Distribution testing in the 21 $^{1/2}$th century

Ryan O'Donnell
Carnegie Mellon University

based on joint work with **Costin Bădescu** (CMU) **& John Wright** (MIT)

Slide 1, in which I get defensive

Quantum. Why should you care?

Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

2. You get to do it all again

3. The math is (even more) beautiful

Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

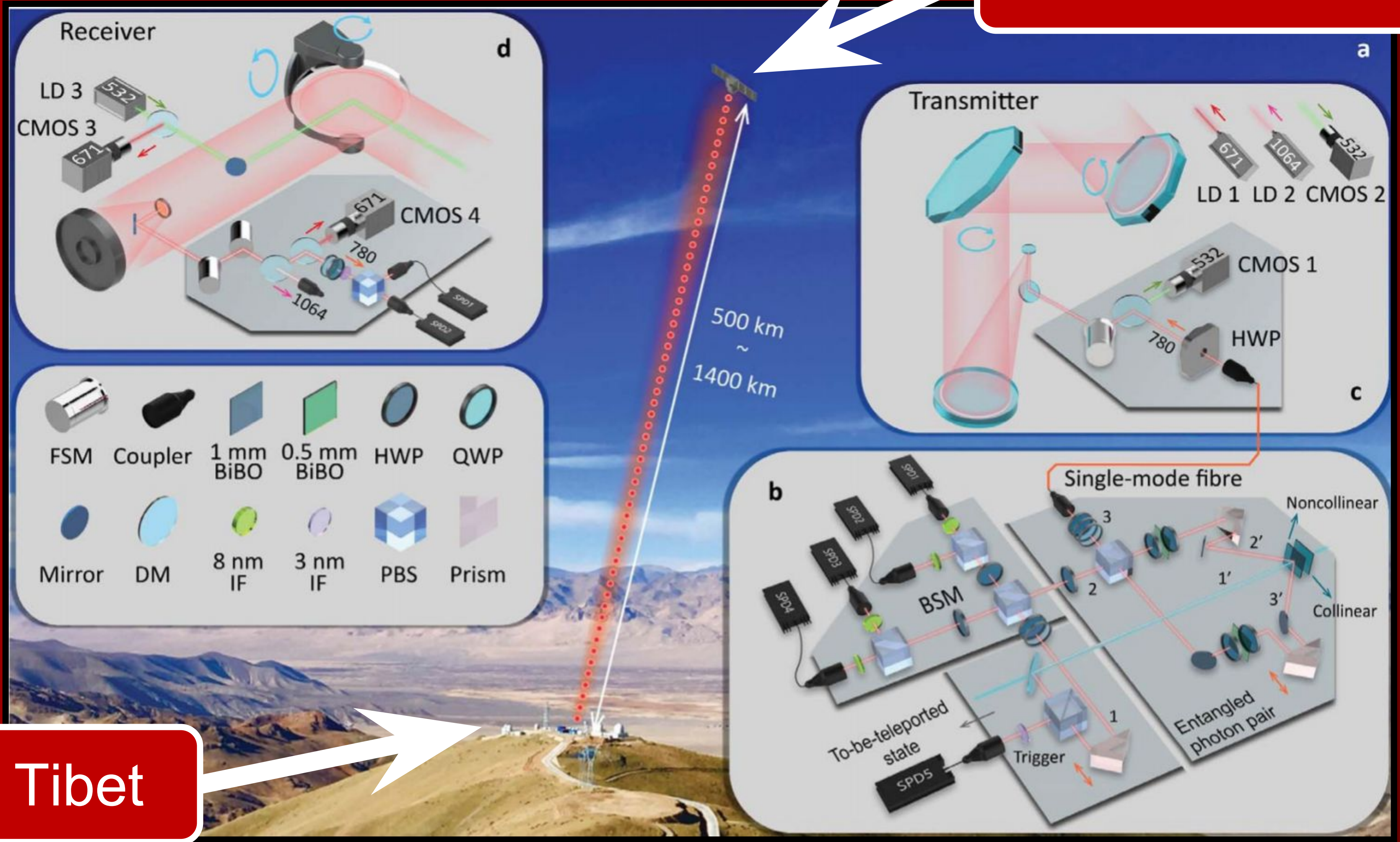
2. You get to do it all again

3. The math is (even more) beautiful

Quantum teleportation, July 2017

Jian-Wei Pan et al.

