Sequential Hypothesis Testing and Variable Length Coding

Graduate Seminar

Shai Ginzach

Supervised by
Igal Sason and Neri Merhav

Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa, Israel

December 25th, 2014

Outline

1. Sequential Hypothesis Testing
   - Sequential Binary Hypothesis Testing
   - Multi-hypothesis Testing
   - Multi-hypothesis Testing with Control

Outline

Outline

Sequential Binary Hypothesis Testing - Setting

- *Observation sequence*:

$$Y_1, Y_2, Y_3, \ldots \overset{\text{i.i.d.}}{\sim} P.$$

Sequential Binary Hypothesis Testing - Setting

- *Observation sequence*:

$$Y_1, Y_2, Y_3, \ldots \overset{\text{i.i.d.}}{\sim} P.$$

- Two *hypotheses*:

$$\begin{cases} H_0 & : P = P_0, \\ H_1 & : P = P_1. \end{cases}$$

Where P_0 and P_1 are completely known distinct probability measures.

Sequential Binary Hypothesis Testing - Setting

- *Observation sequence*:

$$Y_1, Y_2, Y_3, \ldots \overset{\text{i.i.d.}}{\sim} P.$$

- Two *hypotheses*:

$$\begin{cases} H_0 & : P = P_0, \\ H_1 & : P = P_1. \end{cases}$$

Where P_0 and P_1 are completely known distinct probability measures.

- *Priors*: $\mathbb{P}\{H_0\} = \pi_0$, and $\mathbb{P}\{H_1\} = \pi_1 = 1 - \pi_0$.

Sequential Binary Hypothesis Testing - Basic Components

Definition - Sequential Binary Hypothesis Test

A Sequential binary hypothesis test is a pair $\Delta = (N, d)$ where:

- N is the *stopping time* (such that $\{N = n\}, \{N > n\} \in \sigma(Y_1^n)$).
- $d : Y_1^N \to \{H_0, H_1\}$ is the *terminal decision rule*.

Sequential Binary Hypothesis Testing - Basic Components

Definition - Sequential Binary Hypothesis Test

A Sequential binary hypothesis test is a pair $\Delta = (N, d)$ where:

- N is the *stopping time* (such that $\{N = n\}, \{N > n\} \in \sigma(Y_1^n)$).
- $d : Y_1^N \to \{H_0, H_1\}$ is the *terminal decision rule*.

- Two types of errors:
  1. *Type* 1 *error:* Reject the null hypothesis when correct
  $$\alpha \triangleq P_0(d = H_1).$$
  2. *Type* 2 *error:* Accept the null hypothesis when incorrect
  $$\beta \triangleq P_1(d = H_0).$$

Sequential Probability Ratio Test (SPRT)

Sequential Probability Ratio Test (SPRT)

- Define the *Log Likelihood Ratio function (LLR):*

$$L_n\left(y_1^n\right) \triangleq \log\left[\frac{P_0\left(y_1^n\right)}{P_1\left(y_1^n\right)}\right] = \sum_{i=1}^{n} \log\left[\frac{P_0\left(y_i\right)}{P_1\left(y_i\right)}\right]$$

Sequential Probability Ratio Test (SPRT)

- Define the *Log Likelihood Ratio function (LLR)*:

$$L_n\left(y_1^n\right) \triangleq \log\left[\frac{P_0\left(y_1^n\right)}{P_1\left(y_1^n\right)}\right] = \sum_{i=1}^{n} \log\left[\frac{P_0\left(y_i\right)}{P_1\left(y_i\right)}\right]$$

- Select two *boundary values* A and B such that $0 < B < 1 < A$

Sequential Probability Ratio Test (SPRT)

- Define the *Log Likelihood Ratio function (LLR)*:

$$L_n\left(y_1^n\right) \triangleq \log\left[\frac{P_0\left(y_1^n\right)}{P_1\left(y_1^n\right)}\right] = \sum_{i=1}^{n} \log\left[\frac{P_0\left(y_i\right)}{P_1\left(y_i\right)}\right]$$

- Select two *boundary values* A and B such that $0 < B < 1 < A$

Definition - SPRT (Wald 1943)

The Sequential Probability Ratio Test (SPRT) $\Delta_{\mathsf{SPRT}} \triangleq \left(N_{\mathsf{SPRT}}, d_{\mathsf{SPRT}}\right)$ is defined as follows:

$$N_{\mathsf{SPRT}} \triangleq \min\left\{n \in \mathbb{N}\colon L_n\left(Y_1^n\right) \leq \log\left(B\right) \quad \text{or} \quad L_n\left(Y_1^n\right) \geq \log\left(A\right)\right\}$$

$$d_{\mathsf{SPRT}} \triangleq \begin{cases} H_1 & \text{if} \quad L_{N_{\mathsf{SPRT}}}\left(Y_1^{N_{\mathsf{SPRT}}}\right) \leq \log\left(B\right) \\ H_0 & \text{if} \quad L_{N_{\mathsf{SPRT}}}\left(Y_1^{N_{\mathsf{SPRT}}}\right) \geq \log\left(A\right) \end{cases}$$

Basic Properties of the SPRT

Basic Properties of the SPRT

- Define $\bar{N} \triangleq \min\{n \in \mathbb{N} \colon L_n \geq \log(A)\}$. Then,

$$\mathbb{E}_{P_0}[N_{\mathsf{SPRT}}] \leq \mathbb{E}_{P_0}\left[\bar{N}\right] \leq \frac{\log(A)}{D(P_0 \parallel P_1)} \leq \frac{-\log(\alpha)}{D(P_0 \parallel P_1)}.$$

Basic Properties of the SPRT

- Define $\bar{N} \triangleq \min\{n \in \mathbb{N} \colon L_n \geq \log(A)\}$. Then,

$$\mathbb{E}_{P_0}[N_{\mathsf{SPRT}}] \leq \mathbb{E}_{P_0}\left[\bar{N}\right] \leq \frac{\log(A)}{D(P_0 \parallel P_1)} \leq \frac{-\log(\alpha)}{D(P_0 \parallel P_1)}.$$

Similarly, $\mathbb{E}_{P_1}[N_{\mathsf{SPRT}}] \leq \frac{-\log(B)}{D(P_1 \| P_0)} \leq \frac{-\log(\beta)}{D(P_1 \| P_0)}$

Basic Properties of the SPRT

- Define $\bar{N} \triangleq \min\left\{n \in \mathbb{N} \colon L_n \geq \log\left(A\right)\right\}$. Then,

$$\mathbb{E}_{P_0}\left[N_{\mathsf{SPRT}}\right] \leq \mathbb{E}_{P_0}\left[\bar{N}\right] \leq \frac{\log\left(A\right)}{D\left(P_0 \parallel P_1\right)} \leq \frac{-\log\left(\alpha\right)}{D\left(P_0 \parallel P_1\right)}.$$

Similarly, $\mathbb{E}_{P_1}\left[N_{\mathsf{SPRT}}\right] \leq \frac{-\log(B)}{D(P_1 \| P_0)} \leq \frac{-\log(\beta)}{D(P_1 \| P_0)}$

Theorem - Optimality of the SPRT (Wald & Wolfowitz 1953)

Let $\Delta_{\mathsf{SPRT}} = \left(N_{\mathsf{SPRT}}, d_{\mathsf{SPRT}}\right)$ be Wald's SPRT with error probabilities α_{SPRT} and β_{SPRT}, and let $\Delta' = \left(N', d'\right)$ be any other sequential decision rule with finite $\mathbb{E}_{P_1}\left[N'\right], \mathbb{E}_{P_0}\left[N'\right]$ and error probabilities α' and β' satisfying

$$\alpha' < \alpha_{\mathsf{SPRT}} \quad \text{and} \quad \beta' < \beta_{\mathsf{SPRT}}.$$

Then

$$\mathbb{E}_{P_1}\left[N'\right] \geq \mathbb{E}_{P_1}\left[N_{\mathsf{SPRT}}\right], \quad \mathbb{E}_{P_0}\left[N'\right] \geq \mathbb{E}_{P_0}\left[N_{\mathsf{SPRT}}\right].$$

Sequential Multi-hypothesis Testing - Setting

- *Observation sequence*:

$$\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots \overset{\text{i.i.d.}}{\sim} P.$$

where $\mathbf{Y}_i$ is an l-valued random vector $(\mathbf{Y_i} = (Y_{i,1} \ldots, Y_{i,l}))$.

- Define $M\ (>2)$ *hypotheses*:

$$H_i \colon P = P_i, \quad i \in \{0, \ldots, M-1\}\,.$$

where P_i are completely known distinct probability measures.

- *Priors*: $\pi = \{\pi_0 \ldots, \pi_{M-1}\}$ where $\pi_i = \mathbb{P}\{H_i\}$

Sequential Multi-hypothesis Testing - Setting

- *Observation sequence*:

$$\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \ldots \overset{\text{i.i.d.}}{\sim} P.$$

where $\mathbf{Y}_i$ is an l-valued random vector ($\mathbf{Y_i} = (Y_{i,1} \ldots, Y_{i,l})$).

- Define M (> 2) *hypotheses*:

$$H_i \colon P = P_i, \quad i \in \{0, \ldots, M-1\}.$$

where P_i are completely known distinct probability measures.

- *Priors*: $\pi = \{\pi_0 \ldots, \pi_{M-1}\}$ where $\pi_i = \mathbb{P}\{H_i\}$
- A *Multi-hypothesis test* Δ is a pair (N, d) where N is the *stopping time* and d is the *decision rule*.

Sequential Multi-hypothesis Testing - Risk and Error Probabilities

- Let $\alpha_{ji}(\Delta) = P_j(d = i)$ be the probability of accepting the hypothesis H_i when H_j is true (defined for $j \neq i$).

