Binary hypothesis testing with feedback

Y. Polyanskiy and S. Verdú

Princeton University

ITA-2011, UCSD

Introduction

Value of feedback:

- In coding for memoryless channels:
  1. Fixed blocklength: feedback does not improve reliability function (Dobrushin'62).
     Fine print: For symmetric channels and $R > R_c$.
  2. Variable length: feedback helps a lot (Burnashev'76).

Introduction

Value of feedback:

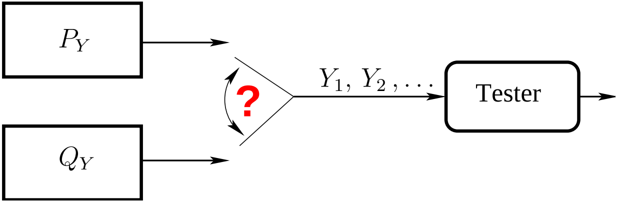
- In coding for memoryless channels:
  1. Fixed blocklength: feedback does not improve reliability function (Dobrushin'62).
     Fine print: For symmetric channels and $R > R_c$.
  2. Variable length: feedback helps a lot (Burnashev'76).
- In hypothesis testing (This talk!):
  1. Fixed sample size: feedback does not improve exponential tradeoff.
  2. Variable sample size: feedback helps a lot.

Disclaimer:

Everything in this talk is discrete, memoryless and asymptotic.

Fixed sample size

Classical binary hypothesis testing



Two types of errors:

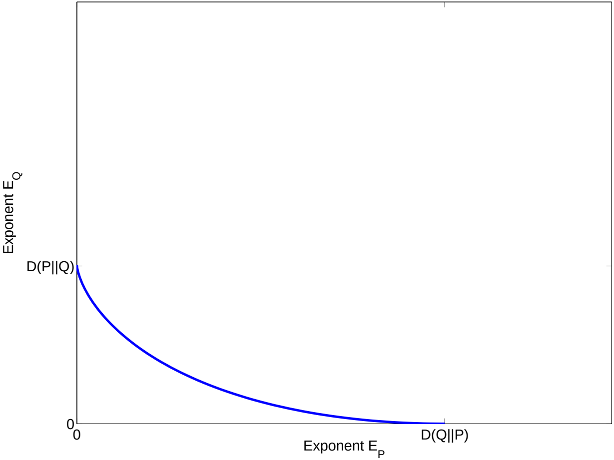
$$\begin{aligned} \bar{\alpha}_n &= \mathbb{P}[Y^n \notin E] \\ \beta_n &= \mathbb{Q}[Y^n \in E] \end{aligned}$$

Exponential performance:

$$\begin{aligned} E_P &= \lim_{n\to\infty} -\frac{1}{n} \log \bar{\alpha}_n \\ E_Q &= \lim_{n\to\infty} -\frac{1}{n} \log \beta_n \end{aligned}$$

Question: What is the achievable region of (E_P, E_Q)?

Exponential tradeoff (Hoeffding'65, Blahut'74)

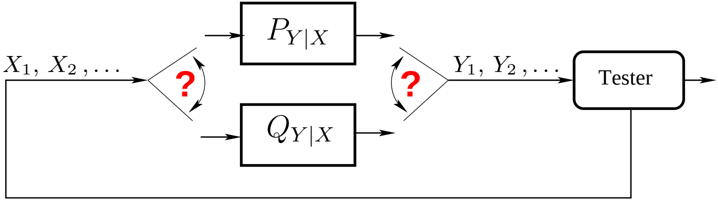


$$\mu(s) = \log \mathbb{E}_P\left[\left(\frac{P_Y(Y)}{Q_Y(Y)}\right)^s\right], \quad -1 \le s \le 0$$

$$E_P(s) = s\mu'(s) - \mu(s)$$

$$E_Q(s) = (s+1)\mu'(s) - \mu(s)$$

Hypothesis testing + control of inputs



Test:

1. A feedback controller:

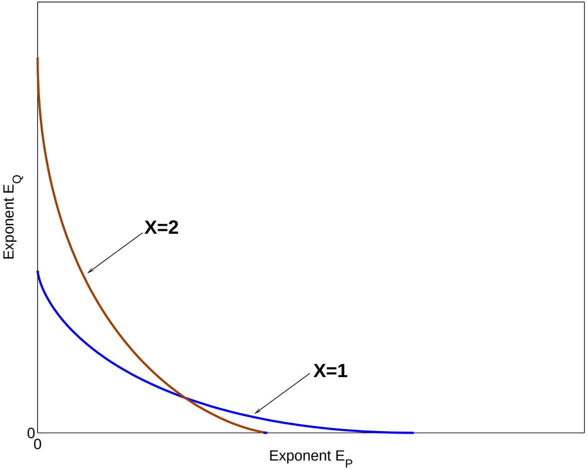
$$X_n = f_n(Y_1, \ldots, Y_{n-1}),$$

2. A critical region E:

$$\{Y^n \in E\} \Longrightarrow \text{declare } P_{Y|X}$$

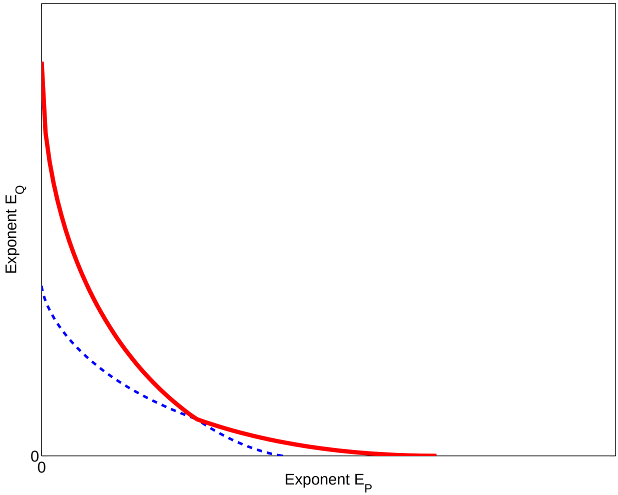
Question: What is the achievable region of (E_P, E_Q)?
Note: $|X| = 1 \iff$ classical setup P_Y vs Q_Y.

Simple strategies: fix $X_j = \mathrm{const}$



Thus, the best non-feedback tradeoff is ...

Optimal open-loop tradeoff



What about feedback?

Theorem

Feedback does not help (the red curve is still optimal).

What about feedback?

Theorem

Feedback does not help (the red curve is still optimal).

Proofs:

1. [Hayashi'09]: upper-bound on the growth of Rényi divergence

$$D_\lambda(P_{Y^n}||Q_{Y^n}) \le n \max_x D_\lambda(P_{Y|X=x}||Q_{Y|X=x})$$

   Note: slick, but requires $P_{Y|X=x} \sim Q_{Y|X=x}$.

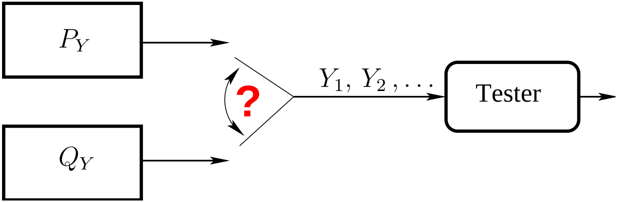
2. [PV'10]: Change measure via a tilted channel

$$V_{Y|X} = c P_{Y|X}^{1+s(X)} Q_{Y|X}^{-s(X)}$$

   and apply martingale arguments to $\log \frac{P_{Y|X}}{Q_{Y|X}}$.
   Note: does not require $P_{Y|X=x} \sim Q_{Y|X=x}$.

Variable sample size

Wald binary hypothesis testing



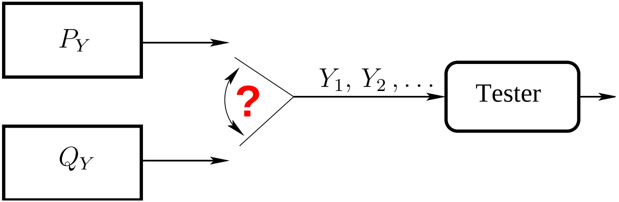
Variable sample size test:

1. A stopping time τ of the filtration $\mathcal{F}_n = \sigma\{Y_1, \ldots, Y_n\}$
2. A critical region $E \in \mathcal{F}_\tau$.

Figures of merit (non-Bayesian!):

$$\bar\alpha = \mathbb{P}[Y^\tau \notin E]$$
$$\beta = \mathbb{Q}[Y^\tau \in E]$$
$$\ell_P = \mathbb{E}_P[\tau]$$
$$\ell_Q = \mathbb{E}_Q[\tau]$$

Wald binary hypothesis testing



Variable sample size test:

1. A stopping time τ of the filtration $\mathcal{F}_n = \sigma\{Y_1, \ldots, Y_n\}$
2. A critical region $E \in \mathcal{F}_\tau$.