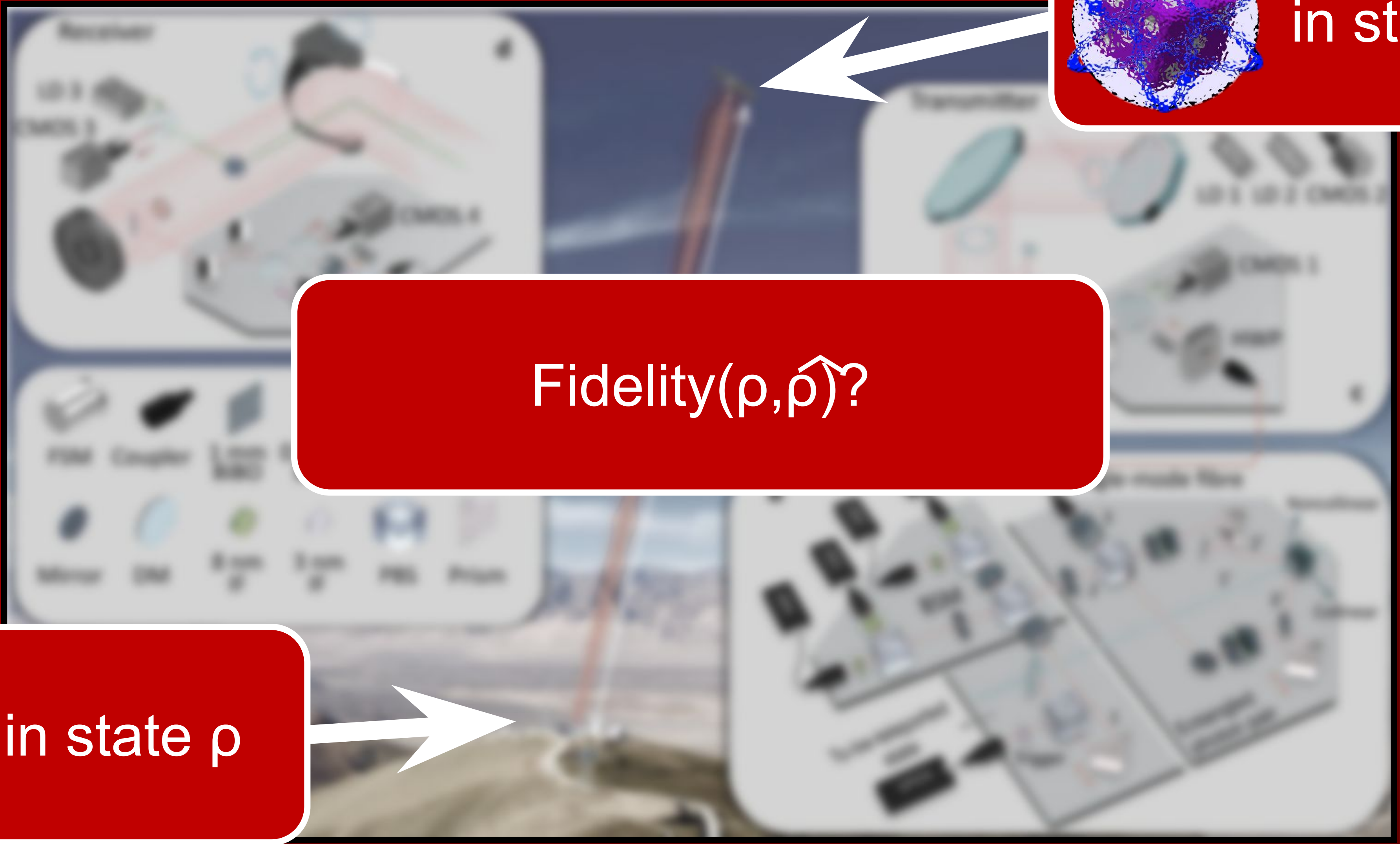
Micius satellite, space



Ngari, Tibet

Quantum teleportation, July 2017

Jian-Wei Pan et al.



in state $\hat{\rho}$

Fidelity$(\rho,\hat{\rho})$?

in state ρ

Quantum teleportation, July 2017

Jian-Wei Pan et al.

in state $\hat{\rho}$

in state ρ

(quantum-) Hellinger^2-Dist(ρ,ρ) $\widehat{=}$?

Quantum teleportation, July 2017

Jian-Wei Pan et al.

... n state $\rho^{\otimes 911}$

(quantum-) Hellinger^2-Dist$(\rho,\rho) \triangleq$?

in state ρ

Quantum teleportation, July 2017

Jian-Wei Pan et al.

… n state $\rho^{\otimes 911}$

(quantum-) Hellinger^2-Dist(ρ,ρ) $\widehat{=}$ 0.21 $\pm.01$

in state ρ

Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

2. You get to do it all again

3. The math is (even more) beautiful

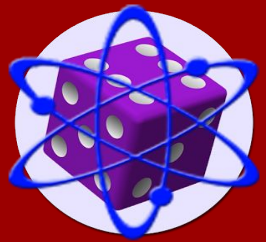
Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

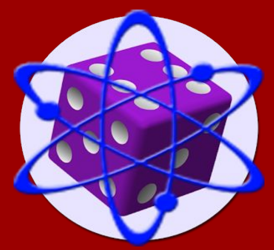
2. You get to do it all again

3. The math is (even more) beautiful

What is classical Probability Density Testing?



Unknown n-outcome
source of randomness ρ



Unknown n-outcome source of randomness p

m samples

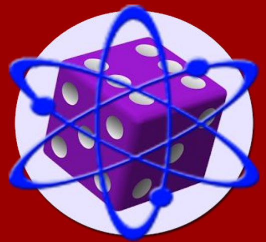
$p^{\otimes m}$

$x \in \mathbb{R}$

Maybe x $\in$ {0,1} is a guess as to whether p is uniformly random.

Maybe x is an estimate of Dist(q,p) for some hypothesis q.

Maybe x is an estimate of Entropy(p).



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

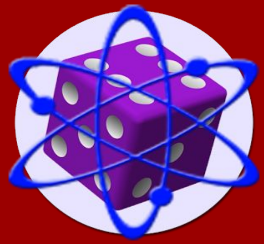
Example:

You have a hope that $p \equiv \mathbf{1}/n$, the uniform distribution.

You want to estimate Dist($\mathbf{1}/n$, p),

where "Dist" ∈ {TV, Hellinger^2, Chi-Squared, , …} ℓ_2^2

Latter two are the same here, so let's choose them.



