Mathematical Physics

Course Instructor: Michael Aizenman

February 6, 2016

Contents

Introduction

Notes from Michael Aizenman's class "Mathematical Physics" at Princeton in Spring 2016.

This class is of interest to both physicists and mathematicians. Several recent Fields medals are for work related to these topics.

I plan to cover the following topics. The focus is on **Topics in Mathematical Statistic Mechanics**.

1. **The statistical mechanic perspective**: systems can be described at the microscopic level with many degrees of freedom. We observe their collective behavior and find emergent behavior.

2. **Thermodynamics principles**: Intellectually preceding statistical mechanics is thermodynamics, a field of physics which emerged through experimental and intellectual work trying to understand what is happening with the transfer of heat. A couple of principles emerged. This framework is more appropriate to the macroscopic description of physical systems. In departure from mechanics, which cares about equalities like $F = ma, E = mc^2$, a unique thing about thermodynamics is that its key principle is in inequality: entropy increases.
$$\Delta S \geq 0$$

3. **The emergence of thermodynamics from statistical mechanics via the equidistribution assumption and the "large deviation theory"**. I would like to discuss the emergence of thermodynamics from statistical mechanics. Mathematicians formalized the theory but the concepts were introduced earlier by physicists. These principles led Boltzmann to introduce these ideas; it took a while for physics to absorb the ideas, which may have led to his premature death via suicide.

4. **Phase transitions**: When you have a system, say H_2O, which you can control with temperature and pressure, you can induce changes in state. Continuous change in control parameters results in discrete jumps in result.

5. **Critical phenomena, critical exponents, universality classes**. Phase transitions are fascinating since there are interesting critical phenomena which are characterized by critical exponents, which turn out (this is one of the suprising discoveries experimentally) to result in universality classes of critical phenomena. Systems are macroscopically different, but the singularities you observe are given by the same power laws.

6. **Exact solution of the 2-D Ising model**. Mathematically, 3 is the hardest dimension to comprehend.

   (a) 1-D is solvable: correlations can be described by Markov chains, and can be computed.

   (b) In 2-D, the conformal group is very important. It gives many constraints on critical behavior, leading to a rich behavior.

   (c) In 3-D, this does not apply except for ongoing work finding consequences in 3-D from results in 2-D.

   (d) Anything with ≥ 4 dimensions gets simpler. High dimensions are characterized by the fact that loop effects do not play a large role. In 1-D a simple random walk is recurrent. In sufficient high dimensions, a simple random walk is not recurrent. The infinite-dimensional case reduces to models on trees; high dimensions exhibit the same behavior as infinite dimensions.

   In coding theory and discrete math people study phase transitions of different graphs and are interested in the same topics.

   The 2D Ising model is usually a highly specialized topic, but we will do it in a way done a bit differently from normal. We would like to do this through "graph zeta functions". What I found fascinating about this topics is that there are a lot of connections to other topics. There are analogues of the Riemann zeta function on graphs, and there is a relationship. It is actually one of the simplest paths to tackle this proof by.

7. **Stochastic geometry behind correlation functions at criticality**: in the Ising model we have a collection of spin variables $\sigma_x = \pm 1$ for $x \in \mathbb{Z}^d$. The spins are correlated in such a way that agreement among neighbors is encouraged. Thus there are correlations which spread through the system. It is interesting and powerful to represent this correlation between spins is via a "shadow system" in which you can play the following game: You decompose at random into connected clusters, and when you see spin values, you don't see who is connected to whom: But for each such decomposition the spins are connected on each cluster. You can think of it as forming cliques in class, and then each clique chooses what to do and votes unanimously. So if you just saw the voting pattern, you would see some cliques, but the nature of correlations among the votes become transparent if you know the clusters. These states for critical Ising models become longer and longer, but become fractal. Fractal geometric objects can be used to explain the structure. We will introduce all of this later in much detail from the ground up. If you have heard of percolation, this topic is related. There is an interesting fractal geometry which tells us about correlation functions.

8. **Scaling limits**: Finally, if time permits, I would like to discuss scaling limits. You may want to know properties of a substance when you have something like a lattice that is very fine-grained, a statistic mechanical system which macroscopically is described

by a myriad of variables. There are techniques of describing the relevant quantities from the macroscopic perspective, and this becomes relevant for describing the critical state. This provides a fascinating link between statistical mechanics and field theory. It's related to quantum field theory, except we are now in the Euclidean regime. There are some related results for quantum spin systems.

