Large Deviations and Sum Rules for Orthogonal Polynomials

CLAPEM XIV

Universidad de Costa Rica, December, 2016

Barry Simon

IBM Professor of Mathematics and Theoretical Physics, Emeritus
California Institute of Technology
Pasadena, CA, U.S.A.

Lecture 2: Meromorphic Herglotz Functions and Proof of Killip Simon Sum Rule

- Lecture 1: OPRL, OPUC and Sum Rules
- Lecture 2: Meromorphic Herglotz Functions and Proof of KS Sum Rule
- Lecture 3: The Theory of Large Deviations
- Lecture 4: GNR Proof of Sum Rules

References

[OPUC2] B. Simon, *Orthogonal Polynomials on the Unit Circle, Part 2: Spectral Theory*, AMS Colloquium Series, **54.2**, American Mathematical Society, Providence, RI, 2005.

[SzThm] B. Simon, *Szegő's Theorem and Its Descendants: Spectral Theory for L^2 Perturbations of Orthogonal Polynomials*, M. B. Porter Lectures, Princeton University Press, Princeton, NJ, 2011.

[CA] B. Simon, *A Comprehensive Course in Analysis, Part 2A, Basic Complex Analysis*, American Mathematical Society, Providence, R.I., 2015.

[HA] B. Simon, *A Comprehensive Course in Analysis, Part 3, Harmonic Analysis*, American Mathematical Society, Providence, R.I., 2015.

Killip–Simon functions

Given a measure $d\mu = w(x)\,dx + d\mu_s$ with pure points $\{E_j^{\pm}\}_{j=1}^{N_\pm}$ outside $[-2, 2]$ (with + above 2 and - below -2) one defines (with $m(z) = \int (x-z)^{-1} d\mu(x)$ and $m(x) \equiv \lim_{y \downarrow 0} m(x + iy)$).

Killip–Simon functions

Given a measure $d\mu = w(x)\,dx + d\mu_s$ with pure points $\{E_j^{\pm}\}_{j=1}^{N_{\pm}}$ outside $[-2,2]$ (with + above 2 and - below -2) one defines (with $m(z) = \int (x-z)^{-1} d\mu(x)$ and $m(x) \equiv \lim_{y\downarrow 0} m(x+iy)$).

$$Q(\mu) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\sin(\theta)}{\operatorname{Im} m(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

Killip–Simon functions

Given a measure $d\mu = w(x)\,dx + d\mu_s$ with pure points $\{E_j^\pm\}_{j=1}^{N_\pm}$ outside $[-2,2]$ (with + above 2 and - below -2) one defines (with $m(z) = \int (x-z)^{-1} d\mu(x)$ and $m(x) \equiv \lim_{y \downarrow 0} m(x+iy)$).

$$Q(\mu) = \frac{1}{2\pi} \int_0^{2\pi} \log\left(\frac{\sin(\theta)}{\operatorname{Im} m(2\cos(\theta))} \right) \sin^2(\theta) d\theta$$

$$G(a) = a^2 - 1 - \log(a^2) \text{ and}$$

Killip–Simon functions

Given a measure $d\mu = w(x)\,dx + d\mu_s$ with pure points $\{E_j^\pm\}_{j=1}^{N_\pm}$ outside $[-2,2]$ (with + above 2 and - below -2) one defines (with $m(z) = \int (x-z)^{-1} d\mu(x)$ and $m(x) \equiv \lim_{y\downarrow 0} m(x+iy)$).

$$Q(\mu) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\sin(\theta)}{\operatorname{Im} m(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

$$G(a) = a^2 - 1 - \log(a^2) \text{ and}$$

$$F(E) \equiv \tfrac{1}{4}[\beta^2 - \beta^{-2} - \log(\beta^4)] \qquad E = \beta + \beta^{-1} \qquad |\beta| > 1$$

Killip–Simon functions

Given a measure $d\mu = w(x)\,dx + d\mu_s$ with pure points $\{E_j^{\pm}\}_{j=1}^{N_\pm}$ outside $[-2,2]$ (with + above 2 and - below -2) one defines (with $m(z) = \int (x-z)^{-1} d\mu(x)$ and $m(x) \equiv \lim_{y\downarrow 0} m(x+iy)$).

$$Q(\mu) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\sin(\theta)}{\operatorname{Im} m(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

$$G(a) = a^2 - 1 - \log(a^2) \text{ and}$$

$$F(E) \equiv \tfrac{1}{4}[\beta^2 - \beta^{-2} - \log(\beta^4)] \qquad E = \beta + \beta^{-1} \qquad |\beta| > 1$$

The gem will come from the fact that $F \geq 0$, vanishes exactly at $E = \pm 2$ and is $\mathrm{O}((|E|-2)^{3/2})$ there and that $G \geq 0$, vanishes exactly at $a = 1$ and is $\mathrm{O}((a-1)^2)$ there.

Killip–Simon sum rule

The Killip–Simon rule says that

$$Q(\mu) + \sum_{j,\pm} F(E_j^{\pm}) = \sum_{n=1}^{\infty} \tfrac{1}{4} b_n^2 + \tfrac{1}{2} G(a_n)$$

Killip–Simon sum rule

The Killip–Simon rule says that

$$Q(\mu) + \sum_{j,\pm} F(E_j^{\pm}) = \sum_{n=1}^{\infty} \tfrac{1}{4} b_n^2 + \tfrac{1}{2} G(a_n)$$

An important point is that it always holds although both sides may be $+\infty$.

Killip–Simon sum rule

The Killip–Simon rule says that

$$Q(\mu) + \sum_{j,\pm} F(E_j^{\pm}) = \sum_{n=1}^{\infty} \tfrac{1}{4} b_n^2 + \tfrac{1}{2} G(a_n)$$

An important point is that it always holds although both sides may be $+\infty$. Our main goal in Lecture 2 will be to sketch a variant of the original proof of this sum rule.

