Exact Channel Simulation under Exponential Cost

Spencer Hill

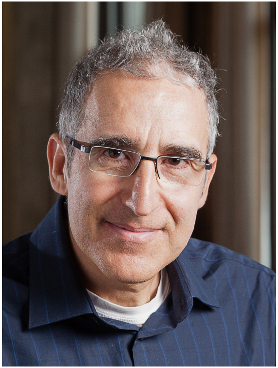
Queen's University, Canada

Allerton Conference

September 18, 2025

Joint work with

Tamás Linder



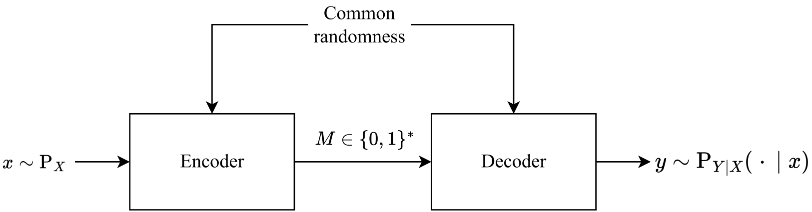
Fady Alajaji

Queen's University

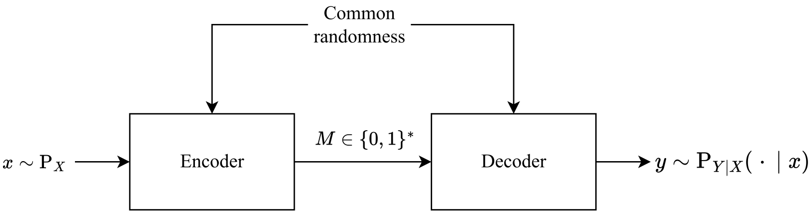
Outline

1. What is channel simulation?
2. Interesting applications
3. Channel simulation algorithms and performance
4. Exponential (Campbell) cost
5. Our results

Channel Simulation

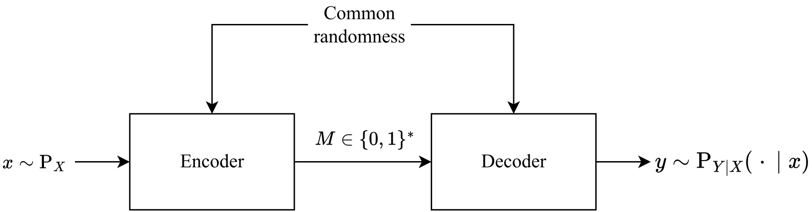


Channel Simulation



- Use noiseless channel to simulate noisy channel $X \to Y$

Channel Simulation



- Use noiseless channel to simulate noisy channel $X \to Y$
- When the goal is to efficiently communicate M, one can achieve

$$\mathbb{E}|M| \approx I(X;Y) \quad \text{bits}$$

Why Care?

Neural Compression via Nonlinear Transform Coding

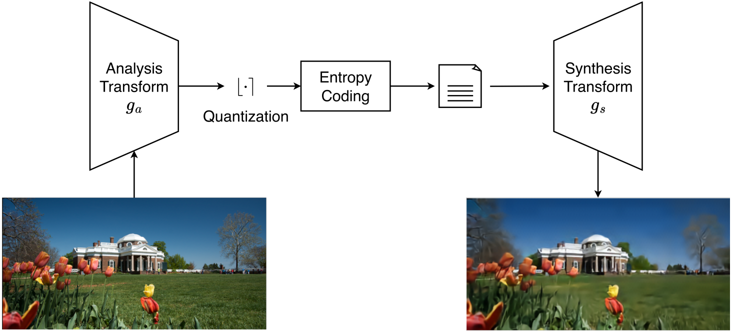
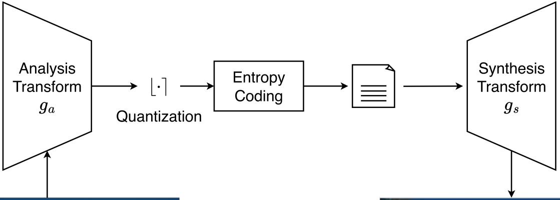
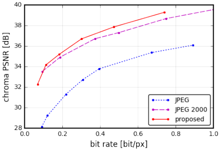


Image credits Ballé et al. (2017).

