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

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

2. Variable-Length Coding with Feedback
   - Unlimited Feedback
   - ARQ Schemes
   - Stop-Feedback Scheme
1 Sequential Hypothesis Testing
   - Sequential Binary Hypothesis Testing
   - Multi-hypothesis Testing
   - Multi-hypothesis Testing with Control

2 Variable-Length Coding with Feedback
   - Unlimited Feedback
   - ARQ Schemes
   - Stop-Feedback Scheme

3 Summary and Conclusions
Observation sequence:

\[ Y_1, Y_2, Y_3, \ldots \overset{i.i.d.}{\sim} P. \]
Sequential Binary Hypothesis Testing - Setting

- **Observation sequence:**
  \[ Y_1, Y_2, Y_3, \ldots \overset{i.i.d.}{\sim} P. \]

- **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:**
  \[
  Y_1, Y_2, Y_3, \ldots \overset{	ext{i.i.d.}}{\sim} P.
  \]

- **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 \( \mathbb{P} \{ H_1 \} = \pi_1 = 1 - \pi_0 \).
A Sequential binary hypothesis test is a pair \( \Delta = (N, d) \) where:

- \( N \) is the *stopping time* (such that \( \{ N = n \} , \{ N > n \} \in \sigma (Y^n_1) \)).
- \( d : Y^N_1 \rightarrow \{ H_0, H_1 \} \) is the *terminal decision rule*. 

Two types of errors:

1. **Type 1 error**: Reject the null hypothesis when correct \( \alpha \equiv P_{H_0}(d = H_1) \).
2. **Type 2 error**: Accept the null hypothesis when incorrect \( \beta \equiv P_{H_1}(d = H_0) \).
### Definition - Sequential Binary Hypothesis Test

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

- $N$ is the *stopping time* (such that $\{N = n\}, \{N > n\} \in \sigma(Y^n_1)$).
- $d : Y^n_1 \to \{H_0, H_1\}$ 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 Hypothesis Testing

Sequential Binary Hypothesis Testing

Sequential Probability Ratio Test (SPRT)

Define the Log Likelihood Ratio function (LLR):

\[ L_n(y_{n+1}) \equiv \log \left[ \frac{P_0(y_n)}{P_1(y_n)} \right] = \sum_{i=1}^{n} \log \left[ \frac{P_0(y_i)}{P_1(y_i)} \right] \]

Select two boundary values \( A \) and \( B \) such that \( 0 < B < A < 1 \)

Definition - SPRT (Wald 1943)

The Sequential Probability Ratio Test (SPRT) \( \Delta_{SPRT} \) is defined as follows:

\[ \Delta_{SPRT} = \min \{ n \in \mathbb{N} : L_n(Y_{n+1}) \leq \log(B) \text{ or } L_n(Y_{n+1}) \geq \log(A) \} \]

\[ d_{SPRT} = \begin{cases} 
H_1 & \text{if } L_{N_{SPRT}}(Y_{N_{SPRT}+1}) \leq \log(B) \\
H_0 & \text{if } L_{N_{SPRT}}(Y_{N_{SPRT}+1}) \geq \log(A) 
\end{cases} \]
Define the \textit{Log Likelihood Ratio function (LLR)}:

\[
L_n(y^n_1) \triangleq \log \left[ \frac{P_0(y^n_1)}{P_1(y^n_1)} \right] = \sum_{i=1}^{n} \log \left[ \frac{P_0(y_i)}{P_1(y_i)} \right]
\]
Sequential Probability Ratio Test (SPRT)

- Define the Log Likelihood Ratio function (LLR):
  \[ L_n(y^n_1) \triangleq \log \left( \frac{P_0(y^n_1)}{P_1(y^n_1)} \right) = \sum_{i=1}^{n} \log \left( \frac{P_0(y_i)}{P_1(y_i)} \right) \]

- 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)*:

\[ L_n(y^n_1) \triangleq \log \left[ \frac{P_0(y^n_1)}{P_1(y^n_1)} \right] = \sum_{i=1}^{n} \log \left[ \frac{P_0(y_i)}{P_1(y_i)} \right] \]

- Select two *boundary values* \( A \) and \( B \) such that \( 0 < B < 1 < A \)

**Definition - SPRT (Wald 1943)**

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

\[ N_{\text{SPRT}} \triangleq \min \left\{ n \in \mathbb{N} : L_n(Y^n_1) \leq \log(B) \quad \text{or} \quad L_n(Y^n_1) \geq \log(A) \right\} \]

\[ d_{\text{SPRT}} \triangleq \begin{cases} H_1 & \text{if} \quad L_{N_{\text{SPRT}}}(Y^{N_{\text{SPRT}}}_1) \leq \log(B) \\ H_0 & \text{if} \quad L_{N_{\text{SPRT}}}(Y^{N_{\text{SPRT}}}_1) \geq \log(A) \end{cases} \]
Basic Properties of the SPRT

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

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

Similarly,

$$
\mathbb{E}_{P_1}[N_{SPRT}] \leq -\log(B) \mathcal{D}(P_1 \parallel P_0) \leq -\log(\beta) \mathcal{D}(P_1 \parallel P_0).
$$

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

Let $\Delta_{SPRT} = (N_{SPRT}, d_{SPRT})$ be Wald’s SPRT with error probabilities $\alpha_{SPRT}$ and $\beta_{SPRT}$, and let $\Delta' = (N', d')$ be any other sequential decision rule with finite $\mathbb{E}_{P_1}[N']$, $\mathbb{E}_{P_0}[N']$ and error probabilities $\alpha'$ and $\beta'$ satisfying $\alpha' < \alpha_{SPRT}$ and $\beta' < \beta_{SPRT}$.

