COMPUTER PROJECT 5

Revision as of Thursday, November 14th, 2024 at 13:51:36 PM

Due: Monday, November 25, 2024

This computer project's focus is on the Sequential Probability Ratio Test (SRPT) of a specific simple null hypothesis $\mathcal{H}_0$ against a specific simple alternative hypothesis $\mathcal{H}_1$. The main question of interest is the performance of the SPRT as a function of the desired type I & II error probabilities, the "distance" between the two hypotheses, etc. The SPRT is due to Abraham Wald (1947) who essentially set fourth the entire field of sequential analysis. One remarkable fact about the SPRT is that its average decision time under either hypothesis is the lowest possible among all reasonable tests with type I & II error probabilities no worse than those of the SPRT. However, this theoretically strong and practically appealing property is extremely "fragile" in the sense that the two hypotheses really do need to be simple; if the hypotheses are composite (e.g., to handle some form of parametric uncertainty), then the otherwise unbeatable performance of the SPRT goes down the drain, and the performance degradation may be severe.

The classical SPRT of the simple hypothesis $\mathcal{H}_0\colon \vartheta = \vartheta_0$ against the simple alternative $\mathcal{H}_1\colon \vartheta = \vartheta_1 (\neq \vartheta_0)$ is based upon a sequence of independent and identically distributed (iid) observations $X_1, X_2, \ldots$ drawn sequentially from a population characterized by the pdf $p(x;\vartheta)$ where ϑ is a parameter. That is, the observations $X_1, X_2, \ldots$ are all iid and their common pdf is $p(x;\vartheta)$.

At each time n the log-likelihood ratio to test $\mathcal{H}_0$ against $\mathcal{H}_1$ using the sample $X_1, X_2, \ldots, X_n$ is

$$\mathcal{L}_{1:n} = \sum_{i=1}^{n} \log \frac{p(X_i;\vartheta_1)}{p(X_i;\vartheta_0)}, \quad n = 1, 2, \ldots,$$

and it can also be rewritten as the random walk $\{S_n\}$ given by

$$S_n = S_{n-1} + Z_n, \quad n = 1, 2, \ldots,$$

where $S_0 = 0$ and $Z_n = \log[p(X_n;\vartheta_1)/p(X_n;\vartheta_0)]$ is the log-likelihood ratio of the n-th data point X_n.

Given $\{S_n\}$, the SPRT for testing $\mathcal{H}_0$ against $\mathcal{H}_1$ is described by the stopping rule

$$N = \inf\{n \geqslant 1\colon S_n \notin (-a_0, a_1)\} \quad \text{with} \quad \inf\{\varnothing\} = \infty,$$

where $a_0, a_1 > 0$, and the terminal decision rule

$$\delta = \begin{cases} 1, & \text{if } S_n \geqslant a_1, \\ 0, & \text{if } S_n \leqslant -a_0, \end{cases}$$

which is the index of the accepted hypothesis.

The performance of an SPRT is traditionally assessed in terms of its Operating Characteristic (OC) function and in terms of its Average Sample Number (ASN) function. Formally, $\mathrm{OC}(\vartheta) := \mathbb{P}_\vartheta(\text{accept } \mathcal{H}_0)$ and $\mathrm{ASN}(\vartheta) := \mathbb{E}_\vartheta[N]$; the subscript ϑ is to indicate that the quantity is computed assuming the value of the parameter is ϑ. Note that $\mathrm{OC}(\vartheta)$ and $\mathrm{ASN}(\vartheta)$ both depend on the stopping bounds $-a_0$ and a_1. Also, since $\mathbb{P}_\vartheta(\text{accept } \mathcal{H}_0) = \mathbb{P}_\vartheta(\delta = 0) = \mathbb{P}_\vartheta(S_N \leqslant -a_0)$, the OC function is nothing but the lower exit probability.

Let $\alpha_0 = \mathbb{P}_{\vartheta_0}(\text{reject } \mathcal{H}_0)$, which is the type I error probability, and $\alpha_1 = \mathbb{P}_{\vartheta_1}(\text{accept } \mathcal{H}_0)$, which is the type II error probability. A type I error is usually associated with a "false alarm", while a type II error can be interpreted as an "oversight". It is direct to see that $\mathrm{OC}(\vartheta_0) = 1 - \alpha_0$ and $\mathrm{OC}(\vartheta_1) = \alpha_1$. Apart from the values of the OC function at ϑ_0 and at ϑ_1, we will also be interested in the values of the ASN function at the same two points, i.e., $\mathrm{ASN}(\vartheta_0)$ and $\mathrm{ASN}(\vartheta_1)$.

