May 9 – July 21, 1985 p. 1–130

May 9, 1985

Let $v = (v_j)$ be an eigenvector belonging to the eigenvalue λ for a matrix $A = (a_{ij})$ with $a_{ij} \geq 0$. Then

$$\lambda v_i = \sum_j a_{ij} v_j \Rightarrow |\lambda||v_i| = \left|\sum_j a_{ij} v_j\right| \leq \sum_j a_{ij} |v_j| \leq \left(\sum_j a_{ij}\right) \max |v_j|$$

Assume A stochastic: $\sum_j a_{ij} = 1$ for all i. Taking i to be such that $|v_i| = \max |v_j|$ we see that all eigenvalues of a stochastic matrix satisfy $|\lambda| \leq 1$.

Next suppose all $a_{ij} > 0$ and choose i as before. The two inequalities above become equalities. The second equality forces $|v_j| = |v_i|$ for all j, and the first forces the numbers $a_{ij} v_j$ for different j to have the same phases. Hence we have $v_j = v_i$ for all j. Thus a strictly positive stochastic matrix has a unique eigenvalue such that $|\lambda| = 1$; it is $\lambda = 1$ and the corresp. eigenvector is $v_j = 1$ for all j.

There is an immediate generalization of this conclusion to the case where A is "primitive", that is, A^n is strictly positive for some n. A permutation matrix is doubly-stochastic and has roots of unity for its eigenvalues. So one can not expect 1 to be the only eigenvalue of abs. value 1 in general.

May 11, 1985

Let $P = (p_{ij})$ be stochastic and such that $p_{ij} > 0$ for all i, j. Suppose $p_{ij} \geq \varepsilon$ with $\varepsilon > 0$. Given $v = (v_j)$ [redacted] then we can show $P^n v \longrightarrow$ constant vector as follows. Set

$$m = \min(v_j) \qquad M = \max(v_j)$$

Then

$$\left(\sum_j p_{ij} v_j\right) - m = \sum_j p_{ij}(v_j - m) \geq \varepsilon(M-m)$$

$$M - \left(\sum_j p_{ij} v_j\right) = \sum_j p_{ij}(M - v_j) \geq \varepsilon(M-m)$$

so

$$\sum_j p_{ij} v_j \quad \begin{array}{l} \geq m + \varepsilon(M-m) \\ \leq M - \varepsilon(M-m) \end{array}$$

which shows that the min and max for Pv is strictly contained in (m, M). In fact the new interval is of length

$$\leq M - \varepsilon(M-m) - (m + \varepsilon(M-m)) = (1-2\varepsilon)(M-m).$$

So we conclude that $P^n v \to$ constant vector exponentially

---

Let $A = (a_{ij})$ with $\overset{\text{all}}{\wedge}\, a_{ij} \geq 0$. If $v = (v_j)$ is an eigenvector $\sum_j a_{ij} v_j = \lambda v_i$ for A with all $v_j > 0$, then A is conjugate to $v_i^{-1} a_{ij} v_j$ which is $\lambda \times$ stochastic. Thus once it is demonstrated that A has a strictly positive eigenvector, [redacted] the rest of the story should follow from the stochastic case.

We will now take $A = (a_{ij})$, all $a_{ij} > 0$ and a vector $v = (v_j)$ with $v_j > 0$. We let

$$m, M = \text{min and max of } \frac{w_i}{v_i} = \frac{\sum_j a_{ij} v_j}{v_i}$$

and we want to show this spread does not increase upon replacing v by w. We have

$$m v_i \leq \underbrace{\sum_j a_{ij} v_j}_{w_i} \leq M v_i$$

or

$$m v_j \leq w_j \leq M v_j \qquad \text{for all } j$$

Multiplying by a_{ij} and summing gives

$$m w_i \leq \sum_j a_{ij} w_j \leq M w_i$$

which gives the assertion.

But something more is true. I would like to show the spread strictly decreases upon replacing v by Av. We have

$$\left(\sum_j a_{ij} w_j\right) - m w_i = \sum_j a_{ij} (w_j - m v_j) \geq 0$$

with equality for some i iff $w_j = m \cdot v_j$ for all j.

Similarly

$$M w_i - \left(\sum_j a_{ij} w_j\right) = \sum_j a_{ij} (M v_j - w_j) \geq 0$$

with equality iff $w_j = M v_j$ for all j. Thus the spread strictly decreases unless v is an eigenvector.

This means that if we can find a v which maximizes $m = \min\limits_i \frac{\sum a_{ij} v_j}{v_i}$ or which minimizes $M = \max_i \dfrac{\sum a_{ij} v_j}{v_i}$, then we have the desired eigenvector. This must be the basis for the variational proof.

What we seek now is a $v = (v_j)$ with all $v_j > 0$ which maximizes

$$m(v) = \min_i \left(\frac{\sum_j a_{ij} v_j}{v_i} \right)$$

From the above we will be able to conclude that v is an eigenvector for A. (Proof. Put $m = m(v)$, ~~whereso~~ and let $w = Av$ so that all $w_i > 0$. By defn. of $m(v)$ one has

$$m v_j \le w_j \qquad \text{for all } j$$

so $$m w_i = m \sum a_{ij} v_j \overset{(*)}{\le} \sum a_{ij} w_j$$ for all i, whence we have $m(w) \ge m$. By the maximal property of v we have $m(w) = m$ so for some i $(*)$ is an equality. So for this i

$$\sum a_{ij} w_j - m w_i = \sum \underbrace{a_{ij}}_{>0} (w_j - m v_j) = 0$$

and we conclude $w = mv$, whence $Av = mv$ as claimed.)

One can establish the existence of v as follows. The set on which one seeks the maximum v has to be closed under A, and one wants to use compactness and continuity of $m(w)$. There are two methods. One would be to work on the simplex $\{w \mid w_j \ge 0, \ \sum w_j = 1\}$, but then one has to worry about the w_j which are zero. Another would be to exhibit a compact subset of the interior which is stable under A. Here I think of rays in $\{w \mid w_j \geq 0\}$ as being the same as points on the simplex. The compact set could be taken to be the image of A of the simplex.

Therefore at this point we have established the existence of a strictly positive eigenvector for A, where A is a strictly positive matrix. The proof goes as follows. We consider the effect of A upon rays in the cone of $v = (v_j)$ such that $v_j \geq 0$ for all j. Such a ray contains a unique point with $\sum_j v_j = 1$, so the space of rays is a simplex.

We pick any ray ρ in this simplex and consider the sequence $\rho, A\rho, A^2\rho, A^3\rho, \dots$ and take a limit point. For this sequence the number

$$m(v) = \min_i \frac{\sum_j a_{ij} v_j}{v_i}$$

is increasing. The function m is continuous on the interior of the simplex which contains A(simplex) which is compact, etc.

May 12, 1985:

Von Neumann L^2-ergodic thm: Let T be a contraction operator on a Hilbert space: $\|Tf\| \le \|f\|$ for all f. Then as $n \to \infty$

$$\frac{1}{n} \sum_{j=0}^{n-1} T^j f$$

converges to the orthogonal projection of f onto $\mathrm{Ker}(T-1)$. Also $\mathrm{Ker}(T^*-1) = \mathrm{Ker}(T-1)$.

First note that if $f \in \overline{\mathrm{Im}(1-T)}$, then the average $\frac{1}{n}\sum_{j=0}^{n-1} T^j f$ converges to zero. Call this average $A_n f$. In effect for any $\varepsilon > 0$ we can find a g with $\|f-(g-Tg)\| < \varepsilon$.

A_n is a contraction operator:

$$\|A_n(f)\| \le \frac{1}{n} \sum_{j=0}^{n-1} \|T^j(f)\| \le \frac{1}{n} n \|f\| = \|f\|.$$

Then $\varepsilon \ge \|A_n f - A_n(1-T)g\| = \|A_n f - \frac{g-T^n g}{n}\|$, so letting

$\ge \|A_n f\| - \|\frac{g-T^n g}{n}\|$

n go to infinity we get $\limsup_n \|A_n f\| \le \varepsilon$, for any $\varepsilon > 0$, so done.

Now $\overline{\mathrm{Im}(1-T)}^{\perp} = \mathrm{Ker}(1-T^*)$, so all we must do is show $\mathrm{Ker}(1-T^*) = \mathrm{Ker}(1-T)$. As T^* is also a contraction op. it will be enough to show $\mathrm{Ker}(1-T^*) \subset \mathrm{Ker}(I-T)$. But if $T^* f = f$, then

$$\|f - Tf\|^2 = \|f\|^2 - \langle f | Tf \rangle - \langle Tf | f \rangle + \|Tf\|^2$$

$$= \|f\|^2 - \langle T^* f | f \rangle - \langle f | T^* f \rangle + \|Tf\|^2$$

$$= -\|f\|^2 + \|Tf\|^2 \le 0$$

so $Tf = f$. QED.

Finally I should review the proof that $\|T^*\| = \|T\|$:

$$\|T\| = \sup_{\|v\|=1} \|Tv\| = \sup_{\|v\|=1} \left(\sup_{\|w\|=1} |\langle w | Tv \rangle| \right) = \sup_{\|v\|=\|w\|=1} |\langle w | Tv \rangle|$$

$$= \sup_{\|v\|=\|w\|=1} |\langle T^* w | v \rangle| = \|T^*\|$$

Consider a Markov chain on a finite set E given by a stochastic matrix $p(j,k)$, $j,k\in E$. One considers the space of sequences (x_n), $n \geq 0$ in E, denote this $\Omega = E^{\mathbb{N}}$. On this space one has probability measures P_x for each $x\in E$ which gives the probability of a sequence (x_n) knowing that it started at x. Thus

$$P_x(x_0 = e_0, \ldots, x_n = e_n) = \delta_x(e_0)\, p(e_0,e_1)\, p(e_1,e_2) \cdots p(e_{n-1},e_n)$$

Now

Stroock was in a situation where for each $n \geq 0$ he had shifting operators θ_k on Ω whose effect on a function $(x_n) \longmapsto f(x_0)$ was to yield the function $(x_n) \longmapsto f(x_k)$. Thus $\theta_k((x_n)) = (x_{k+n})$. The only way I can see how to make make these operators work on the space L^2 as a contraction is to use the following invariant probability measure.

Let's assume $p_{jk} > 0$ for all $j,k\in E$. Then by the Frobenius thm. there is a unique prob. measure μ on E (i.e. $\mu_j \geq 0$, $\sum_j \mu_j = 1$) with $\sum_j \mu_j p_{jk} = \mu_k$. Then we use the measure on Ω such that

$$P(x_0 = e_0, \ldots, x_n = e_n) = \mu_{e_0}\, p_{e_0 e_1} \cdots\cdots p_{e_{n-1} e_n}$$

Then when we consider $\theta_n : \Omega \longrightarrow \Omega$ it is easily seen that $(\theta_n)_* P = P$, and hence

$$\theta_n^* : L^2(\Omega, P) \longrightarrow L^2(\Omega, P)$$

is an isometric map.

May 14, 1985

I want to explain formally how Stroock links large deviations, the ergodic property, and the Sanov theorem. I begin with the law of large nos. For this we need a prob. space (E, ν) and a random variable X on it, by which one means a measurable map $X: E \longrightarrow V$ where V is another space. We form $(E^n, \nu^{\otimes n})$ and the 'identical' r.v.'s $X_j = X \circ pr_j$. We the take the average

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j .$$

For this to make sense the space V has to be a vector space, or at least a convex subspace of a vector space.

We then push forward the measure $\nu^{\otimes n}$ by the map $\bar{X}: E^n \longrightarrow V$ to obtain a measure μ_n on V. The weak law of large numbers says that

$$\mu_n \longrightarrow \delta_m \quad (\text{Dirac measure at } m \in V)$$

where $m = \int_E X \, d\nu$ is the mean of X. Thus if f is a real function on V, say bounded continuous, we have

$$\int_V f d\mu_n = \int_{E^n} f\left(\frac{1}{n}\sum_1^n X_j\right) \nu(dx_1) \cdots \nu(dx_n) \longrightarrow f(m) .$$

The universal random variable on E with values in a (good) vector space is

$$\begin{aligned} E &\longrightarrow \mathcal{M}(E) \\ x &\longmapsto \delta_x \end{aligned}$$

since any map $\varphi: E \longrightarrow V$ extends to measures

by
$$\varphi(\mu) = \int_E \varphi(x)\,\mu(dx)$$

So the law of large numbers for the universal random variable $x \longmapsto \delta_x$ says that if μ_n is the prob. measure on $\mathcal{M}(E)$ obtained by pushing forward ν^n under

$$\begin{aligned} E^n &\longrightarrow \mathcal{M}(E) \\ (x_1, \ldots, x_n) &\longmapsto \tfrac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n}), \end{aligned}$$

then
$$\mu_n \longrightarrow \delta_\nu .$$

I recall that the theory of large deviations improves on the (weak) law of large numbers by giving asymptotics of the convergence of μ_n to δ_m. Thus the Sanov thm. is the large deviation result for the universal r.v. on a prob. space.

May 16, 1985

Law of Large Numbers. Let f be a real r.v. on a prob. space (E, ν). Form $\Omega = (E^{N^*}, \nu \otimes \nu \otimes \cdots)$ and consider the sequence of r.r.v. on Ω

$$\frac{S_n}{n} = \frac{1}{n}\sum_{j=1}^{n} f(x_j)$$

The LLN says that $\frac{S_n}{n}$ converges to the constant r.v. with value the mean:

$$m = \int f \, d\nu$$

There are various kinds of convergence one may consider:

- convergence in measure (or probability)
- convergence in L^P
- convergence a.e.

$f_n \to 0$ in measure means $\forall \varepsilon > 0$, $\mu\{|f_n| \geq \varepsilon\} \to 0$ (equivalently, the distribution of f_n approaches δ_0).

Chebychev's inequality shows L^P conv $\Rightarrow$ conv. in measure:

$$\int |f_n|^P \geq \int_{|f_n| \geq \varepsilon} |f_n|^P \geq \varepsilon^P \mu\{|f_n| \geq \varepsilon\}$$

In the presence of a suitable domination, Lebesgue's dominated conv. thm. ~~shows~~ a.e. conv. $\Rightarrow$ L^P conv.

Finally a.e. conv $\Rightarrow$ conv. in measure. For if $f_n \to 0$ a.e. and $\varepsilon > 0$, then putting $A_n = \{|f_n| \geq \varepsilon\}$ we have

$$\bigcap_n \bigcup_{m \geq n} A_m = \{x \mid |f_n(x)| \geq \varepsilon \text{ infinitely often}\}$$

so $\mu(\quad\underline{\qquad}\quad) = 0$. Thus $\mu\left(\bigcup_{m \geq n} A_m\right) \to 0$

and so $\mu(A_n) \to 0$.

Returning again to the sequence $\frac{S_n}{n}$ we see that if $f \in L^2$, then $\frac{S_n}{n} \to m$ in L^2. (One can suppose $m = 0$, then $\|\frac{S_n}{n}\|^2 = \frac{n\|f\|^2}{n^2} \to 0$.) So Chebychev $\Rightarrow$ $\frac{S_n}{n} \to m$ in prob. which is the weak LLN.

