Lecture 16: Designing Sequential Tests: Part I

23rd October, 2025

Instructor: Shubhanshu Shekhar

TODO: fix the notation.

In our previous lecture, we introduce the three main problems of sequential anytime valid inference, and observed that all of them can be reduced to designing sequential power-one tests. In this lecture, we begin our discussion of some general design principles and technical results that will be used throughout in our discussion of power-one tests. To present our ideas in the most transparent way, in this lecture we focus on the technically simple case of i.i.d. observations on finite alphabets. Subsequent lectures will discuss how similar results can be derived in more general settings.

1 Simple Null and Alternative

Throughout this lecture, we assume that $\mathcal{X} = \{x_1, \ldots, x_m\}$ is a finite alphabet, and $\{\mathbb{P}_\theta = P_\theta^\infty : \theta \in \Delta_m\}$ denotes the set of all i.i.d. distributions on $\mathcal{X}^\infty$. In our notation, P_θ denotes a probability distribution on $\mathcal{X}$ with probability mass function (pmf) $\theta \in \Delta_m$. With this notation, consider the following problem: Given $\{X_n : n \geq 1\} \sim \mathbb{P}_\theta$ and a confidence parameter $\alpha \in (0, 1]$, construct a level-α power-one test to decide between

$$H_0 : \theta = \theta_0, \quad \text{versus} \quad H_1 : \theta = \theta_1.$$

Here, θ_0, θ_1 are two pmfs such that $D_{\mathrm{KL}}(P_{\theta_1} \parallel P_{\theta_0}) > 0$. As we discussed in Example 1 of Lecture 15, a natural approach for designing a level-α test in this case is

$$\tau = \inf\{n \geq 1 : L_n \geq 1/\alpha\}, \quad \text{where} \quad L_0 = 1, \ L_n = L_{n-1} \times \frac{\theta_1[X_n]}{\theta_0[X_n]}, \ \forall n \geq 1.$$

The stochastic process $\{L_n : n \geq 0\}$ adapted to the natural filtration $\{\mathcal{F}_n : n \geq 0\}$ with $\mathcal{F}_0 = \{\emptyset, \mathcal{X}^\infty\}$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ is called the *likelihood ratio* process. It is easy to verify that the likelihood ratio process is a *test martingale*; that is, a non-negative supermartingale with an initial value 1 under the null. That is, each L_n is integrable and for all $n \geq 1$, we have with $Y_n := \theta_1[X_n]/\theta_0[X_n]$:

$$\mathbb{E}_{\theta_0}[L_n \mid \mathcal{F}_{n-1}] = \mathbb{E}_{\theta_0}[L_{n-1} \times Y_n \mid \mathcal{F}_{n-1}] = L_{n-1}\mathbb{E}_{\theta_0}[Y_n \mid \mathcal{F}_{n-1}] = L_{n-1},$$

where the second equality uses the fact that L_{n-1} is $\mathcal{F}_{n-1}$-measurable, and the last inequality uses the assumption that $X_n \perp \mathcal{F}_{n-1}$, and that $\mathbb{E}_{\theta_0}[Y_n] = 1$. This establishes that $\{L_n : n \geq 1\}$ is a test-martingale under the null. Next, we state a time-uniform generalization of Markov's inequality, referred to as Ville's inequality, that will allow us to establish the level-α property of our tests.

Fact 1.1 (Ville (1939)). *Let $\{S_n : n \geq 0\}$ denote a nonnegative supermartingale in a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n \geq 0}, \mathbb{P})$; that is each $S_n \geq 0$ a.s., integrable, and satisfies $\mathbb{E}[S_n \mid \mathcal{F}_{n-}] \leq S_{n-1}$ a.s. Then, for any $\lambda > 0$, we have*

$$\mathbb{P}\left(\sup_{n \geq 0} S_n \geq \lambda\right) \leq \frac{\mathbb{E}[S_0]}{\lambda}.$$

In particular, if $S_0 = 1$ a.s., and $\lambda = 1/\alpha$, we get that $\mathbb{P}(\sup_{n \geq 0} S_n \geq 1/\alpha) \leq \alpha$.

Thus, a direct application of Ville's inequality implies the level-α property of our test τ. We now proceed to the analysis of the performance of this test under the alternative.

