ECE531 Lecture 13: Sequential Detection of Discrete-Time Signals

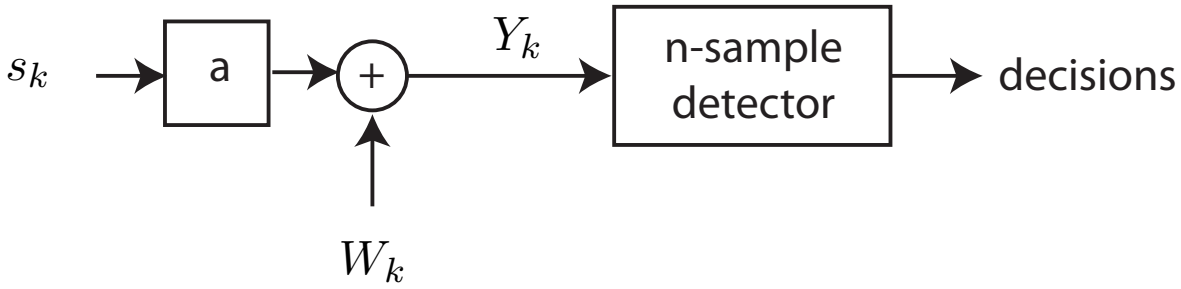
D. Richard Brown III

Worcester Polytechnic Institute

30-Apr-2009

Introduction

- All of the detection problems we have considered in this class have involved a **fixed sample size** n with the decision rule only being applied after all n samples are available.



- We now consider the **sequential detection** problem where we collect samples one at a time until we have enough observations to generate a final/terminal decision.
- There are two approaches to this problem:
  - Bayes: Assign a relative cost to taking an observation and try to minimize the overall average cost.
  - N-P: Keep taking observations until a desired tradeoff can be obtained between P_{fp} and P_D. Good decision rules achieve the desired tradeoff with fewer observations.

Mathematical Model for Sequential Detection

- We consider only simple binary hypotheses here: $\mathcal{H}_0 : X = x_0$ and $\mathcal{H}_1 : X = x_1$.
- We assume that the scalar observations are i.i.d. random variables with density $p_0(z)$ under $\mathcal{H}_0$ and $p_1(z)$ under $\mathcal{H}_1$.
- The set of possible observation sequences can be written as

$$\mathcal{Y} = \{y : \, y = \{y_k\}_{k=1}^{\infty}, y_k \in \mathbb{R}\}$$

- A sequential decision rule (SDR) is a pair of sequences

$$\begin{aligned} \phi &= \{\phi_k\}_{k=0}^{\infty} \text{ (stopping rule)} \\ \delta &= \{\delta_k\}_{k=0}^{\infty} \text{ (terminal decision rule)} \end{aligned}$$

where $\phi_k : \mathbb{R}^k \mapsto \{0, 1\}$ and $\delta_k : \mathbb{R}^k \mapsto \{0, 1\}$. Note that the SDR begins at index $k = 0$ whereas the first sample occurs at index $k = 1$.

Sequential Decision Rules

- SDR procedure:
  - If $\phi_n(y_1, \ldots, y_n) = 0$, we take another sample.
  - If $\phi_n(y_1, \ldots, y_n) = 1$, we stop sampling and make the terminal decision $\delta_n(y_1, \ldots, y_n)$.
- The **stopping time** is the random variable

$$N(\phi) \quad = \quad \min\{n : \phi_n(Y_1, \ldots, Y_n) = 1\}$$

- Under our assumption of simple binary hypotheses, the performance metrics of interest are the conditional risks $R_0(\delta)$ and $R_1(\delta)$ as well as the expected stopping times under each state $\mathrm{E}_0[N(\phi)] := \mathrm{E}[N(\phi) \,|\, X = x_0]$ and $\mathrm{E}_1[N(\phi)] := \mathrm{E}[N(\phi) \,|\, X = x_1]$.
- We would like to minimize all of these quantities but each can be traded off against the other.

Bayes Formulation for Sequential Detection

- We assume the UCA. If we assign a cost of c to each observation, the conditional risks can be written as

$$\begin{aligned} R_0(\phi,\delta) &= \mathrm{Prob}(\text{decide } \mathcal{H}_1 \,|\, X = x_0) + c\mathrm{E}_0[N(\phi)] \\ &= \mathrm{E}_0[\delta(Y_1,\ldots,Y_N)] + c\mathrm{E}_0[N(\phi)] \end{aligned}$$

and

$$\begin{aligned} R_1(\phi,\delta) &= \mathrm{Prob}(\text{decide } \mathcal{H}_0 \,|\, X = x_1) + c\mathrm{E}_1[N(\phi)] \\ &= 1 - \mathrm{E}_1[\delta(Y_1,\ldots,Y_N)] + c\mathrm{E}_1[N(\phi)]. \end{aligned}$$

- Given a prior $\pi_0 = \mathrm{Prob}(X = x_0)$, $\pi_1 = \mathrm{Prob}(X = x_1)$, we can combine the conditional risks into a single Bayes risk

$$r(\phi,\delta,\pi_1) = (1-\pi_1)R_0(\phi,\delta) + \pi_1 R_1(\phi,\delta).$$

- We seek a SDR (ϕ,δ) that minimizes $r(\phi,\delta,\pi_1)$.

Step 0: No Samples Taken Yet

We have three choices: decide 0, decide 1, or take a sample.

If $\phi_0 = 1$, we are choosing to make a decision without any observations. Is this a good idea? Actually, it might be if

- the cost of an observation is very high,
- the prior π_1 is very close to zero or one, or
- the observations are very noisy.

