Constant Composition Distribution Matching

Patrick Schulte
Patrick.Schulte@tum.de

Motivation and Example

- Reliable communication possible if transmission rate is below capacity.
- Shaping gap can be reduced using non-uniform input distributions.
- Distribution matchers transform random processes reversibly.

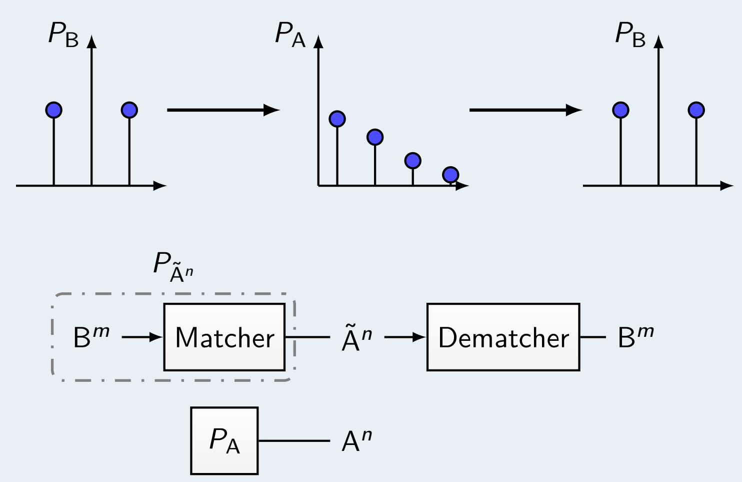


Figure 1 : Matching a data block $B^m = B_1 \ldots B_m$ to output symbols $\tilde{A}^n = \tilde{A}_1 \ldots \tilde{A}_n$ and reconstructing the original sequence at the dematcher. The rate is $\frac{m}{n}\left[\frac{\text{bits}}{\text{output symbol}}\right]$. The matcher can be interpreted as emulating a discrete memoryless source P_A.

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

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

Approaches so far

- Optimal variable length distribution matchers proposed in [1](variable-to-fixed) and [2](fixed-to-variable)
- Optimal fixed length distribution matcher [3]
  → No closed algorithm, codebook needs to be stored. Infeasible for large codebooks.
- Arithmetic distribution matcher creates codebook online [4]
  → variable length approaches have problems like error propagation and huge buffers.
- ϵ-error distribution matchers [5, Sec. 4.8] [6]
  → some sequences are irreversible.
- Adaptive arithmetic distribution matching (aadm) is a fixed length distribution matcher.[7]
  → computationally too complex for practical implementation

Question:
How can we create an invertible fixed length computationally feasible distribution matcher?

Converse

Converse,[8, Proposition 8]

There exists a positive-valued function β with

$$\beta(\alpha) \stackrel{\alpha \to 0}{\longrightarrow} 0$$

such that (1) implies

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

[8, Proposition 8] bounds the maximum rate that can be achieved. Since $\mathbb{H}(B) = 1$ we have

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

Constant Composition Codes

The empirical distribution of a vector $\boldsymbol{c}$ of length n is defined as

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

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

Constant Composition Distribution matching

Idea

Constant composition distribution matchers (ccdm) create only constant composition codewords.

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

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

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

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

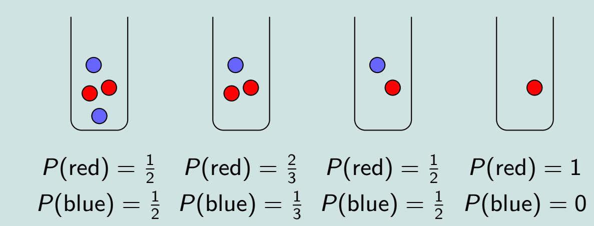
N-type approximation of [9, Proposition 4] guarantees

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

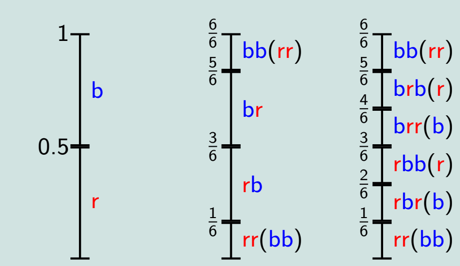
Question:
How can we efficiently index 2^m sequences out of $\mathcal{T}^n_{P_{\bar{A}}}$?

Indexing with Arithmetic Coding

Drawing without replacement



Drawing without replacement from a urn with $n = 4$, $n_{\text{blue}} = 2$, $n_{\text{red}} = 2$. The probability of drawing a red ball changes conditioned on balls that where already drawn.



All sequences of drawn balls are equally probable and they have the same empirical distribution.

