# Variations of the Expectation due to Changes in the Measure

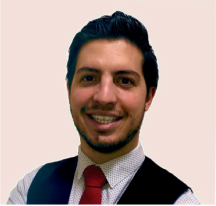
Applications to **Generalization** and **Game Theory**

**Samir M. Perlaza**
samir.perlaza@inria.fr
INRIA, Centre Inria d'Université Côte d'Azur

**Information Theory Seminar**

University of Cambridge
February 26, 2025, Cambridge, UK

**Xinying Zou**
INRIA
France



**Francisco Daunas**
*The University of Sheffield*,
UK

**Yaiza Bermudez**
INRIA
France

**Iñaki Esnaola**
*The University of Sheffield*,
UK

**H. Vincent Poor**
*Princeton University*,
United States

**Gaetan Bisson**
*University of French Polynesia*,
French Polynesia


Inria

# Table of Contents

# Table of Contents

**Comments on Notation:**

- The Borel $\sigma$-algebra on $\mathbb{R}^m$ is denoted by $\mathscr{B}(\mathbb{R}^m)$
- The set of all **probability measures** on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ is denoted by $\triangle(\mathbb{R}^m)$.

**Comments on Notation:**

- The Borel $\sigma$-algebra on $\mathbb{R}^m$ is denoted by $\mathscr{B}(\mathbb{R}^m)$
- The set of all **probability measures** on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ is denoted by $\triangle(\mathbb{R}^m)$.

- Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a **Borel measurable function**.

**Comments on Notation:**

- The Borel $\sigma$-algebra on $\mathbb{R}^m$ is denoted by $\mathscr{B}(\mathbb{R}^m)$
- The set of all **probability measures** on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ is denoted by $\triangle(\mathbb{R}^m)$.

- Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a **Borel measurable function**.
- Given two measures $P_1$ and $P_2$ in $\triangle(\mathbb{R}^m)$, assume that for all $i \in \{1, 2\}$,

$$\int |h(x, y)| \, \mathrm{d}P_i(y) < +\infty, \text{ for some fixed } x.$$

**Comments on Notation:**

- The Borel $\sigma$-algebra on $\mathbb{R}^m$ is denoted by $\mathscr{B}(\mathbb{R}^m)$
- The set of all **probability measures** on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ is denoted by $\triangle(\mathbb{R}^m)$.

- Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a **Borel measurable function**.
- Given two measures $P_1$ and $P_2$ in $\triangle(\mathbb{R}^m)$, assume that for all $i \in \{1, 2\}$,

$$\int |h(x, y)| \,\mathrm{d}P_i(y) < +\infty, \text{ for some fixed } x.$$

- Let $\mathsf{G}_h : \mathbb{R}^n \times \triangle(\mathbb{R}^m) \times \triangle(\mathbb{R}^m) \to \mathbb{R}$ be a functional such that

$$\mathsf{G}_h(x, P_1, P_2) = \int h(x, y)\mathrm{d}P_1(y) - \int h(x, y)\mathrm{d}P_2(y).$$

**Comments on Notation:**

- The Borel $\sigma$-algebra on $\mathbb{R}^m$ is denoted by $\mathscr{B}(\mathbb{R}^m)$
- The set of all **probability measures** on the measurable space $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ is denoted by $\triangle(\mathbb{R}^m)$.

- Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a **Borel measurable function**.
- Given two measures $P_1$ and $P_2$ in $\triangle(\mathbb{R}^m)$, assume that for all $i \in \{1, 2\}$,

$$\int |h(x, y)| \, \mathrm{d}P_i(y) < +\infty, \text{ for some fixed } x.$$

- Let $\mathsf{G}_h : \mathbb{R}^n \times \triangle(\mathbb{R}^m) \times \triangle(\mathbb{R}^m) \to \mathbb{R}$ be a functional such that

$$\mathsf{G}_h(x, P_1, P_2) = \int h(x, y) \mathrm{d}P_1(y) - \int h(x, y) \mathrm{d}P_2(y).$$

The value $\mathsf{G}_h(x, P_1, P_2)$ characterizes the **variation of the expectation of** $h(x, Y)$, when $Y \sim P_2$ changes to $Y \sim P_1$.

# Preliminaries

## Variation of the Expectation due to Changes in the Measure

### Definition

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

$$x \mapsto P_{Y|X=x}(\mathcal{A}) \quad \text{is Borel measurable.}$$

### Definition

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

$$x \mapsto P_{Y|X=x}(\mathcal{A}) \quad \text{is Borel measurable.}$$

### Comments on **Notation:**

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

**Definition**

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

$$x \mapsto P_{Y|X=x}(\mathcal{A}) \quad \text{is Borel measurable.}$$

**Comments on Notation:**

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

- Let $\bar{\mathsf{G}}_h : \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^n) \to \mathbb{R}$ be a functional such that

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

## Definition

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

$$x \mapsto P_{Y|X=x}(\mathcal{A}) \quad \text{is Borel measurable.}$$

## Comments on **Notation:**

- The set of all such conditional probability measures is denoted by $\triangle(\mathbb{R}^m|\mathbb{R}^n)$.
- Let $\bar{\mathsf{G}}_h : \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^m|\mathbb{R}^n) \times \triangle(\mathbb{R}^n) \to \mathbb{R}$ be a functional such that

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

The value $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ characterizes the **variation of the expectation of** $h(X,Y)$, when $(X,Y) \sim P^{(2)}_{Y|X}P_X$ changes to $Y \sim P^{(1)}_{Y|X}, P_X$.

**Special Case:**

**Special Case:**

- Given $P_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$ and $P_X \in \triangle(\mathbb{R}^n)$, let $P_Y \in \triangle(\mathbb{R}^n)$ be such that for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$,

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

**Special Case:**

- Given $P_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$ and $P_X \in \triangle(\mathbb{R}^n)$, let $P_Y \in \triangle(\mathbb{R}^n)$ be such that for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$,

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

-

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

**Special Case:**

- Given $P_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$ and $P_X \in \triangle(\mathbb{R}^n)$, let $P_Y \in \triangle(\mathbb{R}^n)$ be such that for all sets $\mathcal{A} \in \mathscr{B}(\mathbb{R}^m)$,

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

-

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

... **from the joint probability measure** to the **product of the marginals**.

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ and some $x \in \mathbb{R}^n$:

- Denote by $\mathsf{K}_{h,Q,x} : \mathbb{R} \to \mathbb{R}$ the function that satisfies

$$\mathsf{K}_{h,Q,x}(t) = \log\left(\int \exp\left(t\, h(x,y)\right) \mathrm{d}Q(y)\right).$$

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ and some $x \in \mathbb{R}^n$:

- Denote by $\mathrm{K}_{h,Q,x} : \mathbb{R} \to \mathbb{R}$ the function that satisfies

$$\mathrm{K}_{h,Q,x}(t) = \log\left(\int \exp\left(t\, h(x,y)\right) \mathrm{d}Q(y)\right).$$

If $Q$ is a probability measure, $\mathrm{K}_{h,Q,x}$ is the **cumulant generating function** of $h(x, Y)$, with $Y \sim Q$.

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ and some $x \in \mathbb{R}^n$:

- Denote by $\mathsf{K}_{h,Q,x} : \mathbb{R} \to \mathbb{R}$ the function that satisfies

$$\mathsf{K}_{h,Q,x}(t) = \log\left(\int \exp\left(t\, h(x,y)\right) \mathrm{d}Q(y)\right).$$

If $Q$ is a probability measure, $\mathsf{K}_{h,Q,x}$ is the **cumulant generating function** of $h(x, Y)$, with $Y \sim Q$.