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

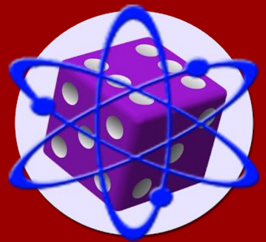
Example:

You have a hope that $p \equiv 1/n$, the uniform distribution.

You want to estimate ℓ_2^2-Dist(1/n, p)

$$= \sum_{i=1}^{n} (p_i - 1/n)^2$$

$$= \sum_{i=1}^{n} p_i^2 - 1/n$$



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

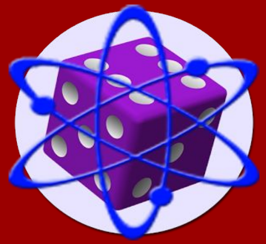
You basically want to estimate $\sum_{i=1}^{n} p_i^2$ (the "collision probability")

Say $m = 2$. What should Algorithm X be?

Algorithm X: Given sample $(a,b) \sim p^{\otimes 2}$, output $\begin{cases} 1 & \text{if } a = b, \\ 0 & \text{else.} \end{cases}$

$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$ 

Var$[X]$ = large 



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

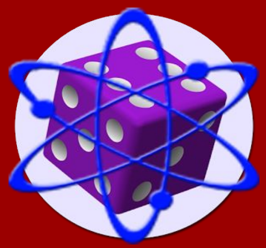
You basically want to estimate $\sum_{i=1}^{n} p_i^2$

Say m > 2. What should Algorithm X be?

Algorithm X: Average the m=2 algorithm over all $\binom{m}{2}$ pairs.

$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$ 

Var[X] = (tedious but straightforward)



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

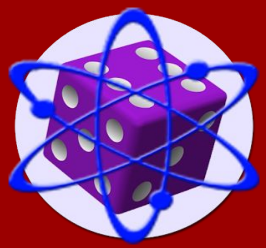
$$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$$

Var[X] = (tedious but straightforward)

Chebyshev ⇒ $m = O(\sqrt{n}/\epsilon^2)$ samples suffice to distinguish

ℓ_2-Dist(**1**/n, p) ≤ $.99\epsilon^2/n$

vs. ℓ_2-Dist(**1**/n, p) ≥ ϵ^2/n whp.



Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

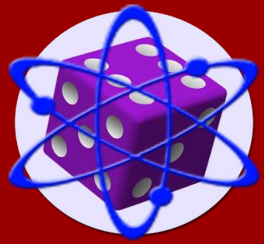
$$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$$

Var[X] = (tedious but straightforward)

Chebyshev $\Rightarrow$ $m = O(\sqrt{n}/\epsilon^2)$ samples suffice to distinguish

ℓ_2-Dist($\mathbf{1}$/n, p) $\leq$ .99ϵ^2/n

vs. TV-Dist($\mathbf{1}$/n, p) $\geq \epsilon$ whp.



Unknown n-outcome source of randomness p

m samples

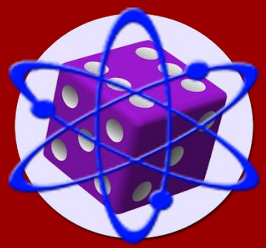
$p^{\otimes m}$

$x \in \mathbb{R}$

Remember two things:

1. The algorithm: Average, over all transpositions $\tau \in S_m$, of 0/1 indicator that τ leaves samples unchanged

2. Any alg. is just a random variable, based on randomness $p^{\otimes m}$



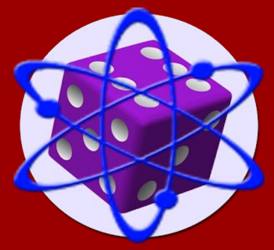
Unknown n-outcome source of randomness p

m samples

$p^{\otimes m}$

$x \in \mathbb{R}$

Classical probability density testing picture, $m=1$:



sample →

$x \in \mathbb{R}$

Unknown n-outcome source of randomness p

Classical probability density testing picture, m=1:

p is an n-dimensional vector $p \geq 0$, $\sum_{i=1}^{n} p_i = 1$

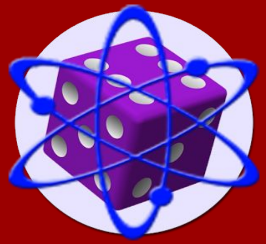
X is an n-dimensional vector

$\mathbf{E}_p[X] = \langle p, X \rangle = \sum_i p_i X_i$

$\mathbf{E}_p[X^2] = \langle p, X^2 \rangle$ "=" $\sum_i p_i X_i^2$

Changing the picture: Classical → Quantum

Replace “vector” with “symmetric matrix” everywhere.
(Hermitian)



Unknown n-outcome source of randomness p

sample $\longrightarrow$

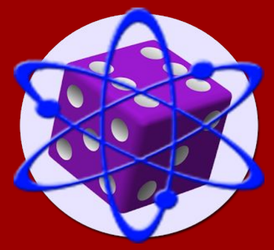
$x \in \mathbb{R}$

Classical probability density testing picture, m=1:

p is an n-dimensional vector $p \geq 0$, $\sum_{i=1}^{n} p_i = 1$

X is an n-dimensional vector

$$\mathbf{E}_p[X] = \langle p, X \rangle = \sum_i p_i X_i \qquad \mathbf{E}_p[X^2] = \langle p, \text{“}X^2\text{”} \rangle = \sum_i p_i X_i^2$$



sample

$x \in \mathbb{R}$

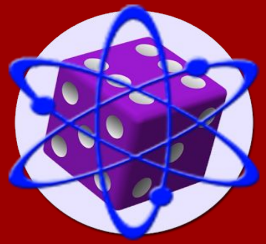
Unknown n-outcome source of randomness ρ

