

Lecture 15: Chernoff bounds and Sequential detection

01 March 2016

1 Chernoff Bounds

1.1 Bayesian Hypothesis Test

A test using log-likelihood ratio statistic has the form,

$$T(Y) = \log L(Y) \underset{<}{\overset{\geq}{\cong}} \tau. \quad (1)$$

Bound-1: The probability of error P_e is bounded as,

$$P_e \leq (\pi_0 + \pi_1 e^\tau) e^{\mu_{T,0}(s_0) - s_0 \tau}, \quad (2)$$

where $\mu_{T,0}(s) = \log \mathbb{E}_0[e^{sT(y)}]$, and $\mu'_{T,0}(s_0) = \tau$.

Bound-2: $\forall s \in [0, 1]$,

$$P_e \leq \max(\pi_0, \pi_1 e^\tau) e^{\mu_{T,0}(s) - s\tau}. \quad (3)$$

Derivation of the above bound: Consider,

$$\begin{aligned} P_e &= \pi_0 P_0(\Gamma_1) + \pi_1 P_1(\Gamma_0), \\ &= \pi_0 \int_{\Gamma_1} dP_0(y) + \pi_1 \int_{\Gamma_0} dP_1(y). \end{aligned} \quad (4)$$

Now we have,

$$\begin{aligned} \Gamma_1 &= \{y : T(y) > \tau\} = \{y : \log L(y) > \tau\}, \\ &= \{y : \exp(s \log L(y)) > \exp(s\tau)\}, \quad \forall s \geq 0. \end{aligned} \quad (5)$$

Therefore we have,

$$\int_{\Gamma_1} dP_0(y) \leq \int_{\Gamma_1} \exp(s \log L(y) - s\tau) dP_0(y), \quad (6)$$

$$= e^{-s\tau} \int_{\Gamma_1} L(y)^s dP_0(y). \quad (7)$$

Similarly for $\int_{\Gamma_0} dP_1(y)$, we have,

$$\int_{\Gamma_0} dP_1(y) \leq e^{(1-s)\tau} \int_{\Gamma_0} L(y)^s dP_0(y). \quad (8)$$

Therefore, we get,

$$\begin{aligned} P_e &\leq \pi_0 e^{-s\tau} \int_{\Gamma_1} L(y)^s dP_0(y) + \pi_1 e^{(1-s)\tau} \int_{\Gamma_0} L(y)^s dP_0(y), \quad (9) \\ &\leq \max(\pi_0, \pi_1 e^\tau) e^{-s\tau} \underbrace{\left[\int_{\Gamma_1} L(y)^s dP_0(y) + \int_{\Gamma_0} L(y)^s dP_0(y) \right]}_{\mathbb{E}_0[\exp(s \log L(y))]}, \end{aligned}$$

here $\mathbb{E}_0[\exp(s \log L(y))] = \exp(\mu_{T,0}(s))$. Therefore,

$$P_e \leq \max(\pi_0, \pi_1 e^\tau) \exp(\mu_{T,0}(s) - s\tau). \quad (10)$$

In particular, consider the minimum probability of error detector,

$$\tau = \log \left(\frac{\pi_0}{\pi_1} \right) \quad (11)$$

For this detector, bound-2 gives, $\forall s \in [0, 1]$,

$$P_e \leq \pi_0^{1-s} \pi_1^s \exp(\mu_{T,0}(s)). \quad (12)$$

1.2 Chernoff bound for Quadratic Detector:

$$H_0 : \underline{Y} = \underline{N}$$

versus

$$H_1 : \underline{Y} = \underline{S} + \underline{N}$$

where $\underline{N} \sim \mathcal{N}(0, \sigma^2 I)$, $\underline{S} \sim \mathcal{N}(0, \Sigma_s)$ and \underline{S} is independent of \underline{N} .

Optimal detector: Let,

$$\Sigma_s = \sum_{k=1}^n \lambda_k \underline{v}_k \underline{v}_k^T, \quad (13)$$

where \underline{v}_k is an orthonormal set.