So what I have just said is that the L^2(LLN) is clear by orthogonality and it implies the weak LLN. If $f \in L^1$, then Birkhoff's ergodic thm gives $\frac{S_n}{n} \to m$ a.e. which is the strong LLN.

So far I have been looking at the LLN for real r.v.'s, but ultimately one wants to consider r.v.'s with values in a topological v.s. V. V should be subject to suitable conditions, maybe completeness so that the mean m might exist. For example V might be a Hilbert space. It seems that the L^2 LLN holds in this case.

Recall that the universal continuous map from E to a topol. vector space (good) should be the map

$$\begin{aligned} E &\longrightarrow \mathcal{M}(E) \\ x &\longmapsto \delta_x \end{aligned}$$

The LLN for this random variable says that the sequence of r.v.'s on Ω

$$\frac{S_n}{n} = \frac{1}{n}\sum_{j=1}^{n} \delta_{x_j} \quad : \quad E^n \longrightarrow \mathcal{M}(E)$$

converges in some sense to the original measure ν.

Notice that the convergence concept must depend on ν because the r.v.'s $\frac{S_n}{n}$ do not. This suggests that the measures on Ω belonging to different ν are mutually singular.

In any case it is suggestive to think that ~~the~~ random measure

$$\frac{1}{n}\left(\delta_{x_1}+\dots+\delta_{x_n}\right) \qquad (x_i)\in E^n$$

approximates ν as $n\to\infty$ when the x_i are varying independently with distribution ν.

Next I want to discuss the generalization to a Markov chain. Here we suppose given transition probabilities $p(x,x')$, $x,x'\in E$, such that $p(x,x')>0$ for all x,x' and that E is finite. Then there is a unique prob. measure ν [redacted] on E with $\sum_x \mu(x)p(x,x')=\mu(x')$, and we knows $\mu(x)>0$ for all x.

We know [redacted] that [redacted] we can put ~~consistent~~ measures $\nu^{(n)}$ on E^n:

$$\nu^{(n)}(x_1,\dots,x_n) = \mu(x_1)\,p(x_1,x_2)\cdots p(x_{n-1},x_n)$$

and that we then get a prob. measure P on Ω such that $\theta_*(P)=P$, where $\theta:\Omega\to\Omega$ is the (backward) shift.

We propose now to use the family $\nu^{(n)}$ on E^n instead of $\nu^{\otimes n}$. We ~~define~~ consider the random measure $\frac{1}{n}\sum_1^n \delta_{x_j}$ on E with $(x_1,\dots,x_n)$ varying subject ~~to~~ $\nu^{(n)}$ and we want the limiting distribution as $n\to\infty$. This means we look at the measure μ_n on $\mathcal{M}(E)$

which is the image of $\nu^{(n)}$ and we want the limit of μ_n. This limit is supposed to be δ_ν.

The only way I know how to even begin to see why $\mu_n \to \delta_\nu$ is to use the Laplace transform. Thus I take a cont. linear function on $\mathcal{M}(E)$, that is, an $f \in C_b(E)$ in general, (but in the case we are considering E is finite). Then we want the transform of μ_n:

$$\int_{\mathcal{M}(E)} e^{\int f \, d\mu} \mu_n(d\mu) = \int_{\mathcal{M}(E)} e^{\langle f, \cdot \rangle} d\mu_n$$

$$= \int_{E^n} e^{\frac{1}{n}(f(x_1) + \cdots + f(x_n))} \nu^{(n)}(dx_1, \ldots, dx_n)$$

$$= \sum_{x_1, \ldots, x_n} e^{\frac{1}{n}(f(x_1) + \cdots + f(x_n))} \mu(x_1)\, p(x_1, x_2) \cdots p(x_{n-1}, x_n)$$

and we want the behavior of this as $n \to \infty$. This should be a consequence of the eigenvalues of the matrix

$$e^{\frac{1}{n} f(x)} p(x, y) \quad \text{or} \quad p(x,y) e^{\frac{1}{n} f(y)}$$

We have

$$\int_{\mathcal{M}(E)} e^{\langle f, \cdot \rangle} d\mu_n = \langle \mu \,|\, e^{\frac{1}{n} f} P e^{\frac{1}{n} f} P \cdots P e^{\frac{1}{n} f} \,|\, \mathbb{1} \rangle$$

Now we will try to guess the answer. But actually we know all the measures

μ_n have the mean μ, so what we want to establish is

$$\underbrace{\lim_{n\to\infty} \langle \mu | e^{\frac{1}{n}f} (P e^{\frac{1}{n}f})^n | \mathbb{1} \rangle}_{\langle \mu | (P e^{\frac{1}{n}f})^{n+1} | \mathbb{1} \rangle} = e^{\int f \, d\mu}$$

$$\langle \mu | (P e^{\frac{1}{n}f})^{n+1} | \mathbb{1} \rangle = \mathrm{tr}\Big(\underbrace{|\mathbb{1}\rangle\langle\mu|}_{\substack{\text{projection on the} \\ \text{dominant eigenspace of } P.}} \cdot (P e^{\frac{1}{n}f})^{n+1} \Big)$$

If I put $P e^{\frac{1}{n}f} = P + \Delta P$, then I could try to understand $(P+\Delta P)^n$ by the expansion

$$P^n + \sum_{i+j=n-1} P^i \, \Delta P \, P^j + \sum_{i+j+k=n-2} P^i \, \Delta P \, P^j \, \Delta P \, P^k + \cdots$$

In thinking about this [redacted] it seems one wants to know <u>P</u> is a <u>contraction operator</u>: $\|P\| \leq 1$.

<u>Prop.</u> A stochastic matrix p_{ij} gives rise to a contraction operator on $L^p(\mu)$ for any invariant probability measure μ (i.e. $\sum_i \mu_i p_{ij} = \mu_j$). (Assume μ_i always >0).

<u>Proof</u>: Equip $E \times E$ with the prob. measure $\mu(x)\, p(x,y)$. Then the two projections

$$\begin{tikzcd} & E \times E \arrow[dl, "pr_1"'] \arrow[dr, "pr_2"] & \\ E & & E \end{tikzcd}$$

are measure preserving. The maps

$$L^p(E,\mu) \xhookrightarrow{pr_2^*} L^p(E\times E, \ldots\ldots) \xrightarrow{pr_{1*}} L^p(E,\mu)$$

are respectively an isometric embedding and a contraction. (In general, one knows conditional expectation is a contraction on L^p.). So $pr_{1*}\, pr_2^*$ is a contraction.

Finally notice that given f on E

$$(pr_2^* f)(x,y) = f(y)$$

$$(pr_{1*}\, pr_2^* f)(x) = \sum_y p(x,y) f(y) = (Pf)(x).$$

QED.

This proof shows that $\|Pf\| = \|f\|$ implies $pr_2^* f$ is in the image of pr_1^*, i.e. that $f(x)$ is constant. Here I am identifying pr_{1*} with orthogonal projection in the L^2 case.

Let's return to $Pe^{\frac{1}{n}f}$ and let's compute how the dominant eigenspace and eigenvalue change for this perturbation of P. Put $\varepsilon = \frac{1}{n}$

$$Pe^{\varepsilon f} = P + \varepsilon Pf + O(\varepsilon^2)$$

$$v = \mathbb{1} + \varepsilon v_1 + \cdots$$

$$\lambda = 1 + \varepsilon \lambda_1 + \cdots$$

$$(Pe^{\varepsilon f} - \lambda)\, v = 0$$

$$\big(P + \varepsilon Pf - (\mathbb{1} + \varepsilon\lambda_1)\big)(\mathbb{1} + \varepsilon v_1 + \cdots) = 0$$

$$(P - \mathbb{1})\, v_1 + (Pf - \lambda_1)\mathbb{1} = 0 .$$ Apply $\langle\mu|$

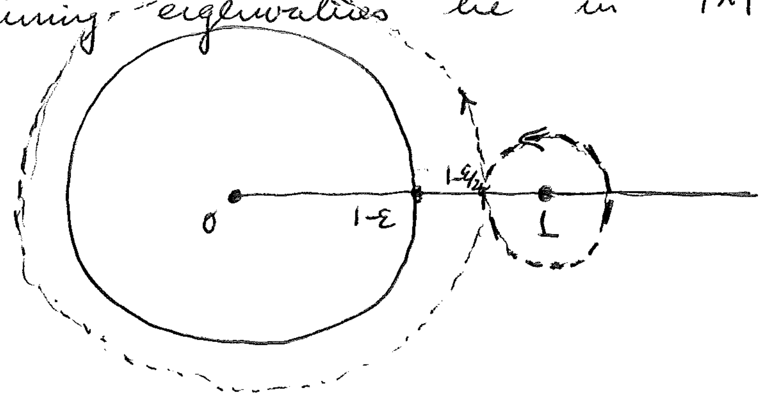
$$\lambda_1 = \langle \mu | Pf \rangle = \langle \mu | f \rangle$$

Thus $\lambda = 1 + \varepsilon\langle\mu|f\rangle = 1 + \frac{1}{n}\langle\mu|f\rangle$, so we expect

$$\langle \mu | (Pe^{\frac{1}{n}f})^{n+1} \mathbb{1} \rangle \sim \left(1 + \tfrac{1}{n}\langle\mu|f\rangle\right)^{n+1} \longrightarrow e^{\langle\mu|f\rangle}$$

Is there a way to make this rigorous?

The spectrum of P consists of the simple eigenvalues $\lambda = 1$ and the rest which lies in $|\lambda| \leq 1-\varepsilon$ for some $\varepsilon > 0$. For n large $P_n = Pe^{\frac{1}{n}f}$ has a simple eigenvalue λ_n inside $|\lambda - 1| = \varepsilon/2$ and the remaining eigenvalues lie in $|\lambda| \leq 1-\varepsilon/2$. We have by contour integration



$$P_n^{n+1} = \underbrace{\frac{1}{2\pi i}\oint_{|\lambda|=1-\varepsilon/2} \frac{1}{\lambda - P_n}\lambda^{n+1}\,d\lambda}_{\substack{\text{dominated by}\\ (1-\varepsilon/2)^{n+1}}} + \underbrace{\left(\frac{1}{2\pi i}\oint_{|\lambda-1|=\varepsilon/2} \frac{1}{\lambda - P_n}\right)}_{\substack{\text{projection on the}\\ \text{dominant eigenspace}\\ \text{call this } E_n}}\lambda_n^{n+1}$$

Thus

$$\mathrm{tr}\left(|1\rangle\langle\mu|\circ P_n^{n+1}\right) = \underbrace{O\left((1-\varepsilon/2)^{n+1}\right)}_{\downarrow\, 0} + \underbrace{\mathrm{tr}\left(|1\rangle\langle\mu|\circ E_n\right)}_{\downarrow\, 1}\underbrace{\lambda_n^{n+1}}_{\downarrow\, e^{\langle\mu|f\rangle}}$$

and so everything is clear.

Therefore at this point we have verified that $\mu_n \longrightarrow \delta_\mu$ in the sense of the Laplace transform. So what I want to do next is to discuss the large **deviations** result for this limit.

Let's recall that one of the consequences of the large deviation theory is to give the leading part of the Laplace asymptotics for an integral of the form

$$\int e^{n\Phi}\, d\mu_n$$

and that the result is roughly to the effect that

$$\int e^{n\Phi}\, d\mu_n \sim e^{n\left(\sup_x \Phi(x) - W(x)\right)}$$

where W is the "rate function". The simplest Φ one might take is a linear function.

(Review the Log. transform theory: Assume μ_n comes from $\frac{S_n}{n}$. Then

$$\int e^{nJx}\, \mu_n(dx) = Z(J)^n = e^{n \log Z(J)}$$

If this integral ↑ $\sim e^{n\left(\sup_x Jx - W(x)\right)}$, then

$$\sup_x Jx - W(x) = \log Z(J)$$

which is just the inverse transform of the defn of W as $\quad W(x) = \sup_J J \cdot x - \log Z(J)$.)

Now let's apply this in the Markov chain case. I take Φ on $\mathcal{M}(E)$ which is linear, i.e. $\Phi = \langle f, \cdot \rangle$ with f a function on E. Then

$$\int e^{n\langle f, \cdot\rangle}\, d\mu_n = \langle \mu | (Pe^f)^{n+1} | 1 \rangle$$

We want the asymptotics of this as $n \to \infty$. It would appear that

$$W(\alpha) = \sup_f \left\{ \int f\, d\alpha - \log \lambda_{max}(Pe^f) \right\}$$

Consider the independent case: $p(x,y) = \mu(y)$. Then $Pe^f = |1\rangle\langle \mu| e^f$ has one eigenvalue $\neq 0$

namely $\langle \mu | e^f | 1 \rangle = \int e^f \, d\mu$, so

$$W(\alpha) = \sup_f \left\{ \int f \, d\alpha - \log \left(\int e^f \, d\alpha \right) \right\}$$

as in the Sanov theorem.

~~[illegible]~~

It appears likely that

$$f \longmapsto \log \lambda_{max} (P e^f)$$

is a convex function of the vector f.

<u>May 25, 1985</u>

Basic map $\quad T\backslash U(n)/T \longrightarrow \left\{\begin{matrix}\text{doubly stochastic}\\ \text{matrices}\end{matrix}\right\}$

$$g \longmapsto (|g_{jk}|^2)$$

I would like to know whether this map is an isomorphism. The Birkhoff - van Neumann thm. says that the set of doubly stochastic matrices is the convex hull of the permutation matrices. One might hope to prove both assertions simultaneously.

Let $\mathcal{P}$ be the set of doubly-stochastic matrices $P = (p_{jk})$. This means $P \cdot \mathbb{1} = \mathbb{1}$ and $\mathbb{1}^t P = \mathbb{1}^t$ and each $p_{jk} \geq 0$.

Let V be the space of $n \times n$ matrices A with $A\mathbb{1} = 0$ and $\mathbb{1}^t A = 0$, i.e. both row and column sums are zero. V has dimension $(n-1)^2$. A natural basis of V consists of

$$(*) \qquad \begin{array}{c c} & j \\ i \rightarrow & \left(\begin{array}{ccc|c} & 1 & & -1 \\ \hline & -1 & & 1 \end{array}\right) \end{array}$$

for $1 \leq i, j \leq n-1$.

Fix $p^{(0)}$ = the doubly stochastic matrix with all entries $\frac{1}{n}$. Then $\mathcal{P}$ is a convex subset of $p^{(0)} + V$, so $\mathcal{P} - p^{(0)}$ is a convex subset of V. Clearly $\mathcal{P} - p^{(0)}$ consists of A in V such that $a_{ij} \geq -\frac{1}{n}$ for all i,j. If we use the basis $(*)$ of V, it is clear that [redacted] $\mathcal{P} - p^{(0)}$ contains an open nbd of 0 in V. Thus $\mathcal{P}$ is a convex set in $p^{(0)} + V$ with a non empty interior.