**Definition (Gibbs Conditional Probability Measure)**

The probability measure $P^{(h,Q,\lambda)}_{Y|X} \in \triangle(\mathbb{R}^m|\mathbb{R}^n)$, with $\lambda \in \mathbb{R}$, is said to be an $(h, Q, \lambda)$-**Gibbs conditional probability measure** if for all $x \in \mathbb{R}^n$,

$$\mathsf{K}_{h,Q,x}(-\lambda) < +\infty;$$

and for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$,

$$\frac{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}{\mathrm{d}Q}(y) = \exp\left(-\lambda h(x,y) - \mathsf{K}_{h,Q,x}(-\lambda)\right).$$

# Preliminaries

## Relative Entropy Regularization and the Gibbs Probability Measure

**Comments on Notation:**

$$\triangle_Q(\mathbb{R}^m) \triangleq \{P \in \triangle(\mathbb{R}^m) : P \ll Q\}$$

**Comments on Notation:**

$$\triangle_Q(\mathbb{R}^m) \triangleq \{P \in \triangle(\mathbb{R}^m) : P \ll Q\}$$

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$; a real $\lambda$; and some $x \in \mathbb{R}^n$:

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

$$\max_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x,y)\mathrm{d}P(y) + \frac{1}{\lambda}\mathrm{D}(P\|Q). \quad (2)$$

# Preliminaries

## Relative Entropy Regularization and the Gibbs Probability Measure

**Comments on Notation:**

$$\triangle_Q(\mathbb{R}^m) \triangleq \{P \in \triangle(\mathbb{R}^m) : P \ll Q\}$$

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$; a real $\lambda$; and some $x \in \mathbb{R}^n$:

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

$$\max_{P \in \triangle_Q(\mathbb{R}^m)} \int h(x, y) \mathrm{d}P(y) + \frac{1}{\lambda} \mathrm{D}(P \| Q). \quad (2)$$

*Lemma (Lemma 1 in [1])*

*Assume that the optimization problem in (1) (respectively, in (2)) admits solutions. Then, if $\lambda > 0$ (respectively, if $\lambda < 0$), the Gibbs probability measure $P^{(h,Q,\lambda)}_{Y|X=x}$ is the unique solution.*

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

# Preliminaries

## Relative Entropy Regularization and the Gibbs Probability Measure

*Lemma (Lemma 2 in [1])*

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

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

*Moreover, if* $\lambda > 0$*,*

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

*Alternatively, if* $\lambda < 0$*,*

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

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

# Variations of the Expectation due to **Deviations from the Gibbs Probability Measure**

*Lemma (Lemma 3 in [1])*

*Consider an $(h, Q, \lambda)$-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 $P \in \triangle_Q(\mathbb{R}^m)$,*

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

# Variations of the Expectation due to **Deviations from the Gibbs Probability Measure**

**Lemma (Lemma 3 in [1])**

*Consider an $(h, Q, \lambda)$-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 $P \in \triangle_Q(\mathbb{R}^m)$,*

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

*Proof:* The proof follows by noticing that for all $P \in \triangle_Q(\mathbb{R}^m)$,

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

■

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.
[2] Yaiza Bermudez, Samir M. Perlaza, Gaetan Bisson, and Iñaki Esnaola. "**Proofs for Folklore Theorems on the Radon-Nikodym Derivative**". arXiv preprint arXiv:2501.18374.

# Variations of the Expectation due to **Changes in the Probability Measure**

Characterization of $\mathsf{G}_h\left(x, P_1, P_2\right)$

Key Observation:

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

Key Observation:

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

*Theorem (Theorem 4 in [1])*

*For all probability measures $P_1$ and $P_2$ in $\triangle_Q(\mathbb{R}^m)$, with $Q$ a $\sigma$-finite measure,*

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

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

Key Observation:

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

**Theorem (Theorem 4 in [1])**

*For all probability measures $P_1$ and $P_2$ in $\triangle_Q(\mathbb{R}^m)$, with $Q$ a $\sigma$-finite measure,*

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

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

- $Q$ can be $P_1$, if $P_2 \ll P_1$; or $P_2$, if $P_1 \ll P_2$
- $Q$ can be the Lebesgue measure, if $P_1$ and $P_2$ has a **probability density function**
- $Q$ can be the counting measure, if $P_1$ and $P_2$ has a **probability mass function**

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

### *Corollary (Corollary 5 in [1])*

*Consider the variation* $\mathsf{G}_h(x, P_1, P_2)$. *If* $P_1 \ll P_2$, *then,*

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

*Alternatively, if* $P_2 \ll P_1$, *then,*

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

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

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

### *Theorem (Theorem 6 in [1])*

*Consider the variation $\bar{\mathsf{G}}_h\left(P^{(1)}_{Y|X}, P^{(2)}_{Y|X}, P_X\right)$ and assume that for all $x \in \mathbb{R}^n$, $P^{(1)}_{Y|X=x} \ll Q$ and $P^{(2)}_{Y|X=x} \ll Q$, with $Q$ a $\sigma$-measure. Then,*

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

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

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

**Mutual information**: $I\left(P_{Y|X}; P_X\right) \triangleq \int \mathrm{D}\left(P_{Y|X=x} \| P_Y\right) \mathrm{d}P_X(x)$; and

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

# Variations of the Expectation due to **Changes in the Probability Measure**



**Mutual information**: $I\left(P_{Y|X}; P_X\right) \triangleq \int \mathrm{D}\left(P_{Y|X=x} \| P_Y\right) \mathrm{d}P_X(x)$; and

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

## *Theorem (Theorem 6 in [1])*

*Consider the expected variation $\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right)$ and assume that, for all $x \in \mathbb{R}^n$:*

(a) $P_Y \ll Q$ *and* $P_{Y|X=x} \ll Q$, *with* $Q$ *a given* $\sigma$*-finite measure; and*

(b) $P_Y \ll\gg P_{Y|X=x}$.

*Then, it follows that*

$$\bar{\mathsf{G}}_h\left(P_Y, P_{Y|X}, P_X\right) = \frac{1}{\lambda}\Bigg( I\left(P_{Y|X}; P_X\right) + L\left(P_{Y|X}; P_X\right) + \int\int \log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right) \mathrm{d}P_Y(y)\mathrm{d}P_X(x) - \int\int \log\left(\frac{\mathrm{d}P_{Y|X=x}}{\mathrm{d}P^{(h,Q,\lambda)}_{Y|X=x}}(y)\right) \mathrm{d}P_{Y|X=x}(y)\mathrm{d}P_X(x)\Bigg),$$

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

### *Corollary (Corollary 9 in [1])*

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

$$P_Y^{(h,Q,\lambda)}\left(\mathcal{A}\right) = \int P_{Y|X=x}^{(h,Q,\lambda)}\left(\mathcal{A}\right) \mathrm{d}P_X(x).$$

*Then,*

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

[1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.

# What is so special about the **Gibbs probability measure**?

Consider a $\sigma$-**finite measure** $Q$ over $(\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$; some $\lambda \in \mathbb{R} \setminus \{0\}$; some $x \in \mathbb{R}^n$; and a convex function $f : [0, +\infty) \to \mathbb{R}$, with $f(1) = 0$.

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

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

**Nothing special:** The solutions to the optimization problems above possess the same properties! [3]

[3] Francisco Daunas, Iñaki Esnaola, Samir M. Perlaza, and Gholamali Aminian. "**Generalization Error of $f$-Divergence Stabilized Algorithms via Duality**". arXiv preprint arXiv:2502.14544.