Then

$$
\mathbb{E}_{P_1}[N'] \geq \mathbb{E}_{P_1}[N_{SPRT}], \quad \mathbb{E}_{P_0}[N'] \geq \mathbb{E}_{P_0}[N_{SPRT}].
$$
Define $\tilde{N} \triangleq \min \{ n \in \mathbb{N} : L_n \geq \log (A) \}$. Then,

$$\mathbb{E}_{P_0} [N_{\text{SPRT}}] \leq \mathbb{E}_{P_0} [\tilde{N}] \leq \frac{\log (A)}{D(P_0 \parallel P_1)} \leq \frac{-\log (\alpha)}{D(P_0 \parallel P_1)}.$$
Define $\bar{N} \triangleq \min \{ n \in \mathbb{N} : L_n \geq \log (A) \}$. Then,

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

Similarly,

$$
\mathbb{E}_{P_1}[N_{\text{SPRT}}] \leq \frac{-\log (B)}{D(P_1 \parallel P_0)} \leq \frac{-\log (\beta)}{D(P_1 \parallel P_0)}.
$$
Basic Properties of the SPRT

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

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

Similarly,

$$\mathbb{E}_{P_1} [N_{SPRT}] \leq \frac{-\log (B)}{D(P_1 \parallel P_0)} \leq \frac{-\log (\beta)}{D(P_1 \parallel P_0)}.$$

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

Let $\Delta_{SPRT} = (N_{SPRT}, d_{SPRT})$ be Wald’s SPRT with error probabilities $\alpha_{SPRT}$ and $\beta_{SPRT}$, and let $\Delta' = (N', d')$ be any other sequential decision rule with finite $\mathbb{E}_{P_1} [N'], \mathbb{E}_{P_0} [N']$ and error probabilities $\alpha'$ and $\beta'$ satisfying

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

Then

$$\mathbb{E}_{P_1} [N'] \geq \mathbb{E}_{P_1} [N_{SPRT}], \quad \mathbb{E}_{P_0} [N'] \geq \mathbb{E}_{P_0} [N_{SPRT}].$$
Sequential Multi-hypothesis Testing - Setting

- **Observation sequence:**
  \[ Y_1, Y_2, Y_3, \ldots \sim \text{i.i.d.} \ P. \]
  where \( Y_i \) is an \( l \)-valued random vector \( (Y_i = (Y_{i,1}, \ldots, Y_{i,l})) \).

- Define \( M (> 2) \) hypotheses:
  \[ H_i: P = P_i, \quad i \in \{0, \ldots, M - 1\} \]
  where \( P_i \) are completely known distinct probability measures.

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

\[ Y_1, Y_2, Y_3, \ldots \overset{i.i.d.}{\sim} P. \]

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

Define \( M (> 2) \) hypotheses:

\[ H_i: P = P_i, \quad i \in \{0, \ldots, M - 1\}. \]

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 \( \Delta \) is a pair \((N, d)\) where \( N \) is the stopping time and \( d \) is the decision rule.
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 $\Delta$, 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 $\alpha_{ji}(\Delta) = P_j(d = i)$ be the probability of accepting the hypothesis $H_i$ when $H_j$ is true (defined for $j \neq i$).

For a multiple hypothesis test $\Delta$, 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 $\overline{R} \triangleq (\overline{R}_0, \overline{R}_1, \ldots, \overline{R}_{M-1})$ be a vector of positive finite numbers and define:

$$\Delta(\overline{R}) \triangleq \{ \Delta : R_i(\Delta) \leq \overline{R}_i, i \in \{0, \ldots, M-1\} \}.$$
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 $\Delta$, 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 $\overline{R} \triangleq (\overline{R}_0, \overline{R}_1, \ldots, \overline{R}_{M-1})$ be a vector of positive finite numbers and define:

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

- Our focus: $\Delta \in \Delta(\overline{R})$ as $R_{\text{max}} \triangleq \max_i \overline{R}_i \to 0$ and $M$ fixed.
Define the $LLR$ of $P_i$ w.r.t. to a (dominating) measure $Q$ by:

$$L_i(n) = \log \left[ \frac{P_i(Y_1, \ldots, Y_n)}{Q(Y_1, \ldots, Y_n)} \right], \quad i \in \{0, \ldots, M - 1\}.$$
Multi-hypothesis SPRT (MSPRT)

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

$$L_i(n) = \log \left[ \frac{P_i(Y_1, \ldots, Y_n)}{Q(Y_1, \ldots, Y_n)} \right], \quad i \in \{0, \ldots, M - 1\}.$$

- 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:

$$L_i(n) = \log \left[ \frac{P_i(Y_1, \ldots, Y_n)}{Q(Y_1, \ldots, Y_n)} \right], \quad i \in \{0, \ldots, M - 1\}.$$ 

- 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:

$$L_i(n) = \log \left[ \frac{P_i(Y_1, \ldots, Y_n)}{Q(Y_1, \ldots, Y_n)} \right], \quad i \in \{0, \ldots, M - 1\}.$$

- 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 - $\Delta_a$ (Baum & Veeravalli 1994, Fishman 1987)

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

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

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

Theorem - Optimality of $\Delta_a$ (Dragalin et al. 2000)

1. For all $i \in \{0, 1, \ldots, M - 1\}$, 
   
   $\inf_{\Delta \in \Delta(P)} \mathbb{E}_i[N] \geq \left[ -\log(R_i) D_i \right] (1 + o(1))$

2. If $a_i = \log[\pi_i R_i]$ then $\mathbb{E}_i[N_{a_i}] \sim -\log(R_i) D_i$ as $R_{\text{max}} \to 0$ for all $m \geq 1$. 

S. Ginzach (Technion)
Define $D_i \triangleq \min_{j \neq i} D(P_i \| P_j)$.

**Theorem - Optimality of $\Delta_a$ (Dragalin et al. 2000)**

1. For all $i \in \{0, 1, \ldots, M - 1\}$
   \[
   \inf_{\Delta \in \Delta(R)} \mathbb{E}_i[N] \geq \left[ \frac{-\log(R_i)}{D_i} \right] (1 + o(1))
   \]

2. If $a_i = \log \left( \frac{\pi_i}{R_i} \right)$ then
   \[
   \mathbb{E}_i[N_a] \sim \frac{-\log(R_i)}{D_i}
   \]
as $R_{\max} \to 0$ 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:** \( \mathbb{P} \{ H_i \} = \pi_i, \quad i \in \{ 0, \ldots, M - 1 \} \).
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: \( U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty. \)
Observation sequence: $Y_1, Y_2, Y_3, \ldots \in \mathcal{Y}^\infty$.