In fact the interior of $\mathcal{P}$ consists of all $p = (p_{jk})$ with $p_{jk} > 0$ for all j, k. To see this note that the linear function $p \mapsto p_{jk}$ is open because it is onto, so the interior of $\mathcal{P}$ must go into an open subset of $\mathbb{R}$, hence $p_{jk} > 0$. On the other hand if $p_{jk} > 0$ for all j, k, then any small addition of a point of V will not change this.

If an interior point of $\mathcal{P}$ comes from $g \in U(n)$, then clearly $g_{jk} \neq 0$ for all j, k. It would be nice to show that the map $T \backslash U(n)/T \longrightarrow \mathcal{P}$ is etale over the interior of $\mathcal{P}$.

The map sends g to $p = (p_{jk})$, $p_{jk} = |g_{jk}|^2$ or

$$
\begin{aligned}
p_{jk} &= \bar{g}_{jk}\, g_{jk} = g^*_{kj}\, g_{jk} = (g^{-1})_{kj} (g)_{jk} \\
&= \langle k | g^{-1} | j \rangle\langle j | g | k \rangle \\
&= \operatorname{tr}\left(g^{-1} E_j\, g\, E_k\right) \qquad\qquad E_j = |j\rangle\langle j| \\
&= \operatorname{tr}\left(E_j \cdot g E_k g^{-1}\right)
\end{aligned}
$$

Thus what the map does is to look at the inner products between the standard flag with projectors E_j and its transform under g.

Now we can compute the variation δp of p corresponding to a variation δg of g. As usual, let $\delta g = gX$ with X skew hermitian. Then

$$\delta p_{jk} = \operatorname{tr}\left(E_j\, g [X, E_k] g^{-1}\right).$$

(This is clearly in V. Note $[X, E_k]$ is hermitian so we have on the right the inner product of two hermitian matrices which is real.

What I want to show is that if $g_{jk} \neq 0$ for all j,k, then the map $X \mapsto \delta_P$ maps $\mathfrak{g}$ onto V. I believe that I understand that the dimensions of $T \backslash U^{(n)}/T$ and P are the same. Let's do it carefully. We've seen P has dim. $(n-1)^2$. Take a point $g \in U(n)$ with $g_{jk} \neq 0$ for all j,k. Calculate the isotropy group for the $T \times T$ action: $(t,t')g = t_j g_{jk} (t'_k)^{-1}$. Then $t_j g_{jk} (t'_k)^{-1} = g_{jk}$ for all j,k, so $t_j = t'_k$ for all j,k, so the stabilizer is just S^1. Thus near this g one sees $T \times T/S^1$ acts freely, and so the dimension of the quotient space is $\dim U(n) - \dim (T \times T/S^1) = n^2 - (2n-1) = (n-1)^2$.

Here seems to be a good idea: Let $\lambda =$ denote the diagonal matrix with entries $(\lambda_1, \dots, \lambda_n)$. Then we have a map

$$U(n)/T \longrightarrow \mathfrak{g}, \quad g \mapsto g\lambda g^{-1}$$

which is essentially the moment map for the $U(n)$ action on $U(n)/T$ and for a suitable symplectic structure depending on λ. Restrict to T action and we get

$$U(n)/T \longrightarrow \mathfrak{t}^*$$

$$g \longmapsto (\mu \mapsto \operatorname{tr}(\mu \cdot g\lambda g^{-1}))$$

The point is that we know a lot about this map. The image is the convex hull of the Weyl orbit. In particular we know the critical points of the function $g \mapsto \operatorname{tr}(\mu \cdot g\lambda g^{-1})$, at least for μ, λ

generic, occur at the points of $N/T = W$.

Counterexample that $T\backslash U(n)/T \twoheadrightarrow P$. Take $n=3$ and consider the problem of finding orthonormal vectors

$$(*) \qquad \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \begin{pmatrix} \alpha \\ \beta e^{i\theta} \\ \gamma e^{i\varphi} \end{pmatrix}$$

where $a, b, c, \alpha, \beta, \gamma > 0$. Better I suppose given a doubly stochastic matrix whose first two columns are are

$$\begin{pmatrix} a^2 & \alpha^2 \\ b^2 & \beta^2 \\ c^2 & \gamma^2 \end{pmatrix},$$

where we suppose $a, b, c, \alpha, \beta, \gamma$ are >0.

Then relative to the action of $T \times T$, one can assume any unitary matrix over this doubly stochastic matrix has the first two columns in the form $(*)$. The phases θ, φ have to be such that

$$a\alpha + b\beta e^{i\theta} + c\gamma e^{i\varphi} = 0$$

and this implies $a\alpha < b\beta + c\gamma$. But now we can get a contradiction by taking c and β to be very small. Thus the doubly stochastic matrix

$$\begin{pmatrix} a^2 & a^2 & 1-2a^2 \\ b^2 & \varepsilon^2 & a^2 \\ \varepsilon^2 & b^2 & a^2 \end{pmatrix} \qquad a^2 + b^2 + \varepsilon^2 = 1$$

where say b is fixed $> \frac{1}{\sqrt{2}}$ so that $a^2 < \frac{1}{\sqrt{2}}$ and $2a^2 < 1$, will not come from a unitary for $\varepsilon = 0$ and hence for small ε.

June 2, 1985

Classical gas theory review. The configurations of a gas with N-particles form the space M^N, and the probability distribution on the configuration space is of the form

$$\nu_N = \frac{e^{-\beta u_N(x_1,\dots,x_N)}\, dx_1 \cdots dx_N}{\int e^{-\beta u_N(x_1,\dots,x_N)}\, dx_1 \cdots dx_N}$$

We suppose u_N symmetric under permutations.

The sort of quantities whose average~~s~~ we are interested in are 1- and 2-particle functions:

$$\sum_{j=1}^{N} f(x_j)$$

$$\frac{1}{2}\sum_{i\neq j} f(x_i,x_j)$$

~~This~~ It seems that these averages can all be explained by the generating function

$$Z_N(f) = \int_{M^N} e^{\sum_1^N f(x_j)}\, \nu_N$$

Assuming this to be so, we see that we are interested in the prob. measure μ_N on $\mathcal{M}(M)$ obtained by pushing ν_N forward under the map

$$M^N \longrightarrow \mathcal{M}(M) \quad , \quad (x_1,\dots,x_N) \longmapsto \sum_1^N \delta_{x_j}$$

($Z_N(f)$ is just the Laplace transform of μ_N)

Now what [redacted] we [redacted] want to do is to let $N \longrightarrow \infty$, or better, say something about what happens when N is large. The typical ~~statistical~~ thing one might want is some way to describe μ_N asymptotically for large N. For example μ_N has a mean which is a density ρ, and a variance, etc.

There are [redacted] two things that one can do [redacted]. The standard thing is to rescale so that $\mu_N \in \mathcal{M}_1(M)$, i.e. use $x_1, \dots, x_N \longmapsto \frac{1}{N} \sum_1^N \delta_{x_j}$. This corresponds to taking $Z_N(f/N)$, and this has a limit [redacted] as $N \longrightarrow \infty$, which is the characteristic function of the measure δ_ρ.

The other thing one can do is to [redacted] consider

$$\lim_{N \to \infty} \frac{1}{N} \log Z_N(f)$$

What would be really nice would be for [redacted] μ_N to have an N-th root in the sense of convolution. [redacted] This means that the interacting gas behaves like a free gas with a strange [redacted] distribution for single particles.

A new ingredient which might enter is the grand ensemble idea.

June 6, 1985

Classical gas models provide examples where large deviation phenomena can be studied. For each n one has a pr. measure ν_n on M^n, where M is the space of configurations for one particle. This pr. meas. can be pushed forward under

$$(x_1, \dots, x_n) \longmapsto \frac{1}{n} \sum \delta_{x_j}$$

to obtain a pr. measure μ_n on the space $\mathcal{M}_1(M)$ of pr. measures on M. One can study the asymptotic behavior of μ_n as $n \to \infty$.

For example, suppose the particles are independent, that is, ν_n is $\nu^{\otimes n}$, where ν is the pr. distribution for one particle. Then the behavior of μ_n as $n \to \infty$ is described by Sanov's theorem.

The Curie-Weiss model for magnetism provides a simple example of an interacting gas. Here $M = \{\pm 1\}$, and the distribution ν_n on $M^n = \{ \vec{s} = (s_j)_{1 \le j \le n} \mid s_j = \pm 1 \}$ is the Boltzmann distribution with the energy

$$E(\vec{s}) = - \left\{ H \sum s_j + \frac{a}{2} \sum_{i \neq j} s_i s_j \right\}$$

where H is the "applied magnetic field".

When $M = \{\pm 1\}$ we can identify $\mathcal{M}_1(M)$ with $[-1, 1]$ by associating to a prob. measure $\nu = a\delta_{-1} + b\delta_1$, $(a, b \geq 0,\ a+b=1)$, the average of s:

$$\int s \, d\nu = b - a$$

Modulo this identification we see that μ_n

becomes the prob. measure on $[-1, 1]$ which is the push-forward of ν_n on M^n under the map:

$$\vec{s} = (s_1, \ldots, s_n) \longmapsto \frac{1}{n}(\delta_{s_1} + \cdots + \delta_{s_n}) \longmapsto \frac{1}{n}(s_1 + \cdots + s_n) = \bar{s}.$$

giving the average spin of the configuration $\vec{s}$. Thus μ_n is the probability distribution of the average spin.

Let's now turn to the calculation of this distribution. Note that

$$\sum_{i \neq j} s_i s_j = \left(\sum s_i\right)^2 - \sum \underbrace{s_i^2}_{1}$$

$$= n^2 \bar{s}^2 - n$$

In fact without changing the Boltzmann measure ν_n we can change the energy by an additive constant and suppose

$$E(\vec{s}) = -\left\{ H n \bar{s} + \frac{a}{2} n^2 \bar{s}^2 \right\}.$$

If $\beta = 1$, then

$$\nu_n(\vec{s}) = e^{Hn\bar{s} + \frac{a}{2} n^2 \bar{s}^2} / \text{normal.}$$

Hence we see that

$$\mu_n = e^{Hn\bar{s} + \frac{a}{2} n^2 \bar{s}^2} \mu_n^0 / \text{norm.}$$

where μ_n^0 is the probability measure on $[-1, 1]$ which is the distribution of $\bar{s} = \frac{1}{n} \sum s_j$ where the 2^n possible $\vec{s}$ are equally likely. Thus μ_n^0 is a binomial distribution

Let's work out the asymptotics of the binomial distribution using Stirling's formula. The probability of k heads is

$$\binom{n}{k} 2^{-n} = \frac{n!}{k!\,(n-k)!}\, 2^{-n}$$

Now

$$\log \binom{n}{k} \doteq \overset{k+(n-k)}{n} \log n - \cancel{n} + \log\sqrt{2\pi n}$$
$$- k\log k + \cancel{k} - \log\sqrt{2\pi k}$$
$$- (n-k)\log(n-k) + \cancel{n-k} - \log\sqrt{2\pi(n-k)}$$

In terms of $x = \frac{k}{n}$ we have

$$\log\binom{n}{k} = -k\log\left(\frac{k}{n}\right) - (n-k)\log\left(\frac{n-k}{n}\right) - \log\sqrt{2\pi} + \log\frac{\sqrt{n}}{\sqrt{nx}\,\sqrt{n(1-x)}}$$

or

$$\log\binom{n}{k}2^{-n} = -n\underbrace{\left\{x\log(x) + (1-x)\log(1-x) + \log(2)\right\}}_{W^\circ(x)} + O(\log n)$$

Notice that this is also what one obtains from Cramér's thm:

$$\mu = \tfrac{1}{2}(\delta_0 + \delta_1)$$

$$Z(J) = \tfrac{1}{2}(1 + e^J)$$

$$x = \frac{d}{dJ}\log Z(J) = \frac{e^J}{1+e^J} = 1 - \frac{1}{1+e^J}$$

$$1 + e^J = \frac{1}{1-x} \qquad e^J = \frac{1}{1-x} - 1 = \frac{x}{1-x}$$

$$W^\circ(x) = xJ - \log Z(J) \qquad \text{where } J = \log\left(\frac{x}{1-x}\right)$$

$$= x\log\left(\frac{x}{1-x}\right) - \log\frac{1}{2(1-x)}$$

$$= x\log x + (1-x)\log(1-x) + \log 2$$

Now I put this on the interval $[-1,1]$ by letting $x = \frac{1+\bar{s}}{2}$ or $\bar{s} = 2x-1$.

$$W^{\circ}(\bar{s}) = \left(\frac{1+\bar{s}}{2}\right)\log(1+\bar{s}) + \left(\frac{1-\bar{s}}{2}\right)\log(1-\bar{s})$$

Thus we conclude

$$\mu_n \sim e^{-n\{W^{\circ}(\bar{s})\} + nH\bar{s} + n^2\frac{a}{2}\bar{s}^2} \Big/ \text{norm}$$

In the Curie-Weiss model one takes $a = \frac{K}{n}$ as $n \to \infty$, so that the field felt at one site due to the other sites is essentially independent of a.

Thus

$$\mu_n \sim e^{-n\,W(\bar{s})}$$

where

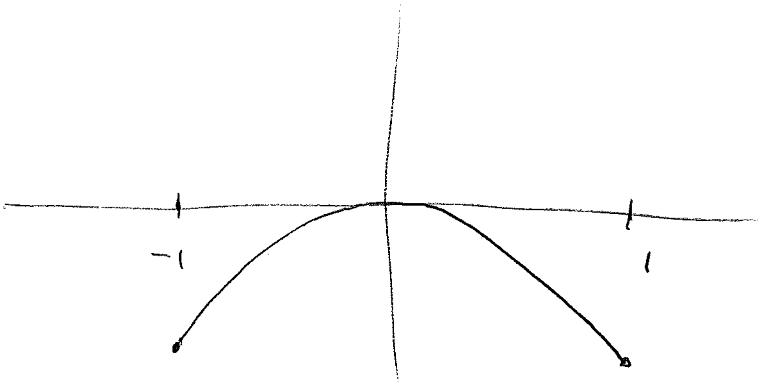
$$W(\bar{s}) = \underbrace{\frac{1+\bar{s}}{2}\log(1+\bar{s}) + \frac{1-\bar{s}}{2}\log(1-\bar{s})} - H\bar{s} - \frac{K}{2}\bar{s}^2$$

W

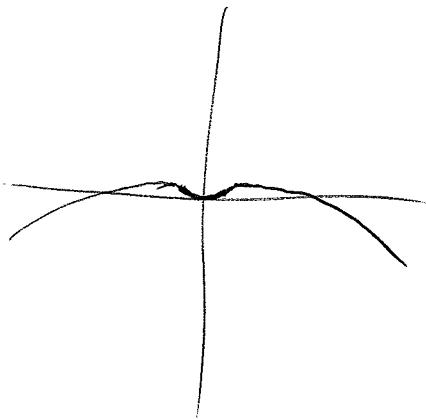
$$\frac{1+\bar{s}}{2}\left(\bar{s} - \frac{\bar{s}^2}{2} + \dots\right) + \frac{1-\bar{s}}{2}\left(-\bar{s} - \frac{\bar{s}^2}{2} + \dots\right) = \frac{\bar{s}^2}{2} + O(\bar{s}^3)$$

Picture of $-W^{\circ}(\bar{s})$:

Take $H=0$



$-W^{\circ}(\bar{s}) + \frac{K}{2}\bar{s}^2$



so that if $K > 1$, there is broken symmetry.