> **Proposition 1.2.** *Suppose the pmfs θ_0, θ_1 are such that there exist constants $-\infty < c_- < c_+ < \infty$ such that $\log(\theta_1[X]/\theta_0[X]) \in [c_-, c_+]$. Then, under the alternative, we have*
>
> $$\mathbb{E}_{\theta_1}[\tau] \leq \frac{\log(1/\alpha) + c_+}{D_{KL}(P_{\theta_1} \parallel P_{\theta_0})}.$$
>
> *As a consequence, we also have the power-one property, $\mathbb{P}_{\theta_1}(\tau < \infty) = 1$.*

Proof. Let us use the notation $S_n = \log L_n = \sum_{i=1}^{n} \log Y_i$, and observe that $\{\tau < n\} = \{S_n/n < \log(1/\alpha/n)\}$. Now, by the strong law of large numbers, we know that $S_n/n \overset{a.s.}{\to} \gamma := D_{\mathrm{KL}}(P_{\theta_1} \parallel P_{\theta_0})$, which implies that $\tau < \infty$ almost surely.

Since $\tau < \infty$ almost surely, and the increments of the process $\{S_n : n \geq 1\}$ are bounded, we can apply Wald's identity to observe

$$\mathbb{E}_{\theta_1}[S_\tau - \tau\gamma] = 0, \quad \text{or} \quad \mathbb{E}_{\theta_1}[S_\tau] = \gamma \mathbb{E}_{\theta_1}[\tau].$$

Finally, by the definition of τ, we know that $S_\tau \leq \log(1/\alpha) + c_+$. Using this in the above equality, we get the required

$$\mathbb{E}_{\theta_1}[\tau] \leq \frac{\log(1/\alpha) + c_+}{\gamma}.$$

This completes the proof. $\qquad\square$

Remark 1.3. We have presented the above statement under bounded log-likelihood ratio assumptions for simplicity. The boundedness assumption on the log-likelihood ratio is rather strong, but it can be relaxed by using more careful arguments to analyze the "overshoot" at the last step; that is, $S_\tau - \log(1/\alpha)$ in the proof of Proposition 1.2.

2 Point Null versus Composite Alternative

We now consider the case of the composite alternative.

$$H_0 : \theta = \theta_0, \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

In this case, we cannot use a single likelihood ratio. Instead, we use ideas from universal compression, to construct a mixture distribution using Jeffreys prior for the numerator. More specifically, define

$$L_0 = 1, \quad L_n = L_{n-1} \times \frac{\hat{\theta}_n[X_n]}{\theta_0[X_n]},$$

where $\widehat{\theta}_n$ is any $\mathcal{F}_{n-1}$-measurable prediction of the unknown pmf. Note that as long as each predictable $\widehat{\theta}_n$ is a pmf, the process $\{L_n : n \geq 0\}$ remains a nonnegative martingale, and hence the level-α property still holds.

Under the alternative however, the performance of the test depends on the quality of the predictions. At this point, we recall a useful fact about the individual sequence regret of the mixture method, or the KT estimator.

Fact 2.1. *Let q_J^n denote the mixture distribution w.r.t. Jeffreys prior for Δ_m. Then, there exists a constant C_m (depends only on the support size m), such that we have the bound*

$$\sup_{x^n \in \mathcal{X}^n} \sup_{p \in \Delta_m} \log \left(\frac{\prod_{i=1}^n p(x_i)}{q_J^n(x^n)} \right) \leq \frac{m-1}{2} \log n + C_m.$$

In particular, for any sequence $x^n \in \mathcal{X}^n$ and any $p \in \Delta_m$, we have

$$\log \left(\frac{1}{q_J^n(x^n)} \right) \leq \log \left(\frac{1}{\prod_{i=1}^n p(x_i)} \right) + r_n, \quad \text{with} \quad r_n := \frac{m-1}{2} \log n + C_m.$$

Now suppose that we select $\widehat{\theta}_n$ according to the mixture w.r.t. Jeffreys prior in the construction of the process $\{L_n : n \geq 1\}$, then for any $n \geq 1$. Then, under the alternative with the true parameter θ_1, Fact 2.1 implies that

$$\log L_n = \log \left(\frac{q_J^n(X^n)}{\theta_0(X^n)} \right) = \log \left(\frac{\theta_1(X^n)}{\theta_0(X^n)} \right) + \log \left(\frac{q_J^n(X^n)}{\theta_0(X^n)} \right)$$
$$\overset{a.s.}{\geq} \log L_n^* - \sup_{x^n} \log \left(\frac{\theta_0(x^n)}{q_J^n(x^n)} \right) \overset{a.s.}{\geq} \log L_n^* - r_n.$$