Figures of merit (non-Bayesian!):

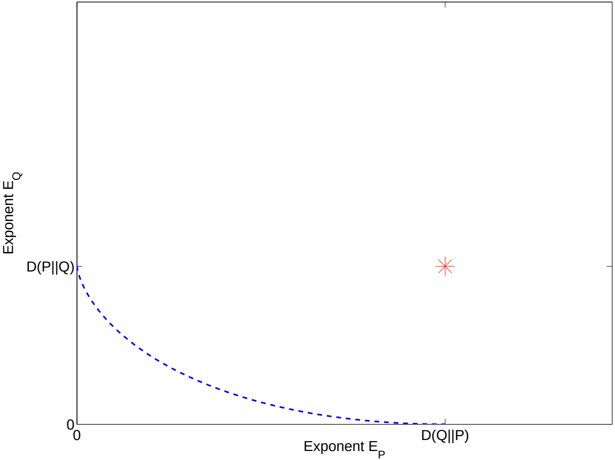
$$\begin{aligned} \bar\alpha &= \mathbb{P}[Y^\tau \notin E] \\ \beta &= \mathbb{Q}[Y^\tau \in E] \\ \ell_P &= \mathbb{E}_P[\tau] \\ \ell_Q &= \mathbb{E}_Q[\tau] \end{aligned} \implies \begin{cases} E_P = \lim_{\ell_P, \ell_Q \to \infty} -\frac{1}{\ell_P} \log \bar\alpha \\ E_Q = \lim_{\ell_P, \ell_Q \to \infty} -\frac{1}{\ell_Q} \log \beta \end{cases}$$

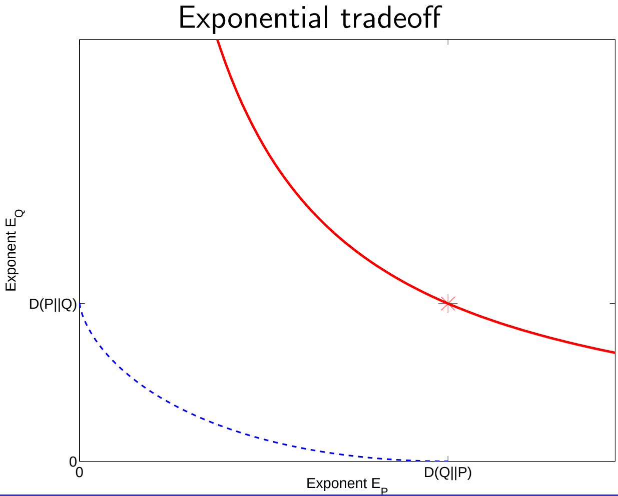
Question: What pairs (E_P, E_Q) are achievable?

Note: If $P_Y \not\ll Q_Y$ or $Q_Y \not\ll P_Y$ then any (E_P, E_Q) is achievable. Thus, assume

$$P_Y \sim Q_Y$$

Exponential tradeoff



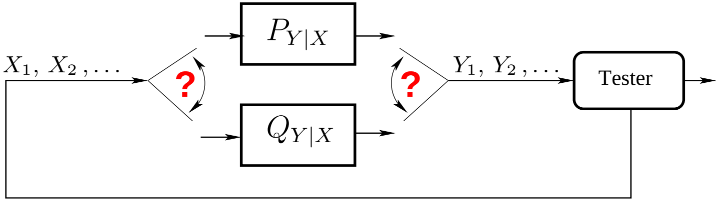


Theorem

The optimal region is $\{E_P E_Q \le D(P_Y||Q_Y)D(Q_Y||P_Y)\}$

Proof: Wald'45 + SPRT + Berk'73.

Hypothesis testing + control of inputs + variable size



Test:

1. A feedback controller:

$$X_n = f_n(Y_1, \ldots, Y_{n-1}),$$

2. A stopping time τ
3. A critical region $E \in \mathcal{F}_\tau$:

$$\{Y^\tau \in E\} \Longrightarrow \text{declare } P_{Y|X}$$

Some general remarks

Let

$$\mathcal{S} = \bigcup_{\text{all tests}} (\bar{\alpha}, \beta, \ell_P, \ell_Q) \subset \mathbb{R}_+^4$$