If we stop, the risks of each possible decision rule are

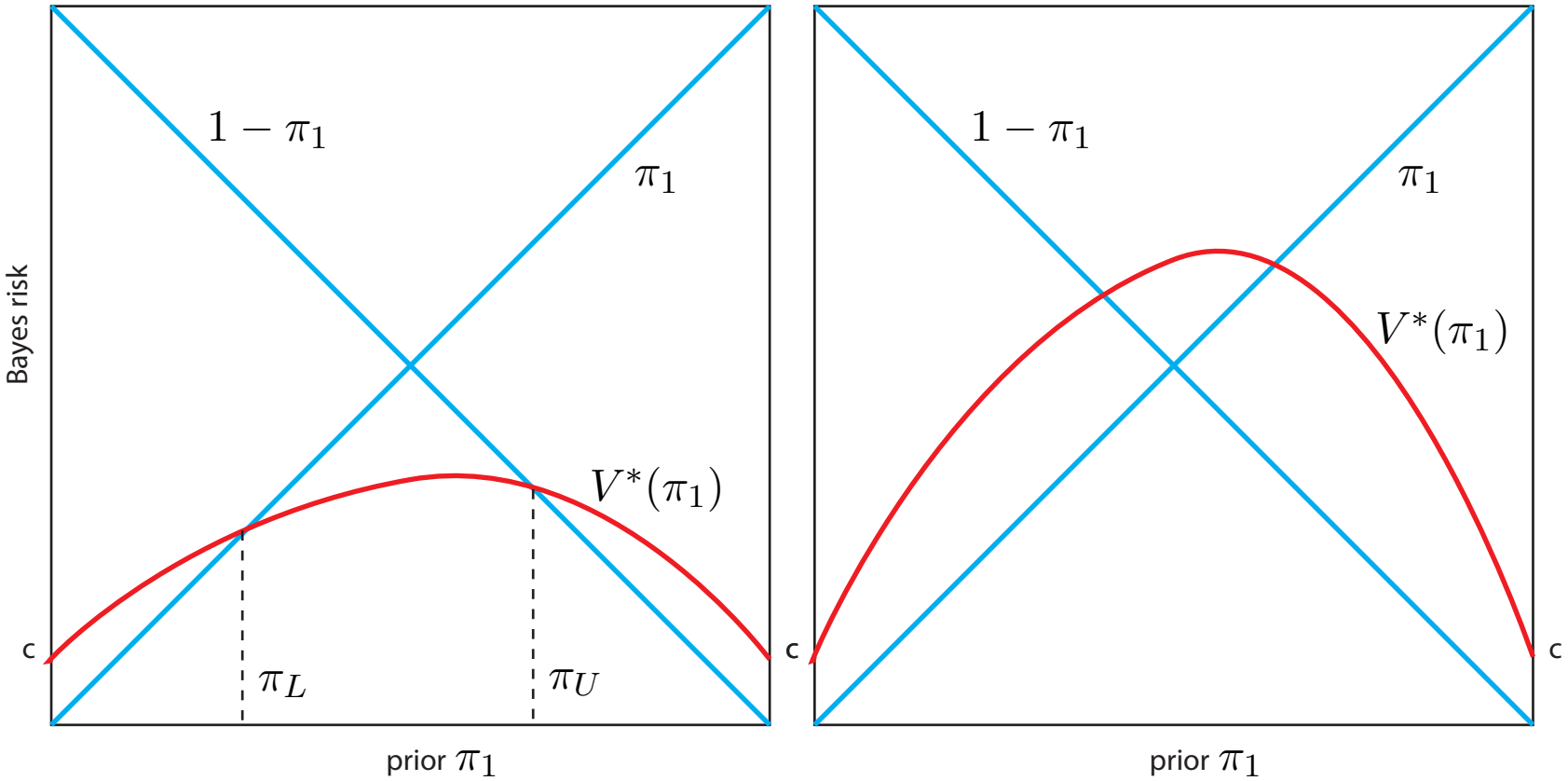
$$\begin{aligned} r(\phi_0 = 1, \delta_0 = 0, \pi_1) &= \pi_1 \\ r(\phi_0 = 1, \delta_0 = 1, \pi_1) &= 1 - \pi_1 \end{aligned}$$

If we choose to take a sample, the minimum Bayes risk will then be

$$V^*(\pi_1) = \min_{\substack{\phi : \phi_0 = 0 \\ \delta}} (1 - \pi_1) R_0(\phi, \delta) + \pi_1 R_1(\phi, \delta) \geq c$$

$V^*(\pi_1)$ is the minimum expected risk over all SDRs that take at least one sample.

Graphical Intuituion: Minimum Risk Function



Step 0: No Samples Taken Yet

It should be clear that our decision to take a sample at step 0 is determined by whether the quantity $V^*(\pi_1)$ is smaller than both π_1 and $1-\pi_1$.

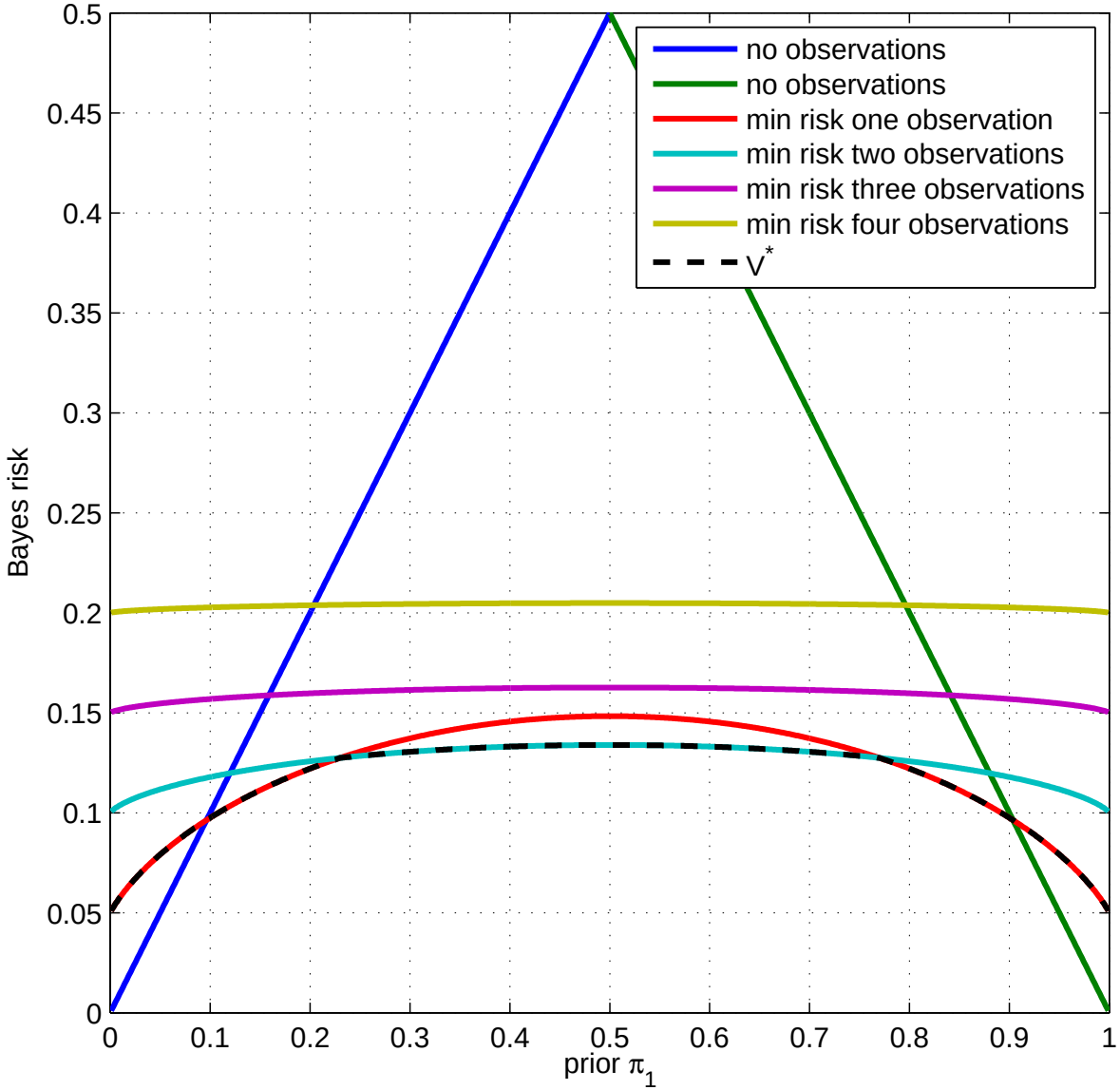
Lemma

When the costs $C_{00} = C_{11} = 0$, then $V^*(0) = V^*(1) = c$ and $V^*(x)$ is a continous concave function on $x \in [0,1]$.

