9. Sequential Testing

ECE 830, Spring 2017

So far we have considered simple hypotheses of the form

$$H_i \; : \; X_1, X_2, \ldots, X_n \overset{iid}{\sim} p_i \; , \; i = 0, 1 \; .$$

The error probabilities decrease as n (the number of iid observations) increases, and we characterized the minimum number n needed to achieve desired levels of error.

Rather than fixing n ahead of time, it is natural to consider a sequential approach to testing which **continues to gather samples until a confident decision can be made**. This idea goes back to Wald '45, and is usually referred to as a **sequential probability ratio test** (SPRT), also called a sequential likelihood ratio test.

The Sequential Probability Ratio Test

The SPRT is based on considering the likelihood ratio as a function of the number of observations. Define

$$\Lambda_k := \prod_{i=1}^{k} \frac{p_1(X_i)}{p_0(X_i)} \ , \ k = 1, 2, \ldots$$

The goal of the SPRT is to decide which hypothesis is correct as soon as possible (i.e., for the smallest value of k). To do this the SPRT requires **two thresholds, $\gamma_1 > \gamma_0$.** The SPRT "stops" as soon as $\Lambda_k \geq \gamma_1$, and we then decide H_1 is correct, or when $\Lambda_k \leq \gamma_0$, and we then decide H_0 is correct.

Algorithm

For $k = 1, 2, \ldots$

1. Observe X_k
2. Compute Λ_k
3. If $\Lambda_k \geq \gamma_1$, declare H_1 and stop collecting data
4. Else if $\Lambda_k \leq \gamma_0$, declare H_0 and stop collecting data
5. Else continue

Setting thresholds

The key is to set the thresholds so that we are guaranteed a certain levels of error. **Making γ_1 larger and γ_0 smaller yields a test that will tend to stop later and produce more accurate decisions.** We will try to set the thresholds to provide desired probabilities of detection P_D and false-alarm P_{FA}.

We can express P_D as follows. To simplify the notation, let $x := (x_1, \dots, x_k)$ and write $p_j(x) := \prod_{i=1}^{k} p_j(x_i)$, $j = 0, 1$. P_D can be written in terms of the decision set $R_1 := \{x : \Lambda_k \geq \gamma_0\}$ as follows

$$
\begin{aligned}
P_D &= \int_{R_1} p_1(x)\, dx = \int_{R_1} \frac{p_1(x)}{p_0(x)}\, p_0(x)\, dx \\
&= \int_{R_1} \Lambda_k\, p_0(x)\, dx \geq \gamma_1 \int_{R_1} p_0(x) = \gamma_1\, P_{FA} ,
\end{aligned}
$$

where we use the fact that $\Lambda_k \geq \gamma_1$ on the set R_1.

Setting thresholds (cont.)

Similarly,

$$\begin{aligned} 1 - P_{FA} &= 1 - \int_{R_1} p_0(x)\,dx \;=\; \int_{R_0} p_0(x)\,dx \\ &= \int_{R_0} \frac{p_0(x)}{p_1(x)}\,p_1(x)\,dx = \int_{R_0} \Lambda_k^{-1}\,p_1(x)\,dx \\ &\geq \gamma_0^{-1} \int_{R_0} p_1(x) \;=\; \gamma_0^{-1}\,(1 - P_D)\ . \end{aligned}$$

Setting thresholds (cont.)

These expressions give us bounds on the thresholds necessary to achieve P_D and P_{FA}:

$$\begin{aligned} \gamma_1 &\leq \frac{P_D}{P_{FA}} , \\ \gamma_0 &\geq \frac{1 - P_D}{1 - P_{FA}} . \end{aligned}$$

Let us err on the side of conservatism and set $\gamma_1 = \frac{P_D}{P_{FA}}$ and $\gamma_0 = \frac{1-P_D}{1-P_{FA}}$. These thresholds guarantee that error probabilities of the test will be at least as small as specified by the choice of P_D and P_{FA}, but they could be too conservative. To gain insight into this issue, let us consider the **expected stopping time.**

Expected Stopping Time of SPRT

Since Λ_k is a random variable, the stopping time of the SPRT is also random. Let K^* denote the random (integer) stopping time. We wish to calculate the expected value of K^*. **To do this, we will relate $\mathbb{E}[K^*]$ to $\mathbb{E}[\log \Lambda_{K^*}]$.**

To simplify notation, we will let $\mathbb{E}_j$ denote the expectation with respect to p_j, $j = 0, 1$. First observe that for any **fixed** time k

$$\begin{aligned}
\mathbb{E}_j[\log \Lambda_k] &= \mathbb{E}_j\left[\sum_{i=1}^{k} \log \frac{p_1(X_i)}{p_0(X_i)}\right] \\
&= \sum_{i=1}^{k} \mathbb{E}_j\left[\log \frac{p_1(X_i)}{p_0(X_i)}\right] \\
&= \begin{cases} k\, D(p_1||p_0)\ , & j = 1 \\ -k\, D(p_0||p_1)\ , & j = 0 \end{cases}
\end{aligned}$$

where $D(p_1||p_0)$ and $D(p_0||p_1)$ are the KL-divergences between p_0 and p_1.