Sequential Multi-hypothesis Testing - Risk and Error Probabilities

- Let $\alpha_{ji}(\Delta) = P_j(d = i)$ be the probability of accepting the hypothesis H_i when H_j is true (defined for $j \neq i$).
- For a multiple hypothesis test Δ, the risk associated with making the decision $d = i$ is defined to be

$$R_i(\Delta) = \sum_{j \neq i} \pi_j \alpha_{ji}(\Delta).$$

Sequential Multi-hypothesis Testing - Risk and Error Probabilities

- Let $\alpha_{ji}(\Delta) = P_j(d = i)$ be the probability of accepting the hypothesis H_i when H_j is true (defined for $j \neq i$).
- For a multiple hypothesis test Δ, the risk associated with making the decision $d = i$ is defined to be

$$R_i(\Delta) = \sum_{j \neq i} \pi_j \alpha_{ji}(\Delta).$$

- Let $\overline{\mathbf{R}} \triangleq \left(\overline{R}_0, \overline{R}_1, \ldots, \overline{R}_{M-1}\right)$ be a vector of positive finite numbers and define:

$$\boldsymbol{\Delta}\left(\overline{\mathbf{R}}\right) \triangleq \left\{\Delta \colon R_i(\Delta) \leq \overline{R}_i, i \in \{0, \ldots, M-1\}\right\}.$$

Sequential Multi-hypothesis Testing - Risk and Error Probabilities

- Let $\alpha_{ji}(\Delta) = P_j(d = i)$ be the probability of accepting the hypothesis H_i when H_j is true (defined for $j \neq i$).
- For a multiple hypothesis test Δ, the risk associated with making the decision $d = i$ is defined to be

$$R_i(\Delta) = \sum_{j \neq i} \pi_j \alpha_{ji}(\Delta).$$

- Let $\overline{\mathbf{R}} \triangleq \left(\overline{R}_0, \overline{R}_1, \ldots, \overline{R}_{M-1}\right)$ be a vector of positive finite numbers and define:

$$\mathbf{\Delta}\left(\overline{\mathbf{R}}\right) \triangleq \left\{\Delta \colon R_i(\Delta) \leq \overline{R}_i, i \in \{0, \ldots, M-1\}\right\}.$$

- Our focus: $\Delta \in \mathbf{\Delta}\left(\overline{\mathbf{R}}\right)$ as $R_{\max} \triangleq \max_i \bar{R}_i \to 0$ and M fixed.

Multi-hypothesis SPRT (MSPRT)

- Define the *LLR* of P_i w.r.t. to a (dominating) measure Q by:

$$L_i(n) = \log\left[\frac{P_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)}{Q(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)}\right], \quad i \in \{0, \ldots, M-1\}.$$

Multi-hypothesis SPRT (MSPRT)

- Define the *LLR* of P_i w.r.t. to a (dominating) measure Q by:

$$L_i(n) = \log\left[\frac{P_i(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)}{Q(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)}\right], \quad i \in \{0,\ldots,M-1\}.$$

- Let $a_i,\ (i \in \{0,\ldots,M-1\})$ be positive threshold values.

Multi-hypothesis SPRT (MSPRT)

- Define the *LLR* of P_i w.r.t. to a (dominating) measure Q by:

$$L_i(n) = \log\left[\frac{P_i(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)}{Q(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)}\right], \quad i \in \{0, \ldots, M-1\}.$$

- Let a_i, $(i \in \{0, \ldots, M-1\})$ be positive threshold values.
- Define the stopping times

$$N_i = \min_{n \geq 0}\left\{L_i(n) \geq a_i + \log\left(\sum_{j \neq i} \exp(L_j(n))\right)\right\},$$

Multi-hypothesis SPRT (MSPRT)

- Define the *LLR* of P_i w.r.t. to a (dominating) measure Q by:

$$L_i(n) = \log\left[\frac{P_i(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)}{Q(\mathbf{Y}_1,\ldots,\mathbf{Y}_n)}\right], \quad i \in \{0,\ldots,M-1\}.$$

- Let $a_i, \ (i \in \{0,\ldots,M-1\})$ be positive threshold values.
- Define the stopping times

$$N_i = \min_{n\geq 0}\left\{L_i(n) \geq a_i + \log\left(\sum_{j\neq i}\exp\left(L_j(n)\right)\right)\right\},$$

Definition - Δ_a (Baum & Veeravalli 1994, Fishman 1987)

Let $\Delta_a = (N_a, d_a)$ be a sequential test defined by:

$$N_a = \min_{0\leq i\leq M-1} N_i, \quad d_a = i^\star \text{ if } N_a = N_{i^\star}.$$

Optimality of Δ_a

- Define $D_i \triangleq \min_{j \neq i} D\left(P_i \parallel P_j\right)$.

Optimality of Δ_a

- Define $D_i \triangleq \min_{j \neq i} D\left(P_i \parallel P_j\right)$.

Theorem - Optimality of Δ_a (Dragalin *et al. 2000*)

1. For all $i \in \{0, 1, \ldots M-1\}$

$$\inf_{\Delta \in \boldsymbol{\Delta}\left(\overline{\mathbf{R}}\right)} \mathbb{E}_i\left[N\right] \geq \left[\frac{-\log\left(\overline{R}_i\right)}{D_i}\right]\left(1 + o\left(1\right)\right)$$

2. If $a_i = \log\left[\frac{\pi_i}{\overline{R}_i}\right]$ then

$$\mathbb{E}_i\left[N_a\right] \sim \frac{-\log\left(\overline{R}_i\right)}{D_i}$$

as $\overline{R}_{\max} \to 0$ for all $m \geq 1$.

Sequential Multi-hypothesis Testing w/ Control - Setting

- *Observation sequence*: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.
- *Hypotheses*: $\{H_i, i = 0, \ldots, M-1\}$.
- *Priors*: $\mathbb{P}\{H_i\} = \pi_i, \quad i \in \{0, \ldots, M-1\}$.

Sequential Multi-hypothesis Testing w/ Control - Setting

- *Observation sequence*: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.
- *Hypotheses*: $\{H_i, i = 0, \ldots, M-1\}$.
- *Priors*: $\mathbb{P}\{H_i\} = \pi_i, \quad i \in \{0, \ldots, M-1\}$.
- *Control sequence*: $U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty$.

Sequential Multi-hypothesis Testing w/ Control - Setting

- *Observation sequence*: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.
- *Hypotheses*: $\{H_i, i = 0, \ldots, M-1\}$.
- *Priors*: $\mathbb{P}\{H_i\} = \pi_i, \quad i \in \{0, \ldots, M-1\}$.
- *Control sequence*: $U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty$.
    - $U_n = q(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1}) \quad \text{[Causality Constraint]}$

Sequential Multi-hypothesis Testing w/ Control - Setting

- *Observation sequence*: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.
- *Hypotheses*: $\{H_i, i = 0, \ldots, M-1\}$.
- *Priors*: $\mathbb{P}\{H_i\} = \pi_i, \quad i \in \{0, \ldots, M-1\}$.
- *Control sequence*: $U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty$.
  - $U_n = q\left(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1}\right) \quad \text{[Causality Constraint]}$
  - Assume: $Y_n \perp \left(Y^{n-1}, U^{n-1}\right)$

Sequential Multi-hypothesis Testing w/ Control - Setting

- *Observation sequence*: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.
- *Hypotheses*: $\{H_i, i = 0, \ldots, M-1\}$.
- *Priors*: $\mathbb{P}\{H_i\} = \pi_i, \quad i \in \{0, \ldots, M-1\}$.
- *Control sequence*: $U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty$.
  - $U_n = q(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1}) \quad \text{[Causality Constraint]}$
  - Assume: $Y_n \perp \left(Y^{n-1}, U^{n-1}\right)$
- *Observation kernel*:

$$p_i^{u_n}(y_n) \triangleq \mathbb{P}(Y_n = y_n \mid H_i, U_n = u_n).$$

Hypothesis Testing with Control - Definitions

- A *Multi-hypothesis test with control* Δ is a triplet (q, N, d) where:

Hypothesis Testing with Control - Definitions

- A *Multi-hypothesis test with control* Δ is a triplet (q, N, d) where:
  1. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.

Hypothesis Testing with Control - Definitions

- A *Multi-hypothesis test with control* Δ is a triplet (q, N, d) where:
  1. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  2. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.

Hypothesis Testing with Control - Definitions

- A *Multi-hypothesis test with control* Δ is a triplet (q, N, d) where:
  1. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  2. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.
- **The Objective:** Find Δ that:

$$\text{minimize } \mathbb{E}\left[N\right] \text{ subject to } P_{\text{er}} \leq \epsilon \quad [\text{Max. Information}]$$

Hypothesis Testing with Control - Definitions

- A *Multi-hypothesis test with control* Δ is a triplet (q, N, d) where:
  1. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  2. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.
- **The Objective:** Find Δ that:

$$\text{minimize } \mathbb{E}\left[N\right] \text{ subject to } P_{\text{er}} \leq \epsilon \quad [\text{Max. Information}]$$

- Let $\mathbb{E}\left[N^\star\right]$ be the minimal expected number of samples required to achieve $P_{\text{er}} \leq \epsilon$.
- **Achievablility:** Chernoff (1960), Veeravalli (2012), Javidi (2013)...

Lower Bounds on $\mathbb{E}[N^\star]$

Theorem (Javidi *et al. 2013*)

For $\frac{\log(M)}{I_{\max}} < w$ and arbitrary $\delta \in (0, 0.5]$:

$$\mathbb{E}[N^\star] \geq (1-\epsilon w)\left[\frac{H(\theta) - [h_2(\delta) + \delta \log(M-1)]}{I_{\max}} + \frac{\log\left(\frac{\delta}{1-\delta}\right) - \log\left(\frac{w^{-1}}{1-w^{-1}}\right)}{D_{\max}} \mathbb{I}\left\{\max_i \pi_i \leq 1-\delta\right\} - \hat{K}'\right]^+$$

where $D_{\max} = \max_{i,ju} D\left(p_i^u \parallel p_j^u\right)$, and $I_{\max} = \max_{u,\tilde{\pi}} I\left(\tilde{\pi}; p_{\tilde{\pi}}^u\right)$.

Lower Bounds on $\mathbb{E}[N^\star]$

Theorem (Javidi *et al.* 2013)

For $\frac{\log(M)}{I_{\max}} < w$ and arbitrary $\delta \in (0, 0.5]$:

$$\mathbb{E}[N^\star] \geq (1 - \epsilon w)\left[\frac{H(\theta) - [h_2(\delta) + \delta \log(M-1)]}{I_{\max}} + \frac{\log\left(\frac{\delta}{1-\delta}\right) - \log\left(\frac{w^{-1}}{1-w^{-1}}\right)}{D_{\max}} \mathbb{I}\left\{\max_i \pi_i \leq 1 - \delta\right\} - \hat{K}'\right]^+$$

where $D_{\max} = \max_{i,ju} D\left(p_i^u \parallel p_j^u\right)$, and $I_{\max} = \max_{u,\tilde{\pi}} I\left(\tilde{\pi}; p_{\tilde{\pi}}^u\right)$.

