An extension of Bayesian methods to the change-point problem in regression

Jason Wyse[1] and Gabrielle E. Kelly[1]

[1] School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4, Ireland

We consider data following a simple linear regression model with two regimes, the change in regime occurring at an unknown point called the change-point. One example of such data is breath-by-breath values of the rate of carbon dioxide output ($\dot{V}CO_2$) regressed on the rate of oxygen intake ($\dot{V}O_2$) for a subject undertaking an incremental exercise test to exhaustion. Both the slope and intercept of the regression line can change when the subject reaches a critical level of $\dot{V}O_2$ called the gas exchange threshold (GET). An estimate and a confidence interval for the time at which this occurs, the change-point, is sought. The GET may be used as a measure of fitness, and thus provides an alternative to more invasive fitness testing procedures such as measurement of the lactate threshold using a blood test. Some illnesses (Cystic Fibrosis, for example) require regular fitness tests to monitor the sufferer, and a non-invasive method is preferable in these cases. A confidence interval for the GET is necessary to monitor when real changes in fitness occur.

Since the data consists of repeated measurements on two variables on the same subject over time, the incorporation of some correlation structure in regression errors is necessary. It is also appropriate to consider heteroscedasticity in regression errors. This is due to a subject's breathing becoming gradually more unstable as they approach exhaustion. Using traditional likelihood and Bayesian methods, such model considerations lead to intractable expressions when one wants to obtain a confidence interval for the change-point. Some likelihood methods have been developed but with quite restrictive assumptions, and extensions of these to more complex models is not trivial. We have, however, extended previous authors' work using Bayesian methods, by an implementation of the Metropolis-Hastings algorithm. Thus we have found a confidence interval for the estimated change-point, by providing an approximation to its posterior distribution.

The breath-by-breath data of three subjects undertaking an incremental exercise test are examined. We present the results of models fitted incorporating correlated errors, heteroscedasticity and a combination of both. We discuss briefly the extension of results to non-normal regression errors.

Another part of the change-point problem is testing if a change-point exists. The use of the Bayesian Information Criterion (BIC) is not theoretically justified for comparing a model with a change-point to a model without. The likelihood ratio test for this problem has been developed under certain assumptions, but is highly technical. For this reason, we examine a method using cross-validation predictive densities and the pseudo-Bayes factor in the case of normal i.i.d. regression errors. We present the form of the cross-validation densities and the pseudo-Bayes factor and discuss how it performs in choosing the correct model for various regime-to-regime slope ratios and error variances using results obtained from a small simulation study. We also report the pseudo-Bayes factor obtained for the three subjects mentioned previously and comment on its performance.