Killip–Simon sum rule

The Killip–Simon rule says that

$$Q(\mu) + \sum_{j,\pm} F(E_j^{\pm}) = \sum_{n=1}^{\infty} \tfrac{1}{4} b_n^2 + \tfrac{1}{2} G(a_n)$$

An important point is that it always holds although both sides may be $+\infty$. Our main goal in Lecture 2 will be to sketch a variant of the original proof of this sum rule. I know of many proofs of Szegő's Theorem but until recently all proofs of the Killip–Simon sum rule were variants of this proof.

Killip–Simon sum rule

The Killip–Simon rule says that

$$Q(\mu) + \sum_{j,\pm} F(E_j^{\pm}) = \sum_{n=1}^{\infty} \tfrac{1}{4}b_n^2 + \tfrac{1}{2}G(a_n)$$

An important point is that it always holds although both sides may be $+\infty$. Our main goal in Lecture 2 will be to sketch a variant of the original proof of this sum rule. I know of many proofs of Szegő's Theorem but until recently all proofs of the Killip–Simon sum rule were variants of this proof. We'll need a Poisson-Jensen formula for certain meromorphic functions, so I start by recalling the classical PJ formula in the subtle form found by Smirnov and Beurling.

Nevanlinna Class and Blaschke Products

I'll remind you of the classical results without proofs which you can find in my *Basic Complex Analysis*, Sections 9.8 and 9.9 and *Harmonic Analysis*, Sections 5.3, 5.6 and 5.7.

Nevanlinna Class and Blaschke Products

I'll remind you of the classical results without proofs which you can find in my *Basic Complex Analysis*, Sections 9.8 and 9.9 and *Harmonic Analysis*, Sections 5.3, 5.6 and 5.7. Let f be an analytic function on the unit disk, $\mathbb{D}$. We say that $f \in N$, the Nevanlinna class, if and only if, $\sup_{0<r<1} \int_0^{2\pi} \log_+ |f(re^{i\theta})| d\theta < \infty$.

Nevanlinna Class and Blaschke Products

I'll remind you of the classical results without proofs which you can find in my *Basic Complex Analysis*, Sections 9.8 and 9.9 and *Harmonic Analysis*, Sections 5.3, 5.6 and 5.7. Let f be an analytic function on the unit disk, $\mathbb{D}$. We say that $f \in N$, the Nevanlinna class, if and only if, $\sup_{0<r<1} \int_0^{2\pi} \log_+ |f(re^{i\theta})| d\theta < \infty$. It is a basic fact that if $\{z_j\}$ is a listing of the zero's of f counting multiplicity, then

$$f \in N \Rightarrow \sum (1 - |z_j|) < \infty$$

Nevanlinna Class and Blaschke Products

I'll remind you of the classical results without proofs which you can find in my *Basic Complex Analysis*, Sections 9.8 and 9.9 and *Harmonic Analysis*, Sections 5.3, 5.6 and 5.7. Let f be an analytic function on the unit disk, $\mathbb{D}$. We say that $f \in N$, the Nevanlinna class, if and only if, $\sup_{0<r<1} \int_0^{2\pi} \log_+ |f(re^{i\theta})| d\theta < \infty$. It is a basic fact that if $\{z_j\}$ is a listing of the zero's of f counting multiplicity, then

$$f \in N \Rightarrow \sum (1 - |z_j|) < \infty$$

and this implies that the Blaschke product $B(z) = \prod_{j=1}^{N} b(z, z_j)$ converges absolutely on the unit disk to an analytic function vanishing precisely at the z_j.

Nevanlinna Class and Blaschke Products

I'll remind you of the classical results without proofs which you can find in my *Basic Complex Analysis*, Sections 9.8 and 9.9 and *Harmonic Analysis*, Sections 5.3, 5.6 and 5.7. Let f be an analytic function on the unit disk, $\mathbb{D}$. We say that $f \in N$, the Nevanlinna class, if and only if, $\sup_{0<r<1} \int_0^{2\pi} \log_+ |f(re^{i\theta})| d\theta < \infty$. It is a basic fact that if $\{z_j\}$ is a listing of the zero's of f counting multiplicity, then

$$f \in N \Rightarrow \sum (1 - |z_j|) < \infty$$

and this implies that the Blaschke product $B(z) = \prod_{j=1}^{N} b(z, z_j)$ converges absolutely on the unit disk to an analytic function vanishing precisely at the z_j. Here:

$$b(z,w) = \begin{cases} z, & \text{if } w = 0 \\ -\frac{|w|(z-w)}{w(1-\bar{w}z)} & \text{if } w \neq 0 \end{cases}$$

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$. Any absolutely convergent Blaschke product, B, lies in H^∞ and has $|B^*(e^{i\theta})| = 1$ for a.e. θ.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$. Any absolutely convergent Blaschke product, B, lies in H^∞ and has $|B^*(e^{i\theta})| = 1$ for a.e. θ. For $p \geq 1$, f^* determines f via a complex Poisson representation but not necessarily if $p < 1$.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$. Any absolutely convergent Blaschke product, B, lies in H^∞ and has $|B^*(e^{i\theta})| = 1$ for a.e. θ. For $p \geq 1$, f^* determines f via a complex Poisson representation but not necessarily if $p < 1$.

For $p \geq 1$, h^p is defined like H^p but its elements, u, are real–valued harmonic, rather than analytic functions.

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$. Any absolutely convergent Blaschke product, B, lies in H^∞ and has $|B^*(e^{i\theta})| = 1$ for a.e. θ. For $p \geq 1$, f^* determines f via a complex Poisson representation but not necessarily if $p < 1$.