So what are the implications of this example? Let's suppose $H = 0$ and that $K > 1$ so that the function $W(\bar{s})$ has two local minima (Check.

$$W(\bar{s}) = \frac{1+\bar{s}}{2} \log(1+\bar{s}) + \frac{1-\bar{s}}{2} \log(1-\bar{s}) - \frac{K}{2} \bar{s}^2$$

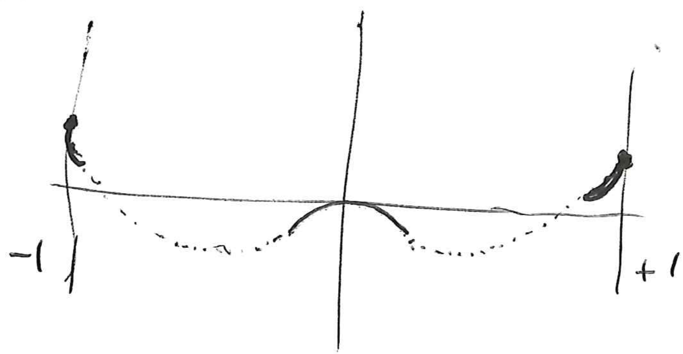
$$= \left(\frac{1-K}{2}\right)\bar{s}^2 + O(\bar{s}^4)$$

$$W(\pm 1) = \log 2 - \frac{K}{2}$$

If $K = 1$, then

$$e^{\log 2 - \frac{1}{2}} = 2/\sqrt{e} > 1$$

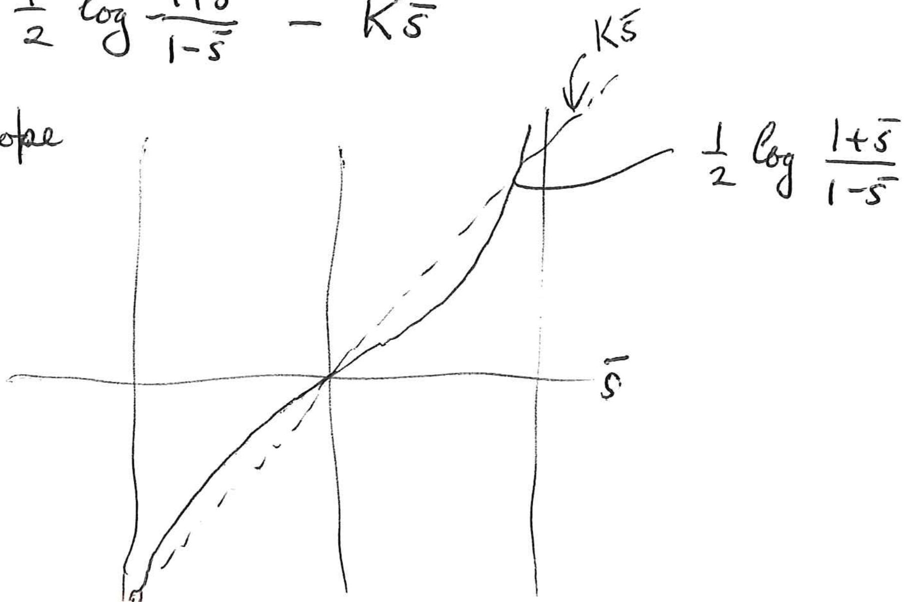
so $\log 2 - \frac{1}{2} > 0$. Thus there is a range of $K > 1$ such that $W(\pm 1) > 0$, and the graph of W has:



$$\frac{dW}{d\bar{s}} = \frac{1}{2}\log(1+\bar{s}) + \cancel{\frac{1}{2}} - \frac{1}{2}\log(1-\bar{s}) - \cancel{\frac{1}{2}} - K\bar{s}$$

$$= \frac{1}{2}\log\frac{1+\bar{s}}{1-\bar{s}} - K\bar{s}$$

Picture of slope



So it's pretty clear that W will have two non-degenerate minimum points for any $K > 1$.)

This should tell us what μ_n approaches as $n \to \infty$, namely, the average of the 2 δ-functions at these minimum points. This result corresponds to the law of large numbers.

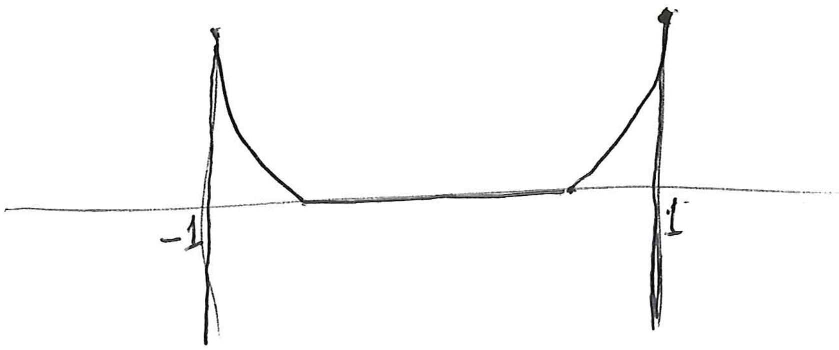
The asymptotic formula for μ_n is the large deviation result in this example. Since the function $W(x)$ is not convex we cannot recover it by a Legendre transform. Thus if we were to consider

$$\lim \frac{1}{n} \log \int e^{nJ\bar{s}} \mu_n = F(J)$$

we expect this to be

$$\sup_{\bar{s}} J\bar{s} - W(\bar{s}) \qquad (W \text{ normalized})$$

which is the Fenchel transform of W. This should transform back to give a convex version of W:



June 7, 1985

I am looking at the Curie-Weiss model in relation with large deviation ideas. In this model ~~the configuration space of one particle is $M = \{\pm 1\}$,~~ there are n sites each of which can be in the states ± 1, so the state space is M^n, where $M = \{\pm 1\}$. The Boltzmann principle determines a probability measure ν_n on M^n, which we can push forward under the map $\vec{s} = (s_1, \ldots, s_n) \longmapsto \frac{1}{n}\sum \delta_{s_j}$ to obtain a distribution μ_n on $\mathcal{M}_1(M)$. We identify $\mathcal{M}_1(M)$ with $[-1,1]$, whence μ_n becomes the pr. meas. on $[-1,1]$ which is the distribution of the random variable $\bar{s} = \frac{1}{n}\sum s_j$.

The energy of the configuration $\vec{s}$ is

$$-E(\vec{s}) = \frac{a}{2}\sum_{i \neq j} s_i s_j = \frac{a}{2}(\sigma^2 - n), \quad \sigma = \sum s_j$$

Taking $\beta = 1$, the different configurations are weighted proportionally to $e^{-E(\vec{s})} = e^{\frac{a}{2}\sigma^2} \cdot \text{const}$ by the Boltzmann principles. So if we put

$$Z_n(J) = \sum_{M^n} e^{J\bar{s} + \frac{a}{2}\sigma^2}$$

then the characteristic function of μ_n is

$$\int_{[-1,1]} e^{J\bar{s}} \mu_n(d\bar{s}) = \int_{M^n} e^{J\bar{s}} \nu_n(\vec{s}) = \frac{Z_n(J)}{Z_n(0)}$$

Next we have

$$e^{\frac{a}{2}\sigma^2} = \int_{-\infty}^{\infty} e^{y\sigma - \frac{y^2}{2a}} \frac{dy}{\sqrt{2\pi a}}$$

so

$$Z_n(J) = \sum_{M^n} \int e^{(\frac{1}{n}J + y)\sigma} \, e^{-\frac{y^2}{2a}} \frac{dy}{\sqrt{2\pi a}}$$

$$= \int \left(e^{\frac{1}{n}J + y} + e^{-(\frac{1}{n}J + y)} \right)^n e^{-\frac{y^2}{2a}} \frac{dy}{\sqrt{2\pi a}}$$

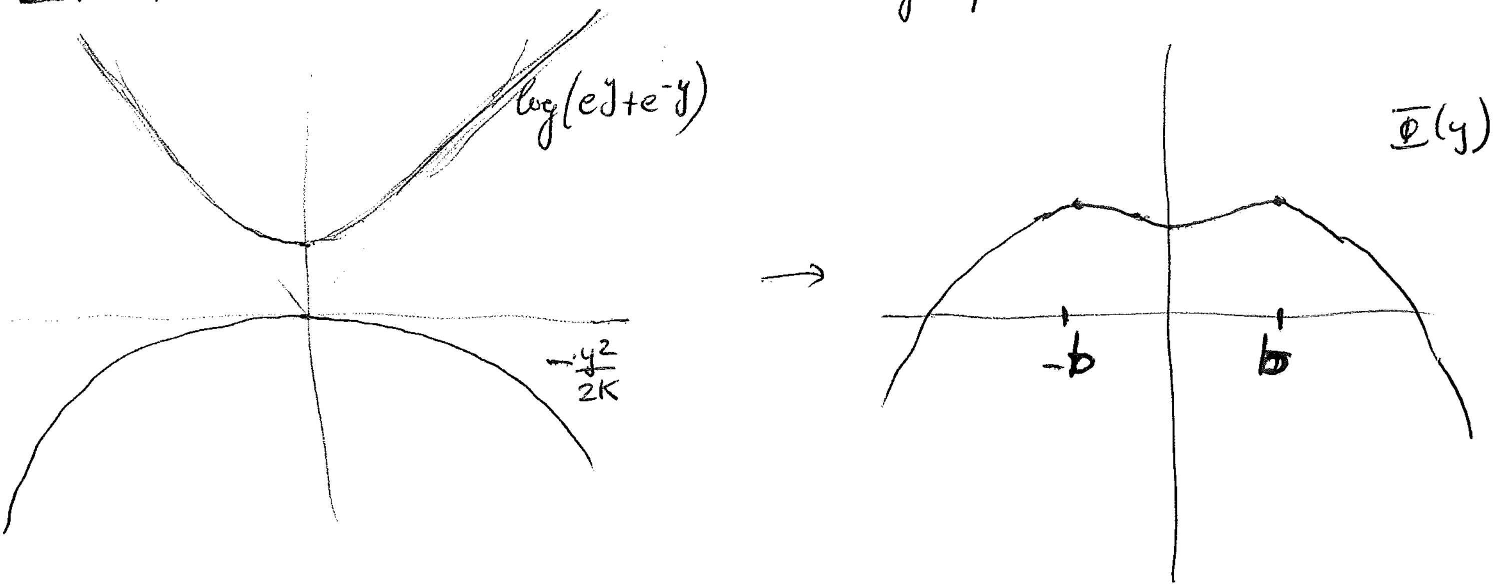
$$= \int \left(e^{y} + e^{-y} \right)^n e^{-\frac{(y - \frac{J}{n})^2}{2a}} \frac{dy}{\sqrt{2\pi a}}$$

$$= \int e^{n\left\{ \log(e^y + e^{-y}) - \frac{1}{2K}(y - \frac{J}{n})^2 \right\}} \frac{dy}{\sqrt{2\pi K/n}}$$

where we put $a = \frac{K}{n}$, K constant. The exponent for $J = 0$ is

$$\Phi(y) = \log(e^y + e^{-y}) - \frac{y^2}{2K}.$$

For $K > 1$ this has the graph



so

$$\frac{Z_n(J)}{Z_n(0)} = \frac{\int e^{n\Phi(y) + \frac{y}{K}J + O(\frac{1}{n})} \frac{dy}{\sqrt{2\pi K/n}}}{\int e^{n\Phi(y)} \frac{dy}{\sqrt{2\pi K/n}}} \longrightarrow \frac{e^{\frac{b}{K}J} + e^{-\frac{b}{K}J}}{2}$$

by the Laplace method. Thus we conclude (rigorously it seems) that

$$\mu_n \longrightarrow \tfrac{1}{2}\left(\delta_\tau + \delta_{-\tau}\right)$$

This is the "large of large numbers" result which [redacted] is usually stated at the beginning of the L.D. theory.

The [redacted] point of the L.D. theory is to describe the asymptotics of μ_n. I think the best sort of description would be a result asserting

$$(\star) \quad [redacted] \quad \int e^{n\Phi}\mu_n \sim e^{n\sup(\Phi - W)}$$

or more precisely

$$\lim_{n\to\infty} \frac{1}{n}\log\left(\int e^{n\Phi}\mu_n\right) = \sup(\Phi - W)$$

Such a result $(\star)$ gives a meaning to $\mu_n \sim e^{-nW}$.