Hypotheses: $\{H_i, i = 0, \ldots, M - 1\}$.

Priors: $\mathbb{P}\{H_i\} = \pi_i$, $i \in \{0, \ldots, M - 1\}$.

Control sequence: $U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty$, $|\mathcal{U}| < \infty$.

$U_n = q(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1})$ [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**: \( U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |U| < \infty. \)

\[ U_n = q(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1}) \quad \text{[Causality Constraint]} \]

- **Assume**: \( Y_n \perp (Y^{n-1}, U^{n-1}) \)
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**: \( U_1, U_2, U_3, \ldots \in \mathcal{U}^\infty, \quad |\mathcal{U}| < \infty \).
  
  \[
  U_n = q(Y_1, \ldots, Y_{n-1}, U_1, \ldots, U_{n-1}) \quad \text{[Causality Constraint]}
  \]
- **Assume**: \( Y_n \perp (Y^{n-1}, U^{n-1}) \)
- **Observation kernel**:
  \[
  p_{i}^{u_n}(y_n) \triangleq \mathbb{P}(Y_n = y_n \mid H_i, U_n = u_n).
  \]
A Multi-hypothesis test with control $\Delta$ is a triplet $(q, N, d)$ where:

1. $q = \{ q_k(Y_k - 1, U_k - 1) \}_{k=1}^N$ is an observation control policy.
2. $N$ is the stopping time, $N \in \sigma(Y_1, ..., Y_n, U_1, ..., U_n)$.
3. $d = d(Y_N, U_N)$ is the decision rule.

The Objective: Find $\Delta$ that:

$$\min_{\Delta} E[N] \text{ subject to } P_{\text{error}} \leq \epsilon$$

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


S. Ginzach (Technion) December 25th, 2014
A Multi-hypothesis test with control $\Delta$ is a triplet $(q, N, d)$ where:

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

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

1. $q = \{q_k (Y^{k-1}, U^{k-1})\}_{k=1,\ldots,N}$ is an observation control policy.
2. $N$ is the stopping time, $N \in \mathcal{F}_n = \sigma (Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.
3. $d = d (Y^N, U^N)$ is the decision rule.

The objective: Find $\Delta$ that:

\[
\minimize \mathbb{E}[N] \text{ subject to } P_{er} \leq \epsilon \quad [\text{Max. Information}]
\]
A *Multi-hypothesis test with control* $\Delta$ is a triplet $(q, N, d)$ where:

1. $q = \{ q_k (Y^{k-1}, U^{k-1}) \}_{k=1,...,N}$ is an *observation control policy*.
2. $N$ is the *stopping time*, $N \in \mathcal{F}_n = \sigma (Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.
3. $d = d (Y^N, U^N)$ is the *decision rule*.

**The Objective:** Find $\Delta$ that:

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

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

**Achievability:** Chernoff (1960), Veeravalli (2012), Javidi (2013)...
Theorem (Javidi et al. 2013)

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

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

where $D_{\text{max}} = \max_{i,j,u} D(p_i^u \parallel p_j^u)$, and $I_{\text{max}} = \max_u I(\hat{\pi}^u; p_{\hat{\pi}}^u)$. 
Lower Bounds on $\mathbb{E} [N^*]$ 

**Theorem (Javidi et al. 2013)**

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

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

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

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

$$
H(\theta \mid \hat{\theta}) \leq h_2(\delta) + \delta \log(M - 1)
$$
Theorem (Javidi et al. 2013)

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

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

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

- Remainder [Fano’s Inequality]: Let \( \delta \) be the error probability of the estimator \( \hat{\theta} \) of \( \theta \). Then

\[
H(\theta | \hat{\theta}) \leq h_2(\delta) + \delta \log(M - 1)
\]
**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.
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: $X_n(\theta) \triangleq X_n(\theta, Y_1, \ldots Y_{n-1}), \forall \theta, Y_1, \ldots Y_{n-1}$
**Variable-Length Coding with Feedback**

**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.
For block codes with fixed block-length $n$:

1. **Rate**: $R \triangleq \frac{\log(M)}{n}$.
2. **Error Exponent**: $E(R) = \limsup_{n \to \infty} \frac{-\log(P_{er})}{n}$.
VL Coding - Performance Indices

For block codes with fixed block-length $n$:

1. Rate: $R \triangleq \frac{\log(M)}{n}$.
2. Error Exponent: $E(R) = \limsup_{n \to \infty} -\frac{\log(P_{er})}{n}$.

For VL codes:

1. Rate: $R \triangleq \frac{\log(M)}{\mathbb{E}[N]}$.
2. Error Exponent: $E(R) = \limsup_{\mathbb{E}[N] \to \infty} -\frac{\log(P_{er})}{\mathbb{E}[N]}$.
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?

S. Ginzach (Technion)  
December 25th, 2014  16 / 30
For block codes with fixed block-length $n$:

1. **Rate**: $R \triangleq \log(M) / n$.
2. **Error Exponent**: $E(R) = \limsup_{n \to \infty} -\log(P_{er}) / n$.

For VL codes:

1. **Rate**: $R \triangleq \log(M) / E[N]$.
2. **Error Exponent**: $E(R) = \limsup_{E[N] \to \infty} -\log(P_{er}) / E[N]$.

