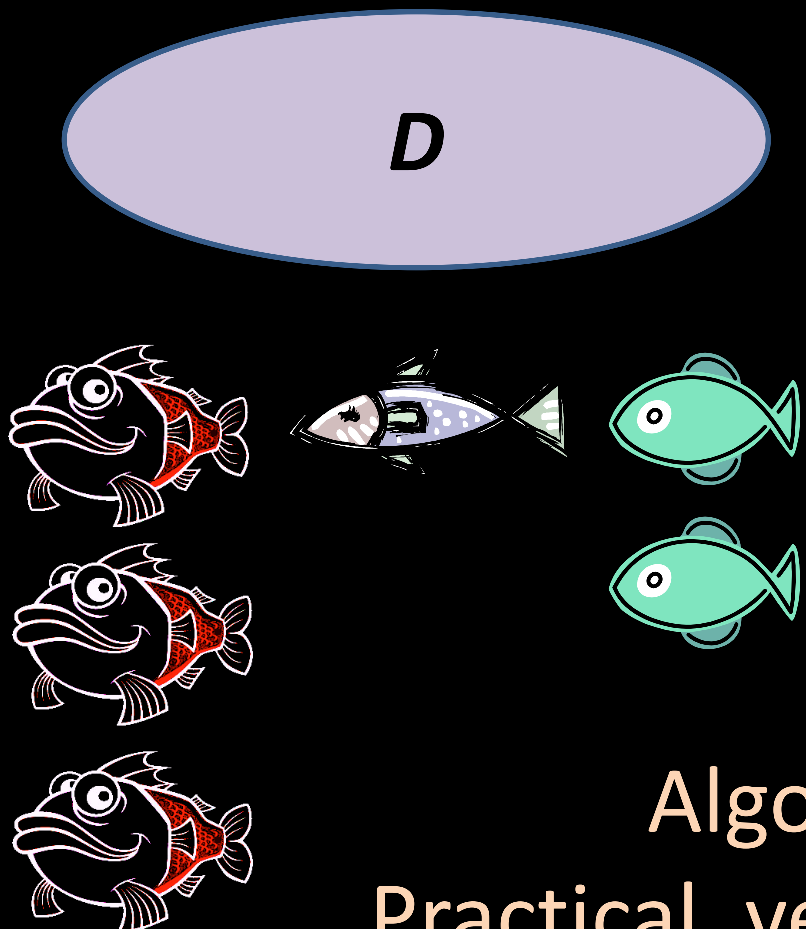
Lower Bound Techniques for Statistical Estimation

