On the deletion channel with small deletion probability

Yashodhan Kanoria

Stanford University

joint work with Andrea Montanari

Outline

Outline

The deletion channel

- n input bits X^n, each bit deleted independently w.p. d
- Output $Y(X^n)$ of length $\text{Binomial}(n, 1-d)$

The deletion channel

- n input bits X^n, each bit deleted independently w.p. d
- Output $Y(X^n)$ of length Binomial$(n, 1-d)$

X^n: 1 1 1 0 1 1 0 0 0 1 0 ...

The deletion channel

- n input bits X^n, each bit deleted independently w.p. d
- Output $Y(X^n)$ of length Binomial$(n, 1-d)$

X^n: 1 1 1 0 1 1 0 0 0 1 0 ...

$Y(X^n)$: 1 1 1 1 0 1 0 ...

Capacity of the deletion channel

$$C_n = \frac{1}{n} \max_{p_{X^n}} I(X^n; Y(X^n))$$

Lemma (Dobrushin '67)

$\lim_{n\to\infty} C_n$ *exists and is equal to* $\inf_{n\geq 1} C_n$.

Define capacity C as this limit.

Capacity of the deletion channel

$$C_n = \frac{1}{n} \max_{p_{X^n}} I(X^n; Y(X^n))$$

Lemma (Dobrushin '67)

$\lim_{n\to\infty} C_n$ *exists and is equal to* $\inf_{n\geq 1} C_n$.

Define capacity C as this limit.

Problems:

- Don't know how to compute capacity
- Optimal p_{X^n} not known for any $d > 0$

What's so hard?

LACK OF SYNCHRONIZATION!

Problems:

- Don't know how to compute capacity
- Optimal p_{X^n} not known for any $d > 0$

What's so hard?

LACK OF SYNCHRONIZATION!

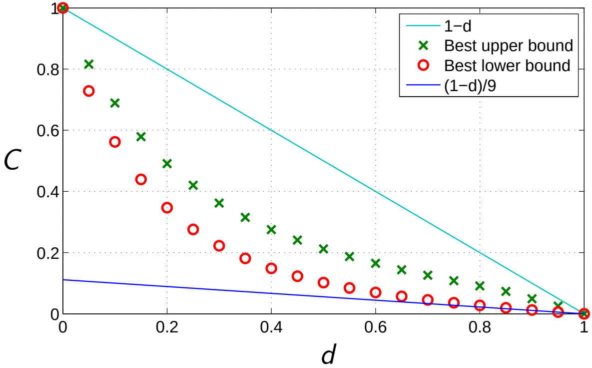
Problems:

- Don't know how to compute capacity
- Optimal p_{X^n} not known for any $d > 0$

What's so hard?

LACK OF SYNCHRONIZATION!

Bounds on Capacity



- $(1-d)/9 \le C \le (1-d)$ [Mitzenmacher et al '06]
- Upper bounds: augmented channels [Diggavi et al '07, Fertonani-Duman '09].
- Best computed lower bds: Markov sources + Jigsaw decoding [Mitzenmacher-Drinea '07]

New Idea

New Idea

I know how to solve $d = 0$!

New Idea

I know how to solve $d = 0$!

Can we expand C for $d \to 0^+$?

Optimal input distribution for small d?

- [Kalai, Mitzenmacher & Sudan, ISIT '10] addresses **same problem**!
- Shows $C = 1 - d\log(1/d) + o(d\log(1/d))$
- Very different proof technique

We obtain in addition:

(i) Order d term (in paper)

(ii) Order d^2 term and optimal coding scheme (updated result)

- [Kalai, Mitzenmacher & Sudan, ISIT '10] addresses **same problem**!
- Shows $C = 1 - d\log(1/d) + o(d\log(1/d))$
- Very different proof technique

We obtain in addition:

(i) Order d term (in paper)

(ii) Order d^2 term and optimal coding scheme (updated result)

- [Kalai, Mitzenmacher & Sudan, ISIT '10] addresses **same problem**!
- Shows $C = 1 - d\log(1/d) + o(d\log(1/d))$
- Very different proof technique

We obtain in addition:

(i) Order d term (in paper)

(ii) Order d^2 term and optimal coding scheme (updated result)

Outline

Main result: Capacity expansion

Result in our ISIT paper:

Theorem

For small d,

1. $$C(d) = 1 - d\log(1/d) - A_1\, d + O(d^{1.4})$$

   where $A_1 = \log(2e) - \sum_{\ell=1}^{\infty} 2^{-\ell-1}\ell\log_2\ell \approx 1.154$

2. *The iid Bernoulli*$(1/2)$ *process achieves rate* $C - O(d^{1.4})$.

