P_{Y|X=x_1},P_{Y|X=x_1}^{(h,Q,\lambda)}\right)\mathrm{d}P_X(x_2)\mathrm{d}P_X(x_1)\\
=&\int\mathsf{G}_h\left(x_1,P_{Y|X=x_1},P_{Y|X=x_1}^{(h,Q,\lambda)}\right)\mathrm{d}P_X(x_1) &\text{(A20)}\\
=&\frac{1}{\lambda}\int D\left(P_{Y|X=x_1}\|P_{Y|X=x_1}^{(h,Q,\lambda)}\right)\mathrm{d}P_X(x_1)+\frac{1}{\lambda}\int D\left(P_{Y|X=x_1}^{(h,Q,\lambda)}\|Q\right)\mathrm{d}P_X(x_1)\\
&-\frac{1}{\lambda}\int D\left(P_{Y|X=x_1}\|Q\right)\mathrm{d}P_X(x_1), &\text{(A21)}
\end{aligned}$$

which completes the proof.

References

1. Gama, J.; Medas, P.; Castillo, G.; Rodrigues, P. Learning with drift detection. In Proceedings of the Proceedings of the 17th Brazilian Symposium on Artificial Intelligence, Sao Luis, Maranhao, Brazil, Oct. 2004; pp. 286–295.
2. Webb, G.I.; Lee, L.K.; Goethals, B.; Petitjean, F. Analyzing concept drift and shift from sample data. *Data Mining and Knowledge Discovery* **2018**, *32*, 1179–1199.
3. Oliveira, G.H.F.M.; Minku, L.L.; Oliveira, A.L. Tackling virtual and real concept drifts: An adaptive Gaussian mixture model approach. *IEEE Transactions on Knowledge and Data Engineering* **2021**, *35*, 2048–2060.
4. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal* **1948**, *27*, 379–423.
5. Shannon, C.E. A mathematical theory of communication. *The Bell System Technical Journal* **1948**, *27*, 623–656.
6. Palomar, D.P.; Verdú, S. Lautum information. *IEEE Transactions on Information Theory* **2008**, *54*, 964–975.
7. Perlaza, S.M.; Esnaola, I.; Bisson, G.; Poor, H.V. On the Validation of Gibbs Algorithms: Training Datasets, Test Datasets and their Aggregation. In Proceedings of the Proceedings of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun. 2023.