Gregory Valiant and Paul Valiant

The Setting

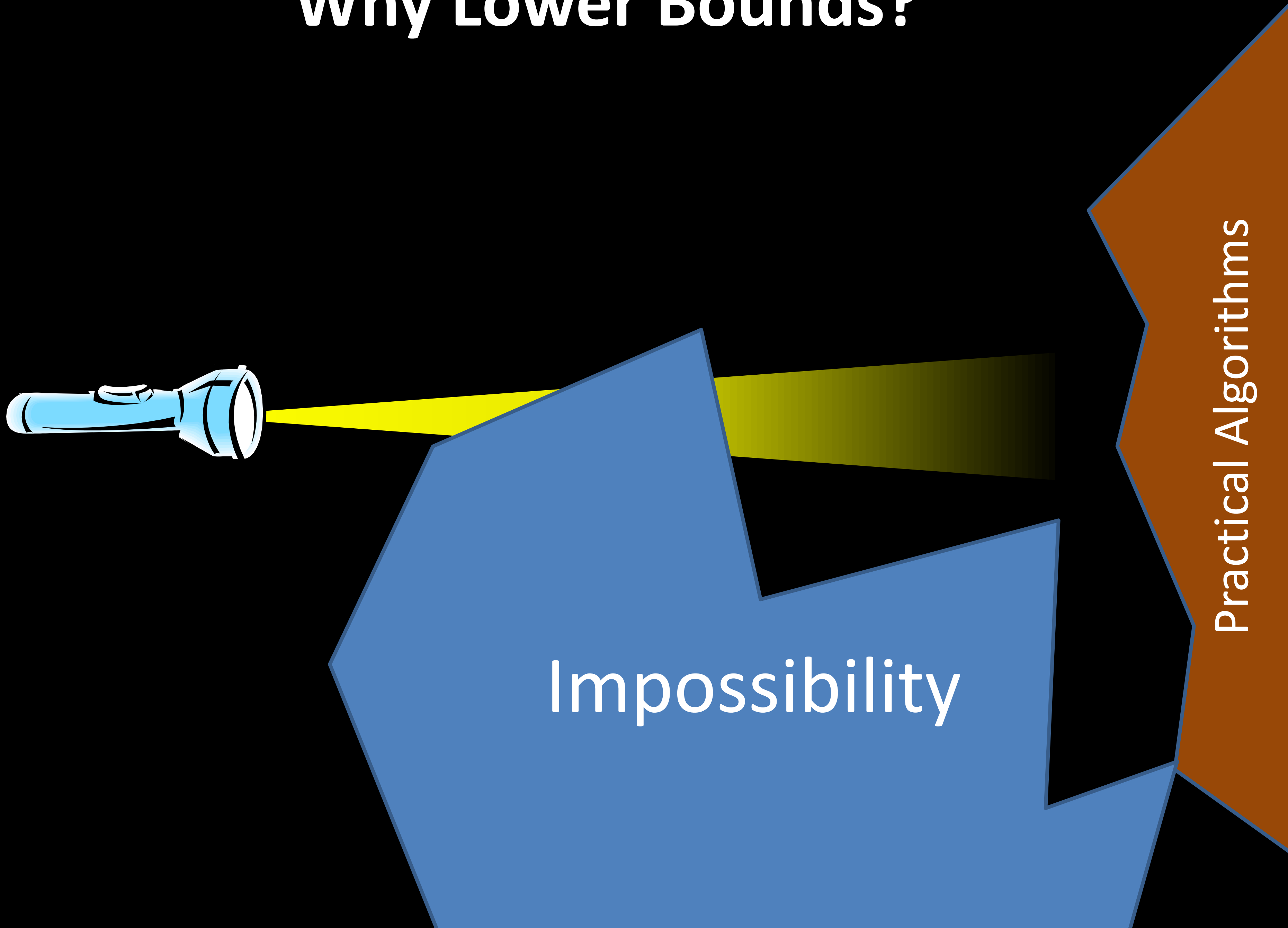
Given independent samples from a distribution (of discrete support):



Estimate # species
Estimate entropy
Hypothesis testing: =D'?
-- close to D'?
Compare 2 distributions
--many metrics

Algorithms:
Practical, yet unexpected!

Why Lower Bounds?



Birthday Paradox

Task: distinguish U_n from $U_{n/2}$ in k samples

Which n/2? Random

1-way	2-way	…
k	0	0

Birthday paradox: for $k = o(\sqrt{n})$, same fingerprints w.h.p.

(▼,▼,▼,▼,▼,▼,▼,▼,▼,▼)

data

▼	▼	▼	▼	▼	▼	▼	▼
2	2	0	0	2	1	2	1

histogram of data

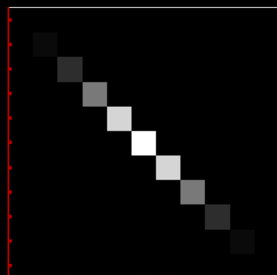
1-way	2-way	…
2	4	0

“fingerprint”
(histogram of histogram)

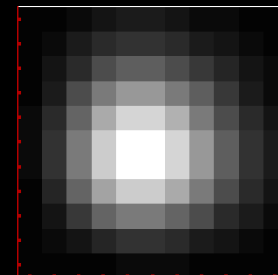
More Generally?

