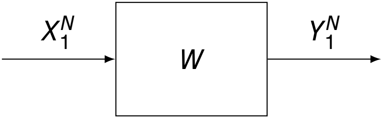
Fast Polarization for Processes with Memory

Joint work with Eren Şaşoğlu and Boaz Shuval

Polar codes in one slide



Polar coding

- **Information vector**: $\tilde{U}_1^k$
- **Padding**: $U_1^N = f(\tilde{U}_1^k)$
- **Encoding**: $X_1^N = U_1^N \cdot G_N^{-1}$
- **Decoding**: Successively, deduce U_i from U_1^{i-1} and Y_1^N

Polar codes in two slides: [Arıkan:09], [ArıkanTelatar:09]

- **Setting**: binary-input, symmetric, memoryless channel
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$

$$X_1^N \quad \text{uniform} \quad \Longleftrightarrow \quad U_1^N \quad \text{uniform}$$

- **Low entropy indices**: Fix $\beta < 1/2$

$$\Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$$

- **Polarization**: Let X_1^N be uniform

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = I(X_1; Y_1)$$

- **Coding scheme**:
  - For $i \in \Lambda_N$, set U_i equal to information bits (uniform)
  - Set remaining U_i to uniform values, reveal to decoder
  - Transmit $X_1^N = U_1^N \cdot G_N^{-1}$ as codeword

In this talk

Setting: binary-input, symmetric, memoryless channel

In this talk

Setting: ~~binary-input~~, symmetric, memoryless channel

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, memoryless channel

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, memoryless ~~channel~~

In this talk

Setting: ~~binary-input~~, ~~symmetric~~, ~~memoryless~~ ~~channel~~

Polar codes: [Şaşoğlu+:09], [KoradaUrbanke:10], [HondaYamamoto:13]

- **Setting**: Memoryless i.i.d. process $(X_i, Y_i)_{i=1}^N$
- **For simplicity**: Assume X_i binary
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$
- **Index sets**:

$$\text{Low entropy:} \quad \Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$$

$$\text{High entropy:} \quad \Omega_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) > \frac{1}{2} - 2^{-N^\beta} \right\}$$

- **Polarization**:

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = 1 - H(X_1 | Y_1)$$

$$\lim_{N \to \infty} \frac{1}{N} |\Omega_N| = H(X_1 | Y_1)$$

Polar codes: [Şaşoğlu+:09], [KoradaUrbanke:10], [HondaYamamoto:13]

Optimal rate for:

- Coding for non-symmetric memoryless channels
- Coding for memoryless channels with non-binary inputs
- (Lossy) compression of memoryless sources

Question

- How to handle memory?

Roadmap

Index sets

$$\text{Low entropy:} \quad \Lambda_N(\epsilon) = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < \epsilon \right\}$$

$$\text{High entropy:} \quad \Omega_N(\epsilon) = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) > \frac{1}{2} - \epsilon \right\}$$

Plan

- Define framework for handling memory
- Establish:
  - Slow polarization: for $\epsilon > 0$ **fixed**,

$$\lim_{N\to\infty} \frac{1}{N} |\Lambda_N(\epsilon)| = 1 - H_\star(X|Y)$$

$$\lim_{N\to\infty} \frac{1}{N} |\Omega_N(\epsilon)| = H_\star(X|Y)$$

$\lim_{N\to\infty} \frac{1}{N} H(X_1^N | Y_1^N)$

  - Fast polarization: also holds for $\epsilon = 2^{-N^\beta}$
  - What is β?