To compute $\mathrm{OC}(\vartheta)$ and $\mathrm{ASN}(\vartheta)$ for a pair of given bounds a_0 and a_1, recall that the SPRT is essentially a stopped random walk. Let

$x \in \mathbb{R}$ be an arbitrary fixed number, and define $\mathrm{OC}(\vartheta, x)$ and $\mathrm{ASN}(\vartheta, x)$ to be, respectively, the OC function and the ASN function of an SPRT whose decision region is the interval $(-a_0 - x, a_1 - x)$; note that we are after $\mathrm{OC}(\vartheta) = \mathrm{OC}(\vartheta, 0)$ and $\mathrm{ASN}(\vartheta) = \mathrm{ASN}(\vartheta, 0)$. Using the Markov property of $\{S_n\}$, it is straightforward to see that

$$\mathrm{OC}(\vartheta, x) = P_\vartheta^Z(-a_0 - x) + \int_{-a_0}^{a_1} \mathcal{K}_\vartheta(x, y)\, \mathrm{OC}(\vartheta, y)\, dy, \tag{1}$$

$$\mathrm{ASN}(\vartheta, x) = 1 + \int_{-a_0}^{a_1} \mathcal{K}_\vartheta(x, y)\, \mathrm{ASN}(\vartheta, y)\, dy, \tag{2}$$

where $P_\vartheta^Z(z) = \mathbb{P}_\vartheta(Z_1 \leqslant z)$ and

$$\mathcal{K}_\vartheta(x, y) = \frac{\partial}{\partial y} P_\vartheta^Z(y - x). \tag{3}$$

Equations (1) and (2) are renewal equations written in the Fredholm integral form of the second type. Such equations rarely permit an exact analytical solution, and therefore are usually solved numerically. One way to treat these equations numerically was proposed by Moustakides et al. (2011), and the idea is as follows.

First, observe that (1) and (2) are each a special case of the more general equation

$$u(x) = v(x) + \int_{-a_0}^{a_1} \mathcal{K}_\vartheta(x, y)\, u(y)\, dy, \tag{4}$$

where the equation's kernel $\mathcal{K}_\vartheta(x, y)$ is as in (3).

To see that (4) is an "umbrella" equation for the equations we are after, observe that, e.g., to obtain equation (1) on the OC function, $\mathrm{OC}(\vartheta, x)$, it is enough to assume $v(x) = P_\vartheta^Z(-a_0 - x)$ for all x; recall that $P_\vartheta^Z(z) = \mathbb{P}_\vartheta(Z_1 \leqslant z)$, i.e., it is the cdf of the log-likelihood ratio $Z_1 = \log[p(X_1; \vartheta_1)/p(X_1; \vartheta_0)]$. Similarly, equation (2) which governs the ASN function, $\mathrm{ASN}(\vartheta, x)$, can be derived from (4) merely by taking $v(x) \equiv 1$ for all x. Thus, solving (4) is the same as solving (1) and (2) – both at once.

Let

$$\mathcal{K}_\vartheta\, \varphi = \int_{-a_0}^{a_1} \mathcal{K}_\vartheta(x, y)\, \varphi(y)\, dy,$$

be the linear operator induced by the kernel $\mathcal{K}_\vartheta(x, y)$, and rewrite (4) equivalently in the operator form as $u = v + \mathcal{K}_\vartheta\, u$.

The numerical scheme of Moustakides et al. (2011) is a *piecewise-constant* (*zero-order polynomial*) *collocation method*. This is a special case of the *piecewise collocation method*, an interpolation-projection type of technique; see, e.g., Atkinson and Han (2009, Chapter 12), (Delves and Walsh, 1974, Chapter 7) and Kress (1998, Chapter 12.4). The method starts with introducing $-a_0 = y_0 < y_1 < \ldots < y_N = a_1$, $N > 0$, a partition of the interval $[-a_0, a_1]$; in general, the nodes $\{y_i\}$ need not be equidistant. Next, the sought function, $u(x)$, is approximated as

$$u_N(x) = \sum_{j=1}^{N} u_{j,N}\, \chi_j(x),$$

where $\{u_{j,N}\}_{1 \leqslant j \leqslant N}$ are constant coefficients to be determined, and $\{\chi_j\}_{1 \leqslant j \leqslant N}$ are suitably chosen (known) basis functions.

