Optimal Reliability over a DMC with Feedback

Tara Javidi

Joint work with M. Naghshvar, O. Shayevitz, M. Wigger

Department of Electrical and Computer Engineering
University of California San Diego
Variable-length Coding with Feedback
Transmitter to communicate a message to the receiver

- $\theta$ belongs to the message set $\Omega := \{1, 2, \ldots, M\}$.
- Uniform prior $\Pr(\theta = i) = \frac{1}{M}$
Transmitter to communicate a message to the receiver
- \( \theta \) belongs to the message set \( \Omega := \{1, 2, \ldots, M\} \).
- Uniform prior \( \Pr(\theta = i) = \frac{1}{M} \).

Over a discrete memoryless channel (DMC) with noiseless feedback
- Input and output sets \( \mathcal{X} \) and \( \mathcal{Y} \), and \( P(Y|X) \)

\[
C = \max_{P_X} I(X; Y), \quad \text{and} \quad C_1 = \max_{x, x' \in \mathcal{X}} \frac{D(P(Y|X = x) || P(Y|X = x'))}{P_X},
\]
Variable-length Coding with Feedback

Transmitter to communicate a message to the receiver
- \( \theta \) belongs to the message set \( \Omega := \{1, 2, \ldots, M\} \).
- Uniform prior \( \Pr(\theta = i) = \frac{1}{M} \)

Over a discrete memoryless channel (DMC) with noiseless feedback
- Input and output sets \( \mathcal{X} \) and \( \mathcal{Y} \), and \( P(Y|X) \)

\[
C = \max_{P_X} I(X; Y), \quad \text{and} \quad C_1 = \max_{x, x' \in \mathcal{X}} D(P(Y|X = x)||P(Y|X = x')).
\]

Goal: conveying the message quickly and accurately
## Variable-length Coding with Feedback

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>(\tau - 1)</th>
<th>(\tau)</th>
</tr>
</thead>
<tbody>
<tr>
<td>channel input</td>
<td>(X_0)</td>
<td>(X_1)</td>
<td>...</td>
<td>(X_{\tau - 1})</td>
<td></td>
</tr>
<tr>
<td>channel output</td>
<td>(Y_0)</td>
<td>(Y_1)</td>
<td>...</td>
<td>(Y_{\tau - 1})</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td>(\hat{\theta})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Variable-length Coding with Feedback

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>channel input</td>
<td>$X_0$</td>
<td>$X_1$</td>
<td>$\ldots$</td>
<td>$X_{\tau-1}$</td>
<td></td>
</tr>
<tr>
<td>channel output</td>
<td>$Y_0$</td>
<td>$Y_1$</td>
<td>$\ldots$</td>
<td>$Y_{\tau-1}$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta}$</td>
</tr>
</tbody>
</table>

- Encoder produces channel inputs for some encoding function $e_t : \Omega \times Y^t \to X$
  $$X_t = e_t(\theta, Y_0, Y_1, \ldots, Y_{t-1}),$$
Variable-length Coding with Feedback

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>channel input</td>
<td>$X_0$</td>
<td>$X_1$</td>
<td>...</td>
<td>$X_{\tau-1}$</td>
<td></td>
</tr>
<tr>
<td>channel output</td>
<td>$Y_0$</td>
<td>$Y_1$</td>
<td>...</td>
<td>$Y_{\tau-1}$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta}$</td>
</tr>
</tbody>
</table>

- Encoder produces channel inputs for some encoding function $e_t : \Omega \times \mathcal{Y}^t \to \mathcal{X}$

  $$X_t = e_t(\theta, Y_0, Y_1, \ldots, Y_{t-1}),$$

- Decoder at the (random) decoding time $\tau$ guesses the message $\theta$ as for some decoding function $d : \mathcal{Y}^\tau \to \Omega$

  $$\hat{\theta} = d(Y_0, Y_1, \ldots, Y_{\tau-1}),$$
Variable-length Coding with Feedback

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>(\tau - 1)</th>
<th>(\tau)</th>
</tr>
</thead>
<tbody>
<tr>
<td>channel input</td>
<td>(X_0)</td>
<td>(X_1)</td>
<td>...</td>
<td>(X_{\tau-1})</td>
<td></td>
</tr>
<tr>
<td>channel output</td>
<td>(Y_0)</td>
<td>(Y_1)</td>
<td>...</td>
<td>(Y_{\tau-1})</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td>(\hat{\theta})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Encoder** produces channel inputs for some encoding function \(e_t : \Omega \times \mathcal{Y}^t \to \mathcal{X}\)

  \[
  X_t = e_t(\theta, Y_0, Y_1, \ldots, Y_{t-1}),
  \]

- **Decoder** at the (random) decoding time \(\tau\) guesses the message \(\theta\) as for some decoding function \(d : \mathcal{Y}^\tau \to \Omega\)

  \[
  \hat{\theta} = d(Y_0, Y_1, \ldots, Y_{\tau-1}),
  \]

**Objective:** Find \(\tau, e_0(\cdot), e_1(\cdot), \ldots, e_{\tau-1}(\cdot), d(\cdot)\) such that

1. probability of error satisfies \(\Pr(\hat{\theta} \neq \theta) \leq \epsilon\); and
2. the expected number of observations \(\mathbb{E}[\tau]\) is minimized
Overview

- Variable-length coding over DMC with feedback

- Capacity [Shannon '56]
  \[ C = \max_{P_X} I(X; Y) \]

- Optimal reliability [Burnashev '74]
  \[ E(R) = C_1(1 - \frac{R}{C}) \]
Overview

- Variable-length coding over DMC with feedback
  - Capacity [Shannon '56] \( C = \max_{P_X} I(X;Y) \)
  - Optimal reliability [Burnashev '74] \( E(R) = C_1(1 - \frac{R}{C}) \)

- Achievable coding schemes
  - Capacity: posterior matching (capacity achieving input distribution)
  - Reliability: two phase coding scheme (capacity code + Ack/Nack)
Overview

- Variable-length coding over DMC with feedback
  - Capacity [Shannon '56] \( C = \max_{P_X} I(X; Y) \)
  - Optimal reliability [Burnashev '74] \( E(R) = C_1(1 - \frac{R}{C}) \)

- Achievable coding schemes
  - Capacity: posterior matching (capacity achieving input distribution)
  - Reliability: two phase coding scheme (capacity code + Ack/Nack)

- How essential these attributes are?
  - Sequential deterministic single-phase coding scheme
  - Optimal reliability (+ achieving capacity)
Prior Work: Optimal Reliability

- The problem was studied by Burnashev (1974 and 1976)
Prior Work: Optimal Reliability

- The problem was studied by Burnashev (1974 and 1976)

**Lower bound:**

$$\mathbb{E}[\tau^*] \geq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}),$$
The problem was studied by Burnashev (1974 and 1976)

**Lower bound:**

\[
\mathbb{E}[\tau^*] \geq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}),
\]

**Upper bound**

\[
\mathbb{E}[\tau^*] \leq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} + O(\log \log \frac{M}{\epsilon})
\]
Prior Work: Optimal Reliability

- The problem was studied by Burnashev (1974 and 1976)

**Lower bound:**

\[ \mathbb{E}[\tau^*] \geq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}), \]

**Upper bound**

\[ \mathbb{E}[\tau^*] \leq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} + O(\log \log \frac{M}{\epsilon}) \]

- Achievability: Two different phases of operation:
Prior Work: Optimal Reliability

- The problem was studied by Burnashev (1974 and 1976)

**Lower bound:**

\[ \mathbb{E}[\tau^*] \geq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}), \]

**Upper bound**

\[ \mathbb{E}[\tau^*] \leq \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} + O(\log \log \frac{M}{\epsilon}) \]

- Achievability: Two different phases of operation:
  - Phase 1 refines the receiver’s belief about \( \theta \) (via random [Burnashev 76] or capacity achieving codes [Yamamoto Itoh 79], [Caire et al 06])
  - Phase 2 verifies the outcome of Phase 1 (via sending Ack/Nack signals)
For coding scheme $\pi$:

$M^\pi(T, \epsilon) := \text{the maximum number of hypothesis that can be identified with } E^\pi[\tau] \leq T \text{ and } Pe^\pi \leq \epsilon$ [Polyanskiy et al '11]
Prior Work: Optimal Reliability

For coding scheme $\pi$:

- $M^\pi(T, \epsilon) :=$ the maximum number of hypotheses that can be identified with $E^\pi[\tau] \leq T$ and $P_e^\pi \leq \epsilon$ [Polyanskiy et al '11]

- Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if

\[
\lim_{T \to \infty} \frac{1}{T} \log M^\pi(T, 2^{-ET}) = R
\]
For coding scheme $\pi$:

- $M^{\pi}(T, \epsilon) :=$ the maximum number of hypotheses that can be identified with $E^{\pi}[T] \leq T$ and $P_{e^{\pi}} \leq \epsilon$ [Polyanskiy et al '11]

- Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if

$$\lim_{T \to \infty} \frac{1}{T} \log M^{\pi}(T, 2^{-ET}) = R$$

Optimal Reliability $E(R) :=$ maximum $E$ at rate $R$. 
Prior Work: Optimal Reliability

For coding scheme $\pi$:

- $M^{\pi}(T, \epsilon) := \text{the maximum number of hypothesis that can be identified with } \mathbb{E}^{\pi}[\tau] \leq T \text{ and } \text{Pe}^{\pi} \leq \epsilon$ [Polyanskiy et al '11]

- Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if
  \[
  \lim_{T \to \infty} \frac{1}{T} \log M^{\pi}(T, 2^{-ET}) = R
  \]

Optimal Reliability $E(R) := \text{maximum } E \text{ at rate } R$.

