Testing Symmetric Properties of Distributions

Paul Valiant, 2008

Presented by Omer Tamuz and Tamar Zondiner

July 8, 2010

A property π

A property of a distribution is a function $\pi : D_n \to \mathbb{R}$, where D_n is the set of probability distributions on $[n]$.

A binary property π_a^b

A property π and pair of real numbers $a < b$ induce a binary property $\pi_a^b : D_n \to \{"yes", "no", \varnothing\}$ defined by:

$$\pi_a^b(p) = \begin{cases} "yes" & \text{if } \pi(p) > b \\ "no" & \text{if } \pi(p) < a \\ \varnothing & \text{otherwise} \end{cases}$$

Let π_a^b be a binary property on D_n.

A tester

An algorithm T is a "π_a^b-**tester with sample complexity** $k(\cdot)$" if, given a sample of size $k(n)$ from a distribution $p \in D_n$, algorithm T will:

- accept with probability greater than $\frac{2}{3}$ if $\pi_a^b(p) = "yes"$, and
- reject with probability greater than $\frac{2}{3}$ if $\pi_a^b(p) = "no"$, and

The tester's behavior is unspecified when $\pi_a^b(p) = \phi$, i.e. when $a \leq \pi(p) \leq b$.

A Symmetric Property

A property π is symmetric if for all distributions p and all permutations σ we have $\pi(p) = \pi(p \circ \sigma)$.

An (ϵ, δ)-weakly continuous property

A property π is (ϵ, δ)-weakly continuous if for all distributions p^+, p^- satisfying $|p^+ - p^-| \leq \delta$ we have $|\pi(p^+) - \pi(p^-)| \leq \epsilon$.

$|x - y|$ denotes the L_1 distance.

Theorem

Distance from the uniform distribution is a symmetric and (δ, δ)*-weakly continuous property.*

Proof.

- Let U_n be the uniform distribution on $[n]$.
- Let $\pi(p) = |U_n - p|$ for $p \in D_n$.
- Let $p^+, p^- \in D_n$ be such that $|p^+ - p^-| < \delta$.
- Assume WLOG that $\pi(p^+) \geq \pi(p^-)$.

$$
\begin{aligned}
|\pi(p^+) - \pi(p^-)| &= |U_n - p^+| - |U_n - p^-| \\
&\leq |U_n - p^-| + |p^+ - p^-| - |U_n - p^-| \\
&= |p^+ - p^-| \leq \delta
\end{aligned}
$$

□

Theorem

The entropy is a symmetric and $\left(1, \frac{1}{2\log n}\right)$*-weakly continuous property.*

Proof.

Easy. □

Consider a sample of size k from distribution p over $[n]$. Let h_i be the number of appearances of i in the sample.

The Canonical Tester with parameter θ

1. Insert the constraint $\sum_i p_i = 1$.
2. For each i such that $h_i > \theta$ insert the constraint $p_i = \frac{h_i}{k}$. Otherwise insert the constraint $p_i \in [0, \frac{\theta}{k}]$.
3. Let P be the set of solutions to these constraints.
4. If the set $\pi_a^b(P)$ (the image of elements of P under π_a^b) contains only "*yes*" and $\varnothing$ return "*yes*". If it contains only "*no*" and $\varnothing$ return "*no*". Otherwise answer arbitrarily.

- It seems plausible that the canonical tester behaves correctly for the high frequency elements.
- The tester effectively discards all information regarding the low frequency elements.
- If we can show that no tester can extract information from these elements then it will follow that the canonical tester is almost optimal.

The Canonical Testing Theorem We Wish For

Not True Theorem

Given a symmetric (ϵ, δ)-weakly continuous property $\pi : D_n \to \mathbb{R}$ and two thresholds $a < b$, such that the Canonical Tester T^θ for $\theta = 600 \log n/\delta^2$ on π_a^b fails to distinguish between $\pi > b$ and $\pi < a$ in k samples, then no tester can distinguish between $\pi > b$ and $\pi < a$ in k samples.

Sadly, this is not true.

Canonical Testing Theorem

Theorem

Given a symmetric (ϵ, δ)-weakly continuous property $\pi : D_n \to \mathbb{R}$ and two thresholds $a < b$, such that the Canonical Tester T^θ for $\theta = 600 \log n / \delta^2$ on π_a^b fails to distinguish between $\pi > b + \epsilon$ and $\pi < a - \epsilon$ in k samples, then no tester can distinguish between $\pi > b - \epsilon$ and $\pi < a + \epsilon$ in $k \cdot \frac{\delta^3}{n^{o(1)}}$ samples.