Quantum probability density testing picture, m=1:

ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii} = 1$

X is an n-dim. symm. matrix

$\mathbf{E}_{\rho}[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$ $\mathbf{E}_{\rho}[X^2] = \langle \rho, X^2 \rangle = \sum_{ij} \rho_{ij} (X^2)_{ij}$



Unknown n-outcome source of randomness ρ

m samples

$\rho^{\otimes m}$

$x \in \mathbb{R}$

Quantum probability density testing picture:

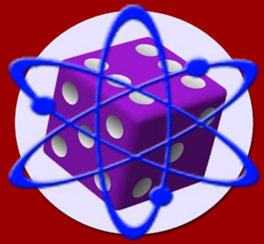
ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii} = 1$

X is an n-dim. symm. matrix

$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$ $\mathbf{E}_\rho[X^2] = \langle \rho, X^2 \rangle = \sum_{ij} \rho_{ij} (X^2)_{ij}$

Changing the picture: Quantum $\rightarrow$ Classical

Let ρ and X be diagonal matrices.



Unknown n-outcome source of randomness ρ

m samples

$\rho^{\otimes m}$

$x \in \mathbb{R}$

Quantum probability density testing picture:

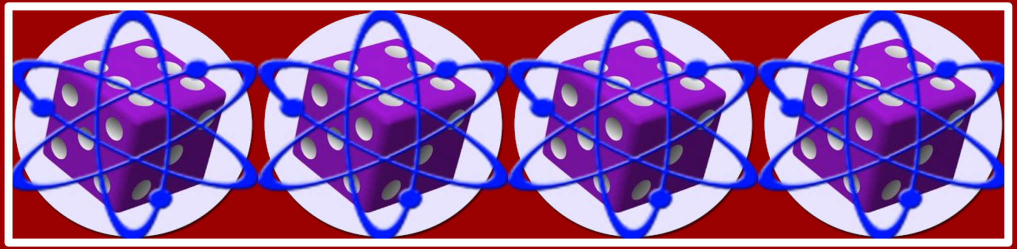
ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii} = 1$

X is an n-dim. symm. matrix

$$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij} \qquad \mathbf{E}_\rho[X^2] = \langle \rho, X^2 \rangle = \sum_{ij} \rho_{ij} (X^2)_{ij}$$

What's going on, physically?

ρ = "state" of a particle-system

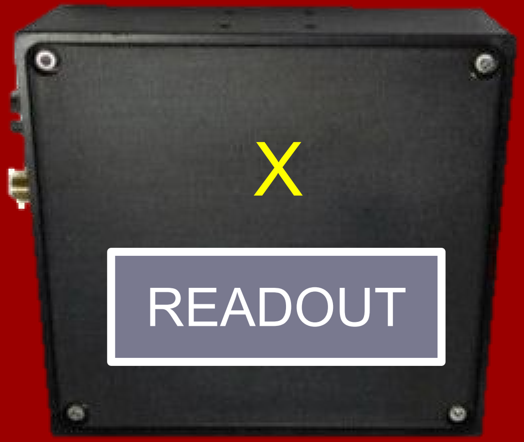


e.g., 4 polarized photons

n = # "basic outcomes"; 2 for a "qubit", 16 for 4 photons

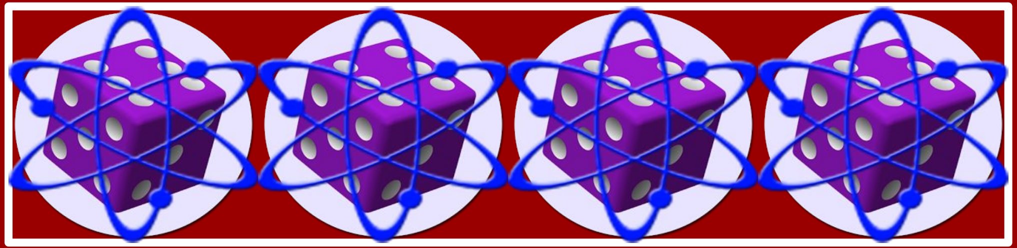
X = "observable" = measuring device

(quantum circuit)



What's going on, physically?

ρ = "state" of a particle-system



e.g., 4 polarized photons

n = # "basic outcomes"; 2 for a "qubit", 16 for 4 photons

X = "observable" = measuring device

(quantum circuit)

What's going on, physically?

ρ = "state" of a particle-system

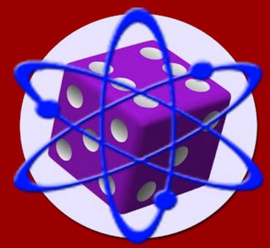
Don't read this:

Readout is λ_i with probability $\langle \phi_i, \rho\phi_i \rangle$

where (λ_i, ϕ_i) are the eigvals/vecs of X.