Characterize (and compare) fingerprint distributions

heads, # tails
given k=10 flips

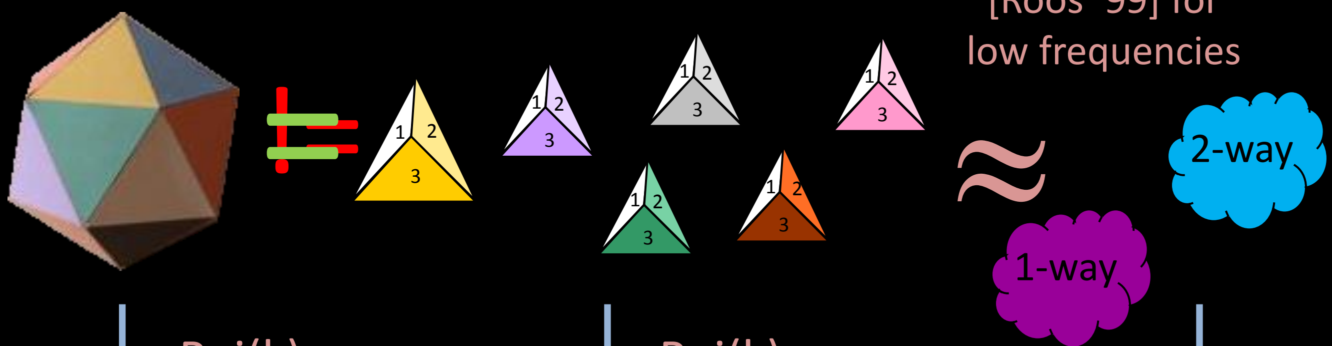


heads, # tails
given Poi(10) flips



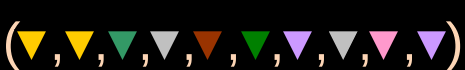
#occurrences of
different elements
are independent!!!

[Roos '99] for
low frequencies

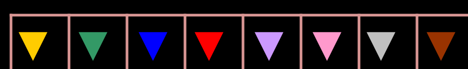


$\times$Poi(k)

$\times$Poi(k)



data



2	2	0	0	2	1	2	1

histogram of data

1-way	2-way	…
2	4	0

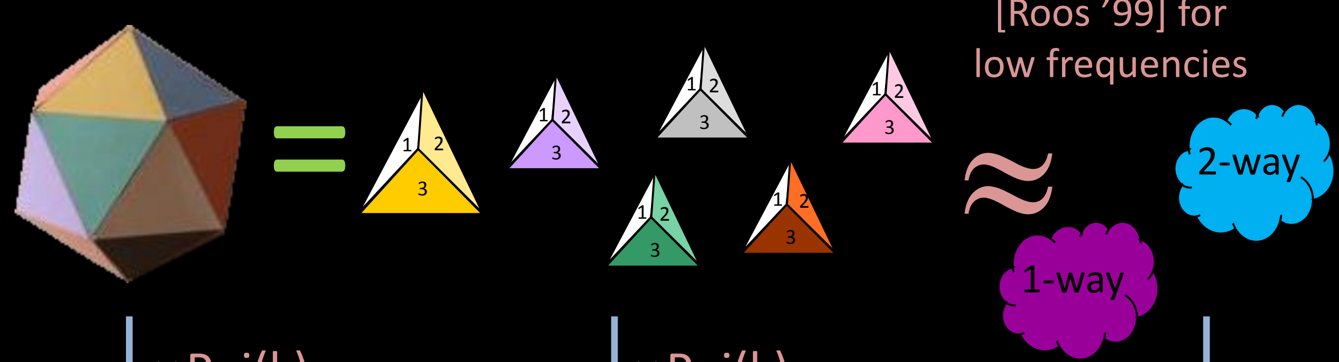
"fingerprint"
(histogram of histogram)

"Double Poissonization"

Introduced in [Raskhodnikova et al. '07], here as in [V '08]

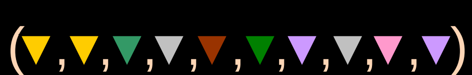
Characterize (and compare) fingerprint distributions

"Fingerprint entries are distributed (almost) like independent Poisson processes"



$\times$Poi(k)

$\times$Poi(k)



data

2	2	0	0	2	1	2	1

histogram of data

1-way	2-way	…
2	4	0

"fingerprint"
(histogram of histogram)

Theorem: Moments Describe All

(for low-weight distributions)

"Fingerprint entries are distributed (almost) like independent Poisson processes"

+

Poisson processes are characterized by their expectation

⬇

Fingerprint distributions are characterized by the expected number of 2-way collisions, 3-way collisions, etc.

$\approx \sum p_i^2$

$\approx \sum p_i^3$, etc.