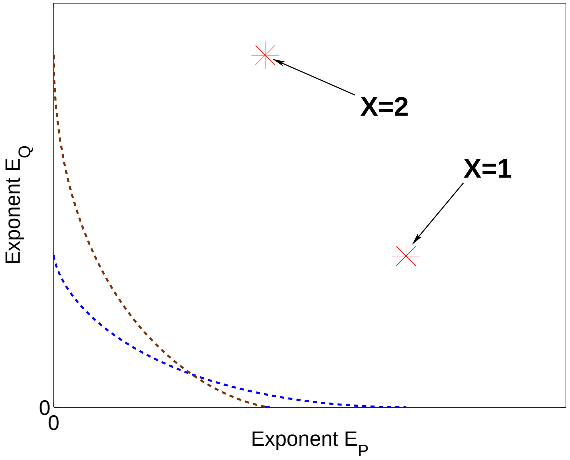
We can show

- $\mathcal{S}$ is convex but **not** closed.
- Every supporting hyperplane touches $\mathcal{S}$ at SPRT.
- … and every such point is a "sharp corner".
- projection of $\mathcal{S}$ on (ℓ_P, ℓ_Q) is $\mathbb{R}_+^2 \setminus \{\text{axes}\}$

For Bayesian setup dynamic programming states:

Optimal controller is a stationary function of $\log \frac{P_{Y^n|X^n}}{Q_{Y^n|X^n}}$.
The optimal test is SPRT.

Open-loop achievable (E_P, E_Q)



Question: What pairs (E_P, E_Q) are achievable?
Note: Without feedback we can easily achieve

$$\{E_P E_Q \le \max_x D(P_{Y|X=x}||Q_{Y|X=x}) D(Q_{Y|X=x}||P_{Y|X=x})\}$$

Question: What pairs (E_P, E_Q) are achievable?
Note: Without feedback we can easily achieve

$$\{E_P E_Q \le \max_x D(P_{Y|X=x}||Q_{Y|X=x})D(Q_{Y|X=x}||P_{Y|X=x})\}$$

Theorem

With feedback the optimal region is

$$\{E_P E_Q \le \max_{x_1} D(P_{Y|X=x_1}||Q_{Y|X=x_1}) \max_{x_2} D(Q_{Y|X=x_2}||P_{Y|X=x_2})\}$$

Proof (converse)

Denote maximal divergences:

$$d_P^* \triangleq \max_x D(P_{Y|X=x}||Q_{Y|X=x})$$
$$d_Q^* \triangleq \max_x D(Q_{Y|X=x}||P_{Y|X=x})$$

Wald's converse:

$$d(1-\bar{\alpha}||\beta) \leq \mathbb{E}_P[\tau]D(P_Y||Q_Y)$$
$$d(\beta||1-\bar{\alpha}) \leq \mathbb{E}_Q[\tau]D(Q_Y||P_Y)$$

Proof (converse)

Denote maximal divergences:

$$\begin{aligned} d_P^* &\stackrel{\triangle}{=} \max_x D(P_{Y|X=x}||Q_{Y|X=x}) \\ d_Q^* &\stackrel{\triangle}{=} \max_x D(Q_{Y|X=x}||P_{Y|X=x}) \end{aligned}$$

Simple generalization of Wald's converse:

$$\begin{aligned} d(1-\bar{\alpha}||\beta) &\le \mathbb{E}_P[\tau] \max_x D(P_{Y|X=x}||Q_{Y|X=x}) \\ d(\beta||1-\bar{\alpha}) &\le \mathbb{E}_Q[\tau] \max_x D(Q_{Y|X=x}||P_{Y|X=x}) \end{aligned}$$

Proof (converse)

Denote maximal divergences:

$$\begin{aligned} d_P^* &\triangleq \max_x D(P_{Y|X=x}||Q_{Y|X=x}) \\ d_Q^* &\triangleq \max_x D(Q_{Y|X=x}||P_{Y|X=x}) \end{aligned}$$

Simple generalization of Wald's converse:

$$d(1-\bar{\alpha}||\beta) \leq d_P^* \ell_P \tag{1}$$

$$d(\beta||1-\bar{\alpha}) \leq d_Q^* \ell_Q \tag{2}$$

Proof (converse)

Denote maximal divergences:

$$d_P^* \triangleq \max_x D(P_{Y|X=x}||Q_{Y|X=x})$$

$$d_Q^* \triangleq \max_x D(Q_{Y|X=x}||P_{Y|X=x})$$

Simple generalization of Wald's converse:

$$d(1-\bar{\alpha}||\beta) \le d_P^* \ell_P \qquad (1)$$

$$d(\beta||1-\bar{\alpha}) \le d_Q^* \ell_Q \qquad (2)$$

Asymptotically, for tests achieving (E_P, E_Q):