Main result: Capacity expansion

Result in our ISIT paper:

Theorem

For small d,

1. $$C(d) = 1 - d\log(1/d) - A_1\, d + O(d^{1.4})$$
   where $A_1 = \log(2e) - \sum_{\ell=1}^{\infty} 2^{-\ell-1}\ell\log_2\ell \approx 1.154$
2. *The iid Bernoulli*$(1/2)$ *process achieves rate* $C - O(d^{1.4})$.

But the expansion can go further!

Main result: Capacity expansion

Updated result:

Theorem

For small d,

1. $$C(d) = 1 - d\log(1/d) - A_1\, d + A_2 d^2 + O(d^{2.9})$$

   where $A_1 = \log(2e) - \sum_{\ell=1}^{\infty} 2^{-\ell-1}\ell\log_2\ell \approx 1.154$

   $A_2 = \ \ldots$ (multi-line expression) ≈ 1.792

2. *The process* X^* (coming up!) *achieves* $C - O(d^{2.9})$.

Implications of expansion

- Previous best upper bound off by $(1/4)d\log(1/d)$
- Previous computed lower bound off by $0.904d^2$:
  Bounds based on Markov sources + Jigsaw decoding
  [Diggavi et al '01, Mitzenmacher-Drinea '07]

Fact

- *The maximum rate achieved by a first order Markov source is*

$$R_{\mathrm{Mkv}} = C - 0.100d^2 + O(d^{2.9})\,.$$

- *'Jigsaw decoding' incurs asymptotic rate loss of* $0.804d^2 + O(d^{2.9})$.

Thus, **previous computed lower bounds off by 0.904d^2 asymptotically**.

Implications of expansion

- Previous best upper bound off by $(1/4)d\log(1/d)$
- Previous computed lower bound off by $0.904d^2$:
  Bounds based on Markov sources + Jigsaw decoding
  [Diggavi et al '01, Mitzenmacher-Drinea '07]

Fact

- *The maximum rate achieved by a first order Markov source is*

$$R_{\mathrm{Mkv}} = C - 0.100d^2 + O(d^{2.9})\,.$$

- *'Jigsaw decoding' incurs asymptotic rate loss of* $0.804d^2 + O(d^{2.9})$.

Thus, **previous computed lower bounds off by $0.904\mathbf{d}^2$ asymptotically**.

Implications of expansion

- Previous best upper bound off by $(1/4)d\log(1/d)$
- Previous computed lower bound off by $0.904d^2$:
  Bounds based on Markov sources + Jigsaw decoding
  [Diggavi et al '01, Mitzenmacher-Drinea '07]

Fact

- *The maximum rate achieved by a first order Markov source is*

$$R_{\mathrm{Mkv}} = C - 0.100d^2 + O(d^{2.9})\,.$$

- *'Jigsaw decoding' incurs asymptotic rate loss of* $0.804d^2 + O(d^{2.9})$.

Thus, **previous computed lower bounds off by $0.904d^2$ asymptotically**.

Implications of expansion

- Previous best upper bound off by $(1/4)d\log(1/d)$
- Previous computed lower bound off by $0.904d^2$:
  Bounds based on Markov sources + Jigsaw decoding
  [Diggavi et al '01, Mitzenmacher-Drinea '07]

Fact

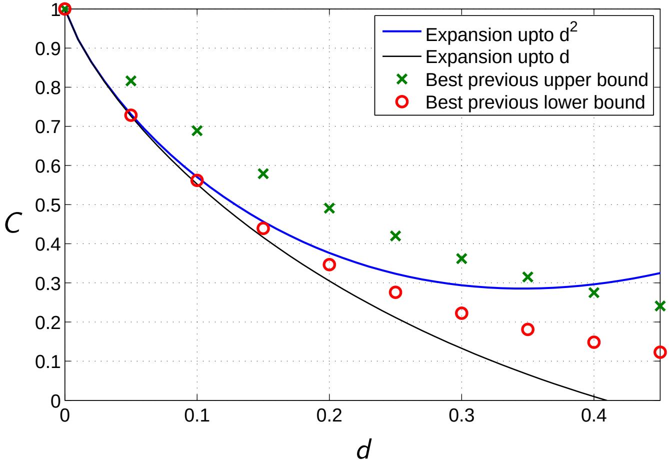
- *The maximum rate achieved by a first order Markov source is*

$$R_{\mathrm{Mkv}} = C - 0.100d^2 + O(d^{2.9})\,.$$

- *'Jigsaw decoding' incurs asymptotic rate loss of* $0.804d^2 + O(d^{2.9})$.