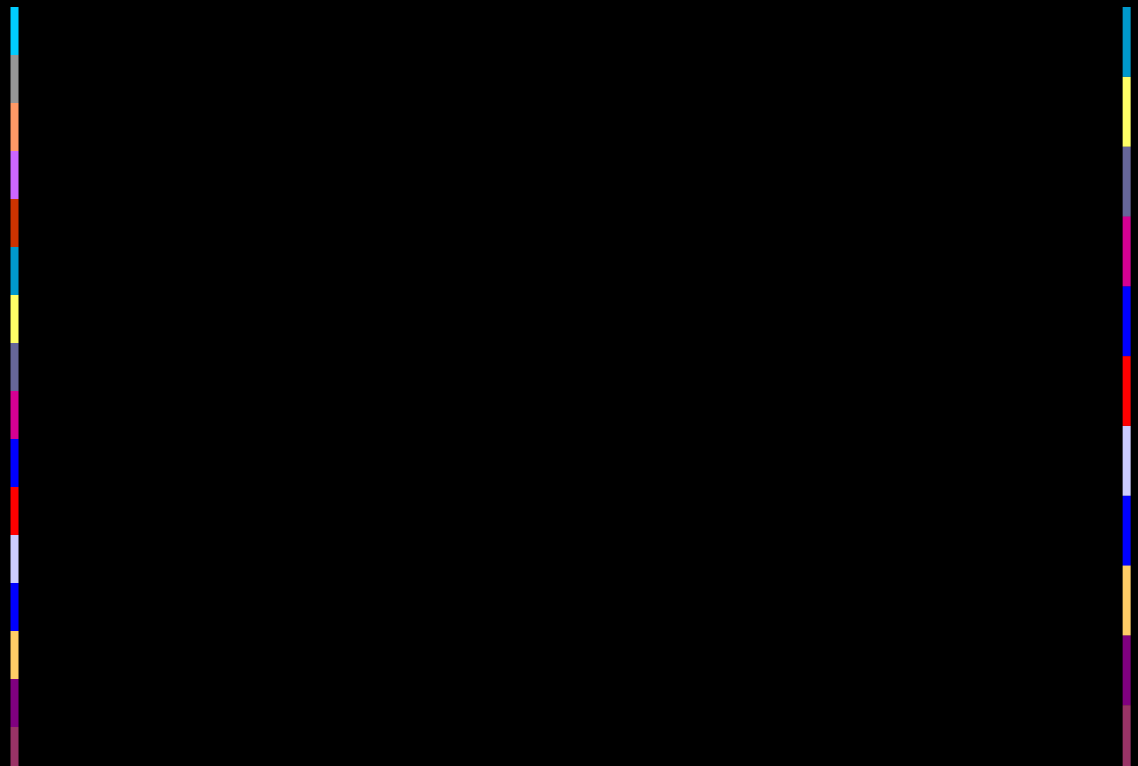
"moments"

Application:
Entropy lower bounds

(Theorem: Moments Don’t Describe Anything)

high entropy

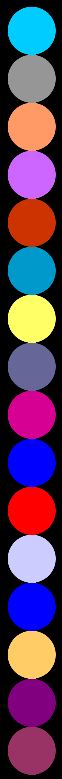
low entropy



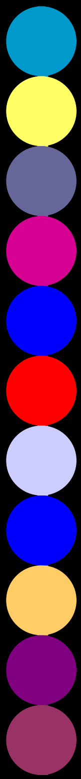
Application:
Entropy lower bounds

(Theorem: Moments Don’t Describe Anything)

