Electronic Communications in Probability, **20(5):1-5, 2015.**

A Lower Bound on the Relative Entropy with Respect to a Symmetric Probability

Raphaël Cerf
ENS Paris

and

Matthias Gorny
Université Paris Sud and ENS Paris

Abstract

Let ρ and μ be two probability measures on $\mathbb{R}$ which are not the Dirac mass at 0. We denote by $H(\mu|\rho)$ the relative entropy of μ with respect to ρ. We prove that, if ρ is symmetric and μ has a finite first moment, then

$$H(\mu|\rho) \geq \frac{\left(\displaystyle\int_{\mathbb{R}} z\, d\mu(z)\right)^2}{2\displaystyle\int_{\mathbb{R}} z^2\, d\mu(z)},$$

with equality if and only if $\mu = \rho$. We give an applicaion to the Curie-Weiss model of self-organized criticality.

AMS 2010 subject classifications: 60E15 60F10 94A17
Keywords: entropy, Cramér transform, Ising Curie-Weiss, SOC

1 Introduction

Given two probability measures μ and ρ on $\mathbb{R}$, the relative entropy of μ with respect to ρ (or the Kullback-Leibler divergence of ρ from μ) is

$$H(\mu|\rho) = \begin{cases} \displaystyle\int_{\mathbb{R}} f(z) \ln f(z)\, d\rho(z) & \text{if } \mu \ll \rho \text{ and } f = \dfrac{d\mu}{d\rho} \\ +\infty & \text{otherwise}\,, \end{cases}$$

where $d\mu/d\rho$ denotes the Radon-Nikodym derivative of μ with respect to ρ when it exists. In this paper, we prove the following theorem:

Theorem 1. *Let ρ and μ be two probability measures on $\mathbb{R}$ which are not the Dirac mass at 0. If ρ is symmetric and if μ has a finite first moment, then*

$$H(\mu|\rho) \geq \frac{\displaystyle\left(\int_{\mathbb{R}} z\, d\mu(z)\right)^2}{\displaystyle 2\int_{\mathbb{R}} z^2\, d\mu(z)}\,,$$

with equality if and only if $\mu = \rho$.

A remarkable feature of this inequality is that the lower bound does not depend on the symmetric probability measure ρ. We found the following related inequality in the literature (see lemma 3.10 of [1]): if ρ is a probability measure on $\mathbb{R}$ whose first moment m exists and such that

$$\exists v > 0 \qquad \forall \lambda \in \mathbb{R} \qquad \int_{\mathbb{R}} \exp(\lambda(z-m))\, d\rho(z) \leq \exp\left(\frac{v\lambda^2}{2}\right)\,,$$

then, for any probability measure μ on $\mathbb{R}$ having a first moment, we have

$$H(\mu|\rho) \geq \frac{1}{2v}\left(\int_{\mathbb{R}} z\, d\mu(z) - m\right)^2\,.$$

Our inequality does not require an integrability condition. Instead we assume that ρ is symmetric.

The proof of the theorem is given in the following section. It consists in relating the relative entropy $H(\,\cdot\,|\rho)$ and the Cramér transform I of (Z, Z^2) when Z is a random variable with distribution ρ. We then use an inequality on I which we proved initially in [2]. We give here a simplified proof of this inequality.

In section 3, we apply the inequality of theorem 1 to the Curie-Weiss model of self-organized criticality we designed in [2]. We prove that, if $(X_n^1, \dots, X_n^n)$ has the distribution

$$d\widetilde{\mu}_{n,\rho}(x_1, \dots, x_n) = \frac{1}{Z_n} \exp\left(\frac{1}{2}\frac{(x_1 + \cdots + x_n)^2}{x_1^2 + \cdots + x_n^2}\right) \mathbb{1}_{\{x_1^2 + \cdots + x_n^2 > 0\}} \prod_{i=1}^n d\rho(x_i)\,,$$

for any $n \geq 1$, and if ρ is symmetric with compact support and such that $\rho(\{0\}) < 1/\sqrt{e}$, then, for any continuous function $f : \mathbb{R} \longrightarrow \mathbb{R}$,

$$\forall \varepsilon > 0 \qquad \lim_{n\to\infty} \widetilde{\mu}_{n,\rho}\left(\left|\frac{1}{n}\sum_{k=1}^n f(X_n^k) - \int_{\mathbb{R}} f(z)\, d\rho(z)\right| \geq \varepsilon\right) = 0\,.$$

2 Proof of the theorem

Let ρ and μ be two probability measures on $\mathbb{R}$ which are not the Dirac mass at 0. We first recall that $H(\mu|\rho) \geq 0$, with equality if and only if $\mu = \rho$.