Hence, the random stopping time τ can be upper bounded by $T := \inf\{n \geq 1 : \log L_n^* \geq \log(1/\alpha) + r_n\}$. The key point is that T_J is easier to analyze as it involves the first boundary crossing ($r_n + \log(1/\alpha)$) of a random walk $\log L_n^* = \sum_{i=1}^n \log Y_i$, where $Y_i = \theta_1[X_i]/\theta_0[X_i]$. In particular, observe that

$$\mathbb{E}_{\theta_1}[\log L_n] \geq \mathbb{E}_{\theta_1}[\log L_n^*] - r_n = n\gamma - \frac{m-1}{2} \log n - C_m, \tag{1}$$

where recall that $\gamma = D_{\mathrm{KL}}(P_{\theta_1} \parallel P_{\theta_0})$. Hence, the process $\log L_n$ is expected to grow at a linear rate (with a negligible logarithmic perturbation by the universal compression regret). This allows us to obtain an upper bound on the expected stopping time as we record in our next result.

> **Proposition 2.2.** *Under the alternative, suppose the true parameter $\theta_1 \neq \theta_0$ is such that there exist constants $-\infty < c_- < c_+ < \infty$ such that $\log(\theta_1[X]/\theta_0[X]) \in [c_-, c_+]$. Then, we have the following with $\gamma := D_{KL}(P_{\theta_1} \parallel P_{\theta_0})$:*
>
> $$\mathbb{E}_{\theta_1}[\tau] = \mathcal{O}\left(\frac{\log(1/\alpha)}{\gamma}\right).$$
>
> *As a consequence, we also have the power-one property, $\mathbb{P}_{\theta_1}(\tau < \infty) = 1$.*

We state the result in the $\mathcal{O}(\cdot)$ form to simplify the calculations, but the statement can be sharpened by a more careful argument.

Proof. Let us use the notation $S_n = \log(L_n)$ and $S_n^* = \log L_n^*$, where L_n is the likelihood function with the mixture distribution in the numerator, and L_n^* is the true likelihood, and define $T = \inf\{n \geq 1 : S_n^* \geq \log(1/\alpha) + r_n\}$. Then, by (1), we know that $T \leq \tau$ almost surely. Thus, to get an upper bound on $\mathbb{E}_{\theta_1}[\tau]$ it suffices to upper bound the expectation of T.

Now, observe the following:

- For $\gamma > 0$, the stopping time T is almost surely finite. This is because $S_n^*/n \overset{a.s.}{\to} \gamma > 0$, while the normalized threshold $(\log(1/\alpha) + r_n)/n \to 0$. Hence, S_n^* crosses the boundary with probability 1.

- Also, by the assumption of the bounded log-likelihood ratio, the process S_n^* has bounded increments.

Because of these conditions, we can apply Wald's identity to the martingale $\{S_n^* - n\gamma : n \geq 1\}$ with the stopping time T, to get

$$0 = \mathbb{E}_{\theta_1}[S_0^* - 0] = \mathbb{E}_{\theta_1}[S_T^* - T\gamma] \quad \implies \quad \mathbb{E}_{\theta_1}[S_T^*] = \gamma\mathbb{E}_{\theta_1}[T]. \tag{2}$$

Next, by the definition of T, and the bounded assumption, we know that $S_T \leq \log(1/\alpha) + r_T + c_+$, which implies that

$$\mathbb{E}_{\theta_1}[S_T] \leq \log(1/\alpha) + c_+ + \log \mathbb{E}_{\theta_1}[r_T]. \tag{3}$$

Recall that $r_T = ((m-1)/2)\log T + C_m$. Combining (2) and (3), we can conclude that

$$\mathbb{E}_{\theta_1}[T] = \mathcal{O}\left(\frac{\log(1/\alpha\gamma)}{\gamma}\right).$$

This completes the proof, as $\mathbb{E}_{\theta_1}[\tau] \leq \mathbb{E}_{\theta_1}[T]$ by construction. $\qquad \square$

3 Composite Null versus Composite Alternative

We now consider the most general case of $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$. In this case, we present a suboptimal and conservative approach to designing sequential tests by using a 'worst-case' point null to construct a test-martingale.

