More on the Sequential Probability Ratio Test

BS2 Statistical Inference, Lecture 15
Michaelmas Term 2004

Steffen Lauritzen, University of Oxford; December 1, 2004

The sequential probability ratio test

The SPRT for a simple hypothesis $H_0 : \theta = \theta_0$ against a simple alternative $H_1 : \theta = \theta_1$ has the following form:

- If $\Lambda_n \geq B$, decide that H_1 is true and stop;
- If $\Lambda_n \leq A$, decide that H_0 is true and stop;
- If $A < \Lambda_n < B$, collect another observation to obtain Λ_{n+1},

where Λ_n is the log-likelihood ratio

$$\Lambda_n = \lambda(X_1, \ldots, X_n) = \log \frac{L(\theta_1; X_1, \ldots, X_n)}{L(\theta_0; X_1, \ldots, X_n)}.$$

Example

Consider X_i independent Bernouilli variables with

$$P(X_i = 1; \theta) = 1 - P(X_i = 0; \theta) = \theta.$$

In this case we get

$$\begin{aligned} \lambda(x_1, \ldots, x_n) &= \log \frac{\theta_1^{\sum x_i}(1-\theta_1)^{n-\sum x_i}}{\theta_0^{\sum x_i}(1-\theta_0)^{n-\sum x_i}} \\ &= \left(\sum x_i\right) \log \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)} + n \log \frac{1-\theta_1}{1-\theta_0}. \end{aligned}$$

If we assume $\theta_1 > \theta_0$ the SPRT thus takes the form that we defer decision if

$$\rho A + \eta \rho n < \sum X_i < \rho B + \eta \rho n,$$

where

$$\rho^{-1} = \log \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}, \quad \eta = \log \frac{1-\theta_0}{1-\theta_1},$$

which has a simple graphical representation, as the boundaries depend on n as parallel straight lines with slope $\rho\eta$ and intercepts $(\rho A, \rho B)$

Limits and error probabilities

Next we will derive the relation between the decision limits A and B and the error probabilities

$$\alpha = P(D_1 \,|\, H_0), \quad \beta = P(D_0 \,|\, H_1),$$

where $P(D_i \,|\, H_j)$ denotes the probability of deciding that H_i is true when in fact H_j is.

Consider a sequence $x_1, \ldots, x_n$ so that D_1 is taken at stage n. For each such sequence we have

$$f(x_1, \ldots, x_n; \theta_1) \geq e^B f(x_1, \ldots, x_n; \theta_0), \tag{1}$$

since this is the condition for deciding that H_1 is true.

Now let D_{1n} denote the set of such sequences. If we assume that a decision is taken at some point with probability one (which we shall prove in a moment), we get

$$\begin{aligned} P(D_1 \mid H_1) &= \sum_n P(D_{1n} \mid H_1) \\ &= \sum_n \int_{D_{1n}} f(x_1, \ldots, x_n; \theta_1)\, dx_1 dx_2 \cdots dx_n \\ &\geq \sum_n \int_{D_{1n}} e^B f(x_1, \ldots, x_n; \theta_0)\, dx_1 dx_2 \cdots dx_n \\ &= e^B P(D_1 \mid H_0), \end{aligned}$$

in other words we have

$$1 - \beta \geq e^B \alpha.$$

Reversing the role of H_0 and H_1 and rewriting the inequalities we obtain

$$B \leq \log \frac{1-\beta}{\alpha}, \quad A \geq \log \frac{\beta}{1-\alpha}.$$

Now, in fact, let us examine how sharp these inequalities were. Suppose the likelihood ratio only changed in very small steps, so that the log-likelihood ratio was in fact almost *equal to B* when H_1 was decided. Then (1) would read

$$f(x_1, \ldots, x_n; \theta_1) \approx e^B f(x_1, \ldots, x_n; \theta_0)$$

which would in turn lead to the approximate relation

$$B \approx \log \frac{1-\beta}{\alpha}, \quad A \approx \log \frac{\beta}{1-\alpha}. \qquad (2)$$

The only error in this approximation is that we have ignored the 'overshoot', i.e. the fact that when the log-likelihood crosses the boundary, it would tend to satisfy $\Lambda_n = B + \delta_n$ rather than $\Lambda_n = B$ and similarly at the other boundary.

In most interesting cases this error is negligible for practical purposes and there is now a long and well-established practice in calculating decision limits by using the relations (2).

Are we certain to make a decision?

Assume that X_i are independent and identically distributed. Then

$$\Lambda_n = \sum_1^n \log \frac{f(X_i; \theta_1)}{f(X_i; \theta_0)} = \sum_1^n Y_i$$

so Λ_n is what is known as a *random walk*.

The function $\log x$ is strictly concave. If we assume $f(\cdot; \theta_1) \neq f(\cdot; \theta_0)$, Jensen's inequality yields

$$\begin{aligned} \mu_0 = \mathbf{E}(Y_i \mid H_0) &= \mathbf{E}\left\{\log \frac{f(X_i; \theta_1)}{f(X_i; \theta_0)} \mid H_0\right\} \\ &< \log \mathbf{E}\left\{\frac{f(X_i; \theta_1)}{f(X_i; \theta_0)} \mid H_0\right\} = 0, \end{aligned}$$

and similarly we get

$$\mu_1 = \mathbf{E}(Y_i \mid H_1) > 0.$$

The strong law of large numbers now yields that if H_0 is true, Λ_n/n will be close to $\mathbf{E}(Y \mid H_0) = \mu_0 < 0$. Thus, if we take $\epsilon = -\mu_0/2$ it holds with probability one for some N that

$$|\Lambda_N/N - \mu_0| < \epsilon \implies \Lambda_N < N(\epsilon + \mu_0) = N\mu_0/2$$

and thus for any A we would have $\Lambda_N < A$ provided N is sufficiently large.

Similarly if H_1 is true, for any B we would have $\Lambda_N > B$ for N sufficiently large.

Exponential families

The simple graphical representation in the binomial case has an easy generalisation to exponential families. In the general (curved) exponential family case we get

$$\Lambda_n = \{a(\theta_1) - a(\theta_0)\}^\top \sum t(X_i) - n\{c(\theta_1) - c(\theta_0)\}$$

so if we let

$$u(x) = \{a(\theta_1) - a(\theta_0)\}^\top t(x),$$

the decision boundaries are again parallel straight lines:

$$A + \eta n < \sum u(X_i) < B + \eta n$$

where $\eta = c(\theta_1) - c(\theta_0)$.