Communication Complexity of Exact Sampling under Rényi Information

Tamás Linder

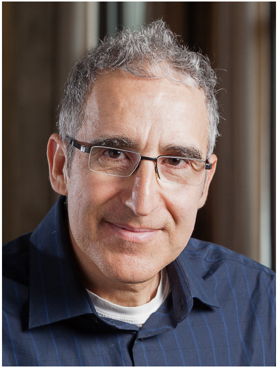
Queen's University, Canada

Hungarian Machine Learning Days

August 14, 2025

Co-authors

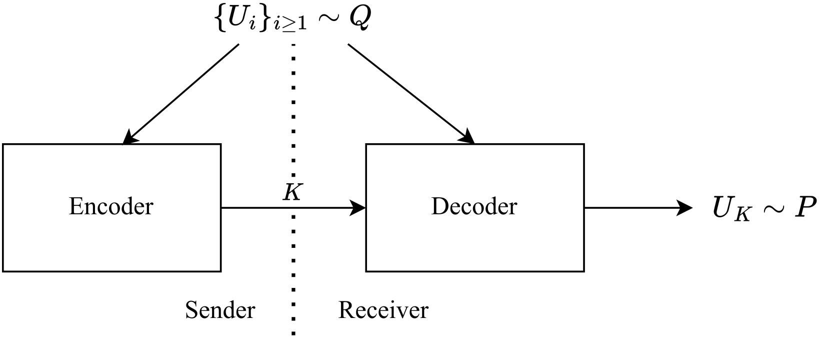
Spencer Hill



Fady Alajaji

Queen's University

Exact Sampling



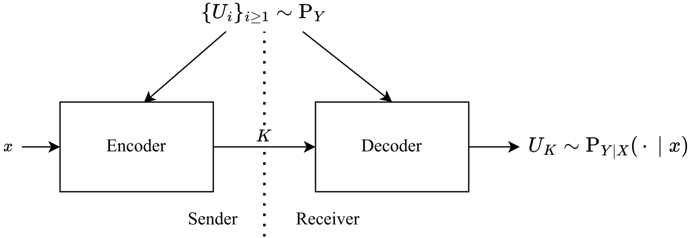
- When the goal is to efficiently communicate K, one can achieve

$$H(K) \approx D(P\|Q) \quad \text{bits}$$

- **Shannon:** K can be losslessly encoded at rate R such that

$$H(K) \leq R < H(K) + 1$$

Channel Simulation from Exact Sampling



X, Y random variables, choose $P = \mathrm{P}_{Y|X}(\,\cdot \mid x)$ and $Q = \mathrm{P}_Y$.

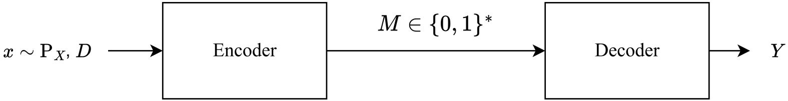
On input $x \sim \mathrm{P}_X$, sampling from P simulates the channel $X \to Y$.

Can simulate the channel with communication cost close to

$$H(K) \approx \mathbb{E}_X[D(\mathrm{P}_{Y|X}(\,\cdot \mid X) \,\|\, \mathrm{P}_Y)] = I(X;Y) \quad \text{bits}$$

Why Care?

Lossy Source Coding

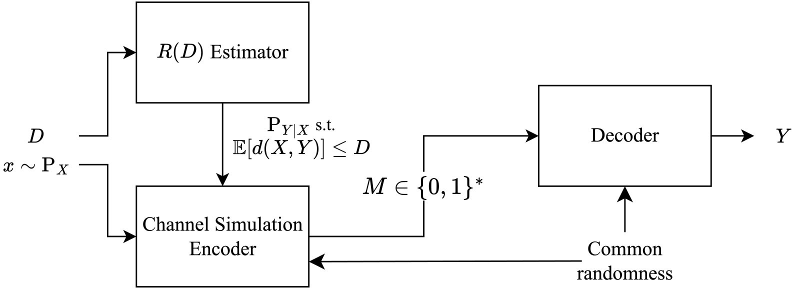


- The encoder encodes the block $(X_1, \ldots, X_n)$
- Decoder reconstructs $(Y_1, \ldots, Y_n)$
- Distortion: $D = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\big[d(X_i, Y_i)\big]$
- Rate: $R = \frac{1}{n}\mathbb{E}|M|$ (expected message length)
- Asymptotically $(n \to \infty)$ optimal performance