Burnashev’s upper and lower bounds imply:

\[
E(R) = C_1 \left( 1 - \frac{R}{C} \right).
\]
Prior Work: Sequential Schemes

Horstein studied the binary input case (in 1963)
Horstein studied the binary input case (in 1963)

Transmit 0 (1) if $\theta < (>)$ median; random if $\theta = $ median
Prior Work: Sequential Schemes

Horstein studied the binary input case (in 1963)

Transmit 0 (1) if \( \theta < (>) \) median; random if \( \theta = \) median

Shayevitz & Feder generalized this scheme for all DMC

- Posterior is “matched” to capacity achieving \( P_X^* \)
- Proved to achieve capacity (some rare exceptions)
- Not known with what reliability (error exponent)
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
Active Hypothesis Testing

- $M$ mutually exclusive Hypothesis: $H_i \Leftrightarrow \{\theta = i\}, \ i = 1, 2, \ldots, M$
Active Hypothesis Testing

- $M$ mutually exclusive Hypothesis: $H_i \iff \{\theta = i\}, \ i = 1, 2, \ldots, M$

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>…</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensing action</td>
<td>$A(0)$</td>
<td>$A(1)$</td>
<td>…</td>
<td>$A(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>$Z(0)$</td>
<td>$Z(1)$</td>
<td>…</td>
<td>$Z(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta} = d(Z^{\tau - 1}, A^{\tau - 1})$</td>
</tr>
</tbody>
</table>
**Active Hypothesis Testing**

- $M$ mutually exclusive Hypothesis: $H_i \Leftrightarrow \{\theta = i\}$, $i = 1, 2, \ldots, M$

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensing action</td>
<td>$A(0)$</td>
<td>$A(1)$</td>
<td>...</td>
<td>$A(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>$Z(0)$</td>
<td>$Z(1)$</td>
<td>...</td>
<td>$Z(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta} = d(Z^{\tau - 1}, A^{\tau - 1})$</td>
<td></td>
</tr>
</tbody>
</table>

- $Z(t)|_{\{\theta = i, A(t) = a\}} \sim q_i^a(\cdot)$: observation density given $a \in A$ and $H_i$
Active Hypothesis Testing

- $M$ mutually exclusive Hypothesis: $H_i \Leftrightarrow \{\theta = i\}, \ i = 1, 2, \ldots, M$

- $\rho(0) = [\rho_1(0), \ldots, \rho_M(0)]$, $\rho_i(0) = P(\theta = i)$

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>$\ldots$</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensing action</td>
<td>$A(0)$</td>
<td>$A(1)$</td>
<td>$\ldots$</td>
<td>$A(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>$Z(0)$</td>
<td>$Z(1)$</td>
<td>$\ldots$</td>
<td>$Z(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta} = d(Z^{\tau-1}, A^{\tau-1})$</td>
<td></td>
</tr>
</tbody>
</table>

- $Z(t)|_{\{\theta = i, A(t) = a\}} \sim q_i^a(\cdot)$: observation density given $a \in A$ and $H_i$

Objective

Find $\tau$, $A(0)$, $A(1)$, $\ldots$, $A(\tau - 1)$, and $d(\cdot)$ that minimize $\mathbb{E}[\tau]$ s.t. $P_e \leq \epsilon$
Active Hypothesis Testing

- \( M \) mutually exclusive Hypothesis: \( H_i \Leftrightarrow \{ \theta = i \}, \ i = 1, 2, \ldots, M \)
- Prior \( \rho(0) = [\rho_1(0), \ldots, \rho_M(0)] \), \( \rho_i(0) = P(\theta = i) \)

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>\ldots</th>
<th>\tau - 1</th>
<th>\tau</th>
</tr>
</thead>
<tbody>
<tr>
<td>sensing action</td>
<td>( A(0) )</td>
<td>( A(1) )</td>
<td>\ldots</td>
<td>( A(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>( Z(0) )</td>
<td>( Z(1) )</td>
<td>\ldots</td>
<td>( Z(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td>( \hat{\theta} = d(Z^{\tau - 1}, A^{\tau - 1}) )</td>
<td></td>
</tr>
</tbody>
</table>

- \( Z(t)|_{\{ \theta = i, A(t) = a \}} \sim q_i^a(\cdot) \): observation density given \( a \in A \) and \( H_i \)

**Objective**

*Find \( \tau, A(0), A(1), \ldots, A(\tau - 1), \) and \( d(\cdot) \) that minimize \( \mathbb{E} [\tau] \) s.t. \( Pe \leq \epsilon \)
Equivalent Dynamical System View

- \( M \) mutually exclusive Hypothesis: \( H_i \Leftrightarrow \{ \theta = i \}, \ i = 1, 2, \ldots, M \)
- Prior \( \rho(0), \ \rho_i(0) = P(\theta = i) \), Belief vector \( \rho(t), \ \rho_i(t) = P(\theta = i|A^{t-1}, Z^{t-1}) \)

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>\ldots</th>
<th>( \tau - 1 )</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>belief vector</td>
<td>( \rho(0) )</td>
<td>( \rho(1) )</td>
<td>\ldots</td>
<td>( \rho(\tau - 1) )</td>
<td>( \rho(\tau) )</td>
</tr>
<tr>
<td>sensing action</td>
<td>( A(0) )</td>
<td>( A(1) )</td>
<td>\ldots</td>
<td>( A(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>( Z(0) )</td>
<td>( Z(1) )</td>
<td>\ldots</td>
<td>( Z(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td>( \hat{\theta} = d(Z^{\tau-1}, A^{\tau-1}) )</td>
<td></td>
</tr>
</tbody>
</table>

- \( Z(t)|_{\{ \theta = i, A(t) = a \}} \sim q_i^a(\cdot) \): observation density given \( a \in A \) and \( H_i \)

Objective

\textit{Find} \( \tau, A(0), A(1), \ldots, A(\tau - 1), \) and \( d(\cdot) \) \textit{that minimize} \( \mathbb{E}[\tau] \) \textit{s.t.} \( Pe \leq \epsilon \)
**Equivalent Dynamical System View**

- **$M$ mutually exclusive Hypothesis:** $H_i \iff \{\theta = i\}, \ i = 1, 2, \ldots, M$

- **Prior $\rho(0), \ \rho_i(0) = P(\theta = i), \ \text{Belief vector} \ \rho(t), \ \rho_i(t) = P(\theta = i|A^{t-1}, Z^{t-1})$$

| time | 0     | 1     | $\ldots$ | $\tau - 1$ | $\tau$ |
|------|-------|-------|-----------|------------|
| belief vector | $\rho(0)$ | $\rho(1)$ | $\ldots$ | $\rho(\tau - 1)$ | $\rho(\tau)$ |
| sensing action | $A(0)$ | $A(1)$ | $\ldots$ | $A(\tau - 1)$ |
| observation | $Z(0)$ | $Z(1)$ | $\ldots$ | $Z(\tau - 1)$ |
| retire-declare action | | | | | $\hat{\theta} = d(\rho(\tau))$ |

- $Z(t)|_{\{\theta=i, A(t) = a\}} \sim q_{i}^a(\cdot)$: observation density given $a \in A$ and $H_i$

---

**Objective**

Find $\tau, A(0), A(1), \ldots, A(\tau - 1), \text{ and } d(\cdot)$ that minimize $\mathbb{E}[\tau]$ s.t. $Pe \leq \epsilon$
**Equivalent Dynamical System View**

- \( M \) mutually exclusive Hypothesis: \( H_i \Leftrightarrow \{ \theta = i \}, \; i = 1, 2, \ldots, M \)
- Prior \( \rho(0), \; \rho_i(0) = P(\theta = i) \), Belief vector \( \rho(t), \; \rho_i(t) = P(\theta = i|A^{t-1}, Z^{t-1}) \)

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>\ldots</th>
<th>( \tau - 1 )</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>belief vector</td>
<td>( \rho(0) )</td>
<td>( \rho(1) )</td>
<td>\ldots</td>
<td>( \rho(\tau - 1) )</td>
<td>( \rho(\tau) )</td>
</tr>
<tr>
<td>sensing action</td>
<td>( A(0) )</td>
<td>( A(1) )</td>
<td>\ldots</td>
<td>( A(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>( Z(0) )</td>
<td>( Z(1) )</td>
<td>\ldots</td>
<td>( Z(\tau - 1) )</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td>( \hat{\theta} = \arg\max_i \rho_i(\tau) )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- \( Z(t)|_{\{ \theta = i, A(t) = a \}} \sim q_i^a(\cdot) \): observation density given \( a \in A \) and \( H_i \)

**Objective**

Find \( \tau, A(0), A(1), \ldots, A(\tau - 1) \) that minimize \( \mathbb{E}[\tau] \) s.t. \( \mathbb{E}[1 - \max_j \rho_j(\tau)] \leq \epsilon \)
**Equivalent Dynamical System View**