Low Frequency Blindness

The crux is to prove that the canonical tester does the "right thing" (i.e., nothing!) for the low frequency elements.

Low Frequency Blindness Theorem

Let π be a symmetric property on distributions on $[n]$ that is (ϵ, δ)-weakly continuous.
Let p^+, p^- be two distributions that are identical for any index occurring with probability at least $\frac{\theta}{k}$ in either distribution, where $\theta = \frac{600 \log n}{\delta^2}$.
If $\pi(p^+) > b$ and $\pi(p^-) < a$, then no tester can distinguish between $\pi > b - \epsilon$ and $\pi < a + \epsilon$ in $k \cdot \frac{\delta^3}{n^{o(1)}}$ samples.

If we could show that such p^+ and p^- exist whenever the canonical tester fails than this would imply the canonical testing theorem.
Example: Entropy

Low Frequency Blindness $\Rightarrow$ Canonical Testing Theorem

Lemma

Given a distribution p and a parameter θ, if we draw k random samples from p then with probability at least $1 - \frac{4}{n}$ the set P constructed by the Canonical Tester will include a distribution $\hat{p}$ such that $|p - \hat{p}| \leq 24\sqrt{\frac{\log n}{\theta}}$.

If $\theta = 600 \log n/\delta^2$ then this reads $|p - \hat{p}| \leq \delta$.

Proof.

"The proof is elementary: use Chernoff bounds on each index i and then apply the union bound to combine the bounds." □

Low Frequency Blindness ⇒ Canonical Testing Theorem

Reminder: the canonical testing theorem states that if the canonical tester fails with k samples then any slightly weaker tester also fails.

Proof: Canonical Testing Theorem

- Assume canonical tester says "no" with probability 1/3 to some p for which $\pi(p) > b + \epsilon$ (so it should have said yes).
- $\Rightarrow$ with probability 1/3 there exists $p^- \in P$ such that $\pi(p^-) < a$.
- By the lemma, P contains some p^+ such that $|p - p^+| < \delta$ with probability $1 - 4/n$. $\pi(p^+) > b$ by continuity.
- $\Rightarrow$ there exists a single P with both of these properties.
- $\Rightarrow$ there exist p^- and p^+ with the same θ-high-frequency elements such that $\pi(p^-) < a$ and $\pi(p^+) > b$.
- $\Rightarrow$ the theorem follows by application of low frequency blindness.

Histogram

The histogram h of a vector $v = (v_1, \dots, v_k)$ is a vector such that h_i is the number of components of v with value i.

Fingerprint

A fingerprint f of a vector v is the histogram of the histogram of v.

Example

Let $v = (3, 1, 2, 2, 5, 1, 2)$. Then:

- Its histogram is $h = (2, 3, 1, 0, 1)$.
- Its fingerprint is $f = (2, 1, 1)$.
- We omit the zero component of f.

A tester for a symmetric distribution π may consider just the fingerprint of the sample and discard the rest of the information.

Poisson Moments Vector

Definition

- Let p be a distribution on $[n]$.
- Let the sample size be k.
- $k_i := \mathrm{E}[h_i] = k \cdot p_i$.

Let $\lambda_a := \sum_i \mathrm{poi}_{k_i}(a)$.
Then $\lambda = \{\lambda_a\}_{a=1}^{\infty}$ is the **Poisson moments vector of p for sample size k**.

- p has histogram h and fingerprint f.
- The distribution of h_i is well approximated by $\mathrm{poi}_{k_i}(\cdot)$.
- $\mathrm{E}[f_a] = \sum_i \mathbf{P}[h_i = a] \approx \lambda_a$.

Coffee Break

Coffee Break

Low Frequency Blindness

Theorem

Let π be a symmetric property on distributions on $[n]$ that is (ϵ, δ)-weakly continuous.
Let p^+, p^- be two distributions that are identical for any index occurring with probability at least $\frac{\theta}{k}$ in either distribution, where $\theta = \frac{600 \log n}{\delta^2}$.
If $\pi(p^+) > b$ and $\pi(p^-) < a$, then no tester can distinguish between $\pi > b - \epsilon$ and $\pi < a + \epsilon$ in $k \cdot \frac{\delta^3}{n^{o(1)}}$ samples.