For $p \geq 1$, h^p is defined like H^p but its elements, u, are real–valued harmonic, rather than analytic functions. If $u \in h^1$, then the measure $u(re^{i\theta})d\theta/2\pi$ has a weak–* limit $d\mu$

Hardy Class and Smirnov–Beurling Representation

For $0 < p < \infty$, the Hardy class, H^p, is the set of f analytic on $\mathbb{D}$ with $\sup_{0<r<1} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta < \infty$ and H^∞ is the bounded analytic functions. For $p < p'$, one has $H^\infty \subset H^{p'} \subset H^p \subset N$, so, in particular, the Blaschke product converges. For any function, f in N and, in particular in any H^p, there is a function, f^*, on $\partial\mathbb{D}$ so for a.e. θ, $f^*(e^{i\theta}) = \lim_{r\uparrow 1} f(re^{i\theta})$. Any absolutely convergent Blaschke product, B, lies in H^∞ and has $|B^*(e^{i\theta})| = 1$ for a.e. θ. For $p \geq 1$, f^* determines f via a complex Poisson representation but not necessarily if $p < 1$.

For $p \geq 1$, h^p is defined like H^p but its elements, u, are real–valued harmonic, rather than analytic functions. If $u \in h^1$, then the measure $u(re^{i\theta})d\theta/2\pi$ has a weak–* limit $d\mu$ and u can be recovered via a Poisson formula.

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part.

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

If $f \in N$, then $f/B \in N$ so $\log(|f/B|) \in h^1$ and we obtain the Poisson–Jensen representation of Smirnov and Beurling

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

If $f \in N$, then $f/B \in N$ so $\log(|f/B|) \in h^1$ and we obtain the Poisson–Jensen representation of Smirnov and Beurling

$$f(z) = \omega B(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \right)$$

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

If $f \in N$, then $f/B \in N$ so $\log(|f/B|) \in h^1$ and we obtain the Poisson–Jensen representation of Smirnov and Beurling

$$f(z) = \omega B(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \right)$$

where $|\omega| = 1$. If $f \in H^p$, then the singular part of the measure is negative and one has Beurling's inner/outer factorization.

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

If $f \in N$, then $f/B \in N$ so $\log(|f/B|) \in h^1$ and we obtain the Poisson–Jensen representation of Smirnov and Beurling

$$f(z) = \omega B(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \right)$$

where $|\omega| = 1$. If $f \in H^p$, then the singular part of the measure is negative and one has Beurling's inner/outer factorization. If $\log(|f/B|)$ is in some h^p, $p > 1$, there is no inner factor and the PJ formula takes the form

Hardy Class and Smirnov–Beurling Representation

Moreover, the a.e. pointwise limit u^* exists and $u^*(e^{i\theta})d\theta/2\pi$ is the a.c. part of $d\mu$ although there can also be a singular part. If $u \in h^p$, $p > 1$, this singular part is absent.

If $f \in N$, then $f/B \in N$ so $\log(|f/B|) \in h^1$ and we obtain the Poisson–Jensen representation of Smirnov and Beurling

$$f(z) = \omega B(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) \right)$$

where $|\omega| = 1$. If $f \in H^p$, then the singular part of the measure is negative and one has Beurling's inner/outer factorization. If $\log(|f/B|)$ is in some h^p, $p > 1$, there is no inner factor and the PJ formula takes the form

$$f(z) = \omega B(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} \log|f(e^{i\theta})| \frac{d\theta}{2\pi} \right)$$

Meromorphic Herglotz Functions

By a *meromorphic Herglotz* function, we mean a function meromorphic on $\mathbb{D}$, real on $(-1, 1)$ with $\operatorname{Im} z > 0 \Rightarrow \operatorname{Im} f(z) > 0$.

Meromorphic Herglotz Functions

By a *meromorphic Herglotz* function, we mean a function meromorphic on $\mathbb{D}$, real on $(-1,1)$ with $\operatorname{Im} z > 0 \Rightarrow \operatorname{Im} f(z) > 0$. It is easy to see that such functions have zeros and poles only on $(-1,1)$ and the zeros and poles are simple and interlace.

Meromorphic Herglotz Functions

By a *meromorphic Herglotz* function, we mean a function meromorphic on $\mathbb{D}$, real on $(-1, 1)$ with $\operatorname{Im} z > 0 \Rightarrow \operatorname{Im} f(z) > 0$. It is easy to see that such functions have zeros and poles only on $(-1, 1)$ and the zeros and poles are simple and interlace. If one looks at the product of Blaschke factors and their inverses for the zeros and poles in $(-r, r)$, it can be shown that they have a limit as $r \uparrow 1$ – an analog of alternating sums converging.

Meromorphic Herglotz Functions

By a *meromorphic Herglotz* function, we mean a function meromorphic on $\mathbb{D}$, real on $(-1,1)$ with $\operatorname{Im} z > 0 \Rightarrow \operatorname{Im} f(z) > 0$. It is easy to see that such functions have zeros and poles only on $(-1,1)$ and the zeros and poles are simple and interlace. If one looks at the product of Blaschke factors and their inverses for the zeros and poles in $(-r,r)$, it can be shown that they have a limit as $r \uparrow 1$ – an analog of alternating sums converging. Let's suppose $f(0) = 0$ and let $B(z)$ be the limiting product of zero and pole Blashcke factors other than the zero at 0.

Meromorphic Herglotz Functions

One can prove that in $\mathbb{D} \cap \mathbb{C}_+$, one has that $|\arg zB(z)| \leq 2\pi$ so that $\arg(f(z)/zB(z))$ is bounded on $\mathbb{D}$.

