

Lecture 16

Sequential Detection

March 03, 2016

1 Introduction

In the last lecture, we introduced the idea of sequential detection. A sequential detector uses a random number of samples based on the observation sequence to achieve a desired level of performance. This is in contrast with fixed-sample-size detectors, which optimize the performance given a fixed number of samples.

Recall that a sequential decision rule is the pair $(\underline{\phi}, \underline{\delta})$, with

$$\begin{aligned} \phi_j(Y_1, \dots, Y_j) &\in \{0, 1\}, \\ \text{and } \delta_j(Y_1, \dots, Y_j) &\in \{H_0, H_1\}. \end{aligned} \quad (1)$$

To find the optimal *Bayesian* sequential decision rule, we assigned priors π_1 and $\pi_0 = 1 - \pi_1$ to hypotheses H_1 and H_0 , respectively. The optimal rule can be summarized as:

$$\begin{aligned} \pi_1 &\leq \pi_L, \text{ stop and declare } H_0, \\ \pi_1 &\geq \pi_U, \text{ stop and declare } H_1, \\ \pi_1 &\in (\pi_L, \pi_U), \text{ take one sample,} \\ \text{Update } \pi_j &: \pi_j(y) = \mathbb{P}(H_j \text{ is true} / Y_1 = y_1), \text{ and recurse.} \end{aligned}$$

The corresponding stopping and decision rules are,

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0, & \text{if } \pi_1(y_1, \dots, y_n) \in (\pi_L, \pi_U), \\ 1, & \text{otherwise.} \end{cases} \quad (2)$$

$$\delta_n(y_1, \dots, y_n) = \begin{cases} 1, & \text{if } \pi_1(y_1, \dots, y_n) \geq \pi_U, \\ 0, & \pi_1(y_1, \dots, y_n) \leq \pi_L. \end{cases} \quad (3)$$

Note that it is generally not easy to compute π_L and π_U . Implementing the optimal test, however, is easy. We do so by first rewriting the optimal Bayes sequential test in terms of the likelihood ratio. Later, we will generalize this idea to what are called

Sequential Probability Ratio Tests. The posterior probability can be computed using Bayes' rule (assuming \mathbb{P}_0 and \mathbb{P}_1 have densities p_0 and p_1 , respectively):

$$\begin{aligned}\pi_1(y_1, \dots, y_n) &= \mathbb{P}(H_1/Y_1 = y_1, \dots, Y_n = y_n) \\ &= \frac{\mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n/H_1)\mathbb{P}(H_1)}{\sum_{i=0,1} \mathbb{P}(Y_1 = y_1, \dots, Y_n = y_n/H_i)\mathbb{P}(H_i)} \\ &= \frac{\pi_1 \prod_{k=1}^n p_1(y_k)}{\sum_{i=0,1} \pi_i \prod_{k=1}^n p_i(y_k)} \\ &= \frac{\pi_1 L_n(y_1, \dots, y_n)}{\pi_1 L_n(y_1, \dots, y_n) + \pi_0}.\end{aligned}$$

We can thus rewrite the optimal Bayes sequential test as

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0, & \text{if } L_n(y_1, \dots, y_n) \in (\underline{\pi}, \bar{\pi}), \\ 1, & \text{otherwise,} \end{cases} \quad (4)$$

and

$$\delta_n(y_1, \dots, y_n) = \begin{cases} 1, & \text{if } L_n(y_1, \dots, y_n) \geq \underline{\pi}, \\ 0, & \text{if } L_n(y_1, \dots, y_n) \leq \bar{\pi}, \end{cases} \quad (5)$$

where $\underline{\pi} = \frac{\pi_0 \pi_L}{\pi_1(1-\pi_L)}$, and $\bar{\pi} = \frac{\pi_0 \pi_U}{\pi_1(1-\pi_U)}$.

2 Sequential Probability Ratio Test

In general, one can define a family of sequential tests called *Sequential Probability Ratio Tests* (also called SPRT, Wald's SPRT), with boundaries a and b ($a \leq 1 \leq b$). In particular, for any two real numbers a and b satisfying $0 < a \leq 1 \leq b < \infty$, the SPRT with boundaries a and b (denoted by $\text{SPRT}(a, b)$), is defined as,

$$\phi_n(y_1, \dots, y_n) = \begin{cases} 0, & \text{if } a < \lambda(y_1, \dots, y_n) < b \\ 1, & \text{otherwise} \end{cases} \quad (6)$$

and,

$$\delta_n(y_1, \dots, y_n) = \begin{cases} 1, & \text{if } \lambda(y_1, \dots, y_n) \geq a \\ 0, & \text{if } \lambda(y_1, \dots, y_n) \leq b. \end{cases} \quad (7)$$

Note that the rule is left arbitrary if $a = b$. Thus the $\text{SPRT}(a, b)$ continues sampling until the likelihood ratio λ_n falls outside the "boundaries" a and b , and then chooses H_1 if $\lambda_n \geq b$ and H_0 if $\lambda_n \leq a$.

