Stat 591 – Homework 01 Due: Wednesday 09/11

Your group should submit a write-up that includes solutions the problems stated below, any relevant pictures, and a print-out of your computer code and relevant output.

An interesting statistical problem is the assessment of agreement. For example, consider a population of items, each of which can be classified as either "good" (=1) or "bad" (=0). Suppose, also, that there are two raters, A and B, to rate a given item as good or bad. The goal is to assess if the A and B ratings agree. To investigate this, we take an independent sample of n items from the given population, to be rated by *both* raters. There are two classifications and two raters, so there are four possible outcomes, and the results are summarized in a 2×2 table:

		Rater B	
		Bad (=0)	Good (=1)
Rater A	Bad (=0)	X_{00}	X_{01}
	Good (=1)	X_{10}	X_{11}

Note that $X_{00} + X_{01} + X_{10} + X_{11} \equiv n$. Here, for example, X_{00} is the number of sample items rated "bad" by both raters. The two raters exactly agree if $X_{00} + X_{11} = n$ or equivalently, if $X_{01} + X_{10} = 0$. However, the two raters could effectively agree, but have had a few chance disagreements for this particular sample of items. That is, there may be no *significant disagreement.*[1]

For a statistical model, consider a corresponding 2×2 table of parameters:

		Rater B	
		Bad (=0)	Good (=1)
Rater A	Bad (=0)	θ_{00}	θ_{01}
	Good (=1)	θ_{10}	θ_{11}

Note that $\theta_{00} + \theta_{01} + \theta_{10} + \theta_{11} \equiv 1$. Here, for example, θ_{00} denotes the probability that both raters would classify an item as "bad." There's nothing special about the 2×2 tabular structure, so let's rewrite the data and parameters as

$$X = (X_{00}, X_{01}, X_{10}, X_{11}) \equiv (X_1, X_2, X_3, X_4)$$
$$\theta = (\theta_{00}, \theta_{01}, \theta_{10}, \theta_{11}) \equiv (\theta_1, \theta_2, \theta_3, \theta_4).$$

For the model, we shall assume $X = (X_1, X_2, X_3, X_4) \sim \mathsf{Mult}_4(n, \theta)$, i.e., a four-dimensional multinomial distribution with sample size n and category probabilities $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$. The corresponding likelihood function is

$$L(\theta) = c(X)\theta_1^{X_1}\theta_2^{X_2}\theta_3^{X_3}(1 - \theta_1 - \theta_2 - \theta_3)^{n - X_1 - X_2 - X_3},$$

where $c(X)$ is a function of data but not the parameter; note that the constraints on X and θ have been explicitly enforced, so that there are effectively only three parameters.

[1] Though this particular setup may not seem so interesting, there are important examples where this kind of problem arises. Our own Professor Hedayat has a recent book (Springer 2012, with L. Lin and W. Wu) on non-Bayesian methods for assessing agreement. There are also a number of nice examples in the book by L. Broemeling (Chapman–Hall 2009) on Bayesian methods for agreement.

For assessing agreement, a quantity of interest is the "kappa" coefficient

$$\kappa = g(\theta) := \frac{(\theta_1 + \theta_4) - \{(\theta_1 + \theta_2)(\theta_1 + \theta_3) + (\theta_3 + \theta_4)(\theta_2 + \theta_4)\}}{1 - \{(\theta_1 + \theta_2)(\theta_1 + \theta_3) + (\theta_3 + \theta_4)(\theta_2 + \theta_4)\}}.$$

If there is exact agreement, then $\theta_1 + \theta_4 = 1$ and κ is exactly 1, and if there is "effective agreement," so that $\theta_2 + \theta_3 \approx 0$, then κ will be approximately 1. So, the extent of agreement can be measured (at least intuitively) by the proximity of κ to 1. Therefore, we are interested in producing a confidence interval for κ.

1. Show that the maximum likelihood estimator $\hat{\theta}$ of θ has $\hat{\theta}_j = X_j/n$, $j = 1, \ldots, 4$. All you need to do is check that $\hat{\theta}$ above satisfies the likelihood equation. [Hint: Think of $L(\theta)$ a function of only $(\theta_1, \theta_2, \theta_3)$, and find the gradient of the log-likelihood.]

2. A Central Limit Theorem holds for $\hat{\theta}$.[2] That is, $n^{1/2}(\hat{\theta} - \theta) \to \mathsf{N}_4(0, \Sigma_\theta)$, in distribution, where Σ_θ is a 4×4 covariance matrix that depends on θ. Find Σ_θ. [Hint: Use the formulae provided in Appendix A.2.5 in the textbook.]

3. For this problem, and those that follow, suppose that the observed data is

$$X = (22, 15, 33, 30),$$

so that $n = 100$ (Broemeling 2009, p. 39).

We can get an approximate variance of $\hat{\kappa} = g(\hat{\theta})$ using the result in Problem 2b above, along with the Delta Theorem. Let $\nabla g(\theta)$ be the (4×1) gradient of $g(\theta)$. Then the Delta Theorem says the asymptotically approximate variance is $\mathsf{V}(\hat{\kappa}) = \frac{1}{n} \nabla g(\theta)^\top \, \Sigma_\theta \, \nabla g(\theta)$. You can estimate $\mathsf{V}(\hat{\kappa})$ by plugging in $\hat{\theta}$ on the right-hand side. Find the estimate of $\mathsf{V}(\hat{\kappa})$ for the data above, and give a 90% confidence interval for κ. [Hint: The matrix multiplication in the variance formula is too messy to do by hand, so you should do this numerically; you can calculate the gradient $\nabla g(\theta)|_{\theta=\hat{\theta}}$ analytically or numerically, whichever you prefer.]

4. Using the data in Problem 3, find a bootstrap 90% confidence interval for κ, and compare with your Delta Theorem interval above. [Hint: In this problem, the parametric and nonparametric bootstrap are the same. Just sample, say, $B = 3000$ tables from a multinomial distribution[3] with parameters n and $\hat{\theta}$; for each sampled table, $X_b^\star = (X_{b1}^\star, X_{b2}^\star, X_{b3}^\star, X_{b4}^\star)$, compute the maximum likelihood estimator, $\hat{\theta}_b^\star$, and then the corresponding $\hat{\kappa}_b^\star = g(\hat{\theta}_b^\star)$.]

5. Draw a histogram of the bootstrap distribution of $\hat{\kappa}$ from Problem 4, with the normal density corresponding to the Delta Theorem approximation in Problem 3 overlaid. Compare the two plots.

6. Do you think there is agreement in this case? Do the two analyses (MLE-based and bootstrap-based) give the same conclusion, or different? Is the result consistent with your intuition based on looking at the observed data?

[2]You can see this intuitively since each $\hat{\theta}_j$ is a nice sample proportion, but we'll not worry about verifying this rigorously. You can prove it using characteristic functions.

[3]Here's a naive way to sample $X = (X_1, X_2, X_3, X_4)$ from the $\mathsf{Mult}_4(n, \theta)$ distribution using R:

```
X <- as.numeric(table(sample(1:4, size=n, prob=theta, replace=TRUE)))
```