The idea of the method is to seek a prescription for choosing the coefficients $\{u_{j,N}\}_{1 \leqslant j \leqslant N}$. For any such choice, substitution of $u_N(x)$ into the equation will give a residual $u_N - \mathcal{K}_\vartheta\, u_N - v$. Unless the true solution $u(x)$ itself is a linear combination of the basis functions $\{\chi_j\}_{1 \leqslant j \leqslant N}$, no choice of the coefficients $\{u_{j,N}\}_{1 \leqslant j \leqslant N}$ will make the residual identically zero. However, by requiring the residual $u_N - \mathcal{K}_\vartheta\, u_N - v$ to be zero at some $\{x_j\}_{1 \leqslant j \leqslant N}$, where $x_j \in [-a_0, a_1]$ for all $j = 1, 2, \ldots, N$, one can achieve a certain level of closeness of the residual to zero. All this leads to the following system of N algebraic equations on the coefficients $u_{j,N}$:

$$\boldsymbol{u}_N = \boldsymbol{v} + \boldsymbol{\mathcal{K}}_\vartheta\, \boldsymbol{u}_N, \tag{5}$$

where $\boldsymbol{u}_N = [u_{1,N}, u_{2,N}, \ldots, u_{N,N}]'$, $\boldsymbol{v} = [v(x_1), v(x_2), \ldots, v(x_N)]'$ with $'$ being the standard transpose operator, and $\boldsymbol{\mathcal{K}}_\vartheta$ is an N-by-N matrix whose (i, j)-th element is

$$\int_{-a_0}^{a_1} \mathcal{K}_\vartheta(x_i, y)\, \chi_j(y)\, dy \quad \text{for} \quad i, j = 1, 2, \ldots, N.$$

For the system (5) to be consistent, the functions $\{\chi_j\}_{1\leqslant j\leqslant N}$ need to be chosen so as to form a basis in the appropriate functional space, i.e., in particular, $\{\chi_j\}_{1\leqslant j\leqslant N}$ need to be *linearly independent*. The specific choice of $\{\chi_j\}_{1\leqslant j\leqslant N}$ made by Moustakides et al. (2011) is $\chi_j(x) = \mathbb{1}_{\{y_{j-1}\leqslant x<y_j\}}$ with $y_j = -a_0 + j(a_0 + a_1)/N$ (i.e., equidistant nodes) for $j = 0, 1, \ldots, N$. For this basis, the (i, j)-th of matrix $\mathcal{K}_\vartheta$ is $P^Z_\vartheta(y_j - x_i) - P^Z_\vartheta(y_{j-1} - x_i)$; note that the kernel $\mathcal{K}_\vartheta(x, y)$ itself is not required to employ this method. As for the collocation points $\{x_i\}_{1\leqslant i\leqslant N}$, Moustakides et al. (2011) suggested to take the middle points $x_j = (y_{j-1} + y_j)/2$, $j = 1, 2, \ldots, N$.

The theory of collocation-type numerical methods is fairly well-developed. In particular, it is well known that the rate of convergence of a collocation method is *exactly equal* to that of the underlying (piecewise) interpolation scheme; see, e.g., (Atkinson and Han, 2009, Theorem 12.1.2) and (Delves and Walsh, 1974, pp. 93–94). Therefore, since in our case the interpolation is of order zero, the rate is *linear*. The choice of the collocation points being the middle points ensures that the constant is the least possible out of all schemes with linear convergence rate.