We assume that ρ is symmetric and that μ has a finite first moment. We denote

$$\mathcal{F}(\mu) = \frac{\left(\displaystyle\int_{\mathbb{R}} z \, d\mu(z) \right)^2}{2 \displaystyle\int_{\mathbb{R}} z^2 \, d\mu(z)} .$$

If $\mu = \rho$ then $\mathcal{F}(\mu) = 0 = H(\mu|\rho)$. From now onwards we suppose that $\mu \neq \rho$. If the first moment of μ vanishes or if its second moment is infinite, then we have $\mathcal{F}(\mu) = 0 < H(\mu|\rho)$. Finally, if μ is such that $H(\mu|\rho) = +\infty$, then Jensen's inequality implies that

$$\mathcal{F}(\mu) \leq 1/2 < H(\mu|\rho).$$

In the following, we suppose that

$$\int_{\mathbb{R}} z \, d\mu(z) \neq 0, \qquad \int_{\mathbb{R}} z^2 \, d\mu(z) < +\infty ,$$

and that $H(\mu|\rho) < +\infty$. This implies that $\mu \ll \rho$ and we set $f = d\mu/d\rho$. It follows from Jensen's inequality that, for any μ-integrable function Φ,

$$\int_{\mathbb{R}} \Phi \, d\mu - H(\mu|\rho) = \int_{\mathbb{R}} \ln \left(\frac{e^\Phi}{f} \right) \, d\mu \leq \ln \int_{\mathbb{R}} \frac{e^\Phi}{f} \, d\mu = \ln \int_{\mathbb{R}} e^\Phi \, d\rho .$$

As a consequence

$$\sup_{\Phi \in \mathrm{L}^1(\mu)} \left\{ \int_{\mathbb{R}} \Phi \, d\mu - \ln \int_{\mathbb{R}} e^\Phi \, d\rho \right\} \leq H(\mu|\rho) .$$

In order to make appear the first and second moments of ρ, we consider functions Φ of the form $z \longmapsto uz + vz^2$, $(u, v) \in \mathbb{R}^2$. This way we obtain

$$I \left(\int_{\mathbb{R}} z \, d\mu(z), \int_{\mathbb{R}} z^2 \, d\mu(z) \right) \leq H(\mu|\rho) ,$$

where

$$\forall (x, y) \in \mathbb{R}^2 \qquad I(x, y) = \sup_{(u,v) \in \mathbb{R}^2} \left\{ ux + vy - \ln \int_{\mathbb{R}} e^{uz + vz^2} \, d\rho(z) \right\} .$$

The function I is the Cramér transform of (Z, Z^2) when Z is a random variable with distribution ρ. In our paper dealing with a Curie-Weiss model of self-organized criticality [2], we proved with the help of the following inequality that, under some integrability condition, the function $(x, y) \longmapsto I(x, y) - x^2/(2y)$ has a unique global minimum on $\mathbb{R} \times]0, +\infty[$ at $\big(0, \int x^2 \, d\rho(x)\big)$.

Proposition 2. *If ρ is a symmetric probability measure which is not the Dirac mass at 0, then*

$$\forall x \neq 0 \quad \forall y \neq 0 \qquad I(x, y) > \frac{x^2}{2y}.$$

We present here a proof of this proposition which is simpler than in [2].

Proof. Let $x \neq 0$ and $y \neq 0$. By definition of $I(x,y)$, we have

$$\begin{aligned} I(x,y) &\geq x \times \frac{x}{y} + y \times \left(-\frac{x^2}{2y^2}\right) - \ln \int_{\mathbb{R}} \exp\left(\frac{xz}{y} - \frac{x^2z^2}{2y^2}\right)\, d\rho(z) \\ &= \frac{x^2}{2y} - \ln \int_{\mathbb{R}} \exp\left(\frac{xz}{y} - \frac{x^2z^2}{2y^2}\right)\, d\rho(z)\,. \end{aligned}$$

Let $(s,t) \in \mathbb{R}^2$. By using the symmetry of ρ, we obtain

$$\begin{aligned} \int_{\mathbb{R}} \exp(sz - tz^2)\, d\rho(z) &= \int_{\mathbb{R}} \exp(-sz - tz^2)\, d\rho(z) \\ &= \frac{1}{2}\left(\int_{\mathbb{R}} \exp(sz - tz^2)\, d\rho(z) + \int_{\mathbb{R}} \exp(-sz - tz^2)\, d\rho(z)\right) \\ &= \int_{\mathbb{R}} \cosh(sz)\, \exp(-tz^2)\, d\rho(z)\,. \end{aligned}$$