- Let $\mathcal{P}^* = \{x \in [0,1] \,:\, V^*(x) \leq \min\{x, 1-x\}\}$.
- If $\mathcal{P}^*$ is empty, then we definitely should not take a sample. We should just decide 1 when $\pi_1 \geq 0.5$, otherwise we should decide 0.
- If $\mathcal{P}^*$ is not empty then $\mathcal{P}^* = [\pi_L, \pi_U] \subseteq [0,1]$ and the best action at step 0 is now clear:

$$\begin{aligned}
\pi_1 \geq \pi_U &: \text{ stop and decide 1 } (\phi_0 = 1, \delta_0 = 1) \\
\pi_1 \leq \pi_L &: \text{ stop and decide 0 } (\phi_0 = 1, \delta_0 = 0) \\
\pi_L < \pi_1 < \pi_U &: \text{ take a sample } (\phi_0 = 0, \delta_0 = ?)
\end{aligned}$$

Minimum Risk Function: Binary Signals in Gaussian Noise



Example: Binary Signaling

- The previous slide plots the risk for the HT problem:

$$\begin{aligned} \mathcal{H}_0 &: Y_k \sim \mathcal{N}(-1, \sigma^2) \\ \mathcal{H}_1 &: Y_k \sim \mathcal{N}(1, \sigma^2) \end{aligned}$$

  where $\sigma^2 = 0.6$ is known and $c = 0.05$.
- According to the minimum risk curves on the previous slide, if our prior was $\pi_1 = 0.5$, we can minimize the risk by taking two samples.
- So the optimum SDR should take two samples, right?
- What if our first sample turned out to be $y_1 = 10$? Should we take another sample?
- The sample $y_1 = 10$ is very strong evidence in favor of $\mathcal{H}_1$. Is another sample likely to lower our risk after observing $y_1 = 10$?
- A sequential decision rule does not decide in advance how many samples it should take. The decision to take another sample is made dynamically and depends on the outcomes of the previous samples.

Step 1: One Sample Taken: What Should We Do?

- Given no samples, we had a prior probability $\pi_1 = \mathrm{Prob}(X = x_1)$.
- If we decided to take one sample and observed $Y_1 = y_1$, we can compute a posterior probability

$$\begin{aligned}
\pi_1(y_1) &:= \mathrm{Prob}(X = x_1 \,|\, Y_1 = y_1) \\
&= \frac{\pi_1 p_1(y_1)}{\pi_0 p_0(y_1) + \pi_1 p_1(y_1)} \\
&= \frac{L(y_1)}{\frac{1-\pi_1}{\pi_1} + L(y_1)}
\end{aligned}$$

- We have three choices at this point: decide 0, decide 1, or take another sample.
- Since we've assumed i.i.d. samples and the cost of taking another sample is the same, the problem at step $n > 0$ is the same problem that we faced in step 0, the only difference being the **updated prior**.

Example: Coin Flipping

Suppose we have the following simple binary HT problem:

$$\begin{aligned} \mathcal{H}_0 &: \text{coin is fair} \\ \mathcal{H}_1 &: \text{coin is unfair with } \mathrm{Prob}(\mathsf{H}) = q > 0.5 \end{aligned}$$

Let $\pi_1 = \mathrm{Prob}(\text{state is } \mathcal{H}_1)$ and assume the UCA. If we have take no samples, the risks will be

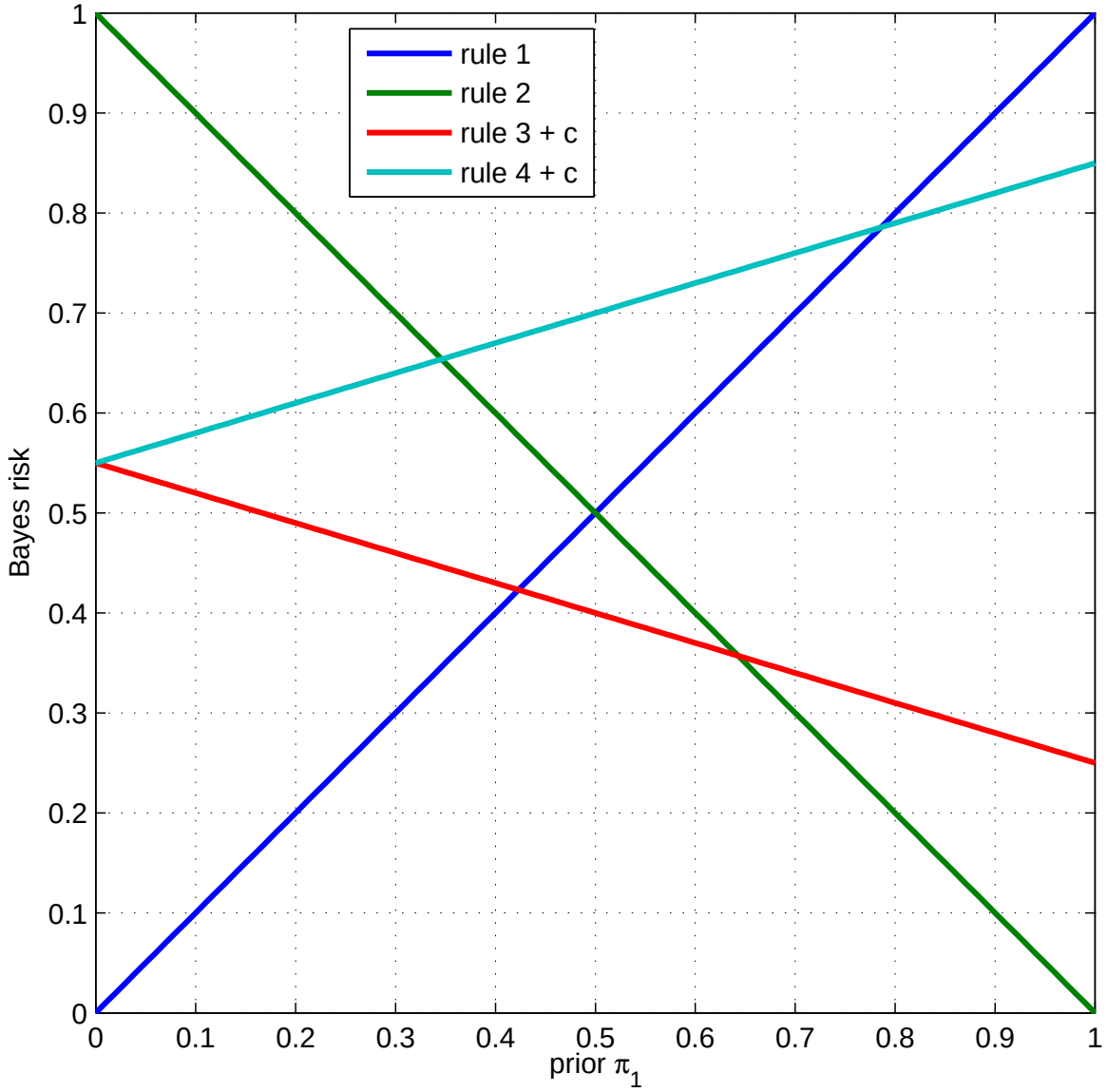
$$\begin{aligned} r(\delta_1, \pi) &= \pi_1 \text{ (always decide fair)} \\ r(\delta_2, \pi) &= 1 - \pi_1 \text{ (always decide unfair)} \end{aligned}$$