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

A: Yes!
i ∈ 0, . . . , M − 1

Encoder \[ X \in \mathcal{X}^\infty \]
DMC \[ Y \in \mathcal{Y}^\infty \]
Decoder \[ \hat{i} = d_N \]

\[ Y_1, Y_2, \ldots, Y_{n-1} \]

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

**Theorem (Javidi et al. 2013)**

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

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

where \( D_{\text{max}} = \max_{i,j,u} D(p_i^u \parallel p_j^u) \), and \( I_{\text{max}} = \max_{u,\tilde{\pi}} I(\tilde{\pi}; p_i^u) \).
Recap:

**Theorem (Javidi et al. 2013)**

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

\[
\mathbb{E}[N^*] \gtrsim \left(1 - \frac{1}{\log(M/\epsilon)}\right) \left[\log(M) - h_2\left(\frac{1}{\log(M/\epsilon)}\right) - \frac{\log(M-1)}{\log(M/\epsilon)} - \hat{K}'\right]
\]

\[
+ \left[-\log\left(\epsilon \log\left(\frac{M}{\epsilon}\right)\right) - \log\left(\log\left(\frac{M}{\epsilon}\right)\right)\right] \left[\frac{1}{M} \leq 1 - \frac{1}{\log(M/\epsilon)}\right]^+ \frac{1}{C_1}
\]
Theorem (Javidi et al. 2013)

For large $M$ and small $\epsilon$,

$$\mathbb{E}[N^*] \gtrsim \frac{\log(M)}{C} + \frac{-\log(P_{er})}{C_1} + O\left(\log\left(\log\left(\frac{M}{\epsilon}\right)\right)\right).$$
Theorem (Javidi et al. 2013)

For large $M$ and small $\epsilon$,

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

- Equivalently,

$$E(R) \lesssim C_1 \left(1 - \frac{R}{C}\right) \triangleq E_B(R).$$
Theorem (Javidi et al. 2013)

For large $M$ and small $\epsilon$,

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

Equivalently,

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

Theorem (Burnashev 1976)

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

$$E(R) = E_B(R).$$
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 $x^{(i)}$ 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 \( x^{(i)} \) be the codeword assigned to the \( i \)'th message.
- Define:

\[
N^i_I = \min_{n \geq 0} \left\{ \sum_{k=1}^{n} \log \left[ \frac{p(y_k \mid x_k^{(i)})}{\Pr(y_k)} \right] \geq (1 + \epsilon) \log (M) \right\}.
\]
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 $x^{(i)}$ be the codeword assigned to the $i$'th message.
- Define:

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

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

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

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

$$\mathbb{E}[N_I] \leq \mathbb{E}[N_I^0] \lesssim \frac{(1 + \epsilon) \log (M)}{C}.$$
Encoder: Sends \((j^*, j^* \ldots)\) if \(d_I\) is correct and \((k^*, k^* \ldots)\) otherwise.
Encoder: Sends \((j^*, j^* \ldots)\) if \(d_I\) is correct and \((k^*, k^* \ldots)\) otherwise.

Decoder: Runs an SPRT with

\[
H_{ACK}: Y_i \sim p(\cdot \mid j^*), \\
H_{NACK}: Y_i \sim p(\cdot \mid k^*). 
\]
Achievability - Phase II (Confirmation)

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

\[
H_{ACK}: Y_i \sim p(\cdot | j^*), \\
H_{NACK}: Y_i \sim p(\cdot | k^*).
\]

- Choose \(j^*\) and \(k^*\) s.t. \(D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1\)
Encoder: Sends \((j^*, j^* \ldots)\) if \(d_I\) is correct and \((k^*, k^* \ldots)\) otherwise.

Decoder: Runs an SPRT with

\[
H_{ACK}: Y_i \sim p(\cdot | j^*), \\
H_{NACK}: Y_i \sim p(\cdot | k^*).
\]

Choose \(j^*\) and \(k^*\) s.t. \(D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1\)

For large \(M\),

\[
\mathbb{E}[N_{II}] \lesssim \frac{-\log(P_e)}{C_1}.
\]
Encoder: Sends \((j^*, j^* \ldots)\) if \(d_I\) is correct and \((k^*, k^* \ldots)\) otherwise.

Decoder: Runs an SPRT with

\[
H_{ACK}: Y_i \sim p(\cdot | j^*),
\]
\[
H_{NACK}: Y_i \sim p(\cdot | k^*).\]

Choose \(j^*\) and \(k^*\) s.t. \(D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1\)

For large \(M\),

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

\[
\Rightarrow \mathbb{E}[N] \approx \mathbb{E}[N_I] + \mathbb{E}[N_{II}] \lesssim \frac{\log(M)}{C} + \frac{-\log(P_e)}{C_1}.
\]
Limited Feedback - "One Shot" Schemes

Example - ARQ scheme:

- $i \in \{0, \ldots, M - 1\}$
- $x^{(i)} \in \{0, 1\}^\infty$
- $y \in \{0, 1\}^\infty$
- $\hat{i} = d_N$

One bit per message
Limited Feedback - "One Shot" Schemes

\[ i \in \{0, \ldots, M - 1\} \quad \xrightarrow{\text{Enc.}} \quad x^{(i)} \in \{0, 1\}^\infty \quad \xrightarrow{\text{BSC}(\epsilon)} \quad y \in \{0, 1\}^\infty \quad \xrightarrow{\text{Dec.}} \quad \hat{i} = d_N \]

One bit per message

Example - ARQ scheme:
- Codebook: \( M \) randomly chosen codewords, each of length \( n \).
Limited Feedback - "One Shot" Schemes

\[ i \in \{0, \ldots, M - 1\} \]
\[ x^{(i)} \in \{0, 1\}^\infty \]
\[ y \in \{0, 1\}^\infty \]
\[ \hat{i} = d_N \]

Example - ARQ scheme:

- **Codebook:** \( M \) randomly chosen codewords, each of length \( n \).
- **Encoding:** Send the \( i \)th codeword periodically to transmit the \( i \)th message.
Limited Feedback - "One Shot" Schemes

$$i \in \{0, \ldots, M - 1\}$$  \hspace{1cm}  $$x^{(i)} \in \{0, 1\}^\infty$$  \hspace{1cm}  $$y \in \{0, 1\}^\infty$$  \hspace{1cm}  $$\hat{i} = d_N$$

Example - ARQ scheme:

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

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

$$\mathcal{R}_M^* = \bigcap_{i=0}^{M-1} (\mathcal{R}_i^*)^c,$$
Define $T > 0$ and for all $i \in \{0, \ldots, M - 1\}$

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

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

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

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

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

$$R + h_2(\delta) - \log(2) = 0.$$
Stop-Feedback Scheme

\[ i \in \{0, \ldots, M - 1\} \]
\[ \mathbf{x}^{(i)} \in \{0, 1\}^\infty \]
\[ \mathbf{y} \in \{0, 1\}^\infty \]
\[ \hat{i} = d_N \]

- 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:

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

- $i \in \{0, \ldots, M - 1\}$
- $x^{(i)} \in \{0, 1\}^\infty$
- $y \in \{0, 1\}^\infty$
- $i = d_N$
- One bit per message

- **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:**

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

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

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

\[ P_i(z) = P_i(x^{(0)}, x^{(1)} \ldots, x^{(M-1)}, y) \triangleq P_{Y|X}(y | x^{(i)}) \prod_{l=0}^{M-1} P_X(x^{(l)}) \]
Error Exponent at $R = 0$

- Define

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

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

- Under each $H_i$ the elements of $z$ are i.i.d.
Error Exponent at $R = 0$

- Define

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

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

- Under each $H_i$ the elements of $z$ are i.i.d.

- Hence, it the limit

$$\inf_{(N,d)} \mathbb{E} [N] = - \log \left( R_i (\Delta) \right) \frac{1}{D_i} = \log M - \log \left( P_{er} \right) \frac{1}{D_i}$$

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

where

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

- Main idea: Decode using $\Delta_a$
Main idea: Decode using $\Delta_a$

Recall we’ve defined

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

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

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

$$N_a = \min_{0 \leq i \leq M-1} N_i, \quad d_a = i^* \text{ if } N_a = N_{i^*}.$$
Lower Bound on $E(R)$

- **Main idea:** Decode using $\Delta_a$
- **Recall we’ve defined**

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

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

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

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

**Result:**

$$\mathbb{E}_0 [N_a] \lesssim \frac{- \log P_{er}}{E_{Forney}(R+\delta)} \Rightarrow E_a(R) \gtrsim E_{Forney}(R)$$
The random-coding error exponent of the stop-feedback communication setup with a binary symmetric forward channel is given by

\[ E(R) = \beta (\delta_{GV}(R) - \delta_{GV}(C)) = E_{sp}(R) + C - R. \]
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.
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
A *Multihypothesis test with control* $\Delta$ is a triplet $(N, q, d)$ where:

1. $N$ is the stopping time, $N \in \sigma(Y_1, ..., Y_n, U_1, ..., U_n)$.
2. $q = \{q_k(Y_k - 1, U_k - 1)\}_{k=1}^N$ is an observation control policy.
3. $d(Y_N, U_N)$ is the decision rule.

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

$$V(\pi) \equiv E[N + wI\{d(U_N, Y_N) \rightarrow \text{error}\}] = E[N] + w\text{Per}.$$

Asymptotic regime: $w \to \infty$. 

S. Ginzach (Technion)
A *Multihypothesis test with control* $\Delta$ is a triplet $(N, q, d)$ where:

1. $N$ is the *stopping time*, $N \in \mathcal{F}_n = \sigma (Y_1, \ldots, Y_n, U_1, \ldots, U_n)$. 

Assume $w > 0$ is the loss associated with making the wrong decision. The objective:

Find a sequential test $\Delta = (N, q, d)$ that minimizes the total cost defined as:

$$V(\pi) \triangleq E[N + wI\{d(U_N, Y_N) \rightarrow \text{error}\}] = E[N] + w\text{Per}.$$

Asymptotic regime: $w \to \infty$. 

S. Ginzach (Technion) 
December 25th, 2014
A Multihypothesis test with control $\Delta$ is a triplet $(N, q, d)$ where:

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

1. $N$ is the *stopping time*, $N \in F_n = \sigma(Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.
2. $q = \{q_k (Y^{k-1}, U^{k-1})\}_{k=1,\ldots,N}$ is an *observation control policy*.
3. $d = d(Y^N, U^N)$ is the *decision rule*. 

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

$$V(\pi) = E[N] + wP_{\text{error}}.$$
A *Multihypothesis test with control* $\Delta$ is a triplet $(N, q, d)$ where:

1. $N$ is the *stopping time*, $N \in \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.
2. $q = \{q_k(Y^{k-1}, U^{k-1})\}_{k=1,\ldots,N}$ is an *observation control policy*.
3. $d = d(Y^N, U^N)$ is the *decision rule*.

Assume $w > 0$ is the *loss* associated with making the wrong decision.
A *Multihypothesis test with control* $\Delta$ is a triplet $(N, q, d)$ where:

1. $N$ is the *stopping time*, $N \in \mathcal{F}_n = \sigma (Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.
2. $q = \{ q_k (Y^{k-1}, U^{k-1}) \}_{k=1,\ldots,N}$ is an *observation control policy*.
3. $d = d (Y^N, U^N)$ 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:

$$V (\pi) \triangleq \mathbb{E} [N + w \mathbb{I} \{d (U^N, Y^N) \rightarrow \text{error}\}] = \mathbb{E} [N] + w \text{Per}.$$
A Multihypothesis test with control $\Delta$ is a triplet $(N, q, d)$ where:

1. $N$ is the stopping time, $N \in \mathcal{F}_n = \sigma(Y_1, \ldots, Y_n, U_1, \ldots, U_n)$.

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:

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

Asymptotic regime: $w \to \infty$. 

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:

$$
\Phi_u(\pi, y) = \left( \pi_0 p_u^{0}(y) \pi_u, \pi_1 p_u^{1}(y) \pi_u, ..., \pi_{M-1} p_u^{M-1}(y) \pi_u \right), \forall u \in U,
$$

where $\pi \equiv (\pi_0, ..., \pi_{M-1})$.