There is a broad spectrum of references, none of which I'll follow exclusively.

- Sacha Friedli and Yvan Velenik's online book project (on mathematical statistical mechanics) at http://www.unige.ch/math/folks/velenik/smbook/index.html. I recommend this for people totally new to the subject.
- David Ruelle formulated models and basic results mathematically (late 70's). His book helped physicists organize their thoughts. It became outdated quickly, but remains a good starting point and reference for the formalism.

Chapter 1

Introduction to statistical mechanics

2-2-16

1 The equivalence principle

1.1 Configurations and ensembles

One way to start is with the axioms of statistical mechanics. Instead I'll take a simple problem, see how it works, and present results in that context. There are simple problems that teach us a lot. ("The elementary problems are the most precious, once you absorb them they are part of your makeup.") The simplest is a lattice gas.

A lattice gas is a substrate where at each lattice site there may or may not be a particle.

The **configuration** is a function $n : \mathbb{Z}^d \to \{0, 1\}$:

$$n_x = \begin{cases} 1, & x \text{ is occupied,} \\ 0, & x \text{ is vacant.} \end{cases}$$

I'll use L to denote the size of the box we are considering, and $\Lambda \subset \mathbb{Z}^d$ to be a region (subset of the system). The **configuration space** is the space of possible n's, $\{0, 1\}^\Lambda$.

We assume conversation in the number of particles, and that particles cannot overlap. We make the **equidistribution assumption**: every configuration with the same number of particles has equal probability. This gives rise to the **microcanonical ensemble**. An ensemble is a probability measure with respect to which you do averages. We have

$$\mathbb{P}(n_\Omega) = \frac{\mathbb{1}[\sum_{x \in \Omega} n_x = N]}{Z}$$

where Z is a normalization constant. Here, $\mathbb{1}[\text{cond}] := \begin{cases} 1, & \text{condition satisfied} \\ 0, & \text{else} \end{cases}$.

For every function $f : \Omega \to \mathbb{R}$ assigning a real number to each configuration, define the **microcanonical ensemble average** by

$$\langle f \rangle_{N,n}^{\text{Can}} = \frac{\sum_{n\in\Omega} \mathbb{1}[n_x = N] f(n)}{\sum_{n\in\Omega} \mathbb{1}[\sum n_x = N]}.$$

Loosely, the **equivalence principle** says that for any "local function", the microcanonical average is approximately the grand canonical ensemble average

$$\langle f \rangle_{N,\Lambda}^{\text{Can}} \approx \langle f \rangle_{\mu,\Lambda}^{\text{Gr.C}}$$

when we take $\mu = \frac{N}{|\Lambda|}$.

The **grand canonical ensemble average** is defined as

$$\langle f \rangle_{\mu,\Lambda}^{\text{Gr.C}} = \frac{\sum_{n\in\Omega} e^{-\mu \sum_{x\in\Lambda} n_x} f(n)}{\sum_{n\in\Omega} e^{-\mu \sum_{x\in\Lambda} n_x}}$$

(Later on we will omit superscripts where it is clear.)

1.2 Equivalence principle

Consider functions which depend only on a system $\Lambda \subseteq \widetilde{\Lambda}$ of much smaller volume, $|\Lambda| \ll |\widetilde{\Lambda}|$. This is the sense in which the averages match up.

The micro-canonical ensemble is draconian: the number of particles is prescribed, all other configurations get weight 0. In the grand canonical ensemble, each configuration contributes. There is a value of μ where the density is the same; at that value the local average of the draconian system is asymptotically the same at that of the more relaxed system. This $\approx$ becomes $=$ when you take the thermodynamic limit,

$$\begin{aligned} \widetilde{\Lambda} &\to \mathbb{Z}^d \\ N &\to \infty \\ \frac{N}{|\widetilde{\Lambda}|} &\to \rho. \end{aligned}$$

What is the induced distribution of the micro-canonical ensemble on Λ? Under the micro-canonical ensemble what is the probability that n_Λ (the restriction to Λ) takes a particular value with $\sum_n n_x = k$? We count the number of ways to complete the configuration in $\Lambda^c = \widetilde{\Lambda} \backslash \Lambda$:

$$\frac{|\{n_{\Lambda^c} : \sum_{x\in\Lambda^c} n_x = N - k\}|}{Z}$$

where Z is a normalization constant.