If we take one sample, the risks will be

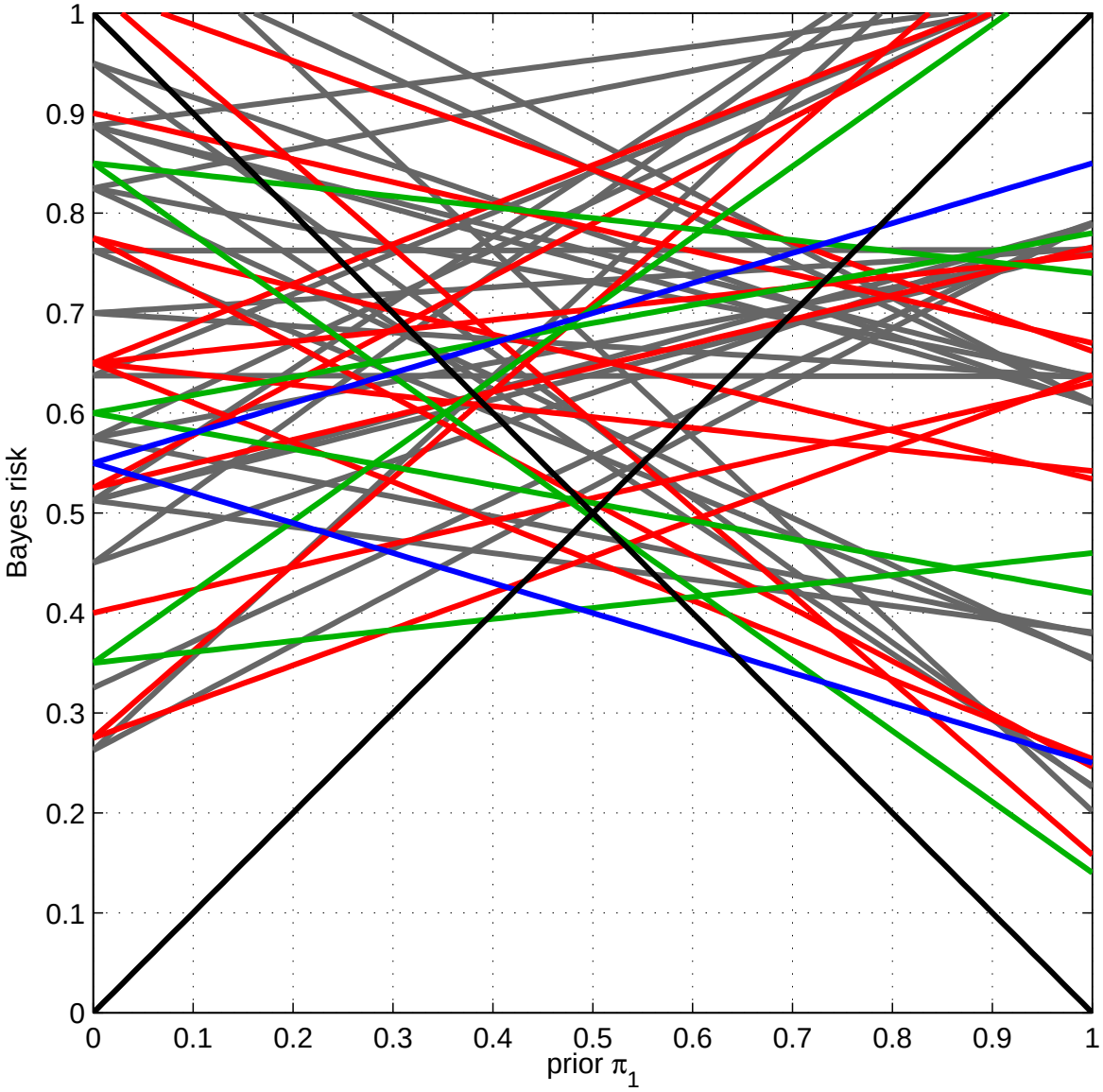
$$\begin{aligned} r(\delta_3, \pi) &= \frac{1-\pi_1}{2} + \pi_1(1-q) + c \text{ ($y_1 =$T, decide fair; $y_1 =$H, decide unfair)} \\ r(\delta_4, \pi) &= \frac{1-\pi_1}{2} + \pi_1 q + c \text{ ($y_1 =$T, decide unfair; $y_1 =$H, decide fair)} \end{aligned}$$

To find V^*, we have to keep doing this for 2, 3, 4, ... samples.

Example: Coin Flipping: $c = 0.05$ and $q = 0.8$



Example: Coin Flipping: $c = 0.05$ and $q = 0.8$



Example: Coin Flipping: $c = 0.05$

- Suppose we have a prior $\pi_1 = 0.5$. What should we do? Take a sample since taking one sample and using rule 3 has a lower Bayes risk at this prior (including the cost of the sample).
- The coin is flipped and you observe a head. What should you do?