Thus, **previous computed lower bounds off by $\mathbf{0.904d^2}$ asymptotically**.

Capacity Expansion



Capacity Expansion

d	LB	C_{exp} upto d^2	C_{exp} upto d	UB
0.05	0.7283	0.7307	0.7262	0.8160
0.10	0.5620	0.5703	0.5524	0.6890
0.15	0.4392	0.4566	0.4163	0.5790

Optimal coding

- 'Runs' of 0s and 1s

$$\begin{array}{ccccccccccccc} & & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & & \mathcal{R}_4 & & \mathcal{R}_5 & & \\ \ldots 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots \end{array}$$

- $L \equiv$ Length of randomly selected run in stationary $\mathbb{X}$
- $\mathbb{X}$ is iid Bernoulli(1/2): $L \sim \text{Geo}(1/2)$, i.e. $p_L(\ell) = 2^{-\ell}$.

Optimal coding

- 'Runs' of 0s and 1s

$$\begin{array}{cccccccccccccc} & & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & & \mathcal{R}_4 & & \mathcal{R}_5 & & \\ \ldots 0 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots \end{array}$$

- $L \equiv$ Length of randomly selected run in stationary $\mathbb{X}$
- $\mathbb{X}$ is iid Bernoulli(1/2): $L \sim \text{Geo}(1/2)$, i.e. $p_L(\ell) = 2^{-\ell}$.

Optimal coding

- 'Runs' of 0s and 1s

$$\ldots 0 \; \overbrace{1 \; 1 \; 1}^{\mathcal{R}_1} \; \overbrace{0}^{\mathcal{R}_2} \; \overbrace{1 \; 1}^{\mathcal{R}_3} \; \overbrace{0 \; 0 \; 0}^{\mathcal{R}_4} \; \overbrace{1}^{\mathcal{R}_5} \; 0 \; \ldots$$

- $L \equiv$ Length of randomly selected run in stationary $\mathbb{X}$
- $\mathbb{X}$ is iid Bernoulli(1/2): $L \sim \text{Geo}(1/2)$, i.e. $p_L(\ell) = 2^{-\ell}$.

Optimal coding

Input distribution to achieve d^2 term:

Theorem

The stationary process $\mathbb{X}^*$ consisting of iid runs with distribution

$$p_L^*(\ell) = 2^{-\ell}\big(1 + d(\ell \ln \ell - c\ell)\big)$$

(where $c = \sum_{\ell=1}^{\infty} 2^{-\ell-1} \ell \ln \ell \approx 0.893.$)
achieves rate $R_* = C - O(d^{2.9})$.

First optimal coding result for deletion channel.

Optimal coding

Input distribution to achieve d^2 term:

Theorem

The stationary process $\mathbb{X}^*$ consisting of iid runs with distribution

$$p_L^*(\ell) = 2^{-\ell}\big(1 + d(\ell \ln \ell - c\ell)\big)$$

(where $c = \sum_{\ell=1}^{\infty} 2^{-\ell-1}\ell \ln \ell \approx 0.893$.)
achieves rate $R_* = C - O(d^{2.9})$.

First optimal coding result for deletion channel.

Outline

Theorem

For small d and any $\epsilon > 0$,

$$C(d) = 1 - d\log(1/d) - 1.154\, d + O(d^{1.4})$$

and the iid Bernoulli$(1/2)$ *process achieves rate* $C - O(d^{1.4})$.

Preliminaries

Lemma

Stationary ergodic sources suffice to achieve C.

Preliminaries

$$I(X^n; Y(X^n)) = H(Y) - H(Y|X^n)$$

Let $D^n \equiv$ channel realization.

$$\begin{aligned} H(Y|X^n) &= H(D^n, Y|X^n) - H(D^n|X^n, Y) \\ &= nh(d) - H(D^n|X^n, Y) \end{aligned}$$

since $Y = f(X^n, D^n)$, and D^n is iid Bernoulli(d) independent of X^n.