We choose now $t = s^2/2$. We have the inequality

$$\forall u \in \mathbb{R}\backslash\{0\} \qquad \cosh(u)\, \exp\big(-u^2/2\big) < 1\,.$$

Since ρ is not the Dirac mass at 0, the above inequality implies that

$$\forall s \neq 0 \qquad \int_{\mathbb{R}} \cosh(sz)\, \exp\left(-\frac{s^2z^2}{2}\right)\, d\rho(z) < 1\,.$$

We finally choose $s = x/y$ and we get

$$\int_{\mathbb{R}} \exp\left(\frac{xz}{y} - \frac{x^2z^2}{2y^2}\right)\, d\rho(z) < 1\,.$$

As a consequence

$$I(x,y) \geq \frac{x^2}{2y} - \ln \int_{\mathbb{R}} \exp\left(\frac{xz}{y} - \frac{x^2z^2}{2y^2}\right)\, d\rho(z) > \frac{x^2}{2y}\,,$$

which is the desired inequality. □

By applying the above proposition with

$$x = \int_{\mathbb{R}} z\, d\mu(z) \neq 0, \qquad y = \int_{\mathbb{R}} z^2\, d\mu(z) \in]0,+\infty[\,,$$

we obtain

$$H(\mu|\rho) \geq I\left(\int_{\mathbb{R}} z\, d\mu(z), \int_{\mathbb{R}} z^2\, d\mu(z)\right) > \mathcal{F}(\mu)\,.$$

This ends the proof of theorem 1.

3 Application to the Curie-Weiss model of SOC

In [2], we designed the following model: Let ρ be a probability measure on $\mathbb{R}$, which is not the Dirac mass at 0. We consider an infinite triangular array of real-valued random variables $(X_n^k)_{1\leq k\leq n}$ such that for all $n\geq 1$, $(X_n^1,\dots,X_n^n)$ has the distribution $\widetilde{\mu}_{n,\rho}$, where

$$d\widetilde{\mu}_{n,\rho}(x_1,\dots,x_n)=\frac{1}{Z_n}\exp\left(\frac{1}{2}\frac{(x_1+\cdots+x_n)^2}{x_1^2+\cdots+x_n^2}\right)\mathbb{1}_{\{x_1^2+\cdots+x_n^2>0\}}\prod_{i=1}^n d\rho(x_i)\,,$$

and Z_n is the renormalization constant. In [2] and [4], we proved that this model exhibits self-organized criticality: for a large class of symmetric distributions, we proved the fluctuations of $S_n=X_n^1+\cdots+X_n^n$ are of order $n^{3/4}$ and the limiting law is $C\exp(-\lambda x^4)\,dx$ for some $C,\lambda>0$.

For any $n\geq 1$, let us introduce the empirical measure

$$M_n=\frac{1}{n}\big(\delta_{X_n^1}+\cdots+\delta_{X_n^n}\big)\,.$$

The inequality of theorem 1 is the key ingredient to prove the following theorem:

Theorem 3. *Let ρ be a symmetric probability measure on $\mathbb{R}$ with compact support and such that $\rho(\{0\})<1/\sqrt{e}$. Then, under $\widetilde{\mu}_{n,\rho}$, the sequence $(M_n)_{n\geq 1}$ converges weakly in probability to ρ, i.e., for any continuous function f from $\mathbb{R}$ to $\mathbb{R}$, we have*

$$\forall\varepsilon>0\qquad \lim_{n\to\infty}\widetilde{\mu}_{n,\rho}\left(\left|M_n(f)-\int_{\mathbb{R}}f\,d\rho\right|\geq\varepsilon\right)=0\,.$$

Let us prove this theorem. We suppose that there exists $L>0$ such that the support of ρ is $[-L,L]$ or $]-L,L[$. We denote by $\mathcal{M}_1^L$ the space of all probability measures on $[-L,L]$ endowed with the topology of weak convergence. Let $\varepsilon>0$ and let f be a continuous function from $\mathbb{R}$ to $\mathbb{R}$. The set

$$\mathcal{U}_\varepsilon=\left\{\mu\in\mathcal{M}_1^L:\left|\int_{\mathbb{R}}f\,d\mu-\int_{\mathbb{R}}f\,d\rho\right|<\varepsilon\right\}\,,$$

is open in $\mathcal{M}_1^L$. Let $n\geq 1$. We denote by $\widetilde{\theta}_{n,\rho}$ the law of $(\delta_{Y_1}+\cdots+\delta_{Y_n})/n$ when $Y_1,\dots,Y_n$ are n independent random variables with distribution ρ. We have

$$\widetilde{\mu}_{n,\rho}(M_n\in\mathcal{U}_\varepsilon^c)=\frac{1}{Z_n}\int_{\mathcal{U}_\varepsilon^c}\exp\left(n\mathcal{F}(\mu)\right)\mathbb{1}_{\mu\neq\delta_0}\,d\widetilde{\theta}_{n,\rho}(\mu)\,.$$