Consider now the following Gaussian-against-Gaussian scenario. Suppose the distribution under the null is Gaussian with mean ϑ_0 and variance $\sigma^2 > 0$. Suppose further that the distribution under the alternative is also Gaussian, but with mean $\vartheta_1 \neq \vartheta_0$ and with the same variance $\sigma^2 > 0$. For this scenario, the "instantaneous" log-likelihood ratio, $Z_n = \log[p(X_n; \vartheta_1)/p(X_n; \vartheta_0)]$, for the n-th data point, X_n, can be seen to be

$$Z_n = \frac{\vartheta_1 - \vartheta_0}{\sigma^2} X_n - \frac{\vartheta_1^2 - \vartheta_0^2}{2\sigma^2},$$

whence one obtains $P^Z_\vartheta(z) = \Phi(z; [(\vartheta_1 - \vartheta_0)/\sigma^2] \cdot [\vartheta - (\vartheta_0 + \vartheta_1)/2], [(\vartheta_1 - \vartheta_0)/\sigma]^2)$, where hereafter $\Phi(x; \mu, \sigma^2)$ is the Gaussian cdf with mean μ and variance σ^2.

Given $P^Z_\vartheta(z)$, and one can employ, e.g., the numerical technique explained above to solve (1) and (2). It can be easily seen that the (i, j)-th element of the matrix $\mathcal{K}_\vartheta$ in this case is

$$\operatorname{sign}(\vartheta_1 - \vartheta_0) \cdot \Big\{ \Phi(y_j - x_i; [(\vartheta_1 - \vartheta_0)/\sigma^2] \cdot [\vartheta - (\vartheta_0 + \vartheta_1)/2], [(\vartheta_1 - \vartheta_0)/\sigma]^2) - \\ - \Phi(y_{j-1} - x_i; [(\vartheta_1 - \vartheta_0)/\sigma^2] \cdot [\vartheta - (\vartheta_0 + \vartheta_1)/2], [(\vartheta_1 - \vartheta_0)/\sigma]^2) \Big\},$$

where we recall that $y_j = -a_0 + j(a_0 + a_1)/N$, $j = 0, 1, \ldots, N$, $x_i = (y_{i-1} + y_i)/2$, $i = 1, 2, \ldots, N$, and $N > 1$.

Write an R script that implements the numerical method presented above to solve (1) and (2) for the Gaussian-against-Gaussian scenario. The script has to be flexible enough to handle any ϑ, ϑ_0, ϑ_1, $\sigma > 0$, N, a_0, a_1, etc.

Assume now that $\alpha_0 = \alpha_1 = 0.001$, $\vartheta_0 = 0$, $\vartheta_1 = 0.5$, and $\sigma = 1$. Use your script to first compute a_0 and a_1. These do have to be computed first to set up the SPRT. It may help to use some approximations to get a preliminary idea as to what a_0 and a_1 are. For example, Wald's approximations discussed in class may come in handy. Report the obtained results. Next, use your script to compute $\mathrm{ASN}(\vartheta)$ as a function of $\vartheta \in [0, 0.5]$. Prepare a plot showing $\mathrm{ASN}(\vartheta)$ as a function of $\vartheta \in [0, 0.5]$. Briefly comment on it.

Suppose now that you also have the classical Neyman–Pearson test to differentiate between $\mathcal{H}_0$ and $\mathcal{H}_1$. Assuming that the Neyman–Pearson test has the same type I and II error probabilities as the SPRT, find the corresponding sample size required by the Neyman–Pearson test. How does it compare to the $\mathrm{ASN}(\vartheta)$ curve for the SPRT?

What happens when $\alpha_0 = 0.001$ but $\alpha_1 = 0.0001$? What happens when $\alpha_0 = 0.0001$ and $\alpha_1 = 0.001$?

Submit all of your scripts, graphics, tables, and comments electronically via EMail.

References

Atkinson, K. and Han, W. (2009). *Theoretical Numerical Analysis: A Functional Analysis Framework*, volume 39 of *Texts in Applied Mathematics*. Springer, third edition.

Delves, L. M. and Walsh, J., editors (1974). *Numerical Solution of Integral Equations*. Oxford University Press.

Kress, R. (1998). *Numerical Analysis*, volume 181 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

Moustakides, G. V., Polunchenko, A. S., and Tartakovsky, A. G. (2011). A numerical approach to performance analysis of quickest change-point detection procedures. *Statistica Sinica*, 21(2).

Wald, A. (1947). *Sequential Analysis*. J. Wiley & Sons, Inc., New York.