$$R(D) = \min_{P_{Y|X} \,:\, \mathbb{E}[d(X,Y)] \le D} I(X;Y).$$

Realization with Channel Simulation

$$R(D) = \min_{P_{Y|X} : \mathbb{E}[d(X,Y)] \leq D} I(X;Y).$$



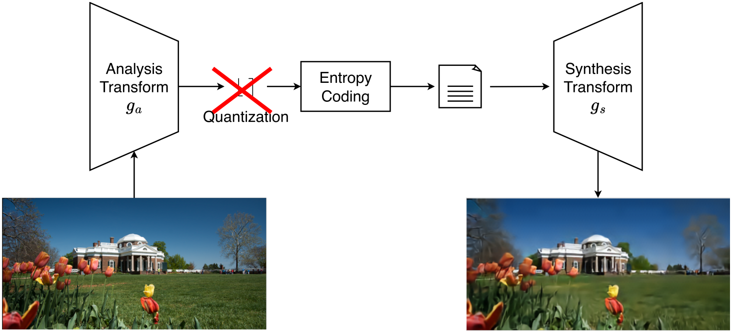
Recent work on neural-estimation of the rate-distortion function and $R(D)$-achieving conditional distribution Lei et al. (2023).

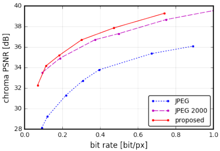
Channel simulation at cost $I(X;Y) \implies$ one-shot code achieving $R(D)$

Other Applications

- **Neural compression via nonlinear transform coding**
- Compression via implicit neural representation
- Rate-distortion-perception tradeoff
- Local differential privacy
- Federated learning, . . .

Neural Compression via Nonlinear Transform Coding

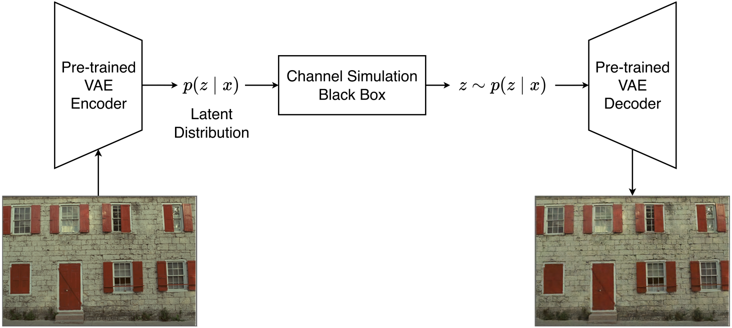




JPEG

Image credits Ballé et al. (2017).

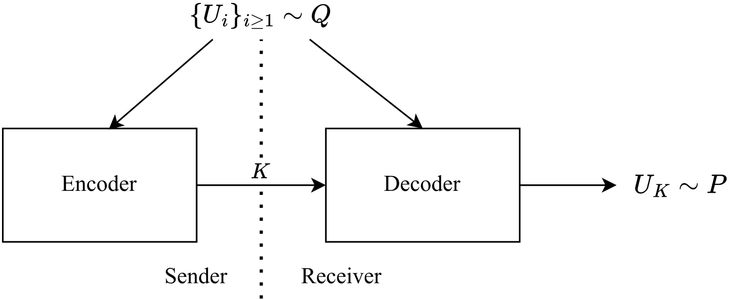
Neural Compression with Channel Simulation



- Fully differentiable end-to-end system!
- Channel simulation $\Longrightarrow$ **Relative entropy coding**

Image credits Flamich et al. (2020).