The most useful Φ are the linear functions, so we are led in trying to establish $(\star)$ to studying

$$\lim \frac{1}{n}\log Z_n(nJ)$$

i.e.

$$Z_n(nJ)^{1/n}$$

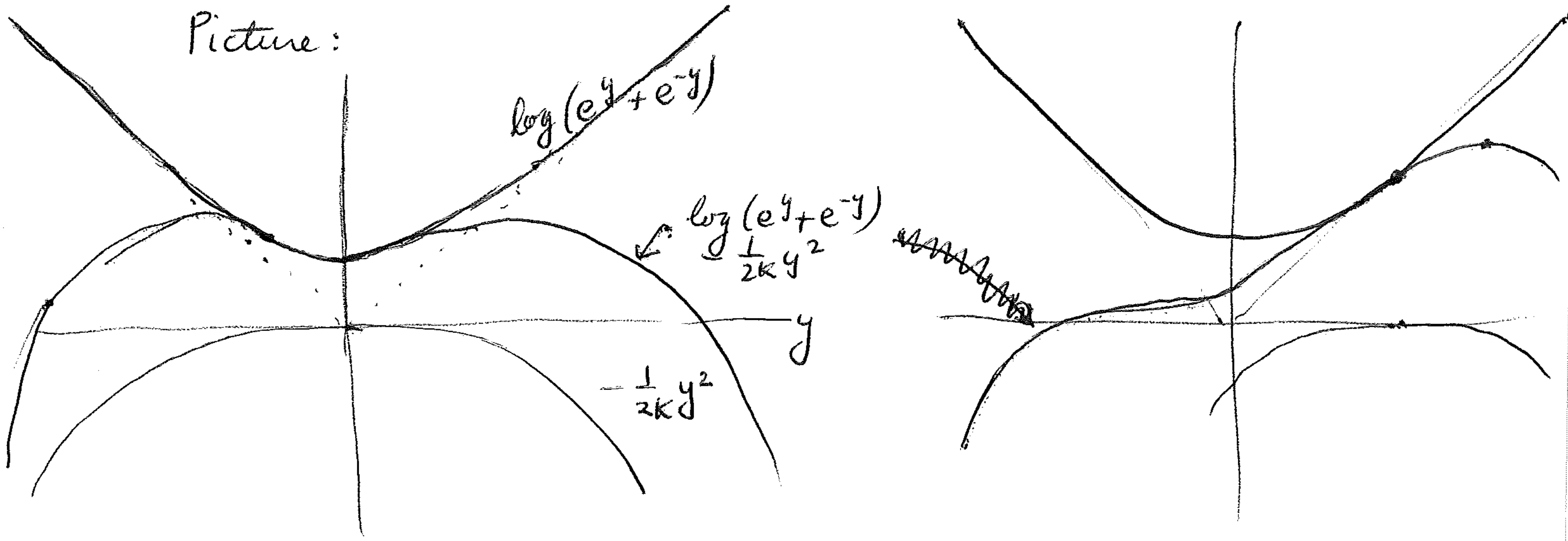
instead of $\lim Z_n(J)$. We can calculate this precisely from the formula for $Z_n(J)$. Thus

$$Z_n(nJ) = \int e^{n\left\{\log(e^y + e^{-y}) - \frac{1}{2K}(y-J)^2\right\}} \, dy/\sqrt{2\pi K/n}$$

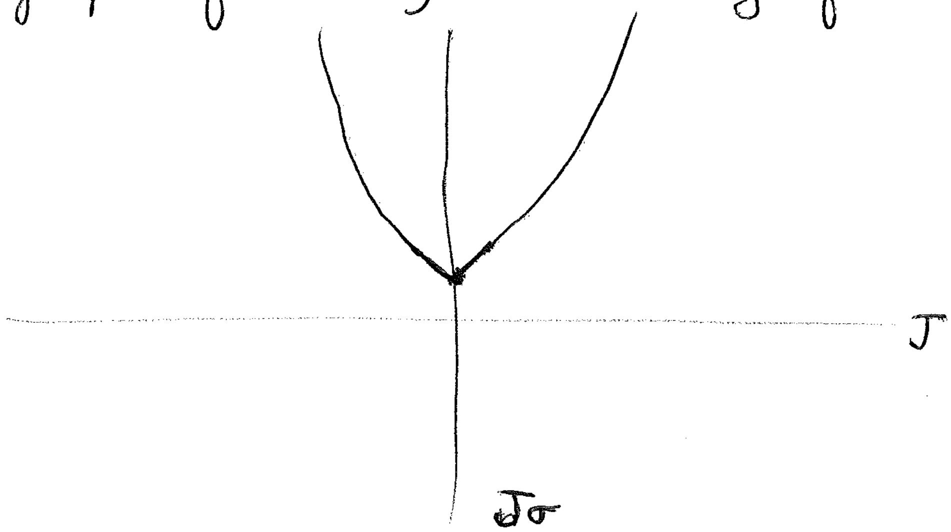
and so

$$F(J) = \lim_{n} \frac{1}{n} \log Z_n(nJ) = \max \left\{ \log(e^y + e^{-y}) - \frac{1}{2K}(y-J)^2 \right\}$$

Picture:



The graph of $F(J)$ is clearly of the form:

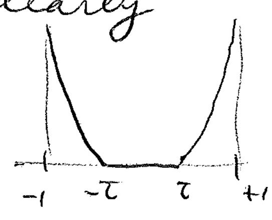


Also $$e^{nF(J)} \sim \int e^{\overset{J\sigma}{nJ\bar{s}} + \frac{a}{2}\sigma^2}$$ so that

$$\frac{d}{dJ}F(J) = \langle \bar{s} \rangle.$$

This agrees with the idea that $\langle \bar{s} \rangle$ should jump from $-\tau$ to $+\tau$ as J crosses 0.

So the effective ~~potential~~ $W(x)$ clearly has the shape illustrated on page 30



Summary:

I first describe some of the ideas in the theory of "large deviations."

We suppose given a sequence μ_n of probability measures on a vector space V. We would like to establish an asymptotic formula of the form

$$(*) \qquad \mu_n \sim e^{-nW}$$

where W is a function on V. ($(*)$ is vague; more precise forms would be

$$\lim_{n\to\infty} \frac{1}{n} \int e^{n\Phi} \mu_n = \sup_V \Phi - W \qquad \text{suitable fns. } \Phi \text{ on } V$$

$$\lim_{n\to\infty} \frac{1}{n} \log \mu_n(G) = \sup_G (-W) \qquad \text{suitable } G \subset V.)$$

Assuming W is l.s.c. and convex, we can find W as follows. ~~Taking Φ to be a linear function $\Phi(x) = J\cdot x$, $J \in V^*$ we have~~ Define $F(J)$, $J \in V^*$ by

$$(**) \qquad F(J) = \lim_{n\to\infty} \frac{1}{n} \log\left(\int e^{n J\cdot x} \mu_n\right)$$

assuming this limit exists. Assuming $(*)$ one has

$$F(J) = \sup_{x\in V} J\cdot x - W(x)$$

so F is the Fenchel transform of W. By Fenchel theory one has, assuming W l.s.c. + convex

$$(***) \qquad W(x) = \sup_J J\cdot x - F(J)$$

In any case $(***)$ is a candidate for W.

Example: Starting from a prob. meas. μ on V let μ_n be the image of $\mu^{\otimes n}$ on V^n under the map $(x_1, \ldots, x_n) \longmapsto \frac{1}{n}\sum x_j$. Then

$$\int e^{nJ.x}\mu_n = Z(J)^n \quad , \quad Z(J) = \int e^{J.x}\mu$$

so $F(J) = \log Z(J)$ in this case and $W(x)$ is the Fenchel transform of $\log Z(J)$. The fact (*) holds is the generalized Cramèr theorem

Classical gas models furnish situations where one can ask "large deviation" questions. In such a model there is a state space M for a single particle, and M^n is the state space for a gas of n particles. On M^n is a prob. measure ν_n given by the Boltzmann prinsiple from the potential energy function. We have the map

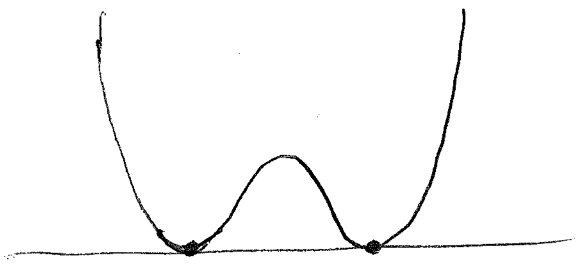
$$M^n \longrightarrow \mathcal{M}(M) \quad , \quad (x_1, \ldots, x_n) \longmapsto \frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n})$$

where $\mathcal{M}(M)$ is the space of (signed) measures on M, (or the space of distributions). Put $V = \mathcal{M}(M)$ and let μ_n be the image of ν_n under the above map. (Instead of $\mathcal{M}(M)$ ~~it~~ might be more natural to take V to be a vector space with a map $M \longrightarrow V$.)

The simplest gas models have state space M^n, where $M = \{\pm 1\}$. In this case the space of pr. measures on M can be identified with $[-1,1]$ via the map $\nu \longmapsto \int s\,\nu = \nu\{+1\} - \nu\{-1\}$. Then $\mu_n =$ pr. meas. on $[-1,1]$ giving the distribution of the average spin $\bar{s} = \frac{1}{n}\sum s_j$.

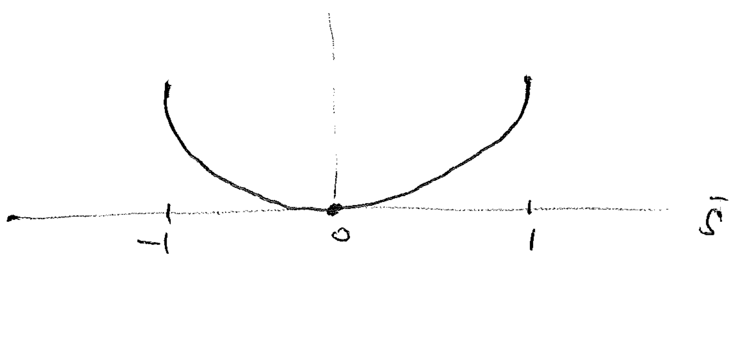
We have three examples of such models. The first is the free case where the particles are independent and it leads to the binomial distribution. The second is the Markov-chain model studied in Stroock's book. (I think it arises from any strictly positive 2×2 matrix.) The third is the Curie–Weiss model.

The Curie–Weiss model gives an example where the function W is not convex. Thus in this example one has the asymptotic behavior $\mu_n \sim e^{-nW}$, however the function W has two minimum points

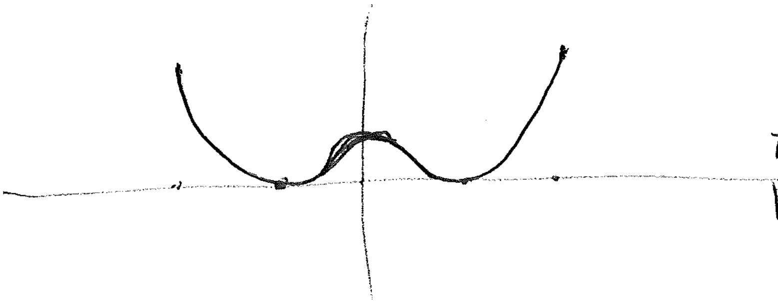


Consequently one can't expect to obtain W from the formulas (**) and (***). Formulas: For the free model, one gets the binomial distribution which has

$$\mu_n^\circ \sim e^{-nW^\circ} \qquad W^\circ(\bar{s}) = \frac{1+\bar{s}}{2}\log(1+\bar{s}) + \frac{1-\bar{s}}{2}\log(1-\bar{s})$$



Then $\mu_n \sim e^{-nW}$, $W = \frac{1+\bar{s}}{2}\log(1+\bar{s}) + \frac{1-\bar{s}}{2}\log(1-\bar{s}) - \frac{K}{2}\bar{s}^2 + C$



C such that the min. value of W is zero

June 8, 1985:

I ought to be able to prove Stroock's large deviation result for finite Markov chains. ~~[illegible]~~ To simplify I suppose there are two states labeled ± 1. One is given a strictly positive matrix $p(s,s')$ and defines a measure on M^n, $M = \{\pm 1\}$, by assigning the mass

$$(1) \qquad p(s_1,s_2)\,p(s_2,s_3)\cdots p(s_n,s_1)$$

to $\vec{s} = (s_1, \ldots, s_n)$. (Think in terms of a cyclic Ising model.) This can be normalized to give a prob. measure ν_n on M^n which we can then push forward under $\vec{s} \mapsto \bar{s}$ to get μ_n on $[-1,1]$. The Laplace transform of μ_n is

$$\int e^{\xi \bar{s}}\,d\mu_n = \frac{\sum p(s_1,s_2)\,e^{\frac{1}{n}\xi s_2}\,p(s_2,s_3)\cdots p(s_n,s_1)\,e^{\frac{1}{n}\xi s_1}}{\sum p(s_1,s_2)\cdots\cdots p(s_n,s_1)}$$

$$= \frac{\operatorname{tr}\left(pe^{\frac{1}{n}\xi s}\right)^n}{\operatorname{tr}(p^n)}$$

However another thing we can do, which is suggested by the magnet model, is to modify the measure on M^n given by (1) to

$$(2) \qquad p(s_1,s_2)\,e^{Js_2}\cdots p(s_n,s_1)\,e^{Js_1}$$

whence one obtains a modified prob. measure $\mu_{n,J}$ with L.T.

$$\int e^{\xi \bar{s}}\,d\mu_{n,J} = \frac{\operatorname{tr}\left(pe^{(J+\frac{1}{n})s}\right)^n}{\operatorname{tr}\left(pe^{Js}\right)^n}$$

The critical observation is that the measure (2) is $e^{nJ\bar{s}}$ times the measure (1).
Hence

$$\mu_{n,J} = e^{nJ\bar{s}}\mu_n \Big/ \int e^{nJ\bar{s}}\mu_n .$$

As a check:

$$\int e^{\xi \bar{s}}\mu_{n,J} \overset{?}{=} \int e^{(\xi+nJ)\bar{s}}\mu_n \Big/ \int e^{nJ\bar{s}}\mu_n$$

$$= \frac{\operatorname{tr}\left(p e^{\frac{1}{n}(\xi+nJ)s}\right)^n}{\operatorname{tr}(p^n)} \Big/ \frac{\operatorname{tr}\left(p e^{\frac{1}{n}(nJ)s}\right)^n}{\operatorname{tr}(p^n)}$$

The point here is that

$$\mu_{n,J} = e^{nJ\bar{s}}\mu_n / \text{norm}$$

is of the same sort of gadget as μ_n; it arises from pe^{Js} instead of p. Consequently we will know a L.N. result for $\mu_{n,J}$ for any J.

Intuitively the L.D. result for μ_n should be roughly equivalent to the L.N. result for $\mu_{n,J}$ and any J. This is what we learned from the proof of Cramèr's thm.

The first remark is to see that in the Cramèr case we do have

$$\mu_{n,J} = e^{nJx}\mu_n / \text{norm}.$$

where μ_n is the distribution of $\bar{x} = \frac{1}{n}\sum x_j$, and the x_j are independent r.v.'s governed by

$$\mu_J = e^{Jx}\mu/\text{norm.}$$

We can do this by generating functions:

$$\int e^{\xi x} \underbrace{\mu_{n,J}}_{(\mu_J)_n} = Z_J(\xi/n)^n$$

$$= \left(\int e^{(\xi/n)x} e^{Jx}\mu/Z(J)\right)^n$$

$$= \left(Z(J+\xi/n)/Z(J)\right)^n$$

$$\frac{\int e^{\xi x} e^{nJx}\mu_n}{\int e^{nJx}\mu_n} = \frac{Z\left(\frac{1}{n}(\xi+nJ)\right)^n}{Z\left(\frac{1}{n}nJ\right)^n}$$

So it works.

Thus $$\mu_{n,J} = \frac{e^{nJx}\mu_n}{Z(J)^n} \longrightarrow \delta_{x_J}$$

as $n \to \infty$ which means intuitively that

$$\mu_n \text{ at } x_J \sim e^{-n\underbrace{(Jx_J - \log Z(J))}_{W(x_J)}}$$

which is the desired L.D. result.

I would like to try to use the same ideas and push thru the L.D. result in the Markov case. Let's set things up by analogy using a common notation.

