A PRACTICAL INTRODUCTION TO THE GSPRT

MICHEL VAN DEN BERGH

1. Description of the GSPRT

Let $p(\underline{\theta}, x)$ be a parametrized distribution and let $L(\underline{\theta}, \underline{x}) = \sum_{i=1}^{N} \log p(\underline{\theta}, x_i)$ be the corresponding log-likelihood function for N independent trials.

Let $\phi(\underline{\theta})$ be a function of the parameters and assume given ϕ_0, $\phi_1 \in \mathbb{R}$. The Generalized Sequential Probability Likelihood Ratio Test [1] for $\phi = \phi_0$ versus $\phi = \phi_1$ is based on monitoring the difference

$$\text{LLR} := L(\hat{\underline{\theta}}_1, \underline{x}) - L(\hat{\underline{\theta}}_0, \underline{x}) \tag{1.1}$$

where $\hat{\underline{\theta}}_i$ is the maximum likelihood estimator for $\underline{\theta}$ subject to the condition $\phi(\hat{\underline{\theta}}_i) = \phi_i$. I.e.

$$\hat{\underline{\theta}}_i = \underset{\phi(\underline{\theta})=\phi_i}{\arg\max}\, L(\underline{\theta}, \underline{x})$$

Like for the SPRT the test stops when LLR leaves the interval $[\log(\beta/(1-\alpha)), \log((1-\beta)/\alpha)]$ where as usual α, β stand for the Type I,II error probabilities.

2. An approximation

Let $\hat{\underline{\theta}}$ be the unconstrained maximum likelihood estimator for $\underline{\theta}$, i.e.

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\arg\max}\, L(\underline{\theta}, \underline{x})$$

and put $\hat{\phi} = \phi(\hat{\underline{\theta}})$. Let $V(\hat{\phi})$ be an estimator for the variance $\text{Var}(\hat{\phi})$ of $\hat{\phi}$ with relative error $O(1/\sqrt{N})$ (thus any standard estimator will do). We claim that under suitable regularity conditions we have the following very convenient approximation for (1.1)

$$\boxed{L(\hat{\underline{\theta}}_1, \underline{x}) - L(\hat{\underline{\theta}}_0, \underline{x}) \cong \frac{1}{2}\frac{(\phi_1 - \phi_0)(2\hat{\phi} - \phi_0 - \phi_1)}{V(\hat{\phi})}} \tag{2.1}$$

This means that the, sometimes quite cumbersome, calculation of the conditional estimates $\hat{\underline{\theta}}_0$, $\hat{\underline{\theta}}_1$ is not actually required.

Example. Assume we take independent trials from a multinomial distribution with probabilities $(p_i)_{i=1,\ldots,n}$. Let $\phi = \sum_{i=1}^{n} a_i p_i$ for given $(a_i)_{i=1,\ldots,n} \in \mathbb{R}$. Assume that after N trials the outcome frequencies are $(N_i)_{i=1,\ldots,n}$ (with $N = \sum_{i=1}^{n} N_i$). Then to calculate (2.1) we calculate first the empiric probabilities $\hat{p}_i := N_i/N$ and then we put

$$\hat{\phi} = \sum_{i=1}^{n} a_i \hat{p}_i$$
$$V(\hat{\phi}) = \frac{1}{N}(-\hat{\phi}^2 + \sum_i a_i^2 \hat{p}_i)$$

This example is relevant for chess engine testing [2] in which case ϕ would stand for the expected score of a match. For the naive trinomial model one takes $(p_1, p_2, p_3) = (w, d, l)$ and $a_1 = 1$, $a_2 = 1/2$, $a_3 = 0$ whereas in the 5-nomial model (for paired games with reversed colors) one takes $(p_1, p_2, p_3, p_4, p_5) = (p_2, p_{3/2}, p_1, p_{1/2}, p_0)$ with $a_1 = 1$, $a_2 = 3/4$, $a_3 = 1/2$, $a_4 = 1/4$, $a_5 = 0$. Note that in the 5-nomial model N is the number of games divided by two (one trial consists of two games).

3. Derivation of the approximation

For simplicity we will give the derivation for one particular choice of $V(\hat{\phi})$. One may check that the relative change in the right hand side of (2.1), when replacing one $V(\hat{\phi})$ by another, goes to zero when N goes to infinity.

In order to verify (2.1) the first mission is to calculate $\hat{\underline{\theta}}_i$. Using Lagrange multipliers we see that we have to solve ($i \in \{0, 1\}$)

$$\nabla_{\underline{\theta}} L(\hat{\underline{\theta}}_i, \underline{x}) = \lambda \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}}_i)$$
$$\phi(\hat{\underline{\theta}}_i) = \phi_i$$

If $\hat{\phi} = \phi_i$ then $\lambda = 0$, $\hat{\underline{\theta}}_i = \hat{\underline{\theta}}$. We will assume that $\hat{\phi}$ is close to ϕ_i so that λ is small. We get in first order (H = Hessian)

$$(H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x}) - \lambda H_{\underline{\theta}} \phi(\hat{\underline{\theta}})) \cdot (\hat{\underline{\theta}}_i - \hat{\underline{\theta}}) = \lambda \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})$$
$$\nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})^t \cdot (\hat{\underline{\theta}}_i - \hat{\underline{\theta}}) = \phi_i - \hat{\phi}$$

and hence

$$\hat{\underline{\theta}}_i - \hat{\underline{\theta}} = \lambda H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x})^{-1} \cdot \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})$$
$$\lambda \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})^t \cdot H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x})^{-1} \cdot \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}}) = \phi_i - \hat{\phi} \tag{3.1}$$

Write $V(\hat{\phi}) = -\nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})^t \cdot H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x})^{-1} \cdot \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})$. It is well-known that $V(\hat{\phi})$ is an approximation for the variance $\mathrm{Var}(\hat{\phi})$ of $\hat{\phi}$. Then we get by eliminating λ from (3.1)

$$\hat{\underline{\theta}}_i - \hat{\underline{\theta}} = -\frac{\phi_i - \hat{\phi}}{V(\hat{\phi})} H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x})^{-1} \cdot \nabla_{\underline{\theta}} \phi(\hat{\underline{\theta}})$$

Now we have (using the fact that $\hat{\underline{\theta}}$ is extremal for $L(\underline{\theta}, \underline{x})$)

$$L(\hat{\underline{\theta}}_i, \underline{x}) \cong L(\hat{\underline{\theta}}, \underline{x}) + \frac{1}{2}(\hat{\underline{\theta}}_i - \hat{\underline{\theta}})^t \cdot H_{\underline{\theta}} L(\hat{\underline{\theta}}, \underline{x}) \cdot (\hat{\underline{\theta}}_i - \hat{\underline{\theta}})$$
$$\cong L(\hat{\underline{\theta}}, \underline{x}) - \frac{1}{2}\frac{(\phi_i - \hat{\phi})^2}{V(\hat{\phi})}$$

Substituting this in (1.1) yields what we want.

References

1. Xiaoou Li, Jingchen Liu, and Zhiliang Ying, *Generalized Sequential Probability Ratio Test for Separate Families of Hypotheses*, http://stat.columbia.edu/~jcliu/paper/GSPRT_SQA3.pdf.
2. Chess Programming WIKI, *Match statistics*, https://chessprogramming.wikispaces.com/Match+Statistics.