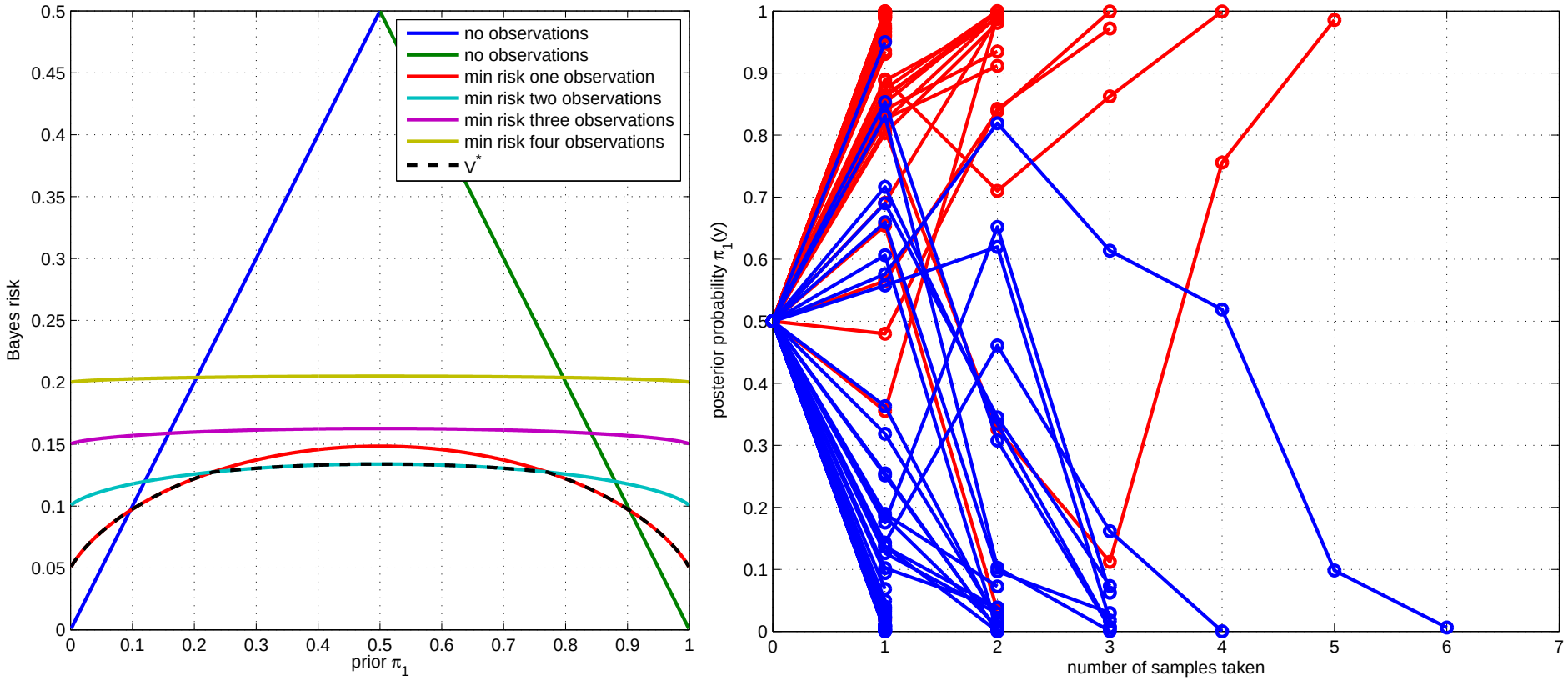
$$\begin{aligned}\pi_1(y_1 = \mathsf{H}) &= \frac{\pi_1 p_1(y_1 = \mathsf{H})}{\pi_0 p_0(y_1 = \mathsf{H}) + \pi_1 p_1(y_1 = \mathsf{H})} \\ &= \frac{0.5 \cdot 0.8}{0.5 \cdot 0.5 + 0.5 \cdot 0.8} \\ &\approx 0.6154\end{aligned}$$

- Now what should we do? It looks like we need to take another sample.
- The coin is flipped and you observe a tail. What should you do?

$$\begin{aligned}\pi_1(\{y_1, y_2\} = \{\mathsf{H},\mathsf{T}\}) &= \frac{\pi_1(y_1 = \mathsf{H}) p_1(y_2 = \mathsf{T})}{\pi_0(y_1 = \mathsf{H}) p_0(y_2 = \mathsf{T}) + \pi_1(y_1 = \mathsf{H}) p_1(y_2 = \mathsf{T})} \\ &\approx 0.3902\end{aligned}$$

- Now what should we do? It is close. Let's stop and decide $\mathcal{H}_0$.

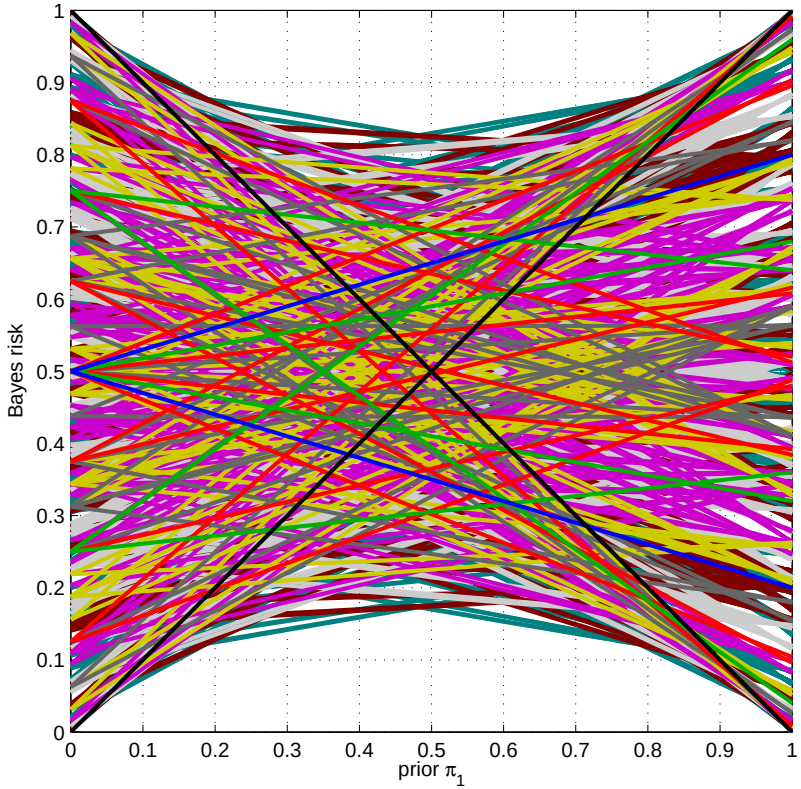
Example: Binary Signals in Gaussian Noise



$$a_1 = -a_0 = 1;\ \sigma^2 = 0.6;\ c = 0.05$$

Bayesian Formulation: Remarks

- When $c > 0$, we stop taking samples when the prior has been updated to a point where the expected minimum risk of taking one or more samples exceeds the risk of just making a decision now.
- If each sample were free, we could keep taking samples and make the minimum risk $V^*(\pi_1)$ go to zero over all priors, e.g. coin flipping:



Neyman-Pearson Formulation

- Samples no longer have an explicit cost.
- We have already investigated the case when the number of samples is fixed. We can design decision rules to achieve a two-way tradeoff between P_{fp} and P_D.
- Now, we would like to design a sequential decision rule that achieves a desired **three-way tradeoff** between P_{fp}, P_D, and the expected number of samples.

Sequential Probability Ratio Test

Given an n-sample observation $\{y_k\}_{k=1}^n$, and a prior π_1, the posterior can be written as

$$\begin{aligned}\pi_1(\{y_k\}_{k=1}^n) &= \frac{\pi_1 p_1(\{y_k\}_{k=1}^n)}{(1-\pi_1)p_0(\{y_k\}_{k=1}^n) + \pi_1 p_1(\{y_k\}_{k=1}^n)} \\ &= \frac{L(\{y_k\}_{k=1}^n)}{\frac{1-\pi_1}{\pi_1} + L(\{y_k\}_{k=1}^n)} = \frac{\lambda_n}{\frac{1-\pi_1}{\pi_1} + \lambda_n}\end{aligned}$$

where

$$\lambda_n := L(\{y_k\}_{k=1}^n) = \frac{p_1(\{y_k\}_{k=1}^n)}{p_0(\{y_k\}_{k=1}^n)} = \prod_{k=1}^n \frac{p_1(y_k)}{p_0(y_k)} = \prod_{k=1}^n L(y_k)$$

- $\pi_1(\{y_k\}_{k=1}^n)$ is a strictly monotone increasing function of λ_n when $0 < \pi_1 < 1$. Hence testing $\pi_1(\{y_k\}_{k=1}^n) \geq \pi_U$ and $\pi_1(\{y_k\}_{k=1}^n) \leq \pi_L$ is equivalent to testing $\lambda_n \geq B$ and $\lambda_n \leq A$.