The basic object will be the generating

function

$$Z_n(J) = \mathrm{Tr}\,(p e^{Js})^n$$

This is the partition function in the physical magnetism setup, and is not the L.T. of μ_n. J is the external field. In the free case

$$Z_n(J) = Z(J)^n$$

We have the L.T. of the measure $\mu_{n,J}$:

$$\int e^{\xi x}\, d\mu_{n,J} = \frac{Z_n(J+\frac{\xi}{n})}{Z_n(J)}$$

From the Frob. theorem we know

$$\lim \frac{1}{n} \log Z_n(J) = \lambda_{max}\left(p e^{Js}\right)$$

Call this limit $F(J)$. Assume that one has an asymptotic expansion

$$\log Z_n(J) = nF(J) + G(J) + O(\tfrac{1}{n})$$

Then $$\log Z_n(J+\tfrac{\xi}{n}) = nF(J+\tfrac{\xi}{n}) + G(J+\tfrac{\xi}{n}) + O(\tfrac{1}{n})$$

$$= nF(J) + \xi F'(J) + G(J) + O(\tfrac{1}{n})$$

and so $$\log \frac{Z_n(J+\frac{\xi}{n})}{Z_n(J)} = \xi F'(J) + O(\tfrac{1}{n})$$

so $$\int e^{\xi x}\, d\mu_{n,J} \longrightarrow e^{\xi F'(J)}$$

June 9, 1985

Question: What is the infinitesimal version of a positive matrix and the corresponding Frobenius theorem?

We want $A = (a_{ij})$ to be such that e^{tA} is positive for $t \geq 0$. Since

$$A = \lim_{t \downarrow 0} \frac{e^{tA} - I}{t}$$

a necessary condition is that $a_{ij} \geq 0$ for $i \neq j$. Conversely if this holds then

$$1 + \frac{t}{n} A \quad \text{is positive for small } t \geq 0$$

and so

$$e^{tA} = \lim_{n \to \infty} \left(1 + \frac{t}{n} A\right)^n$$

is also positive.

Next we want the analogue of a strictly positive or primitive positive matrix. The idea is to form a directed graph out of the indices with an arrow from j to i when $a_{ij} > 0$. A natural condition is that one can go from j to i via a chain of such arrows.

The analogue of a stochastic matrix is an $A = (a_{ij})$ with $a_{ij} \geq 0$ for $i \neq j$ and

$$\sum_j a_{ij} = 0$$

Clearly e^{tA} is stochastic for all $t \geq 0 \iff$ this holds.

Suppose that A is a positive off-diagonal with zero row sums, let λ be an eigenvector and $v = (v_i)$ an eigenvector belonging to λ.

Then
$$\lambda v_i = \sum_j a_{ij} v_j$$
$$(\lambda - a_{ii}) v_i = \sum_{j \neq i} a_{ij} v_j$$

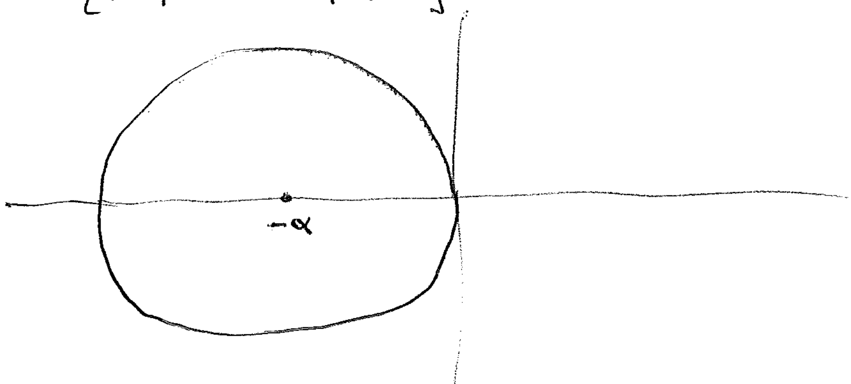
$$|\lambda - a_{ii}| |v_i| = \left| \sum_{j \neq i} a_{ij} v_j \right| \leq \sum_{j \neq i} a_{ij} |v_j|$$

Let $M = \max_j |v_j|$ and choose i such that $|v_i| = M$.

Then
$$|\lambda - a_{ii}| M \leq \sum_{j \neq i} a_{ij} |v_j| \leq \left(\sum_{j \neq i} a_{ij} \right) M$$

so we have $|\lambda + \alpha| \leq \alpha$ where $\alpha = -a_{ii} = \sum_{j \neq i} a_{ij}$.

But $\{\lambda \mid |\lambda + \alpha| \leq \alpha\}$ is the circle



so we see that $Re(\lambda) \leq 0$ with equality iff $\lambda = 0$.

Next let's check the uniqueness of the zero eigenspace assuming that one can get from any index i to another j via a chain of arrows in the directed graph. An eigenvector for $\lambda = 0$ satisfies
$$\left(\sum_{j \neq i} a_{ij} \right) v_i = \sum_{j \neq i} a_{ij} v_j .$$

As this system of equations is real, we can concentrate on v being real. Let $M = \max |v_j|$ and let i be such that $|v_i| = M$. Multiplying by ± 1 we can suppose $v_i = M$. Then from

$$\Big(\sum_{j=i} a_{ij}\Big) \overset{M}{\overset{\shortparallel}{v_i}} = \Big|\sum_{j\neq i} a_{ij} v_j\Big| \leq \sum_{j \neq i} a_{ij} \cdot M$$

we deduce that $v_j = M$ if $a_{ij} \neq 0$. Similarly $v_k = M$ if there are j, k with $a_{ij} \neq 0$, $a_{jk} \neq 0$. By the assumption we can get from i to any other index in this way we see the uniqueness of the zero eigenvector.

Domination argument seems to show that when the states are connected by these chains, then e^{tA} is strictly positive for $t > 0$. It is clear that A dominates a small constant times a cyclic permutation on a subset of the states. But maybe simpler is to take a chain $i \to k \to l \to \cdots \to j$ whence if you label this $1 \to 2 \to 3 \to \cdots \to k$, you see that A dominates

$$\varepsilon \begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & 0 & 1 & & \\ & & & 0 & & \\ & & & & 0 & \\ & & & & & 0 \end{pmatrix}$$

so e^{tA} dominates (say $\varepsilon = 1$).

$$\begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdot & \frac{t^k}{k!} \\ & 1 & t & \ddots & \vdots \\ & & 1 & t & \ddots \\ & & & \ddots & t \\ & & & & 1 \end{pmatrix}$$

showing that $(e^{tA})_{ij} > 0$.

June 11, 1985

Interpretation of effective potential (after Coleman)

Consider the quantum mechanics of a particle moving on the line. This is a 0-space dimensional field theory in which the field φ is the position of the particle. Let H be the Hamiltonian operator and let us consider the modified Hamiltonian

$$H_J = H - J\hat{\varphi}$$

where $\hat{\varphi}$ is the operator of multiplication by φ. (Actually in the following we will be ~~treating~~ $H, \hat{\varphi}$ as an arb. pair of hermitian matrices.)

Let $E(H_J)$ be the ground energy of H_J. First order perturbation theory gives

$$\frac{\partial}{\partial J}\big(-E(H_J)\big) = \langle a|\hat{\varphi}|a\rangle$$

if $|a\rangle$ is the ground state of H_J normalized so that $\langle a|a\rangle = 1$. One can interpret $\langle a|\hat{\varphi}|a\rangle$ as the average value of φ in the ground state. Let's denote this φ_J so

$$\varphi_J = \frac{\partial}{\partial J}\big(-E(H_J)\big)$$

The effective potential is the Legendre transform of $J \mapsto -E(H_J)$:

$$W(\varphi) = J\varphi - (-E(H_J)) \qquad \text{if } \varphi = \varphi_J$$

Thus we have

$$\frac{\partial}{\partial \varphi}W(\varphi) = J \qquad \text{where } \varphi_J = \varphi$$

Physically, we apply a force J to the particle; the new potential is $W(\varphi) - J\varphi$ and the particle is to be found at the minimum.

To summarize, the effective potential is the Legendre transform of $J \longmapsto -E(H_J)$.

Coleman's interpretation of the eff. potential says that $W(\varphi)$ is the minimum of the energy $\langle a|H|a\rangle$ among states $|a\rangle$ satisfying the constraint

$$\langle a|\hat{\varphi}|a\rangle = \varphi$$

To see this we wish to minimize $\langle a|H|a\rangle$ subject to $\langle a|a\rangle = 1$, and $\langle a|\hat{\varphi}|a\rangle = \varphi$, and we use Lagrange multipliers

$$\delta\left\{\langle a|H|a\rangle - E\left(\langle a|a\rangle - 1\right) - J\left(\langle a|\hat{\varphi}|a\rangle - \varphi\right)\right\} = 0$$

$$= \langle \delta a|\, H - E - J\hat{\varphi}\,|a\rangle + \text{c.c.}$$

to get

$$(H - E - J\hat{\varphi})|a\rangle = 0$$

Thus $|a\rangle$ is an eigenstate for H_J with eigenvalue E, so $E = E(H_J)$, and $\varphi = \langle a|\hat{\varphi}|a\rangle = \varphi_J$; the minimum value of $\langle a|H|a\rangle$ is

$$\langle a|H|a\rangle = \langle a|E + J\hat{\varphi}|a\rangle = E(H_J) + J\langle a|\hat{\varphi}|a\rangle$$

$$= E(H_J) + J\varphi_J = W(\varphi_J).$$

To summarize, $W(\varphi)$ is also the minimum of the energy $\langle a|H|a\rangle$ among states $|a\rangle$ in which the operator $\hat{\varphi}$ has the (average) value φ

At the moment I have many links to sort out. Suppose I start out with [redacted] the Markov chain example which I learned from Stroock's book. There is a state space M and a [redacted] compatible family of probability measures ν_n on M^n. We get [redacted] a Hilbert space, in fact, a chain of Hilbert spaces $L^2(M^n, \nu_n)$, which fit together to form the limit space $L^2(M$[redacted]$^{\infty}, \nu_\infty)$.

[redacted]

I recall that in this situation we [redacted] can define martingales, [redacted] and a measure preserving transformation to which the ergodic theorem applies. So what I see is a very complicated (say very infinite Hilbert space) based on the measure ν_∞ on M^∞ [redacted]. Now another thing we can do is to take the doubly-infinite product $M^{\mathbb{Z}}$ and put a probability measure ν on this which projects into ν_n on M^n for each n.

A further ingredient is the set of generating functions for the different measures ν_n:

$$\int_{M^n} e^{\frac{1}{n}\sum f(x_j)}\, d\nu_n = \frac{\langle \nu \text{[redacted]} | (p e^{\frac{1}{n} f})^n | 1 \rangle}{\langle \nu | p^n | 1 \rangle}$$

Another thing one can do, and this is important for the large deviation theory, is to introduce an applied magnetic field. This modifies the sequence of measures [redacted] ν_n.

June 12, 1985

There is a new idea in this Markov chain **business** which I learn from Fried's paper, and which goes back to Ruelle. Roughly the point is that partition functions and zeta functions are related, say by analogy. But zeta functions can be attached to expanding maps and Anosov diffeomorphisms as generating functions for the numbers of periodic points. It seems that by means of "Markov partitions", one can relate these zeta functions to the characteristic polynomial of a "transfer" matrix.

Hence the key thing to understand is how the Markov partitions capture or encode the fixpoint information. A Markov partition is given by a 0-1 matrix, that is, a subset $A \subset I \times I$ where I is the index set. I is assumed to be finite. One lets Σ_A be the subset of $I^{\mathbb{Z}}$ consisting of sequences (i_n) such that $(i_n, i_{n+1}) \in A$ for all n.

Σ_A is a compact space with an action of $\mathbb{Z}$. There is a connection between properties of this action and the 0-1 matrix corresponding to A viewed as a positive matrix. In fact, the ζ function of the shift is supposed to be related to the characteristic polynomial of the matrix.

June 13, 1985

Let A be a non-negative integral matrix with index set I. We construct a correspondence C on I

$$\begin{tikzcd} & C \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\ I & & I \end{tikzcd}$$

as follows. $C = \{(i,j,m) \in I \times I \times \mathbb{N}^* \mid 1 \leq m \leq a_{ij}\}$

$$p_2(i,j,m) = j$$
$$p_1(i,j,m) = i$$

Notice that C gives a_{ij} ways to go from j to i. This correspondence gives rise to an operator on the space of functions on I, namely,

$$f \longmapsto p_{1*} p_2^* f$$

$$(p_{1*} p_2^* f)(i) = \sum_{(i,j,m)\in C} f(j) = \sum_j a_{ij} f(j).$$

Hence this operator is just the matrix A.

Let Ω be the infinite product of the correspondence C over I; this means the subset of $C^{\mathbb{Z}}$ consisting of sequences (γ_n) such that

$$p_2(\gamma_n) = p_1(\gamma_{n+1}) \quad \text{for all } n.$$

This is a compact space on which $\mathbb{Z}$ acts by shifting.

Let f be the shift on Ω. We want to see that

(★) $\quad \mathrm{card}\left(\mathrm{Fix}(f^n)\right) = \mathrm{tr}(A^n).$

To see this we first note that $\mathrm{Fix}(f^n)$ is the number of sequences $\gamma_1, \dots, \gamma_n$ in C such that $\mathrm{source}(\gamma_i) = \mathrm{target}(\gamma_{i+1})$ and $\mathrm{source}(\gamma_n) = \mathrm{target}(\gamma_1)$

$$\begin{tikzcd}[column sep=small]
 & \gamma_1 \arrow[dl] \arrow[dr] & & \gamma_2 \arrow[dl] \arrow[dr] & & \cdots & & \gamma_n \arrow[dl] \arrow[dr] & \\
\iota_n & & \iota_1 & & \iota_2 & \cdots & \iota_{n-1} & & \iota_n
\end{tikzcd}$$

Let's count them as a sum over the different "objects" sequence $(\iota_1, \iota_2, \dots, \iota_n)$. The number of γ-sequences over this object sequense is clearly

$$a_{\iota_1\iota_2}\, a_{\iota_2\iota_3} \cdots a_{\iota_{n-1}\iota_n}\, a_{\iota_n\iota_1}.$$

Summing up then gives the trace of A^n.

An obvious consequense of (★) is that the generating function for the fixpoints

$$\zeta(z) = e^{\sum_1^\infty \frac{z^n}{n} \mathrm{card}(\mathrm{Fix}\, f^n)}$$

is

$$e^{\sum_1^\infty \frac{z^n}{n} \mathrm{tr}(A^n)} = \frac{1}{\det(1 - zA)}$$

Let's try to describe Fried's paper which generalizes Ruelle: Inv. Math 34 (1976). He studies a flow ϕ_t on a compact X together with a lift ψ_t of ϕ_t to a vector bundle E over X.

If γ is a closed orbit, let $\mu(\gamma)$ be its multiplicity $(\gamma = \gamma_0^{\mu(\gamma)}$ where γ_0 is prime closed orbit of ϕ), and denote by $\operatorname{tr}\psi_\gamma$ the trace of $\psi_{\ell(\gamma)} : E_x \to E_x$ for x in the orbit (monodromy is well-defined up to conjugacy). Then the Ruelle zeta fn. is

$$R_\psi(z) = \exp\Big(-\sum_\gamma \frac{1}{\mu(\gamma)} e^{-z\ell(\gamma)} \operatorname{tr}(\psi_\gamma)\Big)$$

$$= \prod_{\gamma \text{ prime}} \det\big(1 - e^{-z\ell(\gamma)} \psi_\gamma\big)$$

In the case of interest X is a basic set Λ for an axiom A flow on M. This means in particular that the vector bundle $TM|\Lambda$ splits into $E^u \oplus E^s \oplus E^c$, u unstable, s stable, E^c 1-diml spanned by $d\phi/dt$.