high entropy



low entropy

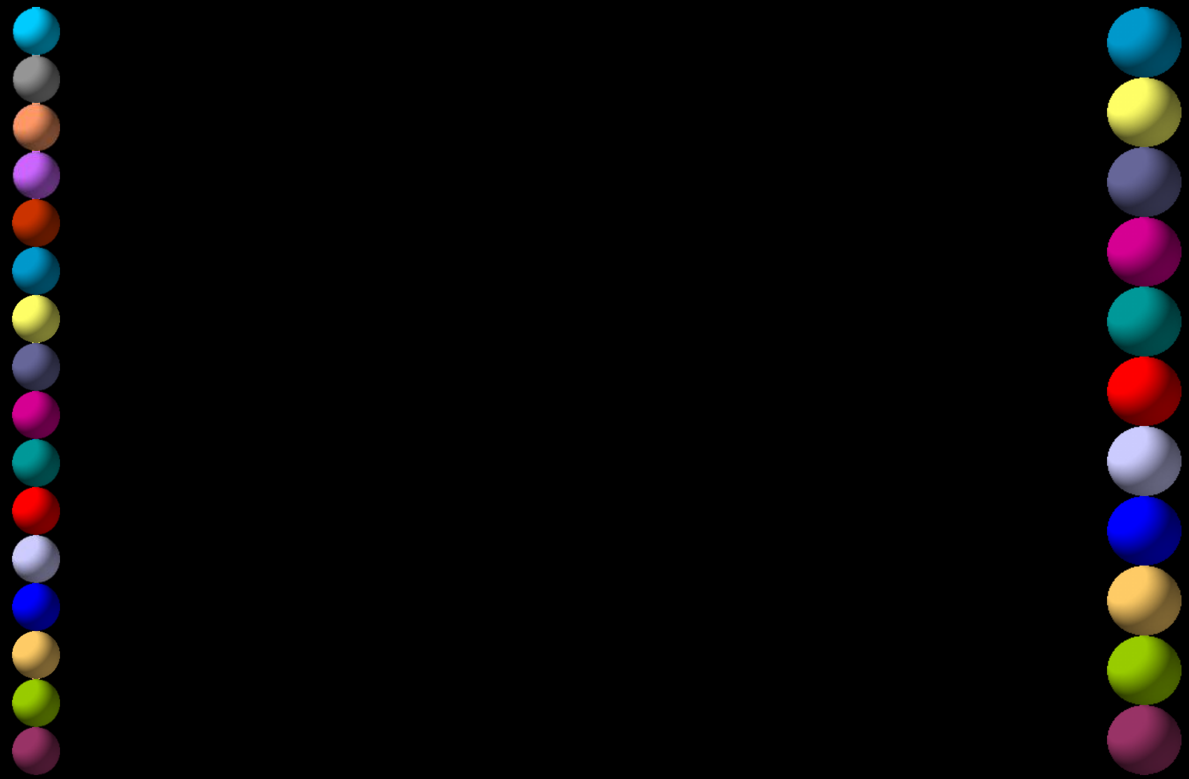


Application:
Entropy lower bounds

(Theorem: Moments Don’t Describe Anything)

high entropy

low entropy



Application:
Entropy lower bounds

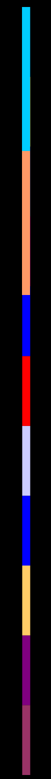
(Theorem: Moments Don’t Describe Anything)

high entropy

low entropy

edit a small portion of the distributions

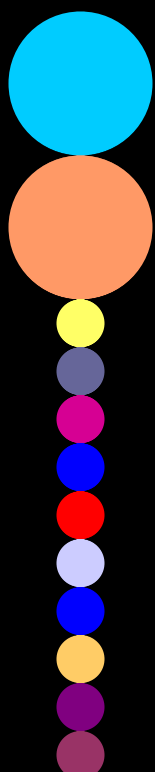
entropy is ≈preserved by continuity



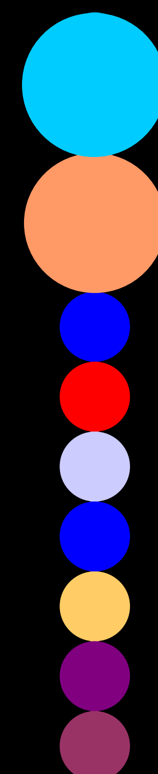
Application:
Entropy lower bounds

(Theorem: Moments Don’t Describe Anything)

high entropy



low entropy



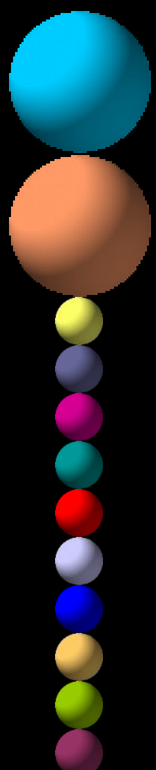
edit a small portion of the distributions

entropy is ≈preserved by continuity

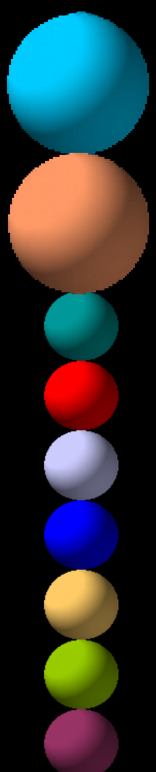
Application: Entropy lower bounds

(Theorem: Moments Don’t Describe Anything)