Test Statistic: $\underline{y}^T Q \underline{y} = \sum_{k=1}^n \bar{y}_k^2$, where $Q = (\Sigma_0^{-1} - \Sigma_1^{-1})$, and,

$$\bar{y}_k = \sqrt{\frac{\lambda_k}{\sigma^2(\sigma^2 + \lambda_k)}} \underline{v}_k^T \underline{y}. \quad (14)$$

$\{\bar{y}_k\}_{k=1}^n$ are independent and zero mean Gaussian random variables with variances $\{\sigma_{j,k}\}_{k=1}^n$,

$$\{\sigma_{j,k}\}_{k=1}^n = \begin{cases} \frac{\lambda_k}{\sigma^2 + \lambda_k} & \text{if } j = 0, \\ \frac{\lambda_k}{\sigma^2} & \text{if } j = 1. \end{cases} \quad (15)$$

Therefore,

$$\begin{aligned} L(\underline{y}) &= \prod_{k=1}^n \frac{\frac{1}{\sigma_{1k}\sqrt{2\pi}} \exp\left(-\frac{\bar{y}_k^2}{2\sigma_{1k}^2}\right)}{\frac{1}{\sigma_{0k}\sqrt{2\pi}} \exp\left(-\frac{\bar{y}_k^2}{2\sigma_{0k}^2}\right)}, \\ &= \prod_{k=1}^n \left(\frac{\sigma_{0k}}{\sigma_{1k}}\right) \exp\left(-\frac{\bar{y}_k^2}{2} \left(\frac{1}{\sigma_{1k}^2} - \frac{1}{\sigma_{0k}^2}\right)\right), \\ &= \prod_{k=1}^n \left(\frac{\sigma_{0k}}{\sigma_{1k}}\right) \exp\left(\frac{\bar{y}_k^2}{2}\right). \end{aligned} \quad (16)$$

with $T = \log L$,

$$\begin{aligned} \mu_{T,0}(s) &= \log \mathbb{E}_0 [\exp(sL)], \\ &= \log \mathbb{E}_0 \left[\prod_{k=1}^n \left(\frac{\sigma_{0k}}{\sigma_{1k}}\right)^s \exp\left(\frac{s\bar{y}_k^2}{2}\right) \right], \\ &= \sum_{k=1}^n s \log\left(\frac{\sigma_{0k}}{\sigma_{1k}}\right) + \sum_{k=1}^n \log \mathbb{E}_0 \left[\exp\left(\frac{s\bar{y}_k^2}{2}\right) \right]. \end{aligned} \quad (17)$$

here, $Y_k^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2\sigma_{0k}^2}\right)$, under H_0 .

$$\mathbb{E}_0 \left[\exp\left(\frac{s\bar{y}_k^2}{2}\right) \right] = \begin{cases} (1 - s\sigma_{0k}^2)^{-\frac{1}{2}} & \text{if } s < \frac{1}{\sigma_{0k}^2}, \\ \infty & \text{if } s \geq \frac{1}{\sigma_{0k}^2}. \end{cases} \quad (18)$$

So, the probability of error is given by,

$$P_e \leq \pi_0^{1-s} \pi_1^s \exp \left\{ \sum_{k=1}^n s \log \left(\frac{\sigma_{0k}}{\sigma_{1k}} \right) + \sum_{k=1}^n (1 - s \sigma_{0k}^2)^{-\frac{1}{2}} \right\} \quad \forall s \in [0, 1]. \quad (19)$$

Minimizing RHS over $s \in [0, 1]$, i.e., $\frac{d}{ds}[RHS] = 0$, we get,

$$\sum_{k=1}^n \frac{\lambda_k}{2(\sigma^2 + (1 - s_0)\lambda_k)} = \sum_{k=1}^n \log \left(1 + \frac{\lambda_k}{\sigma^2} \right) + \log \left(\frac{\pi_1}{\pi_0} \right). \quad (20)$$

2 Sequential Detection:

All the detectors discussed are of fixed size, i.e., number of samples is fixed and we optimized the performance of the detectors. Now we fix the desired performance and vary the number of samples to achieve the performance. The detector which uses random number of samples based on the observation sequence is called “sequential detector”. The observations $\{Y_k, k = 1, 2, \dots\}$ are i.i.d distributed according to

$$\begin{aligned} H_0 : Y_k &\sim P_0, \quad k = 1, 2, \dots \\ \textit{versus} \\ H_1 : Y_k &\sim P_1, \quad k = 1, 2, \dots \end{aligned}$$

where $P_0 \neq P_1$ are distributions on \mathbb{R} .

Definition 2.1. A sequential decision rule is a pair of sequences $(\underline{\phi}, \underline{\delta})$ where

$$\begin{aligned} \underline{\phi} &= \{\phi_0, \phi_1, \phi_2, \dots\}, \\ \underline{\delta} &= \{\delta_0, \delta_1, \delta_2, \dots\}, \end{aligned}$$

where $\underline{\phi}$ is called stopping rule ($\phi_j : \mathbb{R}^j \rightarrow \{0, 1\}, \forall j \geq 0$) and $\underline{\delta}$ is called terminal decision rule ($\delta_j : \mathbb{R}^j \rightarrow \{0, 1\}, \forall j \geq 0$).