Remark 1. By convention, the likelihood ratio with no samples is taken to be one, i.e., $L_0 \equiv 1$.

Remark 2. If $a = 1 < b$, then take no sample and output H_0 .

Remark 3. If $a < b = 1$, then take no sample and output H_1 .

Remark 4. If $a = b = 1$, then take no sample and choose H_0 or H_1 arbitrarily.

Example 2.1. Sequential detection of a constant signal We now consider the problem of detecting a constant signal θ in additive zero-mean Gaussian noise. We have

$$\begin{aligned} H_0 : Y_k &= N_k, \quad k = 1, 2, 3, \dots, \\ H_1 : Y_k &= N_k + \theta, \quad k = 1, 2, 3, \dots, \end{aligned}$$

where $\theta \in \mathbb{R}_+$ is a fixed number and $\{N_k\}_{k=1}^\infty$ is an i.i.d. sequence of noise samples distributed $\mathcal{N}(0, \sigma^2)$.

We can calculate the likelihood ratio for this case as follows,

$$\begin{aligned} L_n(\underline{y}) &= \frac{p_1(\underline{y})}{p_0(\underline{y})} \\ &= \prod_{k=1}^n \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_k - \theta)^2}{2\sigma^2}}}{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y_k^2}{2\sigma^2}}} \\ &= \exp \left\{ \sum_{k=1}^n \frac{\theta(y_k - \frac{\theta}{2})}{\sigma^2} \right\}. \end{aligned}$$

Therefore, $\text{SPRT}(a, b)$ would suggest sampling until

$$\sum_{k=1}^n \frac{\theta(y_k - \frac{\theta}{2})}{\sigma^2} \notin (\log a, \log b),$$

and declare,

$$H_0 : \text{if } \log L_n \leq \log a,$$

$$H_1 : \text{if } \log L_n \geq \log b.$$

In addition to the optimality of $\text{SPRT}(\underline{\pi}, \bar{\pi})$ in the Bayesian setting, $\text{SPRT}(a, b)$ also has an optimality property similar to that in the Neyman-Pearson type setting. To this end, let us first define some notations.

Let

$$\mathbb{P}_F(\phi, \delta) = \mathbb{P}_0[\delta_N(Y_1, \dots, Y_N) = 1], \quad (8)$$

$$\mathbb{P}_M(\phi, \delta) = \mathbb{P}_1[\delta_N(Y_1, \dots, Y_N) = 0], \quad (9)$$

where,

$$\begin{aligned} N \equiv N(\underline{\phi}) &= \text{Random time at which we stop.} \\ &= \min\{n \geq 0, \phi_n(Y_1, \dots, Y_n) = 1\}. \end{aligned}$$

$N(\underline{\phi})$ is also called the sample number of $\underline{\phi}$. We then have the following result.

Theorem 2.1. (Wald-Wolfowitz Theorem '98): Suppose that $(\underline{\phi}^*, \underline{\delta}^*)$ represent the SPRT(a, b) and $(\underline{\phi}, \underline{\delta})$ is any other sequential test for which,

$$\begin{aligned} \mathbb{P}_F(\underline{\phi}, \underline{\delta}) &\leq \mathbb{P}_F(\underline{\phi}^*, \underline{\delta}^*), \\ \text{and } \mathbb{P}_M(\underline{\phi}, \underline{\delta}) &\leq \mathbb{P}_M(\underline{\phi}^*, \underline{\delta}^*) \end{aligned}$$

Then, $\mathbb{E}_j[N(\underline{\phi})] \geq \mathbb{E}_j[N(\underline{\phi}^*)] \forall j \in \{0, 1\}$.

Thus we can see from the above result that for a given level of performance, no sequential decision rule has a smaller expected sample size than the SPRT with that performance. It should be noted here that a fixed sample size detector (no randomness in sample size as in the theorem) is a special case of the sequential decision rule. Hence the theorem asserts that the *average* sample size of an SPRT is no larger than the sample size of a fixed sample size detector with the same performance. The theorem also implies that given expected sample sizes, no sequential decision rule has smaller error probabilities than the SPRT.

