On Channel Simulation with Causal Rejection Samplers

Daniel Goc Gergely Flamich

Computational and Biological Learning Lab
Department of Engineering

UNIVERSITY OF CAMBRIDGE

Channel Simulation vs Lossy Source Coding

Figure 1: Bottom-right image from Careil et al. [2023]

Channel Simulation

1. $X, Y \sim P_{X,Y}$
2. Common randomness S
3. Receive $X \sim P_X$, send code C, decode $Y \sim P_{Y|X}$
4. Efficiency [Li and El Gamal, 2018]:

$$
\begin{aligned}
D_{\mathrm{KL}}[P_{Y|X} \| P_Y] \\
&\leq \mathbb{E}_S\left[|C| \mid X\right] \\
&\qquad \leq D_{\mathrm{KL}}[P_{Y|X} \| P_Y] + \log_2(D_{\mathrm{KL}}[P_{Y|X} \| P_Y] + 1) + 4
\end{aligned}
$$

5. No algorithm with $\mathrm{poly}(D_{\mathrm{KL}}[P_{Y|X} \| P_Y])$ runtime! [Agustsson and Theis, 2020].

Computational framework

Harsha et al. [2007]:

- $X_1, X_2, \ldots, X_N, \ldots \sim P$

- $X_N \sim Q$

Causal rejection sampling:

- N is a stopping time:

$$\{N > n\} \perp X_{n+1}, \ldots$$

Runtime of causal rejection samplers

The expected runtime of any *causal* rejection sampler is lower-bounded by:

$$\exp(D_\infty[Q \| P])$$

New Divergence

Channel Simulation Divergence

For $Q \ll P$:

$$D_{CS}[Q \| P] = -\int_{h=0}^{\infty} w(h) \log_2(w(h))\, dh$$

where:

$$w(h) = \mathbb{P}_{X \sim P}\left[\frac{dQ}{dP}(X) \geq h\right] = P\left(\frac{dQ}{dP} \geq h\right)$$

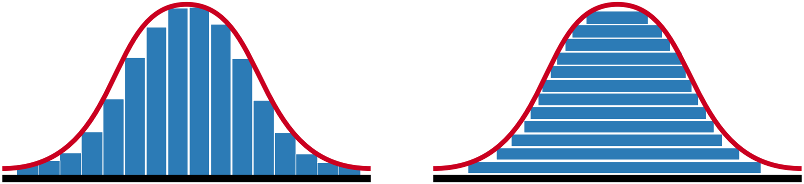

Figure 2: Left: “KL” Right: “CSD”

Comparison with KL Divergence

Properties of channel simulation divergence:

- Non-negativity:

$$D_{CS}[Q \| P] \geq 0$$

- Convexity:

$$D_{CS}[\lambda Q_1 + (1-\lambda) Q_2 \| P] \leq \lambda D_{CS}[Q_1 \| P] + (1-\lambda) D_{CS}[Q_2 \| P]$$

- A sandwich bound:

$$D_{\mathrm{KL}}[Q \| P] \leq D_{CS}[Q \| P] \leq D_{\mathrm{KL}}[Q \| P] + \log_2(D_{\mathrm{KL}}[Q \| P] + 1) + 1$$

Key Result

Entropy of causal rejection samplers

Let (P, Q) be a pair of distributions. Let N be an index returned by causal rejection sampler with proposal P and target Q. Then:

$$D_{CS}[Q \| P] \leq \mathbb{H}[N \mid S]$$

Furthermore, if N is the index returned, then:

$$D_{CS}[Q \| P] \leq \mathbb{H}[N \mid S] \leq D_{CS}[Q \| P] + \log_2(e+1)$$

$$\mathbb{E}_X\left[D_{CS}[P_{Y|X} \| P_Y]\right] \leq \mathbb{H}[Y \mid X, S] \leq \mathbb{H}[Y \mid S] \leq \mathbb{E}_S\left[|C|\right]$$

Numerical Examples

- **(A):** 1D Laplace
  - $P = \mathcal{L}(0, 1)$
  - $Q = \mathcal{L}(0, b)$
  - $D_{CS}[Q \| P] \cdot \ln 2 = b + \psi(1/b) + \gamma - 1$
  - $D_{CS}[Q \| P] - D_{\mathrm{KL}}[Q \| P] \to \gamma \log_2 e$ as $b \to 0$
- **(B):** d iid Gaussians
  - $P = \mathcal{N}(0, 1)^{\otimes d}$
  - $Q = \mathcal{N}(1, 1/4)^{\otimes d}$
  - How does $D_{CS}[Q \| P] - D_{\mathrm{KL}}[Q \| P]$ scale as $d \to \infty$?