- Remainder [Fano's Inequality]: Let δ be the error probability of the estimator $\hat{\theta}$ of θ. Then

$$H\left(\theta \mid \hat{\theta}\right) \leq h_2(\delta) + \delta \log(M-1)$$

Lower Bounds on $\mathbb{E}\left[N^\star\right]$

Theorem (Javidi *et al.* 2013)

For $\frac{\log(M)}{I_{\max}} < w$ and arbitrary $\delta \in (0, 0.5]$:

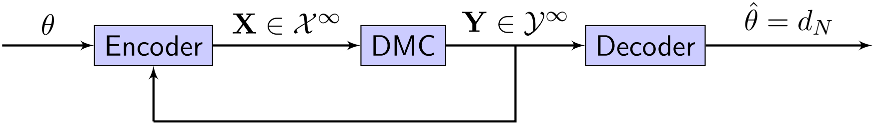
$$\mathbb{E}\left[N^\star\right] \geq (1-\epsilon w)\left[\frac{H(\theta) - \left[h_2(\delta) + \delta \log(M-1)\right]}{I_{\max}} + \frac{\log\left(\frac{\delta}{1-\delta}\right) - \log\left(\frac{w^{-1}}{1-w^{-1}}\right)}{D_{\max}} \mathbb{I}\left\{\max_i \pi_i \leq 1-\delta\right\} - \hat{K}'\right]^+$$

where $D_{\max} = \max_{i,j u} D\left(p_i^u \parallel p_j^u\right)$, and $I_{\max} = \max_{u,\tilde{\pi}} I\left(\tilde{\pi}; p_{\tilde{\pi}}^u\right)$.

- Remainder [Fano's Inequality]: Let δ be the error probability of the estimator $\hat{\theta}$ of θ. Then

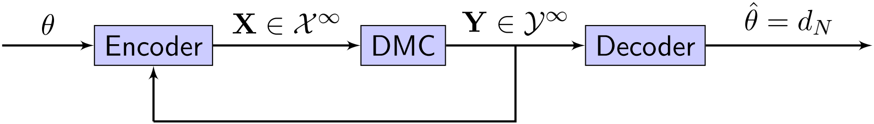
$$H\left(\theta \mid \hat{\theta}\right) \leq h_2(\delta) + \delta \log(M-1)$$

VL Coding with Perfect Feedback



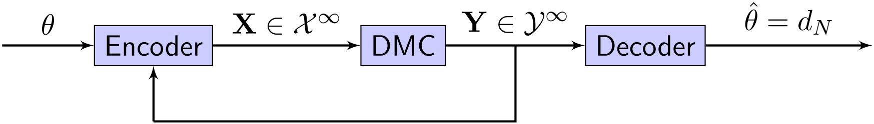
- *Message*: One of M equiprobable symbols $\theta \in \{0 \ldots, M-1\}$.
- *Forward Channel*: $\mathcal{X} = \{1, \ldots, K\}$ and $\mathcal{Y} = \{1, \ldots, L\}$.
- *Feedback channel*: Instantaneous, infinite capacity, noiseless.

VL Coding with Perfect Feedback



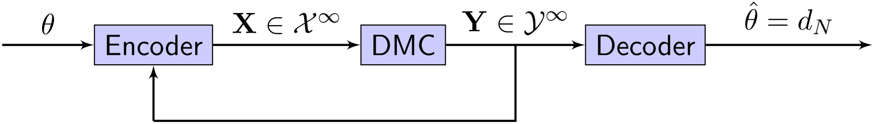
- *Message*: One of M equiprobable symbols $\theta \in \{0 \ldots, M-1\}$.
- *Forward Channel*: $\mathcal{X} = \{1, \ldots, K\}$ and $\mathcal{Y} = \{1, \ldots, L\}$.
- *Feedback channel*: Instantaneous, infinite capacity, noiseless.
- *Coding Algorithm*: $X_n(\theta) \triangleq X_n(\theta, Y_1, \ldots Y_{n-1}), \ \forall \theta, Y_1, \ldots Y_{n-1}$

VL Coding with Perfect Feedback



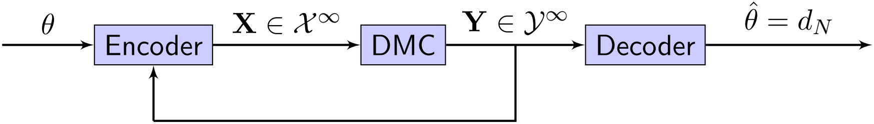
- *Message*: One of M equiprobable symbols $\theta \in \{0 \ldots, M-1\}$.
- *Forward Channel*: $\mathcal{X} = \{1, \ldots, K\}$ and $\mathcal{Y} = \{1, \ldots, L\}$.
- *Feedback channel*: Instantaneous, infinite capacity, noiseless.
- *Coding Algorithm*: $X_n(\theta) \triangleq X_n(\theta, Y_1, \ldots Y_{n-1}), \ \forall \theta, Y_1, \ldots Y_{n-1}$
- *Decoding Criterion*: A pair (N, d), where N is the stopping time and d is the decision function.

VL Coding - Performance Indices



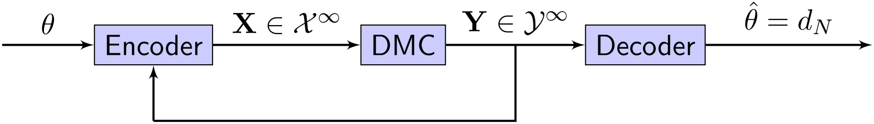
- For block codes with fixed block-length n:
  1. *Rate*: $R \triangleq \frac{\log(M)}{n}$.
  2. *Error Exponent*: $E(R) = \limsup_{n\to\infty} \frac{-\log(P_{\text{er}})}{n}$.

VL Coding - Performance Indices



- For block codes with fixed block-length n:
  1. *Rate*: $R \triangleq \frac{\log(M)}{n}$.
  2. *Error Exponent*: $E(R) = \limsup_{n\to\infty} \frac{-\log(P_{\text{er}})}{n}$.
- For VL codes:
  1. *Rate*: $R \triangleq \frac{\log(M)}{\mathbb{E}[N]}$.
  2. *Error Exponent*: $E(R) = \limsup_{\mathbb{E}[N]\to\infty} \frac{-\log(P_{\text{er}})}{\mathbb{E}[N]}$.

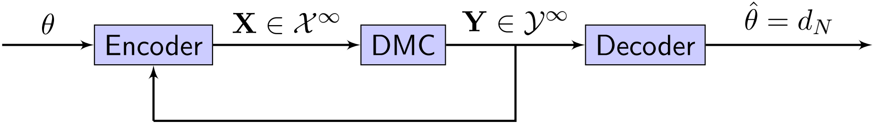
VL Coding - Performance Indices



- For block codes with fixed block-length n:
  1. *Rate*: $R \triangleq \frac{\log(M)}{n}$.
  2. *Error Exponent*: $E(R) = \limsup_{n \to \infty} \frac{-\log(P_{\text{er}})}{n}$.
- For VL codes:
  1. *Rate*: $R \triangleq \frac{\log(M)}{\mathbb{E}[N]}$.
  2. *Error Exponent*: $E(R) = \limsup_{\mathbb{E}[N] \to \infty} \frac{-\log(P_{\text{er}})}{\mathbb{E}[N]}$.

Q: Is the VL coding problem amenable to hypothesis testing analysis?

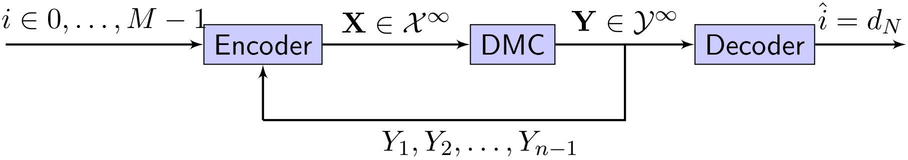
VL Coding - Performance Indices



- For block codes with fixed block-length n:
  1. *Rate*: $R \triangleq \frac{\log(M)}{n}$.
  2. *Error Exponent*: $E(R) = \limsup_{n\to\infty} \frac{-\log(P_{\text{er}})}{n}$.
- For VL codes:
  1. *Rate*: $R \triangleq \frac{\log(M)}{\mathbb{E}[N]}$.
  2. *Error Exponent*: $E(R) = \limsup_{\mathbb{E}[N]\to\infty} \frac{-\log(P_{\text{er}})}{\mathbb{E}[N]}$.

Q: Is the VL coding problem amenable to hypothesis testing analysis?
A: Yes!

VL Coding and Controlled Hypothesis Testing, Cont.



- $\pi = \left[\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M}\right]$.
- $p_i^u(y_n) = p(y_n \mid x_n(i,u))$.
- $D_{\max} = \max_{j,k} D\left(p(\cdot \mid j) \,\|\, p(\cdot \mid k)\right) \triangleq C_1$.
- $I_{\max} = \max_{P_X} I(X;Y) = C$.

VL Coding and Controlled Hypothesis Testing, Cont.

- Recap:

Theorem (Javidi *et al.* 2013)

For $\frac{\log(M)}{I_{\max}} < w < 1/\epsilon$ and arbitrary $\delta \in (0, 0.5]$:

$$\mathbb{E}\left[N^{\star}\right] \geq (1-\epsilon w)\left[\frac{H(\pi) - h_2(\delta) - \delta \log(M-1)}{I_{\max}} + \frac{\log\left(\frac{1-w^{-1}}{w^{-1}}\right) - \log\left(\frac{1-\delta}{\delta}\right)}{D_{\max}} \mathbb{I}\left\{\max_i \pi_i \leq 1-\delta\right\} - \hat{K}'\right]^+$$

where $D_{\max} = \max_{i,ju} D\left(p_i^u \parallel p_j^u\right)$, and $I_{\max} = \max_{u,\tilde{\pi}} I\left(\tilde{\pi}; p_{\tilde{\pi}}^u\right)$.

VL Coding and Controlled Hypothesis Testing, Cont.

- Recap:

Theorem (Javidi *et al.* 2013)

For $w = \frac{1}{\epsilon \log\left(\frac{M}{\epsilon}\right)}$ and $\delta = \frac{1}{\log\left(\frac{M}{\epsilon}\right)}$:

$$\mathbb{E}\left[N^{\star}\right] \gtrsim \left(1 - \frac{1}{\log\left(\frac{M}{\epsilon}\right)}\right)\left[\frac{\log\left(M\right) - h_2\left(\frac{1}{\log\left(\frac{M}{\epsilon}\right)}\right) - \frac{\log(M-1)}{\log\left(\frac{M}{\epsilon}\right)}}{C} - \hat{K}' + \frac{-\log\left(\epsilon \log\left(\frac{M}{\epsilon}\right)\right) - \log\left(\log\left(\frac{M}{\epsilon}\right)\right)}{C_1}\mathbb{I}\left\{\frac{1}{M} \leq 1 - \frac{1}{\log\left(\frac{M}{\epsilon}\right)}\right\}\right]^{+}$$

VL Coding and Controlled Hypothesis Testing, Cont.