# Table of Contents

Consider the following **supervised learning** setting:

- Set $\mathcal{X}$ of **patterns**, set $\mathcal{Y}$ of **labels**, and set $\mathcal{M} \subset \mathbb{R}^d$ of **models** with $d \in \mathbb{N}$.

Consider the following **supervised learning** setting:

- Set $\mathcal{X}$ of **patterns**, set $\mathcal{Y}$ of **labels**, and set $\mathcal{M} \subset \mathbb{R}^d$ of **models** with $d \in \mathbb{N}$.
- **Hypothesis class:** A function $h : \mathcal{M} \times \mathcal{X} \to \mathcal{Y}$.

Consider the following **supervised learning** setting:

- Set $\mathcal{X}$ of **patterns**, set $\mathcal{Y}$ of **labels**, and set $\mathcal{M} \subset \mathbb{R}^d$ of **models** with $d \in \mathbb{N}$.
- **Hypothesis class:** A function $h : \mathcal{M} \times \mathcal{X} \to \mathcal{Y}$.

**Definition (Risk Function)**

Given a **data point** $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the model $\boldsymbol{\theta} \in \mathcal{M}$ induces the *risk* $\ell\left(h(\boldsymbol{\theta}, x), y\right)$, where $\ell : \mathcal{Y} \times \mathcal{Y} \to [0, +\infty)$ is a **risk function.**

Consider the following **supervised learning** setting:

- Set $\mathcal{X}$ of **patterns**, set $\mathcal{Y}$ of **labels**, and set $\mathcal{M} \subset \mathbb{R}^d$ of **models** with $d \in \mathbb{N}$.
- **Hypothesis class:** A function $h : \mathcal{M} \times \mathcal{X} \to \mathcal{Y}$.

### Definition (Risk Function)

Given a **data point** $(x, y) \in \mathcal{X} \times \mathcal{Y}$, the model $\boldsymbol{\theta} \in \mathcal{M}$ induces the *risk* $\ell\left(h(\boldsymbol{\theta}, x), y\right)$, where $\ell : \mathcal{Y} \times \mathcal{Y} \to [0, +\infty)$ is a **risk function.**

### Definition (Empirical Risk)

Given the dataset $\boldsymbol{z} = \left((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\right) \in (\mathcal{X} \times \mathcal{Y})^n$, the *empirical risk* induced by the model $\boldsymbol{\theta} \in \mathcal{M}$ is

$$\mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) = \frac{1}{n} \sum_{i=1}^{n} \ell\left(h(\boldsymbol{\theta}, x_i), y_i\right).$$

## Generalization Error of Machine Learning Algorithms

**Definition (Expected Empirical Risk)**

Let the function $\mathsf{R}_{\boldsymbol{z}} : \triangle(\mathcal{M}, \mathscr{B}(\mathcal{M})) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{z}}(P) = \int \mathsf{L}(\boldsymbol{z}, \boldsymbol{\theta}) \, \mathrm{d}P(\boldsymbol{\theta}).$$

Let the function $\mathsf{R}_{\boldsymbol{\theta}} : \triangle(\mathcal{X} \times \mathcal{Y}) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{\theta}}(Q) = \int \ell(h(\boldsymbol{\theta}, x), y) \, \mathrm{d}Q(x, y).$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

# Generalization Error of Machine Learning Algorithms

## Definition (Expected Empirical Risk)

Let the function $\mathsf{R}_{\boldsymbol{z}} : \triangle\left(\mathcal{M}, \mathscr{B}\left(\mathcal{M}\right)\right) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{z}}\left(P\right) = \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P(\boldsymbol{\theta}).$$

Let the function $\mathsf{R}_{\boldsymbol{\theta}} : \triangle\left(\mathcal{X} \times \mathcal{Y}\right) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{\theta}}\left(Q\right) = \int \ell\left(h(\boldsymbol{\theta}, x), y\right) \mathrm{d}Q(x, y).$$

## Definition (Generalization Gap)

The **generalization gap** induced by the algorithm $P_{\boldsymbol{\Theta}|\boldsymbol{Z}}$, with training and test datasets $\boldsymbol{z}$ and $\boldsymbol{u}$ is

$$\underbrace{\mathsf{R}_{\boldsymbol{u}}\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}\right)}_{\text{Test Expected Risk}} - \underbrace{\mathsf{R}_{\boldsymbol{z}}\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}\right)}_{\text{Training Expected Risk}}$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

# Generalization Error of Machine Learning Algorithms

## Definition (Expected Empirical Risk)

Let the function $\mathsf{R}_{\boldsymbol{z}} : \triangle(\mathcal{M}, \mathscr{B}(\mathcal{M})) \to [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{z}}(P) = \int \mathsf{L}(\boldsymbol{z}, \boldsymbol{\theta}) \, \mathrm{d}P(\boldsymbol{\theta}).$$

Let the function $\mathsf{R}_{\boldsymbol{\theta}} : \triangle(\mathcal{X} \times \mathcal{Y}) \to [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{\theta}}(Q) = \int \ell(h(\boldsymbol{\theta}, x), y) \, \mathrm{d}Q(x, y).$$

## Assumption:

**Training datasets** $\boldsymbol{z} = \left((x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\right) \in (\mathcal{X} \times \mathcal{Y})^n$ and **test datasets** $\boldsymbol{u} = \left((\mu_1, \nu_1), (\mu_2, \nu_2), \ldots, (\mu_n, \nu_n)\right) \in (\mathcal{X} \times \mathcal{Y})^n$ are **i.i.d** according to $P_{\boldsymbol{Z}} \in \triangle((\mathcal{X} \times \mathcal{Y})^n)$.

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

# Generalization Error of Machine Learning Algorithms

## Definition (Expected Empirical Risk)

Let the function $\mathsf{R}_{\boldsymbol{z}} : \triangle\left(\mathcal{M}, \mathscr{B}\left(\mathcal{M}\right)\right) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{z}}\left(P\right) = \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P(\boldsymbol{\theta}).$$

Let the function $\mathsf{R}_{\boldsymbol{\theta}} : \triangle\left(\mathcal{X} \times \mathcal{Y}\right) \rightarrow [0, +\infty)$ be such that

$$\mathsf{R}_{\boldsymbol{\theta}}\left(Q\right) = \int \ell\left(h(\boldsymbol{\theta}, x), y\right) \mathrm{d}Q(x, y).$$

## Generalization Error

The generalization error of the algorithm $P_{\Theta|\boldsymbol{Z}}$ is

$$\overline{\overline{G}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right) \triangleq \int \int \left(\mathsf{R}_{\boldsymbol{u}}\left(P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}\right) - \mathsf{R}_{\boldsymbol{z}}\left(P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}\right)\right) \mathrm{d}P_{\boldsymbol{Z}}\left(\boldsymbol{u}\right) \mathrm{d}P_{\boldsymbol{Z}}\left(\boldsymbol{z}\right).$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

Challenge: Apparently, **generalization error is not** a variation of an expectation

# Generalization Error as a **Variation of an Expectation**

*Lemma (Lemma 3 in [4])*

*The generalization error* $\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right)$ *satisfies*

$$\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right) = \int \left( \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right) - \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta|Z=\boldsymbol{z}}\left(\boldsymbol{\theta}\right) \right) \mathrm{d}P_Z\left(\boldsymbol{z}\right),$$

*where for all measurable subsets* $\mathcal{C}$ *of* $\mathcal{M}$, $P_{\Theta}\left(\mathcal{C}\right) = \int P_{\Theta|Z=\boldsymbol{z}}\left(\mathcal{C}\right) \mathrm{d}P_Z\left(\boldsymbol{z}\right)$.