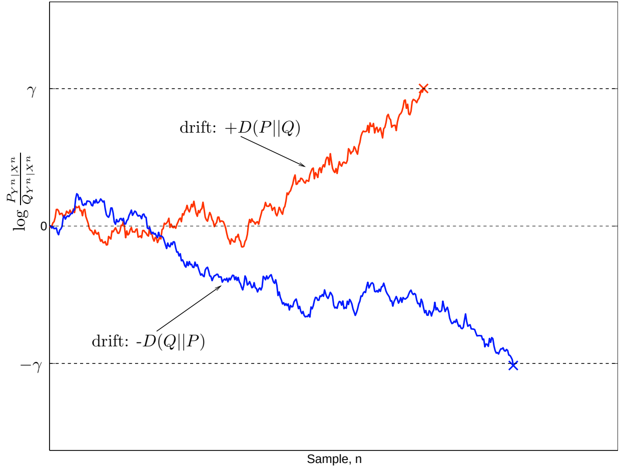
$$d(1-\bar{\alpha}||\beta) = \ell_Q E_Q + o(\ell_Q)$$

$$d(\beta||1-\bar{\alpha}) = \ell_P E_P + o(\ell_P)$$

$(1) \times (2) \Longrightarrow$

$$\ell_P \ell_Q E_P E_Q \le \ell_P \ell_Q d_P^* d_Q^* + o(\ell_P \ell_Q)$$

Proof (achievability)



Analysis of SPRT:

$$\bar{\alpha} \approx e^{-\gamma}$$

$$\beta \approx e^{-\gamma}$$

$$\ell_P \approx \frac{\gamma}{\text{drift under } P}$$

$$\ell_Q \approx \frac{\gamma}{\text{drift under } Q}$$

$$\downarrow$$

$$E_P \approx \text{drift under } P$$

$$E_Q \approx \text{drift under } Q$$

Problem: Choosing x to maximize $D(P_{Y|X=x}||Q_{Y|X=x})$ achieves

$$\begin{aligned} E_P &= d_P^* \\ E_Q &< d_Q^* \end{aligned}$$

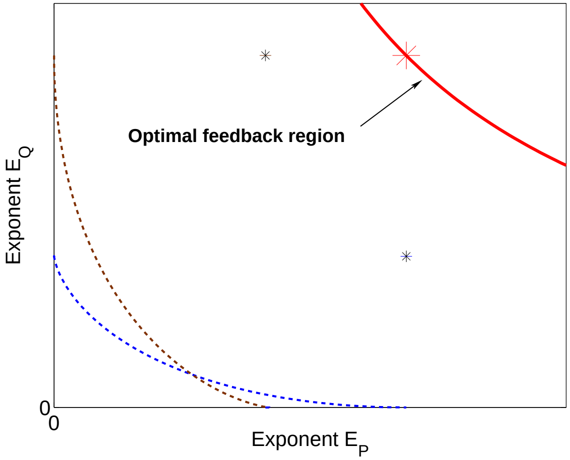
Solution: With feedback can optimize both drifts simultaneously!

$$X_{n+1} = \begin{cases} \mathrm{argmax}_x\, D(P_{Y|X=x}||Q_{Y|X=x}), & \log \frac{P_{Y^n|X^n}}{Q_{Y^n|X^n}} > 0 \\ \mathrm{argmax}_x\, D(Q_{Y|X=x}||P_{Y|X=x}), & \text{o/w} \end{cases}$$

Reason: log-likelihood is positive most of the time (under P).

Note: Same control strategy is asymptotically optimal in a certain Bayesian setting [Naghshvar-Javidi'10].

Optimal (E_P, E_Q) achievable with feedback



Open-loop controller + variable size test

Theorem

The open-loop region is strictly smaller unless for some x

$$\begin{aligned} D(P_{Y|X=x}||Q_{Y|X=x}) &= d_P^* \\ D(Q_{Y|X=x}||P_{Y|X=x}) &= d_Q^* \end{aligned}$$

(one input simultaneously maximizes both divergences).

Proof:

- WLOG use SPRT.
- Then stopping time concentrates at ℓ_P or ℓ_Q (Berk'73).
- For any input sequence a non-vanishing portion of X_j's should be suboptimal for either P or Q.
- Thus accumulated drift cannot be $\ell_P d_P^*$ and $\ell_Q d_Q^*$ simultaneously.

Thank You!