Theorem (Javidi *et al.* 2013)

For large M and small ϵ,

$$\mathbb{E}\left[N^{\star}\right] \gtrsim \frac{\log\left(M\right)}{C} + \frac{-\log\left(P_{\text{er}}\right)}{C_1} + O\left(\log\left(\log\left(\frac{M}{\epsilon}\right)\right)\right).$$

VL Coding and Controlled Hypothesis Testing, Cont.

Theorem (Javidi *et al.* 2013)

For large M and small ϵ,

$$\mathbb{E}\left[N^{\star}\right] \gtrsim \frac{\log\left(M\right)}{C} + \frac{-\log\left(P_{\text{er}}\right)}{C_1} + O\left(\log\left(\log\left(\frac{M}{\epsilon}\right)\right)\right).$$

- Equivalently,

$$E\left(R\right) \lesssim C_1\left(1 - \frac{R}{C}\right) \triangleq E_{\text{B}}\left(R\right).$$

VL Coding and Controlled Hypothesis Testing, Cont.

Theorem (Javidi *et al.* 2013)

For large M and small ϵ,

$$\mathbb{E}\left[N^{\star}\right] \gtrsim \frac{\log\left(M\right)}{C} + \frac{-\log\left(P_{\text{er}}\right)}{C_1} + O\left(\log\left(\log\left(\frac{M}{\epsilon}\right)\right)\right).$$

- Equivalently,

$$E\left(R\right) \lesssim C_1\left(1 - \frac{R}{C}\right) \triangleq E_{\text{B}}\left(R\right).$$

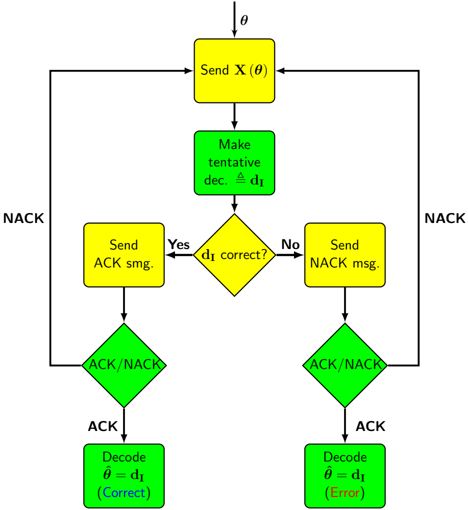
Theorem (Burnashev 1976)

For any transmission method over a DMC with perfect feedback and any $R \in [0, C]$

$$E\left(R\right) = E_{\text{B}}\left(R\right).$$

Achievablity - General Scheme

- Akin to Yamamoto & Itoh (1979).



Direct Statement - Phase I (Tentative Decision)

- Codebook: For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X- i.i.d. sequence.

Direct Statement - Phase I (Tentative Decision)

- Codebook: For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X- i.i.d. sequence.
- Let $\boldsymbol{x}^{(\boldsymbol{i})}$ be the codeword assigned to the i'th message.

Direct Statement - Phase I (Tentative Decision)

- Codebook: For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X- i.i.d. sequence.
- Let $\boldsymbol{x}^{(i)}$ be the codeword assigned to the i'th message.
- Define:

$$N_I^i = \min_{n \geq 0} \left\{ \sum_{k=1}^{n} \log \left[\frac{p\left(y_k \mid x_k^{(i)}\right)}{\Pr\left(y_k\right)} \right] \geq (1+\epsilon) \log(M) \right\}.$$

Direct Statement - Phase I (Tentative Decision)

- Codebook: For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X- i.i.d. sequence.
- Let $\boldsymbol{x}^{(i)}$ be the codeword assigned to the i'th message.
- Define:

$$N_I^i = \min_{n \geq 0} \left\{ \sum_{k=1}^{n} \log \left[\frac{p\left(y_k \mid x_k^{(i)}\right)}{\Pr\left(y_k\right)} \right] \geq (1+\epsilon) \log (M) \right\}.$$

- Decoder: $\Delta_I = (N_I, d_I)$:

$$N_I = \min_{0 \leq i \leq M-1} N_I^i, \quad d_I = i^\star \text{ if } N_I = N_I^{i^\star}.$$

- Assume $\mathbf{x}^{(0)}$ was transmitted. Then

$$\mathbb{E}\left[N_I\right] \leq \mathbb{E}\left[N_I^0\right] \lesssim \frac{(1+\epsilon)\log(M)}{C}.$$

Achievability - Phase II (Confirmation)

- Encoder: Sends $(j^\star, j^\star \ldots)$ if d_I is correct and $(k^\star, k^\star \ldots)$ otherwise.

Achievability - Phase II (Confirmation)

- Encoder: Sends $(j^\star, j^\star \ldots)$ if d_I is correct and $(k^\star, k^\star \ldots)$ otherwise.
- Decoder: Runs an SPRT with

$$H_{ACK} : Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK} : Y_i \sim p\left(\cdot \mid k^\star\right).$$

Achievability - Phase II (Confirmation)

- Encoder: Sends $(j^\star, j^\star \ldots)$ if d_I is correct and $(k^\star, k^\star \ldots)$ otherwise.
- Decoder: Runs an SPRT with

$$H_{ACK}: Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK}: Y_i \sim p\left(\cdot \mid k^\star\right).$$

- Choose $j^\star$ and $k^\star$ s.t. $D\left(p\left(\cdot \mid j^\star\right) \| p\left(\cdot \mid k^\star\right)\right) = C_1$

Achievability - Phase II (Confirmation)

- Encoder: Sends $(j^\star, j^\star \dots)$ if d_I is correct and $(k^\star, k^\star \dots)$ otherwise.
- Decoder: Runs an SPRT with

$$H_{ACK}: Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK}: Y_i \sim p\left(\cdot \mid k^\star\right).$$

- Choose $j^\star$ and $k^\star$ s.t. $D\left(p\left(\cdot \mid j^\star\right) \| p\left(\cdot \mid k^\star\right)\right) = C_1$
- For large M,

$$\mathbb{E}\left[N_{II}\right] \lesssim \frac{-\log\left(P_e\right)}{C_1}.$$

Achievability - Phase II (Confirmation)

- Encoder: Sends $(j^\star, j^\star \ldots)$ if d_I is correct and $(k^\star, k^\star \ldots)$ otherwise.
- Decoder: Runs an SPRT with

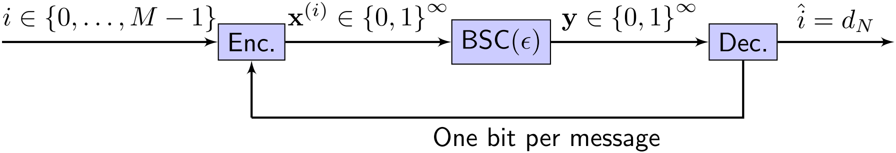
$$H_{ACK} : Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK} : Y_i \sim p\left(\cdot \mid k^\star\right).$$

- Choose $j^\star$ and $k^\star$ s.t. $D\left(p\left(\cdot \mid j^\star\right) \parallel p\left(\cdot \mid k^\star\right)\right) = C_1$
- For large M,

$$\mathbb{E}\left[N_{II}\right] \lesssim \frac{-\log\left(P_e\right)}{C_1}.$$

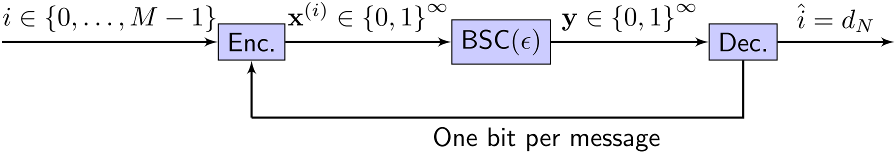
- $\Rightarrow \mathbb{E}\left[N\right] \approx \mathbb{E}\left[N_I\right] + \mathbb{E}\left[N_{II}\right] \lesssim \frac{\log(M)}{C} + \frac{-\log(P_e)}{C_1}.$

Limited Feedback - "One Shot" Schemes



Example - ARQ scheme:

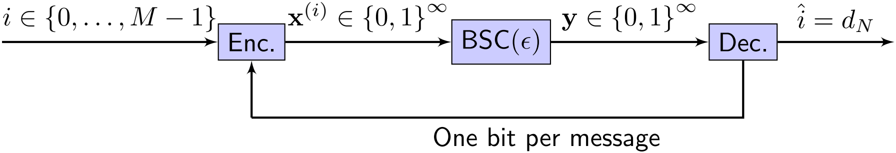
Limited Feedback - "One Shot" Schemes



Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.

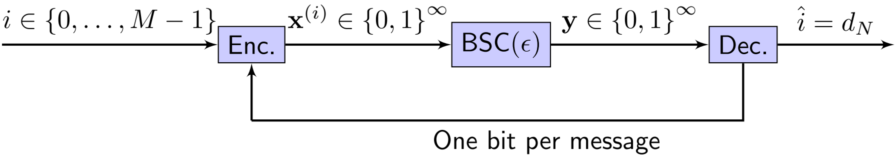
Limited Feedback - "One Shot" Schemes



Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.

Limited Feedback - "One Shot" Schemes



Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.
- Decoding:
  - Partition $\mathbb{R}^n$ into M decision regions and one erasure area.
  - If $\mathbf{Y} \in \bigcup_{i=0}^{M-1} \mathcal{R}_i$ send the stopping bit and decode.
  - Else, wait for the next n symbols and repeat the process.