# Generalization Error as a **Variation of an Expectation**

*Lemma (Lemma 3 in [4])*

*The generalization error* $\overline{\overline{G}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right)$ *satisfies*

$$\overline{\overline{G}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right) = \int \left( \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right) - \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}\left(\boldsymbol{\theta}\right) \right) \mathrm{d}P_{\boldsymbol{Z}}\left(\boldsymbol{z}\right),$$

*where for all measurable subsets* $\mathcal{C}$ *of* $\mathcal{M}$, $P_{\Theta}\left(\mathcal{C}\right) = \int P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}\left(\mathcal{C}\right) \mathrm{d}P_{\boldsymbol{Z}}\left(\boldsymbol{z}\right)$.

**Problem** 1: Empirical Risk Minimization with Relative Entropy Regularization [5]

$$\min_{P \in \triangle_Q(\mathcal{M})} \quad \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P(\boldsymbol{\theta}) + \lambda \mathsf{D}\left(P \| Q\right).$$

# Generalization Error as a **Variation of an Expectation**

*Lemma (Lemma 3 in [4])*

*The generalization error* $\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right)$ *satisfies*

$$\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right) = \int \left( \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right) - \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta|Z=\boldsymbol{z}}\left(\boldsymbol{\theta}\right) \right) \mathrm{d}P_Z\left(\boldsymbol{z}\right),$$

*where for all measurable subsets* $\mathcal{C}$ *of* $\mathcal{M}$, $P_{\Theta}\left(\mathcal{C}\right) = \int P_{\Theta|Z=\boldsymbol{z}}\left(\mathcal{C}\right) \mathrm{d}P_Z\left(\boldsymbol{z}\right)$.

**Problem** 1: Empirical Risk Minimization with Relative Entropy Regularization [5]

$$\min_{P \in \triangle_Q(\mathcal{M})} \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P(\boldsymbol{\theta}) + \lambda \mathrm{D}\left(P \| Q\right).$$

*Theorem (Theorem 3 in [5])*

*The solution to* **Problem** 1 *is unique, denoted by* $P^{(Q,\lambda)}_{\Theta|Z=\boldsymbol{z}}$, *and satisfies for all* $\boldsymbol{\theta} \in \operatorname{supp} Q$,

$$\frac{\mathrm{d}P^{(Q,\lambda)}_{\Theta|Z=\boldsymbol{z}}}{\mathrm{d}Q}\left(\boldsymbol{\theta}\right) = \exp\left(-K_{Q,\boldsymbol{z}}\left(-\frac{1}{\lambda}\right) - \frac{1}{\lambda}\mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right)\right).$$

# Generalization Error as a **Variation of an Expectation**

## *Lemma (Lemma 3 in [4])*

*The generalization error* $\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right)$ *satisfies*

$$\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right) = \int \left( \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right) - \int \mathsf{L}\left(\boldsymbol{z}, \boldsymbol{\theta}\right) \mathrm{d}P_{\Theta|Z=\boldsymbol{z}}\left(\boldsymbol{\theta}\right) \right) \mathrm{d}P_Z\left(\boldsymbol{z}\right),$$

*where for all measurable subsets* $\mathcal{C}$ *of* $\mathcal{M}$, $P_{\Theta}\left(\mathcal{C}\right) = \int P_{\Theta|Z=\boldsymbol{z}}\left(\mathcal{C}\right) \mathrm{d}P_Z\left(\boldsymbol{z}\right)$.

## *Lemma (Lemma 4 in [4])*

*If for all* $\boldsymbol{z} \in \left(\mathcal{X} \times \mathcal{Y}\right)^n$, *the measures* $P_{\Theta|Z=\boldsymbol{z}}$ *and* $P_{\Theta}$ *are absolutely continuous w.r.t. the* $\sigma$*-finite measure* $Q$,

$$\overline{\overline{G}}\left(P_{\Theta|Z}, P_Z\right) = \lambda \int \left( \mathsf{D}\left(P_{\Theta} \| P^{(Q,\lambda)}_{\Theta|Z=\boldsymbol{z}}\right) - \mathsf{D}\left(P_{\Theta|Z=\boldsymbol{z}} \| P^{(Q,\lambda)}_{\Theta|Z=\boldsymbol{z}}\right) + \mathsf{D}\left(P_{\Theta|Z=\boldsymbol{z}} \| Q\right) - \mathsf{D}\left(P_{\Theta} \| Q\right) \right) \mathrm{d}P_Z\left(\boldsymbol{z}\right).$$

# Exact Characterizations of the Generalization Error

*Theorem (Theorem 14 in [4])*

*Assume that for all* $\boldsymbol{z} \in (\mathcal{X} \times \mathcal{Y})^n$*:*

(a) *The probability measures* $P_{\Theta}$ *and* $P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}$ *are both absolutely continuous w.r.t.* $Q \in \triangle(\mathcal{M})$*;*

(b) *The measure* $Q$ *is absolutely continuous with respect to* $P_{\Theta}$*; and*

(c) *The measure* $P_{\Theta}$ *is absolutely continuous with respect to* $P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}$*.*

*Then,*

$$\overline{\overline{G}}(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = \lambda \left( I\left(P_{\Theta|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right) + L\left(P_{\Theta|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right) \right) + \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}}(\boldsymbol{\theta}) \mathrm{d}P_{\Theta}(\boldsymbol{\theta}) \mathrm{d}P_{\boldsymbol{Z}}(\boldsymbol{z}) - \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}}(\boldsymbol{\theta}) \mathrm{d}P_{\Theta|\boldsymbol{Z}=\boldsymbol{z}}(\boldsymbol{\theta}) \mathrm{d}P_{\boldsymbol{Z}}(\boldsymbol{z}).$$

# Exact Characterizations of the Generalization Error

**Theorem (Theorem 14 in [4])**

*Assume that for all $\mathbf{z} \in (\mathcal{X} \times \mathcal{Y})^n$:*

(a) *The probability measures $P_{\Theta}$ and $P_{\Theta|\mathbf{Z}=\mathbf{z}}$ are both absolutely continuous w.r.t. $Q \in \triangle(\mathcal{M})$;*

(b) *The measure $Q$ is absolutely continuous with respect to $P_{\Theta}$; and*

(c) *The measure $P_{\Theta}$ is absolutely continuous with respect to $P_{\Theta|\mathbf{Z}=\mathbf{z}}$.*

*Then,*

$$\overline{\overline{G}}(P_{\Theta|\mathbf{Z}}, P_{\mathbf{Z}}) = \lambda \left( I\left(P_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right) + L\left(P_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right)\right) + \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\mathrm{d}P_{\Theta}(\boldsymbol{\theta})\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}(\boldsymbol{\theta})\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}).$$

**What if...**

$$\lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\,\mathrm{d}P_{\Theta}(\boldsymbol{\theta})\,\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\,\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}(\boldsymbol{\theta})\,\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) = 0.$$

# Exact Characterizations of the Generalization Error

**Theorem (Theorem 14 in [4])**

*Assume that for all* $\mathbf{z} \in (\mathcal{X} \times \mathcal{Y})^n$*:*

(a) *The probability measures* $P_{\Theta}$ *and* $P_{\Theta|\mathbf{Z}=\mathbf{z}}$ *are both absolutely continuous w.r.t.* $Q \in \triangle(\mathcal{M})$*;*

(b) *The measure* $Q$ *is absolutely continuous with respect to* $P_{\Theta}$*; and*

(c) *The measure* $P_{\Theta}$ *is absolutely continuous with respect to* $P_{\Theta|\mathbf{Z}=\mathbf{z}}$*.*