$\underline{\phi}$ tells us when to stop sampling (if $\phi_n(y_1, y_2, \dots, y_n) = 0$, we take one more sample and if $\phi_n(y_1, y_2, \dots, y_n) = 1$, we stop sampling and take decision) and $\underline{\delta}$ tells the decision to be made when we stop sampling. The protocol for sequential detection is shown in Algo. 1.

Define $\phi_0 \in \{0, 1\}, \delta_0 \in \{0, 1\}$. It follows that the sequential decision rule $(\underline{\phi}, \underline{\delta})$ outputs decision $\delta_N(Y_1, Y_2, \dots, Y_N)$, where $N = \min\{n \geq 0 : \phi_n(Y_1, Y_2, \dots, Y_N) = 1\}$ is also a random variable.

Algorithm 1 Protocol for sequential detection

```
1: for  $k = 0, 1, 2, \dots$  do
2:   if  $\phi_k(Y_1, Y_2, \dots, Y_k) = 1$  then,
3:     output decision  $\delta_k(Y_1, Y_2, \dots, Y_k)$ 
4:   end if
5: end for
```

Example 2.2. Fixed sample decision rule for m samples: $\forall j \geq 0$,

$$\phi_j(y_1, y_2, \dots, y_j) = \begin{cases} 0, & \text{if } j \neq m, \\ 1, & \text{if } j = m, \end{cases} \quad (21)$$

$$\text{and } \delta_j(y_1, y_2, \dots, y_j) = \begin{cases} \delta(y_1, y_2, \dots, y_m), & \text{if } j = m, \\ \text{arbitrary}, & \text{otherwise.} \end{cases} \quad (22)$$

2.1 Bayesian Sequential Detection:

The priors π_0 and $\pi_1 = (1 - \pi_0)$ are assigned to the hypothesis H_0 and H_1 respectively. Assign a cost $C > 0$ for every observed sample (assume uniform costs for simplicity). Then conditional risks for a given sequential decision rule are

$$R_0(\underline{\phi}, \underline{\delta}) = \mathbb{E}_0[\delta_N(Y_1, Y_2, \dots, Y_N) + NC], \quad (23)$$

$$R_1(\underline{\phi}, \underline{\delta}) = \mathbb{E}_1[1 - \delta_N(Y_1, Y_2, \dots, Y_N) + NC], \quad (24)$$

and Bayes risk is given by

$$r(\underline{\phi}, \underline{\delta}) = (1 - \pi_1)R_0(\underline{\phi}, \underline{\delta}) + \pi_1R_1(\underline{\phi}, \underline{\delta}). \quad (25)$$

Definition 2.3. (Optimal) Bayesian Sequential decision rule: It is a rule $(\underline{\phi}, \underline{\delta})$ that minimizes $r(\underline{\phi}, \underline{\delta})$ over all sequential rules.

To observe the structure of the optimum decision rule, define the function

$$V^*(\pi_1) = \min_{\substack{(\underline{\phi}, \underline{\delta}) \\ \phi_0=0}} r(\underline{\phi}, \underline{\delta}), \quad \forall \pi_1 \in [0, 1] \quad (\text{optimum risk function}). \quad (26)$$

Here $\phi_0 = 0$ implies that the test does not stop with zero observations. $V^*(\pi_1)$ gives the minimum Bayes risk over all the sequential tests with at least one sample. The function $V^*(\pi_1)$ is a concave (because it is minimum of linear functions), continuous function of π_1 and $V^*(0) = V^*(1) = C$. Fig. 1 also shows the variation of Bayes risk with π_1 for two other sequential rules: one which decides H_1 without taking any samples ($\phi_0 = 1 = \delta_0$) and the other which decides H_0 without taking any samples ($\phi_0 = 1 = 1 - \delta_0$). The abscissa of the intersection of the latter two plots with the plot of $V^*(\pi_1)$ give π_L and π_U respectively.