Forney's Decision Regions (Forney 1968)

- Define $T > 0$ and for all $i \in \{0, \ldots, M-1\}$

$$\mathcal{R}_i^\star = \left\{ y \in \mathcal{Y}^n : \frac{p\left(y \mid x^{(i)}\right)}{\sum_{j \neq i} p\left(y \mid x^{(j)}\right)} \geq \exp\left(nT\right) \right\}, \quad i \in \{0, \ldots, M-1\},$$

$$\mathcal{R}_M^\star = \bigcap_{i=0}^{M-1} \left(\mathcal{R}_i^\star\right)^c,$$

Forney's Decision Regions (Forney 1968)

- Define $T > 0$ and for all $i \in \{0, \dots, M-1\}$

$$\mathcal{R}_i^\star = \left\{ y \in \mathcal{Y}^n : \frac{p\left(y \mid x^{(i)}\right)}{\sum_{j \neq i} p\left(y \mid x^{(j)}\right)} \geq \exp\left(nT\right) \right\}, \quad i \in \{0, \dots, M-1\},$$

$$\mathcal{R}_M^\star = \bigcap_{i=0}^{M-1} \left(\mathcal{R}_i^\star\right)^c,$$

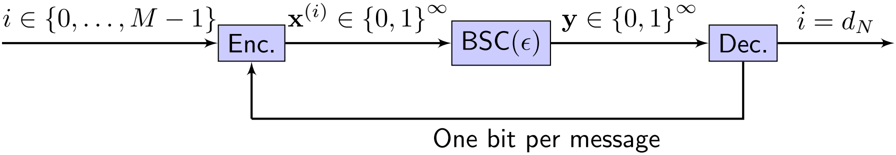
- Achievability: $E\left(R\right) \geq E_{\text{Forney}}\left(R\right)$.

$$E_{\text{Forney}}\left(R\right) \triangleq E_{\text{sp}}\left(R\right) + C - R = \beta\left(\delta_{\text{GV}}\left(R\right) - \delta_{\text{GV}}\left(C\right)\right).$$

where $\beta = \log\left(\frac{1-\epsilon}{\epsilon}\right)$ and $\delta_{\text{GV}}\left(R\right)$ is the smaller solution to

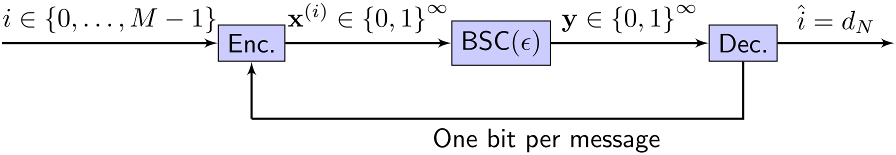
$$R + h_2\left(\delta\right) - \log\left(2\right) = 0$$

Stop-Feedback Scheme



- Main difference: decoding can stop at *any* time.
- Codebook: M i.i.d.-drawn sequences, each assigned to a message.

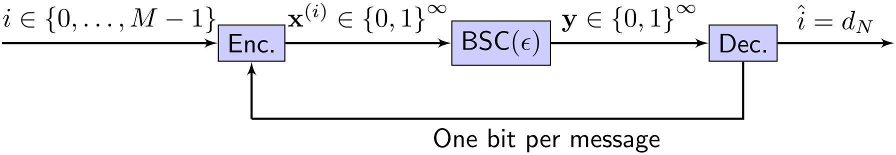
Stop-Feedback Scheme



- Main difference: decoding can stop at *any* time.
- Codebook: M i.i.d.-drawn sequences, each assigned to a message.
- This a multi-hypothesis testing problem:

$$\pi_i = \frac{1}{M}, \quad R_i(\Delta) = \sum_{j=0, j\neq i}^{M-1} \pi_j P_j(d=i) = \frac{P_{\text{er}}(\Delta)}{M}.$$

Stop-Feedback Scheme



- Main difference: decoding can stop at *any* time.
- Codebook: M i.i.d.-drawn sequences, each assigned to a message.
- This a multi-hypothesis testing problem:

$$\pi_i = \frac{1}{M}, \quad R_i(\Delta) = \sum_{j=0, j\neq i}^{M-1} \pi_j P_j(d=i) = \frac{P_{\text{er}}(\Delta)}{M}.$$

- Obstacles:
  1. M is not fixed as $\mathbb{E}[N] \to \infty$.
  2. Observations are not i.i.d. (?)

Error Exponent at R = 0

- Define

$$H_i: \quad \Pr(\mathbf{z}) = P_i(\mathbf{z}), \quad i \in \{0, \ldots, M-1\}$$

$$P_i(\mathbf{z}) = P_i\left(\mathbf{x}^{(0)}, \mathbf{x}^{(1)} \ldots, \mathbf{x}^{(M-1)}, \mathbf{y}\right) \triangleq P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y} \mid \mathbf{x}^{(i)}\right) \prod_{l=0}^{M-1} P_{\mathbf{X}}\left(\mathbf{x}^{(l)}\right)$$

Error Exponent at R = 0

- Define

$$H_i: \quad \Pr(\mathbf{z}) = P_i(\mathbf{z}), \quad i \in \{0, \ldots, M-1\}$$

$$P_i(\mathbf{z}) = P_i\left(\mathbf{x}^{(0)}, \mathbf{x}^{(1)} \ldots, \mathbf{x}^{(M-1)}, \mathbf{y}\right) \triangleq P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y} \mid \mathbf{x}^{(i)}\right) \prod_{l=0}^{M-1} P_{\mathbf{X}}\left(\mathbf{x}^{(l)}\right)$$

- Under each H_i the elements of $\mathbf{z}$ are i.i.d.

Error Exponent at R = 0

- Define

$$H_i: \quad \Pr(\mathbf{z}) = P_i(\mathbf{z}), \quad i \in \{0, \ldots, M-1\}$$

$$P_i(\mathbf{z}) = P_i\left(\mathbf{x}^{(0)}, \mathbf{x}^{(1)} \ldots, \mathbf{x}^{(M-1)}, \mathbf{y}\right) \triangleq P_{\mathbf{Y}|\mathbf{X}}\left(\mathbf{y} \mid \mathbf{x}^{(i)}\right) \prod_{l=0}^{M-1} P_{\mathbf{X}}\left(\mathbf{x}^{(l)}\right)$$

- Under each H_i the elements of $\mathbf{z}$ are i.i.d.
- Hence, it the limit

$$\inf_{(N,d)} \mathbb{E}[N] = \frac{-\log(R_i(\Delta))}{D_i} = \frac{\log M - \log(P_{\text{er}})}{D_i}$$

$$\Leftrightarrow E(0) = \lim \frac{-\log(P_{\text{er}})}{\mathbb{E}[N]} = D_i \triangleq D = E_{\text{Forney}}(0),$$

where

$$D \triangleq \sum_{x^{(0)} \in \mathcal{X}} \sum_{x^{(1)} \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_X(x^{(0)}) P_X(x^{(1)}) p(y \mid x^{(0)}) \log\left[\frac{p\left(y \mid x^{(0)}\right)}{p\left(y \mid x^{(1)}\right)}\right].$$

Lower Bound on $E(R)$

- Main idea: Decode using Δ_a

Lower Bound on $E(R)$

- Main idea: Decode using Δ_a
- Recall we've defined

$$N_i = \min_{n \geq 0} \left\{ L_i(n) \geq a + \log\left(\sum_{j \neq i} \exp\{L_j(n)\} \right) \right\}$$
$$= \min_{n \geq 0} \left\{ \log\left[\frac{P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(i)} \right)}{\sum_{j \neq i} P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(j)} \right)} \right] \geq a \right\}.$$

- $\Delta_a = (N_a, d_a)$ is then defined as follows:

$$N_a = \min_{0 \leq i \leq M-1} N_i, \quad d_a = i^\star \text{ if } N_a = N_{i^\star}.$$

Lower Bound on $E(R)$

- Main idea: Decode using Δ_a
- Recall we've defined

$$N_i = \min_{n \geq 0} \left\{ L_i(n) \geq a + \log\left(\sum_{j \neq i} \exp\{L_j(n)\} \right) \right\}$$
$$= \min_{n \geq 0} \left\{ \log\left[\frac{P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(i)} \right)}{\sum_{j \neq i} P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(j)} \right)} \right] \geq a \right\}.$$

- $\Delta_a = (N_a, d_a)$ is then defined as follows:

$$N_a = \min_{0 \leq i \leq M-1} N_i, \quad d_a = i^\star \text{ if } N_a = N_{i^\star}.$$

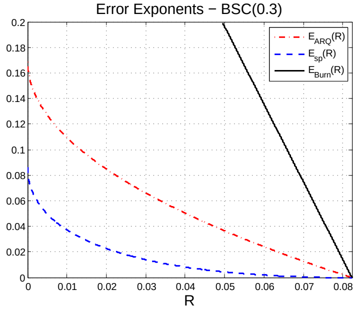
Result: $\mathbb{E}_0[N_a] \lesssim \frac{-\log P_{\text{er}}}{E_{\text{Forney}}(R+\delta)} \Rightarrow E_a(R) \gtrsim E_{\text{Forney}}(R)$

The Stop-Feedback Error Exponent

Theorem - Random Coding Stop-Feedback Error Exponent

The random-coding error exponent of the stop-feedback communication setup with a binary symmetric forward channel is given by

$$E(R) = \beta\left(\delta_{\mathsf{GV}}(R) - \delta_{\mathsf{GV}}(C)\right) = E_{\mathsf{sp}}(R) + C - R.$$



Summary and Conclusions

- We have seen an example as to how hypothesis testing theory can help gain intuition and prove results in coding theory.
- New achievable scheme was given for the unlimited feedback case.
- Results from multi-hypothesis testing were used in order to obtain a tight bound on the error exponent at zero rate.
- An optimal multi-hypothesis test was used in order to prove achievability of an error exponent function for a BSC.
- This lower bound was then shown to be tight for the BSC case.

Summary and Conclusions

- We have seen an example as to how hypothesis testing theory can help gain intuition and prove results in coding theory.
- New achievable scheme was given for the unlimited feedback case.
- Results from multi-hypothesis testing were used in order to obtain a tight bound on the error exponent at zero rate.
- An optimal multi-hypothesis test was used in order to prove achievability of an error exponent function for a BSC.
- This lower bound was then shown to be tight for the BSC case.

Thank you!

Appendix:

DP Formulation | Fictitious Agent | Phase I - Cont.

Phase II | Forney's Exponent | Stop Feedback

More On Multi-hypothesis Testing With Control

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  2. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  2. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  2. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.
- Assume $w > 0$ is the *loss* associated with making the wrong decision.

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma(Y_1, \dots, Y_n, U_1, \dots, U_n)$.
  2. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\dots,N}$ is an *observation control policy*.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.

- Assume $w > 0$ is the *loss* associated with making the wrong decision.
- **The objective:** Find a sequential test $\Delta = (q, N, d)$ that minimizes the total *cost* defined as:

$$V(\pi) \triangleq \mathbb{E}\left[N + w\mathbb{I}\left\{d\left(U^N, Y^N\right) \rightarrow \text{ error}\right\}\right] = \mathbb{E}[N] + wP_{\text{er}}.$$

Hypothesis Testing with Control - Definitions

- A *Multihypothesis test with control* Δ is a triplet (N, q, d) where:
  1. N is the *stopping time*, $N \in \mathcal{F}_n = \sigma\left(Y_1, \ldots, Y_n, U_1, \ldots, U_n\right)$.
  2. $q = \left\{q_k\left(Y^{k-1}, U^{k-1}\right)\right\}_{k=1,\ldots,N}$ is an *observation control policy*.
  3. $d = d\left(Y^N, U^N\right)$ is the *decision rule*.
- Assume $w > 0$ is the *loss* associated with making the wrong decision.
- **The objective:** Find a sequential test $\Delta = (q, N, d)$ that minimizes the total *cost* defined as:

$$V(\pi) \triangleq \mathbb{E}\left[N + w\mathbb{I}\left\{d\left(U^N, Y^N\right) \rightarrow \text{ error}\right\}\right] = \mathbb{E}[N] + wP_{\text{er}}.$$

- Asymptotic regime: $w \to \infty$.