Meromorphic Herglotz Functions

One can prove that in $\mathbb{D} \cap \mathbb{C}_+$, one has that $|\arg zB(z)| \leq 2\pi$ so that $\arg(f(z)/zB(z))$ is bounded on $\mathbb{D}$. Since $\arg(g) = \operatorname{Im}(\log g)$, M. Riesz's Theorem implies that $\log(f(z)/zB(z))$ is in all H^p with $p < \infty$ so it obeys a Poisson–Jensen formula (with no singular inner part). Thus

Meromorphic Herglotz Functions

One can prove that in $\mathbb{D} \cap \mathbb{C}_+$, one has that $|\arg zB(z)| \leq 2\pi$ so that $\arg(f(z)/zB(z))$ is bounded on $\mathbb{D}$. Since $\arg(g) = \operatorname{Im}(\log g)$, M. Riesz's Theorem implies that $\log(f(z)/zB(z))$ is in all H^p with $p < \infty$ so it obeys a Poisson–Jensen formula (with no singular inner part). Thus

$$f(z) = zB(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} \log |f(e^{i\theta})| \frac{d\theta}{2\pi} \right)$$

Meromorphic Herglotz Functions

One can prove that in $\mathbb{D} \cap \mathbb{C}_+$, one has that $|\arg zB(z)| \leq 2\pi$ so that $\arg(f(z)/zB(z))$ is bounded on $\mathbb{D}$. Since $\arg(g) = \operatorname{Im}(\log g)$, M. Riesz's Theorem implies that $\log(f(z)/zB(z))$ is in all H^p with $p < \infty$ so it obeys a Poisson–Jensen formula (with no singular inner part). Thus

$$f(z) = zB(z) \exp\left(\int \frac{e^{i\theta} + z}{e^{i\theta} - z} \log|f(e^{i\theta})| \frac{d\theta}{2\pi} \right)$$

Taking $\log$'s, one gets relations between Taylor coefficients of $\log(f(z)/z)$, certain sums involving $\log$s or powers of zeros and poles and integrals $\cos(n\theta) \log|f(e^{i\theta})|$.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x - z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x-z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2, 2]$ flipping the sign of $\mathrm{Im} z$.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x-z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2,2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2,2]$, $M(z) = -m(z+z^{-1})$ is what we called a meromorphic Herglotz function.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x - z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2, 2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2, 2]$, $M(z) = -m(z + z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x-z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2, 2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2, 2]$, $M(z) = -m(z + z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1. The Taylor coefficients of $\log M(z)$ about zero are related to those of $m(z)$ at infinity and so polynomials in the Jacobi parameters.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x - z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2, 2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2, 2]$, $M(z) = -m(z + z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1. The Taylor coefficients of $\log M(z)$ about zero are related to those of $m(z)$ at infinity and so polynomials in the Jacobi parameters. We use $\pi w(x) = \lim_{\epsilon \downarrow 0} \mathrm{Im}\, m(x + i\epsilon)$.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x-z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2,2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2,2]$, $M(z) = -m(z+z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1. The Taylor coefficients of $\log\, M(z)$ about zero are related to those of $m(z)$ at infinity and so polynomials in the Jacobi parameters. We use $\pi w(x) = \lim_{\epsilon \downarrow 0} \mathrm{Im}\, m(x+i\epsilon)$.

The above procedure thus yields a relation between polynomials of Jacobi parameters, the difference of functions of the eigenvalues of J and J_1 and integral of $\log |M(e^{i\theta})|$.

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x-z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2,2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2,2]$, $M(z) = -m(z+z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1. The Taylor coefficients of $\log M(z)$ about zero are related to those of $m(z)$ at infinity and so polynomials in the Jacobi parameters. We use $\pi w(x) = \lim_{\epsilon \downarrow 0} \mathrm{Im}\, m(x + i\epsilon)$.

The above procedure thus yields a relation between polynomials of Jacobi parameters, the difference of functions of the eigenvalues of J and J_1 and integral of $\log |M(e^{i\theta})|$. Because $m(z)^{-1} = b_1 - z - a_1^2 m_1(z)$, one finds that $|M(e^{i\theta})|^{-2} \mathrm{Im}\, M(e^{i\theta}) = a_1^2 \mathrm{Im}\, M_1(e^{i\theta})$

Case Step–by-Step Sum Rules

Recall that $m(z) = \int d\mu(x)/(x - z)$. It defines a Herglotz function on $\mathbb{C}_+$, real on $\mathbb{R}$. The map $z \mapsto z + z^{-1}$ maps $\mathbb{D}$ to $(\mathbb{C} \cup \infty) \setminus [-2, 2]$ flipping the sign of $\mathrm{Im} z$. Thus, if $\sigma_{ess}(J) \subset [-2, 2]$, $M(z) = -m(z + z^{-1})$ is what we called a meromorphic Herglotz function. Its poles are the eigenvalues of J under the inverse image of the map $z \mapsto z + z^{-1}$ and its zeros are the same for J_1. The Taylor coefficients of $\log M(z)$ about zero are related to those of $m(z)$ at infinity and so polynomials in the Jacobi parameters. We use $\pi w(x) = \lim_{\epsilon \downarrow 0} \mathrm{Im}\, m(x + i\epsilon)$.

The above procedure thus yields a relation between polynomials of Jacobi parameters, the difference of functions of the eigenvalues of J and J_1 and integral of $\log |M(e^{i\theta})|$. Because $m(z)^{-1} = b_1 - z - a_1^2 m_1(z)$, one finds that $|M(e^{i\theta})|^{-2} \mathrm{Im}\, M(e^{i\theta}) = a_1^2 \mathrm{Im}\, M_1(e^{i\theta})$ so the log integral is a $\log$ of ratios of w and w_1.

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters).

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity and Killip and I found that none of the Case rules had the required positivity.

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity and Killip and I found that none of the Case rules had the required positivity.

However, we discovered that $C_0 + \frac{1}{2}C_2$ had the required positivity.

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity and Killip and I found that none of the Case rules had the required positivity.

However, we discovered that $C_0 + \frac{1}{2}C_2$ had the required positivity. We had no explanation of why this was so but observed it.

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity and Killip and I found that none of the Case rules had the required positivity.