*Then,*

$$\overline{\overline{G}}(P_{\Theta|\mathbf{Z}}, P_{\mathbf{Z}}) = \lambda \left( I\left(P_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right) + L\left(P_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right)\right) + \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\mathrm{d}P_{\Theta}(\boldsymbol{\theta})\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}) - \lambda \int\int \log \frac{\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\Theta|\mathbf{Z}=\mathbf{z}}}(\boldsymbol{\theta})\mathrm{d}P_{\Theta|\mathbf{Z}=\mathbf{z}}(\boldsymbol{\theta})\mathrm{d}P_{\mathbf{Z}}(\mathbf{z}).$$

**Corollary (Theorem 1 in [7])**

$$\overline{\overline{G}}(P^{(Q,\lambda)}_{\Theta|\mathbf{Z}}, P_{\mathbf{Z}}) = \lambda \left( I\left(P^{(Q,\lambda)}_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right) + L\left(P^{(Q,\lambda)}_{\Theta|\mathbf{Z}}; P_{\mathbf{Z}}\right)\right).$$

[7] Gholamali Aminian; Yuheng Bu; Laura Toni; Miguel R. D. Rodrigues; Gregory W. Wornell. "**Information-Theoretic Characterizations of Generalization Error for the Gibbs Algorithm**". IEEE Transactions on Information Theory, vol. 70, no. 1, pp. 632 - 655, Jan., 2024.

# Generalization Error as a **Variation of an Expectation**

## *Lemma (Lemma 6 in [4])*

*Assume that* $P_{\boldsymbol{Z}} \in \triangle\left((\mathcal{X} \times \mathcal{Y})^n\right)$ *is a* **product measure** *formed by* $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$*. Then, the generalization error* $\overline{\overline{\mathsf{G}}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right)$ *satisfies*

$$\overline{\overline{\mathsf{G}}}(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = \int \left( \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_{Z}\left(x, y\right) - \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_{Z|\Theta=\boldsymbol{\theta}}\left(x, y\right) \right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right).$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.
[6] Xinying Zou, Samir M. Perlaza, Iñaki Esnaola, Eitan Altman, and H. Vincent Poor. "**The Worst-Case Data-Generating Probability Measure in Statistical Learning**". IEEE Journal on Selected Areas in Information Theory, vol. 5, pp. 175–189, Apr., 2024.

# Generalization Error as a **Variation of an Expectation**

## *Lemma (Lemma 6 in [4])*

*Assume that* $P_{\boldsymbol{Z}} \in \triangle((\mathcal{X} \times \mathcal{Y})^n)$ *is a* **product measure** *formed by* $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$. *Then, the generalization error* $\overline{\overline{\mathsf{G}}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right)$ *satisfies*

$$\overline{\overline{\mathsf{G}}}(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = \int \left( \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_Z\left(x, y\right) - \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}\left(x, y\right) \right) \mathrm{d}P_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}\right).$$

## **Problem** 2: Loss Maximization with Relative Entropy Regularization [6]

$$\max_{P \in \triangle_{P_S}(\mathcal{X} \times \mathcal{Y})} \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P(x, y) - \beta \mathrm{D}\left(P \| P_S\right).$$


[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

[6] Xinying Zou, Samir M. Perlaza, Iñaki Esnaola, Eitan Altman, and H. Vincent Poor. "**The Worst-Case Data-Generating Probability Measure in Statistical Learning**". IEEE Journal on Selected Areas in Information Theory, vol. 5, pp. 175–189, Apr., 2024.

# Generalization Error as a **Variation of an Expectation**

*Lemma (Lemma 6 in [4])*

*Assume that* $P_Z \in \triangle((\mathcal{X} \times \mathcal{Y})^n)$ *is a* **product measure** *formed by* $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$. *Then, the generalization error* $\overline{\overline{\mathsf{G}}}\left(P_{\Theta|Z}, P_Z\right)$ *satisfies*

$$\overline{\overline{\mathsf{G}}}(P_{\Theta|Z}, P_Z) = \int \left( \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_Z\left(x, y\right) - \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_{Z|\Theta=\boldsymbol{\theta}}\left(x, y\right) \right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right).$$

**Problem** 2: Loss Maximization with Relative Entropy Regularization [6]

$$\max_{P \in \triangle_{P_S}(\mathcal{X} \times \mathcal{Y})} \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P(x, y) - \beta \mathrm{D}\left(P \| P_S\right).$$

*Theorem (Theorem 1 in [6]: Worst-Case Data-Generating Probability Measure)*

*The solution to* **Problem** 2 *is unique, denoted by* $P^{(P_S,\beta)}_{\hat{Z}|\Theta=\boldsymbol{\theta}}$, *and satisfies for all* $(x, y) \in \operatorname{supp} P_S$,

$$\frac{\mathrm{d}P^{(P_S,\beta)}_{\hat{Z}|\Theta=\boldsymbol{\theta}}}{\mathrm{d}P_S}(x, y) = \exp\left(\frac{1}{\beta}\ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) - \mathrm{J}_{P_S,\boldsymbol{\theta}}\left(\frac{1}{\beta}\right)\right).$$

# Generalization Error as a **Variation of an Expectation**

**Lemma (Lemma 6 in [4])**

*Assume that $P_{\boldsymbol{Z}} \in \triangle((\mathcal{X} \times \mathcal{Y})^n)$ is a* **product measure** *formed by $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$. Then, the generalization error $\overline{\overline{\mathsf{G}}}\left(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}\right)$ satisfies*

$$\overline{\overline{\mathsf{G}}}(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = \int \left( \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_Z\left(x, y\right) - \int \ell\left(h\left(\boldsymbol{\theta}, x\right), y\right) \mathrm{d}P_{Z|\Theta=\boldsymbol{\theta}}\left(x, y\right) \right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right).$$

**Lemma (Lemma 7 in [4])**

*If for all $\boldsymbol{\theta} \in \mathcal{M}$, the probability measures $P_Z$ and $P_{Z|\Theta=\boldsymbol{\theta}}$ are absolutely continuous w.r.t. the probability measure $P_S \in \triangle(\mathcal{X} \times \mathcal{Y})$,*

$$\overline{\overline{\mathsf{G}}}(P_{\Theta|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = \beta \int \left( \mathrm{D}\left(P_{Z|\Theta=\boldsymbol{\theta}} \| P^{(P_S,\beta)}_{\hat{Z}|\Theta=\boldsymbol{\theta}}\right) - \mathrm{D}\left(P_Z \| P^{(P_S,\beta)}_{\hat{Z}|\Theta=\boldsymbol{\theta}}\right) - \mathrm{D}\left(P_{Z|\Theta=\boldsymbol{\theta}} \| P_S\right) + \mathrm{D}\left(P_Z \| P_S\right) \right) \mathrm{d}P_{\Theta}\left(\boldsymbol{\theta}\right).$$

[4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.