Exact Sampling

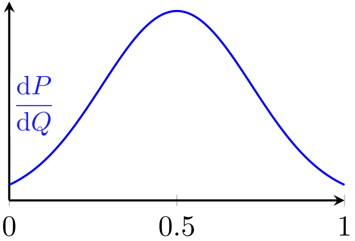


Key Questions:

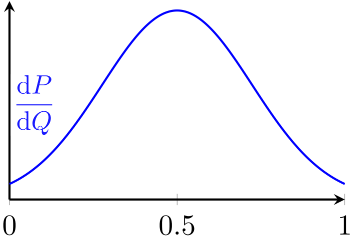
- How can we choose K such that $U_K \sim P$ *exactly*?
- How close can we get to $D(P||Q)$?

Greedy Rejection Sampling

- **Rejection sampling:** Accept U_k with probability $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(U_k)$, $\gamma > 0$ s.t. $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(u) \leq 1$ for all u.
- **Greedy rejection sampling:** Accept U_k with probability $f_k(U_k)$, for function f_k which maximizes the acceptance probability at stage k under the condition that the scheme is exact.



Rejection sampling

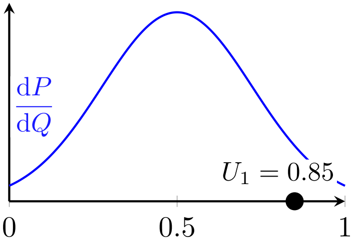


Greedy rejection sampling

$$P = \mathcal{N}(0.5, 0.05)|_{[0,1]},\ Q = \text{Uniform}([0,1]),\ \gamma = 0.55.$$

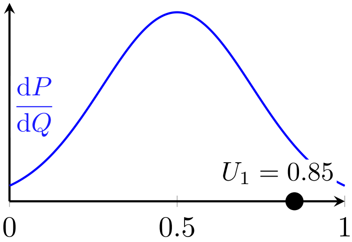
Greedy Rejection Sampling

- **Rejection sampling:** Accept U_k with probability $\gamma \frac{dP}{dQ}(U_k)$, $\gamma > 0$ s.t. $\gamma \frac{dP}{dQ}(u) \leq 1$ for all u.
- **Greedy rejection sampling:** Accept U_k with probability $f_k(U_k)$, for function f_k which maximizes the acceptance probability at stage k under the condition that the scheme is exact.



Rejection sampling

$\mathbb{P}(\text{Accept}) = \gamma \frac{dP}{dQ}(U_1) = 0.275$

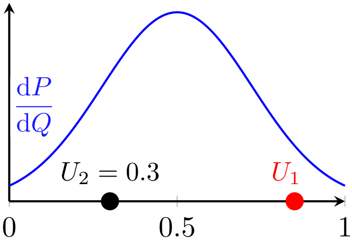


Greedy rejection sampling

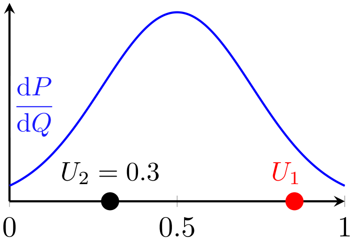
$\mathbb{P}(\text{Accept}) = \left(\frac{dP}{dQ}(U_1) - 0 \right) / 1 = 0.5$

Greedy Rejection Sampling

- **Rejection sampling:** Accept U_k with probability $\gamma\frac{\mathrm{d}P}{\mathrm{d}Q}(U_k)$, $\gamma > 0$ s.t. $\gamma\frac{\mathrm{d}P}{\mathrm{d}Q}(u) \leq 1$ for all u.
- **Greedy rejection sampling:** Accept U_k with probability $f_k(U_k)$, for function f_k which maximizes the acceptance probability at stage k under the condition that the scheme is exact.