Main problem: $H(D^n|X^n, Y)$.

Preliminaries

$$I(X^n; Y(X^n)) = H(Y) - H(Y|X^n)$$

Let $D^n \equiv$ channel realization.

$$H(Y|X^n) = H(D^n, Y|X^n) - H(D^n|X^n, Y)$$
$$= nh(d) - H(D^n|X^n, Y)$$

since $Y = f(X^n, D^n)$, and D^n is iid Bernoulli(d) independent of X^n.

Main problem: $H(D^n|X^n, Y)$.

Preliminaries

$$I(X^n; Y(X^n)) = H(Y) - H(Y|X^n)$$

Let $D^n \equiv$ channel realization.

$$\begin{aligned} H(Y|X^n) &= H(D^n, Y|X^n) - H(D^n|X^n, Y) \\ &= nh(d) - H(D^n|X^n, Y) \end{aligned}$$

since $Y = f(X^n, D^n)$, and D^n is iid Bernoulli(d) independent of X^n.

Main problem: $H(D^n|X^n, Y)$.

Preliminaries

$$I(X^n; Y(X^n)) = H(Y) - H(Y|X^n)$$

Let $D^n \equiv$ channel realization.

$$\begin{aligned} H(Y|X^n) &= H(D^n, Y|X^n) - H(D^n|X^n, Y) \\ &= nh(d) - H(D^n|X^n, Y) \end{aligned}$$

since $Y = f(X^n, D^n)$, and D^n is iid Bernoulli(d) independent of X^n.

Main problem: $H(D^n|X^n, Y)$.

Many possibilities for D^n given $(X^n, Y(X^n))$, e.g.

$$\vdots$$

What is conditional entropy rate $\lim_{n\to\infty} H(D^n|X^n, Y)/n$?

Many possibilities for D^n given $(X^n, Y(X^n))$, e.g.

X^n: 1 1 1 0 1 1 0 0 0 1 0 ...

$Y(X^n)$: 1 1 1 1 0 1 0 ...

$\vdots$

What is conditional entropy rate $\lim_{n\to\infty} H(D^n|X^n, Y)/n$?

Many possibilities for D^n given $(X^n, Y(X^n))$, e.g.

X^n: 1 1 1 0 1 1 0 0 0 1 0 ...

$Y(X^n)$: 1 1 1 1 0 1 0 ...

$\vdots$

What is conditional entropy rate $\lim_{n\to\infty} H(D^n|X^n, Y)/n$?

Many possibilities for D^n given $(X^n, Y(X^n))$, e.g.

X^n: 1 1 1 0 1 1 0 0 0 1 0 . . .

$Y(X^n)$: 1 1 1 1 0 1 0 . . .

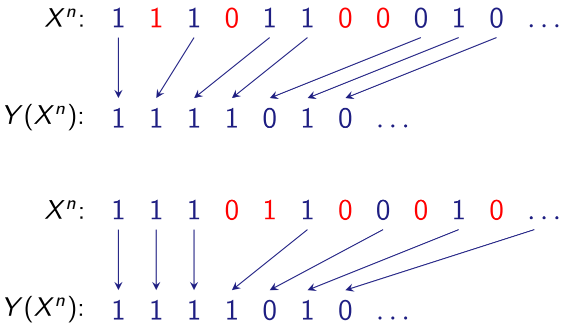
X^n: 1 1 1 0 1 1 0 0 0 1 0 . . .

$Y(X^n)$: 1 1 1 1 0 1 0 . . .

⋮

What is conditional entropy rate $\lim_{n\to\infty} H(D^n|X^n, Y)/n$?

Many possibilities for D^n given $(X^n, Y(X^n))$, e.g.



$\vdots$

What is conditional entropy rate $\lim_{n\to\infty} H(D^n|X^n, Y)/n$?

Key Lemma

Lemma

$$\lim_{n\to\infty} \frac{1}{n} H(D^n | X^n, Y(X^n)) = d \frac{\mathbb{E}[L \log L]}{\mathbb{E}[L]} + O(d^{1.9})$$

provided $\mathbb{E}[L^2 \log L] < d^{-0.05}$.

- Lemma holds uniformly for all processes $\mathbb{X}$.
- Key to showing both achievability and upper bound.

