ABSTRACT

Calibration with Many Checking Rules

Alvaro Sandroni Northwestern University

In each period, one of finitely many possible states will occur. To simplify, assume just two possible states, a and b. We must predict the probability that state a will be realized next period. We must do so without *any* prior knowledge about the underlying process determining the states in each period. Given sufficiently many observations, how accurate can our forecasts be?

We are not interested in the accuracy of the forecasts with respect to a particular realization of the states because we do not know which sequence of states will be realized. So, we seek guarantees on the accuracy of the forecasting method for *every* possible infinite string of data on $\{a, b\}^{\infty}$. However, it is by no means obvious what guarantees can be offered. So, we first formalize different notions under which the forecasts may (or may not) comply with the data.

Let a *history-based checking rule* be an arbitrary function of finite histories to $\{0, 1\}$. Given a history-based checking rule, and a forecasting scheme, let a *forecast-based checking rule* be a function equal to one if the history-based checking rule is one and the forecast of a is p and equal to zero otherwise. History-based checking rules and forecast-based checking rules are said to be *active* (or *inactive*) if they assume the value one (or zero). The checking rules determine the subsequences in which the time average forecasts will be compared with the empirical frequencies.

The *calibration score* of a checking rule (history or forecast based) is the difference between the frequency of a and the time average of the forecasts of a in the periods that the checking rule was active. A forecasting scheme *calibrates* a checking rule if, given any data, the calibration score is eventually zero.

Given this setup three questions are natural. What kinds of checking rules can be calibrated? What sets of checking rules can be calibrated simultaneously? What are the consequences of being able to calibrate many checking rules?