The number of configurations of M particles in volume V is $\binom{V}{M} = \frac{V!}{M!(V-M)!}$. Using Stirling's approximation

$$\ln(M!) = M(\ln M - 1)(1 + o(1)),$$

letting the **entropy** S be the logarithm of the number of configurations and

$$s(\rho) = -[\rho \ln \rho + (1-\rho)\ln(1-\rho)],$$

we have

$$\binom{V}{M} = \frac{V!}{M!(V-M)!} =: e^{S(M,V)} \approx e^{Vs(\rho)}$$

(do this calculation as an exercise).

Shannon also found such a formula for entropy.

This attains maximum of $\ln 2$ at $\frac{1}{2}$ where it has quadratic behavior.

The implication is that if you slightly change the density, the number of configurations changes drastically. In physical substances V may be 10^{23}. The change would then be $e^{10^{23}\Delta s}$. In any average over configurations, only those at the peak contribute: "winner takes all."

What is the probability of observing k particles in the small box given n in the big box?

$$\mathbb{P}(n_\Lambda) \approx \frac{e^{|\Lambda^c| s\left(\frac{N-k}{|\widetilde{\Lambda}|-|\Lambda|}\right)}}{C}$$

$$\frac{N-k}{|\widetilde{\Lambda}|-|\Lambda|} = \rho - \underbrace{\left(\rho - \frac{N}{|\widetilde{\Lambda}|-|\Lambda|}\right)}_{\to 0 \text{ independent of } k} - \frac{k}{|\widetilde{\Lambda}|-|\Lambda|}$$

$$s\left(\frac{N-k}{|\widetilde{\Lambda}|-|\Lambda|}\right) \approx s(\rho) - s'(\rho)\frac{k}{|\Lambda^c|}$$

Changing k by a little bit affects how many particles are outside but not so much the density outside: the correction term is small. Hence for $\sum_{x\in\Lambda} n_x$,

$$\mathbb{P}(n_\Lambda) \approx \frac{e^{|\Lambda^c| s(\rho)} e^{-s'(\rho)k}}{Z}.$$

$e^{|\Lambda^c| s(\rho)}$ is a huge factor but it does not vary with k so we can omit it. Writing out Z as the sum of the numerators over all n_Λ, and letting $\mu = s'(\rho)$, we get that this equals

$$= \frac{e^{-\mu k}}{\sum_{n'\in\Omega_\Lambda} e^{-\mu \sum_{x\in\Lambda} n'_x}},$$

which is the grand canonical ensemble average. The rest of the system acts on the small system as a "particle (heat) bath."

Here we used very explicit machinery, namely, the Stirling formula. We want a general expression that doesn't rely on the Stirling formula because in more complicated models, we will not have the luxury of using the Stirling approximation. We'll do this in the next section.

Then we will be able to consider models where there are more energy constraints. The general procedure is first make a list of energy constraint. There is a generalization of the equivalence principle where

- in the micro-canonical ensemble we average over configurations where the constraints have prescribed values. (We considered the special case where just the number of particles was prescribed.)

- in the grand canonical ensemble, we add an energy term to the exponential: $e^{-\mu N(n)-\beta\mathcal{E}(n)}$ where $N(n)$ is the number of particles, $-\mu N(n)$ is the Gibbs factor, and $\mathcal{E}(n)$ the energy.

How can we construct an alternative method without the Stirling formula?
2-4-16

1.3 Second take

Define the **partition function** for the grand canonical ensemble as the (unnormalized) sum of the likelihoods over all configurations. At each site there are 2 possibilities and they are independent, so the sum factors into a product.[1]

$$\begin{aligned} Z_\Lambda &= \sum_{n\in\Omega} e^{-\mu N(n)} \\ &= \sum_{n\in\Omega}\prod_{x\in\Lambda} e^{-\mu \mathbb{1}[n_x=1]} \\ &= \prod_{x\in\widetilde{\Lambda}} (1+e^{-\mu}) \\ &= (1+e^{-\mu})^{|\Lambda|} \\ \frac{1}{|\Lambda|}\ln Z_\Lambda &= 1+e^{-\mu} \end{aligned} \tag{1.1}$$

