

21-317 MEETING 2: FINITE PROBABILITY AND ENTROPY REVISED STUDENT NOTES REFLECTING THE MATERIAL COVERED

MATTHEW ROSENZWEIG

ABSTRACT. Meeting 2 introduced expectation and variance on finite probability spaces, interpreted variance as squared distance from a function to the constants, analyzed the two-point space, introduced Shannon–Boltzmann entropy, and ended with the additivity of entropy for independent systems. Relative entropy and Gibbs equilibrium begin in Meeting 3.

CONTENTS

1. Finite probability, expectation, and variance	1
1.1. The mean is the best constant approximation	2
1.2. The two-point space	2
2. Shannon–Boltzmann entropy	2
2.1. Independent systems	3
3. Exercises	4
4. Further reading	4
References	4

1. FINITE PROBABILITY, EXPECTATION, AND VARIANCE

Let Ω be a finite set and let μ be a probability law on Ω :

$$(1.1) \quad \mu(x) \geq 0, \quad \sum_{x \in \Omega} \mu(x) = 1.$$

For $f : \Omega \rightarrow \mathbb{R}$, define

$$(1.2) \quad \mathbb{E}_\mu f := \sum_{x \in \Omega} f(x) \mu(x)$$

and

$$(1.3) \quad \text{Var}_\mu(f) := \mathbb{E}_\mu[(f - \mathbb{E}_\mu f)^2].$$

Expanding the square gives

$$(1.4) \quad \text{Var}_\mu(f) = \mathbb{E}_\mu[f^2] - (\mathbb{E}_\mu f)^2.$$

Date: August 26, 2026; revised August 31, 2026.

1.1. The mean is the best constant approximation.

Proposition 1.1. *For every $a \in \mathbb{R}$,*

$$(1.5) \quad \mathbb{E}_\mu[(f - a)^2] = \text{Var}_\mu(f) + (a - \mathbb{E}_\mu f)^2.$$

Consequently,

$$(1.6) \quad \text{Var}_\mu(f) = \min_{a \in \mathbb{R}} \mathbb{E}_\mu[(f - a)^2],$$

and the minimizing constant is $a = \mathbb{E}_\mu f$.

Proof. Let $m = \mathbb{E}_\mu f$. Then

$$\begin{aligned} \mathbb{E}_\mu[(f - a)^2] &= \mathbb{E}_\mu[(f - m + m - a)^2] \\ &= \mathbb{E}_\mu[(f - m)^2] + 2(m - a)\mathbb{E}_\mu[f - m] + (m - a)^2 \\ &= \text{Var}_\mu(f) + (m - a)^2. \end{aligned}$$

□

Interpretation

Variance is the squared $L^2(\mu)$ distance from f to the constant functions:

$$(1.7) \quad \text{Var}_\mu(f) = \inf_{a \in \mathbb{R}} \|f - a\|_{L^2(\mu)}^2.$$

A Poincaré inequality will later control this global fluctuation by a local energy.

1.2. The two-point space. Let $\Omega = \{-1, +1\}$ with the uniform law, and suppose

$$f(-1) = a, \quad f(+1) = b.$$

Then

$$\mathbb{E}f = \frac{a + b}{2}$$

and

$$(1.8) \quad \text{Var}(f) = \frac{1}{4}(a - b)^2.$$

The natural discrete derivative is the finite difference $b - a$. Thus

$$(1.9) \quad \text{Var}(f) = \frac{1}{4}|f(+1) - f(-1)|^2.$$

This is already a complete local-to-global identity.

Remark 1.2 (Biased two-point law). If $\mu(+1) = p$ and $\mu(-1) = 1 - p$, then

$$(1.10) \quad \text{Var}_\mu(f) = p(1 - p)(b - a)^2.$$

2. SHANNON–BOLTZMANN ENTROPY

All logarithms are natural. We use the convention $0 \log 0 = 0$.

Definition 2.1 (Entropy). For a probability law ν on a finite set Ω , define

$$(2.1) \quad H(\nu) := - \sum_{x \in \Omega} \nu(x) \log \nu(x).$$

A concentrated law

If $\nu = \delta_{x_0}$, then

$$(2.2) \quad H(\nu) = 0.$$

A uniform law

If u_M is uniform on a set of cardinality M , then

$$(2.3) \quad H(u_M) = \log M.$$

For a uniform ensemble, entropy is exactly the logarithm of the number of available microscopic states.

Information content

If $X \sim \nu$, then

$$(2.4) \quad H(\nu) = \mathbb{E}_\nu[-\log \nu(X)].$$

The number $-\log \nu(x)$ is the information content, or surprisal, of observing x .

2.1. Independent systems. Let ν be a law on Ω and η a law on Λ . Their independent joint law is

$$(2.5) \quad (\nu \otimes \eta)(x, y) = \nu(x)\eta(y).$$

Proposition 2.2 (Additivity of entropy). *For independent finite systems,*

$$(2.6) \quad H(\nu \otimes \eta) = H(\nu) + H(\eta).$$

Proof. Using (2.5) and the logarithm rule,

$$\begin{aligned} H(\nu \otimes \eta) &= - \sum_{x,y} \nu(x)\eta(y) \log(\nu(x)\eta(y)) \\ &= - \sum_{x,y} \nu(x)\eta(y) \log \nu(x) - \sum_{x,y} \nu(x)\eta(y) \log \eta(y) \\ &= H(\nu) + H(\eta). \end{aligned}$$

□

Why logarithms appear

Multiplicities multiply for independent systems, while entropies add. The logarithm converts products into sums. This additivity, often called extensivity, is fundamental in statistical mechanics and information theory.

Where the course continues

Meeting 3 introduces Jensen's inequality and relative entropy, which compares a probability law with a specified reference law.

3. EXERCISES

Exercise 3.1 (Variance as best approximation). Prove Proposition 1.1 without consulting the proof, and explain why the minimizer is unique.

Exercise 3.2 (Biased two-point variance). Prove (1.10).

Exercise 3.3 (Repeated independent systems). If $\nu^{\otimes N}$ is the law of N independent copies of a system with law ν , prove

$$H(\nu^{\otimes N}) = NH(\nu).$$

Exercise 3.4 (Entropy of a Bernoulli law). For $p \in [0, 1]$, let $\nu_p = (p, 1 - p)$. Compute $H(\nu_p)$ and determine its values at $p = 0, 1/2, 1$.

4. FURTHER READING

The definition and additivity of Shannon entropy are standard; see Cover–Thomas [2]. Chafaï’s short notes connect finite entropy to Gibbs equilibrium and functional inequalities [1]. Tong [4] and Sethna [3] give accessible statistical-mechanical discussions of multiplicity, entropy, and extensivity.

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] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity*, 2nd ed., Oxford University Press, 2021, <https://sethna.lassp.cornell.edu/StatMech/>.
- [4] D. Tong, *Lectures on Statistical Physics*, University of Cambridge, <https://www.damtp.cam.ac.uk/user/tong/statphys.html>.