- $M$ mutually exclusive Hypothesis: $H_i \iff \{\theta = i\}, i = 1, 2, \ldots, M$
- Prior $\rho(0), \rho_i(0) = P(\theta = i)$, Belief vector $\rho(t), \rho_i(t) = P(\theta = i | A^{t-1}, Z^{t-1})$

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>...</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>belief vector</td>
<td>$\rho(0)$</td>
<td>$\rho(1)$</td>
<td>...</td>
<td>$\rho(\tau - 1)$</td>
<td>$\rho(\tau)$</td>
</tr>
<tr>
<td>sensing action</td>
<td>$A(0)$</td>
<td>$A(1)$</td>
<td>...</td>
<td>$A(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>$Z(0)$</td>
<td>$Z(1)$</td>
<td>...</td>
<td>$Z(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\hat{\theta} = \arg \max_i \rho_i(\tau)$</td>
</tr>
</tbody>
</table>

- $Z(t) | \{\theta = i, A(t) = a\} \sim q_i^a (\cdot)$: observation density given $a \in A$ and $H_i$

$$\rho_i(t + 1) = \frac{q_i^{A(t)}(Z(t))}{\sum_j \rho_j(t) q_j^{A(t)}(Z(t))} \rho_i(t)$$

**Objective**

Find $\tau, A(0), A(1), \ldots, A(\tau - 1)$ that minimize $\mathbb{E} [\tau]$ s.t. $\mathbb{E} [1 - \max_j \rho_j(\tau)] \leq \epsilon$
Sequential problem:
- Each action has an effect over the entire decision making horizon
Sequential versus Single-shot

**Sequential problem:**
- Each action has an effect over the entire decision making horizon

**Single-shot problem:**
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

$$\mathcal{IU}(a, \rho, V) = V(\rho) - \mathbb{E}[V(\Phi^a(\rho, Z))]$$
Sequential versus Single-shot

**Sequential problem:**
- Each action has an effect over the entire decision making horizon

**Single-shot problem:**
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

$$\mathcal{IU}(a, \rho, V) = V(\rho) - \mathbb{E}[V(\Phi^a(\rho, Z))]$$

Bayes operator
Sequential versus Single-shot

**Sequential problem:**
- Each action has an effect over the entire decision making horizon

**Single-shot problem:**
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

\[
IU(a, \rho, V) = V(\rho) - \mathbb{E}[V(\Phi^a(\rho, Z))]
\]

Sequential problem reduces to a sequence of single-shots with $V^*$
Sequential versus Single-shot

**Sequential problem:**
- Each action has an effect over the entire decision making horizon

**Single-shot problem:**
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

$$\mathcal{IU}(a, \rho, V) = V(\rho) - \mathbb{E}[V(\Phi^a(\rho, Z))]$$

Bayes operator

**Optimal policy:**
$$\pi^*(\rho) = \arg \max_a \mathcal{IU}(a, \rho, V^*)$$
Sequential versus Single-shot

Sequential problem:
- Each action has an effect over the entire decision making horizon

Single-shot problem:
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

\[ IU(a, \rho, V) = V(\rho) - E[V(\Phi^a(\rho, Z))] \]

Sequential problem reduces to a sequence of single-shots with $V^*$

Optimal policy:
\[ \pi^*(\rho) = \arg \max_a IU(a, \rho, V^*) \]

$V^*$ solves the dynamic programming equation:

\[ V^*(\rho) = \begin{cases} 
1 + \min_a E[V^*(\Phi^a(\rho, Z))] & \text{max}_j \rho_j < 1 - \epsilon \\
0 & \text{max}_j \rho_j \geq 1 - \epsilon 
\end{cases} \]
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds ⇒ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
DP can be used to obtain lower bounds for $V^*$
Implications of Dynamic Programming (I)

- DP can be used to obtain lower bounds for $V^*$
Implications of Dynamic Programming (I)

- DP can be used to obtain lower bounds for $V^*$
  - if $\max_i \rho_i > 1 - l^{-1} \implies$ optimal to stop
DP can be used to obtain lower bounds for $V^*$

- if $\max_i \rho_i > 1 - l^{-1} \Rightarrow$ optimal to stop
- how fast $\rho(t)$ can reach set $\{\rho : \max_i \rho_i \geq 1 - l^{-1}\}$?
Implications of Dynamic Programming (I)

- DP can be used to obtain lower bounds for $V^*$
  - if $\max_i \rho_i > 1 - l^{-1} \Rightarrow$ optimal to stop
  - how fast $\rho(t)$ can reach set $\{\rho : \max_i \rho_i \geq 1 - l^{-1}\}$?

\[
\mathbb{E}\{\tau\} \geq \frac{\mathbb{E}U(\rho(\tau)) - U(\rho)}{\max \mathbb{E}\Delta U}
\]
Implications of Dynamic Programming (I)

- DP can be used to obtain lower bounds for $V^*$
  - if $\max_i \rho_i > 1 - l^{-1} \Rightarrow$ optimal to stop
  - how fast $\rho(t)$ can reach set $\{\rho : \max_i \rho_i \geq 1 - l^{-1}\}$?

  $$\mathbb{E}\{\tau\} \geq \frac{\mathbb{E} U(\rho(\tau)) - U(\rho)}{\max \mathbb{E} \Delta U}$$

- Concise proof using the following lemma
Implications of Dynamic Programming (I)

- DP can be used to obtain lower bounds for $V^*$
  - if $\max_i \rho_i > 1 - l^{-1} \Rightarrow$ optimal to stop
  - how fast $\rho(t)$ can reach set $\{ \rho : \max_i \rho_i \geq 1 - l^{-1} \}$?

$$\mathbb{E}\{\tau\} \geq \frac{\mathbb{E}U(\rho(\tau)) - U(\rho)}{\max \mathbb{E}\Delta U}$$

- Concise proof using the following lemma

**Lemma**

Suppose there exists a functional $V : \mathbb{P}(\Theta) \rightarrow \mathbb{R}_+$ such that for all belief vectors $\rho \in \mathbb{P}(\Theta)$

$$V(\rho) \leq \begin{cases} 1 + \min_a \mathbb{E}[V(\Phi^a(\rho, Z))] & \max_j \rho_j < 1 - \epsilon \\
0 & \max_j \rho_j \geq 1 - \epsilon \end{cases}$$

Then $V(\rho) \leq V^*(\rho)$ for all $\rho \in \mathbb{P}(\Theta)$. 
Assumption 1. Both $C$ and $C_1$ are strictly positive and finite.

Assumption 2. There exists finite constant $\xi < \infty$ such that
\[
\max_{x,x' \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \frac{P(Y = y | X = x)}{P(Y = y | X = x')} < \xi.
\]

Proposition
\[
V^*(\rho) \geq \left[ (1 - \frac{2}{\log \frac{4}{\epsilon}} - \epsilon \log \frac{1}{\epsilon}) H(\rho) - 2 \frac{C}{C'} \right. \\
\left. + \frac{\log \frac{1}{\epsilon} - 2 \log \log \frac{2}{\epsilon} - \log \xi - 4}{C_1} \right]^+. 
\]
Alternative and Concise Proof of Converse

**Assumption 1.** Both $C$ and $C_1$ are strictly positive and finite.

**Assumption 2.** There exists finite constant $\xi < \infty$ such that
\[
\max_{x, x' \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \frac{P(Y = y | X = x)}{P(Y = y | X = x')} < \xi.
\]

**Proposition**
\[
\mathbb{E}[\tau^*] \gtrsim \frac{\log M}{C} + \frac{\log \frac{1}{\epsilon}}{C_1} - O(\log \log \frac{M}{\epsilon}).
\]
Assumption 1. Both $C$ and $C_1$ are strictly positive and finite.

Assumption 2. There exists finite constant $\xi < \infty$ such that
$$\max_{x,x' \in \mathcal{X}} \sup_{y \in \mathcal{Y}} \frac{P(Y=y|X=x)}{P(Y=y|X=x')} < \xi.$$ 

Corollary

At rates higher than $C$, error probability approaches 1. Furthermore,

$$E(R) \leq C_1 \left(1 - \frac{R}{C}\right)$$
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds ⇒ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
DeGroot’s Information Utility

**Sequential problem:**
- Each action has an effect over the entire decision making horizon

**Single-shot problem:**
- Measure of uncertainty $V$ [DeGroot 1962]
- Information utility associated with $V$

$$\mathcal{IU}(a, \rho, V) = V(\rho) - \mathbb{E}[V(\Phi^a(\rho, Z))]$$

Bayes operator

Sequential problem reduces to a sequence of single-shots with $V^*$

**Optimal policy:**
$$\pi^*(\rho) = \arg \max_a \mathcal{IU}(a, \rho, V^*)$$

$V^*$ solves the **dynamic programming equation**:

$$V^*(\rho) = \begin{cases} 
1 + \min_a \mathbb{E}[V^*(\Phi^a(\rho, Z))] & \text{if } \max_j \rho_j < 1 - \epsilon \\
0 & \text{if } \max_j \rho_j \geq 1 - \epsilon
\end{cases}$$
Consider (suboptimal) $\tau = \min \{ t : \max_i \rho_i(t) \geq 1 - \epsilon \}$

Stop transmission; decode to $\hat{\theta} = i$ if $\rho_i(t) \geq 1 - \epsilon$ (satisfies $P_e \leq \epsilon$)
Consider (suboptimal) $\tau = \min\{t : \max_i \rho_i(t) \geq 1 - \epsilon\}$

Stop transmission; decode to $\hat{\theta} = i$ if $\rho_i(t) \geq 1 - \epsilon$ (satisfies $Pe \leq \epsilon$)

Take concave function $W$
Achievability Analysis

- Consider (suboptimal) $\tau = \min \{ t : \max \rho_i(t) \geq 1 - \epsilon \}$
  - Stop transmission; decode to $\hat{\theta} = i$ if $\rho_i(t) \geq 1 - \epsilon$ (satisfies $P_e \leq \epsilon$)
- Take concave function $W$ (bounded $|W(\rho(t+1)) - W(\rho(t))| \leq \Delta$).
Achievability Analysis

Consider (suboptimal) \( \tau = \min\{t : \max_i \rho_i(t) \geq 1 - \epsilon\} \)

- Stop transmission; decode to \( \hat{\theta} = i \) if \( \rho_i(t) \geq 1 - \epsilon \) (satisfies \( P_e \leq \epsilon \))

- Take concave function \( W \) (bounded \( |W(\rho(t+1)) - W(\rho(t))| \leq \Delta \)).