# Exact Characterizations of the Generalization Error

*Theorem (Theorem 29 in [4])*

*Assume that for all $\boldsymbol{\theta} \in \mathcal{M}$:*

$(a)$ *The probability measure $P_{\boldsymbol{Z}}$ is a* **product** *measure formed by some $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$;*

$(b)$ *The probability measures $P_Z$ and $P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}$ are absolutely continuous w.r.t. $P_S$; and*

$(c)$ *The probability measures $P_Z$ and $P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}$ are mutually absolutely continuous.*

*Then,*

$$\overline{\overline{\mathsf{G}}}(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = -\frac{\beta}{\boldsymbol{n}}\left(I\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right) + L\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right)\right)$$

$$+\beta \int\int \log\left(\frac{\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}}{\mathrm{d}P^{(P_S,\beta)}_{\hat{Z}|\boldsymbol{\Theta}=\boldsymbol{\theta}}}(z)\right)\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}\left(z\right)\mathrm{d}P_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}\right) - \beta\int\int \log\left(\frac{\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}}{\mathrm{d}P^{(P_S,\beta)}_{\hat{Z}|\boldsymbol{\Theta}=\boldsymbol{\theta}}}(z)\right)\mathrm{d}P_Z\left(z\right)\mathrm{d}P_{\boldsymbol{\Theta}}\left(\boldsymbol{\theta}\right).$$

# Exact Characterizations of the Generalization Error

*Theorem (Theorem 29 in [4])*

*Assume that for all* $\boldsymbol{\theta} \in \mathcal{M}$*:*

(a) *The probability measure* $P_{\boldsymbol{Z}}$ *is a* **product** *measure formed by some* $P_Z \in \triangle(\mathcal{X} \times \mathcal{Y})$*;*

(b) *The probability measures* $P_Z$ *and* $P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}$ *are absolutely continuous w.r.t.* $P_S$*; and*

(c) *The probability measures* $P_Z$ *and* $P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}$ *are mutually absolutely continuous.*

*Then,*

$$\overline{\overline{\mathsf{G}}}(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}, P_{\boldsymbol{Z}}) = -\frac{\beta}{n}\left(I\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right) + L\left(P_{\boldsymbol{\Theta}|\boldsymbol{Z}}; P_{\boldsymbol{Z}}\right)\right)$$

$$+\beta \int\int \log\left(\frac{\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}}{\mathrm{d}P^{(P_S,\beta)}_{\hat{Z}|\boldsymbol{\Theta}=\boldsymbol{\theta}}}(z)\right)\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}(z)\,\mathrm{d}P_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) - \beta\int\int \log\left(\frac{\mathrm{d}P_{Z|\boldsymbol{\Theta}=\boldsymbol{\theta}}}{\mathrm{d}P^{(P_S,\beta)}_{\hat{Z}|\boldsymbol{\Theta}=\boldsymbol{\theta}}}(z)\right)\mathrm{d}P_Z(z)\,\mathrm{d}P_{\boldsymbol{\Theta}}(\boldsymbol{\theta}).$$

What if...

$$\lambda\int\int \log\frac{\mathrm{d}P_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}}(\boldsymbol{\theta})\,\mathrm{d}P_{\boldsymbol{\Theta}}(\boldsymbol{\theta})\,\mathrm{d}P_{\boldsymbol{Z}}(\boldsymbol{z}) - \lambda\int\int \log\frac{\mathrm{d}P_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}}{\mathrm{d}P^{(Q,\lambda)}_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}}(\boldsymbol{\theta})\,\mathrm{d}P_{\boldsymbol{\Theta}|\boldsymbol{Z}=\boldsymbol{z}}(\boldsymbol{\theta})\,\mathrm{d}P_{\boldsymbol{Z}}(\boldsymbol{z})=0.$$

# Table of Contents

# Zero-Sum Game Formulation

## Classical $m_1 \times m_2$ ZSGs in Normal Form

- **Actions** of Player $k$: $\mathcal{A}_k \triangleq \{a_{k,1}, a_{k,2}, \ldots, a_{k,m_k}\}$.

# Zero-Sum Game Formulation

## Classical $m_1 \times m_2$ ZSGs in Normal Form

- **Actions** of Player $k$: $\mathcal{A}_k \triangleq \{a_{k,1}, a_{k,2}, \ldots, a_{k,m_k}\}$.
- When Player 1 plays $a_{1,i}$ and Player 2 plays $a_{2,j}$, the **payoff** is $u_{i,j}$

$$\underline{\boldsymbol{u}} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m_2} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m_2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m_1,1} & u_{m_1,2} & \cdots & u_{m_1,m_2} \end{pmatrix}.$$

# Zero-Sum Game Formulation

## Classical $m_1 \times m_2$ ZSGs in Normal Form

- **Actions** of Player $k$: $\mathcal{A}_k \triangleq \{a_{k,1}, a_{k,2}, \dots, a_{k,m_k}\}$.
- When Player 1 plays $a_{1,i}$ and Player 2 plays $a_{2,j}$, the **payoff** is $u_{i,j}$

$$\underline{\boldsymbol{u}} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m_2} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m_2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m_1,1} & u_{m_1,2} & \cdots & u_{m_1,m_2} \end{pmatrix}.$$

- A **strategy** for Player $k$ is a probability measure $P_{A_k} \in \triangle(\mathcal{A}_k)$.

# Zero-Sum Game Formulation

## Classical $m_1 \times m_2$ ZSGs in Normal Form

- **Actions** of Player $k$: $\mathcal{A}_k \triangleq \{a_{k,1}, a_{k,2}, \ldots, a_{k,m_k}\}$.
- When Player 1 plays $a_{1,i}$ and Player 2 plays $a_{2,j}$, the **payoff** is $u_{i,j}$

$$\underline{\boldsymbol{u}} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m_2} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m_2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m_1,1} & u_{m_1,2} & \cdots & u_{m_1,m_2} \end{pmatrix}.$$

- A **strategy** for Player $k$ is a probability measure $P_{A_k} \in \triangle(\mathcal{A}_k)$.
- **Expected Payoff** determined by the function $u : \triangle(\mathcal{A}_1) \times \triangle(\mathcal{A}_2) \to \mathbb{R}$:

$$u(P_{A_1}, P_{A_2}) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} P_{A_1}(a_{1,i}) P_{A_2}(a_{2,j}) u_{i,j},$$

# Zero-Sum Game Formulation

## Classical $m_1 \times m_2$ ZSGs in Normal Form

- **Actions** of Player $k$: $\mathcal{A}_k \triangleq \{a_{k,1}, a_{k,2}, \ldots, a_{k,m_k}\}$.
- When Player 1 plays $a_{1,i}$ and Player 2 plays $a_{2,j}$, the **payoff** is $u_{i,j}$

$$\underline{\boldsymbol{u}} = \begin{pmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,m_2} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,m_2} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m_1,1} & u_{m_1,2} & \cdots & u_{m_1,m_2} \end{pmatrix}.$$

- A **strategy** for Player $k$ is a probability measure $P_{A_k} \in \triangle(\mathcal{A}_k)$.
- **Expected Payoff** determined by the function $u : \triangle(\mathcal{A}_1) \times \triangle(\mathcal{A}_2) \to \mathbb{R}$:

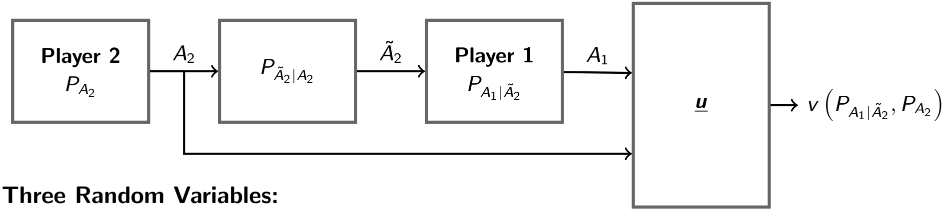
$$u(P_{A_1}, P_{A_2}) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} P_{A_1}(a_{1,i}) P_{A_2}(a_{2,j}) u_{i,j},$$

Player 1 **maximizes**, while Player 2 **minimizes** the payoff.