As is clear from the theorem, SPRTs turn out to be optimal in terms of $\mathbb{P}_F, \mathbb{P}_M$ and the number of samples if optimized altogether. Hence the natural question to ask is how to set a and b in an SPRT to achieve a desired performance (*i.e.*, $\mathbb{P}_F, \mathbb{P}_M$). This question is the topic of interest for the next section.

2.1 Wald's Approximation

Suppose $(\underline{\phi}, \underline{\delta}) \equiv SPRT(a, b)$, $a < 1 < b$ that achieves,

$$\alpha = \mathbb{P}_F(\underline{\phi}, \underline{\delta}) \tag{10}$$

$$\gamma = 1 - \beta = \mathbb{P}_M(\underline{\phi}, \underline{\delta}). \tag{11}$$

The rejection region of $(\underline{\phi}, \underline{\delta})$ is by definition,

$$\begin{aligned} \Gamma_1 &= \{y \in \mathbb{R}^\infty | L_N(y_1, \dots, y_N) \geq b\}, \\ &= \bigcup_{n=1}^{\infty} \{y \in \mathbb{R}^\infty | N = n, L_n(y_1, \dots, y_n) \geq b\}. \end{aligned} \tag{12}$$

Let $\{Q_n\} \triangleq \{y \in \mathbb{R}^\infty | N = n, L_n(y_1, \dots, y_n) \geq b\}$. Clearly, $\{Q_n\}$ are mutually exclusive. We thus have,

$$\begin{aligned} \alpha &= \mathbb{P}_0(\Gamma_1) \\ &= \sum_{n=1}^{\infty} \mathbb{P}_0(Q_n) \\ &= \sum_{n=1}^{\infty} \int_{Q_n} \prod_{k=1}^n [p_0(y_k) d(y_k)] \\ &\leq \sum_{n=1}^{\infty} \int_{Q_n} \prod_{k=1}^n [p_1(y_k) d(y_k)] \times \frac{1}{b}. \end{aligned}$$

The last inequality follows from the fact that in Q_n we have $\prod_{k=1}^n p_0(y_k) \leq \frac{1}{b} \times \prod_{k=1}^n p_1(y_k)$. Hence,

$$\begin{aligned} \alpha &\leq \frac{1}{b} \times \mathbb{P}_1[L_N(Y_1, \dots, Y_N) \geq b], \\ &= \frac{1}{b} \times \mathbb{P}_1[\Gamma_1], \\ &= \frac{1}{b} \times (1 - \gamma), \end{aligned}$$

i.e.,

$$\alpha \leq \frac{1 - \gamma}{b}. \quad (13)$$

Letting $\{A_n\} \triangleq \{y \in \mathbb{R}^\infty | N = n, L_n(y_1, \dots, y_n) \leq a\}$, we have

$$\begin{aligned} \gamma &= \mathbb{P}_1(\mathsf{T}_0), \\ &= \sum_{n=1}^{\infty} \mathbb{P}_1(A_n), \\ &= \sum_{n=1}^{\infty} \int_{A_n} \prod_{k=1}^n [p_1(y_k) d(y_k)], \\ &\leq \sum_{n=1}^{\infty} \int_{A_n} \prod_{k=1}^n [p_0(y_k) d(y_k)] \times a. \end{aligned}$$

The last inequality follows from the fact that in A_n we have $\prod_{k=1}^n p_1(y_k) \leq a \times \prod_{k=1}^n p_0(y_k)$. Hence,

$$\begin{aligned} \gamma &\leq a \times \mathbb{P}_0[L_N(Y_1, \dots, Y_N) \leq b], \\ &= a \times \mathbb{P}_0[\mathsf{T}_0], \\ &= \frac{1}{b} \times (1 - \alpha), \end{aligned}$$

i.e.,

$$\gamma \leq \frac{(1 - \alpha)}{b} \quad (14)$$

Hence, from equations (13) and (14), we have,

$$b \leq \frac{1 - \gamma}{\alpha} \quad (15)$$

$$\text{and, } a \geq \frac{\gamma}{1 - \alpha}. \quad (16)$$

Note that these are inequalities, which mean that this is a conservative guideline to get $\mathbb{P}_F \leq \alpha$, and $\mathbb{P}_M \leq \gamma$, if one sets the corresponding a and b as given by equations (3) and (4). Moreover, if we assume that whenever $L_n(y_1, \dots, y_n)$ crosses a or b and stops, the overshoot $(L_n - b)$ or $(a - L_n)$ is negligible, then we get

$$b \approx \frac{1 - \gamma}{\alpha} \text{ and } a \approx \frac{\gamma}{1 - \alpha}. \quad (17)$$

This is called *Wald's approximation*.