Low Frequency Blindness
(simplified)

We'll limit our analysis to distributions with low frequencies. Suppose **all** elements have probability $< \frac{\theta}{k}$ where $\theta = \frac{600 \log n}{\delta^2}$.

Lemma

Let π be a symmetric property on distributions on $[n]$ that is (ϵ, δ)-weakly continuous.
Let p^+, p^- be two distributions for which all indices occur with probability at most $\frac{\theta}{k}$, where $\theta = \frac{600 \log n}{\delta^2}$.
If $\pi(p^+) > b$ and $\pi(p^-) < a$, then no tester can distinguish between $\pi > b - \epsilon$ and $\pi < a + \epsilon$ in $k \cdot \frac{\delta^3}{n^{o(1)}}$ samples.

Proof Sketch

Let p^+ and p^- be **low frequency distributions** such that $\pi(p^+) > b$ and $\pi(p^-) < a$.

1. We construct $\hat{p}^+$ and $\hat{p}^-$ such that
   - $|\hat{p}^\pm - p^\pm| < \delta$, and therefore $\pi(\hat{p}^+) > b - \epsilon$ and $\pi(\hat{p}^-) < a + \epsilon$.
   - $\hat{p}^+$ and $\hat{p}^-$ have similar **Poisson moments vector** for sample size $\hat{k} = k\frac{\delta^3}{n^{o(1)}}$.
2. For any sample size for which two distributions have similar Poisson moments vectors, they also have similar **fingerprints**.
3. We now have two distributions with similar fingerprints; one has the property and the other doesn't. It is therefore impossible to test for π_a^b with $\hat{k}$ samples.

Steps two and three are the **"Wishful Thinking Theorem"**.

- Each component of the fingerprint is a sum of many indicators. For example, f_3 is the sum of the indicators of the events $h_i = 3$.
- **Wishfully** assume that the h_is are independent and distributed Poisson with parameter $k_i = k \cdot p_i$. Then $\mathrm{E}[f_a] = \mathrm{Var}[f_a] = \lambda_a$.
- **Wishfully** assume that the f_as are independent and distributed Poisson with parameter λ_a.
- If for p^+ and p^- and each a we have that $|\lambda_a^- - \lambda_a^+|$ is smaller than $\sqrt{\lambda_a^+}$ then we expect the distributions' fingerprints to be indistinguishable.
- If $\pi(p^+) > b$ and $\pi(p^-) < a$ then no tester can test π_a^b.

Wishful Thinking Theorem

Given an integer $\hat{k} > 0$, let p^+ and p^- be two distributions, all of whose frequencies are at most $\frac{1}{500\hat{k}}$. Let λ^+ and λ^- be their Poisson moments vectors for sample size $\hat{k}$. If it is the case that

$$\sum_a \frac{|\lambda_a^+ - \lambda_a^-|}{\sqrt{1 + \max\{\lambda_a^+, \lambda_a^-\}}} < \frac{1}{25}$$

then it is impossible to test any symmetric property that is true for p^+ and false for p^- in $\hat{k}$ samples.

Reminder: whenever the canonical tester fails we are guaranteed to have such p^+ and p^-.

1. Show $h_i \approx \text{poi}_{k_i}$ (and $h \approx \text{Poi}(kp)$).
2. Show $f_a \approx \text{poi}_{\lambda_a}$ (and $f \approx \text{Poi}(\lambda)$).
3. Bound $|\text{Poi}(\lambda^+) - \text{Poi}(\lambda^-)|$.
4. Deduce a bound on $|f^+ - f^-|$.
5. Finally, conclude that since the fingerprints are indistinguishable (even though the distributions might not be), then the property can't be tested.

Poissonization

Poissonization

A k-Poissonized tester T is a function that correctly classifies a property on a distribution p with probability $7/12$ on input samples generated in the following way:

- Draw $k' \leftarrow \mathrm{poi}_k$.
- Return k' samples from p.

Lemma

If there exists a k-sample tester T for a property π_a^b then there exists a k-Poissonized tester T' for π_a^b.

Poissonization

- After Poissonization, the histogram component h_i is distributed poi_{k_i}, and the different h_is are independent.
- By additivity of expectations and variances $\text{E}[f_a] = \text{Var}[f_a] = \sum_i \text{poi}_{k_i}(a) = \lambda_a$.
- However, the different f_as aren't independent.