Arithmetic Coding

- Links all possible input sequences in $\{0,1\}^m$ uniquely to sequences in $\mathcal{T}^n_{P_{\bar{A}}}$
- Associates an interval to each sequence.
- Interval's size equal to according probability of input (Bernoulli$\left(\frac{1}{2}\right)$) and output (urn) model, respectively.
- The intervals are ordered lexicographically.
- All input and output intervals range from 0 to 1.
- Codebook implicitly defined by the matcher.

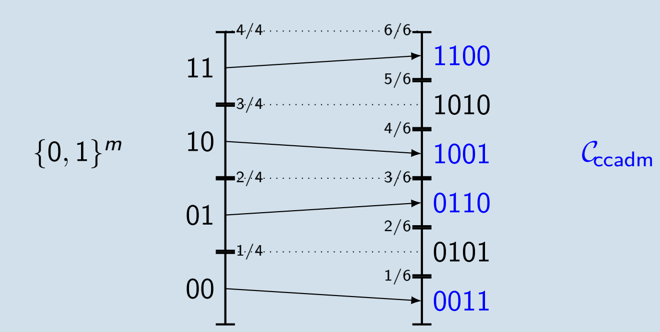


Figure 2 : Diagram of a constant composition arithmetic encoder with $P_{\bar{A}}(0) = P_{\bar{A}}(1) = 0.5$, $m = 2$ and $n = 4$.

Advantages

- On the fly matching and dematching
- Low memory resources needed
- Asymptotically optimal
- Very long code blocks are possible.
- low computational complexity

Simulation Results

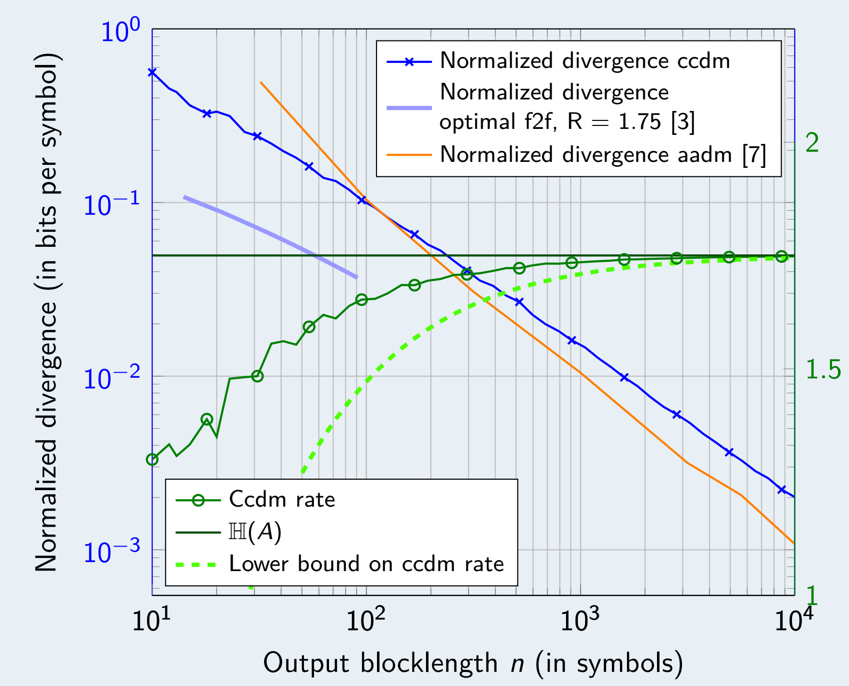


Figure 3 : Normalized divergence and rate of ccdm over output blocklength. $P_A = (0.0722; 0.1654; 0.3209; 0.4415)$.

Future Directions

- Good performance finite length distribution matcher
- Applications for distribution matching

References

[1] G. Böcherer and R. Mathar, "Matching dyadic distributions to channels," in *Proc. Data Compression Conf.*, 2011, pp. 23–32.

[2] R. A. Amjad and G. Böcherer, "Fixed-to-variable length distribution matching," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2013, pp. 1511–1515.

[3] G. Böcherer and R. A. Amjad, "Block-to-block distribution matching," Feb. 2013. [Online]. Available: http://arxiv.org/abs/1302.1020

[4] S. Baur and G. Böcherer, "Arithmetic distribution matching," in *Int. ITG Conf. Syst. Commun. Coding*, Feb. 2015.

[5] R. A. Amjad, "Algorithms for simulation of discrete memoryless sources," Master's thesis, Technische Universität München, 2013.

[6] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," *Proc. Allerton Conf. Commun., Contr., Comput.*, Oct. 2014.

[7] P. Schulte, "Zero error fixed length distribution matching," Master's thesis, Technische Universität München, 2014.

[8] 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.

[9] G. Böcherer and B. C. Geiger, "Optimal quantization for distribution synthesis," *arXiv preprint*, 2014. [Online]. Available: http://arxiv.org/abs/1307.6843

Institute for Communications Engineering

Unterstützt von / Supported by
Alexander von Humboldt Stiftung/Foundation

Technische Universität München