ADAPTIVE GREEDY REJECTION SAMPLING

GERGELY FLAMICH AND LUCAS THEIS

26/06/2023

GERGELY-FLAMICH.GITHUB.IO

1. IN COLLABORATION WITH

Lucas Theis

2. BACKGROUND

2.1. CHANNEL SIMULATION

2.1. CHANNEL SIMULATION

Correlated r.v.s $X, Y \sim P_{X,Y}$

2.1. CHANNEL SIMULATION

Correlated r.v.s $X, Y \sim P_{X,Y}$

Alice receives $Y \sim P_Y$

2.1. CHANNEL SIMULATION

Correlated r.v.s $X, Y \sim P_{X,Y}$

Alice receives $Y \sim P_Y$

Bob wants to simulate $X \sim P_{X|Y}$

2.1. CHANNEL SIMULATION

Correlated r.v.s $X, Y \sim P_{X,Y}$

Alice receives $Y \sim P_Y$

Bob wants to simulate $X \sim P_{X|Y}$

How many bits does Alice need to send to Bob?

2.1. CHANNEL SIMULATION

Correlated r.v.s $X, Y \sim P_{X,Y}$

Alice receives $Y \sim P_Y$

Bob wants to simulate $X \sim P_{X|Y}$

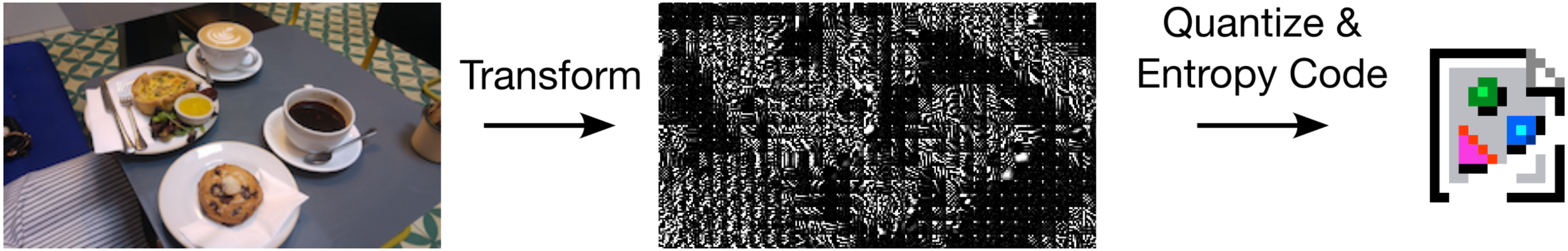
How many bits does Alice need to send to Bob?

When common randomness S available (SFRL):

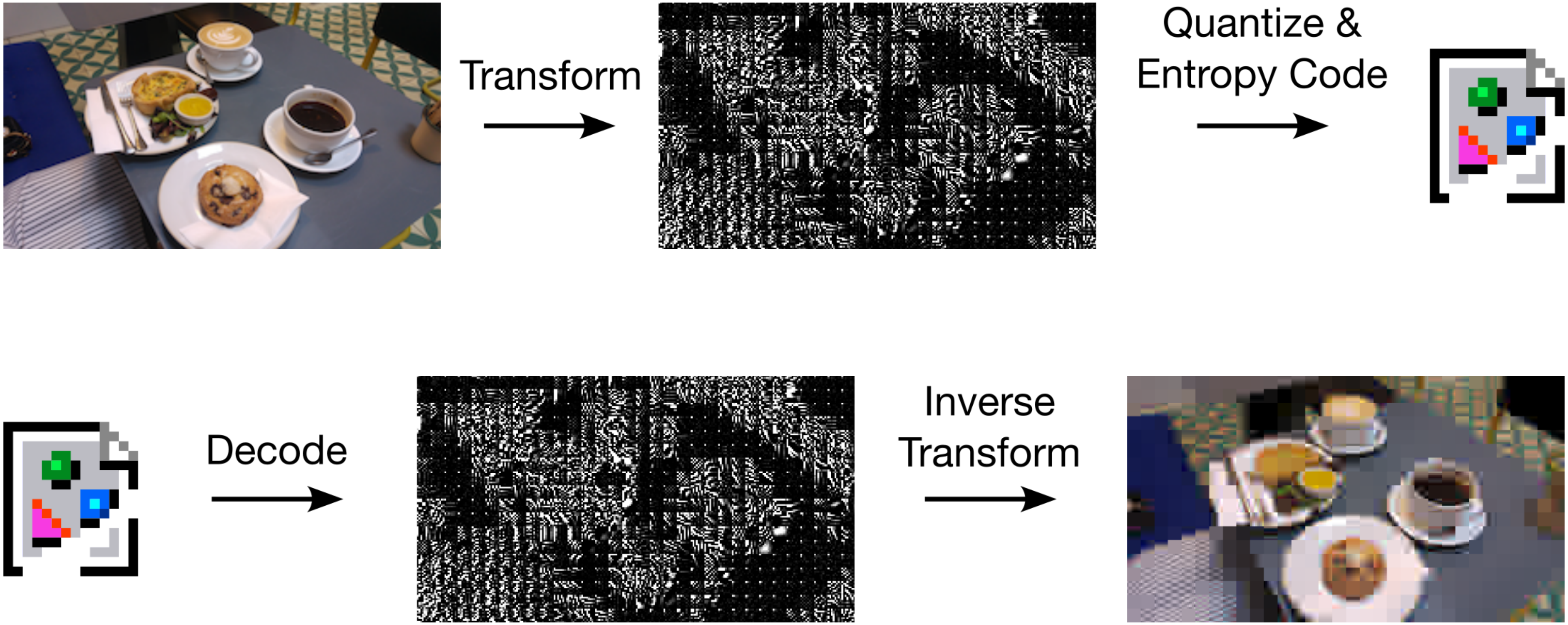
$$H[X \mid S] \leq I[X;Y] + \log(I[X;Y] + 1) + 4.$$

2.2. LOSSY COMPRESSION

2.2. LOSSY COMPRESSION



2.2. LOSSY COMPRESSION



3. GREEDY REJECTION SAMPLING

3.1. GREEDY REJECTION SAMPLING

3.1. GREEDY REJECTION SAMPLING

Originally proposed by Harsha et al. (2006) for discrete distributions.