2.2 Calculation of expected number of samples in a SPRT

Lemma 2.2 (Wald's Lemma). Let Z_1, Z_2, \dots be i.i.d. random variables with mean μ . Let $K \geq 0$ be any integer valued random variable with finite expectation, $\mathbb{E}[K] < \infty$ and such that the event $\{K = k\}$ is completely determined by $(Z_1, \dots, Z_k) \forall k$. Then,

$$\mathbb{E} \left[\sum_{i=1}^K Z_i \right] = [\mathbb{E}K]\mu. \quad (18)$$

K-L divergence Given two distributions P_1 and P_0 on Γ , the K-L divergence between P_1 and P_0 is,

$$D(P_1 || P_0) = \mathbb{E}_1 \left[\log \frac{P_1(Y)}{P_0(Y)} \right].$$

Recall that $N = \min\{n \geq 0, \phi_n(Y_1, \dots, Y_n) = 1\}$. Clearly, the event $\{N = n\}$ just depends on Y_1, \dots, Y_n . Additionally, if we assume that $\mathbb{E}[N] < \infty$, then we can write

$$\begin{aligned} \mathbb{E}_0[\log L_N(Y_1, \dots, Y_n)] &= \mathbb{E}_0 \left[\sum_{k=1}^N \log \frac{p_1(Y_k)}{p_0(Y_k)} \right] \\ &= \mathbb{E}_0[N] \cdot \mathbb{E}_0 \left[\log \frac{p_1(Y_k)}{p_0(Y_k)} \right] \\ &= -\mathbb{E}_0[N] \cdot D[P_0 || P_1]. \end{aligned} \quad (19)$$

Here the second equality follows from a direct consequence Wald's lemma on the function $g_k(Y_k) \triangleq \log \frac{p_1(Y_k)}{p_0(Y_k)}$ and the second inequality follows from the definition of K-L divergence.

Again, we can alternatively use the Wald's approximation which essentially says that at the stopping time either $L_n(Y_1, \dots, Y_n) \approx a$ or $\approx b$. If we take the expectation of the likelihood over hypothesis H_0 ,

$$\mathbb{E}_0[\log L_n(Y_1, \dots, Y_n)] \approx (\log a)(1 - \alpha) + (\log b)\alpha. \quad (20)$$

This is because under hypothesis H_0 , the likelihood ratio can take the value a with probability $(1 - \alpha)$ which is essentially $\mathbb{P}_0(\Gamma_0)$, or value b with probability (α) which is essentially $\mathbb{P}_0(\Gamma_1)$. From (19) and (20) we can now write,

$$\begin{aligned} \mathbb{E}_0[N] &\approx \frac{(1-\alpha) \log a + \alpha \log b}{-D[P_0||P_1]}, \\ &= \frac{(1-\alpha) \log \frac{1-\alpha}{\gamma} + \alpha \log \frac{\alpha}{1-\gamma}}{D[P_0||P_1]} \end{aligned} \quad (21)$$

Similarly, we can repeat the exercise to calculate the expected number of samples under hypothesis H_1 .

$$\begin{aligned} \mathbb{E}_1[\log L_N(Y_1, \dots, Y_n)] &= \mathbb{E}_1 \left[\sum_{k=1}^N \log \frac{p_1(Y_k)}{p_0(Y_k)} \right] \\ &= \mathbb{E}_1[N] \cdot \mathbb{E}_1 \left[\log \frac{p_1(Y_k)}{p_0(Y_k)} \right] \\ &= \mathbb{E}_1[N] \cdot D[P_1||P_0]. \end{aligned} \quad (22)$$

Here the second equality follows from a direct consequence Wald's lemma on the function $g_k(Y_k) \triangleq \log \frac{p_1(Y_k)}{p_0(Y_k)}$, and the second inequality follows from the definition of K-L divergence.

Again, we can alternatively use the Wald's approximation which essentially says that at the stopping time either $L_n(Y_1, \dots, Y_n) \approx a$ or $\approx b$. If now we take the expectation of the likelihood over hypothesis H_1 ,

$$\mathbb{E}_1[\log L_n(Y_1, \dots, Y_n)] \approx (\log a)(\gamma) + (\log b)(1 - \gamma). \quad (23)$$

This is because under hypothesis H_1 , the likelihood ratio can take the value a with probability γ which is essentially $\mathbb{P}_1(\Gamma_0)$, or value b with probability $(1 - \gamma)$ which is essentially $\mathbb{P}_1(\Gamma_1)$.

From (22) and (23) we can now write,

$$\mathbb{E}_1[N] \approx \frac{(1-\gamma) \log \frac{1-\gamma}{\alpha} + \gamma \log \frac{\gamma}{1-\alpha}}{D[P_1||P_0]}. \quad (24)$$