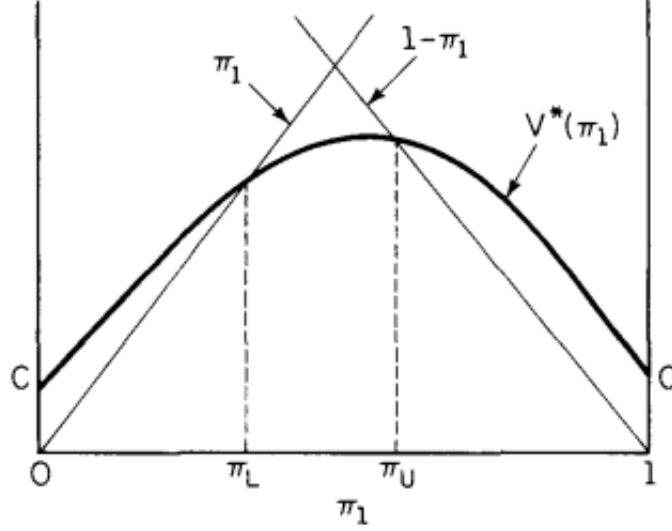


Figure 1: Relationships yielding the Bayes sequential rule for uniform costs of errors and cost C per sample.

Rules: It follows that

1. if $\pi_1 \leq \pi_L$, then the optimum sequential rule is $\phi_0 = 1 = 1 - \delta_0$
2. if $\pi_1 \geq \pi_U$, then the optimum sequential rule is $\phi_0 = 1 = \delta_0$
3. if $\pi_L < \pi_1 < \pi_U$, then the optimum sequential rule consumes $N \geq 1$ samples.

After taking one sample y_1 , it is as if our prior on $\{H_0, H_1\}$ has changed to a posterior distribution $\pi_1(y_1) = \mathbb{P}[H_1 \text{ is true} | Y_1 = y_1]$. Therefore, we can recursively apply the original rules (The shape of V^* is not affected by having the knowledge of Y_1 as the samples are independent).

Optimum Bayesian sequential decision rule: The test continues sampling until $\pi_1(y_1, \dots, y_n) = \mathbb{P}[H_1 \text{ is true} | Y_1 = y_1, \dots, Y_n = y_n]$ goes out of the interval (π_L, π_U) and chooses H_0 or H_1 based on the value of $\pi_1(y_1, \dots, y_n)$. The stopping rule and the terminal decision rule are obtained as

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0, & \text{if } \pi_1(y_1, \dots, y_n) \in [\pi_L, \pi_U], \\ 1, & \text{if } \pi_1(y_1, \dots, y_n) \notin [\pi_L, \pi_U], \end{cases} \quad (27)$$

$$\text{and } \delta_n(y_1, \dots, y_n) = \begin{cases} 1, & \text{if } \pi_1(y_1, \dots, y_n) \geq \pi_U, \\ 0, & \text{if } \pi_1(y_1, \dots, y_n) \leq \pi_L, \\ \text{arbitrary}, & \text{otherwise,} \end{cases} \quad (28)$$

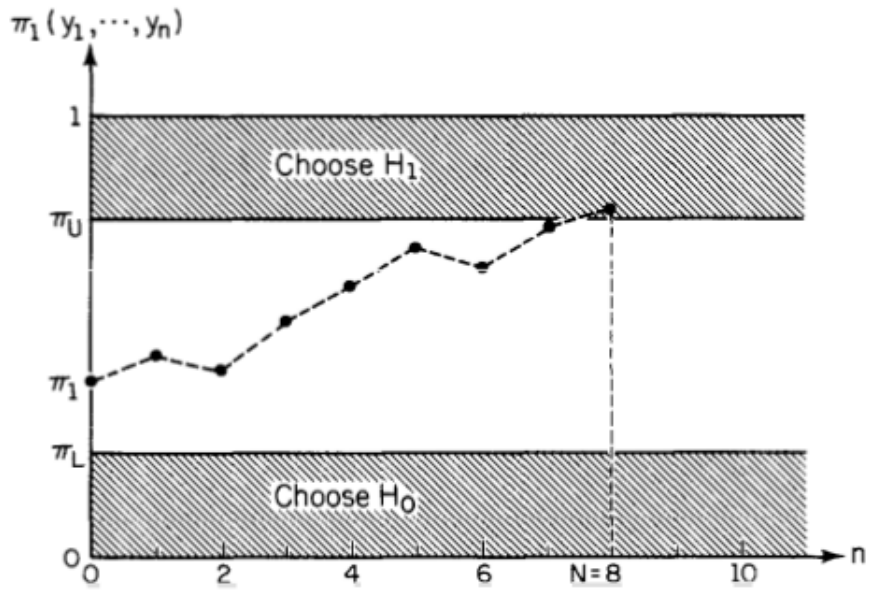


Figure 2: Depiction of realization of a Bayes sequential test

where $\pi_1(y_1, \dots, y_n) = \mathbb{P}[H_1 | Y_1 = y_1, \dots, Y_n = y_n]$.

The test terminates (under mild conditions) as $\pi_1(y_1, \dots, y_n)$ converges to 1 under H_1 or 0 under H_0 almost surely.