A framework for memory

- **Process**:

$$(X_i, Y_i, S_i)_{i=1}^{N}$$

  - **Finite number of states**: $S_i \in \mathcal{S}$, where $|\mathcal{S}| < \infty$
  - **Hidden state**: S_i is unknown to encoder and decoder
- **Probability distribution**:

$$P(x_i, y_i, s_i | s_{i-1})$$

  - **Stationary**: same for all i
  - **Markov**:

$$P(x_i, y_i, s_i | s_{i-1}) = P(x_i, y_i, s_i | \{x_j, y_j, s_j\}_{j<i})$$

  - **State sequence**: aperiodic and irreducible Markov chain

Example 1

- **Model**: Finite state channel

$$P_s(y|x)\ , \quad s \in \mathcal{S}$$

- **Input distribution**: X_i i.i.d. and independent of state
- **State transition**:

$$\pi(s_i|s_{i-1})$$

- **Distribution**:

$$P(x_i, y_i, s_i|s_{i-1}) = P(x_i)\pi(s_i|s_{i-1})P_{s_i}(y_i|x_i)$$

Example 2

- **Model**: ISI + noise

$$Y_i = h_0 X_i + h_1 X_{i-1} + \cdots + h_m X_{i-m} + \text{noise}$$

- **Input**: X_i has memory

$$P(x_i | x_{i-1}, x_{i-2}, \ldots, x_{i-m}, x_{i-m-1})$$

- **State**:

$$S_i = \begin{bmatrix} X_i & X_{i-1} & \cdots & X_{i-m} \end{bmatrix}$$

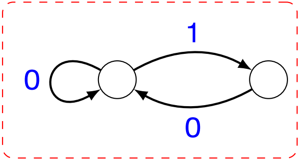
- **Distribution**: For x_i, s_i, s_{i-1} compatible,

$$P(x_i, y_i, s_i | s_{i-1}) = P_{\text{noise}}(y_i | h^T s_i) \cdot P(x_i | s_{i-1})$$

Example 3

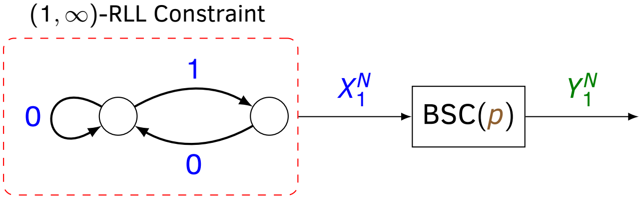
- **Model**: (d, k)-RLL constrained system with noise

$(1, \infty)$-RLL Constraint



Example 3

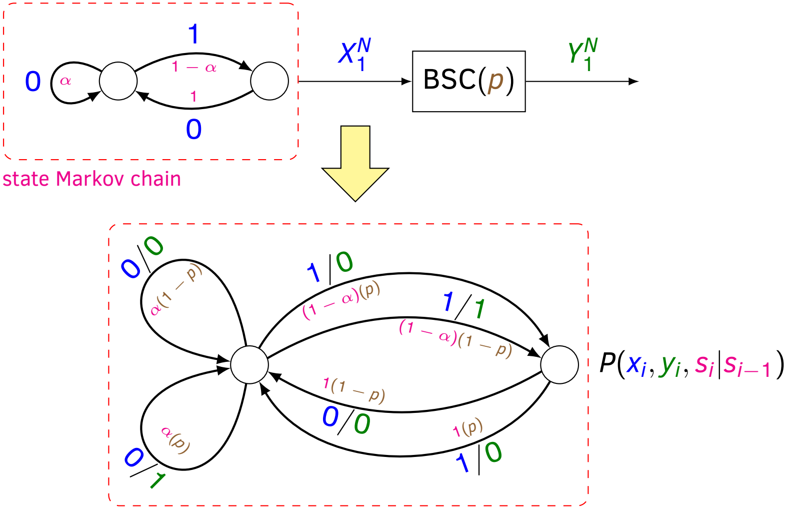
- **Model**: (d, k)-RLL constrained system with noise



Example 3

- **Model**: (d, k)-RLL constrained system with noise

$(1, \infty)$-RLL Constraint



Example 4

- **Model**: Lossy compression of a source with memory

$$Y_1^N \xrightarrow{\text{Lossy compression}} X_1^N$$

- **Source distribution**:

$$P_s(y)\ , \quad s \in \mathcal{S}$$

- **State transition**:

$$\pi(s_i|s_{i-1})$$

- **Distortion**: test channel $P(x|y)$
- **Distribution**:

$$P(x_i, y_i, s_i|s_{i-1}) = \pi(s_i|s_{i-1})P_{s_i}(y_i)P(x_i|y_i)$$

Polar codes: [Şaşoğlu:11], [ŞaşoğluTal:16], [ShuvalTal:17]