X = "observable" = measuring device

(quantum circuit)



Unknown n-outcome source of randomness ρ

sample

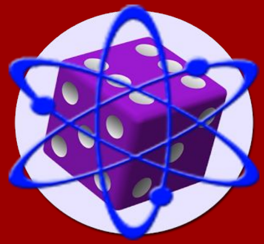
$x \in \mathbb{R}$

Quantum probability density testing picture:

ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii} = 1$

X is an n-dim. symm. matrix

$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$ $\mathbf{E}_\rho[X^2] = \langle \rho, X^2 \rangle = \sum_{ij} \rho_{ij} (X^2)_{ij}$



sample

$x \in \mathbb{R}$

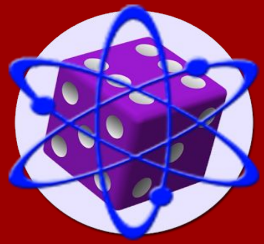
Unknown n-outcome source of randomness ρ

<u>Baseline</u>: **Learning ρ** (“quantum tomography”)

ρ is an n-dim. symm. matrix

X is an n-dim. symm. matrix

$$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$$



Unknown n-outcome source of randomness ρ

sample $\longrightarrow$

$x \in \mathbb{R}$

Baseline: **Learning ρ**

ρ is an n-dim. symm. matrix

X is an n-dim. symm. matrix

$$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$$

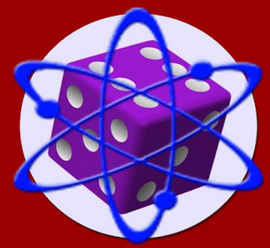
Naive method:

Use X's like

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & .5 & 0 & 0 \\ .5 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

to learn each ρ_{ij} separately.

$O(n^4)$ samples.



sample $\longrightarrow$

$x \in \mathbb{R}$

Unknown n-outcome source of randomness ρ

<u>Baseline</u>: **Learning ρ**

ρ is an n-dim. symm. matrix

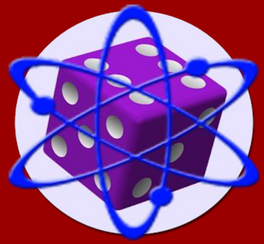
X is an n-dim. symm. matrix

$$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$$

Theorem: [HHJWY15,OW15]

It is necessary & sufficient to have $m = \Theta(n^2/\epsilon^2)$ samples to learn ρ to ϵ-accuracy in "trace (ℓ_1) distance".

A better way to think about the scenario



Unknown n-outcome source of randomness ρ

m samples

$\rho^{\otimes m}$

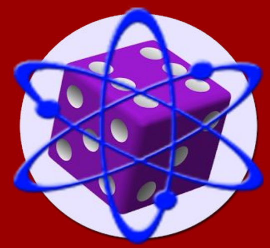
$x \in \mathbb{R}$

Quantum probability density testing picture:

ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii} = 1$

X is an n-dim. symm. matrix

$\mathbf{E}_\rho[X] = \langle \rho, X \rangle = \sum_{ij} \rho_{ij} X_{ij}$ $\mathbf{E}_\rho[X^2] = \langle \rho, X^2 \rangle = \sum_{ij} \rho_{ij} (X^2)_{ij}$



m samples

$\rho^{\otimes m}$

$x \in \mathbb{R}$

Unknown n-outcome
source of randomness ρ

Quantum probability density testing picture:

ρ is an n-dim. symm. matrix, $\rho \geq 0$, $\sum_{i=1}^{n} \rho_{ii}$

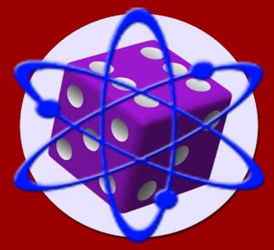
ρ is PSD

$\Leftrightarrow$ ρ's eigenvalues are ≥ 0

trace(ρ) = 1

$\Leftrightarrow$ ρ's eigenvalues sum to 1

$\Leftrightarrow$ ρ's eigenvalues form a probability distribution! Call it $p_1, \ldots, p_n$



Unknown n-outcome source of randomness ρ

m samples

$\rho^{\otimes m}$

$x \in \mathbb{R}$

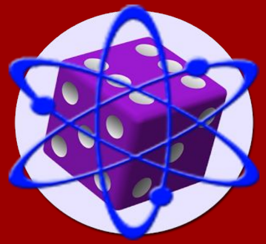
Quantum probability density testing picture:

ρ's eigenvalues form a probability distribution! Call it $p_1, \ldots, p_n$

"Over" ρ's orthonormal eigenvectors in $\mathbb{C}^n$. Call them $v_1, \ldots, v_n$

Think of ρ as emitting v_i with probability p_i.

<u>Exercise</u>: Conditioned on v_i, $\mathbf{E}[X] = \langle X v_i, v_i \rangle$.



Unknown n-outcome source of randomness ρ

m samples

$\rho^{\otimes m}$

$x \in \mathbb{R}$

Quantum probability density testing picture:

ρ's eigenvalues form a probability distribution! Call it $p_1, \ldots, p_n$

"Over" ρ's orthonormal eigenvectors in $\mathbb{C}^n$. Call them $v_1, \ldots, v_n$

Think of ρ as emitting v_i with probability p_i.