Define the operator $T_u, u \in U$, such that for any measurable function $g: \Delta \rightarrow \mathbb{R}$:

$$
(T_u g)(\pi) = \int g(\Phi_u(\pi, y)) p_u^{\pi}(y) dy.
$$
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

$$
\Phi^u(\pi, y) = \left( \pi_0 \frac{p^u_0(y)}{p^u_\pi(y)}, \pi_1 \frac{p^u_1(y)}{p^u_\pi(y)}, \ldots, \pi_{M-1} \frac{p^u_{M-1}(y)}{p^u_\pi(y)} \right), \forall u \in U,
$$

where

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


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

$$
\Phi^u(\pi, y) = \left( \pi_0 \frac{p^u_0(y)}{p^\pi(y)}, \pi_1 \frac{p^u_1(y)}{p^\pi(y)}, \ldots, \pi_{M-1} \frac{p^u_{M-1}(y)}{p^\pi(y)} \right), \forall u \in U,
$$

where

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

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

$$
(\mathbb{T}^u g)(\pi) = \int g(\Phi^u(\pi, y)) p^u_\pi(y) dy.
$$
Lower Bounds on $V^*(\pi)$

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

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

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

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

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

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

Theorem (Javidi et al. 2013)

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

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

$$V^*(\pi) \geq \left[ \frac{H(\pi) - h_2(\delta) - \delta \log(M-1)}{I_{\text{max}}} \right] + \log \left( \frac{1-w^{-1}}{w^{-1}} \right) - \log \left( \frac{1-\delta}{\delta} \right) \frac{D_{\text{max}}}{I_{\text{max}}} \max_{i} \left\{ \pi_i \leq 1 - \delta \right\} - \tilde{K}'.$$
Information Acquisition Problem

Find a test 

\[ \Delta = (N, q, d) \]

with the object to minimize 

\[ E[N] \]

subject to 

\[ \text{Per} \leq \epsilon \]

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 

\[ E[N^* \epsilon] \]

be the minimal expected number of samples required to achieve 

\[ \text{Per} \leq \epsilon \]. Then 

\[ E[N^* \epsilon] \geq (1 - \epsilon w) (V^*(\pi) - 1) \]

where

\[ V^*(\pi) \]

is the optimal solution to Problem (P) for a prior \( \pi \) and cost for wrong decision \( w \).
Find a test $\Delta = (N, q, d)$ with the object to minimize $\mathbb{E}[N]$ subject to $P_{er} \leq \epsilon$.

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

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

where $V^{\star}(\pi)$ is the optimal solution to Problem (P) for a prior $\pi$ and cost for wrong decision $w$. 
Find a test $\Delta = (N, q, d)$ with the object to

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

This is the primal (constrained) version of the previous problem
Information Acquisition Problem

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

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

- 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 $\mathbb{E}[N^*_\epsilon]$ be the minimal expected number of samples required to achieve $P_{er} \leq \epsilon$. Then

$$\mathbb{E}[N^*_\epsilon] \geq (1 - \epsilon w)(V^*(\pi) - 1)$$

where $V^*(\pi)$ is the optimal solution to Problem (P) for a prior $\pi$ and cost for wrong decision $w$. 
More On Fictitious Agent
VL Coding and Controlled Hypothesis Testing

\[ \theta \xrightarrow{e_n(\theta)} X \in \mathcal{X}^\infty \xrightarrow{\text{Channel}} Y \in \mathcal{Y}^\infty \xrightarrow{\text{Decoder}} \hat{\theta} = d_N \]

Fictitious Agent

\[ Y_1, Y_2 \ldots, Y_{n-1} \]

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

Q: Is the VL coding problem amenable to hypothesis testing analysis?
A: Yes!
VL Coding and Controlled Hypothesis Testing, Cont.

\[ e_n(\theta) \xrightarrow{X \in \mathcal{X}^\infty} \text{Channel} \xrightarrow{Y \in \mathcal{Y}^\infty} \text{Decoder} \xrightarrow{\hat{\theta} = d_N} \]

- \( \pi = \left[ \frac{1}{M}, \frac{1}{M}, \ldots, \frac{1}{M} \right] \).
- \( p_i^u(k) = p(k \mid e(i)) \).
- \( D_{\text{max}} = \max_{i,k} \max_{u \in \mathcal{U}} D \left( p_i^u \parallel p_j^u \right) = \max_{j,k} D \left( p(\cdot \mid j) \parallel p(\cdot \mid k) \right) \triangleq C_1 \).
- \( I_{\text{max}} = \max_{u \in \mathcal{U}} \max_{\tilde{\pi} \in \Delta_M} I(\tilde{\pi}; p_{\tilde{\pi}}^u) = C \).
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.
For each message $i \in \{0, \ldots, M - 1\}$ randomly draw an infinite $P_X$-i.i.d. sequence. Let $x(i)$ be the codeword assigned to the $i$’th message.
For each message \( i \in \{0, \ldots, M - 1\} \) randomly draw an infinite \( P_X \)-i.i.d. sequence.

Let \( x^{(i)} \) be the codeword assigned to the \( i \)'th message.

For each \( i \in \{0, \ldots, M - 1\} \), define the following two hypotheses:

\[
H^i_0: \quad \Pr \left( x^{(i)}, y \right) = p \left( y \mid x^{(i)} \right) P_X \left( x^{(i)} \right),
\]

\[
H^i_1: \quad \Pr \left( x^{(i)}, y \right) = \Pr \left( y \right) P_X \left( x^{(i)} \right).
\]
For each message $i \in \{0, \ldots, M - 1\}$ randomly draw an infinite $P_X$-i.i.d. sequence.

Let $x^{(i)}$ be the codeword assigned to the $i$'th message.

For each $i \in \{0, \ldots, M - 1\}$, define the following two hypotheses:

$$H_0^i: \Pr (x^{(i)}, y) = p (y | x^{(i)}) P_X (x^{(i)}),$$

$$H_1^i: \Pr (x^{(i)}, y) = \Pr (y) P_X (x^{(i)}).$$

Define

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

$$= \inf_{n \geq 0} \left\{ \sum_{j=1}^{n} \log \left[ \frac{p (y_j | x_j^{(i)})}{\Pr (y_j)} \right] \geq (1 + \epsilon) \log (M) \right\}.$$
More On Phase II (Burnashev Achievability)
Let $m_1$ be the message chosen at Phase I.
Let $m_1$ be the message chosen at Phase I.