eq:z1

We calculate Z_Λ a different way, as follows. Note that the only thing that matters in the summand is the number of particles in n, so let's group the summands by this. Letting $\rho = \frac{N}{|\Lambda|}$, and using the fact that the number of states with density ρ is $e^{|\Lambda|s(\rho)}$,[2]

$$\begin{aligned} Z_\Lambda &= \sum_{n\in\Omega} e^{-\mu N(n)} \\ &\approx \sum_{\rho\in\frac{1}{|\Lambda|}\mathbb{Z}} e^{-\mu N(n)} e^{|\Lambda|s(\rho)} \\ &= \sum_{\rho\in\frac{1}{|\Lambda|}\mathbb{Z}} e^{|\Lambda|[s(\rho)-\mu\rho]} \\ &\approx e^{|\Lambda|\max_{\rho\in[0,1]}[s(\rho)-\mu\rho]}. \end{aligned} \tag{1.2}$$

eq:z2

[1] Note that Λ in the last section is the micro-canonical ensemble, but here it is the grand canonical ensemble.

[2] Rather than getting this with $s(\rho) = -[\rho\ln(\rho) - (1-\rho)\ln(1-\rho)]$ by the Stirling approximation in the last section, we can take this as the *definition* of s here, and use it to solve for s.

To get the last step, note that the error we are making by focusing on the maximal point rather than counting with multiplicity the near-maximal value is at most a factor equal to the volume. But $\frac{1}{|\Lambda|}\ln|\Lambda| \to 0$. (Details to follow.)

The expression $\max_{\rho\in[0,1]}[s(\rho)-\mu\rho]$ is called the **Legendre transform** of the entropy and denoted by $s^*(\rho)$. Matching (1.1) and (1.2),

$$s^*(\mu) := \max_\rho[s(\rho)-\mu\rho] = \ln(1+e^{-\mu}) \tag{1.3}$$

eq:s-leg

Using this, we can derive the expression the formula for $s(\rho)$, using the inverse Legendre transform.

To check that taking the maximum in (1.2) was legit, note that[3]

$$\max_\rho[s(\rho)-\mu\rho] \le \frac{1}{|\Lambda|}\ln Z_\Lambda \le \max_\rho[s(\rho)-\mu\rho] + \underbrace{\frac{1}{|\Lambda|}\ln|\Lambda|}_{\to 0}$$

$$\implies \lim_{|\Lambda|\to\infty}\frac{1}{|\Lambda|}\ln Z_\Lambda = \sup_{0\le\rho\le 1}[s(\rho)-\rho\mu].$$

We can verify that $s(\rho) = -[\rho\ln\rho + (1-\rho)\ln(1-\rho)]$ does indeed make (1.3) hold. We find the maximum (critical point) or $s(\rho)-\mu\rho$ by setting the derivative to 0. We use the trick

$$[x(\ln x - 1)]' = \ln x.$$

We can subtract 1 from each of the logs changing the expression by a constant. Thus

$$s'(\rho) = -\ln\rho + \ln(1-\rho) - \mu \implies \frac{\rho}{1-\rho} = e^{-\mu}.$$

We can use this to solve for ρ and find $s^*(\mu)$ (do this yourself).

1.4 Legendre transform

In general, a micro-canonical ensemble specifies all the conserved quantities: Particles, energies, and whatever else there is. The grand canonical ensemble also generalizes by changing the factor to $e^{-\mu N}e^{\beta H}$. Such factors are referred to as **Gibbs factors** or Gibbs measures. Our next topics are:

- Legendre transform
- Convexity/concavity
- A variational characterization of Gibbs states
- First order phase transitions (thermodynamics and statistical mechanics perspectives)

[3] Mathematicians are paranoid, so we use sup instead of max. For many practical purposes they are the same.

Definition 1.1.1: A function on $\mathbb{R}^k$ is **concave** (**convex**) if for any $x_0, x_1 \in \mathbb{R}^k$, $0 \leq \lambda \leq 1$,

$$F(\lambda x_1 + (1-\lambda)x_0) \begin{matrix} \leq \\ \geq \end{matrix} \lambda F(x_1) + (1-\lambda)F(x_0).$$

Concavity (convexity) means if you draw a chord between two points, it will lie below (above) the curve.