Also: $\rho^{\otimes 5}$ emits $v_3 \otimes v_1 \otimes v_4 \otimes v_1 \otimes v_n$ with probability $p_3 \cdot p_1 \cdot p_4 \cdot p_1 \cdot p_n \ldots$

Classical World	Quantum World
$(p_1, \ldots, p_n)$	no analogue
$\{p_1, \ldots, p_n\}$	$\{p_1, \ldots, p_n\}$
no analogue	$v_1, \ldots, v_n$
support-size(p) = # nonzero p_i's	rank(ρ) = # nonzero p_i's
Entropy(p) = $\sum_i p_i \log(1/p_i)$	vN-Entropy(ρ) = $\sum_i p_i \log(1/p_i)$
uniform distribution, $p = 1/n$	maximally mixed state, $\rho = I/n$
total variation distance	trace distance
Hellinger^2 distance	infidelity
ℓ_2^2 -distance	Frobenius^2-distance

Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

2. You get to do it all again

3. The math is (even more) beautiful

Quantum Distribution Testing: Why care?

1. Practically relevant problems
   at the vanguard of computing

2. You get to do it all again

3. The math is (even more) beautiful

Estimating ℓ_2^2-distance between ρ and $\mathbb{I}/n$

(AKA testing if ρ is the maximally mixed state)

[Bădescu-O-Wright'17]

$$\|\mathbb{I}/n - \rho\|_F^2 = \sum_{i=1}^{n} p_i^2 - 1/n$$

(Same as in classical case!)

Estimating ℓ_2^2 -distance between ρ and I/n

You basically want to estimate $\sum_{i=1}^{n} p_i^2$ (the "quantum purity")

Say $m = 2$. What should Algorithm X be?

ρ emits $v_a \otimes v_b \in (\mathbb{C}^n)^{\otimes 2}$ with probability $p_a \cdot p_b$

The "algorithm" should be an operator X on $(\mathbb{C}^n)^{\otimes 2}$

Doesn't know $v_1, \ldots, v_n$, but does understand "tensor structure"

Let X act by swapping tensor components: $X\, e_a \otimes e_b = e_b \otimes e_a$

$\Rightarrow X\, v_a \otimes v_b = v_b \otimes v_a$

Estimating ℓ_2^2-distance between ρ and $\mathbb{I}/n$

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

Say $m = 2$. What should Algorithm X be?

Let X act by swapping tensor components: $X\, e_a \otimes e_b = e_b \otimes e_a$

$\Rightarrow X\, v_a \otimes v_b = v_b \otimes v_a$

Conditioned on ρ emitting $v_a \otimes v_b$, $\mathbf{E}[X] = \langle X\, v_a \otimes v_b,\, v_a \otimes v_b \rangle$

$$= \langle v_b \otimes v_a,\, v_a \otimes v_b \rangle$$

$$= \begin{cases} 1 & \text{if } a = b, \\ 0 & \text{else.} \end{cases}$$

$\therefore \mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$ 😄

Estimating ℓ_2^2-distance between ρ and $\mathbb{I}/n$

Let X act by swapping tensor components: $X\, e_a \otimes e_b = e_b \otimes e_a$

$\Rightarrow X\, v_a \otimes v_b = v_b \otimes v_a$

$$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$$

$$\mathbf{Var}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

Conditioned on ρ emitting $v_a \otimes v_b$, $\mathbf{E}[X^2] = \langle X^2\, v_a \otimes v_b,\, v_a \otimes v_b \rangle$

$$= \langle v_a \otimes v_b,\, v_a \otimes v_b \rangle$$

$$= 1$$

$\therefore \mathbf{Var}[X]$ = extra large

Estimating ℓ_2^2-distance between ρ and I/n

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

Say $m > 2$. What should Algorithm X be?

X = avg { $R(\tau)$: transpositions $\tau \in S_m$ }

where, in general, $R(\pi)$ acts on $(\mathbb{C}^n)^{\otimes m}$ by permuting tensor components according to π

$\mathbf{E}[X] = \sum_{i=1}^{n} p_i^2$  For **Var**[X], need to compute X^2.

Estimating ℓ_2^2 -distance between ρ and $\mathbb{I}/n$

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

X = avg { $R(\tau)$: transpositions $\tau \in S_m$ }

X^2 = avg {$R(\sigma)\, R(\tau)$: transpositions $\sigma,\tau \in S_m$ }

= avg { $R(\sigma\tau)$: transpositions $\sigma,\tau \in S_m$ }

$R : S_m \to$ {Matrices acting on $(\mathbb{C}^n)^{\otimes m}$}

is a *group representation*!

Estimating ℓ_2^2 -distance between ρ and I/n

You basically want to estimate $\displaystyle\sum_{i=1}^{n} p_i^2$

$$X = \text{avg} \{ R(\tau) : \text{transpositions } \tau \in S_m \}$$

$$X^2 = \text{avg} \{R(\sigma)\, R(\tau) : \text{transpositions } \sigma,\tau \in S_m \}$$

$$= \text{avg} \{ \quad R(\sigma\tau) \quad : \text{transpositions } \sigma,\tau \in S_m \}$$

$$\begin{aligned} = \quad & c_1 \text{ avg} \{ R(\pi) : \text{cycleType}(\pi) = (1) \} \\ & + c_2 \text{ avg} \{ R(\pi) : \text{cycleType}(\pi) = (2,2) \} \\ & + c_3 \text{ avg} \{ R(\pi) : \text{cycleType}(\pi) = (3) \} \end{aligned}$$

for some straightforward but slightly annoying

to compute coefficients c_1, c_2, c_3

Estimating ℓ_2^2 -distance between ρ and I/n

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

$$\mathbf{E}[X^2] = c_1 \text{ avg } \{ R(\pi) : \text{cycleType}(\pi) = (1) \}$$
$$+ c_2 \text{ avg } \{ R(\pi) : \text{cycleType}(\pi) = (2,2) \}$$
$$+ c_3 \text{ avg } \{ R(\pi) : \text{cycleType}(\pi) = (3) \}$$

for some straightforward but slightly annoying

to compute coefficients c_1, c_2, c_3

<u>Exercise</u>: Let $A_{(7,4,2)} = \text{avg } \{ R(\pi) : \text{cycleType}(\pi) = (7,4,2) \}$.

