Fixed Length Distribution Matching

Patrick Schulte

Chair for Communications Engineering
Technische Universität München
patrick.schulte@tum.de

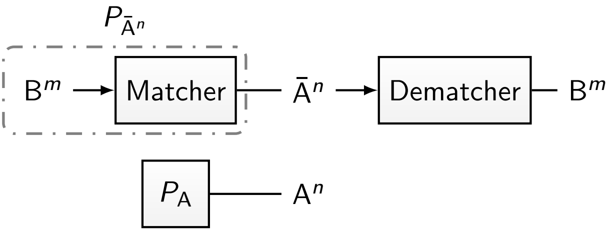
May 26, 2015
Doktorandenseminar

Outline

- Motivation: Fixed Length Distribution Matching
- Contribution: Constant Composition Distribution Matcher
- New Challange: unnormalized Divergence

Distribution Matching

Emulating discrete memoryless source P_A:



- B_i iid Bernoulli(1/2).
- $\bar{A}_i$ approximately iid $\sim P_A$ in the sense of

$$\frac{D(P_{\bar{A}^n} \| P_A^n)}{n} \overset{n\to\infty}{\to} 0, \qquad \frac{m}{n} \overset{n\to\infty}{\to} H(A)$$

- invertible: B^m can be recovered from $\bar{A}^n$ with zero error.

Definition: Achievable Rate

A matching rate $R = m/n$ is achievable for a distribution P_{A} if for any $\alpha > 0$ and sufficiently large n there is an invertible mapping $f : \{0, 1\}^m \to \mathcal{A}^n$ for which

$$\frac{\mathbb{D}\left(P_{f(\mathsf{B}^m)} || P_{\mathsf{A}}^n\right)}{n} \leq \alpha. \tag{1}$$

Converse,[1]

There exists a positive-valued function β with

$$\beta(\alpha) \xrightarrow{\alpha \to 0} 0$$

such that (1) implies

$$\frac{m}{n} \leq \frac{\mathbb{H}(A)}{\mathbb{H}(B)} + \beta(\alpha).$$

The converse[1] bounds the maximum rate that can be achieved. Since $\mathbb{H}(B) = 1$ we have

$$R \leq \mathbb{H}(A)$$

[1]Georg, Ali "Informational divergence and entropy rate on rooted trees with probabilities"

Optimal Fixed Length Distribution Matcher[2]

Algorithm

- select m near $m = nR$
- take 2^m most probable sequences (accoding to P_a^n) out of $\{0,1\}^n$ and index them by $\{0,1\}^m$
- perform line search on m with respect to best divergence.

$\Rightarrow$ achieves the converse bound

$\Rightarrow$ Codebook needs to be stored.

[2] Ali, "Algorithms for simulation of discrete memoryless sources"

Constant Composition Codes

The empirical distribution of a output sequence $\mathbf{c}$ of length n is

$$P_{\mathsf{A},\mathbf{c}}(a) := \frac{n_a(\mathbf{c})}{n},$$

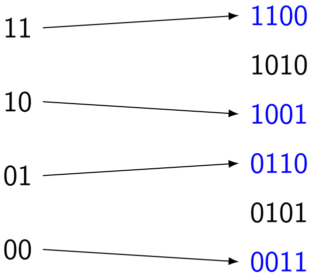
- $n_a(\mathbf{c}) = |\{i : c_i = a\}|$ number of occurences of symbol a in $\mathbf{c}$
- $P_{\mathsf{A},\mathbf{c}}$ is the *type* of $\mathbf{c}$
- $\mathcal{T}^n_{P_{\mathsf{A}}}$ is set of all n-type P_{A} sequences
- Codebook $\mathcal{C}_{\text{ccdm}} \subseteq \mathcal{A}^n$ is called a *constant composition code* if all codewords are of the same type, i.e. $\mathcal{C}_{\text{ccdm}} \subseteq \mathcal{T}^n_{P_{\mathsf{A}}}$

Constant Composition Distribution Matching

Idea

Use codewords of the same type

- Choose sequence length of the matcher output n
- Choose n-type approximation $P_{\bar{A}}$ of the target distribution P_A
- Choose $m = \left\lfloor \log_2 \left|\mathcal{T}^n_{P_{\bar{A}}}\right| \right\rfloor$
- Construct a unique mapping $\{0,1\}^m \to \mathcal{T}^n_{P_{\bar{A}}}$



Constant Composition Distribution Matching

We can show for any mapping to constant composition codewords for normalized divergence and Rate

$$\frac{\mathbb{D}\left(P_{\bar{\mathsf{A}}^n}||P_{\mathsf{A}}^n\right)}{n} = \mathbb{H}(\bar{\mathsf{A}}) - R + \mathbb{D}\left(P_{\bar{\mathsf{A}}}||P_{\mathsf{A}}\right)$$