Minimization of the Cost Using DP

- Problem (P): Find $\Delta = (N, q, d)$ that minimizes $V(\pi)$.

Minimization of the Cost Using DP

- Problem (P): Find $\Delta = (N, q, d)$ that minimizes $V(\pi)$.
- Solution - Dynamic Programming:
- Assume control u has been taken and Y has been observed.

Minimization of the Cost Using DP

- Problem (P): Find $\Delta = (N, q, d)$ that minimizes $V(\pi)$.
- Solution - Dynamic Programming:
- Assume control u has been taken and Y has been observed.
- Then the posterior distribution of the hypotheses, $\Phi^u(\pi, y)$ is given by

Minimization of the Cost Using DP

- Problem (P): Find $\Delta = (N, q, d)$ that minimizes $V(\pi)$.
- Solution - Dynamic Programming:
- Assume control u has been taken and Y has been observed.
- Then the posterior distribution of the hypotheses, $\Phi^u(\pi, y)$ is given by

$$\Phi^u(\pi, y) = \left(\pi_0 \frac{p_0^u(y)}{p_\pi^u(y)}, \pi_1 \frac{p_1^u(y)}{p_\pi^u(y)} \ldots, \pi_{M-1} \frac{p_{M-1}^u(y)}{p_\pi^u(y)}\right), \ \forall u \in \mathcal{U},$$

where

1. $\pi \triangleq (\pi_0, \ldots, \pi_{M-1})$.
2. $p_\pi^u(y) = \sum_{i=0}^{M-1} \pi_i p_i^u(y)$.

Minimization of the Cost Using DP

- Problem (P): Find $\Delta = (N, q, d)$ that minimizes $V(\pi)$.
- Solution - Dynamic Programming:
- Assume control u has been taken and Y has been observed.
- Then the posterior distribution of the hypotheses, $\Phi^u(\pi, y)$ is given by

$$\Phi^u(\pi, y) = \left(\pi_0 \frac{p_0^u(y)}{p_\pi^u(y)}, \pi_1 \frac{p_1^u(y)}{p_\pi^u(y)} \ldots, \pi_{M-1} \frac{p_{M-1}^u(y)}{p_\pi^u(y)}\right), \; \forall u \in \mathcal{U},$$

where

1. $\pi \triangleq (\pi_0, \ldots, \pi_{M-1})$.
2. $p_\pi^u(y) = \sum_{i=0}^{M-1} \pi_i p_i^u(y)$.

- Define the operator $\mathbb{T}^u, u \in \mathcal{U}$, such that for any measurable function $g\colon \triangle_M \to \mathbb{R}$:

$$(\mathbb{T}^u g)(\pi) = \int g(\Phi^u(\pi, y))\, p_\pi^u(y)\, dy.$$

Lower Bounds on $V^\star(\pi)$

Fact - Solution to Problem (P) (Bertsekas, Shereve 2007)

The optimal value function $V^\star$ satisfies the fixed point equation:

$$V^\star(\pi) = \min\left\{1 + \min_{u\in\mathcal{U}} (\mathbb{T}^u V^\star)(\pi), \min_{j\in\{0,\ldots,M-1\}} (1-\pi_j)\, w\right\}.$$

Lower Bounds on $V^\star(\pi)$

Fact - Solution to Problem (P) (Bertsekas, Shereve 2007)

The optimal value function $V^\star$ satisfies the fixed point equation:

$$V^\star(\pi) = \min\left\{1 + \min_{u\in\mathcal{U}} (\mathbb{T}^u V^\star)(\pi), \min_{j\in\{0,\ldots,M-1\}} (1-\pi_j)\, w\right\}.$$

Theorem (Javidi *et al. 2013*)

Define $D_{\max} = \max_{i,j\in\{0,\ldots,M-1\}} \max_{u\in\mathcal{U}} D\left(p_i^u \,\|\, p_j^u\right)$, and $I_{\max} = \max_{u\in\mathcal{U}} \max_{\tilde{\pi}\in\triangle_M} I\left(\tilde{\pi}; p_{\tilde{\pi}}^u\right)$.

For $w > \frac{\log(M)}{I_{\max}}$ and arbitrary $\delta \in (0, 0.5]$,

$$V^\star(\pi) \geq \left[\frac{H(\pi) - h_2(\delta) - \delta\log(M-1)}{I_{\max}} + \frac{\log\left(\frac{1-w^{-1}}{w^{-1}}\right) - \log\left(\frac{1-\delta}{\delta}\right)}{D_{\max}}\mathbb{I}\left\{\max_i \pi_i \leq 1-\delta\right\} - \hat{K}'\right]^+$$

Information Acquisition Problem

Information Acquisition Problem

- Find a test $\Delta = (N, q, d)$ with the object to

$$\text{minimize } \mathbb{E}\left[N\right] \text{ subject to } P_{\text{er}} \leq \epsilon$$

Information Acquisition Problem

- Find a test $\Delta = (N, q, d)$ with the object to

$$\text{minimize } \mathbb{E}\left[N\right] \text{ subject to } P_{\text{er}} \leq \epsilon$$

- This is the primal (constrained) version of the previous problem

Information Acquisition Problem

- Find a test $\Delta = (N, q, d)$ with the object to

$$\text{minimize } \mathbb{E}[N] \text{ subject to } P_{\text{er}} \leq \epsilon$$

- This is the primal (constrained) version of the previous problem

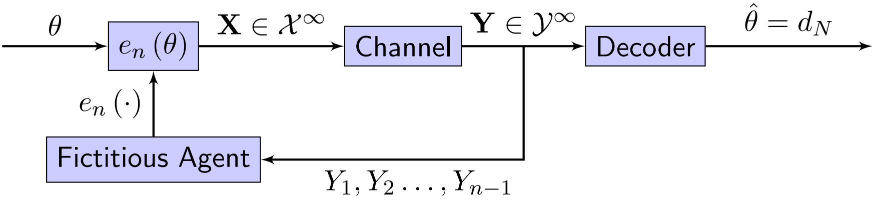
Theorem - The relation between the stopping time and the value function (Javidi *et al. 2013*)

Let $\mathbb{E}[N_\epsilon^\star]$ be the minimal expected number of samples required to achieve $P_{\text{er}} \leq \epsilon$. Then

$$\mathbb{E}[N_\epsilon^\star] \geq (1 - \epsilon w)(V^\star(\pi) - 1)$$

where $V^\star(\pi)$ is the optimal solution to Problem (P) for a prior π and cost for wrong decision w.

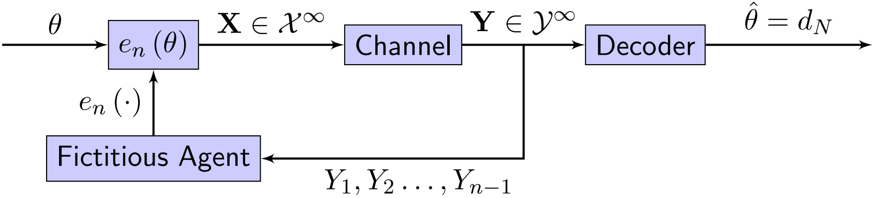
More On Fictitious Agent



- *Error Probability*: $P_{\text{er}} \triangleq \frac{1}{M} \sum_{i=0}^{M-1} \mathbb{P}\left(d_N \neq i \mid \theta = i\right)$.
- *Expected Transmission Time*: $\mathbb{E}\left[N\right] = \frac{1}{M} \sum_{i=0}^{M-1} \mathbb{E}\left[N \mid \theta = i\right]$.
- *Rate*: $R \triangleq \frac{\log(M)}{\mathbb{E}[N]}$.
- *Error exponent*: $E\left(R\right) = \limsup_{\mathbb{E}[N] \to \infty} \frac{-\log(P_{\text{er}})}{\mathbb{E}[N]}$.

Q: Is the VL coding problem amenable to hypothesis testing analysis?

A: Yes!



- $\pi = \left[\frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M}\right]$.
- $p_i^u(k) = p(k \mid e(i))$.
- $D_{\max} = \max_{i,k} \max_{u \in \mathcal{U}} D\left(p_i^u \parallel p_j^u\right) = \max_{j,k} D(p(\cdot \mid j) \parallel p(\cdot \mid k)) \triangleq C_1$.
- $I_{\max} = \max_{u \in \mathcal{U}} \max_{\tilde{\pi} \in \triangle_M} I(\tilde{\pi}; p_{\tilde{\pi}}^u) = C$.

More On Phase I (Burnashev Achievability)

- For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X-i.i.d. sequence.

Direct Statement - Phase I (Tentative Decision)

- For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X-i.i.d. sequence.
- Let $\boldsymbol{x}^{(\boldsymbol{i})}$ be the codeword assigned to the i'th message.

Direct Statement - Phase I (Tentative Decision)

- For each message $i \in \{0, \ldots, M-1\}$ randomly draw an infinite P_X-i.i.d. sequence.
- Let $\boldsymbol{x}^{(i)}$ be the codeword assigned to the i'th message.
- For each $i \in \{0, \ldots, M-1\}$, define the following two hypotheses:

$$
\begin{aligned}
H_0^i:\ \Pr\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}\right) &= p\left(\boldsymbol{y} \mid \boldsymbol{x}^{(i)}\right) P_X\left(\boldsymbol{x}^{(i)}\right), \\
H_1^i:\ \Pr\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}\right) &= \Pr\left(\boldsymbol{y}\right) P_X\left(\boldsymbol{x}^{(i)}\right).
\end{aligned}
$$

- For each message $i \in \{0, \dots, M-1\}$ randomly draw an infinite P_X-i.i.d. sequence.
- Let $\boldsymbol{x}^{(i)}$ be the codeword assigned to the i'th message.
- For each $i \in \{0, \dots, M-1\}$, define the following two hypotheses:

$$H_0^i: \ \Pr\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}\right) = p\left(\boldsymbol{y} \mid \boldsymbol{x}^{(i)}\right) P_X\left(\boldsymbol{x}^{(i)}\right),$$
$$H_1^i: \ \Pr\left(\boldsymbol{x}^{(i)}, \boldsymbol{y}\right) = \Pr\left(\boldsymbol{y}\right) P_X\left(\boldsymbol{x}^{(i)}\right).$$

- Define

$$N_{I,k}^i = \inf_{n \geq 0} \left\{ \log\left[\frac{p\left([\boldsymbol{y}]_n \mid \left[\boldsymbol{x}^{(i)}\right]_n\right)}{\Pr\left([\boldsymbol{y}]_n\right)}\right] \geq (1+\epsilon)\log(M) \right\}$$
$$= \inf_{n \geq 0} \left\{ \sum_{j=1}^{n} \log\left[\frac{p\left(y_j \mid x_j^{(i)}\right)}{\Pr\left(y_j\right)}\right] \geq (1+\epsilon)\log(M) \right\}.$$

More On Phase II (Burnashev Achievability)

Achievability - Phase II (Confirmation)

- Let m_1 be the message chosen at Phase I.