Rejection sampling

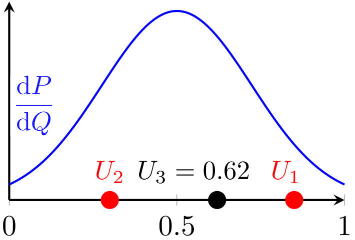


Greedy rejection sampling

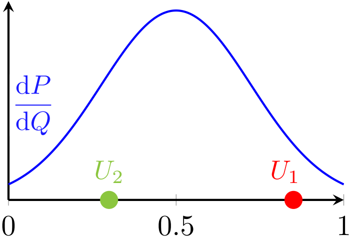
$\mathbb{P}(\text{Accept}) = \gamma\frac{\mathrm{d}P}{\mathrm{d}Q}(U_2) = 0.67$ $\quad$ $\mathbb{P}(\text{Accept}) = \frac{1}{0.255}\left(\frac{\mathrm{d}P}{\mathrm{d}Q}(U_2) - 1\right) = 0.89$

Greedy Rejection Sampling

- **Rejection sampling:** Accept U_k with probability $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(U_k)$, $\gamma > 0$ s.t. $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(u) \leq 1$ for all u.
- **Greedy rejection sampling:** Accept U_k with probability $f_k(U_k)$, for function f_k which maximizes the acceptance probability at stage k under the condition that the scheme is exact.



Rejection sampling

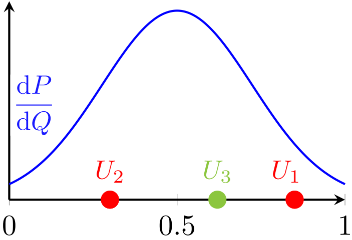


Greedy rejection sampling

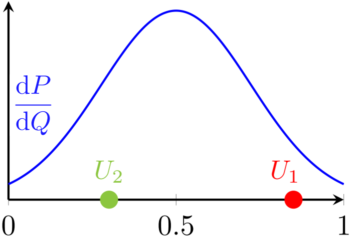
$\mathbb{P}(\text{Accept}) = \gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(U_2) = 0.87$

Greedy Rejection Sampling

- **Rejection sampling:** Accept U_k with probability $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(U_k)$, $\gamma > 0$ s.t. $\gamma \frac{\mathrm{d}P}{\mathrm{d}Q}(u) \leq 1$ for all u.
- **Greedy rejection sampling:** Accept U_k with probability $f_k(U_k)$, for function f_k which maximizes the acceptance probability at stage k under the condition that the scheme is exact.



Rejection sampling



Greedy rejection sampling

GRS: $D(P||Q) \leq \mathbb{E}[|M|] \leq D(P||Q) + \log_2(D(P||Q) + 1) + 4$

Recent Tighter Bounds

Recently, Goc and Flamich (2024) showed a tight bound on the expected message length:

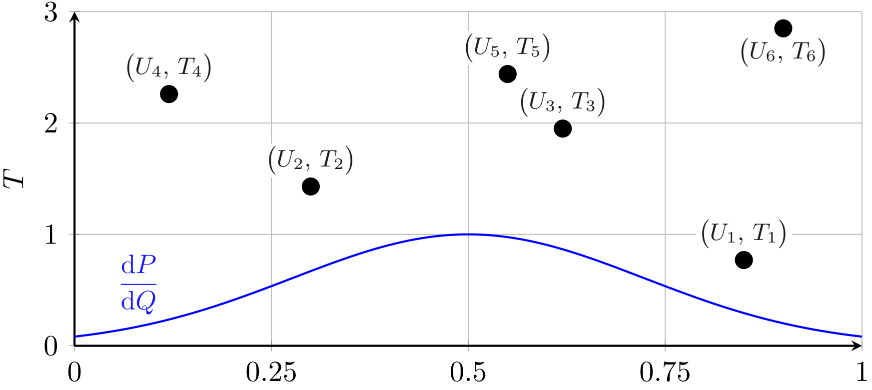
$$D(P||Q) \leq D_{CS}(P||Q) \leq \mathbb{E}[|M|] \leq D_{CS}(P||Q) + \log_2(e+1) + 1$$

for $D_{CS}(P||Q)$ the *channel simulation divergence.*

The upper bound on $\mathbb{E}[|M|]$ is achieved using greedy rejection sampling.