Numerical Results

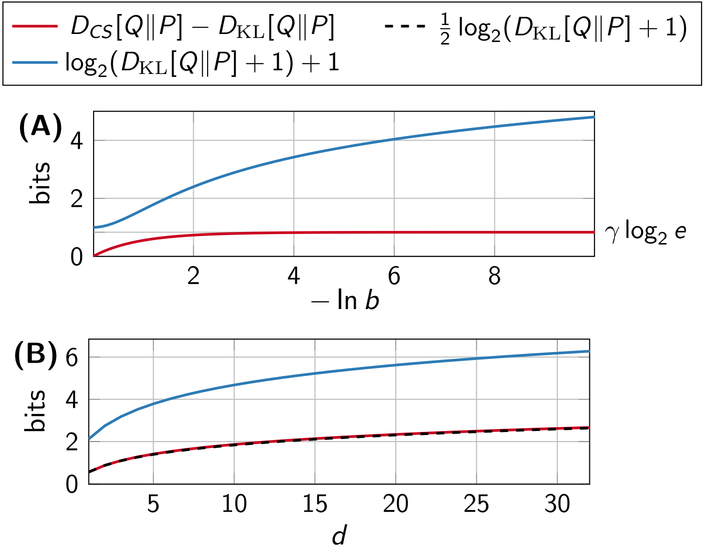

Figure 3: **A)** $Q = \mathcal{L}(0, b), P = \mathcal{L}(0, 1)$. **B)** $Q = \mathcal{N}(1, 1/4)^{\otimes d}, P = \mathcal{N}(0, 1)^{\otimes d}$

Contributions

- Shown that for any causal rejection sampler

$$\exp(D_\infty[Q\|P]) \leq \mathbb{E}\left[N\right]$$

- Defined and analysed new statistical distance $D_{CS}[Q\|P]$.

- Shown that

$$D_{CS}[Q\|P] \leq \mathbb{H}[N \mid S] \leq D_{CS}[Q\|P] + \log_2(e+1)$$

- Demonstrated non-trivial lower-bounds.

Shout-out: For upper-bounds: Optimal Redundancy in Exact Channel Synthesis by Sriramu and Wagner [2024].

References

E. Agustsson and L. Theis. Universally quantized neural compression. *Advances in Neural Information Processing Systems*, 33, 2020.

M. Careil, M. J. Muckley, J. Verbeek, and S. Lathuilière. Towards image compression with perfect realism at ultra-low bitrates. In *The Twelfth International Conference on Learning Representations*, 2023.

P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan. The communication complexity of correlation. In *Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 10–23. IEEE, 2007.

C. T. Li and A. El Gamal. Strong functional representation lemma and applications to coding theorems. *IEEE Transactions on Information Theory*, 64(11):6967–6978, 2018.

S. M. Sriramu and A. B. Wagner. Optimal redundancy in exact channel synthesis. *arXiv preprint arXiv:2401.16707*, 2024.

Relation to Excess Functional Information

Li and El Gamal (2017) defined:

$$\Psi(X \to Y) = \inf_{Z:\, Z \perp\!\!\!\perp X,\, H(Y|X,Z)=0} I(X; Z|Y)$$

Excess functional information

Let X, Y be two correlated variables. Then:

$$0 \leq \Psi(X \to Y) \leq \log_2(I(X;Y) + 1) + 4$$

We show that:

$$\mathbb{E}_X \left[D_{CS}[P_{Y|X} \| P_Y] \right] \leq \Psi(Y \to X) + I(X;Y)$$