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:



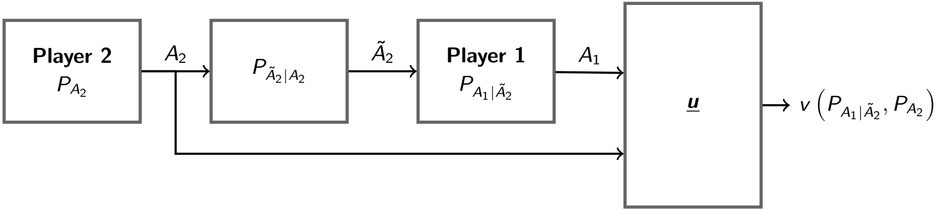

- **Three Random Variables:**
  - Action of Player 1: $A_1$
  - Action of Player 2: $A_2$
  - Noisy Observation of the Action of Player 2: $\tilde{A}_2$.
  - Joint Probability Distribution:

$$P_{A_1\tilde{A}_2A_2} = P_{A_1|\tilde{A}_2}P_{\tilde{A}_2|A_2}P_{A_2}.$$

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:



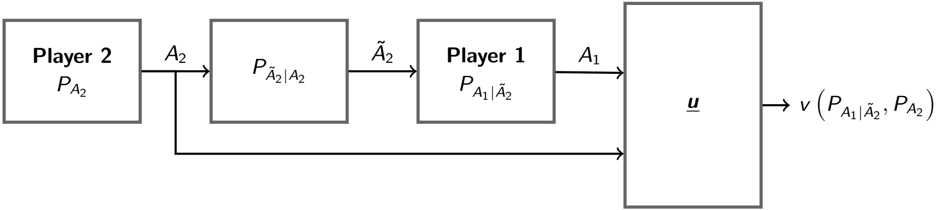

$$P_{A_1\tilde{A}_2A_2} = P_{A_1|\tilde{A}_2}P_{\tilde{A}_2|A_2}P_{A_2}.$$

- **Strategy of Player 1:** $P_{A_1|\tilde{A}_2} \in \triangle\left(\mathcal{A}_1|\tilde{\mathcal{A}}_2\right)$
- **Strategy of Player 2:** $P_{A_2} \in \triangle\left(\mathcal{A}_2\right)$
- **Channel** Output: $P_{\tilde{A}_2|A_2} \in \triangle\left(\tilde{\mathcal{A}}_2|\mathcal{A}_2\right)$

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:



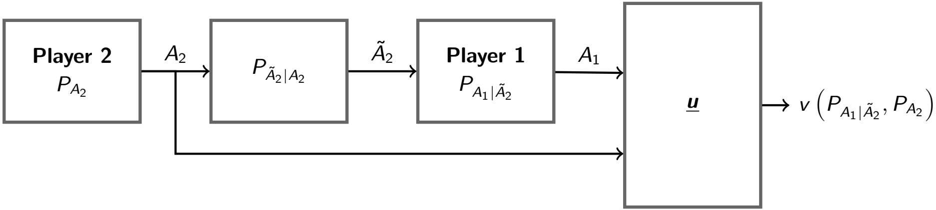

The **expected payoff** is determined by $v : \triangle\left(\mathcal{A}_1|\tilde{\mathcal{A}}_2\right) \times \triangle\left(\mathcal{A}_2\right) \rightarrow \mathbb{R}$,

$$v\left(P_{A_1|\tilde{A}_2}, P_{A_2}\right) = \int \left( \int \left( \int u_{a,b} \mathrm{d}P_{A_1|\tilde{A}_2=\tilde{b}}(a) \right) \mathrm{d}P_{\tilde{A}_2|A_2=b}(\tilde{b}) \right) \mathrm{d}P_{A_2}(b)$$

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:




The **expected payoff** is determined by $v : \triangle\left(\mathcal{A}_1|\tilde{\mathcal{A}}_2\right) \times \triangle\left(\mathcal{A}_2\right) \to \mathbb{R}$,

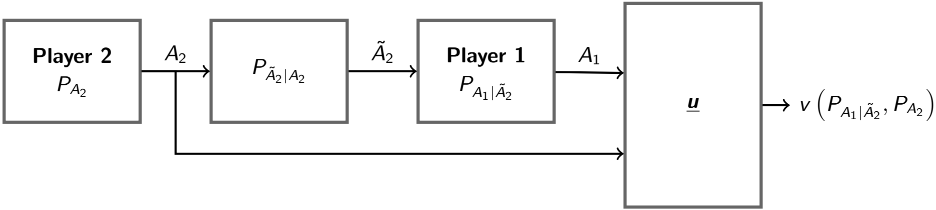
$$v\left(P_{A_1|\tilde{A}_2}, P_{A_2}\right) = \int \left( \sum_{i=1}^{m_1} P_{A_1|\tilde{A}_2=\tilde{b}}(a_{1,i}) \left( \sum_{j=1}^{m_2} u_{i,j} P_{A_2}(a_{2,j}) \frac{\mathrm{d}P_{\tilde{A}_2|A_2=a_{2,j}}}{\mathrm{d}P_{\tilde{A}_2|A_2=a_{2,k}}}(\tilde{b}) \right) \right) \mathrm{d}P_{\tilde{A}_2|A_2=a_{2,k}}(\tilde{b}),$$

for some $k \in \{1, 2, \ldots, m_2\}$.

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

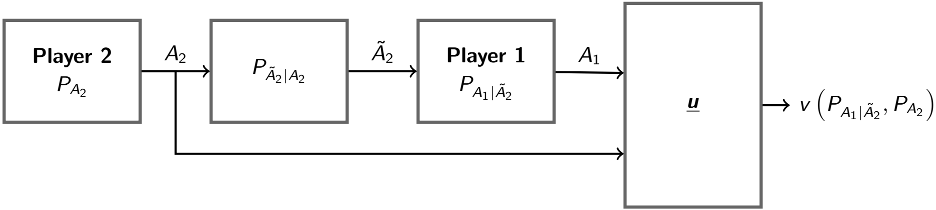
Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:




**Definition (Best Response)**

$$\mathrm{BR}_1(P, \tilde{b}) \triangleq \arg \max_{Q \in \triangle(\mathcal{A}_1)} \sum_{i=1}^{m_1} Q(a_{1,i}) \left( \sum_{j=1}^{m_2} u_{i,j} P(a_{2,j}) \frac{\mathrm{d}P_{\tilde{A}_2|A_2=a_{2,j}}}{\mathrm{d}P_{\tilde{A}_2|A_2=a_{2,k}}}(\tilde{b}) \right).$$

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:




**Definition (Cost of Player 2)**

Let the function $\hat{v} : \triangle(\mathcal{A}_2) \to \mathbb{R}$ be

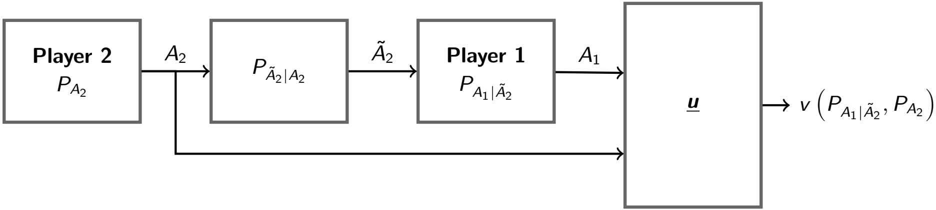
$$\hat{v}(P) = v(Q_{A_1|\tilde{A}_2}, P),$$

where for all $\tilde{b} \in \tilde{\mathcal{A}}_2$, it holds that $Q_{A_1|\tilde{A}_2=\tilde{b}} \in \mathrm{BR}_1(P, \tilde{b})$.