high entropy



low entropy



moments are dominated by those of the (arbitrary!) edits

left and right moments can be made to match

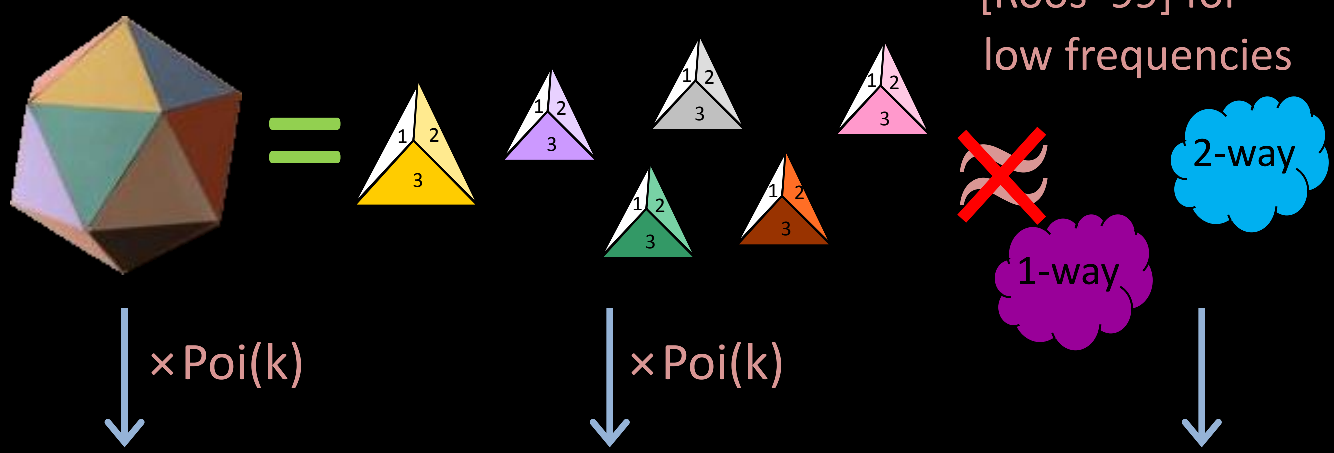
moments don’t determine entropy, or any other continuous property!

“Double Poissonization”

Characterize (and compare) fingerprint distributions

“Fingerprint entries are distributed (almost) like independent Poisson processes”

[Roos ’99] for low frequencies



(▼,▼,▼,▼,▼,▼,▼,▼,▼,▼) →

▼	▼	▼	▼	▼	▼	▼	▼
2	2	0	0	2	1	2	1

→

1-way	2-way	…
2	4	0

data

histogram of data

“fingerprint”
(histogram of histogram)

Generalized Multinomial Distributions

Generalizes binomial, multinomial distributions, and any sums of such distributions.

Parameterized by matrix

0	1
.3	.7
.3	.7
.3	.7
.3	.7
1	3

Generalized Multinomial Distributions

Generalizes binomial, multinomial distributions, and any sums of such distributions.

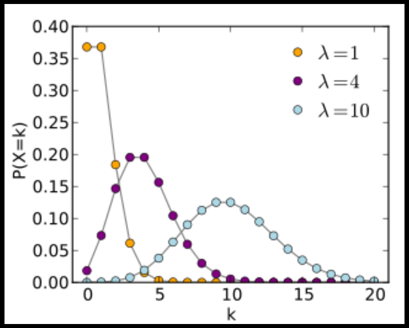
Parameterized by matrix

0	1	2
.5	.3	.2
.8	.1	.1
.3	.3	.4
.2	.8	0
1	2	1

Aim: Characterize these distributions so we can realize when two are close

So far: Characterization by Poissons in the "rare events" regime

Poisson vs ...?



Poi(λ) ≈ # heads in 1000λ coin flips,
where 1/1000 prob. heads

Poi(λ_1, λ_2, ...) : produce each dimension separately

~~Nice features:~~

Variance = mean

Each dimension is independent

New idea: multivariate ^ Gaussian

discretized

Parameterized by matrix → more degrees of freedom

Gaussian Characterization of Generalized Multinomial Distributions

Convolution/Deconvolution

Earthmover CLT

Unimodality [Gurvitz 2008]

Thm: Given multinomial distribution M with n rows, j columns, mean μ covariance Σ:

$$d_{TV}\left(M, G^{disc}(\mu,\Sigma)\right) \le \frac{j^{4/3}}{\sigma^{1/3}} \cdot 2.2 \cdot (3.1 + 0.83 \log n)^{2/3},$$

σ^2 min eigenvalue of Σ, G_d “Discretized” Gaussian

Characterizing Fingerprints

Fingerprint expectations
~~+ covariance~~ very similar

⇕

Dist. of fingerprints close

Richer characterization… too rich?