- **Setting**: Process $(X_i, Y_i, S_i)_{i=1}^N$ with memory, as above
- **Hidden state**: State unknown to encoder and decoder
- **Polar transform**: $U_1^N = X_1^N \cdot G_N$
  U_1^N are neither independent, nor identically distributed
- **Index sets**: Fix $\beta < 1/2$

$$\text{Low entropy:} \quad \Lambda_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) < 2^{-N^\beta} \right\}$$

$$\text{High entropy:} \quad \Omega_N = \left\{ i : P_{\text{error}}(U_i | U_1^{i-1}, Y_1^N) > \frac{1}{2} - 2^{-N^\beta} \right\}$$

- **Polarization**:

$$\lim_{N \to \infty} \frac{1}{N} |\Lambda_N| = 1 - H_\star(X|Y)$$

$$\lim_{N \to \infty} \frac{1}{N} |\Omega_N| = H_\star(X|Y)$$

$\boxed{\lim_{N \to \infty} \frac{1}{N} H(X_1^N | Y_1^N)}$

Achievable rate

- **Achievable rate**: In all examples, R approaches

$$I_{\star}(X;Y) = \lim_{N\to\infty} \frac{1}{N} I(X_1^N; Y_1^N)$$

- **Successive cancellation**: [Wang+:15]

Mixing

Consider the process (X_i, Y_i) — hidden state

$$\begin{array}{cccccccccc} X_1 & X_2 & \cdots & X_L & X_{L+1} & \cdots & X_M & X_{M+1} & X_{M+2} & \cdots & X_N \\ Y_1 & Y_2 & \cdots & Y_L & Y_{L+1} & \cdots & Y_M & Y_{M+1} & Y_{M+2} & \cdots & Y_N \end{array}$$

Then, there exist $\psi(k)$, $k \geq 0$, such that

$$P_{\mathsf{X}_1^L, \mathsf{Y}_1^L, \mathsf{X}_{M+1}^N, \mathsf{Y}_{M+1}^N} \leq \psi(M-L) \cdot P_{\mathsf{X}_1^L, \mathsf{Y}_1^L} \cdot P_{\mathsf{X}_{M+1}^N, \mathsf{Y}_{M+1}^N}$$

where:

- $\psi(0) < \infty$
- $\psi(k) \to 1$

Three parameters

- Joint distribution $P(x, y)$
- For simplicity: $X \in \{0, 1\}$
- **Parameters**:

Entropy	$H(X\|Y) = -\sum_{x,y} P(x,y) \log P(x\|y)$
Bhattacharyya	$Z(X\|Y) = 2\sum_y \sqrt{P(0,y)P(1,y)}$
T.V. distance	$K(X\|Y) = \sum_y \|P(0,y) - P(1,y)\|$

- **Connections**:

$$H \approx 0 \iff Z \approx 0 \iff K \approx 1$$

$$H \approx 1 \iff Z \approx 1 \iff K \approx 0$$

Three processes

For $n = 1, 2, \ldots$

- $N = 2^n$
- $U_1^N = X_1^N G_N$
- Pick $B_n \in \{0, 1\}$ uniform, i.i.d.
- **Random index** from $\{1, 2, \ldots, N\}$

$$\boldsymbol{i} = 1 + \langle B_1 \, B_2 \, \cdots \, B_n \rangle_2$$

- **Processes**:

Entropy	$H_n = H(U_{\boldsymbol{i}} \vert U_1^{\boldsymbol{i}-1}, Y_1^N)$
Bhattacharyya	$Z_n = Z(U_{\boldsymbol{i}} \vert U_1^{\boldsymbol{i}-1}, Y_1^N)$
T.V. distance	$K_n = K(U_{\boldsymbol{i}} \vert U_1^{\boldsymbol{i}-1}, Y_1^N)$

$$\{X_i, Y_i, S_i\}$$
$$\Downarrow$$
$$\{X_i, Y_i\}$$

Proof — memoryless case

Slow polarization

$$H_n \in (\epsilon, 1-\epsilon)$$

$$\Downarrow$$

$$|H_{n+1} - H_n| > 0$$

Fast polarization

$$Z_{n+1} \leq \begin{cases} 2Z_n & B_{n+1} = 0 \\ Z_n^2 & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Lambda_N| \xrightarrow[n\to\infty]{} 1 - H(X_1|Y_1)$$