Then $\mathbf{E}[A_{(7,4,2)}] = \left(\sum_{i=1}^{n} p_i^7\right)\left(\sum_{i=1}^{n} p_i^4\right)\left(\sum_{i=1}^{n} p_i^2\right)$

Estimating ℓ_2^2-distance between ρ and I/n

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

Can now compute the variance,
bound it asymptotically,
use Chebyshev, etc. etc.,
just like in the classical case.

(It's actually somewhat cleaner.)

Exercise: Let $A_{(7,4,2)}$ = avg { $R(\pi)$: cycleType(π) = $(7,4,2)$ }.

Then $\mathbf{E}[A_{(7,4,2)}] = \left(\sum_{i=1}^{n} p_i^7\right)\left(\sum_{i=1}^{n} p_i^4\right)\left(\sum_{i=1}^{n} p_i^2\right)$

Estimating ℓ_2^2-distance between ρ and $\mathbf{I}$/n

Long story short:

m = $O(n/\epsilon^2)$ samples suffice to distinguish

ℓ_2^2-Dist($\mathbf{I}$/n, ρ) ≤ $.99\epsilon^2/n$

vs. ℓ_1-Dist($\mathbf{I}$/n, ρ) ≥ ϵ whp.

Also in [Bădescu-O-Wright'17]:

Same for distinguishing closeness of two unknown ρ, q

[O-Wright'15]: m = $\Omega(n/\epsilon^2)$ samples needed for testing "ρ = $\mathbf{I}$/n"

Estimating ℓ_2^2 -distance between ρ and I/n

You basically want to estimate $\sum_{i=1}^{n} p_i^2$

I promised you
"beautiful math"...

cleType(π) = (1) }

e(π) = (2,2) }

e(π) = (3) }

for some straightforward but slightly annoying

to compute coefficients c_1, c_2, c_3

Exercise: Let $A_{(7,4,2)}$ = avg { R(π) : cycleType(π) = (7,4,2) }.

Then $\mathbf{E}[A_{(7,4,2)}] = \left(\sum_{i=1}^{n} p_i^7\right)\left(\sum_{i=1}^{n} p_i^4\right)\left(\sum_{i=1}^{n} p_i^2\right)$