Neural Compression via Nonlinear Transform Coding

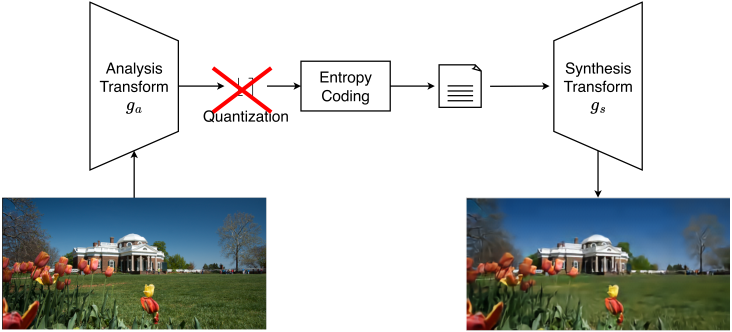


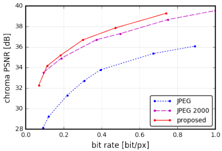


JPEG

Image credits Ballé et al. (2017).

Neural Compression via Nonlinear Transform Coding





JPEG

Image credits Ballé et al. (2017).

Neural Compression with Channel Simulation

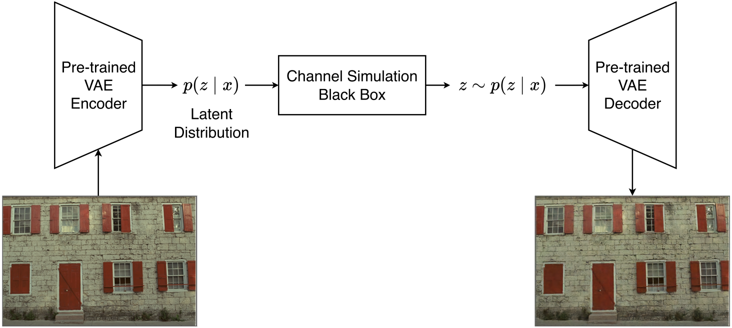
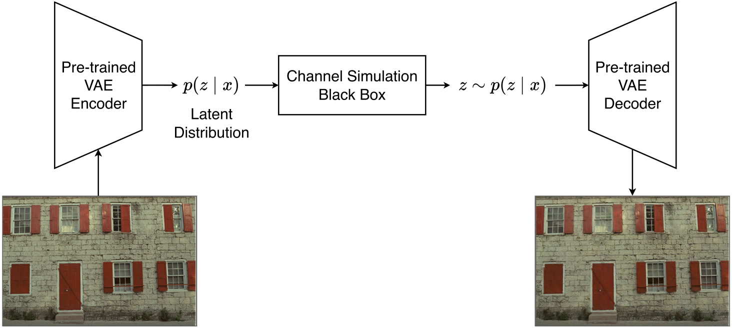


Image credits Flamich et al. (2020).

Neural Compression with Channel Simulation



- Fully differentiable end-to-end system trained via the reparameterization trick!

Image credits Flamich et al. (2020).

Other Applications

- Lossy source coding

Other Applications

- Lossy source coding
- Rate-distortion-perception tradeoff

Other Applications

- Lossy source coding
- Rate-distortion-perception tradeoff
- Compression via implicit neural representation

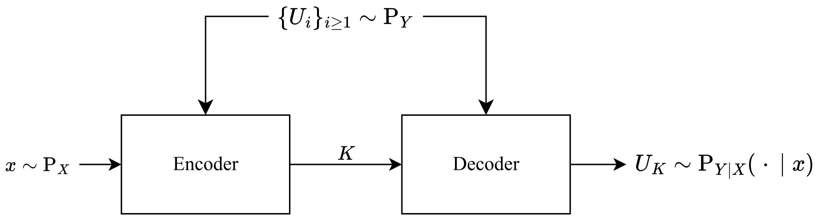
Other Applications

- Lossy source coding
- Rate-distortion-perception tradeoff
- Compression via implicit neural representation
- Local differential privacy