However, we discovered that $C_0 + \frac{1}{2}C_2$ had the required positivity. We had no explanation of why this was so but observed it. We called this the P_2 sum rule (P for positive) and it is now known as the Killip–Simon sum rule.

P_2 Sum Rule

What results is a step–by–step sum rule which if iterated with boundary terms dropped yields the formal sum rules stated by Case (although, unlike Case, Killip and I had explicit formulae for the polynomials in the Jacobi parameters). These C_n step–by–step sum rules, especially C_0 have turned out to be useful in spectral theory, but to get a gem, one needs positivity and Killip and I found that none of the Case rules had the required positivity.

However, we discovered that $C_0 + \frac{1}{2}C_2$ had the required positivity. We had no explanation of why this was so but observed it. We called this the P_2 sum rule (P for positive) and it is now known as the Killip–Simon sum rule. The rather complicated functions F and G just arose by taking the functions from the Case sum rule and combining them.

Step–by–Step P_2 Sum Rule

The explicit result is:

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\frac{1}{4}b_1^2 + \frac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm} \left[F(E_j^\pm(J)) - F(E_j^\pm(J_1)) \right]$$

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\tfrac{1}{4}b_1^2 + \tfrac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm} \Big[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_1)) \Big]$$

where

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\tfrac{1}{4}b_1^2 + \tfrac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm} \left[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_1))\right]$$

where

$$Q(J|J_1) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\operatorname{Im} M_1(2\cos(\theta))}{\operatorname{Im} M(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\frac{1}{4}b_1^2 + \frac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm} \left[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_1)) \right]$$

where

$$Q(J|J_1) = \frac{1}{2\pi} \int_0^{2\pi} \log \left(\frac{\operatorname{Im} M_1(2\cos(\theta))}{\operatorname{Im} M(2\cos(\theta))} \right) \sin^2(\theta) d\theta$$

By interlacing, the sum of F terms is always convergent. And one can prove that the integral defining Q is always convergent.

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\tfrac{1}{4}b_1^2 + \tfrac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm}\left[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_1))\right]$$

where

$$Q(J|J_1) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\operatorname{Im} M_1(2\cos(\theta))}{\operatorname{Im} M(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

By interlacing, the sum of F terms is always convergent. And one can prove that the integral defining Q is always convergent. The $\log$ terms in F and G comes from the leading, i.e. C_0 rule.

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\frac{1}{4}b_1^2 + \frac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm}\left[F(E_j^\pm(J)) - F(E_j^\pm(J_1))\right]$$

where

$$Q(J|J_1) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\operatorname{Im} M_1(2\cos(\theta))}{\operatorname{Im} M(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

By interlacing, the sum of F terms is always convergent. And one can prove that the integral defining Q is always convergent. The $\log$ terms in F and G comes from the leading, i.e. C_0 rule. $\sin^2\theta$ enters as $\frac{1}{2}[1-\cos(2\theta)]$, the 1 from C_0 and $\cos(2\theta)$ from C_2. Iterating

Step–by–Step P_2 Sum Rule

The explicit result is:

$$\tfrac{1}{4}b_1^2 + \tfrac{1}{2}G(a_1) = Q(J|J_1) + \sum_{j,\pm}\left[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_1))\right]$$

where

$$Q(J|J_1) = \frac{1}{2\pi}\int_0^{2\pi} \log\left(\frac{\operatorname{Im} M_1(2\cos(\theta))}{\operatorname{Im} M(2\cos(\theta))}\right)\sin^2(\theta)d\theta$$

By interlacing, the sum of F terms is always convergent. And one can prove that the integral defining Q is always convergent. The $\log$ terms in F and G comes from the leading, i.e. C_0 rule. $\sin^2\theta$ enters as $\frac{1}{2}[1-\cos(2\theta)]$, the 1 from C_0 and $\cos(2\theta)$ from C_2. Iterating

$$\sum_{j=1}^{n} \tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j) = Q(J|J_n) + \sum_{j,\pm}\left[F(E_j^{\pm}(J)) - F(E_j^{\pm}(J_n))\right]$$

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j)\right] \leq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] \le Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove.

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j)\right] \le Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove. If both are finite, the same is true for $Q(J_n)$ and $\sum_{j,\pm} F(E_j^{\pm}(J_n))$

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] \leq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove. If both are finite, the same is true for $Q(J_n)$ and $\sum_{j,\pm} F(E_j^{\pm}(J_n))$ so in the iterated step–by–step sum rule, we can write $Q(J|J_n) = Q(J) - Q(J_n)$ and move both J_n terms to the other side

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j)\right] \leq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove. If both are finite, the same is true for $Q(J_n)$ and $\sum_{j,\pm} F(E_j^{\pm}(J_n))$ so in the iterated step–by–step sum rule, we can write $Q(J|J_n) = Q(J) - Q(J_n)$ and move both J_n terms to the other side and drop them to get the inequality for $\sum_{j=1}^{n}$

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j)\right] \leq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove. If both are finite, the same is true for $Q(J_n)$ and $\sum_{j,\pm} F(E_j^{\pm}(J_n))$ so in the iterated step–by–step sum rule, we can write $Q(J|J_n) = Q(J) - Q(J_n)$ and move both J_n terms to the other side and drop them to get the inequality for $\sum_{j=1}^{n}$ and then take $n \to \infty$.

Upper Bound on Jacobi Sum

We'll prove the sum rule by proving two inequalities. First that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4}b_j^2 + \tfrac{1}{2}G(a_j)\right] \le Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

If either $Q(J)$ or $\sum_{j,\pm} F(E_j^{\pm}(J))$ is infinite, then there is nothing to prove. If both are finite, the same is true for $Q(J_n)$ and $\sum_{j,\pm} F(E_j^{\pm}(J_n))$ so in the iterated step–by–step sum rule, we can write $Q(J|J_n) = Q(J) - Q(J_n)$ and move both J_n terms to the other side and drop them to get the inequality for $\sum_{j=1}^{n}$ and then take $n \to \infty$. The fact that all functions are positive is clearly needed in the dropping argument.