Define $j^*$ and $k^*$ by $D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1$
Achievability - Phase II (Confirmation)

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

Define $j^*$ and $k^*$ by $D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1$

Encoder: Sends $(j^*, j^*\ldots)$ if $m_1$ is correct and $(k^*, k^*\ldots)$ o.w.

Decoder: Runs an SPRT with

$$H_{ACK}: Y_i \sim p(\cdot | j^*),$$

$$H_{NACK}: Y_i \sim p(\cdot | k^*).$$
Let $m_1$ be the message chosen at Phase I.

Define $j^*$ and $k^*$ by $D(p(\cdot | j^*) \parallel p(\cdot | k^*)) = C_1$

Encoder: Sends $(j^*, j^* \ldots)$ if $m_1$ is correct and $(k^*, k^* \ldots)$ o.w.

Decoder: Runs an SPRT with

$$H_{ACK}: Y_i \sim p(\cdot | j^*),$$
$$H_{NACK}: Y_i \sim p(\cdot | k^*).$$

For large $M$,

$$\mathbb{E}[N_{II,k}] = \pi_A \mathbb{E}_{PA}[N_{II,k}] + \pi_N \mathbb{E}_{PN}[N_{II,k}] \approx \mathbb{E}_{PA}[N_{II,k}] \lesssim \frac{-\log(P_e)}{C_1}.$$ 

$$\Rightarrow \mathbb{E}[N] \approx \mathbb{E}[N_{I,1}] + \mathbb{E}[N_{II,1}] \lesssim \frac{\log(M)}{C} + \frac{-\log(P_e)}{C_1}.$$
Forney’s Error Exponent
Example - ARQ scheme:

Codebook: \( M \) randomly chosen codewords, each of length \( n \).

Encoding: Send the \( i \)th codeword periodically to transmit the \( i \)th message.

Decoding:

Partition \( R_n \) into \( M \) decision regions and one erasure area.

\[
R_0 \bigcup R_1 \bigcup R_2 \bigcup R_3 \cup \cdots \bigcup R_{M-1}
\]

If \( Y(n+1) \in M-1 \bigcup \bigcup_{i=0}^{M-1} R_i \), send the stopping bit and decode.

Else, wait for the next \( n \) symbols and repeat the process.
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 $i$th codeword periodically to transmit the $i$th message.
Example - ARQ scheme:

- **Codebook:** $M$ randomly chosen codewords, each of length $n$.
- **Encoding:** Send the $i$th codeword periodically to transmit the $i$th message.
- **Decoding:** Partition $R_n$ into $M$ decision regions and one erasure area.

If $Y(n+1) \in M - 1 \cup \bigcup_{i=0}^{M-1} R_i$ sent the stopping bit and decode. Else, wait for the next $n$ symbols and repeat the process.

S. Ginzach (Technion)
Example - ARQ scheme:

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

If $Y(n+1) \in M - 1 \cup \bigcup_{i=0}^{M-1} R_i$ sent the stopping bit and decode. Else, wait for the next $n$ symbols and repeat the process.
Example - ARQ scheme:

- **Codebook**: \( M \) randomly chosen codewords, each of length \( n \).
- **Encoding**: Send the \( i \)th codeword periodically to transmit the \( i \)th 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 $i$th codeword periodically to transmit the $i$th 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.
Let

\[ E_{1,k} = \{ \text{Not making the right decision on the } k\text{th round} \}, \]
\[ E_{2,k} = \{ \text{Making an undetected error on the } k\text{th round} \}. \]
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 \( \mathbb{P}(\text{errasure}) = \mathbb{P}(R_M) = \mathbb{P}(\mathcal{E}_1) - \mathbb{P}(\mathcal{E}_2) . \)
Let

\[ E_{1,k} = \{ \text{Not making the right decision on the } k\text{th round} \}, \]
\[ E_{2,k} = \{ \text{Making an undetected error on the } k\text{th round} \}. \]

Then \( P(\text{errasure}) = P(\mathcal{R}_M) = P(E_1) - P(E_2). \)

More results:

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

\[ R = \frac{\log (M)}{E[N]} = \frac{\log (M)}{n} (1 - P(\mathcal{R}_M)) = \tilde{R} (1 - P(\mathcal{R}_M)) \]

\[ P_{er} = \sum_{k=1}^{\infty} (P(\mathcal{R}_M))^{k-1} P(E_2) \]
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 \( P(\text{errasure}) = P(\mathcal{R}_M) = P(\mathcal{E}_1) - P(\mathcal{E}_2) \to 0. \)

- More results:

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

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

\[ P_{er} = \sum_{k=1}^{\infty} (P(\mathcal{R}_M))^{k-1} P(\mathcal{E}_2) \to P(\mathcal{E}_2). \]

\[ E_{\text{Forney}}(R) = -\frac{\log(P_{er})}{\mathbb{E}[N]} \to -\frac{1}{n} \log(P(\mathcal{E}_2)). \]
Forney’s Decision Regions (Forney 1968)

Define

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

\[ \mathcal{R}_M^* = \bigcap_{m=0}^{M-1} (\mathcal{R}_m^*)^c, \]
Define

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

and

\[ \mathcal{R}^*_M = \bigcap_{m=0}^{M-1} (\mathcal{R}^*_m)^c, \]

and \( e_i(R, T) \triangleq \limsup_{n \to \infty} \left[ -\frac{1}{n} \log(\mathbb{P}(\mathcal{E}_i)) \right]. \)
Forney’s Decision Regions (Forney 1968)

- Define
  \[
  R^*_m = \left\{ y \in Y^n : \frac{p(y | x_m)}{\sum_{m' \neq m} p(y | x_m)} \geq \exp(nT) \right\}, \quad m \in \{0, \ldots, M - 1\},
  \]
  \[
  R^*_M = \bigcap_{m=0}^{M-1} (R^*_m)^c,
  \]
  and \( e_i(R, T) \triangleq \limsup_{n \to \infty} \left[ -\frac{1}{n} \log(P(E_i)) \right] \).