# Zero-Sum Games with **Noisy Observations**

## Game Formulation – Player 1 maximizes – Player 2 minimizes

Consider the $m_1 \times m_2$ ZSG with payoff matrix $\underline{\boldsymbol{u}}$:




### Definition (Equilibrium)

The strategies $P^{\dagger}_{A_1|\tilde{A}_2} \in \Delta(\mathcal{A}_1|\tilde{\mathcal{A}}_2)$ and $P^{\dagger}_{A_2} \in \triangle(\mathcal{A}_2)$ form an equilibrium if

$$P^{\dagger}_{A_2} \in \arg \min_{P \in \triangle(\mathcal{A}_2)} \hat{v}(P),$$

and for all $\tilde{b} \in \tilde{\mathcal{A}}_2$

$$P^{\dagger}_{A_1|\tilde{A}_2 = \tilde{b}} \in \mathrm{BR}_1(P^{\dagger}_{A_2}, \tilde{b}).$$

# Zero-Sum Games with **Noisy Observations**

Equilibrium – Player 1 maximizes – Player 2 minimizes

*Theorem*

*The game (with noisy observations) possesses a unique equilibrium.*

# Zero-Sum Games with **Noisy Observations**

Equilibrium – Player 1 maximizes – Player 2 minimizes

## *Theorem*

*The game (with noisy observations) possesses a unique equilibrium.*

## *Lemma*

*Let the probability measures* $(P^{\dagger}_{A_1|\tilde{A}_2}, P^{\dagger}_{A_2})$ *form an equilibrium of the game (with noisy observations); and let the pair of strategies* $\left(P^{\star}_{A_1}, P^{\star}_{A_2}\right) \in \triangle(\mathcal{A}_1) \times \triangle(\mathcal{A}_2)$ *be an Nash Equilibrium of the game (without observations). Then,*

$$v\left(P^{\star}_{A_1}, P^{\star}_{A_2}\right) \leq v\left(P^{\dagger}_{A_1|\tilde{A}_2}, P^{\dagger}_{A_2}\right) \leq \min_j \max_i u_{i,j}.$$

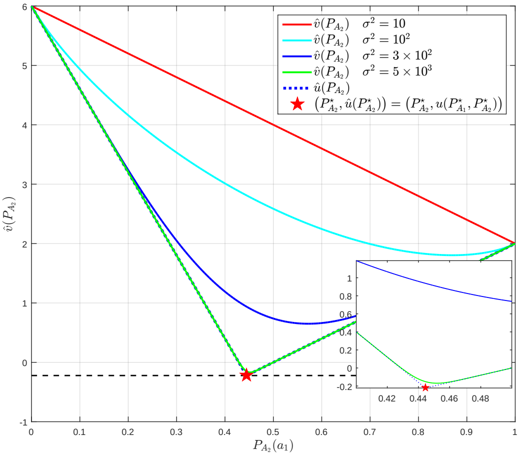
[7] Ke Sun, Samir M. Perlaza, and Alain Jean-Marie. **“2 x 2 Zero-Sum Games with Commitments and Noisy Observations”**. In Proc. of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun., 2023.
[8] Emmanouil-Marios Athanasakos and Samir M. Perlaza. **“Leveraging Noisy Observations in Zero-Sum Games”**. In Proc. of the IEEE Information Theory Workshop (ITW), Shenzhen, China, Nov. 2024.

Challenge: What is the gain/loss due to the **noisy observation of the actions** ?

# Zero-Sum Games with **Noisy Observations**

## Numerical Results

- **Two Actions**: $|\mathcal{A}_1| = |\mathcal{A}_2| = 2$
- **Gaussian Channel**: $\tilde{A}_2 = A_2 + W$, with $W \sim \mathcal{N}\left(0, \sigma^2\right)$, for some given $\sigma^2 > 0$.



*Lemma*

*Let the probability measures $(P^{\dagger}_{A_1|\tilde{A}_2}, P^{\dagger}_{A_2})$ form an equilibrium of the game (with noisy observations); and let the pair of strategies $(P^{\star}_{A_1}, P^{\star}_{A_2}) \in \triangle(\mathcal{A}_1) \times \triangle(\mathcal{A}_2)$ be an Nash Equilibrium of the game (without observations). Then,*

$$v\left(P^{\star}_{A_1}, P^{\star}_{A_2}\right) \leq v\left(P^{\dagger}_{A_1|\tilde{A}_2}, P^{\dagger}_{A_2}\right) \leq \min_j \max_i u_{i,j}.$$

$$v\left(P^{\dagger}_{A_1|\tilde{A}_2}, P^{\dagger}_{A_2}\right) - v\left(P^{\star}_{A_1}, P^{\star}_{A_2}\right) = \int u_{a,b} \mathrm{d} P^{\dagger}_{A_1 A_2}(a,b) - \int u_{a,b} \mathrm{d} P^{\star}_{A_1} P^{\star}_{A_2}(a,b),$$

where for all measurable subsets $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2$ of $\mathcal{A}_1 \times \mathcal{A}_2$,

$$P^{\dagger}_{A_1 A_2}(\mathcal{C}) = \int_{\mathcal{C}_1} \int \int_{\mathcal{C}_2} \mathrm{d} P^{\dagger}_{A_1|\tilde{A}_2 = \tilde{b}}(a) \mathrm{d} P_{\tilde{A}_2|A_2=b}(\tilde{b}) \mathrm{d} P^{\dagger}_{A_2}(b). \tag{3}$$

# Table of Contents

# Thank you for your attention!

Questions/Comments/Typos: samir.perlaza@inria.fr


- [1] Samir M. Perlaza and Gaetan Bisson. "**Variations on the Expectation Due to Changes in the Probability Measure**". arXiv preprint arXiv:2502.02887.
- [2] Yaiza Bermudez, Samir M. Perlaza, Gaetan Bisson, and Iñaki Esnaola. "**Proofs for Folklore Theorems on the Radon-Nikodym Derivative**". arXiv preprint arXiv:2501.18374.
- [3] Francisco Daunas, Iñaki Esnaola, Samir M. Perlaza, and Gholamali Aminian. "**Generalization Error of $f$-Divergence Stabilized Algorithms via Duality**". arXiv preprint arXiv:2502.14544.
- [4] Samir M. Perlaza and Xinying Zou. "**The Generalization Error of Machine Learning Algorithms**". arXiv preprint arXiv:2411.12030.
- [5] Samir M. Perlaza, Gaetan Bisson, Iñaki Esnaola, Alain Jean-Marie, and Stefano Rini. "**Empirical Risk Minimization with Relative Entropy Regularization**". IEEE Transactions on Information Theory, vol. 70, no. 7, pp. 5122 – 5161, July, 2024.
- [6] Xinying Zou, Samir M. Perlaza, Iñaki Esnaola, Eitan Altman, and H. Vincent Poor. "**The Worst-Case Data-Generating Probability Measure in Statistical Learning**". IEEE Journal on Selected Areas in Information Theory, vol. 5, pp. 175–189, Apr., 2024.
- [7] Ke Sun, Samir M. Perlaza, and Alain Jean-Marie. **"2 x 2 Zero-Sum Games with Commitments and Noisy Observations"**. In Proc. of the IEEE International Symposium on Information Theory (ISIT), Taipei, Taiwan, Jun., 2023.
- [8] Emmanouil-Marios Athanasakos and Samir M. Perlaza. **"Leveraging Noisy Observations in Zero-Sum Games"**. In Proc. of the IEEE Information Theory Workshop (ITW), Shenzhen, China, Nov. 2024.