For technical reasons I don't yet understand having to do with holomorphic Lefschetz formulas, it is easier to work with Selberg zeta functions. These involve the monodromy of the <u>stable</u> bundle E^s; denote this S_γ, it is well defined up to conjugacy. The Selberg zeta function is

$$S_\psi(z) = \exp\Big(-\sum_\gamma \frac{1}{\mu(\gamma)} \frac{e^{-z\ell(\gamma)} \operatorname{tr}(\psi_\gamma)}{\det(1 - S_\gamma)}\Big)$$

$$= \prod_{\gamma \text{ prime}} \prod_{j \ge 0}^{\infty} \det\big(1 - e^{-z\ell(\gamma)} \psi_\gamma \otimes S_\gamma^j\big)$$

The main result is to prove these are meromorphic functions of z assuming that everything is analytic.

I now want to describe the methods which go into the proof. The analyticity enters in order to use the following holomorphic fixpt formula:

see p 55

June 14, 1985

[redacted] I would like to get to the bottom of Ruelle's zeta functions, and especially I would like to understand the statistical mechanical motivations. There are a number of nice ideas.

[redacted] One of these ideas is to take apart the manifold into pieces which can be complexified and such that the glueing is via contractive maps. This enables one to define [redacted] holomorphic transfers $\mathcal{L}(z)$ [redacted] which are a nuclear operators. Then the Grothendieck theory [redacted] yields an entire fn. $\det(1-\mathcal{L}(z))$.

Let's now consider the case of an expanding map $g: M \longrightarrow M$ and suppose one is given over M a vector bundle E and a lift $L: g^*E \to E$ of g to E. ~~Then we get an operator $\mathcal{L} = Lg^*$ on $\Gamma(M,E)$.~~

One defines a zeta function

$$R(t) = \exp \sum \frac{t^n}{n} a_n$$

$$a_n = \sum_{g^n(x)=x} \operatorname{tr}\left(E_x \xleftarrow{L} E_{gx} \xleftarrow{L} \cdots \xleftarrow{L} E_{\underset{x}{\overset{\shortparallel}{g^n x}}}\right)$$

Assuming everything is analytic, Ruelle proves $R(t)$ is meromorphic.

Example: E trivial line bdle, $L = id$. Then $a_n = |\mathrm{Fix}\, g^n|$, so $R(t)$ is the zeta function. Assuming no orientation problems, the Lefschetz formula shows R is rational. For example take $M = S^1$ and $g(z) = z^2$. Then

$$\mathrm{Fix}\, g^n = \{z \mid z^{2^n-1} = 1\} \quad \text{has card } 2^n - 1$$

and $$\exp\left(\sum_1^\infty \frac{t^n}{n}(2^n-1)\right) = \frac{1-t}{1-2t}$$

Lefschetz formula:

$$\sum_{f(P)=P} \operatorname{sgn}(\det 1-df_P) = \sum (-1)^q \operatorname{tr} f^* \text{ on } H^q$$

For $f(z)=z^k$, or $f(\theta)=k\theta$, one has $df = k\,d\theta$ so $1-df$ is mult. by $1-k<0$. Thus

$$-(k-1) = \underbrace{\operatorname{tr} f^* \text{ on } H^0}_{1} - \underbrace{\operatorname{tr} f^* \text{ on } H^1}_{k}$$

which checks. $\blacksquare$

The way to prove the zeta function is meromorphic is to relate it to characteristic series of homomorphic transfer operators:

$$\det(1-z\mathcal{L}).$$

The first candidate for $\mathcal{L}$ is the natural operator T on $\Gamma(M,E)$ given by $s \longmapsto L g^* s$. This is a "geometric endomorphism" in the sense of Atiyah and Bott, and because the graph of g is transversal to the diagonal, they define a trace for it and prove a fixpoint formula. However this does not seem to yield a proof that the ζ function is meromorphic.

<u>Lemma:</u> Let $\mathcal{U}$ be the closure of a bounded open set in $\mathbb{C}^d$, let $F: \mathcal{U} \to \mathrm{Int}(\mathcal{U})$ be holomorphic and a contraction. Let E be a holom. v.b. over $\mathcal{U}$ and $L: F^*(E) \to E$ a holomorphic v.b. map. Let $\mathcal{B}$ be the Banach space of continuous sections of E over $\mathcal{U}$ which are holomorphic in the interior. Let $\mathcal{L}$ be the operator on $\mathcal{B}$ given by $\mathcal{L}s = L F^*(s)$. Then $\mathcal{L}$ has a well-defined trace $\mathrm{Tr}\,\mathcal{L} = \sum \lambda_i$ where λ_i are the eigenvalues. The trace is given by

$$\mathrm{Tr}\,\mathcal{L} = \frac{\mathrm{tr}\,(L \text{ on } E_p)}{\det\,(1 - df_p)}$$

where p is the unique fixpt of F and df_p is regarded as a $\mathbb{C}$-linear endomorphism of $T_p\mathcal{U}$.

June 16, 1985

Here is the example of a Markov partition in the case of the expanding map $z \mapsto z^2$ on S^1. I look for a partition of $M = S^1$ into intervals R_i such that the different branches of $\sqrt{z}$ carry R_i into another R_j. Such a map $R_i \to R_j$ will be contracting.

The simplest choice for the partition is to use the interval $[0,1]$ to cover $S^1 = \mathbb{R}/\mathbb{Z}$. Then we have the two branches

$$[0,1] \rightrightarrows [0,1]$$
$$\theta \longmapsto \frac{\theta}{2} \text{ and } \frac{\theta}{2} + \frac{1}{2}$$

One should remember that the point of the Markov partitions is to furnish a correspondence C on $\coprod R_i$, so that the different components of C are graphs of contracting maps. One then replaces C and $\coprod R_i$ by the sets of components thereby obtaining a correspondence A on $I \times I$. From this correspondence one constructs the symbolic dynamical system Ω.

So in the example under consideration we see that I consists of 1 point and A has 2 points, so that the one sided iterated correspondence defined by A is just $A^{\mathbb{N}} = \{0,1\}^{\mathbb{N}}$. We now want to construct a map

$$\pi : A^{\mathbb{N}} \longrightarrow M$$

compatible with the backword shift in $A^{\mathbb{N}}$ and the expanding map on $M = S^1$.

To do this we first look at the one-sided iterated correspondence associated to $R = [0,1]$ and C. This means we want to give $\theta_0 \in R$ and an element c_1 of C transport θ_0 to a θ_1, then an element c_2 transporting θ_1 to a θ_2, etc. It is clear that c_1 is specified by giving the branch of $\sqrt{z}$ one is using. So c_1, c_2, etc are determined by elements $a_1, a_2, \dots \in \{0,1\}$ and

$$\theta_1 = \frac{\theta_0 + a_1}{2}, \quad \theta_2 = \frac{\theta_1 + a_2}{2}$$

Thus we see that the (one-sided) iterated correspondance of C is the product

$$R \times \{0,1\}^{N}$$

We can map this to M (?)

I seem to have the arrows wrong. What I want to get ultimately is the map

$$\{0,1\}^{N^+} \xrightarrow{\pi} S^1 = \mathbb{R}/\mathbb{Z}$$

$$(a_n) \longmapsto \sum_{n=1}^{\infty} a_n 2^{-n} \qquad \text{mod } 1.$$

which is compatible with the backward shift and multiplication by 2. This map is onto and it is 1-1 except over $\bigcup (\mathbb{Z}\cdot 2^{-n})/\mathbb{Z}$ where it is 2-1. Thus a finite 2-adic decimal can be represented in 2 ways.

$$a_1 a_2 \dots \overset{1}{a_k} 0 0 0 = a_1 a_2 \dots a_{k-1} 0 1 1 1 1 1 \dots$$

Next we note that as far as fixpts are concerned, the fixpoints of 2^k is the subgroup killed by $2^k - 1$. This intersects the subgroup of 2-power-torsion points where the map π is

2-1 only at the identity 0. Thus we get a one-one correspondence between fixpoints of order dividing k except for the identity. This explains why the 2^k becomes $2^k - 1$ under π. Note also that the exceptional point occurs at the boundary of R.

So what I want to get straight is the iterated correspondence. Somehow I have got involved with the solenoid of S^1 and the map of mult. by 2. This is the inverse limit of

$$\xrightarrow{2} S^1 \xrightarrow{2} S^1 \xrightarrow{2} S^1 \xrightarrow{2} S^1 .$$

Call it L. Then the kernel of $L \to S^1$ is the inverse limit of

$$\begin{tikzcd} \mu_8 \arrow[r] & \mu_4 \arrow[r] & \mu_2 \arrow[r] \arrow[d, equal] & 1 \\ \mathbb{Z}/8 \arrow[u] & \mathbb{Z}/4 & \mathbb{Z}/2 & \end{tikzcd}$$

which is the 2-adics $\mathbb{Z}_2$.

I have been forming the iterated correspondence in the wrong direction. What I want is the analogue of the space of sequences $(x, f(x), f^2(x), \ldots)$ with $x \in M$, but when x is in the cover $[0,1]$. Thus I want to have $x = \theta_0 \in [0,1]$ and its image under the map:

$$2\theta_0 = \theta_1 + a_1 \qquad a_1 \in \{0,1\}$$

and so forth

$$2\theta_1 = \theta_2 + a_2 \qquad a_2 \in \text{"} .$$

Thus

$$\theta_0 = \tfrac{1}{2}a_1 + \tfrac{1}{2}\left(\tfrac{1}{2}a_2 + \cdots\right) = \sum_{n=1}^{\infty} a_n 2^{-n}$$

Note that unless $\theta_0 = 1/2$, a_1 is uniquely determined. In other words the correspondence

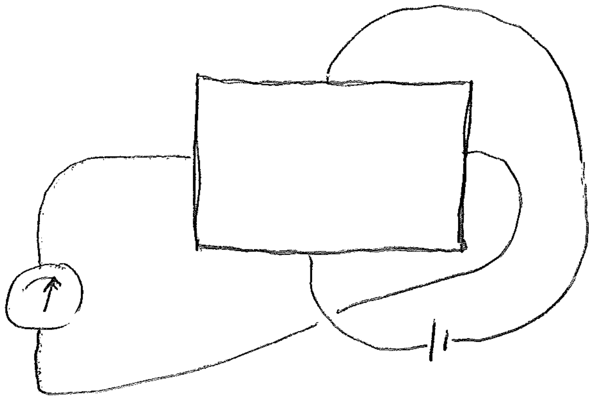
$$\begin{tikzcd}[column sep=small] & C \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\ R & & R \end{tikzcd} \qquad \begin{tikzcd}[column sep=small] & {[0,1] \amalg [0,1]} \arrow[dl, "p_1"'] \arrow[dr, "p_2"] & \\ {[0,1]} & & {[0,1]} \end{tikzcd}$$

where $p_1 = id + id$ and $p_2 = (\theta \mapsto \frac{1}{2}\theta) + (\theta \mapsto \frac{1}{2}(\theta + 1))$ is not a graph. Note p_2 is not bijective, but covers the point $\frac{1}{2}$ twice.

It is clear from the formula $\theta_k = \sum_{n>k} a_n 2^{k-n}$ that the iterated correspondence is homeomorphic to $\{0,1\}^{N^*}$.

June 20, 1985

Quantum Hall Effect:

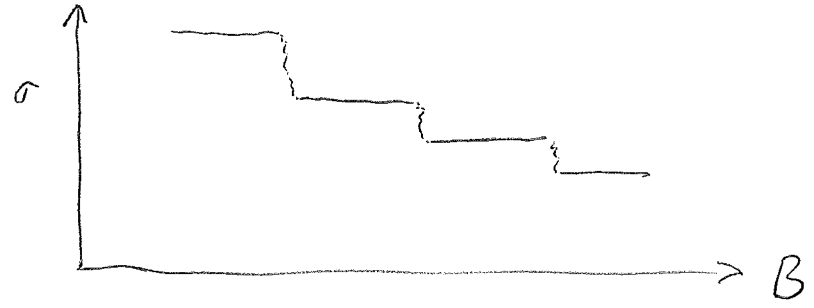


Plane region = interface between layers in a transistor

Magnetic field B $\perp$ to plane

Battery produces an electric field. Galvanometer measures the Hall current.

The physical fact to be explained is the conductivity curve:



and why the plateaus have <u>integral</u> values.

<u>Mathematical Model.</u> For a given B there are two parameters ϕ_1, ϕ_2 in the Hamiltonian which represent the fluxes thru the battery and the galvanometer loops. Really ϕ_1, ϕ_2 belong to a torus, for there are gauge transformations relating $\phi + \lambda$ where $\lambda = (\lambda_1, \lambda_2)$ is a lattice point.

In other words, we have a Hilbert space bundle over the torus of ϕ and a self-adjoint positive definite Hamiltonian family $H(\phi)$ in the fibres.

By genericness arguments, the ground state of each Hamiltonian $H(\phi)$ will be non-degenerate for most B. Thus we will get a line bundle over the torus formed by the family of ground states. By a suitable version of the adiabatic thm. one can identify the conductivity:

$$\sigma = \frac{1}{2\pi i} \int_{\text{Torus}} \operatorname{tr} P\,dP\,dP$$

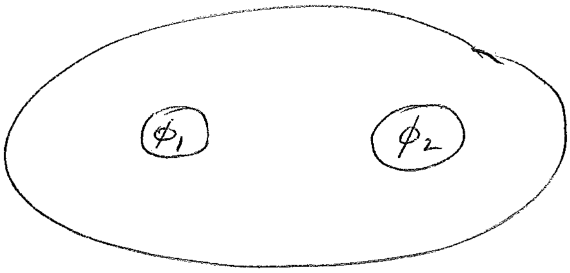
where P is the projector on the ground state. Thus σ is [redacted] essentially the first Chern class of the line bundle.

The jumps in σ occur at B where the ground state become degenerate. An interesting question is why σ decreases as B increases

Picture of the Hamiltonian

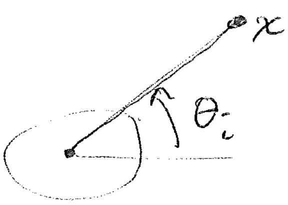
$$H = \sum_{i=1}^{n} \left(p_i + A_i + \phi_1 a_1(x_i) + \phi_2 a_2(x_i) \right)^2 + \underset{\text{interaction}}{W}$$

The 1-particle Hilbert space, which represents the configurations of the electron in the square + two loops, is something like functions on the plane region



with Dirichlet bdry conditions, where ϕ_i are thought of as fluxes through the hole. Somehow

$$a_i(x) = \operatorname{grad} \theta_i$$



7 June 22, 1985 (45 years old today)

Quantum Hall Effect:

Let's recall the Bohm-Aharonov idea: One considers a quantum mechanical situation where the configuration space M is not simply-connected. This means that there are non-trivial flat line bundles over M which can be used to twist the standard quantization given by $L^2(M)$ with $H = -\Delta + V$.