Theorem - Forney’s error exponents for the BSC(\( \epsilon \)) (Somekh-Baruch, Merhav 2011)

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

- \( \Rightarrow \) If \( 0 < e_1(R, T) \to 0 \) then and \( E_{\text{Forney}}(R) = e_2(R, T) \approx T \)
Forney’s Achievable Error Exponent (Forney 1968)

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

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

$$\Leftrightarrow E_{\text{Forney}}(R) = T \approx \beta (\delta_{GV}(R) - \delta_{GV}(C)) = E_{\text{sp}}(R) + C - R.$$
Define $\delta_{GV}(R) = \{\delta : h_2(\delta) = \log(2) - R\}$.

Then,

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

$\Leftrightarrow E_{Forney}(R) = T \approx \beta (\delta_{GV}(R) - \delta_{GV}(C)) = E_{sp}(R) + C - R.$
Reliability Func. (Stop Feedback)
Lower Bound on $E(R)$

- Main idea: Decode using $\Delta_\alpha$
Main idea: Decode using $\Delta_a$

Recall we’ve defined

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

$$= \min_{n \geq 0} \left\{ \log \left[ \frac{P_i([z]_n)}{\sum_{j \neq i} P_j([z]_n)} \right] \geq a \right\}$$

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

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

$$N_a = \min_{0 \leq i \leq M-1} N_i, \quad d_a = i^* \text{ if } N_a = N_i^*.$$
It holds that $a \leq -\log P_{er}$.

$$E_0[N_a] = \sum_{n=0}^{\infty} P_0(N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0(N_0 \geq n).$$
It holds that $a \leq - \log P_{er}$.

$$E_0 [N_a] = \sum_{n=0}^{\infty} P_0 (N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0 (N_0 \geq n).$$

For arbitrarily small $\delta > 0$, take

$$\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log (2) - h_2 \left( \epsilon + \frac{a}{n^\beta} \right) - \delta \right\}.$$
It holds that \( a \leq -\log P_{er} \).

\[
\mathbb{E}_0[N_a] = \sum_{n=0}^{\infty} P_0(N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0(N_0 \geq n).
\]

For arbitrarily small \( \delta > 0 \), take
\[
\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log (2) - h_2 \left( \epsilon + \frac{a}{n\beta} \right) - \delta \right\}.
\]

Then:

1. \( \bar{n} \to \infty \) as \( M \to \infty \).
It holds that $a \leq -\log P_{er}$.

$E_0 [N_a] = \sum_{n=0}^{\infty} P_0 (N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0 (N_0 \geq n)$.

For arbitrarily small $\delta > 0$, take

$\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log (2) - h_2 \left( \epsilon + \frac{a}{n\beta} \right) - \delta \right\}$.

Then:

1. $\bar{n} \to \infty$ as $M \to \infty$.
2. $\bar{n} \leq \frac{a}{E_{\text{Forney}}(R+\delta)} \leq \frac{-\log P_{er}}{E_{\text{Forney}}(R+\delta)}$. 
It holds that \( a \leq -\log P_{er} \).

\[
E_0[N_a] = \sum_{n=0}^{\infty} P_0(N_0 \geq n) \leq \bar{n} + \sum_{n=\bar{n}+1}^{\infty} P_0(N_0 \geq n).
\]

For arbitrarily small \( \delta > 0 \), take
\[
\bar{n} \triangleq \max_{n \in \mathbb{N}} \left\{ \frac{\log M}{n} \geq \log(2) - h_2(\epsilon + \frac{a}{n\beta}) - \delta \right\}.
\]

Then:
1. \( \bar{n} \to \infty \) as \( M \to \infty \).
2. \( \bar{n} \leq \frac{a}{E_{\text{Forney}}(R+\delta)} \leq \frac{-\log P_{er}}{E_{\text{Forney}}(R+\delta)} \).
3. For any \( n_0 \geq \bar{n} + 1 \)

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

\( \Rightarrow \) Asymptotically, \( E_0[N_a] \lesssim \frac{-\log P_{er}}{E_{\text{Forney}}(R+\delta)} \Rightarrow E_a(R) \gtrsim E_{\text{Forney}}(R) \).
Define $\Lambda_i(N) \triangleq \log \left[ \frac{p(y|x^{(i)})}{\sum_{j=0, j \neq i}^{M-1} p(y|x^{(j)})} \right]$ and $\Omega_{i, \bar{n}} \triangleq \{d = i, N \leq \bar{n}\}$.
Upper Bound on $E(R)$ - Proof Outline

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

- On the one hand $P_{er}(\Delta) = \sum_{j=0,j\neq i}^{M-1} P_j(d = i)$.
Upper Bound on $E(R)$ - Proof Outline

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

- On the one hand $P_{er}(\Delta) = \sum_{j=0, j \neq i}^{M-1} P_j(d = i)$.

- On the other hand:

$$
\sum_{j=0, j \neq i}^{M-1} P_j(d = i) = \sum_{j=0, j \neq i}^{M-1} \sum_{z} \mathbb{I}\{z : d = i\} P_j(z)
$$

$$
= \mathbb{E}_i \left[ \mathbb{I}\{z : d = i\} \frac{\sum_{j=0, j \neq i}^{M-1} P_j(z)}{P_i(z)} \right]
$$

$$
\geq e^{-a} P_i \left( \Omega_{i,\bar{n}}, \sup_{n \leq \bar{n}} \{ \Lambda_i(n) < a \} \right).
$$

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

Using Markov ineq. we obtaine

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

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

- 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

$$P_i \left( \sup_{n \leq \bar{n}} \{ \Lambda_i(n) > a \} \right) \to 0.$$
Using Markov ineq. we obtain

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

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 \( E (R) \leq E_{\text{Forney}} (R) \).