Achievability - Phase II (Confirmation)

- Let m_1 be the message chosen at Phase I.
- Define $j^\star$ and $k^\star$ by $D\left(p\left(\cdot \mid j^\star\right) \| \, p\left(\cdot \mid k^\star\right)\right) = C_1$

Achievability - Phase II (Confirmation)

- Let m_1 be the message chosen at Phase I.
- Define $j^\star$ and $k^\star$ by $D\left(p\left(\cdot \mid j^\star\right) \| p\left(\cdot \mid k^\star\right)\right) = C_1$
- Encoder: Sends $(j^\star, j^\star \ldots)$ if m_1 is correct and $(k^\star, k^\star \ldots)$ o.w.

Achievability - Phase II (Confirmation)

- Let m_1 be the message chosen at Phase I.
- Define $j^\star$ and $k^\star$ by $D\left(p\left(\cdot \mid j^\star\right) \| \, p\left(\cdot \mid k^\star\right)\right) = C_1$
- Encoder: Sends $(j^\star, j^\star \ldots)$ if m_1 is correct and $(k^\star, k^\star \ldots)$ o.w.
- Decoder: Runs an SPRT with

$$H_{ACK} : Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK} : Y_i \sim p\left(\cdot \mid k^\star\right).$$

Achievability - Phase II (Confirmation)

- Let m_1 be the message chosen at Phase I.
- Define $j^\star$ and $k^\star$ by $D\left(p\left(\cdot \mid j^\star\right) \| p\left(\cdot \mid k^\star\right)\right) = C_1$
- Encoder: Sends $(j^\star, j^\star \ldots)$ if m_1 is correct and $(k^\star, k^\star \ldots)$ o.w.
- Decoder: Runs an SPRT with

$$H_{ACK}: Y_i \sim p\left(\cdot \mid j^\star\right),$$
$$H_{NACK}: Y_i \sim p\left(\cdot \mid k^\star\right).$$

- For large M,

$$\mathbb{E}\left[N_{II,k}\right] = \pi_A \mathbb{E}_{P_A}\left[N_{II,k}\right] + \pi_N \mathbb{E}_{P_N}\left[N_{II,k}\right] \approx \quad \mathbb{E}_{P_A}\left[N_{II,k}\right] \lesssim \frac{-\log\left(P_e\right)}{C_1}.$$

- $\Rightarrow \mathbb{E}[N] \approx \mathbb{E}\left[N_{I,1}\right] + \mathbb{E}\left[N_{II,1}\right] \lesssim \frac{\log(M)}{C} + \frac{-\log(P_e)}{C_1}.$

Forney's Error Exponent

Limited Feedback - "One Shot" Schemes

Example - ARQ scheme:

Limited Feedback - "One Shot" Schemes

Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.

Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.

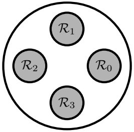
Limited Feedback - "One Shot" Schemes

Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.
- Decoding:

Example - ARQ scheme:

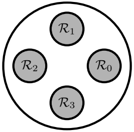
- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.
- Decoding:
  - Partition $\mathbb{R}^n$ into M decision regions and one erasure area.



Limited Feedback - "One Shot" Schemes

Example - ARQ scheme:

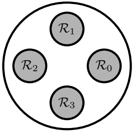
- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.
- Decoding:
  - Partition $\mathbb{R}^n$ into M decision regions and one erasure area.



  - If $Y_{nk+1}^{(n+1)k} \in \bigcup_{i=0}^{M-1} \mathcal{R}_i$ sent the stopping bit and decode.

Example - ARQ scheme:

- Codebook: M randomly chosen codewords, each of length n.
- Encoding: Send the ith codeword periodically to transmit the ith message.
- Decoding:
  - Partition $\mathbb{R}^n$ into M decision regions and one erasure area.



  - If $Y_{nk+1}^{(n+1)k} \in \bigcup_{i=0}^{M-1} \mathcal{R}_i$ sent the stopping bit and decode.
  - Else, wait for the next n symbols and repeat the process.

ARQ Scheme - Analysis

- Let

$$\mathcal{E}_{1,k} = \{\text{Not making the right decision on the } k\text{th round}\},$$
$$\mathcal{E}_{2,k} = \{\text{Making an undetected error on the } k\text{th round}\}.$$

ARQ Scheme - Analysis

- Let

$$\mathcal{E}_{1,k} = \{\text{Not making the right decision on the } k\text{th round}\},$$
$$\mathcal{E}_{2,k} = \{\text{Making an undetected error on the } k\text{th round}\}.$$

- Then $\mathbb{P}\left(\text{errasure}\right) = \mathbb{P}\left(\mathcal{R}_M\right) = \mathbb{P}\left(\mathcal{E}_1\right) - \mathbb{P}\left(\mathcal{E}_2\right).$

ARQ Scheme - Analysis

- Let

$$\mathcal{E}_{1,k} = \{\text{Not making the right decision on the } k\text{th round}\},$$
$$\mathcal{E}_{2,k} = \{\text{Making an undetected error on the } k\text{th round}\}.$$

- Then $\mathbb{P}(\text{errasure}) = \mathbb{P}(\mathcal{R}_M) = \mathbb{P}(\mathcal{E}_1) - \mathbb{P}(\mathcal{E}_2)$.
- More results:

$$\mathbb{E}[N] = n\sum_{k=1}^{\infty} k\mathbb{P}(\text{stop after } k \text{ rounds}) = \frac{n}{1-\mathbb{P}(\mathcal{R}_M)}$$
$$R = \frac{\log(M)}{\mathbb{E}[N]} = \frac{\log(M)}{n}(1-\mathbb{P}(\mathcal{R}_M)) = \tilde{R}(1-\mathbb{P}(\mathcal{R}_M))$$
$$P_{\text{er}} = \sum_{k=1}^{\infty} (\mathbb{P}(\mathcal{R}_M))^{k-1}\,\mathbb{P}(\mathcal{E}_2)$$

- Let

$$\mathcal{E}_{1,k} = \{\text{Not making the right decision on the } k\text{th round}\}$$
$$\mathcal{E}_{2,k} = \{\text{Making an undetected error on the } k\text{th round}\}$$

- Then $\mathbb{P}\left(\text{errasure}\right) = \mathbb{P}\left(\mathcal{R}_M\right) = \mathbb{P}\left(\mathcal{E}_1\right) - \mathbb{P}\left(\mathcal{E}_2\right) \to 0.$
- More results:

$$\mathbb{E}\left[N\right] = n\sum_{k=1}^{\infty} k\mathbb{P}\left(\text{stop after } k \text{ rounds}\right) = \frac{n}{1-\mathbb{P}\left(\mathcal{R}_M\right)} \to n.$$

$$R = \frac{\log\left(M\right)}{\mathbb{E}\left[N\right]} = \frac{\log\left(M\right)}{n}\left(1-\mathbb{P}\left(\mathcal{R}_M\right)\right) \to \frac{\log\left(M\right)}{n} \triangleq \tilde{R}.$$

$$P_{\text{er}} = \sum_{k=1}^{\infty}\left(\mathbb{P}(\mathcal{R}_M)\right)^{k-1}\mathbb{P}\left(\mathcal{E}_2\right) \to \mathbb{P}\left(\mathcal{E}_2\right).$$

$$E_{\text{Forney}}\left(R\right) = \frac{-\log\left(P_{\text{er}}\right)}{\mathbb{E}\left[N\right]} \to -\frac{1}{n}\log\left(\mathbb{P}\left(\mathcal{E}_2\right)\right).$$

Forney's Decision Regions (Forney 1968)

- Define

$$\mathcal{R}_m^\star = \left\{ y \in \mathcal{Y}^n : \frac{p\left(y \mid x_m\right)}{\sum_{m' \neq m} p\left(y \mid x_m\right)} \geq \exp\left(nT\right) \right\}, \quad m \in \{0, \ldots, M-1\},$$

$$\mathcal{R}_M^\star = \bigcap_{m=0}^{M-1} \left(\mathcal{R}_m^\star\right)^c,$$

Forney's Decision Regions (Forney 1968)

- Define

$$\mathcal{R}_m^\star = \left\{ y \in \mathcal{Y}^n : \frac{p\left(y \mid x_m\right)}{\sum_{m' \neq m} p\left(y \mid x_m\right)} \geq \exp\left(nT\right) \right\}, \quad m \in \{0, \ldots, M-1\},$$

$$\mathcal{R}_M^\star = \bigcap_{m=0}^{M-1} \left(\mathcal{R}_m^\star\right)^c,$$

and $e_i\left(R, T\right) \triangleq \limsup_{n \to \infty} \left[-\frac{1}{n} \log\left(\mathbb{P}\left(\mathcal{E}_i\right)\right) \right]$.

Forney's Decision Regions (Forney 1968)

- Define

$$\mathcal{R}_m^\star = \left\{ y \in \mathcal{Y}^n : \frac{p\left(y \mid x_m\right)}{\sum_{m' \neq m} p\left(y \mid x_m\right)} \geq \exp\left(nT\right) \right\}, \quad m \in \{0, \ldots, M-1\},$$

$$\mathcal{R}_M^\star = \bigcap_{m=0}^{M-1} \left(\mathcal{R}_m^\star\right)^c,$$

and $e_i\left(R,T\right) \triangleq \limsup_{n\to\infty} \left[-\frac{1}{n}\log\left(\mathbb{P}\left(\mathcal{E}_i\right)\right)\right]$.

Theorem - Forney's error exponents for the BSC(ϵ) (Somekh-Baruch,Merhav 2011)

Let $\beta \triangleq \log\left(\frac{1-\epsilon}{\epsilon}\right)$. If $R < \log\left(2\right) - h_2\left(\epsilon + \frac{T}{\beta}\right)$ then $e_1\left(R,T\right) > 0$ and $e_2\left(R,T\right) = e_1\left(R,T\right) + T$. Otherwise $e_1\left(R,T\right) = 0$.