 doi:10.20944/preprints202507.2126.v1

8. Perlaza, S.M.; Bisson, G.; Esnaola, I.; Jean-Marie, A.; Rini, S. Empirical Risk Minimization with Relative Entropy Regularization. *IEEE Transactions on Information Theory* **2024**, *70*, 5122 – 5161.
9. Zou, X.; Perlaza, S.M.; Esnaola, I.; Altman, E. Generalization Analysis of Machine Learning Algorithms via the Worst-Case Data-Generating Probability Measure. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, Canada, Feb. 2024.
10. Zou, X.; Perlaza, S.M.; Esnaola, I.; Altman, E.; Poor, H.V. The Worst-Case Data-Generating Probability Measure in Statistical Learning. *IEEE Journal on Selected Areas in Information Theory* **2024**, *5*, 175 – 189.
11. Perlaza, S.M.; Zou, X. The Generalization Error of Machine Learning Algorithms. *arXiv preprint arXiv:2411.12030* **2024**.
12. Chentsov, N.N. Nonsymmetrical distance between probability distributions, entropy and the theorem of Pythagoras. *Mathematical notes of the Academy of Sciences of the USSR* **1968**, *4*, 686–691.
13. Csiszár, I.; Matus, F. Information projections revisited. *IEEE Transactions on Information Theory* **2003**, *49*, 1474–1490.
14. Müller, A. Integral probability metrics and their generating classes of functions. *Advances in applied probability* **1997**, *29*, 429–443.
15. Zolotarev, V.M. Probability metrics. *Teoriya Veroyatnostei i ee Primeneniya* **1983**, *28*, 264–287.
16. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A Kernel Two-Sample Test. *Journal of Machine Learning Research* **2012**, *13*, 723–773.
17. Villani, C. *Optimal transport: Old and new*, first ed.; Springer: Berlin, Germany, 2009.
18. Liu, W.; Yu, G.; Wang, L.; Liao, R. An Information-Theoretic Framework for Out-of-Distribution Generalization with Applications to Stochastic Gradient Langevin Dynamics. *arXiv preprint arXiv:2403.19895* **2024**.
19. Liu, W.; Yu, G.; Wang, L.; Liao, R. An Information-Theoretic Framework for Out-of-Distribution Generalization. In Proceedings of the Proceedings of the IEEE International Symposium on Information Theory (ISIT), Athens, Greece, July 2024; pp. 2670–2675.
20. Agrawal, R.; Horel, T. Optimal Bounds between f-Divergences and Integral Probability Metrics. *Journal of Machine Learning Research* **2021**, *22*, 1–59.
21. Rahimian, H.; Mehrotra, S. Frameworks and results in distributionally robust optimization. *Open Journal of Mathematical Optimization* **2022**, *3*, 1–85.
22. Xu, C.; Lee, J.; Cheng, X.; Xie, Y. Flow-based distributionally robust optimization. *IEEE Journal on Selected Areas in Information Theory* **2024**, *5*, 62 – 77.
23. Hu, Z.; Hong, L.J. Kullback-Leibler divergence constrained distributionally robust optimization. *Optimization Online* **2013**, *1*, 9.
24. Radon, J. *Theorie und Anwendungen der absolut additiven Mengenfunktionen*, first ed.; Hölder: Vienna, Austria, 1913.
25. Nikodym, O. Sur une généralisation des intégrales de MJ Radon. *Fundamenta Mathematicae* **1930**, *15*, 131–179.
26. Aminian, G.; Bu, Y.; Toni, L.; Rodrigues, M.; Wornell, G. An Exact Characterization of the Generalization Error for the Gibbs Algorithm. *Advances in Neural Information Processing Systems* **2021**, *34*, 8106–8118.
27. Perlaza, S.M.; Bisson, G.; Esnaola, I.; Jean-Marie, A.; Rini, S. Empirical Risk Minimization with Relative Entropy Regularization: Optimality and Sensitivity. In Proceedings of the Proceedings of the IEEE International Symposium on Information Theory (ISIT), Espoo, Finland, Jul. 2022; pp. 684–689.
28. Jiang, W.; Tanner, M.A. Gibbs posterior for variable selection in high-dimensional classification and data mining. *The Annals of Statistics* **2008**, *36*, 2207–2231.
29. Perlaza, S.M.; Esnaola, I.; Bisson, G.; Poor, H.V. On the Validation of Gibbs Algorithms: Training Datasets, Test Datasets and their Aggregation. In Proceedings of the Proceedings of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun. 2023.
30. Alquier, P.; Ridgway, J.; Chopin, N. On the properties of variational approximations of Gibbs posteriors. *Journal of Machine Learning Research* **2016**, *17*, 8374–8414.
31. Bu, Y.; Aminian, G.; Toni, L.; Wornell, G.W.; Rodrigues, M. Characterizing and understanding the generalization error of transfer learning with Gibbs algorithm. In Proceedings of the Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS), Virtual Conference, Mar. 2022; pp. 8673–8699.
32. Raginsky, M.; Rakhlin, A.; Tsao, M.; Wu, Y.; Xu, A. Information-theoretic analysis of stability and bias of learning algorithms. In Proceedings of the Proceedings of the IEEE Information Theory Workshop (ITW), Cambridge, UK, Sep. 2016; pp. 26–30.

 doi:10.20944/preprints202507.2126.v1

33. Zou, B.; Li, L.; Xu, Z. The Generalization Performance of ERM algorithm with Strongly Mixing Observations. *Machine Learning* **2009**, *75*, 275–295.
34. He, H.; Aminian, G.; Bu, Y.; Rodrigues, M.; Tan, V.Y. How Does Pseudo-Labeling Affect the Generalization Error of the Semi-Supervised Gibbs Algorithm? In Proceedings of the Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS), Valencia, Spain, Apr. 2023; pp. 8494–8520.
35. Hellström, F.; Durisi, G.; Guedj, B.; Raginsky, M. Generalization Bounds: Perspectives from Information Theory and PAC-Bayes. *Foundations and Trends® in Machine Learning* **2025**, *18*, 1–223.
36. Jaynes, E.T. Information Theory and Statistical Mechanics I. *Physical Review Journals* **1957**, *106*, 620–630.
37. Jaynes, E.T. Information Theory and Statistical Mechanics II. *Physical Review Journals* **1957**, *108*, 171–190.
38. Kapur, J.N. *Maximum Entropy Models in Science and Engineering*, first ed.; Wiley: New York, NY, USA, 1989.
39. Bermudez, Y.; Bisson, G.; Esnaola, I.; Perlaza, S.M. Proofs for Folklore Theorems on the Radon-Nikodym Derivative. Technical Report RR-9591, INRIA, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France, 2025.
40. Donsker, M.D.; Varadhan, S.S. Asymptotic evaluation of certain Markov process expectations for large time, I. *Communications on pure and applied mathematics* **1975**, *28*, 1–47.
41. Heath, T.L. *The Thirteen Books of Euclid's Elements*, 2nd revised edition ed.; Dover Publications, Inc.: New York, 1956.