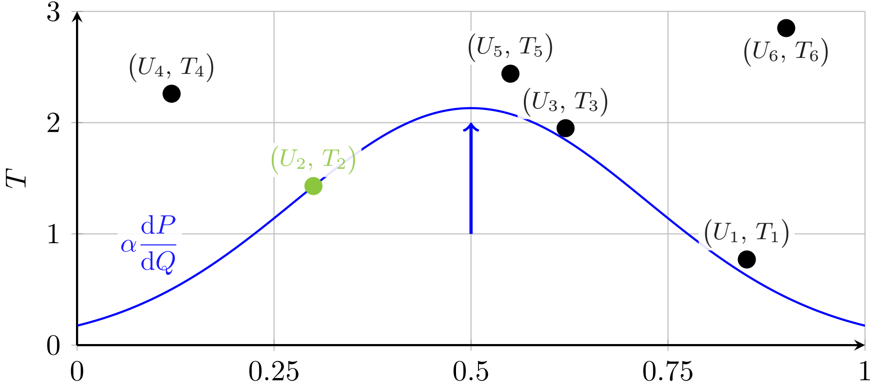
Poisson Functional Representation

For $\{T_i\}_{i \geq 1}$ a rate-one Poisson process, choose $K = \arg\min_{i \geq 1} \frac{T_i}{\frac{dP}{dQ}(U_i)}$, Li and El-Gamal (2018).



Poisson Functional Representation

For $\{T_i\}_{i\geq 1}$ a rate-one Poisson process, choose $K = \arg\min_{i\geq 1} \frac{T_i}{\frac{dP}{dQ}(U_i)}$, Li and El-Gamal (2018).



$$D(P||Q) \leq \mathbb{E}[|M|] \leq D(P||Q) + \log_2(D(P||Q) + 2) + 3$$

Our Setup: Exponential Cost and Rényi's entropy

- The previous results are for the expected message length (number of bits) $\mathbb{E}[|M|]$.
- What are the fundamental limits of exact sampling under a cost which is *exponential* in the message lengths? Can these limits be (almost) achieved by existing algorithms?

Campbell Cost $L(t)$

For uniquely decodable binary encoding $M \in \{0,1\}^*$ of K having length $|M|$ and for $t > 0$,

$$L(t) = \frac{1}{t} \log\Big(\mathbb{E}[2^{t|M|}]\Big).$$

Facts:

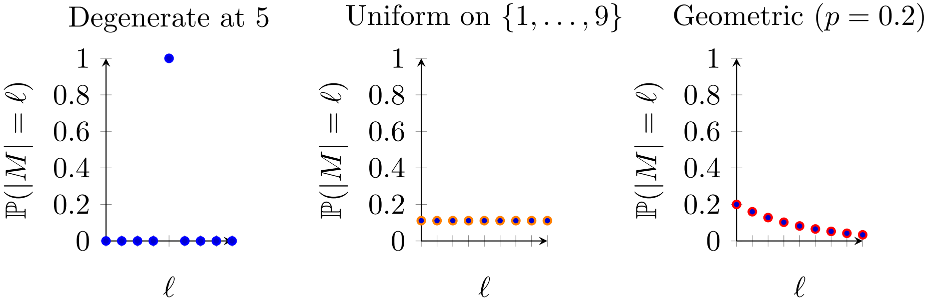
$$\lim_{t \to 0} L(t) = \mathbb{E}[|M|] \qquad \text{and} \qquad \lim_{t \to \infty} L(t) = \max_{\ell \in \mathbb{N}\,:\,\mathbb{P}(|M| = \ell) > 0} \ell$$

For a random variable K with Rényi entropy $H_\alpha(K)$ encoded optimally into message M, Campbell (1965) showed

$$H_\alpha(K) \le L(t) < H_\alpha(K) + 1$$

with $\alpha = \frac{1}{1+t}$.