3.1. GREEDY REJECTION SAMPLING

Originally proposed by Harsha et al. (2006) for discrete distributions.

We extend it to Borel probability measures over Polish spaces.

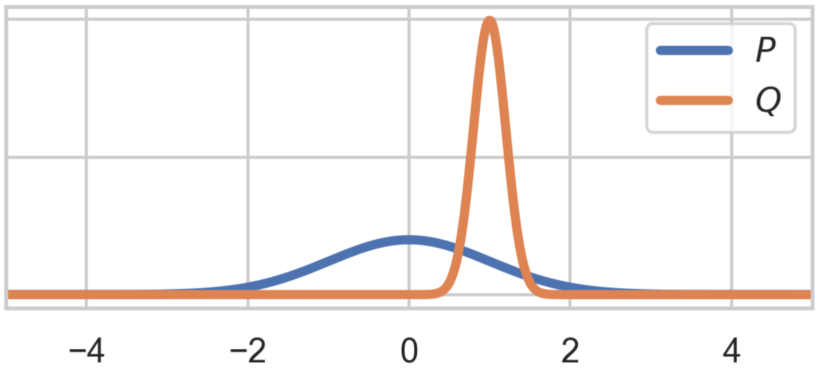
3.2. GRS EXAMPLE

3.2. GRS EXAMPLE

Target $Q = \mathcal{N}(\mu, \sigma^2)$, proposal $P = \mathcal{N}(0, 1)$

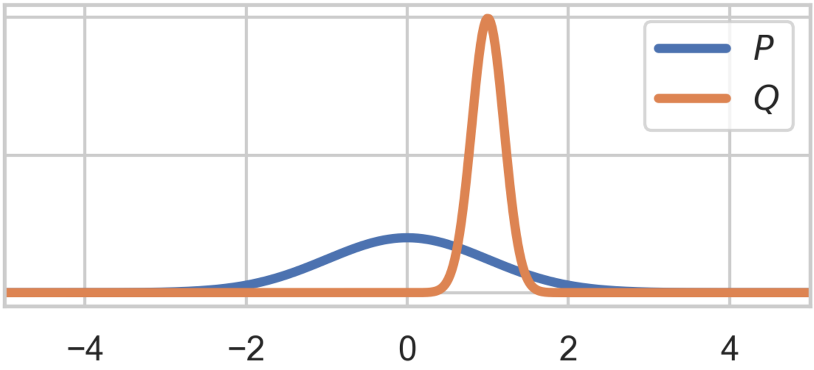
3.2. GRS EXAMPLE

Target $Q = \mathcal{N}(\mu, \sigma^2)$, proposal $P = \mathcal{N}(0, 1)$



3.2. GRS EXAMPLE

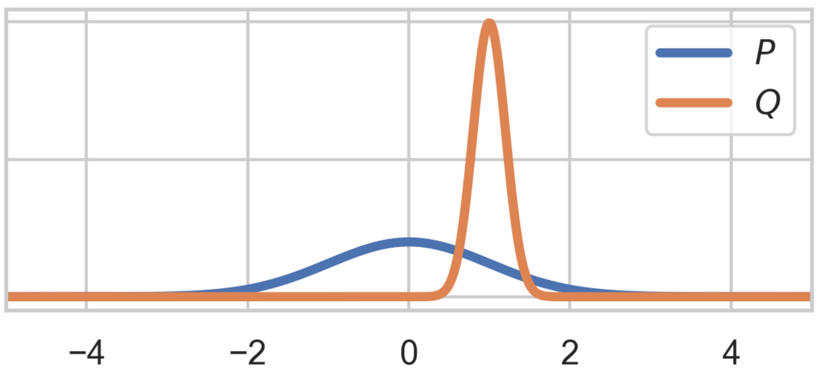
Target $Q = \mathcal{N}(\mu, \sigma^2)$, proposal $P = \mathcal{N}(0, 1)$



