Meeting Summary for the Live Professor Q&A on Statistical Inference at the 2025 Astrostatistics Summer School (06/02/2025)

1 Quick recap

In this QA session following a recording lecture, Hyungsuk discussed various statistical concepts including likelihood ratio tests, model selection criteria, and nested models. The discussion covered topics such as the relationship between Hubble constant, distance, and velocity, the use of Hessian matrices in statistical analysis, and the construction of confidence intervals. Hyungsuk also addressed questions about statistical significance, likelihood functions, and the reliability of asymptotic normality in constructing confidence intervals, providing clarifications on various technical aspects throughout the session.

2 Summary

2.1 QA Session and Audience Questions

Hyungsuk conducted a QA session following a recording lecture and invited questions from attendees. He mentioned receiving a personal message but did not elaborate. Hyungsuk requested that any questions be shared aloud so everyone could hear them.

2.2 Equivalence of Testing Procedures

Hyungsuk explained that the likelihood ratio test and the test based on the Neyman-Pearson Lemma are equivalent procedures, with the latter being a simplified version that ultimately leads to the same testing procedure. He clarified that the rejection region for the MLE part is consistent with the direction of the alternative hypothesis, and there is a formal proof for this relationship.

2.3 Model Selection Criteria: AIC vs BIC

Hyungsuk explained the differences between the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), noting that while both criteria penalize model complexity, BIC penalizes more heavily due to its log(N) coefficient, where N is the data size. He clarified that when both criteria agree, it provides strong evidence for a preferred

model, but when they differ, one should consider the expected model complexity and data size to determine the more beneficial approach. Hyungsuk also defined nested models, explaining that one model is a nested model of another when it is a special case of the other, and provided examples, including a normal distribution with unknown mean and variance, to illustrate this concept.

2.4 Hubble Constant and Statistical Models

Shlok asked Hyungsuk about the relationship between the Hubble constant, distance, and velocity, and Hyungsuk explained that if epsilon (measurement error) follows a normal distribution, then the velocity will also be normally distributed, with the mean shifted by H0 times distance. Hyungsuk clarified that this statistical model can incorporate physical knowledge, and provided an example of Poisson distribution in astrophysics. Jay inquired about the use of the Hessian matrix in estimating variance, and Hyungsuk clarified that he used it to estimate the standard error of H0, not the covariance matrix of the data set.

2.5 Confidence Interval and Alpha Discussion

Jay and Hyungsuk discussed the process of determining the confidence interval and the significance level, or alpha, which is typically set by researchers. Hyungsuk explained that alpha is often set at 0.05 for a 95% confidence level, but researchers can choose their own threshold based on their tolerance for false positives. They also touched on the concept of hypothesis testing and the importance of considering alternative explanations for observed signals in scientific research.

2.6 Statistical Significance of Test Results

Hyungsuk explained the statistical significance of the data, noting that the test statistics were 5.1 and 4.6 sigma, indicating strong evidence against the null hypothesis. He clarified that these values account for the standard error and uncertainty of the estimators, allowing for an apple-to-apples comparison of the evidence.

2.7 Correcting Likelihood Function Notation

Hyungsuk explained to Kyle that the likelihood function should be written as a function of the parameter, not the data, to maintain mathematical accuracy and correct notation.

He clarified that while some papers use the alternative notation, it's important to understand that likelihood is fundamentally a function of the parameter. Kyle acknowledged and understood Hyungsuk's explanation.

2.8 Understanding Confidence Interval Estimation

Hyungsuk discussed the reliability of using asymptotic normality for constructing confidence intervals when data do not follow a normal distribution, explaining that the central limit theorem allows for this approximation when data size is large. He also clarified the difference between the Hessian matrix and covariance matrix, explaining that the Hessian measures the curvature of the likelihood function and is used to estimate the uncertainty of parameter estimates, while the covariance matrix represents the variability of the estimated parameters themselves. Shlok asked a follow-up question about the non-diagonal elements of a 2x2 Hessian matrix, which Hyungsuk confirmed represent the covariance between the two parameters being estimated.