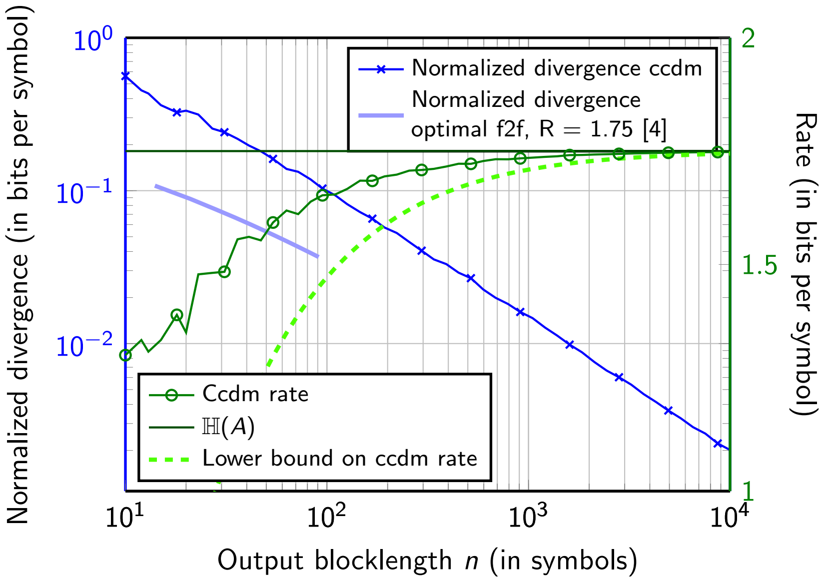
N-type approximation of Gerog and Bernhard [3] guarantees

$$\mathbb{D}\left(P_{\bar{\mathsf{A}}}||P_{\mathsf{A}}\right) < \frac{|\mathcal{A}|}{\min\limits_{a \in \operatorname{supp} P_{\mathsf{A}}} P_{\mathsf{A}}(a) n^2}$$

We obtain

$$\lim_{n\to\infty} R = \mathbb{H}(\mathsf{A}) = \mathbb{H}(\bar{\mathsf{A}})$$

[3] Georg, Bernhard, "Optimal quantization for distribution synthesis"


Normalized divergence ccdm
Normalized divergence optimal f2f, R = 1.75 [4]
Ccdm rate
ℍ(A)
Lower bound on ccdm rate
Normalized divergence (in bits per symbol)
Rate (in bits per symbol)
Output blocklength n (in symbols)
10^0
10^−1
10^−2
2
1.5
1
10^1
10^2
10^3
10^4

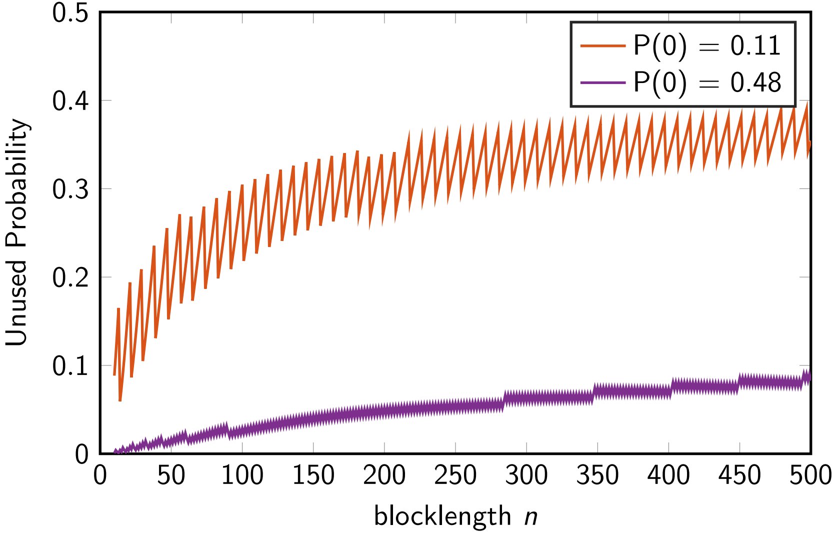
However ... What if we want unnormalized divergence to approach zero?

Tighter Constraints

- CCDM can't drive unnormalized divergence to zero
- Can the optimal block-to-block distribution matcher?

⇒ We don't know yet

Some evidence - Unassigned Probability



Conclusions

- Constant Composition Distribution Matching:
  normalized divergence goes to zero
  unnormalized divergence diverges from zero
- Optimal fixed length matcher:
  only normalized divergence behaviour known

References

- G. Böcherer and R. A. Amjad, "Informational divergence and entropy rate on rooted trees with probabilities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Sep. 2014, pp. 176–180.
- R. A. Amjad, "Algorithms for simulation of discrete memoryless sources," Master's thesis, Technische Universität München, 2013.
- G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," *arXiv preprint*, 2014. [Online]. Available: http://arxiv.org/abs/1307.6843
- G. Böcherer and R. A. Amjad, "Block-to-block distribution matching," Feb. 2013. [Online]. Available: http://arxiv.org/abs/1302.1020