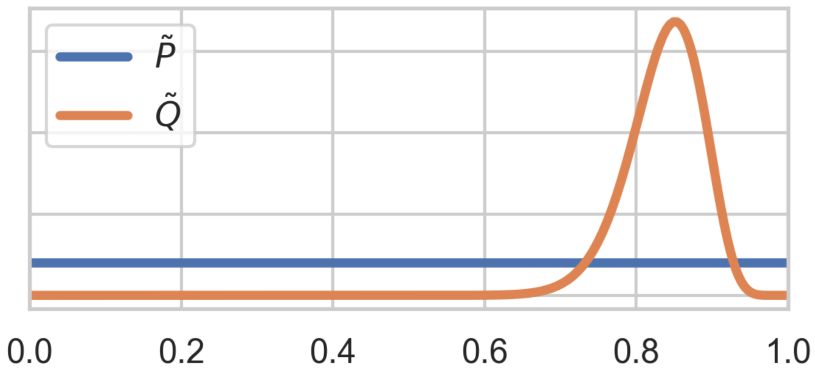
Change of variables: $U = F_P(X)$

3.2. GRS EXAMPLE

Target $Q = \mathcal{N}(\mu, \sigma^2)$, proposal $P = \mathcal{N}(0, 1)$

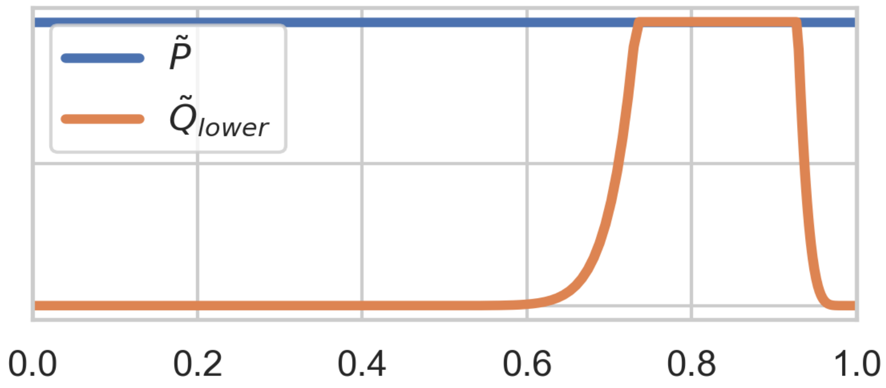


Change of variables: $U = F_P(X)$

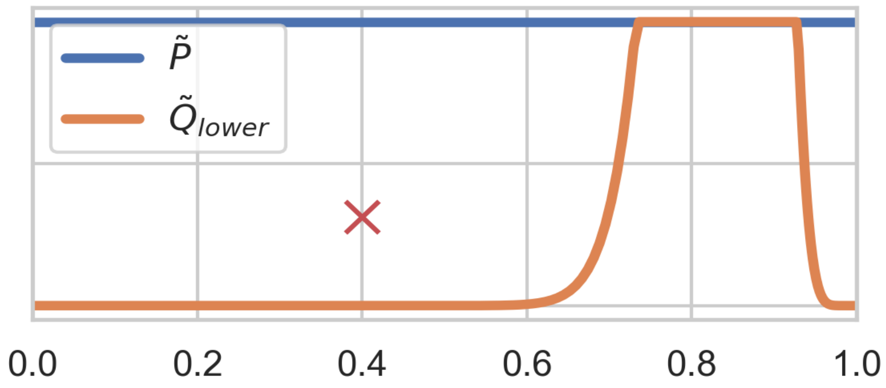


3.3. GRS EXAMPLE CONT'D

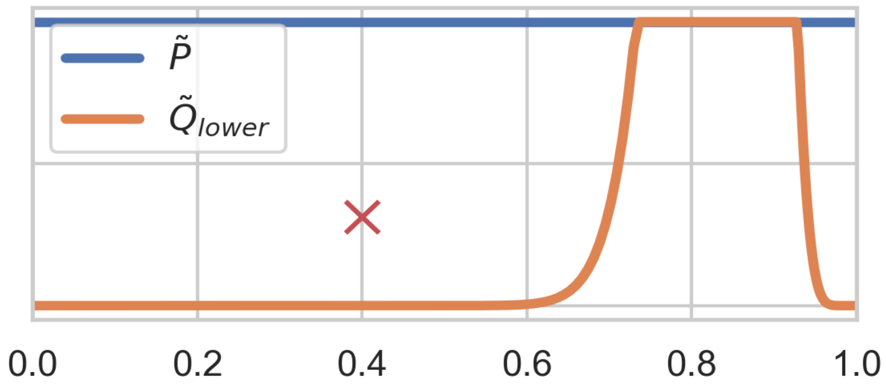
3.3. GRS EXAMPLE CONT'D



3.3. GRS EXAMPLE CONT'D

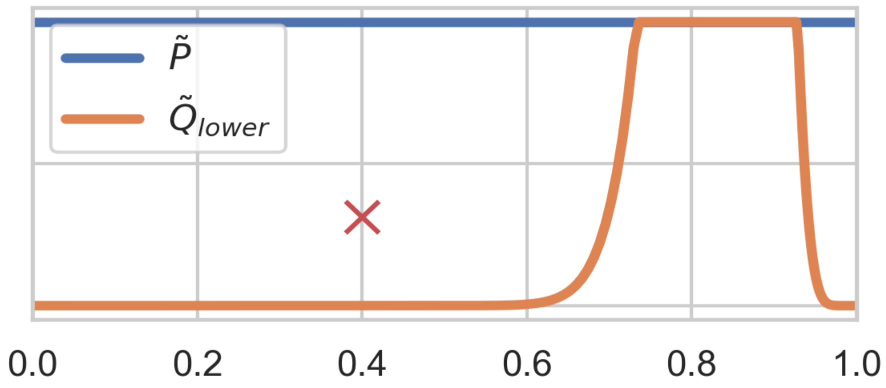


3.3. GRS EXAMPLE CONT'D

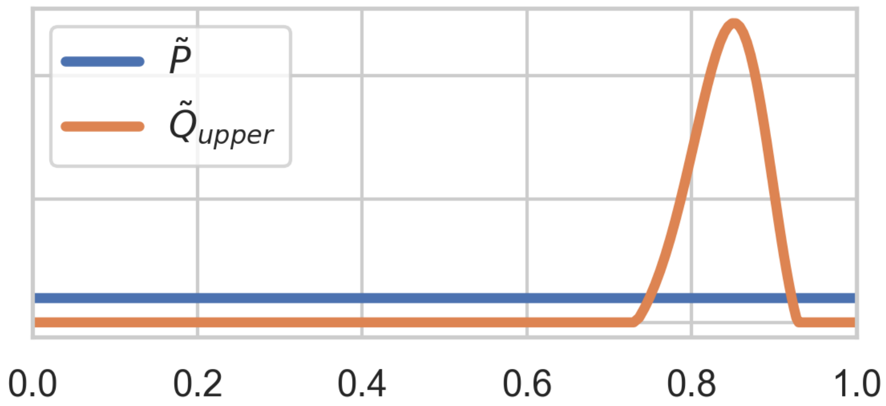


If rejected, normalize upper half and repeat:

3.3. GRS EXAMPLE CONT'D



If rejected, normalize upper half and repeat:



3.4. GRS CODING

3.4. GRS CODING

To encode sample:

3.4. GRS CODING

To encode sample:

Simulate $X_i \sim P$ with common randomness S

3.4. GRS CODING

To encode sample:

Simulate $X_i \sim P$ with common randomness S

Count number of rejections K

3.4. GRS CODING

To encode sample:

Simulate $X_i \sim P$ with common randomness S

Count number of rejections K

Transmit K using a Zeta coding distribution:

$$\zeta(k \mid \alpha) \propto k^{-\alpha}$$

3.5. GRS ANALYSIS

3.5. GRS ANALYSIS

For fixed Q and P:

$$H[K] \leq D_{KL}[Q \,\|\, P] + \log(D_{KL}[Q \,\|\, P] + 1) + 4$$

3.5. GRS ANALYSIS

For fixed Q and P:

$$H[K] \leq D_{KL}[Q \,\|\, P] + \log(D_{KL}[Q \,\|\, P] + 1) + 4$$

When $Q \leftarrow P_{X|Y}$ and $P \leftarrow P_X$, average over Y:

$$H[K] \leq I[X;Y] + \log(I[X;Y] + 1) + 4$$

3.6. GRS ANALYSIS CONT'D

3.6. GRS ANALYSIS CONT'D

$$\mathbb{E}[K+1] = 2^{D_\infty[Q \,\|\, P]}$$

where

$$D_\infty[Q \,\|\, P] = \log\left(\operatorname{ess\,sup}_{x \in \Omega}\left\{\frac{dQ}{dP}(x)\right\}\right)$$

3.7. THE GAUSSIAN CASE

3.7. THE GAUSSIAN CASE

$$X \mid \mu \sim \mathcal{N}(\mu, \rho^2 I), \quad \mu \sim \mathcal{N}(0, \sigma^2 I)$$

3.7. THE GAUSSIAN CASE

$$X \mid \mu \sim \mathcal{N}(\mu, \rho^2 I), \quad \mu \sim \mathcal{N}(0, \sigma^2 I)$$

Then:

$$P_X = \mathcal{N}(0, (\sigma^2 + \rho^2) I)$$

3.7. THE GAUSSIAN CASE

$$X \mid \mu \sim \mathcal{N}(\mu, \rho^2 I), \quad \mu \sim \mathcal{N}(0, \sigma^2 I)$$

Then:

$$P_X = \mathcal{N}(0, (\sigma^2 + \rho^2) I)$$

Now, we find:

$$\mathbb{E}[K+1] = \infty$$

3.8. THE GAUSSIAN CASE: OVERDISPERSION

3.8. THE GAUSSIAN CASE: OVERDISPERSION

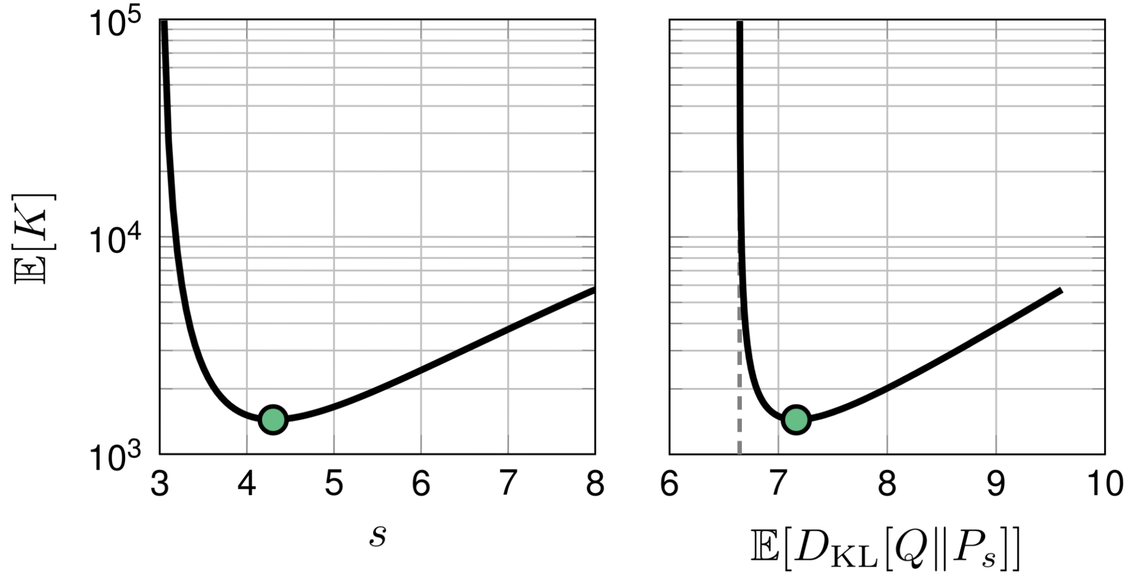
Use overdispersed marginal:

$$P_s = \mathcal{N}(0, (s^2 + \rho^2)I), \quad \sigma^2 < s^2$$

3.8. THE GAUSSIAN CASE: OVERDISPERSION

Use overdispersed marginal:

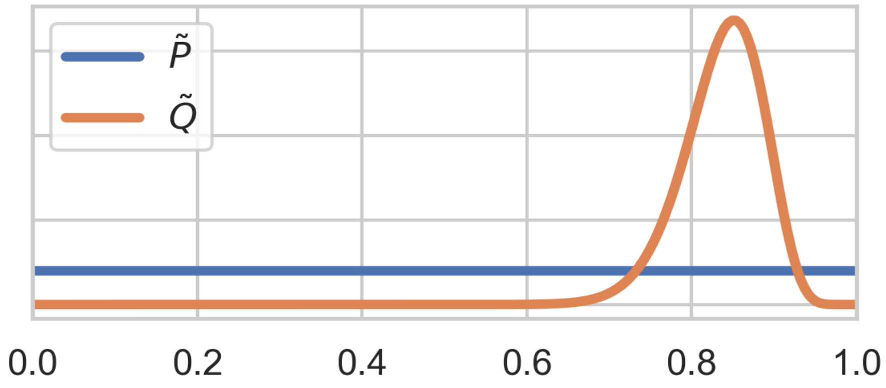
$$P_s = \mathcal{N}(0, (s^2 + \rho^2)I), \quad \sigma^2 < s^2$$