Low entropy set

$\boxed{\text{New}}$

$$K_{n+1} \leq \begin{cases} K_n^2 & B_{n+1} = 0 \\ 2K_n & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Omega_N| \xrightarrow[n\to\infty]{} H(X_1|Y_1)$$

High entropy set

Proof — memory~~less~~ case

$$\{X_i, Y_i, S_i\} \Downarrow \{X_i, Y_i\}$$

Slow polarization

$$H_n \in (\epsilon, 1-\epsilon)$$

$$\Downarrow$$

$$|H_{n+1} - H_n| > 0$$

$$\boxed{\begin{array}{l} \psi = \psi(0) = \max_s \dfrac{1}{\pi(s)} \\ \pi\text{: stationary state distribution} \end{array}}$$

Fast polarization

$$Z_{n+1} \le \begin{cases} 2\psi Z_n & B_{n+1} = 0 \\ \psi Z_n^2 & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Lambda_N| \xrightarrow[n\to\infty]{} 1 - H_\star(X|Y)$$

Low entropy set

$$\hat{K}_{n+1} \le \begin{cases} \psi \hat{K}_n^2 & B_{n+1} = 0 \\ 2\hat{K}_n & B_{n+1} = 1 \end{cases}$$

$$\Downarrow$$

$$\frac{1}{N}|\Omega_N| \xrightarrow[n\to\infty]{} H_\star(X|Y)$$

High entropy set

Notation

- **Two consecutive blocks**: (X_1^N, Y_1^N) and $(X_{N+1}^{2N}, Y_{N+1}^{2N})$.
- **Polar transform**:

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$

- **Random index**:

$$i = 1 + \langle B_1 \, B_2 \, \cdots \, B_n \rangle_2$$

- **Notation**:

$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

Slow polarization

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$
$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

- H_n is a supermartingale

$$H_n = H(U_i|Q_i) = H(V_i|R_i)$$

$$H_{n+1} = \begin{cases} H(U_i + V_i|Q_i, R_i) & B_{n+1} = 0 \\ H(V_i|U_i + V_i, Q_i, R_i) & B_{n+1} = 1 \end{cases}$$

By the chain rule:

$$\begin{aligned} \mathbb{E}\left[H_{n+1}|H_n, \ldots\right] &= \frac{1}{2}\Big(H(U_i + V_i|Q_i, R_i) + H(V_i|U_i + V_i, Q_i, R_i)\Big) \\ &= \frac{1}{2}H(U_i + V_i, V_i|Q_i, R_i) \\ &= \frac{1}{2}H(U_i, V_i|Q_i, R_i) \\ &\leq \frac{1}{2}H(U_i|Q_i) + \frac{1}{2}H(V_i|R_i) = H_n \end{aligned}$$

Slow polarization

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$
$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

Convergence

- H_n is a supermartingale
- $0 \leq H_n \leq 1$

H_n converges a.s. and in L^1 to H_∞

Polarization

- $H_\infty \in [0,1]$
- We need: $H_\infty \in \{0,1\}$
- Easy if (U_1^N, Q_1^N) and (V_1^N, R_1^N) were independent
- They are not: $Y_N \in Q_1^N$ and $Y_{N+1} \in R_1^N$
- But: for almost all i, we have $I(U_i; V_i | Q_i, R_i) < \epsilon$
- Enough? No. Need to show that Q_i and R_i can't cooperate to stop polarization

Fast polarization to low entropy set Λ_N

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$
$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

- Recall:

$$P_{X_1^N, Y_1^N, X_{N+1}^{2N}, Y_{N+1}^{2N}} \leq \psi \cdot P_{X_1^N, Y_1^N} \cdot P_{X_{N+1}^{2N}, Y_{N+1}^{2N}}$$