Other Applications

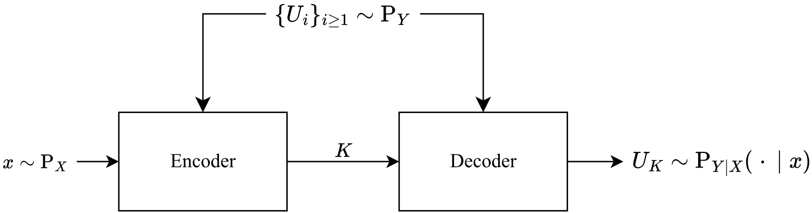
- Lossy source coding
- Rate-distortion-perception tradeoff
- Compression via implicit neural representation
- Local differential privacy
- Federated learning, . . .

Channel Simulation from Sampling



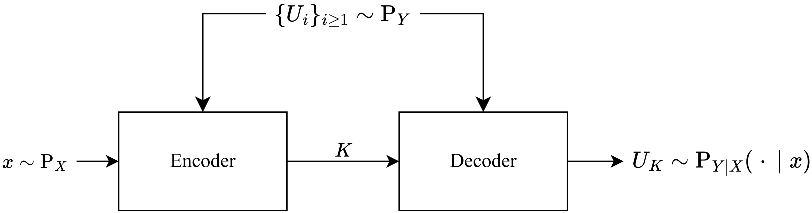
- Common randomness is i.i.d. sequence $\{U_i\}_{i\geq 1} \sim \mathrm{P}_Y$

Channel Simulation from Sampling



- Common randomness is i.i.d. sequence $\{U_i\}_{i\geq 1} \sim \mathrm{P}_Y$
- Transmit index K such that $U_K \sim \mathrm{P}_{Y|X}$

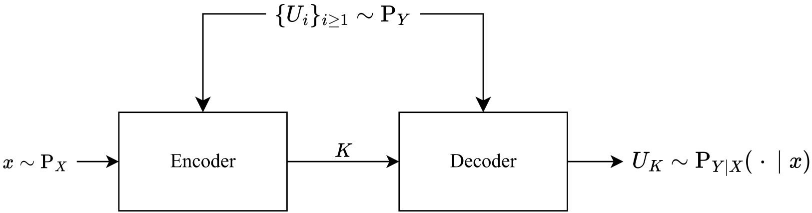
Channel Simulation from Sampling



- Common randomness is i.i.d. sequence $\{U_i\}_{i\geq 1} \sim \mathrm{P}_Y$
- Transmit index K such that $U_K \sim \mathrm{P}_{Y|X}$
- **Sampling:** for sampling from general P given access to sequence from Q, one can achieve

$$\mathbb{E}|M| \approx D(P||Q) \quad \text{bits}$$

Channel Simulation from Sampling



- Common randomness is i.i.d. sequence $\{U_i\}_{i\geq 1} \sim \mathrm{P}_Y$
- Transmit index K such that $U_K \sim \mathrm{P}_{Y|X}$
- **Sampling:** for sampling from general P given access to sequence from Q, one can achieve

$$\mathbb{E}|M| \approx D(P||Q) \quad \text{bits}$$

- Sampling can simulate $X \to Y$ with communication cost

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

Sampling Algorithms

- **Greedy Rejection Sampling:** Accept/reject each U_k sequentially, maximizing the acceptance probability each stage under the condition that the scheme is exact.

Sampling Algorithms

- **Greedy Rejection Sampling:** Accept/reject each U_k sequentially, maximizing the acceptance probability each stage under the condition that the scheme is exact.

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

Sampling Algorithms

- **Greedy Rejection Sampling:** Accept/reject each U_k sequentially, maximizing the acceptance probability each stage under the condition that the scheme is exact.

