Cramer's Theorem in Probability and Statistical Mechanics

An Exercise in Sign Shenanigans

OWEN LYNCH, Universiteit Utrecht

1 Introduction

The purpose of this document is to figure out exactly what is going on with Cramer's theorem, and the difference in sign conventions of the partition function in probability and in statistical mechanics.

2 Cramer's Theorem

From [2] we have the following statement of Cramer's theorem

THEOREM 1. (CRAMER'S THEOREM) Let $X_1, X_2, \ldots$ be i.i.d. random variables with finite logarithmic moment generating function

$$\Lambda(t) := \log \mathbb{E}[\mathrm{e}^{tX_1}]$$

Let

$$\Lambda^*(x) := \sup_{t \in \mathbb{R}} (tx - \Lambda(t))$$

be the Legendre transform of Λ. Then for every $x \geq \mathbb{E}[X_1]$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[\sum_{j=1}^{n} X_j \geq xn\right] = -\Lambda^*(x)$$

In statistical mechanics, we use a different convention for our "partition function"; we define

$$\Lambda(\beta) := \log \mathbb{E}[\mathrm{e}^{-\beta H}]$$

where H is a random variable giving the energy of the system. We will show how to adopt Cramer's theorem to this situation.

Let $Y_1, Y_2, \ldots$ be i.i.d. random variables distributed as H, and let $X_i = -Y_i$. Then we can apply Cramer's theorem to the X_i to get that for every $x \geq \mathbb{E}[X_1]$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[\sum_{j=1}^{n} X_j \geq xn\right] = -\Lambda^*(x)$$

Now, suppose that $x \leq \mathbb{E}[Y_1]$, so that $-x \geq \mathbb{E}[X_1]$. Then

$$\mathbb{P}\left[\sum_{j=1}^{n} Y_j \leq xn\right] = \mathbb{P}\left[\sum_{j=1}^{n} X_j \geq (-x)n\right]$$

Thus,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[\sum_{j=1}^{n} Y_j \leq xn\right] = -\Lambda^*(-x)$$

If we expand out the definition of $-\Lambda^*(-x)$, we get

$$\begin{aligned} -\Lambda^*(-x) &= -\sup_{t \in \mathbb{R}} (t(-x) - \Lambda(t)) \\ &= -\sup_{t \in \mathbb{R}} (-tx - \Lambda(t)) \\ &= \inf_{t \in R} (tx + \Lambda(t)) \end{aligned}$$

as when we move the negative sign across the supremum it becomes an infimum. Thus, we can state an alternative, statistical mechanical version of Cramer's thereom, adding in a purely cosmetic replacement of t with β and x with u. This is the version that [1] uses.

THEOREM 2. (CRAMER'S THEOREM, STATISTICAL MECHANICAL VERSION) Suppose that $Y_1, Y_2, \ldots$ are i.i.d. random variables, all distributed as H. Let $\Phi(\beta)$ be the logarithmic partition function, defined by

$$\Phi(\beta) = \log \mathbb{E}[\mathrm{e}^{-\beta H}]$$

Define the (not quite Legendre transform of Φ) by

$$\varphi(u) = -\inf_{\beta \in \mathbb{R}} (\beta u + \Lambda(\beta))$$

Then, for every $u < \mathbb{E}[H]$,

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{P}\left[\sum_{j=1}^{n} Y_j \leq u\, n\right] = -\varphi(u)$$

3 Relation to Classical Thermodynamics

We now try and investigate exactly what Φ and φ mean in the context of classical thermodynamics. Let $\mathfrak{X}$ be a countable set, and let $H\colon \mathfrak{X} \to \mathbb{R}$ be any function. Then let X be a variable taking values in $\mathfrak{X}$, with distribution

$$\mathbb{P}(X = i) = \frac{\mathrm{e}^{-\beta H(i)}}{Z(\beta)}$$

where

$$Z(\beta) = \sum_{i \in \mathfrak{X}} \mathrm{e}^{-\beta H(i)}$$

is called the *partition function*. Surprisingly, the *log* of partition function,

$$\Phi(\beta) = \log Z(\beta)$$

can be used to calculate quantities related to the system fairly easily. For instance,

$$\begin{aligned}
\frac{\partial \Phi}{\partial \beta} &= \frac{\partial}{\partial \beta} \log Z(\beta) \\
&= \frac{\sum_{i \in \mathfrak{X}} \frac{\partial}{\partial \beta} \mathrm{e}^{-\beta H(i)}}{Z(\beta)} \\
&= \sum_{i \in \mathfrak{X}} -H(i) \frac{\mathrm{e}^{-\beta H(i)}}{Z(\beta)} \\
&= -\sum_{i \in \mathfrak{X}} H(i)\, \mathbb{P}(X = i) \\
&= -\mathbb{E}[H(X)]
\end{aligned}$$

In physics, we often write this last quantity as $-\langle H \rangle$.