- Suppose policy \( \pi \) selects encoder \( e_t, t < \tau \), such that
Consider (suboptimal) $\tau = \min\{t : \max_i \rho_i(t) \geq 1 - \epsilon\}$

Stop transmission; decode to $\hat{\theta} = i$ if $\rho_i(t) \geq 1 - \epsilon$ (satisfies $Pe \leq \epsilon$)

Take concave function $W$ (bounded $|W(\rho(t+1)) - W(\rho(t))| \leq \Delta$).

Suppose policy $c$ selects encoder $e_t$, $t < \tau$, such that

$\mathcal{I}(e(t), \rho(t), W) \geq \alpha$, for some positive $\alpha$. 
Consider (suboptimal) \( \tau = \min \{ t : \max_i \rho_i(t) \geq 1 - \epsilon \} \)

Stop transmission; decode to \( \hat{\theta} = i \) if \( \rho_i(t) \geq 1 - \epsilon \) (satisfies \( P_e \leq \epsilon \))

Take concave function \( W \) (bounded \( |W(\rho(t+1)) - W(\rho(t))| \leq \Delta \)).

Suppose policy \( \xi \) selects encoder \( e_t, t < \tau \), such that
\[
I(e(t), \rho(t), W) \geq \alpha, \text{ for some positive } \alpha.
\]

Then,
\[
\tau^* \lesssim \frac{W(\rho) - W([1 - \epsilon, \epsilon])}{\alpha} + \frac{\Delta}{\alpha}.
\]
Heuristic Policies

Entropy, $H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$, measures of uncertainty:¹

Heuristic Policies

Entropy, $H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$, measures of uncertainty: \footnote{Chaloner Verdinelli 1995, Lindley 1956, MacKay 1992, Paninski 2005, Branson 2010}

$$IU(e, \rho, H) = H(\rho) - \mathbb{E}(H(\Phi^e(\rho, Y)))$$
Heuristic Policies

Entropy, \( H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i} \), measures of uncertainty:\(^1\)

\[
\mathcal{IU}(e, \rho, H) = H(\rho) - \mathbb{E}(H(\Phi^e(\rho, Y)))
\]

\[= I(\theta; Y^e), \quad \text{where } Y^e \sim q^e_\rho = \sum_{i=1}^{M} \rho_i P(Y | X = e(i)).\]

---

\(^1\)[Chaloner Verdinelli 1995], [Lindley 1956], [MacKay 1992], [Paninski 2005], [Branson 2010]
Heuristic Policies

Entropy, $H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$, measures of uncertainty:

$$I_U(e, \rho, H) = H(\rho) - \mathbb{E}(H(\Phi^e(\rho, Y)))$$

$$= I(\theta; Y^e), \quad \text{where } Y^e \sim q^e_\rho = \sum_{i=1}^{M} \rho_i P(Y | X = e(i)).$$

$$I(\theta; Y^e) = \sum_{i=1}^{M} \rho_i D \left( P(Y | X = e(i)) \parallel q^e_\rho(Y) \right)$$

---

Heuristic Policies

Entropy, \( H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i} \), measures of uncertainty:

\[ I(\theta; Y^e) = H(\rho) - \mathbb{E}(H(\Phi^e(\rho, Y))) \]

= \( I(\theta; Y^e) \), where \( Y^e \sim q^e_{\rho} = \sum_{i=1}^{M} \rho_i P(Y|X = e(i)) \).

\[ I(\theta; Y^e) = \sum_{i=1}^{M} \rho_i D \left( P \left( Y|X = e(i) \right) \parallel q^e_{\rho}(Y) \right) \]

Jensen-Shannon divergence [Lin 1991]

\[ \text{References:} \ [\text{Chaloner Verdinelli 1995}], [\text{Lindley 1956}], [\text{MacKay 1992}], [\text{Paninski 2005}], [\text{Branson 2010}] \]
Heuristic Policies

Entropy, $H(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1}{\rho_i}$, measures of uncertainty: ¹

$$IU(e, \rho, H) = H(\rho) - \mathbb{E}(H(\Phi^e(\rho, Y)))$$

$$= I(\theta; Y^e), \quad \text{where } Y^e \sim q^e_\rho = \sum_{i=1}^{M} \rho_i P(Y|X = e(i)).$$

$$I(\theta; Y^e) = \sum_{i=1}^{M} \rho_i D\left( P(Y|X = e(i)) \parallel q^e_\rho(Y) \right)$$

Jensen-Shannon divergence [Lin 1991]

As $\rho_i \to 1$, $I(\theta; Y^e) \to D\left( P(Y|X = e(i)) \parallel P(Y|X = e(i)) \right) = 0$

Extrinsic Jensen-Shannon Divergence

The Extrinsic Jensen-Shannon (EJS) divergence among probability density functions $q_1, q_2, \ldots, q_M$ with respect to $\rho = [\rho_1, \rho_2, \ldots, \rho_M]$ is defined as

$$EJS(\rho; q_1, q_2, \ldots, q_M) = \sum_{i=1}^{M} \rho_i D(q_i \parallel \sum_{k \neq i} \frac{\rho_k}{1-\rho_i} q_k)$$
The Extrinsic Jensen-Shannon (EJS) divergence among probability density functions $q_1, q_2, \ldots, q_M$ with respect to $\rho = [\rho_1, \rho_2, \ldots, \rho_M]$ is defined as

$$EJS(\rho; q_1, q_2, \ldots, q_M) = \sum_{i=1}^{M} \rho_i D(q_i \| \sum_{k \neq i} \frac{\rho_k}{1-\rho_i} q_k)$$

Generalization of J-divergence [Jefferys 73]
New Measure of Information Utility

**Extrinsic Jensen-Shannon Divergence**

The Extrinsic Jensen-Shannon (EJS) divergence among probability density functions \( q_1, q_2, \ldots, q_M \) with respect to \( \rho = [\rho_1, \rho_2, \ldots, \rho_M] \) is defined as

\[
EJS(\rho; q_1, q_2, \ldots, q_M) = \sum_{i=1}^{M} \rho_i D(q_i || \sum_{k \neq i} \frac{\rho_k}{1-\rho_i} q_k)
\]

Generalization of J-divergence [Jefferys 73]

**Lemma**

EJS is the information utility of the average likelihood function:

\[
EJS(\rho; q_1^a, \ldots, q_M^a) = \mathcal{U}(a, \rho, U), \text{ where } U(\rho) = \sum_{i=1}^{M} \rho_i \log \frac{1-\rho_i}{\rho_i}
\]
Achievability: Extrinsic Jensen–Shannon Coding

- Given the belief vector $\rho(t)$, and encoding function $e(\cdot)$
  - The achieved EJS is nothing but
    \[
    EJS(\rho(t), e) := EJS(\rho(t); P_{e(1)}, \ldots, P_{e(M)}).
    \]
- Extended to randomized encoding $\Gamma$: $EJS(\rho(t), \Gamma)$
Achievability: Extrinsic Jensen–Shannon Coding

- Given the belief vector $\rho(t)$, and encoding function $e(\cdot)$
  - The achieved EJS is nothing but
    \[
    EJS(\rho(t), e) := EJS(\rho(t); P_{e(1)}, \ldots, P_{e(M)}).
    \]
  
- Extended to randomized encoding $\Gamma$: $EJS(\rho(t), \Gamma)$

- Generalized Horstein:

  \[
  \gamma_{GHBZ}(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M
  \end{cases}
  \]

  where $k^* := \arg\min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right|$. 


Given the belief vector $\rho(t)$, and encoding function $e(\cdot)$

The achieved EJS is nothing but

$$EJS(\rho(t), e) := EJS(\rho(t); P_e(1), \ldots, P_e(M)).$$

Extended to randomized encoding $\Gamma$: $EJS(\rho(t), \Gamma)$

**Generalized Horstein:**

$$\gamma_{GHBZ}(i) = \begin{cases} 0 & 1 \leq i \leq k^* \\ 1 & k^* < i \leq M \end{cases} \text{ where } k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right|. $$

If $\pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2}$
Achievability: Extrinsic Jensen–Shannon Coding

- Given the belief vector $\rho(t)$, and encoding function $e(\cdot)$
  - The achieved EJS is nothing but
    \[
    EJS(\rho(t), e) := EJS(\rho(t); P_{e(1)}, \ldots, P_{e(M)}).
    \]

- Extended to randomized encoding $\Gamma$: $EJS(\rho(t), \Gamma)$

- Generalized Horstein:
  \[
  \gamma^{\text{GHBZ}}(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M 
  \end{cases}
  \]
  where $k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right|$. 

