MATH 585 – TOPICS IN MATHEMATICAL PHYSICS – FALL 2006

MATHEMATICS OF MEAN FIELD SPIN GLASSES AND THE REPLICA METHOD

APPENDIX TO LECTURE 3′:: ADDENDUM ON ENTROPY

S. STARR
MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER

CONTENTS

1. DEFINITION

We mentioned the relative entropy function, in passing, in proving the Gibbs variational principle. Let us now return to that topic alone and state some of its most important properties.

Suppose Ω is a finite sample space[1], and let $\mathscr{F}$ be the discrete σ-algebra associated to it. Let $\mathscr{M}_1(\Omega,\mathscr{F})$ be the set of probability measures. If $\mu,\nu \in \mathscr{M}_1(\Omega,\mathscr{F})$ then $\nu \ll \mu$ will indicate that ν is absolutely continuous with respect to ν: in other words, $\operatorname{supp}(\nu) \subseteq \operatorname{supp}(\mu)$, where $\operatorname{supp}(\mu) := \{\xi \in \Omega : \mu(\{\xi\}) \neq 0\}$. The relative entropy is a function of two variables, $S(\cdot|\cdot) : \mathscr{M}_1(\Omega,\mathscr{F}) \times \mathscr{M}_1(\Omega,\mathscr{F}) \to \mathbb{R} \cup \{-\infty\}$. If $\nu \ll \mu$ then

$$S(\mu|\nu) \;:=\; \int_\Omega u\left(\frac{d\nu}{d\mu}\right)\, d\mu$$

where $u : [0,\infty) \to \mathbb{R}$ is the function

$$u(t) \;=\; \begin{cases} -t\log(t) & \text{if } t>0,\\ 0 & \text{if } t=0. \end{cases}$$

If $\nu \not\ll \mu$ then $S(\mu|\nu)$ is defined to be $-\infty$.

2. SIMPLEST PROPERTIES

The function u is continuous and strictly concave. Strict concavity means that, given $t_1, t_2 \in [0,\infty)$, unequal, and given $\theta \in (0,1)$, there is strict inequality

$$u(\theta t_1 + [1-\theta] t_2) \;>\; \theta u(t_1) + (1-\theta) u(t_2)\,.$$

Date: September 15, 2006.

[1] Many things are simpler for finite sample spaces, which is all we need. But if your own research/reading leads you further, keep in mind some of these definitions need to be generalized for countable or continuous Ω.

(Why is u strictly concave?) The following lemma is an easy exercise for you:

Lemma 2.1 *Let $f : \mathcal{K} \to \mathbb{R}$ be a strictly concave function, where $\mathcal{K}$ is a convex subset of a vector space. If f attains its maximum on $\mathcal{K}$, then that maximum is unique.*

Here is a lemma we will prove.

Lemma 2.2 *Let $\mu \in \mathscr{M}_1(\Omega, \mathscr{F})$ and define $D_\mu := \{\nu \in \mathscr{M}_1(\Omega, \mathscr{F}) \ : \ \nu \ll \mu\}$. Then $S(\cdot|\mu) \restriction D_\mu$ is strictly concave.*

Proof. Note that the definition of $S(\nu|\mu)$, can be rewritten as

$$S(\nu|\mu) \,=\, \sum_{\xi \in \text{supp}(\mu)} u\left(\frac{\nu(\{\xi\})}{\mu(\{\xi\})}\right)\, \mu(\{\xi\})\,,$$

when $\nu \ll \mu$. Also, the map $\nu \in \mathscr{M}_1(\Omega, \mathscr{F}) \mapsto \frac{\nu(\xi)}{\mu(\xi)}$ is affine (read "linear") for each $\xi \in \text{supp}(\mu)$. Therefore, the map

$$\nu \in \mathscr{M}_1(\Omega, \mathscr{F}) \,\mapsto\, u\left(\frac{\nu(\xi)}{\mu(\xi)}\right)\,,$$

is concave for each $\xi \in \text{supp}(\mu)$. Finally, because a convex combination of concave functions is concave, we see that $S(\nu|\mu)$ is concave in the variable ν, for $\nu \in D_\mu$. In fact, $S(\nu|\mu)$ is strictly concave, because if $\nu_1, \nu_2 \in D_\mu$ and $\nu_1 \neq \nu_2$ then there is some $\xi \in \text{supp}(\mu)$ where $\nu_1(\{\xi\}) \neq \nu_2(\{\xi\})$. Then the fact, that

$$S(\theta\nu_1 + (1-\theta)\nu_2|\mu) \,>\, \theta S(\nu_1|\mu) + (1-\theta)S(\nu_2|\mu)\,,$$

for $\theta \in (0,1)$, is implied by strict concavity of $\nu \mapsto u\left(\frac{\nu(\{\xi\})}{\mu(\{\xi\})}\right)$. □

Corollary 2.3 *For any $\mu, \nu \in \mathscr{M}_1(\Omega, \mathscr{F})$,*

$$S(\nu|\mu) \,\leq\, 0\,,$$

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

Before stating this simple fact, let us recall Jensen's inequality, probably the most basic and important inequality in all of analysis: for a concave function u and a probability measure μ and any (measurable) function $f : \text{supp}(\mu) \to \mathbb{R}$ such that $\text{Ran}(f) \subseteq \text{Dom}(u)$,

$$\int u \circ f\, d\mu \,\leq\, u\left(\int f\, d\mu\right)\,.$$

If any readers have never seen this result before, they should quickly look it up or prove it for themselves.

Proof. Note that

$$\int_\Omega \frac{d\nu}{d\mu}\, d\mu \,=\, \int_\Omega d\nu \,=\, 1\,.$$

So Jensen's inequality gives

$$\int_\Omega u\left(\frac{d\nu}{d\mu}\right)\, d\mu \,\leq\, u\left(\int_\Omega \frac{d\nu}{d\mu}\, d\mu\right) \,=\, u(1) \,=\, 0\,.$$

Now observe that $S(\mu|\mu) = 0$ because $\frac{d\mu}{d\mu} = 1$. Since $S(\cdot|\mu)$ is strictly concave, μ is the unique maximizer: for every other ν, $S(\nu|\mu)$ is strictly less than 0. □

Let us note another property. Suppose that $\nu, \mu_1, \mu_2 \in \mathscr{M}_1(\Omega, \mathscr{F})$, and $\nu \in D_{\mu_1} \cap D_{\mu_2}$. Then

$$S(\nu|\mu_1) - S(\nu|\mu_2) \,=\, \int_{\mathrm{supp}(\nu)} \log\left(\frac{d\mu_1}{d\mu_2}\right)\, d\nu\,.$$

Indeed, this follows just because of the alternative representation

$$S(\nu|\mu) \,=\, -\int_{\mathrm{supp}(\nu)} \log\left(\frac{d\nu}{d\mu}\right)\, d\nu\,,$$

and properties of the logarithm.

Remark 2.4 We use the convention for relative entropy common in classical spin systems. But often it is defined with the opposite sign.

MATHEMATICS DEPARTMENT, UNIVERSITY OF ROCHESTER, ROCHESTER, NY 14627
E-mail address: sstarr@math.rochester.edu