We can also calculate the entropy using the partition function

$$\begin{aligned}
S(\beta) &= -\sum_{i \in \mathfrak{X}} P_i \log P_i \\
&= -\sum_{i \in \mathfrak{X}} \frac{\mathrm{e}^{-\beta H(i)}}{Z(\beta)} \log\left(\frac{\mathrm{e}^{-\beta H(i)}}{Z(\beta)}\right) \\
&= -\sum_{i \in \mathfrak{X}} -\beta H(i) P_i + \sum_{i \in \mathfrak{X}} P_i \log(Z(\beta)) \\
&= \beta \langle H \rangle + \Phi(\beta)
\end{aligned}$$

We can now see if this matches up with the classical definition of entropy for the Gibbs ensemble. Classically, entropy is given by the formula

$$F = U - TS$$

where F is the “free energy”. We want to identify F with $-\frac{\log(Z(\beta))}{\beta} = -T\log(Z(\beta))$, let's see if that ends up agreeing with our calculation for entropy above.

$$\begin{aligned} S &= \beta\langle H\rangle + \Phi(\beta) \\ TS &= \langle H\rangle + T\Phi(\beta) \\ TS &= \langle H\rangle - F \\ F &= \langle H\rangle - TS \end{aligned}$$

It does!

Now, if we instead make S a function of u, by letting $\beta = \beta(u)$, then we get

$$S(\beta(u)) = \beta(u)\,u + \Phi(\beta(u))$$

To find this β, we are looking for β such that $\frac{\partial\Phi}{\partial\beta} = -u$. This β is the minimizer of

$$\beta u + \Phi(\beta)$$

Thus, we have an expression

$$S(u) = \inf_{\beta}\{\beta u + \Phi(\beta)\}$$

So, $\varphi(u) = -S(u)$.

However, this is not as much of a direct connection as one might hope. This is because in Cramer's thereom, we begin with a random variable H that is given to us, and its distribution can be anything. In the canonical ensemble on the other hand, we assume that the distribution is something specific.

To understand what's going on, we must shift our perspective on the statistical mechanics. Suppose that we are given some variable H. We think of the distribution of H as the distribution of our system in the infinite-temperature limit. Now, consider an i.i.d. sequence of variables $Y_1, Y_2, \ldots$, and let $\hat{\mu}_N \in M(\mathbb{R})$ be the empirical distribution observed of the first N variables, where $M(\mathbb{R})$ is the space of measures on $\mathbb{R}$.

The first question that we ask is: conditional on observing that $\sum_{n=1}^{N} Y_n \le NU$, what is $\hat{\mu}_N$? The theory of large deviations (need citation) should say something like

$$\lim_{N\to\infty} \mathbb{E}\left[\hat{\mu}_N(W) \,\middle|\, \sum_{n=1}^{N} Y_n \le NU\right] = \frac{\int_W e^{-\beta u} dP_H(u)}{\int_{\mathbb{R}} e^{-\beta u} dP_H(u)} \tag{1}$$

where β is such that

$$\int_{\mathbb{R}} u\, e^{-\beta u} dP_H(u) = U$$

Define the measure $\hat{\mu}$ by

$$\hat{\mu}(W) = \lim_{N\to\infty} \mathbb{E}\left[\hat{\mu}_N(W) \,\middle|\, \sum_{n=1}^{N} Y_n \le NU\right]$$

Then using Eq. 1, we get

$$\frac{d\hat{\mu}}{dP_H}(u) = \frac{e^{-\beta u}}{\int_{\mathbb{R}} e^{-\beta u} dP_H(u)}$$

If dP_H is absolutely continuous with respect to Lebesgue measure, with Radon-Nikodym derivative

$$\frac{dP_H}{d\lambda}(u) = \Omega(u)$$

then this gives us

$$\frac{d\hat{\mu}}{d\lambda} = \frac{\Omega(u)\, e^{-\beta u}}{Z(\beta)}$$

which is familiar as the pdf of a canonical distribution.

BIBLIOGRAPHY

[1] Jeffrey Commons, Ying-Jen Yang, and Hong Qian. Duality symmetry, two entropy functions, and an eigenvalue problem in gibbs' theory. 2021.

[2] Achim Klenke. *Probability Theory: A Comprehensive Course*. Springer-Verlag, London, 2014.