- If $\pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2}$
  \[
  EJS(\rho(t), \gamma^{\text{GHBZ}}) = \sum_{i=1}^{k^*} \rho_i(t) D \left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} \right\| P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right)
  + \sum_{i=k^*+1}^{M} \rho_i(t) D \left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right\| \right). 
  \]
**Analytic Results (II): Achievability**

**Assumption 1:** \( C_2 = \max_{i,j} \sup_y \frac{P(Y=y|X=i)}{P(Y=y|X=j)} < \infty \)

**Theorem 4**

*Suppose policy \( \varsigma \) selects actions \( A(t), t < \tau \) such that*

\[
\mathbb{E}[J_{\varsigma}(\rho(t), A(t))] \geq \alpha
\]

*for some positive value \( \alpha \). Then under Assumption 1,*

\[
\mathbb{E}[\tau^c] \leq \frac{\log M + \max\{\log \log M, \log \frac{1}{\epsilon}\}}{\alpha} + 4C_2 + 1.
\]
Analytic Results (II): Achievability

Theorem

Let

\[ \tilde{\rho} = 1 - \frac{1}{1 + \max\{\log M, \log \frac{1}{\epsilon}\}}. \]

Suppose policy \( \mathcal{c} \) selects actions \( A(t), t < \tau \) such that

\[ EJS(\rho(t), A(t)) \geq \begin{cases} 
\alpha & \text{if } \max_i \rho_i(t) < \tilde{\rho} \\
\tilde{\rho} \beta & \text{otherwise}
\end{cases}, \]

for some positive values \( \alpha \) and \( \beta \). Then under Assumption 1

\[ \mathbb{E}[\tau^c] \leq \frac{\log M + \max\{\log \log M, \log \log \frac{1}{\epsilon}\}}{\alpha} + \frac{\log \frac{1}{\epsilon}}{\beta} + \frac{6(4C_2)^2}{\alpha \beta} + 1. \]
Analytic Results (II): Achievability

Theorem

Let $\tilde{\rho} = \frac{1}{1 + \max\{\log M, \log \frac{1}{\epsilon}\}}$.

Suppose policy $c$ selects actions $A(t)$, $t < \tau$, such that $E_{JS}(\rho(t), A(t)) \geq \alpha$ if $\max_i \rho_i(t) < \tilde{\rho}$, and $\tilde{\rho}^\beta$ otherwise, for some positive values $\alpha$ and $\beta$.

Then under Assumption 1, $E[\tau_c] \leq \log M + \max\{\log \log M, \log \log \frac{1}{\epsilon}\} + \alpha + \log \frac{1}{\epsilon} \beta + 6(4C_2^2)^{\alpha \beta + 1}$.

Achievable Reliability

- $C_1$
- $\beta$
Sketch of the Proof

Upper bound for $\mathbb{E}[\tau]$
Sketch of the Proof

Upper bound for $\mathbb{E}[\tau]$

- Constructing a submartingale from $U(\rho) = \sum_i \rho_i \log \frac{\rho_i}{1-\rho_i}$
Sketch of the Proof

Upper bound for \( E[\tau] \)

- Constructing a submartingale from \( U(\rho) = \sum_i \rho_i \log \frac{\rho_i}{1-\rho_i} \)
- Recall that EJS is nothing but the mean drift and bounded by \( \alpha \)
Sketch of the Proof

Upper bound for $\mathbb{E}[\tau]$

- Constructing a submartingale from $U(\rho) = \sum_i \rho_i \log \frac{\rho_i}{1 - \rho_i}$
- Recall that EJS is nothing but the mean drift and bounded by $\alpha$

Lemma

Consider submartingale $\{\zeta_n\}$, $n = 0, 1, \ldots$ wrt $\{\mathcal{F}_n\}$ and stopping time $\tau_B = \min\{n : \zeta_n \geq B\}$, $B > 0$. If there exist positive constants $K_1 \leq K_2 \leq K_3$ such that

\[
\begin{align*}
\mathbb{E}[\zeta_{n+1} | \mathcal{F}_n] &\geq \zeta_n + K_1 \quad \text{if } \zeta_n < 0, \\
\mathbb{E}[\zeta_{n+1} | \mathcal{F}_n] &\geq \zeta_n + K_2 \quad \text{if } \zeta_n \geq 0, \\
|\zeta_{n+1} - \zeta_n| &\leq K_3, \quad \text{when } \zeta_n \geq 0,
\end{align*}
\]

then

\[
\mathbb{E}[\tau_B] \leq \frac{B - \zeta_0}{K_2} + \zeta_0 \mathbf{1}_{\{\zeta_0 < 0\}} \left( \frac{1}{K_2} - \frac{1}{K_1} \right) + 3 \frac{K_3^2}{K_1 K_2}.
\]
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
Generalized Horstein-Burnashev-Zigangarov

Generalize Horstein’s for binary input symmetric channels:

\[ \gamma_{\text{GHBZ}}(i) = \begin{cases} 
0 & 1 \leq i \leq k^* \\
1 & k^* < i \leq M 
\end{cases} \]

where \( k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).
Generalize Horstein’s for binary input symmetric channels:

- **Generalized Horstein:**
  \[
  \gamma_{GHBZ}^*(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M 
  \end{cases}
  \]
  where \( k^* := \arg\min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).

- If \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)
  \[
  EJS(\rho(t), \gamma_{GHBZ}^*) = \sum_{i=1}^{k^*} \rho_i(t)D \left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right\| \right) 
  \]
  \[
  + \sum_{i=k^*+1}^{M} \rho_i(t)D \left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right\| \right)
  \]
Generalized Horstein-Burnashev-Zigangarov

- **Generalized Horstein:**

  \[
  \gamma^{\text{GH}BZ}(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M 
  \end{cases}
  \]

  where \( k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).

- **If** \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)

  \[
  EJS(\rho(t), \gamma^{\text{GH}BZ}) = \sum_{i=1}^{k^*} \rho_i(t) D\left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} \right\| P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right) 
  \]

  \[
  + \sum_{i=k^*+1}^{M} \rho_i(t) D\left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} \right\| P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right)
  \]
Generalized Horstein-Burnashev-Zigangarov

- Generalized Horstein:
  \[
  \gamma_{GHBZ}(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M 
  \end{cases}
  \]

  where \( k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).

- If \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)

  \[
  EJS(\rho(t), \gamma_{GHBZ}) = \sum_{i=1}^{k^*} \rho_i(t)D\left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right\| \right) 
  \]

  \[
  + \sum_{i=k^*+1}^{M} \rho_i(t)D\left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right\| \right) 
  \]

  \[
  \geq \sum_{i=1}^{k^*} \rho_i(t)D\left( P_0 \left\| \pi_0(t)P_0 + \pi_1(t)P_1 \right\| \right) 
  \]

  \[
  + \sum_{i=k^*+1}^{M} \rho_i(t)D\left( P_1 \left\| \pi_0(t)P_0 + \pi_1(t)P_1 \right\| \right) 
  \]
Generalized Horstein-Burnashev-Zigangarov

- Generalized Horstein:
  
  \[
  \gamma_{GHBZ}^i = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M 
  \end{cases}
  \]

  where \( k^* := \arg\min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).

- If \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)

  \[
  EJS(\rho(t), \gamma_{GHBZ}^i) = \sum_{i=1}^{k^*} \rho_i(t) D\left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right\| \right)
  \]

  \[
  + \sum_{i=k^*+1}^{M} \rho_i(t) D\left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right\| \right)
  \]

  \[
  \geq \pi_0(t) D\left( P_0 \left\| \pi_0(t) P_0 + \pi_1(t) P_1 \right\| \right)
  \]

  \[
  + \pi_1(t) D\left( P_1 \left\| \pi_0(t) P_0 + \pi_1(t) P_1 \right\| \right)
  \]
Generalized Horstein-Burnashev-Zigangarov

- **Generalized Horstein:**

  \[
  \gamma_{\text{GHBZ}}(i) = \begin{cases} 
  0 & 1 \leq i \leq k^* \\
  1 & k^* < i \leq M
  \end{cases}
  \]
  where \( k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right|. \)

- If \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)

  \[
  EJS(\rho(t), \gamma_{\text{GHBZ}}) = \sum_{i=1}^{k^*} \rho_i(t) D \left( P_0 \left\| \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} \right\| P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right) \\
  + \sum_{i=k^*+1}^{M} \rho_i(t) D \left( P_1 \left\| \frac{\pi_0(t)}{1 - \rho_i(t)} \right\| P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right) \\
  \geq \pi_0(t) D \left( P_0 \left\| \pi_0(t) P_0 + \pi_1(t) P_1 \right\| \right) \\
  + \pi_1(t) D \left( P_0 \left\| \pi_0(t) P_1 + \pi_1(t) P_0 \right\| \right)
  \]
Generalized Horstein-Burnashev-Zigangarov

- Generalized Horstein:

\[ \gamma^{GHBZ}(i) = \begin{cases} 0 & 1 \leq i \leq k^* \\ 1 & k^* < i \leq M \end{cases} \quad \text{where } k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right|. \]

