Cecilia Albert-Black
STAT 563
11/4/2025
Project 2

Conclusion & Discussion

Q1: Visualization and interpretation of the logistic distribution

By plotting the density function of the logistic distribution under different parameter settings, I was able to visualize how the location (μ) and scale (s) parameters influence its shape. When holding the location constant at 0 and varying the scale (s = 0.5, 1, 2), I observed that increasing the scale parameter caused the bell curve to flatten and the tails to become heavier. Conversely, smaller scale values produced a sharper peak around the mean, indicating that most of the probability mass was concentrated near the location constant. When holding the scale constant at 1 and shifting the location (μ = -2, 0, 2), the entire curve moved horizontally without changing shape—illustrating how the location parameter controls the center of the distribution, which also represents both the mean and median.

The variance of the standard logistic distribution ($\pi^2/3 \approx 3.29$) is notably larger than that of the standard normal distribution (1), which reflects its heavier tails and greater probability of extreme values. This property makes the logistic distribution more robust to outliers, as it accommodates larger deviations from the mean with less distortion.

Beyond its statistical interpretation, the logistic distribution plays a foundational role in modeling probability-based outcomes. In statistics, it forms the basis of generalized linear models (GLMs) through the logit link function, which is widely used to model binary outcomes such as presence/absence or success/failure. Although I am still new to machine learning, I can see how the logistic function extends widely into this field—particularly for classification tasks that involve predicting whether a condition or feature is present or absent. It is fascinating that such a simple and elegant distribution underlies so many modern predictive modeling techniques, and I look forward to exploring its applications further as I continue learning.

Q2: Finding the Maximum Likelihood Estimates (MLEs)

The likelihood function represents the joint probability (or probability density) of the observed data, viewed as a function of the model parameters rather than the data itself. Its purpose is to quantify how likely the observed data are for different parameter values. Maximizing the likelihood function therefore means finding the parameter values (e.g., μ) that make the observed data most probable under the assumed model.

As previously discussed, the logistic distribution is defined by a location parameter (μ) and a scale (s) parameter. The log-likelihood function is used to estimate these

parameters through maximum likelihood estimation (MLE). Mathematically, the likelihood function multiplies the individual probabilities (densities) of all observations given μ and s. Taking the logarithm of this product transforms it into a sum—simplifying the calculations—while preserving the location of the maximum. The MLEs of μ and s are the values that maximize this log-likelihood function.

One approach to computing these MLEs is the Fisher scoring algorithm, which iteratively updates parameter estimates using information from the data. In this method, each observation is assigned a weight/score between 0 and 1, reflecting its influence on the parameter estimates. Observations closer to the current estimate of μ (1) receive higher weights, indicating they are more "trustworthy," while outlying observations receive lower weights (0). This weighting makes the MLEs robust to outliers, as extreme values have less impact on the final estimates.

Alternatively, MLEs can be estimated using non-parametric, empirical methods, such as bootstrapping. This approach does not assume a specific parametric form of the underlying distribution and instead determines parameter estimates by resampling the observed data. While it is less efficient than Fisher scoring when the true distribution is known, bootstrapping is more robust for datasets where the underlying distribution is unknown or uncertain.

When I calculated confidence intervals (CIs) for μ and s using both methods, I found that the Fisher information-based CIs contained the true parameter values, whereas the non-parametric bootstrap CIs did not. This difference likely arises because the Fisher scoring method incorporates the known logistic distribution, allowing it to account for the mean being different from the median. In contrast, the non-parametric method implicitly assumes symmetry between the mean and median, which can reduce accuracy when the data follow a specific distribution.

Maximum likelihood estimation has broad applications in both statistics and machine learning. In statistics, MLEs are used to fit parametric models and conduct hypothesis testing using likelihood-based methods. In machine learning, MLEs underlie logistic regression, neural networks, and probabilistic models, forming a foundation for Bayesian inference and posterior estimation.

Q3: Wald Confidence Interval and Percentile Bootstrap Confidence Interval for the Mean

The Wald method is a parametric approach for estimating CIs that relies on a mathematical formula assuming the sample mean follows a normal distribution. This assumption generally holds well for large samples that are not heavily skewed or affected by outliers. The Wald CI is fast and straightforward to compute, as it uses the estimated variance of the mean and a critical value from the standard normal distribution.

In contrast, the percentile bootstrap CI does not rely on any distributional assumptions. Instead, it repeatedly resamples the observed data with replacement to

generate a distribution of sample means. The 2.5th and 97.5th percentiles of these bootstrap means then form the 95% CI. This method reflects the variability inherent in the actual dataset rather than assuming a specific theoretical distribution.

For data following a logistic distribution, the lower bounds of the Wald and bootstrap CIs differed. Because the logistic distribution has longer tails than a normal distribution, the Wald method — which relies on normality — produced a lower CI that was smaller than the bootstrap lower bound, while the bootstrap CI extended higher. This illustrates a key difference: the Wald method "trusts theory," assuming the mean's sampling distribution is normal, whereas the bootstrap method "trusts the data," adapting to its observed shape.

Overall, the bootstrap CI proved to be more flexible and is often more reliable, particularly for small samples or skewed data, while the Wald CI is quicker and simpler but can be misleading if its underlying assumptions do not hold.

Q4: Bootstrap Distribution of the Correlation Matrix

In this section of the project, I investigated both the analytical and bootstrap distributions of the correlation matrix to understand the variability and uncertainty in estimated correlations. The analytical distribution of correlations relies on theoretical results from probability and statistics. For example, when two variables are jointly normal, the sampling distribution of their Pearson correlation can be approximated using Fisher's z-transformation, which provides CIs under assumptions of normality and large sample size. These CIs are convenient and widely used, but they depend heavily on the validity of these assumptions and do not account for potential non-normality, outliers, or dependencies among multiple correlations in a matrix.

By contrast, the bootstrap distribution offers an empirical, data-driven alternative. By repeatedly resampling the observed dataset with replacement, the correlation matrix reflects the actual variability in the data. From these bootstrap samples, the CIs are computed directly from the empirical distribution without relying on strong parametric assumptions. This approach also allows visualization of the shape of the distribution—revealing skewness, multimodality, or other features that the analytical approach may overlook.

To explore the uncertainty within the correlation matrix, we applied slice sampling to a single correlation element after bootstrapping the full matrix. I did not understand how this worked and what its purpose was. Thus, I conducted a little background research and found that compared to Fisher-z confidence intervals, slice-sampled marginals provide a more truthful representation of the variability in the correlation matrix, capturing both the empirical uncertainty and the structural geometry imposed by the matrix constraints.

Conclusion

In this project, I explored multiple aspects of statistical inference, from understanding the logistic distribution to estimating parameters and quantifying uncertainty in correlation matrices. Visualizing the logistic distribution helped me see how the location and scale parameters shape its form and highlighted its heavier tails compared to the normal distribution, emphasizing robustness to outliers. Using maximum likelihood estimation, both via Fisher scoring and bootstrap methods, I learned how assumptions about the underlying distribution influence parameter estimates and confidence intervals, with Fisher scoring providing greater accuracy when the parametric form is known and bootstrapping offering a flexible, assumption-free alternative. Comparing Wald and bootstrap confidence intervals for the mean further illustrated the trade-off between theoretical simplicity and empirical reliability. Finally, analyzing the correlation matrix through both analytical and bootstrap distributions, complemented by slice sampling, allowed me to appreciate the complexity of uncertainty in high-dimensional data, including how structural constraints impact the variability of individual correlations. Overall, the project reinforced the importance of combining theoretical understanding with empirical methods to obtain robust and insightful statistical conclusions.