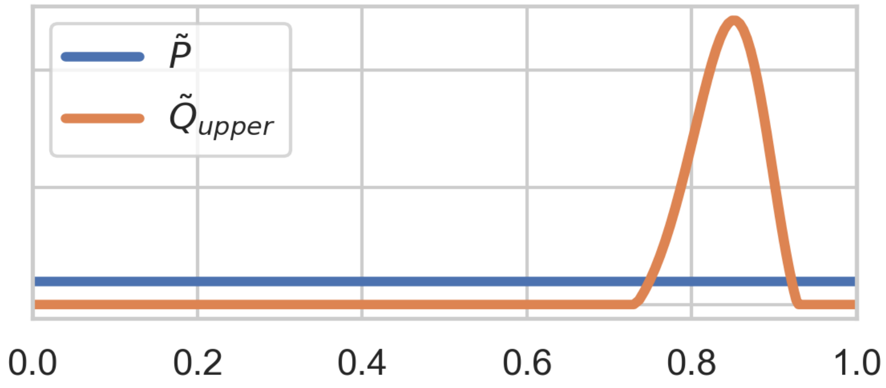
4. ADAPTIVE GREEDY REJECTION SAMPLING

4.1. GRS EXAMPLE REVISITED

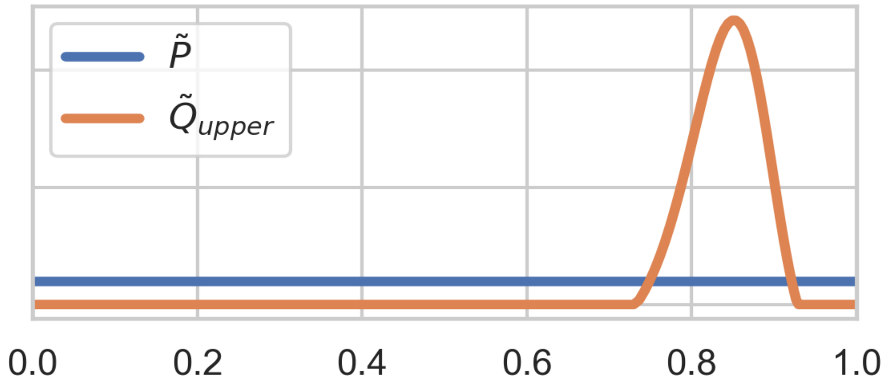
4.1. GRS EXAMPLE REVISITED



4.1. GRS EXAMPLE REVISITED

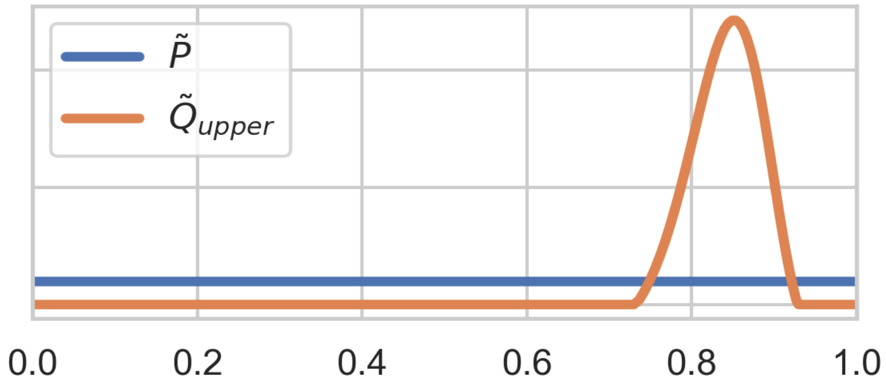


4.1. GRS EXAMPLE REVISITED

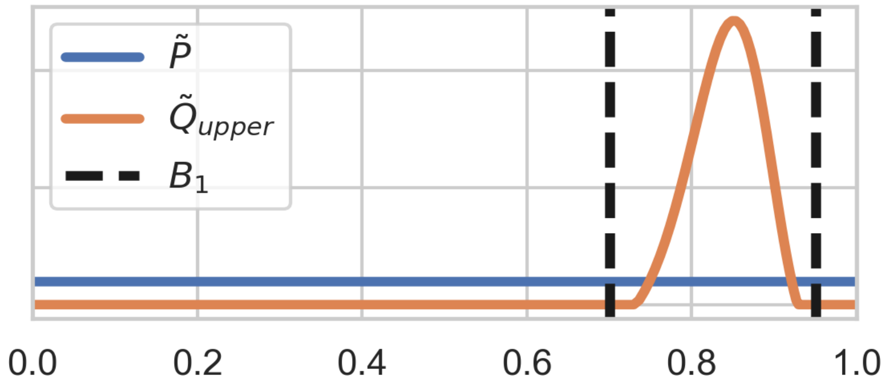


Most of sample space useless, *adapt proposal*:

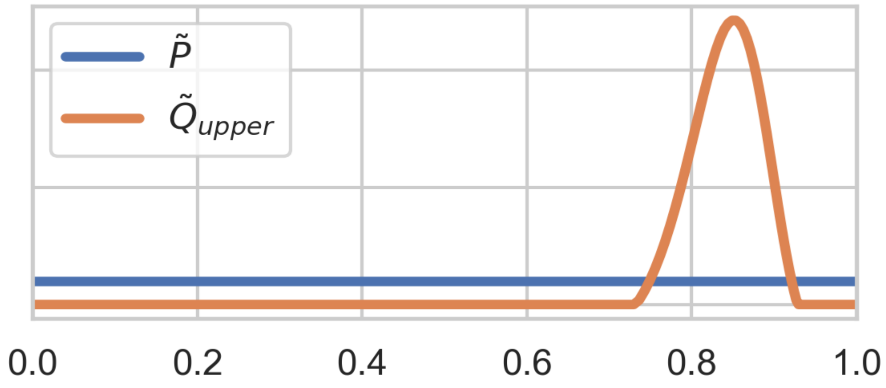
4.1. GRS EXAMPLE REVISITED



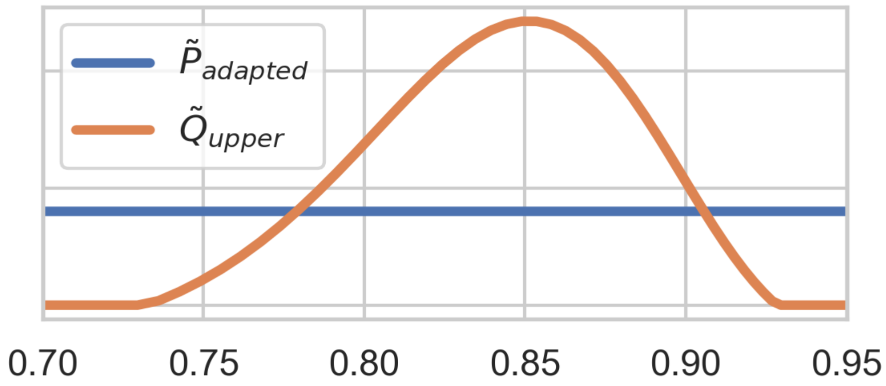
Most of sample space useless, *adapt proposal*:



4.1. GRS EXAMPLE REVISITED



Most of sample space useless, *adapt proposal*:



4.2. AGRS CODING

4.2. AGRS CODING

Need suitable sequence of bounds $B_1, B_2, \ldots$.

4.2. AGRS CODING

Need suitable sequence of bounds $B_1, B_2, \ldots$.

Need to communicate K, B_K!

4.2. AGRS CODING

Need suitable sequence of bounds $B_1, B_2, \ldots$.

Need to communicate K, B_K!

Theorem:

$$H[K] \leq C + \log(C+1) + 3.63,$$

where $C = I[X;Y] + \mathbb{E}[\log P(B_{m_K})]$.

4.3. AGRS WITH DITHERED QUANTIZATION

4.3. AGRS WITH DITHERED QUANTIZATION

$$\lfloor c + U \rceil - U \stackrel{d}{=} c + U'$$

where $U, U' \sim \mathrm{Unif}(-1/2, 1/2)$

4.3. AGRS WITH DITHERED QUANTIZATION

$$\lfloor c + U \rceil - U \overset{d}{=} c + U'$$

where $U, U' \sim \mathrm{Unif}(-1/2, 1/2)$

If $c \in [1/2, M - 3/2)$, then $\lfloor c + U \rceil \in [0 : M - 1]$.

4.3. AGRS WITH DITHERED QUANTIZATION

$$\lfloor c + U \rceil - U \overset{d}{=} c + U'$$

where $U, U' \sim \mathrm{Unif}(-1/2, 1/2)$

If $c \in [1/2, M - 3/2)$, then $\lfloor c + U \rceil \in [0 : M - 1]$.

$$\frac{\lfloor c + U \rceil - U}{M} \sim \mathrm{Unif}\left(\frac{c}{M} - \frac{1}{2M}, \frac{c}{M} + \frac{1}{2M}\right).$$

4.4. AGRS WITH DITHERED QUANTIZATION CONT'D

4.4. AGRS WITH DITHERED QUANTIZATION CONT'D

With DQ, we can encode any bound with size $1/M$.

4.4. AGRS WITH DITHERED QUANTIZATION CONT'D

With DQ, we can encode any bound with size $1/M$.

What about bounds with arbitrary rational sizes?

4.4. AGRS WITH DITHERED QUANTIZATION CONT'D

With DQ, we can encode any bound with size $1/M$.

What about bounds with arbitrary rational sizes?

DQ + Bits-back: Bits-back Quantization (BBQ)