- If \( \pi_0(t) = 1 - \pi_1(t) = \sum_{i=1}^{k^*} \rho_i(t) \geq \frac{1}{2} \)

\[
EJS(\rho(t), \gamma^{GHBZ}) = \sum_{i=1}^{k^*} \rho_i(t) D \left( P_0 \parallel \frac{\pi_0(t) - \rho_i(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t)}{1 - \rho_i(t)} P_1 \right) \\
+ \sum_{i=k^*+1}^{M} \rho_i(t) D \left( P_1 \parallel \frac{\pi_0(t)}{1 - \rho_i(t)} P_0 + \frac{\pi_1(t) - \rho_i(t)}{1 - \rho_i(t)} P_1 \right) \\
\geq \pi_0(t) D \left( P_0 \parallel \pi_0(t) P_0 + \pi_1(t) P_1 \right) \\
+ \pi_1(t) D \left( P_0 \parallel \pi_0(t) P_1 + \pi_1(t) P_0 \right) \\
\geq D \left( P_0 \parallel \frac{1}{2} P_0 + \frac{1}{2} P_1 \right) \geq C
\]
Generalize Horstein-Burnashev-Zigangarov

Generalize Horstein’s for binary input symmetric channels:

\[
\gamma_{\text{GHBZ}}(i) = \begin{cases} 
0 & 1 \leq i \leq k^* \\
1 & k^* < i \leq M 
\end{cases}
\]

where \( k^* := \arg \min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).
Generalized Horstein-Burnashev-Zigangarov

Generalize Horstein’s for binary input symmetric channels:

\[ \gamma_{\text{GHBZ}}(i) = \begin{cases} 
0 & 1 \leq i \leq k^* \\
1 & k^* < i \leq M 
\end{cases} \]

where \( k^* := \arg\min_{k \in \Omega} \left| \sum_{i=1}^{k} \rho_i(t) - \frac{1}{2} \right| \).

**Proposition**

Consider the GHBZ scheme over a symmetric binary-input DMC. For every \( t = 0, 1, \ldots, \tau \) and all possible output sequences \( y^{t-1} \),

\[ EJS(\rho(t), \gamma_{\text{GHBZ}}) \geq C. \]
Modify Horstein’s scheme by allowing a reordering of messages:
Modify Horstein’s scheme by allowing a reordering of messages:

- Encoding function \( \gamma^* \) such that for all \( i \in \{ j \in \Omega : \gamma(j) = 0 \} \),

\[
0 \leq \sum_{j \in \Omega : \gamma(j) = 0} \rho_j(t) - \sum_{j \in \Omega : \gamma(j) = 1} \rho_j(t) < \rho_i(t).
\]
Modify Horstein’s scheme by allowing a reordering of messages:

- Encoding function $\gamma^*$ such that for all $i \in \{j \in \Omega: \gamma(j) = 0\}$,

$$0 \leq \sum_{j \in \Omega: \gamma(j) = 0} \rho_j(t) - \sum_{j \in \Omega: \gamma(j) = 1} \rho_j(t) < \rho_i(t).$$

- When all $\{\rho_i(t)\}_{i \in \Omega}$ small $\Rightarrow$ prob of sending 0 and 1 approx equal
Modify Horstein’s scheme by allowing a reordering of messages:

- Encoding function $\gamma^*$ such that for all $i \in \{ j \in \Omega : \gamma(j) = 0 \}$,

$$0 \leq \sum_{j \in \Omega : \gamma(j) = 0} \rho_j(t) - \sum_{j \in \Omega : \gamma(j) = 1} \rho_j(t) < \rho_i(t).$$

- When all $\{ \rho_i(t) \}_{i \in \Omega}$ small $\Rightarrow$ prob of sending 0 and 1 approx equal
- When $\max_{i \in \Omega} \rho_i(t) > 1/2 \Rightarrow$ send 0 iff $\theta = \arg \max_{i \in \Omega} \rho_i(t)$
Optimized Horstein-Burnashev-Zigangarov

Modify Horstein’s scheme by allowing a reordering of messages:
- Encoding function $\gamma^*$ such that for all $i \in \{j \in \Omega: \gamma(j) = 0\}$,

$$0 \leq \sum_{j \in \Omega: \gamma(j)=0} \rho_j(t) - \sum_{j \in \Omega: \gamma(j)=1} \rho_j(t) < \rho_i(t).$$

- When all $\{\rho_i(t)\}_{i \in \Omega}$ small $\Rightarrow$ prob of sending 0 and 1 approx equal
- When $\max_{i \in \Omega} \rho_i(t) > 1/2 \Rightarrow$ send 0 iff $\theta = \arg \max_{i \in \Omega} \rho_i(t)$

**Proposition**

*Consider the optimized coding scheme over a symmetric binary-input DMC. For every $t = 0, 1, \ldots, \tau$ and all possible output sequences $y^{t-1}$,*

$$EJS(\rho(t), \gamma^*) \geq C.$$  

*Furthermore, when $\max_{i} \rho_i(t) \geq \tilde{\rho}$*

$$EJS(\rho(t), \gamma^*) \geq \tilde{\rho}C_1.$$
Outline of the Rest of the Talk

- Connection to Active Sequential Hypothesis Testing
- From Sequential to Single-shot: DeGroot’s Information
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
  - Based on the capacity achieving input distribution $\pi^*$
Proposed by Shayevitz and Feder in 2007
- Randomization will likely be necessary
- Generalizes Horstein’s scheme for DMCs
- Based on the capacity achieving input distribution $\pi^*$
- Given posterior $\rho(t)$, ensures input distribution $\pi^*$
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
  - Based on the capacity achieving input distribution $\pi^*$
  - Given posterior $\rho(t)$, ensures input distribution $\pi^*$
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
  - Based on the capacity achieving input distribution $\pi^*$
  - Given posterior $\rho(t)$, ensures input distribution $\pi^*$

- If $\theta = i$
Proposed by Shayevitz and Feder in 2007
- Randomization will likely be necessary
- Generalizes Horstein’s scheme for DMCs
- Based on the capacity achieving input distribution $\pi^*$
- Given posterior $\rho(t)$, ensures input distribution $\pi^*$

If $\theta = i$
the input $X(t)$ takes value in the set

$$\mathcal{X}_i(t) := \left\{ x \in \mathcal{X} : \sum_{i' = 1}^{i-1} \rho_{i'}(t) < \sum_{x' \leq x} \pi^*_{x'} \text{ and } \sum_{x' < x} \pi^*_{x'} \leq \sum_{i' = 1}^{i} \rho_{i'}(t) \right\};$$
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
  - Based on the capacity achieving input distribution $\pi^*$
- Given posterior $\rho(t)$, ensures input distribution $\pi^*$

If $\theta = i$
the input $X(t)$ takes value in the set

$$X_i(t) := \left\{ x \in \mathcal{X} : \sum_{i' = 1}^{i-1} \rho_{i'}(t) < \sum_{x' \leq x} \pi^*_{x'} \text{ and } \sum_{x' < x} \pi^*_{x'} \leq \sum_{i' = 1}^{i} \rho_{i'}(t) \right\};$$

where each value $x \in X_i(t)$ is taken with probability

$$\min \left\{ \sum_{i' = 1}^{i} \rho_{i'}(t), \sum_{x' \leq x} \pi^*_{x'} \right\} - \max \left\{ \sum_{i' = 1}^{i-1} \rho_{i'}(t), \sum_{x' < x} \pi^*_{x'} \right\}$$

$$\frac{\rho_i(t)}{\rho_i(t)}.$$
General DMC: Variable-length Posterior Matching

- Proposed by Shayevitz and Feder in 2007
  - Randomization will likely be necessary
  - Generalizes Horstein’s scheme for DMCs
  - Based on the capacity achieving input distribution $\pi^*$
  - Given posterior $\rho(t)$, ensures input distribution $\pi^*$
Proposed by Shayevitz and Feder in 2007
- Randomization will likely be necessary
- Generalizes Horstein’s scheme for DMCs
- Based on the capacity achieving input distribution $\pi^*$
- Given posterior $\rho(t)$, ensures input distribution $\pi^*$

**Proposition**

Consider the posterior matching coding scheme over a DMC. For every $t = 0, 1, \ldots, \tau$ and all possible output sequences $y^{t-1}$,

$$EJS(\rho(t), \gamma^{PM}) \geq C.$$
Proposed by Shayevitz and Feder in 2007

Randomization will likely be necessary

Generalizes Horstein's scheme for DMCs

Based on the capacity achieving input distribution $\pi^*$

Given posterior $\rho(t)$, ensures input distribution $\pi^*$

**Proposition**

Consider the posterior matching coding scheme over a DMC. For every $t = 0, 1, \ldots, \tau$ and all possible output sequences $y^{t-1}$,

$$EJS(\rho(t), \gamma^{PM}) \geq C.$$
So far we analyzed $\alpha > 0$ and $\beta > 0$ for schemes
So far we analyzed $\alpha > 0$ and $\beta > 0$ for schemes.

Instead search for maximum $\alpha > 0$ and $\beta > 0$. 
So far we analyzed $\alpha > 0$ and $\beta > 0$ for schemes.

Instead search for maximum $\alpha > 0$ and $\beta > 0$.