The function $\mathcal{F}$ is continuous on $\mathcal{M}_1^L\backslash\{\delta_0\}$. Next, since $\mathcal{F}(\delta_{1/k})=1/2$ for any $k\geq 1$, we notice that putting $\mathcal{F}(\delta_0)\geq 1/2$ is necessary to ensure that $\mathcal{F}$ is upper semi-continuous. As a consequence we extend the definition of $\mathcal{F}$ on $\mathcal{M}_1^L$ by putting $\mathcal{F}(\delta_0)=1/2$. We suppose that $\rho(\{0\})<1/\sqrt{e}$ so that

$$\mathcal{F}(\delta_0)=1/2<-\ln\rho(\{0\})=H(\delta_0|\rho)\,.$$

If $\mu\in\mathcal{M}_1^L\backslash\{\delta_0\}$ then theorem 1 implies that $\mathcal{F}(\mu)\leq H(\mu|\rho)$ with equality if and only if $\mu=\rho$. Hence the function $\mathcal{F}-H(\,\cdot\,|\rho)$ has a unique maximum on $\mathcal{M}_1^L$ at ρ.

Sanov's theorem (theorem 6.2.10 of [3]) states that $(\widetilde{\theta}_{n,\rho})_{n\geq 1}$ satisfies the large deviation principle in $\mathcal{M}_1^L$ with speed n and governed by the good rate function $H(\,\cdot\,|\rho)$. As a consequence

$$\liminf_{n\to+\infty}\frac{1}{n}\ln Z_n \geq \liminf_{n\to+\infty}\frac{1}{n}\ln\widetilde{\theta}_{n,\rho}(\{\delta_0\}^c) \geq -\inf_{\mu\neq\delta_0} H(\mu|\rho)=0\,.$$

Since $\mathcal{F}$ is bounded (by 1/2) and is upper semi-continuous on $\mathcal{M}_1^L$, Varadhan's lemma (see section 4.3 of [3]) implies that

$$\begin{aligned}\limsup_{n\to+\infty}\frac{1}{n}\ln\widetilde{\mu}_{n,\rho}\left(M_n\in\mathcal{U}_\varepsilon^c\right) &\leq \limsup_{n\to+\infty}\frac{1}{n}\ln\int_{\mathcal{U}_\varepsilon^c}e^{n\mathcal{F}(\mu)}\,d\widetilde{\theta}_{n,\rho}(\mu)-\liminf_{n\to+\infty}\frac{1}{n}\ln Z_n\\ &\leq \sup\left\{\,\mathcal{F}(\mu)-H(\mu|\rho):\mu\in\mathcal{U}_\varepsilon^c\,\right\}.\end{aligned}$$

Since $H(\,\cdot\,|\rho)$ is a good rate function, $\mathcal{F}$ is upper semi-continuous and $\mathcal{U}_\varepsilon^c$ is a closed subset of $\mathcal{M}_1^L$ which does not contain ρ, the unique maximum of the function $\mathcal{F}-H(\,\cdot\,|\rho)$, we get

$$\sup\left\{\,\mathcal{F}(\mu)-H(\mu|\rho):\mu\in\mathcal{U}_\varepsilon^c\,\right\}<0\,.$$

As a consequence, there exists $c_\varepsilon>0$ and $n_\varepsilon\geq 1$ such that

$$\forall n\geq n_\varepsilon\qquad \widetilde{\mu}_{n,\rho}\left(M_n\in\mathcal{U}_\varepsilon^c\right)\leq\exp(-nc_\varepsilon).$$

This implies the convergence in theorem 3.

References

[1] Stéphane Boucheron, Gàbor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[2] Raphaël Cerf and Matthias Gorny. A Curie–Weiss model of Self–Organized Criticality. *The Annals of Probability*, to appear, 2013.

[3] Amir Dembo and Ofer Zeitouni. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, 2010.

[4] Matthias Gorny. The Cramér Condition for the Curie–Weiss model of SOC. *preprint*, 2013.

[5] Matthias Gorny. A Curie–Weiss model of Self–Organized Criticality : The Gaussian Case. *Markov processes and related fields*, 20(3):563–576, 2014.