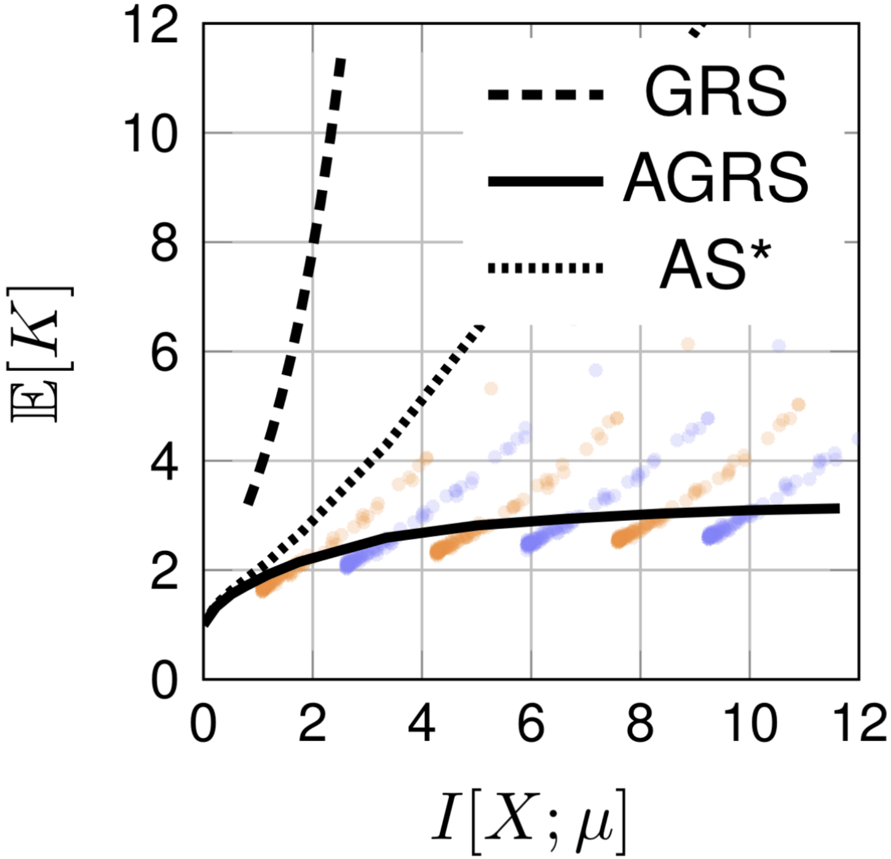
4.5. AGRS FOR 1D GAUSSIANS

4.5. AGRS FOR 1D GAUSSIANS

Apply AGRS to our Gaussian example:

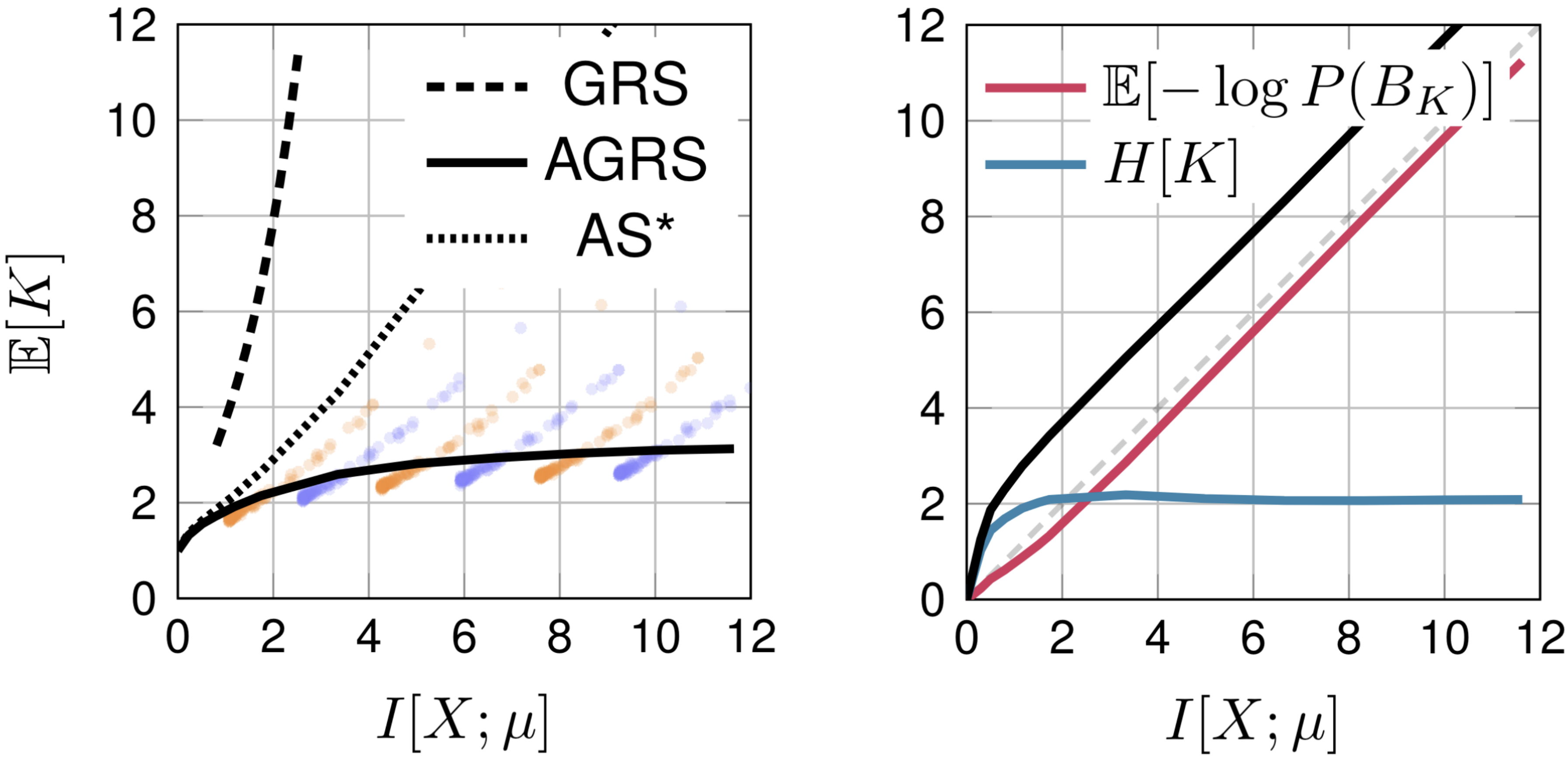
4.5. AGRS FOR 1D GAUSSIANS

Apply AGRS to our Gaussian example:



4.5. AGRS FOR 1D GAUSSIANS

Apply AGRS to our Gaussian example:



5. FUTURE DIRECTIONS

5. FUTURE DIRECTIONS

1. Is there a sampling algorithm with $\mathcal{O}\left(2^{D_{KL}[Q \,\|\, P]}\right)$ or is $2^{D_\infty[Q \,\|\, P]}$ tight?