Nice Lemma:

Fingerprint expectations
determine fingerprint covariance

Expectations→Covariance

Lemma: Fingerprint expectations robustly *determine* fingerprint covariance

$$E[f_i] = \sum_{x:h(x)\neq 0} Poi(kx,i)\, h(x)$$

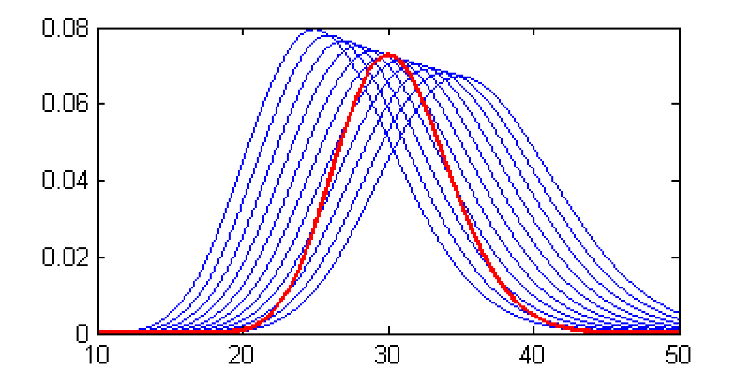
$$Cov[f_i, f_j] = - \sum_{x:h(x)\neq 0} Poi(kx,i)\, Poi(kx,j)\, h(x)$$

$$Poi(\lambda,i)*Poi(\lambda,j) = Poi(2\lambda,i+j)\binom{i+j}{j}\, 2^{-(i+j)}$$

Approximating "Thin" Poissons as Linear Combinations of Poissons

For fixed m but *all* λ, can we approximate Pr[Poi(2λ)=m] as a linear combination over j of Pr[Poi(λ)=j]?

$$\frac{e^{-2\lambda}(2\lambda)^m}{m!} \approx \sum_j \alpha_{j,m} \frac{e^{-\lambda}\lambda^j}{j!}$$



These look like Gaussians! Same answer as for Gaussians!

Result: can approximate to within ε using coefficients at most 1/ε

Application: Two Lower Bounds

Tool:

Fingerprint expectations
~~+ covariance~~ very similar
⇕
Dist. of fingerprints close

1. Explicit construction of distributions p^+, p^- where p^+ is "close" to $U_{n/2}$, and p^- is "close" to U_n , that have very close k-sample expected fingerprints, for k=c n/log n

2. Too hard! Throw everything into a linear program and let linear programming do all the work for us

Lower Bound Construction

Goal: distributions p^+, p^- where p^+ is "close" to $U_{n/2}$, and p^- is "close" to U_n, that have very close k-sample expected fingerprints, for k=c n/log n

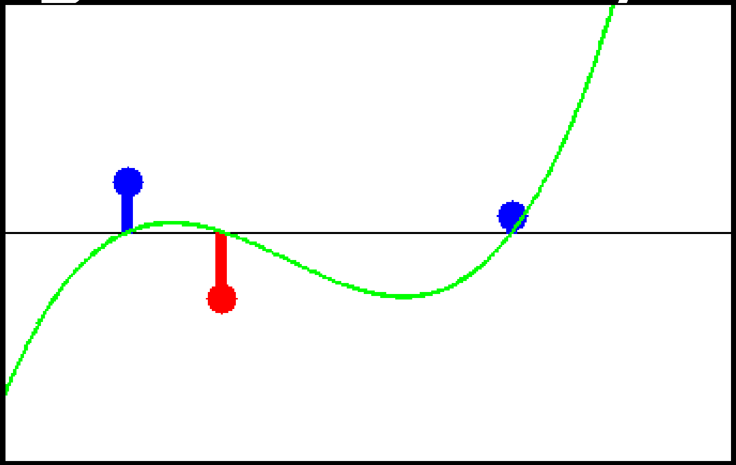
For all j, $\Sigma_x(h^+(x) - h^-(x)) \dfrac{(kx)^j e^{-xk}}{j!} \approx 0$