- $\Rightarrow$ If $0 < e_1\left(R,T\right) \to 0$ then and $E_{\text{Forney}}\left(R\right) = e_2\left(R,T\right) \approx T$

Forney's Achievable Error Exponent (Forney 1968)

- Define $\delta_{\mathsf{GV}}(R) = \{\delta\colon h_2(\delta) = \log(2) - R\}$.
- Then,

$$R \approx \log(2) - h_2\left(\epsilon + \frac{T}{\beta}\right)$$

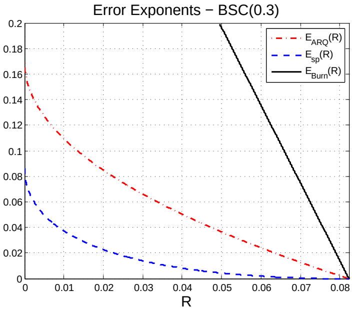
$$\Leftrightarrow E_{\mathsf{Forney}}(R) = T \approx \beta\left(\delta_{\mathsf{GV}}(R) - \delta_{\mathsf{GV}}(C)\right) = E_{\mathsf{sp}}(R) + C - R.$$

Forney's Achievable Error Exponent (Forney 1968)

- Define $\delta_{\mathsf{GV}}(R) = \{\delta \colon h_2(\delta) = \log(2) - R\}$.
- Then,

$$R \approx \log(2) - h_2\left(\epsilon + \frac{T}{\beta}\right)$$

$$\Leftrightarrow E_{\mathsf{Forney}}(R) = T \approx \beta\left(\delta_{\mathsf{GV}}(R) - \delta_{\mathsf{GV}}(C)\right) = E_{\mathsf{sp}}(R) + C - R.$$



Reliability Func. (Stop Feedback)

Lower Bound on $E(R)$

- Main idea: Decode using Δ_a

Lower Bound on $E(R)$

- Main idea: Decode using Δ_a
- Recall we've defined

$$N_i = \min_{n \geq 0} \left\{ L_i(n) \geq a + \log \left(\sum_{j \neq i} \exp \{ L_j(n) \} \right) \right\}$$
$$= \min_{n \geq 0} \left\{ \log \left[\frac{P_i([\mathbf{z}]_n)}{\sum_{j \neq i} P_j([\mathbf{z}]_n)} \right] \geq a \right\}$$
$$= \min_{n \geq 0} \left\{ \log \left[\frac{P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(i)}\right)}{\sum_{j \neq i} P_{\mathbf{Y}|\mathbf{X}}\left([\mathbf{y}]_n \mid [\mathbf{x}]_n^{(j)}\right)} \right] \geq a \right\}.$$

- $\Delta_a = (N_a, d_a)$ is then defined as follows:

$$N_a = \min_{0 \leq i \leq M-1} N_i, \quad d_a = i^\star \text{ if } N_a = N_{i^\star}.$$

Lower Bound - Proof Outline, Cont.

- It holds that $a \leq -\log P_{\text{er}}$.
- $\mathbb{E}_0 [N_a] = \sum_{n=0}^{\infty} P_0 (N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0 (N_0 \geq n)$.

Lower Bound - Proof Outline, Cont.

- It holds that $a \leq -\log P_{\text{er}}$.
- $\mathbb{E}_0 [N_a] = \sum_{n=0}^{\infty} P_0 (N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0 (N_0 \geq n)$.
- For arbitrarily small $\delta > 0$, take $\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log(2) - h_2 \left(\epsilon + \frac{a}{n\beta} \right) - \delta \right\}$.

Lower Bound - Proof Outline, Cont.

- It holds that $a \leq -\log P_{\text{er}}$.
- $\mathbb{E}_0\left[N_a\right] = \sum_{n=0}^{\infty} P_0\left(N_0 \geq n\right) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0\left(N_0 \geq n\right)$.
- For arbitrarily small $\delta > 0$, take
  $\bar{n} \triangleq \max_{n\in\mathbb{N}} \left\{ \frac{\log M}{n} \geq \log\left(2\right) - h_2\left(\epsilon + \frac{a}{n\beta}\right) - \delta \right\}$.
- Then:
  1. $\bar{n} \to \infty$ as $M \to \infty$.

Lower Bound - Proof Outline, Cont.

- It holds that $a \leq -\log P_{\text{er}}$.
- $\mathbb{E}_0 \left[N_a \right] = \sum_{n=0}^{\infty} P_0 \left(N_0 \geq n \right) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0 \left(N_0 \geq n \right).$
- For arbitrarily small $\delta > 0$, take
  $\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log(2) - h_2 \left(\epsilon + \frac{a}{n\beta} \right) - \delta \right\}.$
- Then:
  1. $\bar{n} \to \infty$ as $M \to \infty$.
  2. $\bar{n} \leq \frac{a}{E_{\text{Forney}}(R+\delta)} \leq \frac{-\log P_{\text{er}}}{E_{\text{Forney}}(R+\delta)}$.

Lower Bound - Proof Outline, Cont.

- It holds that $a \leq -\log P_{\text{er}}$.
- $\mathbb{E}_0\left[N_a\right] = \sum_{n=0}^{\infty} P_0\left(N_0 \geq n\right) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0\left(N_0 \geq n\right).$
- For arbitrarily small $\delta > 0$, take
  $\bar{n} \triangleq \max_{n\in\mathbb{N}}\left\{\frac{\log M}{n} \geq \log(2) - h_2\left(\epsilon + \frac{a}{n\beta}\right) - \delta\right\}$.
- Then:
  1. $\bar{n} \to \infty$ as $M \to \infty$.
  2. $\bar{n} \leq \frac{a}{E_{\text{Forney}}(R+\delta)} \leq \frac{-\log P_{\text{er}}}{E_{\text{Forney}}(R+\delta)}$.
  3. For any $n_0 \geq \bar{n}+1$

$$P_0\left(N_0 \geq n_0\right) \leq P_0\left(\log\left[\frac{P_{\mathbf{Y}|\mathbf{X}}\left(\left[\mathbf{y}\right]_{n_0} \mid \left[\mathbf{x}\right]_{n_0}^{(i)}\right)}{\sum_{j\neq i} P_{\mathbf{Y}|\mathbf{X}}\left(\left[\mathbf{y}\right]_{n_0} \mid \left[\mathbf{x}\right]_{n_0}^{(j)}\right)}\right] < a\right) \leq \quad e^{-n\delta}$$

- $\Rightarrow$ Asymptotically, $\mathbb{E}_0\left[N_a\right] \lesssim \frac{-\log P_{\text{er}}}{E_{\text{Forney}}(R+\delta)} \Rightarrow E_a(R) \gtrsim E_{\text{Forney}}(R)$.

Upper Bound on $E(R)$ - Proof Outline

- Define $\Lambda_i(N) \triangleq \log\left[\frac{p\left(y|x^{(i)}\right)}{\sum_{j=0,j\neq i}^{M-1} p\left(y|x^{(j)}\right)}\right]$ and $\Omega_{i,\bar{n}} \triangleq \{d = i, N \leq \bar{n}\}$.

- Define $\Lambda_i(N) \triangleq \log\left[\frac{p\left(y|x^{(i)}\right)}{\sum_{j=0,j\neq i}^{M-1} p\left(y|x^{(j)}\right)}\right]$ and $\Omega_{i,\bar{n}} \triangleq \{d = i, N \leq \bar{n}\}$.
- On the one hand $P_{\text{er}}(\Delta) = \sum_{j=0,j\neq i}^{M-1} P_j(d = i)$.

Upper Bound on $E(R)$ - Proof Outline

- Define $\Lambda_i(N) \triangleq \log\left[\frac{p\left(y|x^{(i)}\right)}{\sum_{j=0,j\neq i}^{M-1} p\left(y|x^{(j)}\right)}\right]$ and $\Omega_{i,\bar{n}} \triangleq \{d=i, N \leq \bar{n}\}$.
- On the one hand $P_{\text{er}}(\Delta) = \sum_{j=0,j\neq i}^{M-1} P_j(d=i)$.
- On the other hand:

$$\sum_{j=0,j\neq i}^{M-1} P_j(d=i) = \sum_{j=0,j\neq i}^{M-1} \sum_{\mathbf{z}} \mathbb{I}\{\mathbf{z}: d=i\} P_j(\mathbf{z})$$
$$= \mathbb{E}_i\left[\mathbb{I}\{\mathbf{z}: d=i\} \frac{\sum_{j=0,j\neq i}^{M-1} P_j(\mathbf{z})}{P_i(\mathbf{z})}\right]$$
$$\geq e^{-a} P_i\left(\Omega_{i,\bar{n}}, \sup_{n\leq\bar{n}} \{\Lambda_i(n) < a\}\right).$$

- $\Rightarrow p_e(\Delta) e^a \geq 1 - p_e(\Delta) - P_i(N \geq \bar{n}) - P_i\left(\sup_{n\leq\bar{n}} \{\Lambda_i(n) > a\}\right)$.

- Using Markov ineq. we obtaine

$$\frac{\mathbb{E}[N]}{\bar{n}} \geq 1 - p_e(\Delta)(e^a - 1) - P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right).$$

Upper bound on $E(R)$ - Proof Outline, Cont.

- Using Markov ineq. we obtaine

$$\frac{\mathbb{E}[N]}{\bar{n}} \geq 1 - p_e(\Delta)(e^a - 1) - P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right).$$

- Take $a \triangleq -(1-\delta_1)\log(p_e(\Delta))$.

Upper bound on $E(R)$ - Proof Outline, Cont.

- Using Markov ineq. we obtaine

$$\frac{\mathbb{E}[N]}{\bar{n}} \geq 1 - p_e(\Delta)(e^a - 1) - P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right).$$

- Take $a \triangleq -(1-\delta_1)\log(p_e(\Delta))$.
- Take $\bar{n} = (1+\delta_2)\mathbb{E}[N]$, assume by contradiction that $E(R) > E_{\text{Forney}}(R)$ and get that

$$P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right) \to 0.$$

Upper bound on $E(R)$ - Proof Outline, Cont.

- Using Markov ineq. we obtaine

$$\frac{\mathbb{E}[N]}{\bar{n}} \geq 1 - p_e(\Delta)(e^a - 1) - P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right).$$

- Take $a \triangleq -(1-\delta_1)\log(p_e(\Delta))$.
- Take $\bar{n} = (1+\delta_2)\mathbb{E}[N]$, assume by contradiction that $E(R) > E_{\text{Forney}}(R)$ and get that

$$P_i\left(\sup_{n \leq \bar{n}} \{\Lambda_i(n) > a\}\right) \to 0.$$

- Conclude that $E(R) \leq E_{\text{Forney}}(R)$.