Key Lemma

Lemma

$$\lim_{n\to\infty} \frac{1}{n} H(D^n|X^n, Y(X^n)) = d\frac{\mathbb{E}[L \log L]}{\mathbb{E}[L]} + O(d^{1.9})$$

provided $\mathbb{E}[L^2 \log L] < d^{-0.05}$.

- Lemma holds uniformly for all processes $\mathbb{X}$.
- Key to showing both achievability and upper bound.

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

- No run disappears completely

$$\begin{array}{lccccccccccccc} & & \mathcal{R}_1 & & & \mathcal{R}_2 & & \mathcal{R}_3 & & & \mathcal{R}_4 & & \\ X^n & : & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & \ldots \\ Y(X^n): & & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots & & \\ & & \mathcal{S}_1 & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & & \end{array}$$

$$\mathcal{S}_i \leftarrow \mathcal{R}_i$$

Given $(X^n, Y(X^n))$:

- We know which runs have (one) deletion
- Each bit of run equally likely candidate for deletion

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

- No run disappears completely

$$\begin{array}{lccccccccccccc} & \mathcal{R}_1 & & & \mathcal{R}_2 & & \mathcal{R}_3 & & & \mathcal{R}_4 & & & \\ X^n \quad : & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & \ldots \\ Y(X^n): & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots & & \\ & \mathcal{S}_1 & & \mathcal{S}_2 & & & \mathcal{S}_3 & & \mathcal{S}_4 & & & & \end{array}$$

$$\mathcal{S}_i \leftarrow \mathcal{R}_i$$

Given $(X^n, Y(X^n))$:

- We know which runs have (one) deletion
- Each bit of run equally likely candidate for deletion

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

- No run disappears completely

$$\begin{array}{lccccccccccccc} & & \mathcal{R}_1 & & & \mathcal{R}_2 & & \mathcal{R}_3 & & & \mathcal{R}_4 & & \\ X^n: & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & \dots \\ Y(X^n): & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \dots \\ & \mathcal{S}_1 & & \mathcal{S}_2 & & & \mathcal{S}_3 & & \mathcal{S}_4 & & \end{array}$$

$$\mathcal{S}_i \leftarrow \mathcal{R}_i$$

Given $(X^n, Y(X^n))$:

- We know which runs have (one) deletion
- Each bit of run equally likely candidate for deletion

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

- No run disappears completely

$$
\begin{array}{lllllllllllll}
 & & \mathcal{R}_1 & & & \mathcal{R}_2 & & & \mathcal{R}_3 & & & \mathcal{R}_4 & \\
X^n & : & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \; \ldots \\
Y(X^n): & & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & \ldots & \\
 & & \mathcal{S}_1 & & \mathcal{S}_2 & & & \mathcal{S}_3 & & \mathcal{S}_4 & & &
\end{array}
$$

$$\mathcal{S}_i \leftarrow \mathcal{R}_i$$

Given $(X^n, Y(X^n))$:

- We know which runs have (one) deletion
- Each bit of run equally likely candidate for deletion

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

- No run disappears completely

$$\begin{array}{ll} & \quad\ \mathcal{R}_1 \qquad\ \mathcal{R}_2 \qquad \mathcal{R}_3 \qquad \mathcal{R}_4 \\ X^n: & 0\ 0\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 1\ 0\ \ldots \\ Y(X^n): & 0\ 0\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ \ldots \\ & \ \mathcal{S}_1 \quad\ \mathcal{S}_2 \qquad \mathcal{S}_3 \quad \mathcal{S}_4 \end{array}$$

$$\mathcal{S}_i \leftarrow \mathcal{R}_i$$

Given $(X^n, Y(X^n))$:

- We know which runs have (one) deletion
- Each bit of run equally likely candidate for deletion

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

Run of length ℓ has:

- probability $p_L(\ell)$ of occurring.
- probability $\approx \ell d$ of suffering a deletion.
- Contribution $\log \ell$ to $H(D^n|X^n, Y)$ if deletion

$$\lim_{n\to\infty} \frac{1}{n} H(D^n|X^n, Y) \approx \frac{d}{\mathbb{E}[L]} \sum_{\ell=2}^{\infty} p_L(\ell)\ell \log \ell = d \frac{\mathbb{E}[L \log L]}{\mathbb{E}[L]}$$

Proof of Key Lemma: sparsity of deletions

Simple case: no runs of length 1.