For a strictly concave function, a maximum, whenever it exists, is unique.

Theorem 1.1.2: For any concave function on $\mathbb{R}$,

1. The directional derivatives $F'(x \pm 0)$ exist at all $x \in \mathbb{R}$. The **directional derivative** is defined as
$$F'(x+0) = \lim_{\varepsilon \to 0^+} \frac{F(x+\varepsilon) - F(x)}{\varepsilon}$$
$$F'(x-0) = \lim_{\varepsilon \to 0^-} \frac{F(x+\varepsilon) - F(x)}{\varepsilon}.$$

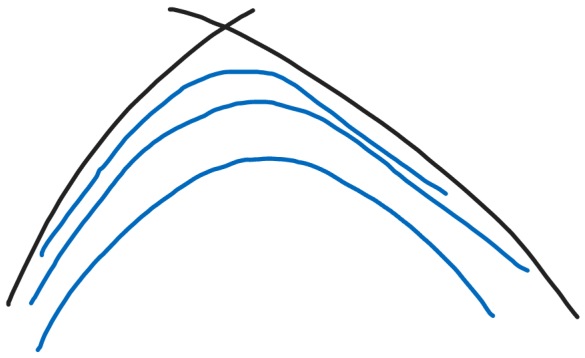
2. $F'(x-0) \geq F'(x+0)$ and $F'(x \pm 0)$ are decreasing.

3. For all but countably many values $x \in \mathbb{R}$, $F'(x-0) = F'(x+0)$, i.e., F is differentiable at x.

4. Let F_n be a sequence of concave functions which converge pointwise: for all x, $\lim_{N\to\infty} F_N(x) =: \widetilde{F}(x)$ exists. Then

   (a) $\widetilde{F}$ is concave.

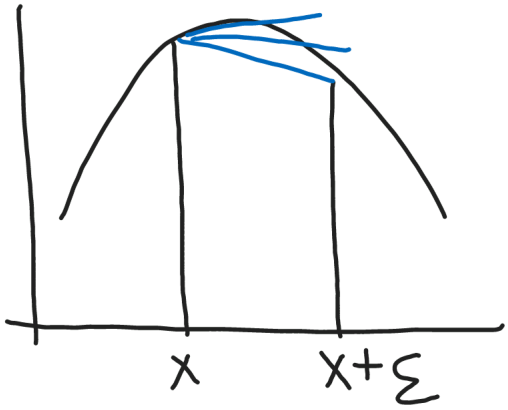
   (b) At points of differentiability of $\widetilde{F}$, the derivatives also converge,[4]
   $$F'(x \pm 0) \to \widetilde{F}'(x).$$



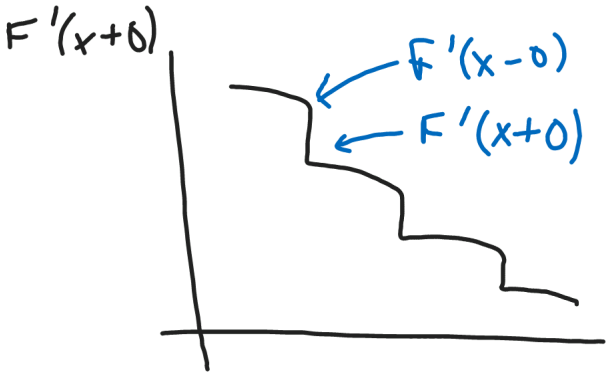
[4]Finite energy functions are always smooth, but their limit can have discontinuous derivative.

Much of this generalizes to directional derivatives in n dimensions.

Proof. For (1) note that concavity implies that the slope of the line between $x, x+\varepsilon$ is increasing as $\varepsilon \to 0^+$.