So: $g(xk):=(h^+(x) - h^-(x))e^{-xk}$ orthogonal to low deg polys

We want a *signed measure* g on the positive reals that:

- Is orthogonal to low degree polynomials
- Decays exponentially fast
- Its positive portion has most of its mass at $\dfrac{2c}{\log n}$
- Its negative portion has most of its mass at $\dfrac{c}{\log n}$

Orthogonal to Polynomials



Fact: If P is a degree j polynomial with distinct real roots $\{x_i\}$, then the signed measure g_P having point mass $1/P'(x_i)$ at each root x_i is orthogonal to all polynomials of degree $\leq j-2$

Task: find P such that $P'(x_i)$ grows exponentially in x_i

LAGUERRE POLYNOMIALS

Application: Two Lower Bounds

Tool:

Fingerprint expectations
~~+ covariance~~ similar
very
Dist. of fingerprints close

1. Explicit construction of distributions p^+, p^- where p^+ is "close" to $U_{n/2}$, and p^- is "close" to U_n , that have very close k-sample expected fingerprints, for k=c n/log n

2. Too hard! Throw everything into a linear program and let linear programming do all the work for us

Dual to "Optimal" Estimator

$$\min_{z,\epsilon} \epsilon + k^{-c} \cdot \sum_i |z(i)|$$

s.t. for all dists. p over $[n]$

$$\left|\mathbb{E}\left[z\left(p^k\right)\right] - H(p)\right| \le \epsilon$$

"Find estimator **z**:
Minimize ε, s.t.
expectation of estimator z applied to
k samples from p is within ε of H(p)"

$$\max_{y^+,y^-} H(y^+) - H(y^-)$$

→for y^+,y^- dists. over $[n]$

s.t. for all i, $E[f^+_i] - E[f^-_i] \le k^{1-c}$

"Find lower bound instance y^+,y^-
Maximize $H(y^+)-H(y^-)$ s.t.
expected fingerprint entries
given k samples from y^+,y^-
match to within k^{1-c}"

Take-home Ideas

Poissonization – crucial first step

"Fingerprints" capture all the information (for symmetric property testing)

→ More generally, "forgetting" and reconstructing sample

Two general characterizations of the fingerprint distribution:

Poisson regime → domain elements each seen <<1 time in sample
Tight characterization, can be leveraged even in extreme situations like >>1 distribution

Gaussian regime → fingerprints have high variance in every direction

Much more flexible, leads to tighter bounds, but more intricate

Fun with orthogonal polynomials! (Hermite, Laguerre, Chebyshev)

Approximating "Thin" Gaussians as Linear Combinations of Gaussians

What do we convolve a Gaussian with to approximate a *thinner* Gaussian?

(Other direction is easy, since convolving Gaussians adds their variances)

"Blurring is easy, unblurring is hard" → can only do it approximately

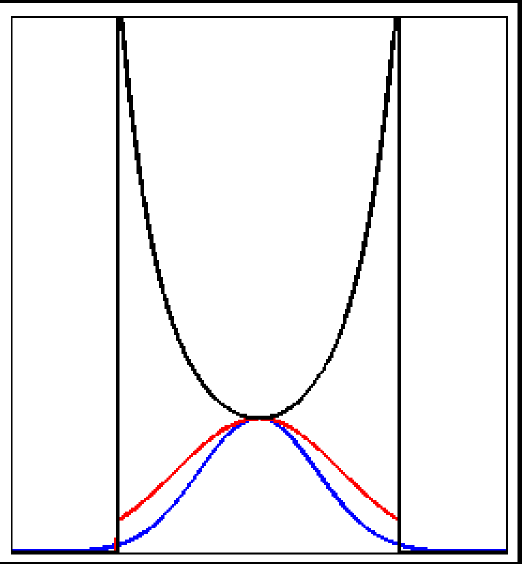
How to analyze? Fourier transform! Convolution becomes multiplication

Now: what do we *multiply* a Gaussian with to approximate a *fatter* Gaussian?

$e^{-x^2} \cdot ??? = e^{-x^2/2}$

$??? = e^{x^2/2}$

Problem: blows up



Answer: approximate to within ε => truncate where $e^{-x^2/2} = \varepsilon$

Result: can approximate to within ε using coefficients at most $1/\varepsilon$