Imagine *no run in X^n has > 1 deletion in total.*

Run of length ℓ has:

- probability $p_L(\ell)$ of occurring.
- probability $\approx \ell d$ of suffering a deletion.
- Contribution $\log \ell$ to $H(D^n|X^n, Y)$ if deletion

$$\lim_{n\to\infty} \frac{1}{n} H(D^n|X^n, Y) \approx \frac{d}{\mathbb{E}[L]} \sum_{\ell=2}^{\infty} p_L(\ell)\ell \log \ell = d \frac{\mathbb{E}[L \log L]}{\mathbb{E}[L]}$$

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

X^n : 0 0 1 1 0 1 1 1 0 1 1 1 0 . . .

$Y(X^n)$: 0 0 1 1 1 1 1 0 1 1 0 . . .

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$\mathcal{R}_1$ $\mathcal{R}_2$ $\mathcal{R}_3$ $\mathcal{R}_4$ $\mathcal{R}_5$ $\mathcal{R}_6$

X^n : 0 0 1 1 0 1 1 1 0 1 1 1 0 ...

$Y(X^n)$: 0 0 1 1 1 1 1 0 1 1 0 ...

$\mathcal{S}_1$ $\mathcal{S}_2$ $\mathcal{S}_3$ $\mathcal{S}_4$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{ll} & \mathcal{R}_1 \quad \mathcal{R}_2 \quad \mathcal{R}_3 \quad \mathcal{R}_4 \quad \mathcal{R}_5 \quad \mathcal{R}_6 \\ X^n: & 0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ \ldots \\ Y(X^n): & 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ \ldots \\ & \mathcal{S}_1 \quad \mathcal{S}_2 \quad \mathcal{S}_3 \quad \mathcal{S}_4 \end{array}$$

$|\mathcal{S}_1| = |\mathcal{R}_1|$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$
\begin{array}{lccccccccccccccc}
 & & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & \\
X^n & : & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\
Y(X^n): & & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\
 & & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & &
\end{array}
$$

$$|\mathcal{S}_1| = |\mathcal{R}_1| \qquad \Rightarrow \qquad \mathcal{S}_1 \leftarrow \mathcal{R}_1$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lcccccccccccccc} & \mathcal{R}_1 & & \mathcal{R}_2 & & \mathcal{R}_3 & \mathcal{R}_4 & \mathcal{R}_5 & \mathcal{R}_6 \\ X^n: & 0 & 0 & 1 & 1 & 0 & 1\,1\,1 & 0 & 1\,1\,1 & 0 & \ldots \\ Y(X^n): & 0 & 0 & 1 & 1 & 1\,1\,1 & 0 & 1\,1 & 0 & \ldots \\ & \mathcal{S}_1 & & & \mathcal{S}_2 & & \mathcal{S}_3 & \mathcal{S}_4 \end{array}$$

$|\mathcal{S}_2| > |\mathcal{R}_2|$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lcccccccccccccc}
 & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & & \\
X^n: & 0\ 0 & & 1\ 1 & 0 & & 1\ 1\ 1 & & 0 & & 1\ 1\ 1 & & 0 & & \ldots \\
Y(X^n): & 0\ 0 & & & 1\ 1\ 1\ 1\ 1 & & & & 0 & & 1\ 1 & & 0 & & \ldots \\
 & \mathcal{S}_1 & & & \mathcal{S}_2 & & & & \mathcal{S}_3 & & \mathcal{S}_4 & & & &
\end{array}$$

$$|\mathcal{S}_2| > |\mathcal{R}_2| \qquad \Rightarrow \qquad \mathcal{S}_2 \leftarrow \mathcal{R}_2 + \mathcal{R}_3 + \mathcal{R}_4$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n | X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lccccccccccccccc}
 & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & \\
X^n: & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\
Y(X^n): & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\
 & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & & &
\end{array}$$

$$|\mathcal{S}_3| = |\mathcal{R}_5|$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\mathcal{R}_1 \quad \mathcal{R}_2 \quad \mathcal{R}_3 \quad \mathcal{R}_4 \quad \mathcal{R}_5 \quad \mathcal{R}_6$$

$$X^n \quad : \quad 0\ 0\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ 1\ 1\ 1\ 0\ \ldots$$

$$Y(X^n): \quad 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0\ 1\ 1\ 0\ \ldots$$

$$\mathcal{S}_1 \quad \mathcal{S}_2 \quad \mathcal{S}_3 \quad \mathcal{S}_4$$

$$|\mathcal{S}_3| = |\mathcal{R}_5| \qquad \Rightarrow \qquad \mathcal{S}_3 \leftarrow \mathcal{R}_5$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$
\begin{array}{lccccccccccccccc}
 & & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & \\
X^n & : & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\
Y(X^n) & : & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\
 & & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & &
\end{array}
$$