Sequential Probability Ratio Test (SPRT)

A Sequential Probability Ratio Test, denoted as SPRT(A, B), is a sequential decision rule of the form

$$(\phi_n(\{y_k\}_{k=1}^n), \delta_n(\{y_k\}_{k=1}^n)) = \begin{cases} (1,1) & \lambda_n \geq B \text{ (stop and decide 1)} \\ (1,0) & \lambda_n \leq A \text{ (stop and decide 0)} \\ (0,?) & A < \lambda_n < B \text{ (take a sample)} \end{cases}$$

How are A and B related to π_L and π_U? Set $\pi_1(\{y_k\}_{k=1}^n)$ equal to π_L or π_U and solve for λ_n ...

$$\begin{aligned} A &= \frac{(1-\pi_1)\pi_L}{\pi_1(1-\pi_L)} \\ B &= \frac{(1-\pi_1)\pi_U}{\pi_1(1-\pi_U)} \end{aligned}$$

Wald-Wolfowitz Theorem

Theorem

Let (ϕ, δ) *be an SPRT with parameters* (A, B) *and let* (ϕ', δ') *be any sequential decision rule. If*

$$\begin{aligned} P_{fp}(\phi', \delta') &\leq P_{fp}(\phi, \delta) \text{ and} \\ P_D(\phi', \delta') &\geq P_D(\phi, \delta) \end{aligned}$$

then

$$\mathrm{E}_j[N(\phi')] \geq \mathrm{E}_j[N(\phi)], \; j = 0 \text{ and } 1.$$

The theorem says that no sequential decision rule can achieve a better false positive probability **and** a better detection probability than an SPRT without also requiring more samples, on average, than the SPRT.

Wald-Wolfowitz Theorem: SPRT vs. n-Sample Detector

- We can reformulate our standard n-sample detector as a sequential decision rule (ϕ', δ') with fixed observation size n:

$$\begin{aligned} \phi'_n(\{y_k\}_{k=1}^n) &= \{0, \ldots, 0, 1\} \text{ where the 1 occurs in the } n\text{th position} \\ \delta'_n(\{y_k\}_{k=1}^n) &= \{?, \ldots, ?, \delta(\{y_k\}_{k=1}^n)\} \text{ where the } \delta \text{ occurs in the } n\text{th position} \end{aligned}$$

- Suppose we design an n-sample N-P decision rule δ' with $P_{\text{fp}} = \alpha$ and $P_D = \beta$. What is $\mathrm{E}_j[N(\phi')]$ for $j = 0$ and 1?
- Now consider any sequential decision rule of the SPRT form that achieves $P_{\text{fp}} \geq \alpha$ and $P_D \leq \beta$. Then the Wald-Wolfowitz theorem says that $\mathrm{E}_j[N(\phi)] \leq n$ for $j = 0$ and 1.
- Hence, the expected number of samples required by an SPRT (under each hypothesis) is no larger than that of a fixed decision rule with the same performance.

How to Select A and B to Achieve the Desired Tradeoff

We define the sets

$$\begin{aligned}\mathcal{S}_n^1 &= \{y = \{y_k\}_{k=1}^{\infty} : \forall j = 1, \ldots, n-1, A < \lambda_j < B \text{ and } \lambda_n \geq B\} \\ \mathcal{S}_n^0 &= \{y = \{y_k\}_{k=1}^{\infty} : \forall j = 1, \ldots, n-1, A < \lambda_j < B \text{ and } \lambda_n \leq A\}\end{aligned}$$

Note that $\mathcal{S}_n^i$ is the set of all observation sequences on which SPRT(A, B) decides $\mathcal{H}_i$ after exactly n samples.

How to Select A and B to Achieve the Desired Tradeoff

The probability of false positive of SPRT(A, B) is then

$$\begin{aligned}
P_{\mathsf{fp}} &= \sum_{n=0}^{\infty} \mathrm{Prob}(\{y_k\}_{k=1}^n \in \mathcal{S}_n^1 \,|\, X = x_0) \\
&= \sum_{n=0}^{\infty} \int_{\mathcal{S}_n^1} \prod_{j=1}^{n} p_0(y_j)\, dy \\
&\leq B^{-1} \sum_{n=0}^{\infty} \int_{\mathcal{S}_n^1} \prod_{j=1}^{n} p_1(y_j)\, dy \\
&= B^{-1} \sum_{n=0}^{\infty} \mathrm{Prob}(\{y_k\}_{k=1}^n \in \mathcal{S}_n^1 \,|\, X = x_1) \\
&= B^{-1} P_D
\end{aligned}$$

where the inequality results from the fact that $\lambda_n = \frac{\prod_{j=1}^n p_1(y_j)}{\prod_{j=1}^n p_0(y_j)} \geq B$.

How to Select A and B to Achieve the Desired Tradeoff

The probability of missed detection (false negative) of SPRT(A, B) can also be written as

$$\begin{aligned}
P_{\mathsf{fn}} &= \sum_{n=0}^{\infty} \mathrm{Prob}(\{y_k\}_{k=1}^n \in \mathcal{S}_n^0 \,|\, X = x_1) \\
&= \sum_{n=0}^{\infty} \int_{\mathcal{S}_n^0} \prod_{j=1}^{n} p_1(y_j)\, dy \\
&\leq A \sum_{n=0}^{\infty} \int_{\mathcal{S}_n^0} \prod_{j=1}^{n} p_0(y_j)\, dy \\
&= A \sum_{n=0}^{\infty} \mathrm{Prob}(\{y_k\}_{k=1}^n \in \mathcal{S}_n^0 \,|\, X = x_0) \\
&= A(1 - P_{\mathsf{fp}})
\end{aligned}$$

where the inequality results from the fact that $\lambda_n = \frac{\prod_{j=1}^n p_1(y_j)}{\prod_{j=1}^n p_0(y_j)} \leq A$.