Why Care About $L(t)$?



t	**Degenerate** $L(t)$	**Uniform** $L(t)$	**Geometric** $L(t)$
0	5	5	5
0.2	5	5.65	11.83
1	5	7.26	∞
5	5	8.56	∞
∞	5	9	∞

Lower Bound

Theorem 1 For any sampling algorithm and $t > 0$, with $\alpha = \frac{1}{1+t}$,

$$L(t) \geq D_{\frac{1}{\alpha}}(P||Q) + \frac{\alpha}{1-\alpha}\log_2(\alpha) - 1. \tag{LB}$$

As $t \to 0$, we recover the lower bound

$$\mathbb{E}[|M|] \geq D(P||Q) - \frac{1}{\ln(2)} - 1.$$

Upper Bounds via Poisson Functional Representation

Theorem 2 For K chosen using the Poisson functional representation, for any $\epsilon > 0$ there exists a uniquely decodable encoding of K such that

$$L(t) \leq (1+\epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P||Q) + c(\alpha, \epsilon), \tag{UB$_1$}$$

with $c(\alpha, \epsilon)$ a constant and $\alpha = \frac{1}{1+t}$.

Upper Bounds via Poisson Functional Representation

Theorem 3 Encoding K (generated by the PFR) using the Elias omega code gives, for any $0 < t < 1/2$ and $\epsilon \le \frac{1}{2t} - 1$,

$$L(t) \le D_{\frac{2-\alpha}{\alpha}}(P||Q) + (1+\epsilon)\log_2(D(P||Q)+1) + c_\epsilon. \qquad (\mathrm{UB}_2)$$

Recovers the bound

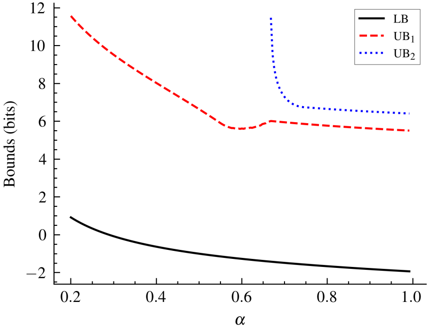
$$\mathbb{E}[|M|] \le D(P||Q) + (1+\epsilon)\log_2(D(P||Q)+1) + c_\epsilon$$

of Harsha et al. (2010) as $t \to 0$.

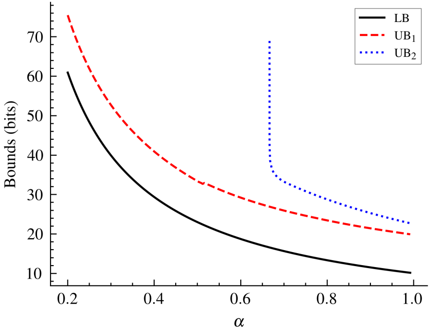
Proof Techniques

- For $\mathbb{E}[|M|]$ and $L(t)$ the lower bounds $D(P\|Q)$ resp. $D_{\frac{1}{\alpha}}(P\|Q)$ are simple to prove.
- The upper bound(s) on $\mathbb{E}[|M|]$ are derived through *sharply* bounding $\mathbb{E}[\,\log_2 K\,]$.
- The upper bounds on $L(t)$ are derived (more or less) by bounding $\mathbb{E}[K^t]$.

Gaussian Examples



$P = \mathcal{N}(0,1)$ and $Q = \mathcal{N}(1,1)$



$P = \mathcal{N}(0,1)$ and $Q = \mathcal{N}(5,1)$

Asymptotic Results

- We want to use the channel n-times with i.i.d. input $X_1, \ldots, X_n$. Thus we sample from the product distribution $P^{\otimes n}$ using samples from $Q^{\otimes n}$.
- We can now fully characterize the optimal $L(t)/n$ as $n \to \infty$:

Theorem 4 For any $t > 0$, let $L_n^*(t)$ be the *minimum Campbell cost* for target $P^{\otimes n}$ and common randomness $\{U_i\}_{i \geq 1} \sim Q^{\otimes n}$. Then, with $\alpha = \frac{1}{1+t}$,

$$\lim_{n \to \infty} \frac{L_n^*(t)}{n} = D_{\frac{1}{\alpha}}(P||Q).$$

This generalizes known results: for the *minimum bits/sample* rate R_n^* for the n-dimensional product distributions,

$$\lim_{n \to \infty} \frac{R_n^*}{n} = D(P||Q).$$

Causal vs. Noncausal Sampling

- A *causal* sampler accepts/rejects each candidate one-at-a-time (K is a stopping time w.r.t. $\{U_i\}_{i \geq 1}$).
- Greedy rejection sampling ✓ Poisson functional representation **X**
- GRS and the PFR **both achieve** bits/sample rate $D(P||Q)$ as $n \to \infty$.

Theorem 5 For any $t > 0$ let $L_n^*(t)$ be the minimum Campbell cost over *causal* samplers between $P^{\otimes n}$ and $Q^{\otimes n}$. Then, with $\alpha = \frac{1}{1+t}$,

$$\liminf_{n \to \infty} \frac{L_n^*(t)}{n} \geq D_\beta(P||Q), \quad \text{where } \beta = \begin{cases} \frac{\alpha}{2\alpha - 1}, & \alpha \in (1/2, 1) \\ \infty, & \alpha \in (0, 1/2]. \end{cases}$$

- $D_\beta(P||Q) > D_{\frac{1}{\alpha}}(P||Q)$ in general!!!
- Greedy rejection sampling does **strictly worse** in the **exponential cost regime**, and the gap is often significant.

Main Takeaways

- Exact sampling is one (highly general) way to perform channel simulation at a near-optimal encoding cost, and has wide applications.
- The Campbell cost $L(t)$ generalizes the expected message length and can be made more sensitive to the tails of the distribution.
- The Poisson functional representation is nearly optimal for exact sampling (typically within 5-10 bits) for the Campbell cost.
- Causal samplers (such as greedy rejection sampling, greedy Poisson rejection sampling, etc.) do **strictly worse** than noncausal samplers in the asymptotic Campbell cost.

References

J. Ballé, V. Laparra, and E. P. Simoncelli, "End-to-end optimized image compression," *5th International Conference on Learning Representations*, 2017.

L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.

G. Flamich, M. Havasi, and J. M. Hernández-Lobato, "Compressing images by encoding their latent representations with relative entropy coding," *Advances in Neural Information Processing Systems*, vol. 33, pp. 16 131–16 141, 2020.

P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 438–449, 2010.

E. Lei, H. Hassani, S. S. Bidokhti, "Neural estimation of the rate-distortion function with applications to operational source coding," *IEEE Journal on Selected Areas in Information Theory*, no. 4, pp. 674-686, 2023.

D. Goc and G. Flamich, "On channel simulation with causal rejection samplers," in *IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2024, pp. 1682–1687.

C. T. Li and A. El-Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 4, pp. 2583–2592, 2018.

References

G. Flamich, S. M. Sriramu, and A. B. Wagner, "The redundancy of non-singular channel simulation," *arXiv preprint arXiv:2501.14053*, 2025.

C. T. Li, *Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy.* Now Publishers, Inc., 2024, vol. 21, no. 6.

G. Flamich, "Greedy Poisson rejection sampling," *Advances in Neural Information Processing Systems*, vol. 36, pp. 37 089–37 127, 2023.

J. Liu and S. Verdú, "Rejection sampling and noncausal sampling under moment constraints," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 1565–1569.

G. Flamich and L. Theis, "Adaptive greedy rejection sampling," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 454–459.

C. T. Li and V. Anantharam, "A unified framework for one-shot achievability via the Poisson matching lemma," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2624–2651, 2021.

S. Hill, F. Alajaji, and T. Linder, "Communication complexity of exact sampling under Rényi information," *arXiv preprint arXiv:2506.12219*, 2025.