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \mid \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \,|\, \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

One has $H(\nu \,|\, \mu) \geq 0$ with equality only if $\mu = \nu$.

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \,|\, \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

One has $H(\nu \,|\, \mu) \geq 0$ with equality only if $\mu = \nu$. The name Kullback–Leibler (KL) divergence is used by statisticians and has been taken over in the probability literature.

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \mid \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

One has $H(\nu \mid \mu) \geq 0$ with equality only if $\mu = \nu$. The name Kullback–Leibler (KL) divergence is used by statisticians and has been taken over in the probability literature. In the information theory literature it is called the relative entropy although in some of the statistical mechanics literature its negative is often called by that name.

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \,|\, \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

One has $H(\nu \,|\, \mu) \geq 0$ with equality only if $\mu = \nu$. The name Kullback–Leibler (KL) divergence is used by statisticians and has been taken over in the probability literature. In the information theory literature it is called the relative entropy although in some of the statistical mechanics literature its negative is often called by that name.

Notice that the OPUC Szegő integral is precisely $-H(\frac{d\theta}{2\pi} \,|\, \mu)$

Lower Semicontinuity of the KL Divergence

Given a pair of probability measures, μ and ν on the same space, one defines their *Kullback–Leibler (KL) divergence* by

$$H(\nu \,|\, \mu) = \begin{cases} \int \log\left(\frac{d\nu}{d\mu}\right)\, d\nu, & \text{if } \nu \text{ is } \mu\text{–a.c.} \\ \infty, & \text{otherwise.} \end{cases}$$

One has $H(\nu \,|\, \mu) \geq 0$ with equality only if $\mu = \nu$. The name Kullback–Leibler (KL) divergence is used by statisticians and has been taken over in the probability literature. In the information theory literature it is called the relative entropy although in some of the statistical mechanics literature its negative is often called by that name.

Notice that the OPUC Szegő integral is precisely $-H(\frac{d\theta}{2\pi} \,|\, \mu)$ and what we called $Q(\mu)$ in the KS sum rule is precisely $H(\nu \,|\, \mu)$ where $d\nu(x) = (2\pi)^{-1}(4-x^2)^{1/2}\chi_{[-2,2]}(x)dx$.

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity.

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \,|\, \mu) = \sup_f \left(-\int f \, d\mu(x) + \int [1 + \log(f(x))] \, d\nu(x) \right)$$

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \mid \mu) = \sup_f \left(-\int f \, d\mu(x) + \int [1 + \log(f(x))] \, d\nu(x) \right)$$

where the $\sup$ is taken over all strictly positive continuous functions.

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \mid \mu) = \sup_f \left(-\int f \, d\mu(x) + \int [1 + \log(f(x))] \, d\nu(x) \right)$$

where the $\sup$ is taken over all strictly positive continuous functions. If $d\nu = g \, d\mu$ with g continuous and strictly positive, then the quantity in the $\sup$ when $f = g$ is H and Jensen's inequality implies the $\sup$ is always great than H.

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \mid \mu) = \sup_f \left(-\int f\, d\mu(x) + \int [1 + \log(f(x))]\, d\nu(x) \right)$$

where the $\sup$ is taken over all strictly positive continuous functions. If $d\nu = g\, d\mu$ with g continuous and strictly positive, then the quantity in the $\sup$ when $f = g$ is H and Jensen's inequality implies the $\sup$ is always great than H. Some approximation arguments complete the proof of the variation principle.

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \mid \mu) = \sup_f \left(-\int f \, d\mu(x) + \int [1 + \log(f(x))] \, d\nu(x) \right)$$

where the $\sup$ is taken over all strictly positive continuous functions. If $d\nu = g \, d\mu$ with g continuous and strictly positive, then the quantity in the $\sup$ when $f = g$ is H and Jensen's inequality implies the $\sup$ is always great than H. Some approximation arguments complete the proof of the variation principle. The variational principle says H is a $\sup$ of continuous linear functionals so

Lower Semicontinuity of the KL Divergence

An important property of the KL divergence is lower semicontinuity. One proves the following variational principle

$$H(\nu \mid \mu) = \sup_f \left(-\int f \, d\mu(x) + \int [1 + \log(f(x))] \, d\nu(x) \right)$$

where the $\sup$ is taken over all strictly positive continuous functions. If $d\nu = g \, d\mu$ with g continuous and strictly positive, then the quantity in the $\sup$ when $f = g$ is H and Jensen's inequality implies the $\sup$ is always great than H. Some approximation arguments complete the proof of the variation principle. The variational principle says H is a $\sup$ of continuous linear functionals so

$H(\nu \mid \mu)$ is jointly convex and jointly lower semicontinuous in μ and ν

Lower Bound on Jacobi Sum

We defined J_n as what one gets by striping off the first n Jacobi pairs, i.e. subtracting from the left. Complementary is $J^{(n)}$ which builds up by adding on the right, i.e. it has Jacobi parameters:

Lower Bound on Jacobi Sum

We defined J_n as what one gets by striping off the first n Jacobi pairs, i.e. subtracting from the left. Complementary is $J^{(n)}$ which builds up by adding on the right, i.e. it has Jacobi parameters:

$$a_k^{(n)}, \quad b_k^{(n)} = \begin{cases} a_k, \quad b_k, & \text{if } k = 1, \dots, n \\ 0, \quad 1, & \text{if } k \geq n \end{cases}$$

Lower Bound on Jacobi Sum

We defined J_n as what one gets by striping off the first n Jacobi pairs, i.e. subtracting from the left. Complementary is $J^{(n)}$ which builds up by adding on the right, i.e. it has Jacobi parameters:

$$a_k^{(n)}, \quad b_k^{(n)} = \begin{cases} a_k, & b_k, & \text{if } k = 1, \ldots, n \\ 0, & 1, & \text{if } k \geq n \end{cases}$$

