Model Selection Properties

1 Consistency - Probability of over-fitting

In this exercise we would like to prove that the Mallow C_p criterion is over-fitting asymptotically, thus not consistent in the strong sense. In order to understand the more general implications of this result, we work with a λ_n parameter which, in the special C_p case, is equal to 2. We limit our reasoning to the linear model case where, by default, AIC and C_p expressions coincide. After having read the section of the e-book devoted to model selection properties, perform the following steps to complete the proof:

(a) Derive the small small sample distribution of the quantity $(C_{p,L+K} - C_{p,K})$ where $C_{p,K}$ is the C_p value for the supposed true model and $C_{p,L+K}$ is the C_p value of a generic over-fitted model which selects L regressors more than the true number (i.e. K). For doing so, use the fact that: $C_{p,K} = \frac{SSE_K}{s^2_{k^\star}} - n + \lambda_n K$ and $s^2_{k^\star} = \frac{SSE_{k^\star}}{n-k^\star}$ meaning that we are estimating the variance of some larger model with $k^\star$ regressors (e.g. full model).

(Hint: Use the Cochran theorem to derive the distribution of SSE and then a well-known result on the ratio of specific independent random variables)

(b) Thanks to the previous step, retrieve the small sample probability of over-fitting which has to depend on λ_n.

(c) Let $n \to +\infty$ and derive the asymptotic probability of over-fitting thanks to the Slutsky theorem. *(Hint: Use the fact that $s^2_{k^\star}$ is a consistent estimator of σ^2)*

(d) For $\lambda_n = [0\ 2\ log(n)]$ derive the asymptotic probabilities of over-fitting. Show that Mallow C_p is not strong consistent and that BIC is not over-fitting asymptotically. What can you conclude on the role of the penalty in this specific situation?