$|\mathcal{S}_4| < |\mathcal{R}_6|$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$
\begin{array}{lcccccccccccccc}
 & & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & \\
X^n & : & 0\ 0 & 1\ 1 & & 0 & 1\ 1\ 1 & & 0 & & 1\ 1\ 1 & & 0 & \ldots \\
Y(X^n): & & 0\ 0 & 1\ 1\ 1\ 1\ 1 & & & & & 0 & 1\ 1 & & 0 & \ldots & \\
 & & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & &
\end{array}
$$

$$|\mathcal{S}_4| < |\mathcal{R}_6| \quad \Rightarrow \quad \mathcal{S}_4 \leftarrow \mathcal{R}_6$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lccccccccccccccc}
 & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & & \mathcal{R}_4 & & \mathcal{R}_5 & & \mathcal{R}_6 & & & & \\
X^n \;: & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\
Y(X^n): & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\
 & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & & &
\end{array}$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lccccccccccccc} & \mathcal{R}_1 & & \mathcal{R}_2 & \mathcal{R}_3 & \mathcal{R}_4 & & & \mathcal{R}_5 & \mathcal{R}_6 & & & & \\ X^n \ : & 0 & 0 & 1\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\ Y(X^n): & 0 & 0 & 1\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\ & \mathcal{S}_1 & & & \mathcal{S}_2 & & & \mathcal{S}_3 & \mathcal{S}_4 & & & & & \end{array}$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n|X^n, Y)$.

Proof of Key Lemma: sparsity of deletions

General case

Imagine *no three consecutive runs in X^n have > 1 deletion in total.*

- Only runs of length 1 can disappear
- Runs in Y unambiguously associated with runs in X^n

$$\begin{array}{lcccccccccccccc}
 & \multicolumn{2}{c}{\mathcal{R}_1} & \multicolumn{2}{c}{\mathcal{R}_2} & \mathcal{R}_3 & \multicolumn{3}{c}{\mathcal{R}_4} & \mathcal{R}_5 & \multicolumn{3}{c}{\mathcal{R}_6} & \\
X^n\ : & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & \ldots \\
Y(X^n): & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & \ldots & & \\
 & \multicolumn{2}{c}{\mathcal{S}_1} & \multicolumn{5}{c}{\mathcal{S}_2} & \mathcal{S}_3 & \multicolumn{2}{c}{\mathcal{S}_4} & & & &
\end{array}$$

- We know $\mathcal{R}_3$ and $\mathcal{R}_6$ have deletion.
- $\mathcal{R}_3$ has no contribution to $H(D^n | X^n, Y)$.

Upper bound

Achievability of $1 - d\log(1/d)$

$$\Downarrow$$

$$H(\mathbb{X}_{\mathrm{opt}}) > 1 - d^{1-\epsilon}\,.$$

$$\Downarrow$$

$\mathbb{X}_{\mathrm{opt}}$ is 'close' to Bernoulli(1/2) process.

Conclusion

We obtained for deletion channel with small d:

- Asymptotic expansion of capacity upto order d^2.
- Optimal coding scheme.

Further directions:

- Explicit upper and lower bounds.
- Next terms in expansion.
- Is the series convergent?
- Series expansion for other 'hard' channels ... ?

Conclusion

We obtained for deletion channel with small d:

- Asymptotic expansion of capacity upto order d^2.
- Optimal coding scheme.

Further directions:

- Explicit upper and lower bounds.
- Next terms in expansion.
- Is the series convergent?
- Series expansion for other 'hard' channels ... ?

Conclusion

We obtained for deletion channel with small d:

- Asymptotic expansion of capacity upto order d^2.
- Optimal coding scheme.

Further directions:

- Explicit upper and lower bounds.
- Next terms in expansion.
- Is the series convergent?
- Series expansion for other 'hard' channels ... ?

THANK YOU!