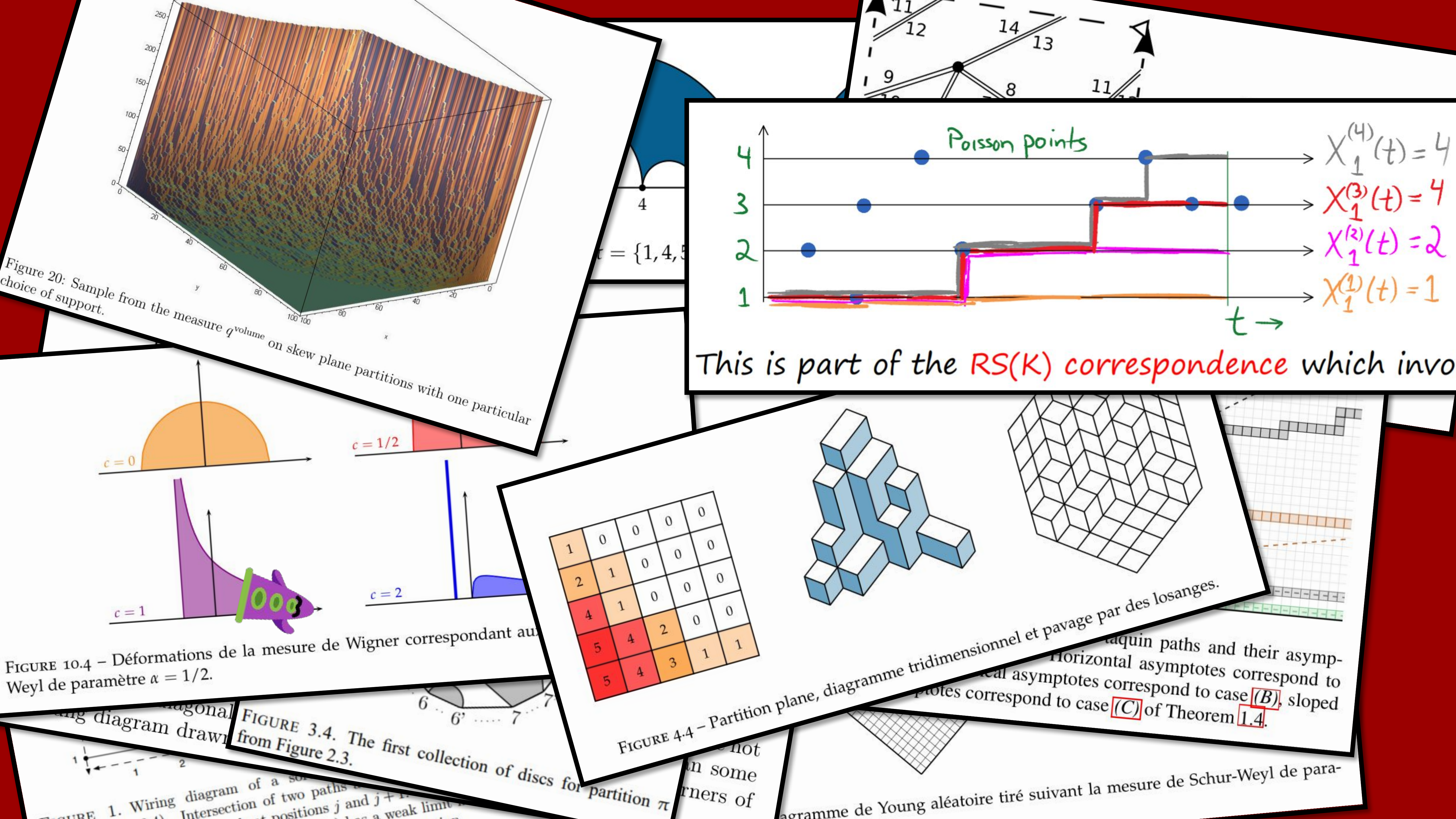


Figure 20: Sample from the measure q^{volume} on skew plane partitions with one particular choice of support.
$= \{1, 4, 5$
4
Poisson points
4
3
2
1
$X_1^{(4)}(t) = 4$
$X_1^{(3)}(t) = 4$
$X_1^{(2)}(t) = 2$
$X_1^{(1)}(t) = 1$
t →
This is part of the RS(K) correspondence which invo
$c = 0$
$c = 1/2$
$c = 1$
$c = 2$
FIGURE 10.4 – Déformations de la mesure de Wigner correspondant au
Weyl de paramètre $\alpha = 1/2$.
FIGURE 4.4 – Partition plane, diagramme tridimensionnel et pavage par des losanges.
aquin paths and their asymp-
Horizontal asymptotes correspond to
al asymptotes correspond to case (B), sloped
ptotes correspond to case (C) of Theorem 1.4.
FIGURE 3.4. The first collection of discs for partition π from Figure 2.3.
agramme de Young aléatoire tiré suivant la mesure de Schur-Weyl de para-

Free Probability!

Tracy-Widom Distributions!

Donald Knuth!

Geometric Complexity Theory!

Longest Increasing Subsequences!

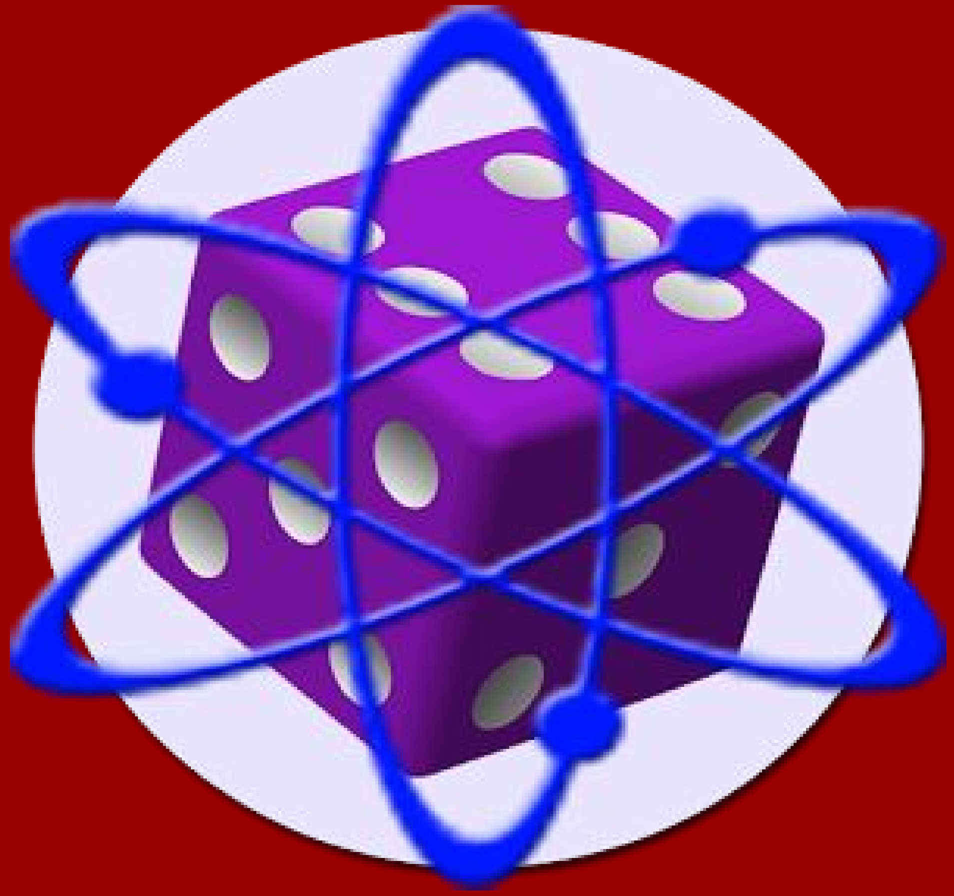
Sorting Networks!

Queueing Theory!

Traffic Models!

For more: www.cs.cmu.edu/~odonnell/papers/tomography-survey.pdf

Thanks!



For more: www.cs.cmu.edu/~odonnell/papers/tomography-survey.pdf