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

- **Poisson Functional Representation:** For $\{T_i\}_{i\geq 1}$ a rate-one Poisson process, choose $K = \arg\min_{i\geq 1} \frac{T_i}{\frac{\mathrm{d}P}{\mathrm{d}Q}(U_i)}$.

Sampling Algorithms

- **Greedy Rejection Sampling:** Accept/reject each U_k sequentially, maximizing the acceptance probability each stage under the condition that the scheme is exact.

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

- **Poisson Functional Representation:** For $\{T_i\}_{i \geq 1}$ a rate-one Poisson process, choose $K = \arg\min_{i \geq 1} \frac{T_i}{\frac{\mathrm{d}P}{\mathrm{d}Q}(U_i)}$.

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

Our Setup: Exponential Cost and Rényi's entropy

- The previous results are for the expected message length (number of bits) $\mathbb{E}|M|$.

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 and channel simulation 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 $t > 0$,

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

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:

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

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:

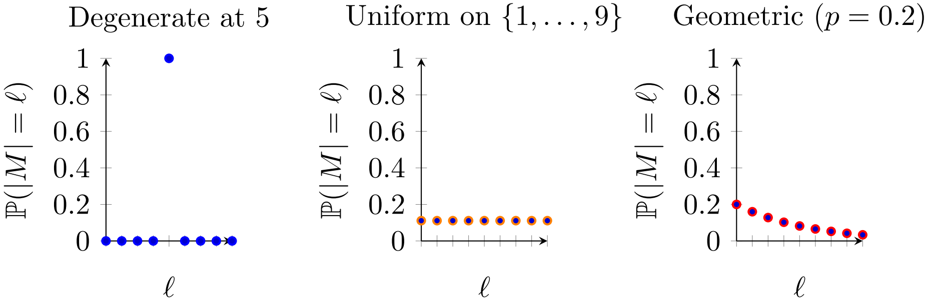
$$\lim_{t \to 0} L(t) = \mathbb{E}|M| \qquad \text{and} \qquad \lim_{t \to \infty} L(t) = \max_{\ell \in \mathbb{N} \,:\, \mathbb{P}(|M| = \ell) > 0} \ell$$

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

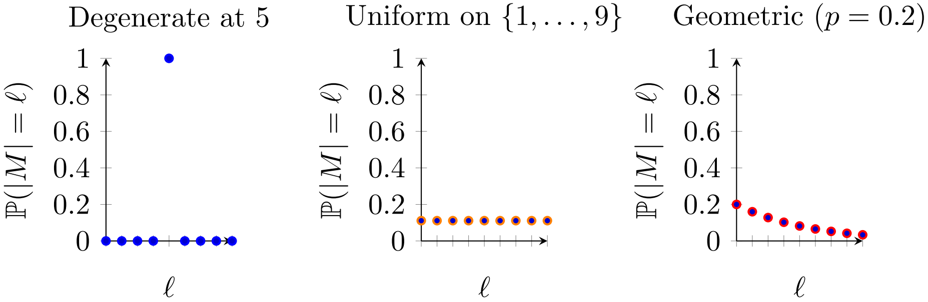
$$H_\alpha(K) \leq L(t) < H_\alpha(K) + 1$$

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

Why Care About $L(t)$?



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 $\alpha = \frac{1}{1+t}$,

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

Lower Bound

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

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

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

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

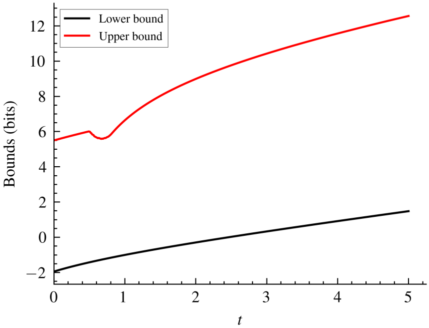
Upper Bound via Poisson Functional Representation

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

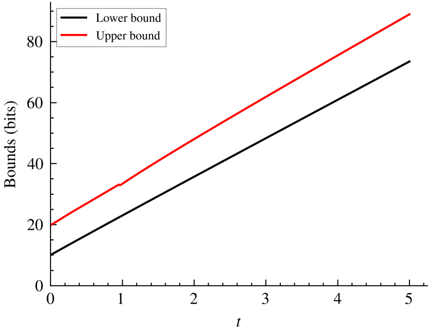
$$L(t) \leq (1+\epsilon) D_{\frac{1+\epsilon(1-\alpha)}{\alpha}}(P||Q) + c(\alpha, \epsilon), \tag{2}$$

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

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 $Q^{\otimes n}$.

