

21-317 MEETING 3: RELATIVE ENTROPY AND GIBBS EQUILIBRIUM REVISED STUDENT NOTES REFLECTING THE MATERIAL COVERED

MATTHEW ROSENZWEIG

ABSTRACT. Meeting 3 began from additivity of entropy for independent systems. We introduced Jensen's inequality, used it to prove the Gibbs inequality for relative entropy, derived the maximum-entropy principle for a uniform reference law, and introduced Gibbs equilibrium and free energy. The meeting ended with the statement of the finite Gibbs variational principle. Its proof begins Meeting 4.

CONTENTS

1. Starting point	1
2. Jensen's inequality	1
3. Relative entropy	2
3.1. Exercises on relative entropy	3
4. Gibbs equilibrium and free energy	3
5. Further reading	4
References	4

1. STARTING POINT

For independent finite systems with laws ν and η , Meeting 2 established

$$(1.1) \quad H(\nu \otimes \eta) = H(\nu) + H(\eta).$$

Entropy describes a single probability law. To compare a trial law with a specified equilibrium or reference law, we need relative entropy.

2. JENSEN'S INEQUALITY

Let Ω be finite, let μ be a probability measure on Ω , and let $f : \Omega \rightarrow I$ for an interval $I \subset \mathbb{R}$.

Theorem 2.1 (Jensen's inequality). *If $\Phi : I \rightarrow \mathbb{R}$ is convex, then*

$$(2.1) \quad \Phi(\mathbb{E}_\mu f) \leq \mathbb{E}_\mu[\Phi(f)].$$

If Φ is strictly convex and $\mu(x) > 0$ for every state, equality holds if and only if f is constant.

Date: August 28, 2026; revised August 31, 2026.

Proof. Write the positive masses as $p_1, \dots, p_n$ and the values of f as $a_1, \dots, a_n$. The case $n = 2$ is the definition of convexity. For the induction step, put

$$q = p_1 + \dots + p_{n-1}, \quad m' = \frac{1}{q} \sum_{i=1}^{n-1} p_i a_i.$$

Then

$$\begin{aligned} \Phi \left(\sum_{i=1}^n p_i a_i \right) &= \Phi(qm' + p_n a_n) \\ &\leq q\Phi(m') + p_n \Phi(a_n) \\ &\leq \sum_{i=1}^n p_i \Phi(a_i). \end{aligned}$$

Strict convexity forces all a_i to agree in the equality case. □

The function $r \log r$

The principal convex function for entropy is

$$(2.2) \quad \Phi(r) = r \log r, \quad r \geq 0,$$

with $0 \log 0 = 0$. Since $\Phi''(r) = 1/r > 0$ for $r > 0$, it is strictly convex. Its tangent line at $r = 1$ gives

$$(2.3) \quad r \log r \geq r - 1,$$

with equality only at $r = 1$.

3. RELATIVE ENTROPY

Let ν and μ be probability laws on the same finite set Ω .

Definition 3.1 (Relative entropy). If $\mu(x) > 0$ for every x , define

$$(3.1) \quad D(\nu \| \mu) := \sum_{x \in \Omega} \nu(x) \log \frac{\nu(x)}{\mu(x)}.$$

If ν assigns positive mass to a state where μ has zero mass, set $D(\nu \| \mu) = +\infty$.

When $\mu > 0$, define the density ratio

$$(3.2) \quad h(x) = \frac{\nu(x)}{\mu(x)}.$$

Then

$$(3.3) \quad \mathbb{E}_\mu h = 1$$

and

$$(3.4) \quad D(\nu \| \mu) = \mathbb{E}_\mu [h \log h].$$

Theorem 3.2 (Gibbs inequality). For finite probability measures ν and μ ,

$$(3.5) \quad D(\nu \| \mu) \geq 0.$$

If $\mu > 0$, equality holds if and only if $\nu = \mu$.

Proof. Apply Jensen's inequality (2.1) to h and the convex function $r \mapsto r \log r$. Using (3.3),

$$D(\nu \parallel \mu) = \mathbb{E}_\mu[h \log h] \geq (\mathbb{E}_\mu h) \log(\mathbb{E}_\mu h) = 0.$$

Strict convexity implies that equality requires h to be constant. Since its mean is one, $h \equiv 1$, hence $\nu = \mu$. $\square$

Relative entropy is a divergence

In general,

$$(3.6) \quad D(\nu \parallel \mu) \neq D(\mu \parallel \nu),$$

and relative entropy does not satisfy a triangle inequality. The order matters: $D(\nu \parallel \mu)$ compares the trial law ν with the chosen reference law μ .

Corollary 3.3 (Maximum entropy). *If u_M is uniform on a finite set of cardinality M , then*

$$(3.7) \quad H(\nu) \leq \log M,$$

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

Proof. Since $u_M(x) = 1/M$,

$$(3.8) \quad D(\nu \parallel u_M) = \log M - H(\nu).$$

Apply Theorem 3.2. $\square$

3.1. Exercises on relative entropy.

Exercise 3.4 (Bernoulli laws). For $p, q \in (0, 1)$, show that

$$D(p \parallel q) = p \log \frac{p}{q} + (1 - p) \log \frac{1 - p}{1 - q}.$$

Use one numerical pair to verify that relative entropy is not symmetric.

Exercise 3.5 (Maximum entropy). Derive (3.8) directly and identify all equality cases in

$$0 \leq H(\nu) \leq \log M.$$

4. GIBBS EQUILIBRIUM AND FREE ENERGY

Let $E : \Omega \rightarrow \mathbb{R}$ be an energy and let $\beta > 0$ be inverse temperature. Define

$$(4.1) \quad Z_\beta := \sum_{x \in \Omega} e^{-\beta E(x)}$$

and

$$(4.2) \quad \mu_\beta(x) := \frac{e^{-\beta E(x)}}{Z_\beta}.$$

The normalization Z_β is the partition function.

For a trial probability law ν , define the Helmholtz free energy

$$(4.3) \quad \mathcal{F}_\beta(\nu) := \sum_x E(x) \nu(x) - \frac{1}{\beta} H(\nu).$$

The first term rewards low energy, while the negative entropy term rewards spreading mass among many states.

Theorem 4.1 (Finite Gibbs variational principle). *For every probability law ν ,*

$$(4.4) \quad D(\nu \| \mu_\beta) = \beta (\mathcal{F}_\beta(\nu) - \mathcal{F}_\beta(\mu_\beta)).$$

Moreover,

$$(4.5) \quad \mathcal{F}_\beta(\mu_\beta) = -\frac{1}{\beta} \log Z_\beta.$$

Consequently, the Gibbs law is the unique minimizer of free energy.

Stopping point

Meeting 3 ended with the statement of Theorem 4.1. Its proof begins Meeting 4.

5. FURTHER READING

The use of Jensen’s inequality to prove Gibbs inequality is standard; see Cover–Thomas [2]. Chafaï’s mini-course [1] and Tong’s notes [3] connect relative entropy, Gibbs laws, and free energy.

REFERENCES

- [1] D. Chafaï, *Mini-course on logarithmic Sobolev inequalities*, lecture notes, 2024, <https://djalil.chafai.net/docs/chafai-mini-lsi-2024.pdf>.
- [2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.
- [3] D. Tong, *Lectures on Statistical Physics*, University of Cambridge, <https://www.damtp.cam.ac.uk/user/tong/statphys.html>.