How to Select A and B to Achieve the Desired Tradeoff

Putting it all together, we can write two linear inequality constraints

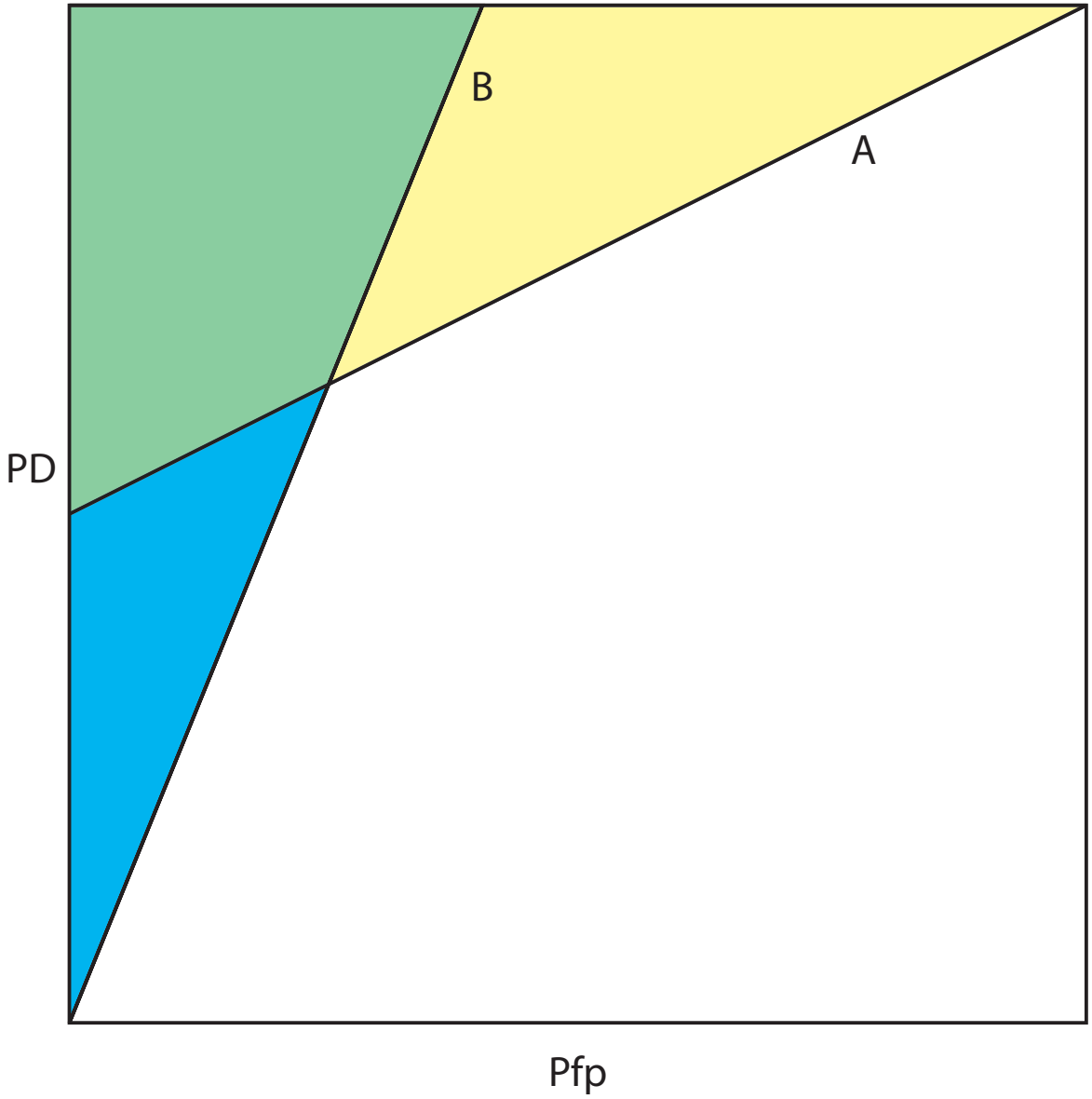
$$\begin{aligned} BP_{\text{fp}} + P_{\text{fn}} &\leq 1 \\ AP_{\text{fp}} + P_{\text{fn}} &\leq A \end{aligned}$$

or, in terms of $P_D = 1 - P_{\text{fn}}$,

$$\begin{aligned} BP_{\text{fp}} - P_D &\leq 0 \\ AP_{\text{fp}} - P_D &\leq A - 1 \end{aligned}$$

- These inequalities constrain P_{fp} and P_D in terms of our SPRT design parameters A and B.
- Note that we certainly want $B > A$ and typically $A < 1$ and $B > 1$.
- Inequalities are necessary here since λ_n may not exactly hit the boundary (A or B) but may jump over it. Under the assumption that each sample provides only a small amount of information, the amount of "overshooot", e.g. $\lambda_n = B + \epsilon$ or $\lambda_n = A - \epsilon$, will usually be small.

Linear Inequality Constraints



Solving for A and B in terms of P_{fp} and P_D

Our inequalities on A and B in terms of P_{fp} and P_D:

$$\begin{aligned} A &\geq \frac{1-P_D}{1-P_{\text{fp}}} \\ B &\leq \frac{P_D}{P_{\text{fp}}} \end{aligned}$$

Under the "small overshoot" assumption, we can assume that these inequalities are approximate equalities. Hence, if we require a N-P decision rule with $P_{\text{fp}} = \alpha$ and $P_D = \beta$, we should choose the SPRT parameters

$$A \approx \frac{1-\beta}{1-\alpha} \tag{1}$$

$$B \approx \frac{\beta}{\alpha} \tag{2}$$

Note that the lines $y = (1-A) + Ax$ and $y = Bx$ intersect at the point (α, β) when A and B are chosen according to (1) and (2).

Example: Sequential Detection of Unfair Coins

Suppose we have the following simple binary HT problem:

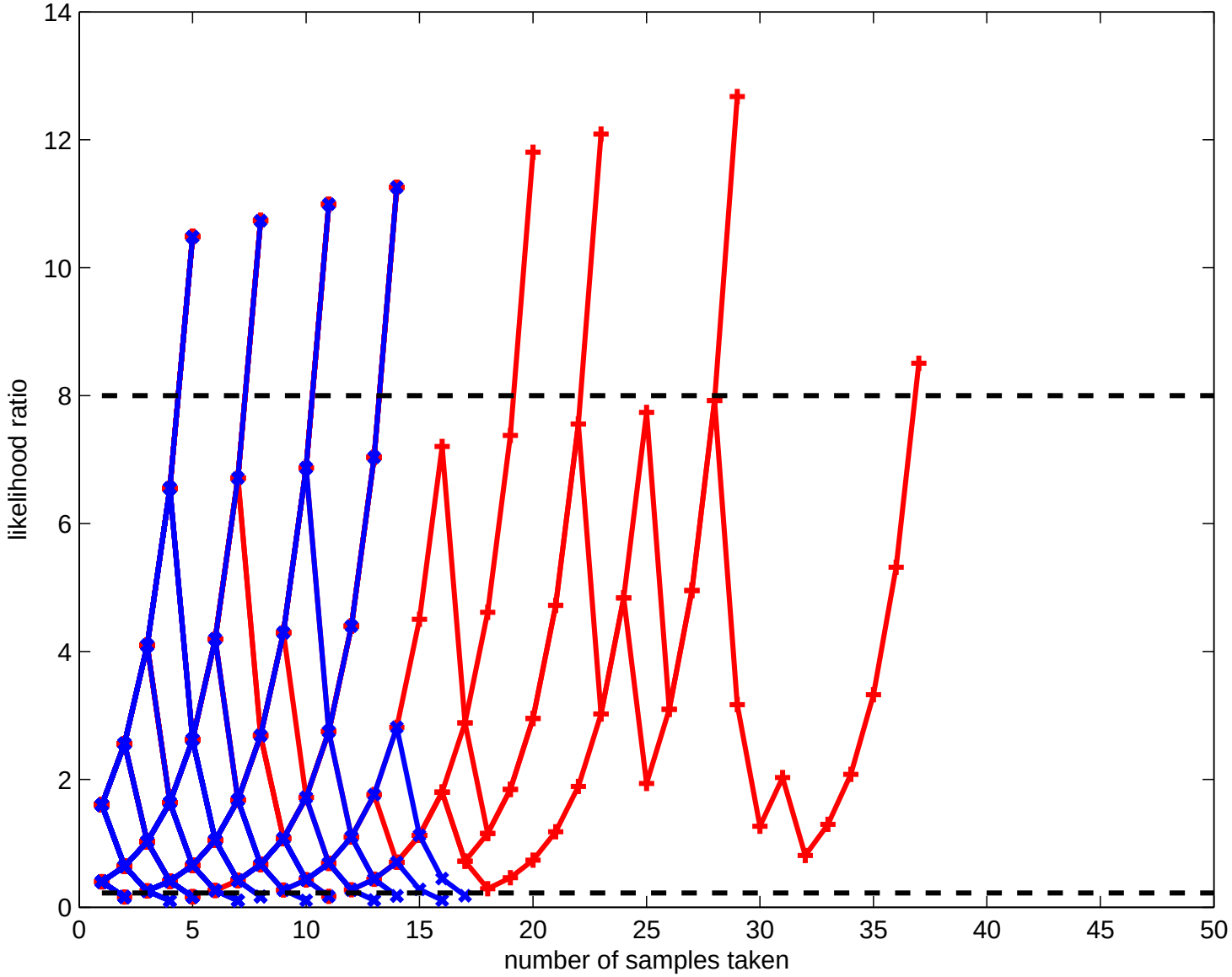
$$\begin{aligned} \mathcal{H}_0 &: \text{coin is fair} \\ \mathcal{H}_1 &: \text{coin is unfair with } \mathrm{Prob}(\mathsf{H}) = q > 0.5 \end{aligned}$$