The point is that $\left[J^{(n)}\right]_n$ is the free Jacobi matrix, which has no eigenvalues outside $[-2, 2]$, no non–trivial Jacobi parameters and the the free m–function. By looking at the n–times iterated sum rule for $J^{(n)}$, we find that

Lower Bound on Jacobi Sum

We defined J_n as what one gets by striping off the first n Jacobi pairs, i.e. subtracting from the left. Complementary is $J^{(n)}$ which builds up by adding on the right, i.e. it has Jacobi parameters:

$$a_k^{(n)}, \quad b_k^{(n)} = \begin{cases} a_k, \quad b_k, & \text{if } k = 1, \dots, n \\ 0, \quad 1, & \text{if } k \geq n \end{cases}$$

The point is that $\left[J^{(n)}\right]_n$ is the free Jacobi matrix, which has no eigenvalues outside $[-2, 2]$, no non–trivial Jacobi parameters and the the free m–function. By looking at the n–times iterated sum rule for $J^{(n)}$, we find that

$$\sum_{j=1}^{n} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] = Q(J^{(n)}) + \sum_{j,\pm} F(E_j^{\pm}(J^{(n)}))$$

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \ldots, K$ for K fixed.

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive).

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \ldots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$.

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$. For each fixed j, one can prove that $E_j^{\pm}(J^{(n)})$ converges to $E_j^{\pm}(J)$ so the sums up to K converge.

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \ldots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$. For each fixed j, one can prove that $E_j^{\pm}(J^{(n)})$ converges to $E_j^{\pm}(J)$ so the sums up to K converge. Then taking $K \to \infty$, we get that

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$. For each fixed j, one can prove that $E_j^{\pm}(J^{(n)})$ converges to $E_j^{\pm}(J)$ so the sums up to K converge. Then taking $K \to \infty$, we get that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] \geq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$. For each fixed j, one can prove that $E_j^{\pm}(J^{(n)})$ converges to $E_j^{\pm}(J)$ so the sums up to K converge. Then taking $K \to \infty$, we get that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] \geq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

This complements the upper bound and proves the full KS sum rule.

Lower Bound on Jacobi Sum

We get a lower bound on these equal term by replacing the full eigenvalue sum which might have more and more terms as n increases by the sum for $j = 1, \dots, K$ for K fixed. Having done that, take $n \to \infty$. The left side converges to the infinite sum (which might be infinite but the summands are positive). One can show the spectral measures for $J^{(n)}$ converge weakly to that of J so, by lower semincontinity, $Q(J)$ is smaller than the $\liminf$. For each fixed j, one can prove that $E_j^{\pm}(J^{(n)})$ converges to $E_j^{\pm}(J)$ so the sums up to K converge. Then taking $K \to \infty$, we get that

$$\sum_{j=1}^{\infty} \left[\tfrac{1}{4} b_j^2 + \tfrac{1}{2} G(a_j)\right] \geq Q(J) + \sum_{j,\pm} F(E_j^{\pm}(J))$$

This complements the upper bound and proves the full KS sum rule. Note the importance of the positivity of F and G in proving the lower bound.

Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

1. Why are there any positive combinations?

Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

1. Why are there any positive combinations?
2. It is easy to understand the $(4 - x^2)^{-1/2}\,dx$ of the Szegő condition. It is $d\theta$ under $x = \cos(\theta)$. Equivalently, it is the potential theoretic equilibrium measure for $[-2, 2]$

Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

1. Why are there any positive combinations?
2. It is easy to understand the $(4 - x^2)^{-1/2}\, dx$ of the Szegő condition. It is $d\theta$ under $x = \cos(\theta)$. Equivalently, it is the potential theoretic equilibrium measure for $[-2, 2]$ but where the heck does the $(4 - x^2)^{1/2}\, dx$ come from?

Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

1. Why are there any positive combinations?
2. It is easy to understand the $(4-x^2)^{-1/2}\,dx$ of the Szegő condition. It is $d\theta$ under $x = \cos(\theta)$. Equivalently, it is the potential theoretic equilibrium measure for $[-2,2]$ but where the heck does the $(4-x^2)^{1/2}\,dx$ come from?
3. What does the function
$$G(a) = a^2 - 1 - \log(a^2)$$
mean?

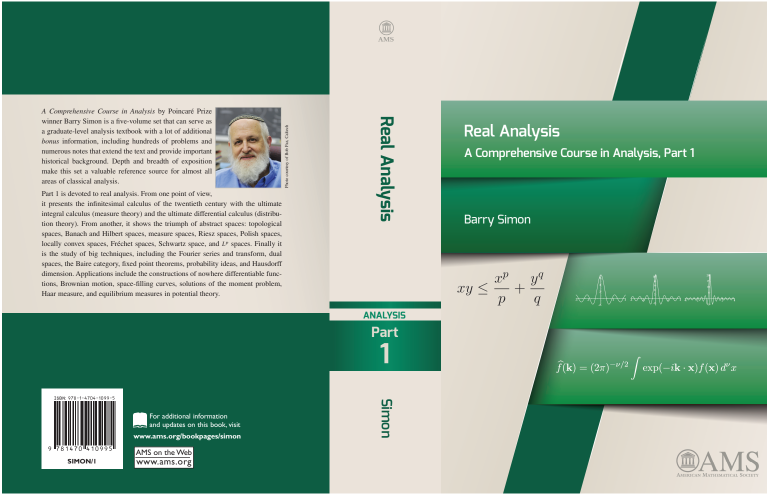
Mysteries

While the gem one gets from the P_2 sum rule is simple and elegant, the proof has lots of mysteries:

1. Why are there any positive combinations?
2. It is easy to understand the $(4-x^2)^{-1/2}\,dx$ of the Szegő condition. It is $d\theta$ under $x=\cos(\theta)$. Equivalently, it is the potential theoretic equilibrium measure for $[-2,2]$ but where the heck does the $(4-x^2)^{1/2}\,dx$ come from?
3. What does the function
$$G(a)=a^2-1-\log(a^2)$$
mean?
4. What does the function
$$F(E)=\tfrac{1}{4}[\beta^2-\beta^{-2}-\log\beta^4];\quad E=\beta+\beta^{-1}$$
mean?

