Stat 314, Problem Set 2
Autumn 2013/14

Due in class on paper, Thursday November 8, 2013
This problem set can be done in groups of up to three students.

1. Problems from the text: 3.11, 3.16, 3.22

   (3.22 is a followup to 3.21, so reading 3.21 might make 3.22 easier to understand)

2. Lemma 4.1 shows that if the hulls $\mathcal{H}_n^+$ and $\mathcal{H}_n^-$ intersect, then $0 \in \mathcal{H}_n$. Does $0 \in \mathcal{H}_n$ imply that $\mathcal{H}_n^+ \cap \mathcal{H}_n^- \neq \varnothing$?

3. Consider linear logistic regression of $Y \in \{0, 1\}$ on $X \in \mathbb{R}$, with

$$\Pr(Y = 1 \mid X = x) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)}.$$

   (a) Find the estimating equations for β_0 and β_1.

   (b) Find their estimating equations if we use Mallows' robust likelihood and $\varepsilon = 0.01$.

   (c) You can get the iris data in R by typing `iris`. Find the coefficients in a logistic regression of $Y \in \{\text{versicolor}, \text{virginica}\}$ versus $X = $ Sepal.Length. For definiteness, take virginica $= 1$. Plot confidence contours for (β_0, β_1) using empirical likelihood on this data. Show contours for 90, 95, 99 and 99.5 percent confidence using the $\chi_{(2)}^2$ calibration.

   (d) Compute contours for the robust likelihood on the same grid. (It might not include the entire outer region.)

   (e) Add some label error: change the flower with smallest Petal Length to be a virginica and the one with the largest Petal Length to be a versicolor. If Mallow's method is working well, the estimating equations from the robust likelihood should be a lot less affected. Are they?

Note: Petal Length perfectly separates setosa from the other two flower types. Usually perfect separation is numerically bad for logistic regression. The MLE is unbounded and confidence intervals are affected. That probably happens with the estimating equations for Mallows too. But maybe the confidence regions come out more interesting. Exploring this issue might make a suitable project.