Suppose we require $P_{\mathsf{fp}} = 0.1$ and $P_D = 0.8$. We can solve for the SPRT parameters

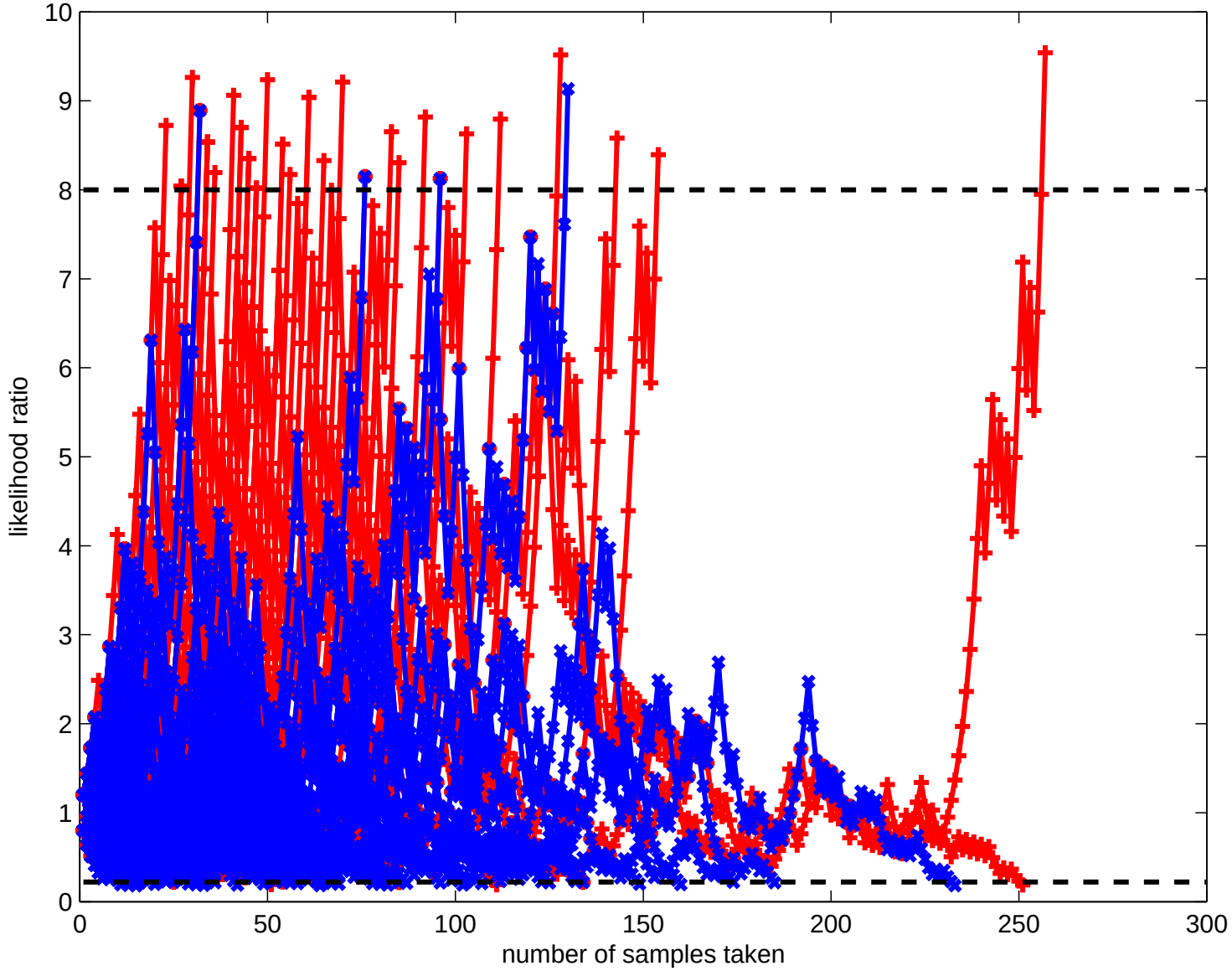
$$\begin{aligned} A &= \frac{2}{9} \\ B &= 8 \end{aligned}$$

Hence, our SPRT should take samples until $\lambda_n \geq 8$, in which case we decide $\mathcal{H}_1$, or until $\lambda_n \leq 2/9$, in which case we decide $\mathcal{H}_0$.
Note that $\lambda_n = p_1(\{y_j\}_{j=1}^n)/p_0(\{y_j\}_{j=1}^n) = q^k(1-q)^{n-k}/0.5^n$ where k is the number of heads and n is the number of samples/flips.

Example: Sequential Detection of Unfair Coins $q = 0.8$



Example: Sequential Detection of Unfair Coins $q = 0.6$



Final Remarks

- Two approaches to sequential detection:
  - Bayes: Assign a cost of c to each sample and a prior π to the states. Take samples until enough evidence is obtained such that the expected cost of additional samples is larger than the expected risk reduction.
  - N-P: Three-way tradeoff between P_D, P_{fp}, and the number of samples.
- Please read example III.D.2 for a comparison of SPRT and fixed-sample-size (FSS) detectors.
- Comments on page 110 of your textbook are also enlightening.
  - Even though the **average** number of samples for an SPRT is usually much less than a FSS detector, the **actual** number of samples for a particular sample sequence is unbounded!
  - Sequential decision rules have limited application when the samples are not i.i.d.