To prove 3, think of the graph of the derivative at the right, $F'(x+0)$. It is decreasing. Any monotonic function has a countable number of discontinuity points because the sum of any uncountable number of steps is infinite.[5] [6]



□

Definition 1.1.3: The **Legendre transform** of a function is defined as

$$(TG)(y) = \inf_x [y \cdot x - G(x)] = -\sup_x [G(x) - y \cdot x]$$

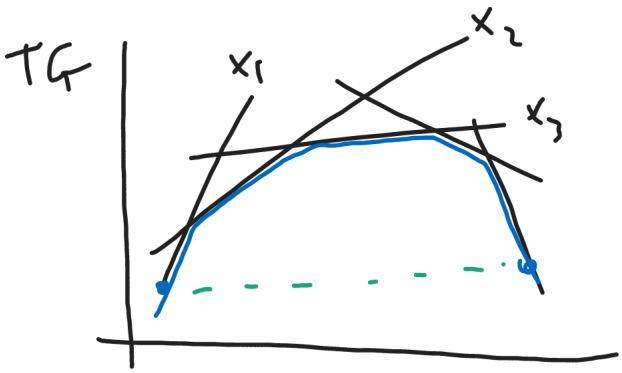
Theorem 1.1.4: For any function G, TG is concave.

[5] More rigorously, for every nonzero interval $[F'(x+0), F'(x-0)]$, we can associate with it a rational number. Then note $\mathbb{Q}$ is countable.

[6] The set of discontinuities can be dense, ex. all the rational numbers. In physics it used to be thought that this can't occur, but there are materials whose free energy discontinuities are dense in certain areas.

Proof. An efficient way to think about this is the following: for each value of the parameter x, as a function of y this is a linear function. For each x we get a linear function. Define the transform by taking the infimum over that.

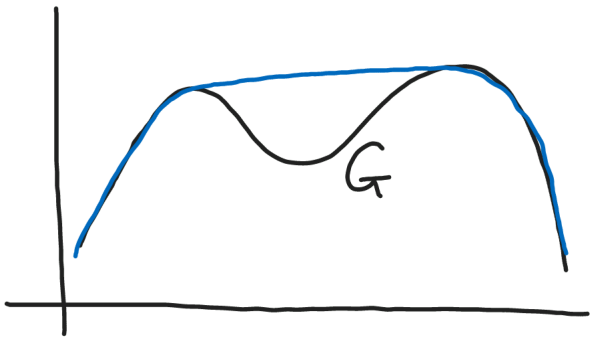
Take 2 points and draw the chord between them. For each linear function the chord lies below it.[7]



□

Definition 1.1.5: The **concave hull** $\widetilde{G}$ of G is defined as the smallest concave function that is at least the function value at every point:

$$\widetilde{G}(x) = \inf \{F(x) : F \text{ concave}, \forall u,\, F(u) \geq G(u)\}\,.$$



Theorem 1.1.6: For *concave* G,

$$T(TG) = G.$$

In general, $T(TG)$ is the concave hull of G.

[7] I.e., we use the following : Let $\mathcal{F}$ be a collection of concave (e.g. linear) functions. Then $\inf_{f\in\mathcal{F}} f(x)$ is concave.

Proof for G differentiable. Use the fact that if G is differentiable, then

$$\inf_x [y \cdot x - G(x)]$$

occurs at $y = G'(x)$. □

Note that we plotted the function and the dual function on the same graph. However, they have inverse units, for example, energy and inverse temperature. You get a lot of insights into physics if you keep track of the units.

Recall $\rho = \frac{N}{|\Lambda|}$. If all $2^{|\Lambda|}$ configurations are given equal weight, the typical value or ρ, the particles per unit volume is $\frac{1}{2}$. The Law of Large Numbers says that with probability 1 the ratio tends to $\frac{1}{2}$. Is it possible that the density is $\frac{1}{3}$? Yes, but the probability of such a density is given by the entropy: it's exponentially small, $e^{|\Lambda|[s\left(\frac{1}{3}\right)-\ln 2]}$. Anything other than $\frac{1}{2}$ is a large deviation; they occur with exponentially small probability. We want to quantify the probability of large deviation events. Here is a language that people found useful.

Definition 1.1.7: A sequence of probability measures on $\mathbb{R}$ is said to satisfy a **large deviation principle** with speed $\{a_N\}$, and rate function I if for each $x \in \mathbb{R}$, $\varepsilon > 0$,

$$-\inf_{|u-x|<\varepsilon} I(u) \leq \underline{\lim}_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}_N((x-\varepsilon, x+\varepsilon]) \leq \overline{\lim}_{N\to\infty} \frac{1}{a_N} \ln \mathbb{P}_N((x-\varepsilon, x+\varepsilon]) \leq -\inf_{|x-u|\leq\varepsilon} I(u)$$

For us, $I(x) = \ln 2 - s(x)$.