5. FUTURE DIRECTIONS

1. Is there a sampling algorithm with $\mathcal{O}\left(2^{D_{KL}[Q \,\|\, P]}\right)$ or is $2^{D_\infty[Q \,\|\, P]}$ tight?
2. Connection to Poisson Functional Representation (Li and El Gamal, 2017)? See Greedy Poisson Rejection Sampling (F., 2023)

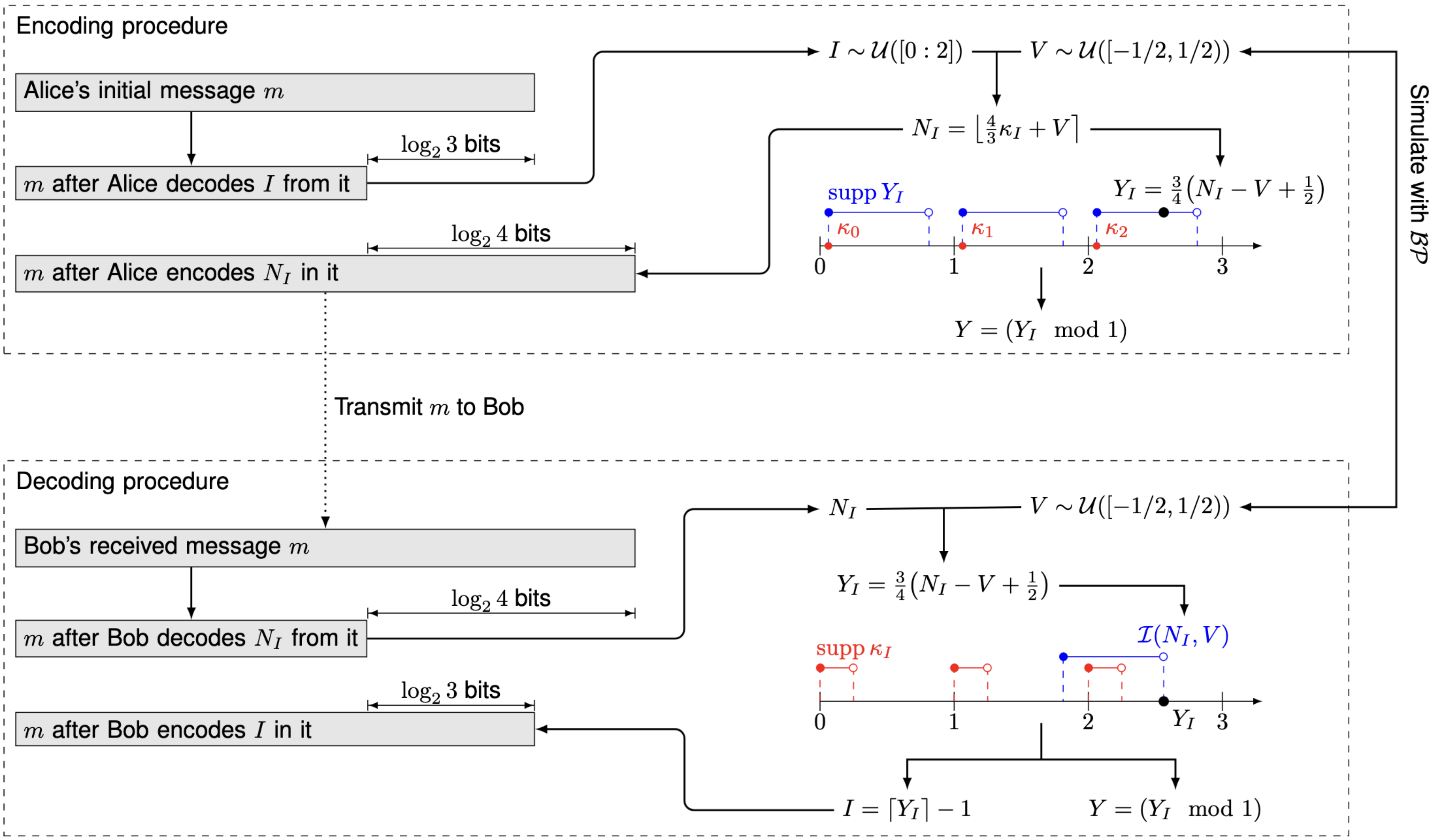
5. FUTURE DIRECTIONS

1. Is there a sampling algorithm with $\mathcal{O}\left(2^{D_{KL}[Q \,\|\, P]}\right)$ or is $2^{D_\infty[Q \,\|\, P]}$ tight?
2. Connection to Poisson Functional Representation (Li and El Gamal, 2017)? See Greedy Poisson Rejection Sampling (F., 2023)
3. Specialized algorithms for Gaussians?

6. REFERENCES

- 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.
- C. T. Li and A. El Gamal, “Strong functional representation lemma and applications to coding theorems,” IEEE Transactions on Information Theory, vol. 64, no. 11, pp. 6967–6978, 2018.
- F. “Greedy Poisson Rejection Sampling,” arXiv preprint arXiv:2305.15313, 2023.

7. BBQ



Encoding procedure
Alice's initial message m
$\log_2 3$ bits
m after Alice decodes I from it
$\log_2 4$ bits
m after Alice encodes N_I in it
$I \sim \mathcal{U}([0:2])$
$V \sim \mathcal{U}([-1/2, 1/2))$
$N_I = \lfloor \frac{4}{3}\kappa_I + V \rceil$
$Y_I = \frac{3}{4}(N_I - V + \frac{1}{2})$
supp Y_I
κ_0
κ_1
κ_2
0 1 2 3
$Y = (Y_I \mod 1)$
Simulate with $\mathcal{BP}$
Transmit m to Bob
Decoding procedure
Bob's received message m
$\log_2 4$ bits
m after Bob decodes N_I from it
$\log_2 3$ bits
m after Bob encodes I in it
N_I
$V \sim \mathcal{U}([-1/2, 1/2))$
$Y_I = \frac{3}{4}(N_I - V + \frac{1}{2})$
supp κ_I
$\mathcal{I}(N_I, V)$
0 1 2 Y_I 3
$I = \lceil Y_I \rceil - 1$
$Y = (Y_I \mod 1)$