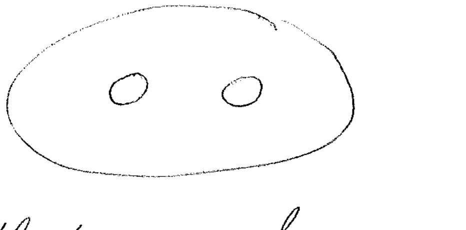
In the path integral picture, the path space breaks into components according to the elements of $\pi_1 M$. We can alter by a character χ of $\pi_1 M$ the path integral as follows

$$\int \mathcal{D}x \ \ \chi(x)\, e^{-S(x)}$$

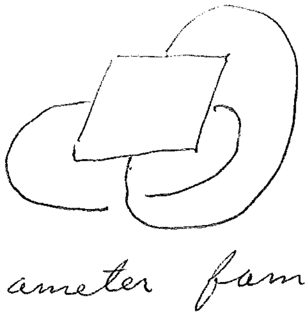
so as to obtain a new quantization with the same equations of motion.

Examples of this are: 1) Quantizing a rotor, where $M = SO(3)$. 2) Original Bohm-Aharonov situation where one has electrons moving outside a "solenoid", so that $M = \mathbb{R}^3 - \mathbb{R}$ or $\mathbb{R}^2 - pt$.

Now in the case of the Hall effect one has a configuration space M with $H_1(M) = \mathbb{Z} \times \mathbb{Z}$, i.e.



or



so that one has a two parameter family of flat line bundles over M. Better, there is a family of flat line bundles parameterized by a 2-torus T; there is a Poincaré line bundle

$$\begin{array}{c} L \\ \downarrow \\ T \times M \end{array}$$

I can now twist $H = -\Delta + V$ on $L^2(M)$ by each of the line bundles in this family. Then I get a family of Hilbert spaces over the torus and in each a self adjoint operator. Assuming that the ground states are non-degenerate we get a line bundle of ground states over T.

Further observation: Suppose we introduce the covering space $\tilde{M}$ of M corresponding to the quotient $\mathbb{Z}^2$ of $\pi_1 M$. I claim that if we pull the line bundle L back to $T \times \tilde{M}$ it becomes trivial. Better: what I want is for the family of Hilbert spaces + operators to be trivial over $\tilde{M}$. Each operator on M should come by descending a fixed operator on $\tilde{M}$.

Let's be more precise. I am considering a family of flat line bundles containing the trivial bundle. So all the line bundles are trivial, and choosing a trivialization we can represent the family by a family of connection forms which are closed.

In fact the isomorphism classes of flat line bundles is

$$H^1(M, S^1)$$

which fits into an exact sequence

$$H^1(M,\mathbb{Z}) \to H^1(M,\mathbb{R}) \longrightarrow H^1(M,S^1) \longrightarrow H^2(M,\mathbb{Z}) \to H^2(M,\mathbb{R})$$

whence $\quad 0 \longrightarrow H^1(M,\mathbb{R})/H^1(M,\mathbb{Z}) \longrightarrow H^1(M,S^1) \longrightarrow H^2(M,\mathbb{Z})_{tors} \to 0$

and the identity component is $H^1(M,\mathbb{R})/H^1(M,\mathbb{Z})$. Call this T, and put $V = H^1(M,\mathbb{R})$.

We lift V back into Ker d on $\Omega^1(M)_{\mathbb{R}}$. so that now each $\phi \in V$ gives us a closed 1-form α_ϕ on M which is linear in ϕ. Let us write this $\alpha_\phi = \phi\alpha = \phi_1\alpha_1 + \phi_2\alpha_2$. Now each α_i becomes cohomologically trivial on $\tilde{M}$, so we have $\alpha_\phi = df_\phi$ where $f_\phi = \phi_1 f_1 + \phi_2 f_2$ over $\tilde{M}$. Thus over $\tilde{M}$ things become gauge trivial.

It seems we have the following picture. We have an operator H on $L^2(\tilde{M})$ which is invariant under the action of $\mathbb{Z}\times\mathbb{Z}$. We also have the functions f_1, f_2 on $\tilde{M}$ which change under $\mathbb{Z}^2$ by constants (since $df_i = \alpha_i$ comes from $\tilde{M}$). Think of $\tilde{M}$ as a 2-torus, i.e. $\tilde{M}/\Gamma$ and $H = -\Delta + U$ where U is a Γ-periodic potential on $\tilde{M}$. Then for each choice of linear function on $\tilde{M}$, we get boundary conditions down on M. So what really is at stake is this operation of descending the operator on $\tilde{M}$.

Let us consider the following situation, namely, where one has a periodic potential V on a torus $M = \tilde{M}/\Gamma$. Then we have a unitary representation of Γ on $L^2(\tilde{M})$, so this can be regarded as a unitary module over the functions on the torus $\hat{\Gamma}$.

Thus we are thinking of $\tilde{M}$ as a bundle over M with the fibres Γ, so somehow $L^2(\tilde{M})$ becomes sections of a vector bundle over $\tilde{M}$ with fibres $\ell^2(\Gamma)$. Now the thing we are after is how H looks in each of these fibres. Let's consider some examples.

Let $\dim M = 1$, say $M = \mathbb{R}/\mathbb{Z}$. and let's take $H = -\frac{d^2}{dx^2}$. The eigenfunctions of H are e^{ikx} and this satisfies the periodicity condition

$$f(x+1) = e^{ia} f(x)$$

iff $k - a \in 2\pi\mathbb{Z}$. Thus the eigenvectors of H_ϕ, $\phi = e^{ia}$, are the e^{ikx} $k \in a + 2\pi\mathbb{Z} \in \mathbb{R}/2\pi\mathbb{Z}$. The eigenvalue is k^2.

Start e^{ia} out at 1, so $a = 0$, and begin to increase a. There is a unique ground state, namely for $k = a$, until e^{ia} reaches -1, whence we have two ground states $k = \pm\pi$. As a increases we jump from $k=a$ to $k=a-2\pi$, and the energy decreases down to zero.

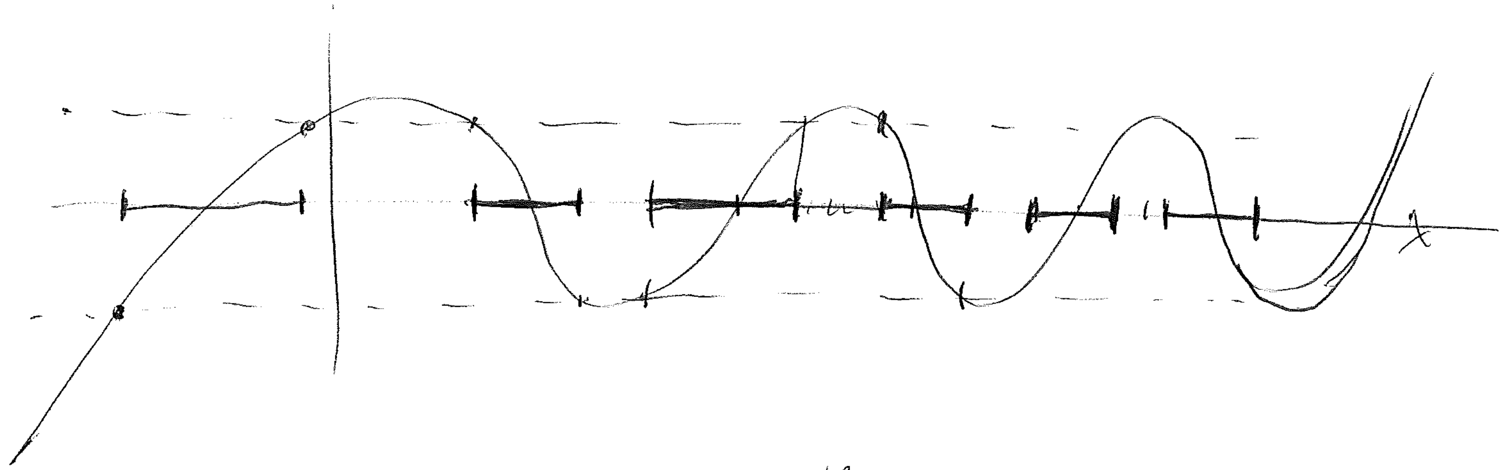
This example clearly shows the crossing of the energy levels.

On the other hand suppose that we take a periodic potential which has a band structure. For each real λ we propagate solutions of

$$\left(-\frac{d^2}{dx} + u\right)\psi = \lambda\psi$$

from $x=0$ to $x=1$. This gives a 2×2 matrix $S(\lambda)$

which is an entire function of λ, ~~and~~ is ~~be~~ real for λ real, and has determinant 1. The spectrum of $H = -\frac{d^2}{dx^2} + u$ on $L^2(\mathbb{R})$ occurs when the eigenvalues of $S(\lambda)$ are on the unit circle, i.e. when $-2 \leq \operatorname{tr} S(\lambda) \leq 2$. If we plot $\operatorname{tr} S(\lambda)$ we get bands:



If we are *inside* a band, then we have two distinct eigenvalues $e^{\pm i\alpha}$ which gives a unique eigenstate of H with energy lying in the band and having the periodicity $\phi = e^{i\alpha}$. Thus it seems fairly clear, but not rigorously, that within a band we have a unique eigenstate, that we don't get the crossing phenomenon encountered with the zero potential.

June 24, 1985

It seems that there is a link between the quantum Hall effect and the irrational rotation algebra described by Connes. Atiyah says this algebra enters when one considers a magnetic field which is irrational relative to the period lattice.

Let's look at this from the beginning. We consider a 2-torus $M = \mathbb{R}^2/\underbrace{\mathbb{Z}^2}_{\Gamma}$, and let $iB\,dx^1dx^2$ be an invariant real 2-form. Then we can choose a unique connection A in the trivial line $/\mathbb{R}^2$ bundle with curvature $iB\,dx^1dx^2$ which is radially around 0. Put $D = d + A$, or $D_\mu = \partial_\mu + A_\mu$ for the covariant derivative, whence

$$[D_1, D_2] = iB.$$

The algebra of operators on functions on $\mathbb{R}^2$ generated by D_1, D_2, id is a Heisenberg algebra. The group of unitary operators is a central extension

$$1 \longrightarrow S^1 \longrightarrow \widetilde{\mathbb{R}}^2 \longrightarrow \mathbb{R}^2 \longrightarrow 1.$$

Now we consider the induced central extension of Γ:

$$1 \longrightarrow S^1 \longrightarrow \tilde{\Gamma} \longrightarrow \Gamma \longrightarrow 1$$

The operators corresponding to $\tilde{\Gamma}$ form an irrational rotation algebra, assuming B is irrational. Precisely if $(1,0), (0,1) \in \mathbb{Z}^2$ generate Γ, then $\tilde{\Gamma}$ is generated by e^{D_1}, e^{D_2} which satisfy:

$$e^{D_1} e^{D_2} e^{-D_1} e^{-D_2} = e^{iB},$$

[redacted] Now suppose that B is integral (i.e. $B \in 2\pi\mathbb{Z}$). whence $\tilde{\Gamma} \simeq \Gamma \times S^1$ is abelian. The sort of thing we have been doing in the Q. Hall effect is to take the Hilbert space of L^2 functions on $\mathbb{R}^2$, regard it as a representation of Γ and then to decompose with respect to the characters of Γ. The piece belonging to a character ϕ of Γ is the space of sections of the line bundle over $M = \mathbb{R}^2/\Gamma$ associated to this character. This description seems to be linked to $B=0$, but really one can start with a fixed line bundle over M having the curvature $iB\,dx^1dx^2$, and then twist by characters of Γ.

[redacted] Now I ought to be able to describe the spectrum of $H = -(D_1^2 + D_2^2)$ in any of these representations. Let's leave this calculation for a while.

Again suppose we start with (correct on the center) a representation $\mathcal{H}$ of the Heisenberg algebra, for example, the L^2 sections of the connected line bundle with curvature $iB\,dx^1dx^2$ [redacted] over $\mathbb{R}^2$. Then we can regard $\mathcal{H}$ as a representation of $\tilde{\Gamma}$. In the integral case $\mathcal{H}$ becomes a repn. of Γ and so fibres over $\hat{\Gamma} = T$. There is inside of the Heisenberg group algebra the operator $e^{-\beta H}$. I see now that I need to have both a left and right multiplication; I want $\tilde{\Gamma}$ to act on the right and to commute with multiplication by $e^{-\beta H}$ on the left.

The natural question is whether the operator $e^{-\beta H}$ on $\mathcal{H}$ viewed as a module over $\tilde{\Gamma}$ is of trace class in the **relative sense**. In this case its trace is an element of the rotation algebra generated by $\tilde{\Gamma}$. I can probably describe this element in general. Probably it is a sum of Gaussian factors indexed by the different elements of Γ.

Could put the Dirac operator in instead and compute $tr_s(e^{-\beta H})$, where $H = -\not{D}^2$. This should be independent of β.

Poincaré line bundle for the irrational rotation algebra?

June 25, 1985

Let's consider the constant 2-form $F = \frac{1}{i} B\,dx^1 dx^2$ on $\mathbb{R}^2$ and the family of connections

$$A_\phi = i\,d(\phi_1 x_1 + \phi_2 x_2) + \frac{B}{2i}(x^1 dx^2 - x^2 dx^1)$$

with curvature F. Suppose first of all that F is integral relative to the lattice $\Gamma \subset \mathbb{R}^2$. Then the connection defines a lifting of translation by Γ to the line bundle L. For example if $\Gamma = \mathbb{Z}^2$, then

$$e^{D_1} e^{D_2} e^{-D_1} e^{-D_2} = e^{-iB}$$

so that if $B \in 2\pi\mathbb{Z}$, we get an action of Γ on sections of L.

So we can divide out by Γ an obtain a 'connected' line bundle over $M = \mathbb{R}^2/\Gamma$, which we denote L_ϕ. Then we obtain a family of Hamiltonians $-D_\mu^2$ or $-\not{D}^2$ as ϕ varies, and because of gauge transformations the family is a well-defined as $\phi \in \mathbb{R}^2/\hat{\Gamma}$.

The Dirac operator is

$$\not{D} = \begin{pmatrix} 0 & D_1 - iD_2 \\ D_1 + iD_2 & 0 \end{pmatrix}$$

and one has

$$\not{D}^2 = D_\mu^2 + \underbrace{\tfrac{1}{2}\gamma^\mu\gamma^\nu F_{\mu\nu}}_{-i\gamma^1\gamma^2 B \;=\; \varepsilon B}$$

Thus $$-\not{D}^2 = -D_\mu^2 - \varepsilon B \geqslant -\varepsilon B.$$

From the fact that the non-zero eigenvalues of $-\not{D}^2$ on $S^\pm$ are the same, we see that the eigenvalues of $-D_\mu^2$ are stable under adding $2B$. In fact one knows the spectrum of $-D_\mu^2$ up over $\mathbb{R}^2$, since this can be easily related to the harmonic oscillator.