Among all coding functions (finite) $\mathcal{E}$ select

$$\gamma^* := \arg \max_{\gamma \in \mathcal{E}} EJS(\rho(t), \gamma).$$
So far we analyzed $\alpha > 0$ and $\beta > 0$ for schemes

- Instead search for maximum $\alpha > 0$ and $\beta > 0$
- Among all coding functions (finite) $\mathcal{E}$ select

$$\gamma^* := \arg \max_{\gamma \in \mathcal{E}} EJS(\rho(t), \gamma).$$

**Proposition**

Consider the MaxEJS coding scheme over a DMC. For every $t$

$$EJS(\rho(t), \gamma^*) \geq C,$$

and furthermore,

$$EJS(\rho(t), \gamma^*) \geq \tilde{\rho}C_1 \quad \text{if } \max_{i \in \Omega} \rho_i(t) \geq \tilde{\rho}. \quad (1)$$
So far we analyzed $\alpha > 0$ and $\beta > 0$ for schemes.

And furthermore,

$$EJS(\rho(t), \gamma^*) \geq \tilde{\rho}C_1 \quad \text{if} \quad \max_{i \in \Omega} \rho_i(t) \geq \tilde{\rho}.$$  

(1)
Sketch of the Proof

Consider the drift

\[
\Delta(t) = \sum_{k=1}^{\lvert X \rvert} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)\|P(\tilde{Y}_{-i}))
\]

where \( \tilde{Y}_{-i}, i \in \Omega \) is output induced by the the extrinsic input \( \tilde{X}_{-i} \)

\[
P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) \setminus \{i\}} \rho_j(t)}{1 - \rho_i(t)}
\]
Sketch of the Proof

- Consider the drift

\[
\Delta(t) = \sum_{k=1}^{||\mathcal{X}||} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k) || P(\tilde{Y}_{-i}))
\]

where \(\tilde{Y}_{-i}, i \in \Omega\) is output induced by the extrinsic input \(\tilde{X}_{-i}\)

\[
P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}
\]

- Consider the drift under a random code where each message \(i \in \Omega\) is assigned to input \(k \in \mathcal{X}\) with \(P_{X^*}(k)\)
Sketch of the Proof

Consider the drift

\[ \Delta(t) = \sum_{k=1}^{\left|\mathcal{X}\right|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X=k)||P(\tilde{Y}_i)) \]

where \( \tilde{Y}_i, i \in \Omega \) is output induced by the the extrinsic input \( \tilde{X}_i \)

\[ \mathbb{P}(\tilde{X}_i = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)} \]

Consider the drift under a random code where each message \( i \in \Omega \) is assigned to input \( k \in \mathcal{X} \) with \( P_{X^*}(k) \)

- \( X^* \) is the capacity achieving input
Sketch of the Proof

Consider the drift

$$\Delta(t) = \sum_{k=1}^{\|X\|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)||P(\tilde{Y}_{-i}))$$

where $\tilde{Y}_{-i}, i \in \Omega$ is output induced by the extrinsic input $\tilde{X}_{-i}$

$$P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}$$

Consider the drift under a random code where each message $i \in \Omega$ is assigned to input $k \in X$ with $P_{X^*}(k)$

- $X^*$ is the capacity achieving input
- $\Delta^*(t) = \sum_{k=1}^{\|X\|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)||P(Y^*)) = C$
Sketch of the Proof

Consider the drift

$$\Delta(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k) \| P(\tilde{Y}_{-i}))$$

where $\tilde{Y}_{-i}$, $i \in \Omega$ is output induced by the extrinsic input $\tilde{X}_{-i}$

$$P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}$$

Consider the drift under a random code where each message $i \in \Omega$ is assigned to input $k \in \mathcal{X}$ with $P_{X^*}(k)$

- $X^*$ is the capacity achieving input
- $\Delta^*(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k) \| P(Y^*)) = C$

(i) By construction $\Delta(t) \geq \Delta^*(t) = C$
Consider the drift

\[
\Delta(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k) \parallel P(\tilde{Y}_{-i}))
\]

where \( \tilde{Y}_{-i}, i \in \Omega \) is output induced by the extrinsic input \( \tilde{X}_{-i} \)

\[
P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}
\]
Sketch of the Proof

- Consider the drift

\[ \Delta(t) = \sum_{k=1}^{\left| \mathcal{X} \right|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)\|P(\tilde{Y}_i)) \]

where \( \tilde{Y}_{-i}, i \in \Omega \) is output induced by the extrinsic input \( \tilde{X}_{-i} \)

\[ \mathbb{P}\left( \tilde{X}_{-i} = k \right) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)} \]

- If \( U \geq \log \frac{2M}{\epsilon} - 1 \), then \( \exists i \in \Omega \) for which \( \rho_i(t) \geq \log \frac{2M}{\epsilon} - 1 \)
Sketch of the Proof

- Consider the drift

\[
\Delta(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)\|P(\tilde{Y}_{-i}))
\]

where \(\tilde{Y}_{-i}, i \in \Omega\) is output induced by the extrinsic input \(\tilde{X}_{-i}\)

\[
P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}
\]

- If \(U \geq \log \frac{2M}{\epsilon} - 1\), then \(\exists i \in \Omega\) for which \(\rho_i(t) \geq \log \frac{2M}{\epsilon} - 1\)

- If \(i\) is assigned to \(x\) and all other messages to \(x'\)
Sketch of the Proof

- Consider the drift

$$\Delta(t) = \sum_{k=1}^{\left|\mathcal{X}\right|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)||P(\tilde{Y}_{-i}))$$

where $\tilde{Y}_{-i}, i \in \Omega$ is output induced by the extrinsic input $\tilde{X}_{-i}$

$$\mathbb{P}(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}$$

- If $U \geq \log \frac{2M}{\epsilon} - 1$, then $\exists i \in \Omega$ for which $\rho_i(t) \geq \log \frac{2M}{\epsilon} - 1$

- If $i$ is assigned to $x$ and all other messages to $x'$
  - $D(P(Y|X = x)||P(Y|X = x')) = C_1$
Sketch of the Proof

- Consider the drift

\[
\Delta(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)||P(\tilde{Y}_{-i}))
\]

where \(\tilde{Y}_{-i}, i \in \Omega\) is output induced by the extrinsic input \(\tilde{X}_{-i}\)

\[
P(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)}
\]

- If \(U \geq \log \frac{2M}{\epsilon} - 1\), then \(\exists i \in \Omega\) for which \(\rho_i(t) \geq \log \frac{2M}{\epsilon} - 1\)

- If \(i\) is assigned to \(x\) and all other messages to \(x'\)
  - \(D(P(Y|X = x)||P(Y|X = x')) = C_1\)
  - \(\hat{\Delta}(t) \geq \rho_i(t) D(P(Y|X = x)||P(Y|X = x')) = \rho_i(t) C_1\)
Sketch of the Proof

- Consider the drift

\[ \Delta(t) = \sum_{k=1}^{|\mathcal{X}|} \sum_{i \in S_k(t)} \rho_i(t) D(P(Y|X = k)||P(\tilde{Y}_{-i})) \]

where \( \tilde{Y}_{-i}, i \in \Omega \) is output induced by the the extrinsic input \( \tilde{X}_{-i} \)

\[ \mathbb{P}(\tilde{X}_{-i} = k) = \frac{\sum_{j \in S_k(t) - \{i\}} \rho_j(t)}{1 - \rho_i(t)} \]

- If \( U \geq \log \frac{2M}{\epsilon} - 1 \), then \( \exists i \in \Omega \) for which \( \rho_i(t) \geq \log \frac{2M}{\epsilon} - 1 \)

- If \( i \) is assigned to \( x \) and all other messages to \( x' \)
  - \( D(P(Y|X = x)||P(Y|X = x')) = C_1 \)
  - \( \hat{\Delta}(t) \geq \rho_i(t) D(P(Y|X = x)||P(Y|X = x')) = \rho_i(t)C_1 \)

(ii) By construction \( \Delta(t) \geq \hat{\Delta}(t) \geq C_1 (\log \frac{2M}{\epsilon} - 1) \)
In Summary...