- "Force" block independence:

$$(\tilde{X}_1^{2N}, \tilde{Y}_1^{2N}) \sim P_{X_1^N, Y_1^N} \cdot P_{X_{N+1}^{2N}, Y_{N+1}^{2N}}$$

- Thus,

$$P_{X_1^N, Y_1^N, X_{N+1}^{2N}, Y_{N+1}^{2N}} \leq \psi \cdot P_{\tilde{X}_1^N, \tilde{Y}_1^N, \tilde{X}_{N+1}^{2N}, \tilde{Y}_{N+1}^{2N}}$$

- With $\tilde{U}_i, \tilde{V}_i, \tilde{Q}_i, \tilde{R}_i$ as above

$$P_{U_1^N, Q_1^N, V_1^N, R_1^N} \leq \psi \cdot P_{\tilde{U}_1^N, \tilde{Q}_1^N, \tilde{V}_1^N, \tilde{R}_1^N}$$

Polarization of Z_N

$$
\begin{aligned}
& Z(U_i + V_i | Q_i, R_i) \\
&= 2 \cdot \sum_{q,r} \sqrt{P_{U_i+V_i,Q_i,R_i}(0,q,r) \cdot P_{U_i+V_i,Q_i,R_i}(1,q,r)} \\
&\leq 2 \cdot \sum_{q,r} \sqrt{\psi P_{\tilde{U}_i+\tilde{V}_i,\tilde{Q}_i,\tilde{R}_i}(0,q,r) \cdot \psi P_{\tilde{U}_i+\tilde{V}_i,\tilde{Q}_i,\tilde{R}_i}(1,q,r)} \\
&= \psi \cdot Z(\tilde{U}_i + \tilde{V}_i | \tilde{Q}_i, \tilde{R}_i) \\
&\leq \psi \cdot 2Z(\tilde{U}_i | \tilde{Q}_i) \\
&= \psi \cdot 2Z(U_i | Q_i)
\end{aligned}
$$

In a similar manner, we show

$$Z(V_i | U_i + V_i, Q_i, R_i) \leq \psi \cdot Z(U_i | Q_i)^2$$

Fast polarization to high entropy set Ω_N

- **Memoryless case**:
  - Proof hinges on independence:

$$P(x_1^{2N}, y_1^{2N}) = P(x_1^N, y_1^N) \cdot P(x_{N+1}^{2N}, y_{N+1}^{2N})$$

- **Memory case**:
  - Force independence: condition on middle state S_N

$$P(x_1^{2N}, y_1^{2N} | s_N) = P(x_1^N, y_1^N | s_N) \cdot P(x_{N+1}^{2N}, y_{N+1}^{2N} | s_N)$$

- **New processes**:

$$\hat{H}_n = H(U_i | U_1^{i-1}, Y_1^N, S_0, S_N)$$
$$\hat{K}_n = K(U_i | U_1^{i-1}, Y_1^N, S_0, S_N)$$

Tying things together

$$\begin{tikzcd}
Z_n \arrow[d] & H_\infty \arrow[l] \\
\Lambda_N &
\end{tikzcd}$$

Tying things together

$$
\begin{tikzcd}
Z_n \arrow[d] & \boxed{H_\infty \overset{\text{a.s.}}{=} \hat{H}_\infty} \arrow[l] \arrow[r] & \hat{K}_n \arrow[d] \\
\Lambda_N & & \Omega_N
\end{tikzcd}
$$

Polarization of K_n (memoryless case)

$$U_1^N = X_1^N \cdot G_N$$
$$V_1^N = X_{N+1}^{2N} \cdot G_N$$
$$Q_i = (U_1^{i-1}, Y_1^N)$$
$$R_i = (V_1^{i-1}, Y_{N+1}^{2N})$$

- **Memoryless assumption**:

$$P(u_i, v_i, q_i, r_i) = P(u_i, q_i) \cdot P(v_i, r_i)$$

- **Notation**:

$$T_i = U_i + V_i$$

- **One step polarization**:

$$K_{n+1} = \begin{cases} K(T_i | Q_i, R_i) & B_{n+1} = 0 & \text{`$-$' transform} \\ K(V_i | T_i, Q_i, R_i) & B_{n+1} = 1 & \text{`$+$' transform} \end{cases}$$