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 fully characterize the optimal $L(t)/n$ as $n \to \infty$:

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 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 $\alpha = \frac{1}{1+t}$,

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

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 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. $\{U_i\}_{i \geq 1}$).

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 ✗

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 $n \to \infty$.

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 $\alpha = \frac{1}{1+t}$,

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

Causal vs. Noncausal Sampling

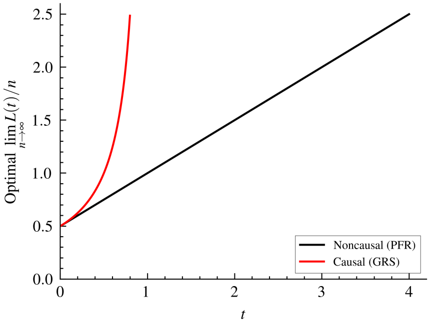
- A *causal* sampler accepts/rejects each candidate one-at-a-time (K is a stopping time w.r.t. $\{U_i\}_{i\geq 1}$).
- Greedy rejection sampling ✓ Poisson functional representation X
- GRS and the PFR **both achieve** bits/sample rate $D(P||Q)$ as $n \to \infty$.

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 $\alpha = \frac{1}{1+t}$,

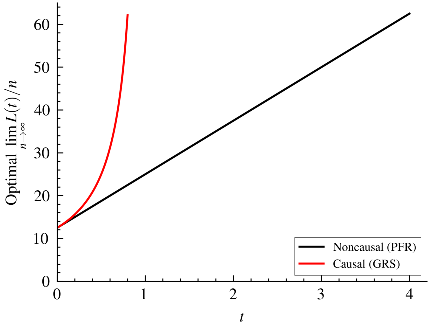
$$\liminf_{n\to\infty} \frac{L_n^*(t)}{n} \geq D_\beta(P||Q), \quad \text{where } \beta = \begin{cases} \frac{\alpha}{2\alpha-1}, & \alpha \in (1/2, 1) \\ \infty, & \alpha \in (0, 1/2]. \end{cases}$$

- $D_\beta(P\|Q) > D_{\frac{1}{\alpha}}(P\|Q)$ in general!!!

Asymptotic Gaussian Examples



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



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

Greedy rejection sampling does **strictly worse** in the **exponential cost regime**, and the gap is often significant.

Main Takeaways

Main Takeaways

- Channel simulation is a theoretically and practically interesting problem.

Main Takeaways

- Channel simulation is a theoretically and practically interesting problem.
- Sampling is a (highly general) way to perform channel simulation at a near-optimal encoding cost.

Main Takeaways

- Channel simulation is a theoretically and practically interesting problem.
- Sampling is a (highly general) way to perform channel simulation at a near-optimal encoding cost.
- The Campbell cost $L(t)$ generalizes the expected message length and can be made more sensitive to the tails of the distribution.

Main Takeaways

- Channel simulation is a theoretically and practically interesting problem.
- Sampling is a (highly general) way to perform channel simulation at a near-optimal encoding cost.
- The Campbell cost $L(t)$ generalizes the expected message length and can be made more sensitive to the tails of the distribution.
- Under the Campbell cost, the Poisson functional representation is nearly optimal for exact sampling.

Main Takeaways

- Channel simulation is a theoretically and practically interesting problem.
- Sampling is a (highly general) way to perform channel simulation at a near-optimal encoding cost.
- The Campbell cost $L(t)$ generalizes the expected message length and can be made more sensitive to the tails of the distribution.
- Under the Campbell cost, the Poisson functional representation is nearly optimal for exact sampling.
- 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.