Now suppose that M is a positive integer-valued random variable, **independent** of $X_1, X_2, \ldots$. Then by conditioning on M we have

$$\begin{aligned}
\mathbb{E}_j[\log \Lambda_M] &= \mathbb{E}_j\left[\mathbb{E}_j[\log \Lambda_M \mid M]\right] \\
&= \mathbb{E}_j\left[\sum_{i=1}^{M} \mathbb{E}_j\left[\log \frac{p_1(X_i)}{p_0(X_i)} \mid M\right]\right] \\
&= \begin{cases} \mathbb{E}_j[M]\, D(p_1||p_0)\ , & j=1 \\ -\mathbb{E}_j[M]\, D(p_0||p_1)\ , & j=0 \end{cases}
\end{aligned}$$

The stopping time K^* is random, but it is also a function of $X_1, X_2, \ldots$ so we cannot apply the simple conditioning argument used for M above. However, a more delicate argument shows that **a similar result holds with M is replaced with K^*.**

Wald's Identity

Let $Y_1, Y_2, \ldots$ be independent and identically distributed random variables with mean μ. Let K be any integer-valued random variable such that $\mathbb{E}[K] < \infty$ and $K = k$ is an event determined by $Y_1, \ldots, Y_k$ and independent of Y_i, $i > k$. Then $\mathbb{E}[\sum_{i=1}^{K} Y_i] = \mu\, \mathbb{E}[K]$.

Proof: Write

$$\mathbb{E}\left[\sum_{i=1}^{K} Y_i\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} \mathbf{1}_{\{K \geq i\}} Y_i\right] = \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{K \geq i\}} Y_i]$$

where $\mathbf{1}_{\{K \geq i\}}$ is the indicator of the event $\{K \geq i\}$ (the interchange of expectation and summation is justified by the monotone convergence theorem).

Proof of Wald's identity (cont.)

Note that the event

$$\{K \geq i\} = \left(\bigcup_{j=1}^{i-1} \{K = j\} \right)^c$$

where the superscript c denotes the complement. Thus, the event is independent of $Y_i, Y_{i+1}, \ldots$ (since it is determined by $Y_1, \ldots, Y_{i-1}$). Therefore,

$$\begin{aligned} \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{K \geq i\}} Y_i] &= \mathbb{E}[Y_i] \sum_{i=1}^{\infty} \mathbb{E}[\mathbf{1}_{\{K \geq i\}}] \\ &= \mu \sum_{i=1}^{\infty} \mathbb{P}(K \geq i) \; = \; \mu \, \mathbb{E}[K] \end{aligned}$$

which proves the identity. □

Applying Wald's Identity

So, by Wald's Identity we have

$$\mathbb{E}_j[\log \Lambda_{K^*}] \quad = \quad \begin{cases} \mathbb{E}_j[K^*]\, D(p_1||p_0)\ , & j = 1 \\ -\mathbb{E}_j[K^*]\, D(p_0||p_1)\ , & j = 0 \end{cases}$$

Now to obtain an expression for $\mathbb{E}_j[K^*]$ we will derive another formula for $\mathbb{E}_j[\log \Lambda_{K^*}]$. Let us assume the value of the likelihood ratio is approximately equal to a threshold level when the SPRT terminates. The value of the likelihood ratio will typically be just slightly greater/lower than the upper/lower threshold level. Using this approximation we can write

$$\begin{aligned} \mathbb{E}_0[\log \Lambda_{K^*}] &\approx P_{FA} \log(\gamma_1) \ + \ (1 - P_{FA}) \log(\gamma_0)\ , \\ &= P_{FA} \log\left(\frac{P_D}{P_{FA}}\right) \ + \ (1 - P_{FA}) \log\left(\frac{1 - P_D}{1 - P_{FA}}\right)\ , \\ \mathbb{E}_1[\log \Lambda_{K^*}] &\approx P_D \log\left(\frac{P_D}{P_{FA}}\right) \ + \ (1 - P_D) \log\left(\frac{1 - P_D}{1 - P_{FA}}\right)\ . \end{aligned}$$