For the given null class, define $\bar{\theta}_0$ as the unnormalized distribution with $\bar{\theta}_0[x] = \sup_{\theta \in \Theta_0} \theta[x]$ for all $x \in \mathcal{X}$. Then, for any $n \geq 1$, and with q_J^n denoting the mixture distribution (with Jeffreys prior), we define the process

$$L_n = \frac{q_J^n(X^n)}{\prod_{i=1}^n \bar{\theta}_0[X_i]}, \quad \text{where} \quad q_J^n(x^n) = \int_{\Delta_m} \theta[x^n] \pi_J(\theta) d\theta, \ \forall x^n \in \mathcal{X}^n.$$

Level-α property. Now, under the null if the true parameter is some $\theta_0 \in \Theta_0$, then due to our conservative choice of $\bar{\theta}_0$ that dominates any θ_0 pointwise (i.e., $\theta_0[x] \leq \bar{\theta}_0[x]$ for all x, we have that

$$L_n \leq L'_n = \frac{q_J^n(X^n)}{\theta_0[X^n]} \text{ for all } n \geq 1.$$

In other words, L_n is upper-bounded by L'_n which is exactly the process we constructed in Section 2. Since $\{L'_n : n \geq 0\}$ is a test-martingale under $\theta = \theta_0$, we can conclude that $\tau = \inf\{n \geq 1 : L_n \geq 1/\alpha\}$ satisfies the level-α property.

Expected Stopping Time. The expected stopping time analysis follows a similar pattern as the previous two cases. In particular, introduce the terms

$$\bar{\gamma} = \sum_{x \in \mathcal{X}} \theta_1[x] \log \left(\frac{\theta_1[x]}{\bar{\theta}_0[x]} \right), \quad \text{and} \quad \bar{S}_n = \log \frac{\prod_{i=1}^n \theta_1[X_i]}{\prod_{i=1}^n \bar{\theta}_0[X_i]}.$$

Then, by using the universal compression regret, we can show that

$$S_n = \log L_n \geq \bar{S}_n - r_n, \quad \text{with} \quad r_n = \frac{m-1}{2} \log n + C_m.$$

Then, as before, we can upper-bound τ with the stopping time $T = \inf\{n \geq 1 : \bar{S}_n \geq \log(1/\alpha) + r_n\}$, and show that

$$\mathbb{E}_{\theta_1}[\tau] = \mathcal{O} \left(\frac{\log(1/l\alpha\bar{\gamma})}{\bar{\gamma}} \right).$$

The key point of difference is that $\bar{\gamma}$ is a more conservative measure of hardness of the problem. A correct complexity measure should be

$$\gamma_{\theta_1} = \inf_{\theta_0 \in \Theta_0} D_{\text{KL}}(P_{\theta_1} \| P_{\theta_0}).$$

By using less conservative process $\{L_n : n \geq 1\}$ and under certain assumptions, we can construct stopping times that depend on γ_{θ_1}.

A Proof of Fact 1.1

Introduce the (possibly infinite) stopping time $\tau = \inf\{n \geq 1 : S_n \geq \lambda\}$, and defined the "stopped process" $\{S_n^\tau : n \geq 1\}$, with $S_n^\tau = S_{\tau \wedge n} = S_{\min\{\tau, n\}}$. Then it is a standard fact that stopped supermartingales are also supermartingales. In particular, this implies that

$$\mathbb{E}[S_n^\tau] \leq \mathbb{E}[S_0^\tau] = \mathbb{E}[S_0].$$

Next, we consider the two events $\{\tau \leq n\}$ and $\{\tau > n\}$.

- Under $\{\tau \leq n\}$, we know that $\min\{\tau, n\} = \tau$, and hence $S_n^\tau = S_\tau \geq \lambda$ (by definition).

- Under $\{\tau > n\}$, we use the trivial bound that $S_n^\tau = S_n \geq 0$ almost surely.

Together, these two observations can be summarized as

$$S_n^\tau \geq \lambda \mathbf{1}_{\tau \leq n} \quad \implies \quad \mathbb{P}\left(\tau \leq n\right) \leq \frac{1}{\lambda}\mathbb{E}[S_n^\tau] \leq \frac{1}{\lambda}\mathbb{E}[S_0]$$

Since the event $\{\tau \leq n\}$ increases to $\{\tau < \infty\}$ as $n \uparrow \infty$, continuity of probability implies $\mathbb{P}(\tau < \infty) = \lim_{n \to \infty} \mathbb{P}(\tau \leq n)$. Thus, we get the required

$$\mathbb{P}(\tau < \infty) = \mathbb{P}(\exists n < \infty : S_n \geq \lambda) \leq \frac{\mathbb{E}[S_0]}{\lambda}.$$

This completes the proof. $\qquad\square$

References

J. Ville. *Étude critique de la notion de collectif.* PhD thesis, Université de Paris, 1939.