

21-317 FUNCTIONAL INEQUALITIES: FROM ANALYSIS TO PROBABILITY
ROUND I ALTERNATE TOPIC PACKET 09
PINSKER INEQUALITY AND ENTROPY-TO-TOTAL-VARIATION
CONVERGENCE

MATTHEW ROSENZWEIG

ABSTRACT. Organizing question: How does relative entropy control probabilistic distinguishability? This packet fixes the mathematical scope and normalization, identifies the required proof content, and provides a curated reading route. It is intended both for informed preference ranking and for preparation after assignment.

AT A GLANCE

Status	Alternate Round I topic
Setting	Information theory on finite probability spaces
Primary tools	Relative entropy, log-sum inequality, coarse-graining, convexity
Background	Finite probability; single-variable calculus
Relative difficulty	2/4
Principal perspective	Information theory and convergence to equilibrium

1. WHY THIS TOPIC MAY APPEAL TO YOU

This alternate is a focused information-theoretic topic with a complete elementary proof. It explains how an analytic entropy estimate becomes an operational probabilistic statement about events and mixing.

2. FIXED CONVENTIONS

For probability measures ν, μ on a finite set with $\mu(x) > 0$, use natural logarithms and define

$$D(\nu\|\mu) = \sum_x \nu(x) \log \frac{\nu(x)}{\mu(x)}, \quad \|\nu - \mu\|_{\text{TV}} = \frac{1}{2} \sum_x |\nu(x) - \mu(x)|.$$

3. PRINCIPAL THEOREM

Prove Pinsker's inequality

$$\|\nu - \mu\|_{\text{TV}} \leq \sqrt{\frac{1}{2} D(\nu\|\mu)}.$$

Then deduce from

$$D(\nu_t\|\mu) \leq e^{-\alpha t} D(\nu_0\|\mu)$$

that

$$\|\nu_t - \mu\|_{\text{TV}} \leq \frac{1}{\sqrt{2}} e^{-\alpha t/2} \sqrt{D(\nu_0 \| \mu)}.$$

4. REQUIRED PROOF ROUTE

- (1) Prove the log-sum inequality and use it to show relative entropy decreases under a two-set coarse-graining.
- (2) Let $A = \{x : \nu(x) \geq \mu(x)\}$ and prove $\nu(A) - \mu(A) = \|\nu - \mu\|_{\text{TV}}$.
- (3) Reduce to Bernoulli parameters $p = \nu(A)$ and $q = \mu(A)$.
- (4) By Taylor's theorem and $[p(1-p)]^{-1} \geq 4$, prove

$$p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \geq 2(p-q)^2.$$

5. INTERPRETATION, EXERCISES, AND PITFALLS

Explain total variation as the maximal discrepancy on an event and relative entropy as a nonsymmetric information discrepancy. Exercises may check the two-point case, prove the event characterization of total variation, or translate an entropy-decay rate into a mixing estimate. The orientation of $D(\nu \| \mu)$ matters; Pinsker is one-way; and changing the $1/2$ convention in total variation changes the constant.

6. SUGGESTED READING ROUTE AND ATTRIBUTION

- (1) R. van Handel, *Probability in High Dimension*, Section 4.1, especially the discussion around pp. 80–82 of Pinsker and transportation-type inequalities ([public course notes](#)).
- (2) I. Csiszar, “Information-type measures of difference of probability distributions and indirect observations,” *Studia Scientiarum Mathematicarum Hungarica* 2 (1967), 299–318, for the broader divergence framework and data-processing viewpoint.

7. BEFORE THE REQUIRED CONSULTATION

Bring a completed one-page preparation document containing: the exact theorem statements; a dependency map of the proof; the portion you will prove in full; planned examples; two exercises with solutions; exact source locations; anticipated questions; and any point you do not yet understand. The consultation is for resolving mathematical and expository issues, not for choosing a topic from scratch.

8. AI-USE DISCLOSURE

The preparation document must disclose any use of generative artificial-intelligence tools. Identify the tool, the task for which it was used, material incorporated into the presentation, and how mathematical assertions and references were independently verified. You remain responsible for every statement in the talk.

This is an instructor-authored assignment guide, not a substitute for reading the listed sources. All normalizations in the talk should agree with this packet unless a change is explicitly approved in advance.