With these approximations we obtain expressions for $\mathbb{E}_j[K^*]$:

$$\begin{aligned}
\mathbb{E}_0[K^*] &\approx \frac{P_{FA}\log\left(\frac{P_D}{P_{FA}}\right) \ + \ (1-P_{FA})\log\left(\frac{1-P_D}{1-P_{FA}}\right)}{-D(p_0||p_1)}\ , \\
&= \frac{(1-P_{FA})\log\left(\frac{1-P_{FA}}{1-P_D}\right) \ - \ P_{FA}\log\left(\frac{P_D}{P_{FA}}\right)}{D(p_0||p_1)} \\
\mathbb{E}_1[K^*] &\approx \frac{P_D\log\left(\frac{P_D}{P_{FA}}\right) \ + \ (1-P_D)\log\left(\frac{1-P_D}{1-P_{FA}}\right)}{D(p_1||p_0)}\ , \\
&= \frac{P_D\log\left(\frac{P_D}{P_{FA}}\right) \ - \ (1-P_D)\log\left(\frac{1-P_{FA}}{1-P_D}\right)}{D(p_1||p_0)}\ .
\end{aligned}$$

Since we are only interested in cases where $P_D > 1/2$ and $P_{FA} < 1/2$, the final expressions are non-negative in both cases. Note that the expected stopping times increase as the KL divergences decreases (as the two densities become less distinguishable). Increasing P_D or decreasing P_{FA} also increases the expected stopping time.

Optimality of SPRT

The expected stopping time of the SPRT that we determined above is optimal. No other test can achieve the same P_D and P_{FA} with a smaller expected number of samples, under either hypothesis, as the following result shows.

Lower bound on expected stopping time of any testing procedure (Wald, 1948)

Let P_{FA} and P_D be given and consider any test with probabilities $P'_{FA} \leq P_{FA}$ and $P'_D \geq P_D$. Then the expected stopping times for the test are bounded as follows:

$$\begin{aligned} \mathbb{E}_0[K^*] &\geq \frac{(1-P_{FA})\log\frac{1-P_{FA}}{1-P_D} - (P_{FA})\log\frac{P_D}{P_{FA}}}{D(p_0||p_1)}, \\ \mathbb{E}_1[K^*] &\geq \frac{P_D\log\frac{P_D}{P_{FA}} + (1-P_D)\log\frac{1-P_D}{1-P_{FA}}}{D(p_1||p_0)}. \end{aligned}$$

The lemma shows that if no other test can have error levels as small or smaller than the SPRT and have expected stopping times less than the values computed above for the SPRT.

Proof: By definition

$$\begin{aligned} \mathbb{E}_1[\log \Lambda_{K^*}] &= \int_\Omega p_1(x) \log \frac{p_1(x)}{p_0(x)} dx \\ &= \int_R p_1(x) \log \frac{p_1(x)}{p_0(x)} dx + \int_{R^c} p_1(x) \log \frac{p_1(x)}{p_0(x)} dx \end{aligned}$$

where R is any subset Ω, with compliment R^c. We can bound the terms above using the **log-sum inequality,** a consequence of Jensen's inequality:

$$\int p_1(x) \log \frac{p_1(x)}{p_0(x)} dx \geq \left(\int p_1(x) dx \right) \left(\log \frac{\int p_1(x) dx}{\int p_0(x) dx} \right)$$

This implies

$$
\begin{aligned}
\mathbb{E}_1[\log \Lambda_{K^*}] \quad \geq \quad & \left(\int_{R} p_1(x)dx\right)\left(\log \frac{\int_{R} p_1(x)dx}{\int_{R} p_0(x)dx}\right) \\
& + \left(\int_{R^c} p_1(x)dx\right)\left(\log \frac{\int_{R^c} p_1(x)dx}{\int_{R^c} p_0(x)dx}.\right)
\end{aligned}
$$

Defining R to be the acceptance region of hypothesis 1, we have:

$$
\mathbb{E}_1[\log \Lambda_{K^*}] \quad \geq \quad P_D \log \frac{P_D}{P_{FA}} + (1 - P_D) \log \frac{1 - P_D}{1 - P_{FA}}
$$

As we showed for any test with random stopping time, by Wald's identity:

$$\mathbb{E}_1[\log \Lambda_{K^*}] \quad = \quad \mathbb{E}_1[K^*]\, D(p_1||p_0).$$

Combining the two, gives the lower bound

$$\mathbb{E}_1[K^*] \quad \geq \quad \frac{P_D \log \frac{P_D}{P_{FA}} + (1-P_D) \log \frac{1-P_D}{1-P_{FA}}}{D(p_1||p_0)}$$

Following the same argument, we can derive a lower bound for the expected number of measurements under hypothesis H_0:

$$\mathbb{E}_0[K^*] \geq \frac{(1-P_{FA}) \log \frac{1-P_{FA}}{1-P_D} - (P_{FA}) \log \frac{P_D}{P_{FA}}}{D(p_0||p_1)}$$

□