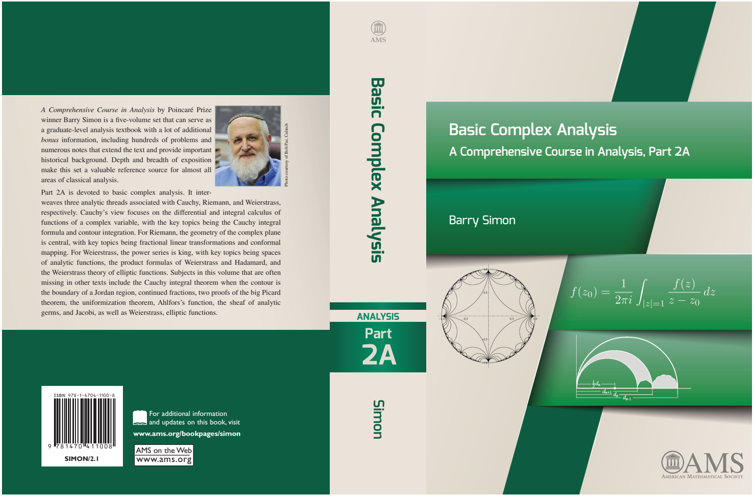
And Now a Word from Our Sponsor

And Now a Word from Our Sponsor



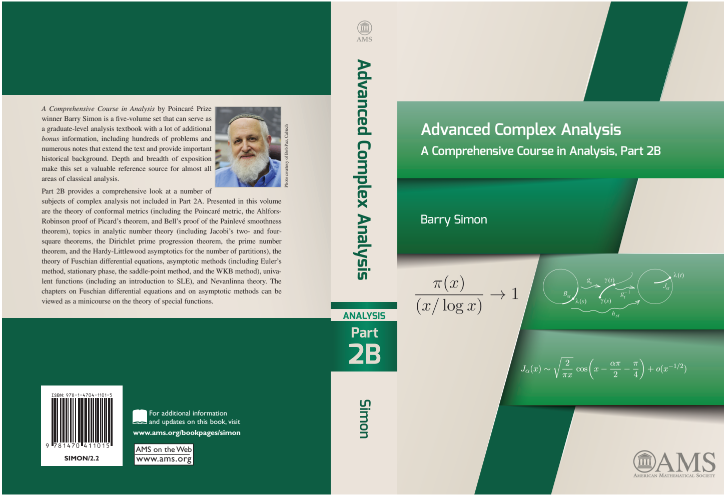
Google *simon comprehensive course*

And Now a Word from Our Sponsor



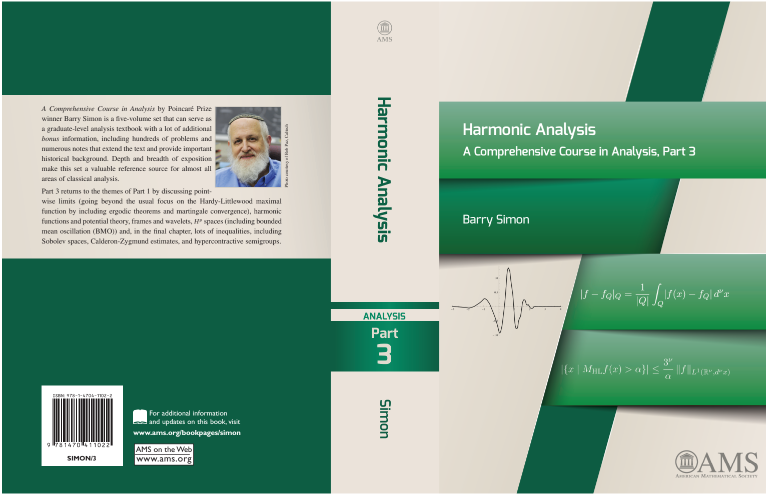
Google *simon comprehensive course*

And Now a Word from Our Sponsor



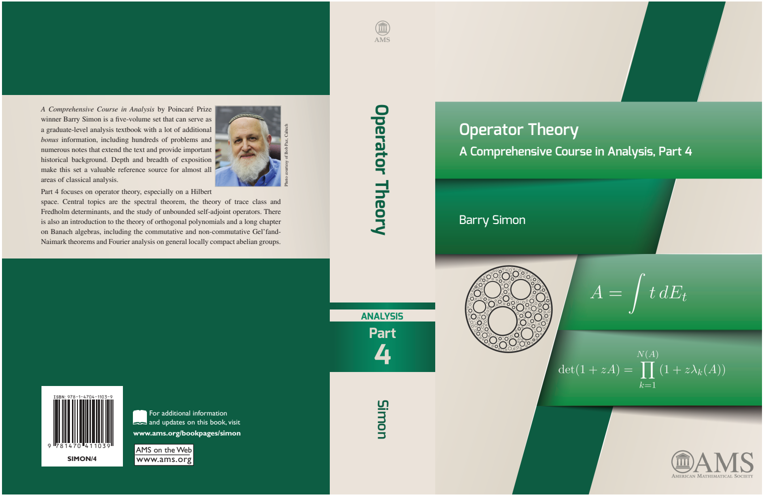
Google *simon comprehensive course*

And Now a Word from Our Sponsor



Google *simon comprehensive course*

And Now a Word from Our Sponsor



Google *simon comprehensive course*