- **Recall**:

$$K(X|Y) = \sum_y |P(0, y) - P(1, y)|$$

Polarization of K_n (memoryless case), '$-$' transform

$$
\begin{aligned}
K_{n+1} &= \sum_{q,r} \left| P_{\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i}(0,q,r) - P_{\mathsf{T}_i,\mathsf{Q}_i,\mathsf{R}_i}(1,q,r) \right| \\
&= \sum_{q,r} \left| \sum_{v=0}^{1} P(v,r)\big(P(v,q) - P(v+1,q)\big) \right| \\
&= \sum_{q,r} \left| \Big(P(0,q) - P(1,q)\Big)\Big(P(0,r) - P(1,r)\Big) \right| \\
&= \sum_{q,r} \left| P(0,q) - P(1,q) \right| \cdot \left| P(0,r) - P(1,r) \right| \\
&= \sum_{q} \left| P(0,q) - P(1,q) \right| \cdot \sum_{r} \left| P(0,r) - P(1,r) \right| \\
&= \mathrm{K}_n^2,
\end{aligned}
$$

Polarization of K_n (memoryless case), '$+$' transform

$$
\begin{aligned}
\mathsf{K}_{n+1} &= \sum_{t,q,r} \left| P_{\mathsf{T}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i}(t,0,q,r) - P_{\mathsf{T}_i,\mathsf{V}_i,\mathsf{Q}_i,\mathsf{R}_i}(t,1,q,r) \right| \\
&= \sum_{t,q,r} |P(t,q)P(0,r) - P(t+1,q)P(1,r)| \\
&\overset{(*)}{\leq} \frac{1}{2} \sum_{t,q,r} P(q)\,|P(0,r) - P(1,r)| + P(r)\,|P(t,q) - P(t+1,q)| \\
&= \frac{1}{2} \sum_{t,r} |P(0,r) - P(1,r)| + \frac{1}{2} \sum_{t,q} |P(t,q) - P(t+1,q)| \\
&= 2\mathsf{K}_n,
\end{aligned}
$$

Identity for $(*)$: For any a, b, c, d:

$$
ab - cd = \frac{(a+c)(b-d) + (b+d)(a-c)}{2}
$$

Polarization of $\hat{K}_n$ (memory)

- Follows steps of memoryless case
- Requires additional inequalities
  - **Inequality I**: For states $s_0, s_N, s_{2N} \in \mathcal{S}$,

$$P(s_0, s_N, s_{2N}) = \frac{P(s_0, s_N) \cdot P(s_N, s_{2N})}{P(s_N)}$$

$$\leq \psi \cdot P(s_0, s_N) \cdot P(s_N, s_{2N})$$

    where

$$\psi = \max_s \frac{1}{\pi(s)}$$

  - **Inequality II**: For $f, g \geq 0$,

$$\sum_{s_N} f(s_N) g(s_N) \leq \sum_{s_N} f(s_N) \sum_{s'_N} g(s'_N)$$

Connections

Extreme Values

$H \approx 0 \Leftrightarrow Z \approx 0 \Leftrightarrow K \approx 1$

$H \approx 1 \Leftrightarrow Z \approx 1 \Leftrightarrow K \approx 0$

also for $(\hat{\cdot})$ processes

Ordering

$\hat{H}_n \leq H_n$

$\hat{Z}_n \leq Z_n$

$\hat{K}_n \geq K_n$

All six processes $(H_n, \hat{H}_n, Z_n, \hat{Z}_n, K_n, \hat{K}_n)$ polarize fast both to 0 and 1 with any $\beta < 1/2$

Summary

- A general framework for memory:

$$P(x_i, y_i, s_i | s_{i-1})$$

  - Memory allowed in both source and channel
  - State sequence S_i
    - Hidden
    - Stationary
    - Finite state Markov
    - Aperiodic and irreducible
- Achieve rate $I_\star(X;Y)$ through polar codes
- No change to polarization exponent ($\beta < 1/2$)