Doubly-Competitive Distribution Estimation

Yi Hao and Alon Orlitsky

Department of Electrical and Computer Engineering
University of California, San Diego

ICML, June 11, 2019

Distribution Estimation

- p - unknown distribution over $\{1, 2, \ldots, k\}$
- $X^n := X_1, X_2, \ldots, X_n \sim p$ independently
- q_{X^n} - estimate based on X^n
- Loss: Kullback-Leibler divergence

$$\ell(p, q_{X^n}) := \sum_{x=1}^{k} p(x) \log \frac{p(x)}{q_{X^n}(x)}$$

Competitive Distribution Estimation

- All reasonable estimators are natural
  - Same probability to symbols appearing same # times
  - $q_{abbc}(a) = q_{abbc}(c)$
- Goal: Estimate every p as well as best natural estimator
- Genie-estimator: knows p, but natural, hence incurs a loss

$$\mathbf{Opt}(p, X^n) := \min_{q\text{ - natural}} \ell(p, q_{X^n})$$

- (Orlitsky & Suresh, 2015) **Good-Turing variation q^{GT}**
  - For every p, with high probability

$$\ell(p, q^{\mathrm{GT}}_{X^n}) \leq \mathbf{Opt}(p, X^n) + \mathcal{O}\left(\frac{1}{\sqrt{n}} \wedge \frac{k}{n}\right)$$

Doubly-Competitive Distribution Estimation

- $D_\Phi :=$ # of distinct frequencies of symbols in X^n

$$X^n = a\ b\ a\ c\ d\ e \implies a \text{ appeared twice, } b\ c\ d\ e \text{ appeared once}$$
$$\implies D_\Phi = 2$$

- Single estimator $q^\star$ achieving (w.h.p.)

$$\ell(p, q^\star_{X^n}) \le \mathbf{Opt}(p, X^n) + \mathcal{O}\left(\frac{D_\Phi}{n}\right)$$

- **Uniform bound**: $D_\Phi \le \sqrt{2n} \wedge k \implies$ (Orlitsky & Suresh, 2015)
- **Better bounds for many distribution classes:**
  - T-step: $D_\Phi \lesssim T \cdot n^{\frac{1}{3}}$; Uniform: $D_\Phi \lesssim n^{\frac{1}{3}}$
  - Log-concave with SD $\approx \sigma$: $D_\Phi \lesssim \sigma \wedge \left(\frac{n^2}{\sigma}\right)^{\frac{1}{3}}$
  - Enveloped power-law $\{p : p(x) \lesssim x^{-\alpha}\}$: $D_\Phi \lesssim n^{-\frac{\alpha}{\alpha+1}}$
  - Log-convex distribution families, etc.

Estimator Construction

- $\Phi(t) := \#$ of symbols appearing t times
- Good-Turing Estimator

$$q^{\mathsf{GT}}(x) := \frac{t+1}{n} \cdot \frac{\Phi(t+1)}{\Phi(t)}$$

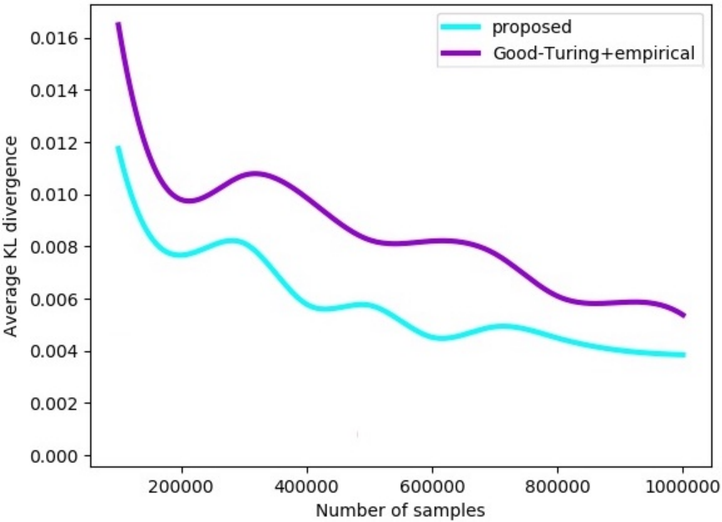
- **Observation**: For $\boldsymbol{x}$ appearing $\boldsymbol{t \gtrsim \log n}$ times, and $\boldsymbol{\Phi(t) \gtrsim \log^2 n}$ q^{GT} has sub-optimal variance in estimating $p(x)$
- **Averaging unbiased estimators reduces the variance**
  $\mathcal{D}(t) :=$ weighted average of $\Phi(t')$ for $|t' - t| \lesssim \sqrt{t/\log n}$

$$q^{\star}(x) := \frac{t+1}{n} \cdot \frac{\mathcal{D}(t+1)}{\mathcal{D}(t)},$$

- **For other $\boldsymbol{x}$**, use Good-Turing or empirical

Experimental Results

Two-step distribution



Thank You