- Variable Length Coding and Active Sequential Hypothesis Testing
- DeGroot’s Information Utility
  - Lower bounds $\Rightarrow$ Rate–reliability outerbound
  - Achievability: Extrinsic Jensen–Shannon Divergence
- Example: symmetric binary input channel
  - Generalize Horstein-Burnashev-zigangarov
  - Optimized sequential (deterministic) scheme
- General DMC
  - Posterior matching: rate–reliability trade-off
  - Optimal sequential (deterministic) scheme: MaxEJS
Active Hypothesis Testing

- $M$ mutually exclusive Hypothesis: $H_i \Leftrightarrow \{\theta = i\}, i = 1, 2, \ldots, M$
- Prior $\rho(0), \rho_i(0) = P(\theta = i)$, Belief vector $\rho(t), \rho_i(t) = P(\theta = i|A^{t-1}, Z^{t-1})$

<table>
<thead>
<tr>
<th>time</th>
<th>0</th>
<th>1</th>
<th>$\ldots$</th>
<th>$\tau - 1$</th>
<th>$\tau$</th>
</tr>
</thead>
<tbody>
<tr>
<td>belief vector</td>
<td>$\rho(0)$</td>
<td>$\rho(1)$</td>
<td>$\ldots$</td>
<td>$\rho(\tau - 1)$</td>
<td>$\rho(\tau)$</td>
</tr>
<tr>
<td>sensing action</td>
<td>$A(0)$</td>
<td>$A(1)$</td>
<td>$\ldots$</td>
<td>$A(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>observation</td>
<td>$Z(0)$</td>
<td>$Z(1)$</td>
<td>$\ldots$</td>
<td>$Z(\tau - 1)$</td>
<td></td>
</tr>
<tr>
<td>retire-declare action</td>
<td>$\hat{\theta} = \arg \max_i \rho_i(\tau)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$Z(t)|\{\theta = i, A(t) = a\} \sim q_i^a(\cdot)$: observation density given $a \in A$ and $H_i$

$$\rho_i(t + 1) = \frac{q_i^{A(t)}(Z(t))}{\sum_j \rho_j(t) q_j^{A(t)}(Z(t))} \rho_i(t)$$

Objective

Find $\tau, A(0), A(1), \ldots, A(\tau - 1)$ that minimize $\mathbb{E}[\tau]$ s.t. $\mathbb{E}[1 - \max_j \rho_j(\tau)] \leq \epsilon$
Sample size and sensing actions can be selected online or off-line

Sequential versus non-sequential

- Non-sequential: Fixed number of samples $\tau$
- Sequential: $\tau$ is selected based on observed samples

Adaptive versus non-adaptive

- Non-adaptive: Fixed (randomized) choice of actions
- Adaptive: Choice of action is adapted to observed samples

How much gain, if any?
Minimum expect cost: $\mathbb{E}\{\tau\}$

Non–sequential Non–adaptive

\[ \geq \frac{-2 \log \epsilon}{\max_{\lambda} \min_{i} D(i, \lambda)} \]

Sequential Non–adaptive

\[ \simeq \frac{-\log \epsilon}{\max_{\lambda} D(\lambda)} \]

Sequential Adaptive

\[ \simeq \frac{-\log \epsilon}{D^*} \]

\[ \bar{D}^* \geq \max_{\lambda} \bar{D}(\lambda) > \max_{\lambda} \min_{i} D(i, \lambda) \]

\[ D(f \| g) = \int f(y) \log \frac{f(y)}{g(y)} \, dy \]

\[ D(i, \lambda) = \min_{j \neq i} \sum_{a \in A} \lambda_a D(q^a_i \| q^a_j) \]
Minimum expect cost: $\mathbb{E}\{\tau\}$

Non-sequential Non-adaptive
\[ \succ \frac{-2 \log \epsilon}{\max \min_i D(i, \lambda)} \]

Sequential Non-adaptive
\[ \simeq \frac{-\log \epsilon}{\max D(\lambda)} \]

Sequential Adaptive
\[ \simeq \frac{-\log \epsilon}{D^*} \]

\[ D^* \geq \max_{\lambda} \bar{D}(\lambda) > \max \min_i D(i, \lambda) \]

\[ D(i, \lambda) = \min_{j \neq i} \sum_{a \in A} \lambda_a D(q^a_i \| q^a_j) \quad \text{and} \quad D^* = \frac{M}{\sum_{i=1}^{M} \frac{1}{D(i, \lambda^*_i)}} \]

How fast $\lambda$ distinguishes $H_i$ from $H_j$
\[ \lambda^*_i = \arg \max_{\lambda} D(i, \lambda) \]
Minimum expect cost: $\mathbb{E}\{\tau\}$

<table>
<thead>
<tr>
<th>Type</th>
<th>Cost Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-sequential Non-adaptive</td>
<td>$\geq \frac{-2 \log \epsilon}{\max \min_{i} D(i, \lambda)}$</td>
</tr>
<tr>
<td>Sequential Non-adaptive</td>
<td>$\simeq \frac{-\log \epsilon}{\max \lambda D(\lambda)}$</td>
</tr>
<tr>
<td>Sequential Adaptive</td>
<td>$\simeq \frac{-\log \epsilon}{\bar{D}^*}$</td>
</tr>
</tbody>
</table>

\[
\bar{D}^* \geq \max_{\lambda} \bar{D}(\lambda) > \max_{\lambda} \min_{i} D(i, \lambda)
\]

\[
D(i, \lambda) = \min_{j \neq i} \sum_{a \in A} \lambda_a D(q_i^a || q_j^a) \quad \text{and} \quad \bar{D}^* = \frac{M}{\sum_{i=1}^{M} \frac{1}{D(i, \lambda^*_i)}}
\]

How fast $\lambda$ distinguishes $H_i$ from $H_j$

$\lambda^*_i = \arg \max_{\lambda} D(i, \lambda)$
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among \textit{a few} hypotheses with \textit{high} accuracy
- discriminate among \textit{many} hypotheses with \textit{low} accuracy
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among \textit{a few} hypotheses with \textit{high accuracy}
- discriminate among \textit{many} hypotheses with \textit{low accuracy}

A good characterization/strategy must account for $M$ as well as $l$
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among a few hypotheses with high accuracy
- discriminate among many hypotheses with low accuracy

A good characterization/strategy must account for $M$ as well as $l$

Given testing strategy $\pi$:
- $M^\pi(T, \epsilon) :=$ the maximum number of hypothesis that can be identified with $E^\pi[\tau] \leq T$ and $Pe^\pi \leq \epsilon$
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among *a few* hypotheses with *high accuracy*
- discriminate among *many* hypotheses with *low accuracy*

A good characterization/strategy must account for $M$ as well as $l$

Given testing strategy $\pi$:
- $M^\pi(T, \epsilon) :=$ the maximum number of hypothesis that can be identified with $\mathbb{E}^\pi[\tau] \leq T$ and $\text{Pe}^\pi \leq \epsilon$
- Achieves *information acquisition rate* $R > 0$ with *reliability* $E > 0$ if
  $$M^\pi(T, 2^{-ET}) \sim 2^{TR}$$
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among a few hypotheses with high accuracy
- discriminate among many hypotheses with low accuracy

A good characterization/strategy must account for $M$ as well as $l$

Given testing strategy $\pi$:
- $M^\pi(T, \epsilon) :=$ the maximum number of hypothesis that can be identified with $E^\pi[\tau] \leq T$ and $P^\pi \leq \epsilon$
- Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if
  $$M^\pi(T, 2^{-ET}) \sim 2^{TR}$$

Chernoff’s scheme (and extensions) achieve $\bar{D}^*$ when $R = 0$
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among a few hypotheses with high accuracy
- discriminate among many hypotheses with low accuracy

A good characterization/strategy must account for $M$ as well as $l$

Given testing strategy $\pi$:
- $M^\pi(T, \epsilon) :=$ the maximum number of hypothesis that can be identified with $E^\pi[\tau] \leq T$ and $P_e^\pi \leq \epsilon$
- Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if

$$M^\pi(T, 2^{-ET}) \sim 2^{TR}$$

Chernoff’s scheme (and extensions) achieve $\bar{D}^*$ when $R = 0$

Is it possible to achieve $R, E > 0$? $R_{\text{max}}$? $E_{\text{max}}(R)$?
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among *a few* hypotheses with *high accuracy*
- discriminate among *many* hypotheses with *low accuracy*

A good characterization/strategy must account for $M$ as well as $l$

Given testing strategy $\pi$:

$$M_\pi(T, \epsilon) := \text{the maximum number of hypotheses that can be identified with } E_\pi[\tau] \leq T \text{ and } Pe_\pi \leq \epsilon$$

Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if $M_\pi(T, 2^{-ET}) \approx 2TR$!

Chernoff’s scheme (and extensions) achieve $\bar{D}^*$ when $R = 0$?

Is it possible to achieve $R, E > 0$? $R_{\max}$? $E_{\max}(R)$?
Chernoff’s asymptotic optimality neglects the tension between
- discriminate among a few hypotheses with high accuracy
- discriminate among many hypotheses with low accuracy

A good characterization/strategy must account for

Given testing strategy $\pi$, 

\[ M_\pi(T, \epsilon) := \text{the maximum number of hypotheses that can be identified with } E_\pi[\tau] \leq T \text{ and } P_{\pi} \leq \epsilon \]

Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if $M_\pi(T, 2^{-ET}) \approx 2T^R$!

Chernoff’s scheme (and extensions) achieve $D^*$ when $K = 0$?

Is it possible to achieve $R, E > 0$? $R_{\max}$? $E_{\max}(R)$?
Chernoff's asymptotic optimality neglects the tension between
- discriminate among a few hypotheses with high accuracy
- discriminate among many hypotheses with low accuracy

A characterization/strategy must account for $M$ as well as $l$.

Given testing strategy $\pi$:

$$M_{\pi}(T, \epsilon) := \text{the maximum number of hypotheses that can be identified with } \epsilon \leq T \text{ and } \text{Pe}_{\pi} \leq \epsilon$$

Achieves information acquisition rate $R > 0$ with reliability $E > 0$ if

$$M_{\pi}(T, 2^{-ET}) \simeq 2TR$$

Chernoff's scheme (and extensions) achieve $\bar{D}^*$ when $R = 0$.

Is it possible to achieve $R, E > 0$? $R_{\text{max}}$? $E_{\text{max}}(R)$?
Current and Future Work

- Joint Source–Channel Coding with and without Cost Constraints
  - Quasi-static sources
  - Slowly mixing Markov sources
  - Estimation and quantization

- Active Hidden Markov Models
  - Tracking and estimation
  - Acquisition and utilization of information


References


References