Generalized Multinomial Distribution

Definition: M^{ρ}, the generalized multinomial distribution(ρ)

- Let ρ be a matrix with n rows, such that row ρ_i represents a distribution.
- From each such row, draw one column according to the distribution.
- Return a row vector recording the total number of samples falling into each column (the histogram of the samples).

Lemma

The distribution of fingerprints of poi(k) samples from p (the distribution of f after Poissonization) is the generalized multinomial distribution, M^{ρ}, when using $\rho_i(a) = \text{poi}_{k_i}(a)$ to define the rows ρ_i.

Roos's Theorem

Roos's theorem

Given a matrix ρ, letting $\lambda_a = \sum_i \rho_i(a)$ be the vector of column sums, we have

$$|M^\rho - \mathrm{Poi}(\lambda)| \leq 8.8 \sum_a \frac{\sum_i \rho_i(a)^2}{\sum_i \rho_i(a)}.$$

So, the multivariate Poisson distribution is a good approximation for the fingerprints, if ρ is small enough.

Bounding ρ using the low-frequencies

Suppose that for some $0 < \epsilon \leq \frac{1}{2}$ it holds that $p_i \leq \frac{\epsilon}{k}$. Then $\rho_i(a) = \mathrm{poi}_{k_i}(a) = \frac{e^{-k_i} k_i^a}{a!} = \frac{e^{-k \cdot p_i}(k \cdot p_i)^a}{a!} \leq (k \cdot p_i)^a \leq \epsilon^a$.

Thus:

$$\sum_a \frac{\sum_i \rho_i(a)^2}{\sum_i \rho_i(a)} \leq \sum_a \max_i \rho_i(a) \leq \sum_a \epsilon^a \leq 2\epsilon$$

and by Roos's theorem:

$$|M^\rho - \mathrm{Poi}(\lambda)| \leq 2 \cdot 8.8\epsilon.$$

Multivariate Poisson Statistical Distance

Bounding the statistical distance between λ^+ and λ^-

The statistical distance between two multivariate Poisson distributions with parameters λ^+, λ^- is bounded by

$$|\mathrm{Poi}(\lambda^+) - \mathrm{Poi}(\lambda^-)| \leq 2 \sum_a \frac{|\lambda_a^+ - \lambda_a^-|}{\sqrt{1 + \max\{\lambda_a^+, \lambda_a^-\}}}.$$

Hence, by the theorem's hypothesis:

$$|\mathrm{Poi}(\lambda^+) - \mathrm{Poi}(\lambda^-)| \leq \frac{2}{25}.$$

Wishful Thinking Theorem
(reminder)

Wishful Thinking Theorem

Given an integer $\hat{k} > 0$, let p^+ and p^- be two distributions, all of whose frequencies are at most $\frac{1}{500\hat{k}}$. Let λ^+ and λ^- be their Poisson moments vectors for sample size $\hat{k}$. If it is the case that

$$\sum_a \frac{|\lambda_a^+ - \lambda_a^-|}{\sqrt{1 + \max\{\lambda_a^+, \lambda_a^-\}}} < \frac{1}{25}$$

then it is impossible to test any symmetric property that is true for p^+ and false for p^- in $\hat{k}$ samples.

Proof.

- $f^{\pm} \sim M^{\rho^{\pm}}$.
- Combining Roos's theorem with the bound on ρ, and assuming that $p_i^{\pm} \leq \frac{1}{500k}$, we get that $|M^{\rho^{\pm}}$ - $\text{Poi}(\lambda^{\pm})| \leq \frac{2 \cdot 8.8}{500} < \frac{1}{25}$.
- The theorem's hypothesis implies $|\text{Poi}(\lambda^{+}) - \text{Poi}(\lambda^{-})| \leq \frac{2}{25}$.
- Using the triangle inequality, we get that the statistical distance between the distributions of fingerprints of $\text{Poi}(k)$ samples from p^{+} versus p^{-} is at most $\frac{4}{25} < \frac{1}{6}$.
- A k-tester (poissonized) must have a gap$> \frac{1}{6}$ (succeed with probability $\frac{7}{12}$). This is impossible if $|p^{+} - p^{-}| < 1/6$.
- If a k-Poissonized tester doesn't exist, then neither does a k-tester.

$\Rightarrow$ it is impossible to test any symmetric property that is true for p^{+} and false for p^{-} in k samples. □

Questions?

Thanks!