MATH297 Report

Kuangyi Yang, A53083212

Problem

This problem is sent from Mr Dart Lindsley, he provides following graphs, the blue line and box plots represents the results of a model run multiple times ad two different variables("Pool" and "Nuclei"). The Horizontal lines represents experimental results. He needs to be able to say how well each of the monte-carlo results matches the experimental results. More specifically, what are the odds that each scenario could have generated the experimental results.

This is a biomedical problem, our understanding about the problem in statistics is, now we have two variable$\theta_1$andθ_2("Pool" and "Nuclei"), we are not sure about the exact form of function f, but we have got lots of simulations of f and real data. We want to find out the (θ_1, θ_2) such that the probability of function applied in real data based on θ_1, θ_2 is the largest, in other words, we want to find out the

$$(\theta_1, \theta_2) = argmax P(y_{real}|\theta_1, \theta_2)$$

Solution

This is a maximum likelihood estimation problem, since we have $y_1, y_2, ..., y_9$ so we consider they are independent, first we can estimate each y_i separately, for each y_i, for a group of (θ_1, θ_2), do simulation for n times and obtain the pseudo $y_i's$. Then use kernel density estimation to obtain the estimated density of these $ny_i's$ and compute the probability under this estimated density.

Since each y_i is independent so we can get their "joint probability" by multiplying all "marginal probability", which means we can get

$$P(y_{real}|\theta_1, \theta_2) = P(y_{1,real}|\theta_1, \theta_2)P(y_{2,real}|\theta_1, \theta_2)...P(y_{3,real}|\theta_1, \theta_2)$$

and compare probability based on different (θ_1, θ_2), pick out the highest probability, the corresponding pair of θ_1, θ_2 is the best.

Conclusion

This method is discussed in MATH 297 class on June 1st by Professor David Meyer and students from MathStorm Group.