STAT 665 - Assignment 3 - due date is on course outline

From Chapter 6:

1. 1.5
2. Prove "Claim 2" in the asymptotic normality proof for U-statistics.
3. 2.2 (Make the correction $h\left(\hat{F}_n\right) = n^{-1}\sum\left|X_i - \bar{X}\right|$.)
4. In Example 6.3.2 check that $\dot{h}(F_0; F_1) = E_{F_1}\left[\psi(X)\right]$ for some function ψ, necessarily $= IF$.
5. Suppose that a functional $\theta(F)$ is defined implicitly by an equation $\lambda_F\left(\theta\right) = 0$. Show that the influence function is

$$IF\left(x; F, \theta\right) = -\frac{\frac{d}{d\varepsilon}\lambda_{F_\varepsilon}\left(\theta\right)_{|\varepsilon=0, F_1=\delta x}}{\lambda'_{F_0}\left(\theta\right)}.$$

6. 5.1
7. 5.4 Use $a = \Phi^{-1}(.75)$. Present the average of the 10 values obtained, for each (ε, τ) combination. This is best done on R, in which case I would like you to e-mail me your program. And – although there are built-in routines which will do bootstrapping, I want you to program this yourself. I should be able to run your .r file and, without doing anything else, see output consisting of the 4 averages arranged in a a 2×2 matrix, with rows and columns labelled appropriately. (As a check on your program first try $\tau = 1$, in which case $\lambda_n\left(F\right) =..?$)

From Chapter 7:

8. 2.2
9. 3.5
10. 5.10 (There is a misprint in (7.9.8.))
11. 7.2 (iii) Also, show that the acceptance region of the LR test is an interval of values of $m_2 = \sum X_i^2/n$, and characterize the endpoints of this interval.
12. 7.8. "Show directly ... " means that you should verify that $R_n^2 = W_n^2 =$ Pearson's χ^2 from Lecture 14.

...over

13. Consider the two-sample Poisson problem, for which the Wald and LR tests were derived in class.
(i) Show that an application of (7.7.68) gives, as the scores statistic for *known* λ,

$$R_n^2(\lambda) = \frac{1}{m+n} \dot{l}_1'(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2)\, \mathbf{I}_{11}^{-1}(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2)\, \dot{l}_1(\boldsymbol{\theta}_{1,0}, \boldsymbol{\theta}_2) = \left(\frac{n}{(m+n)p}\right) Z_n^2(\lambda, \lambda),$$

where $Z_n(\lambda_1, \lambda_2) = \frac{\sqrt{n}(\bar{Y} - \lambda_1)}{\sqrt{\lambda_2}}$. Thus under the null hypothesis $Z_n(\lambda, \lambda) \xrightarrow{\mathcal{L}} N(0,1)$ and $R_n^2 \xrightarrow{\mathcal{L}} \chi_1^2$.
(ii) In practice one would of course want to replace λ_1, λ_2 by consistent estimates $\hat{\lambda}_1, \hat{\lambda}_2$. Find such estimates for which $Z_n\left(\hat{\lambda}_1, \hat{\lambda}_2\right) \xrightarrow{\mathcal{L}} N(0,1)$. Prove that the asymptotic normality holds. Also, show that $\hat{\lambda}_1, \hat{\lambda}_2$ can be chosen in such a way that the Scores test coincides with either version of Wald's test derived in class.

14. Let $\mathcal{F}$ be a family of d.f.s F, symmetric around 0, with differentiable densities $f(x) = f(-x)$. Suppose that we observe a sample $\{X_i\}_{i=1}^n$, where $P(X_i - \theta \le x) = F(x)$ for some $F \in \mathcal{F}$. Then θ is a "location parameter". An "M-estimate" $\hat{\theta}_n = h(\hat{F}_n)$ is defined as a solution to the equation

$$\sum_{i=1}^n \psi(X_i - \theta) = 0.$$

Thus the estimand $\theta = h(F)$ is defined implicitly through the equation

$$E_F[\psi(X - h(F))] = 0.$$

(i) What choice of ψ is required if $h(F)$ is to be (a) the population mean? (b) the population median? (c) the MLE?

(ii) Assume that ψ is differentiable. Fix $F_0 \in \mathcal{F}$. Write out the equation defining $h(F_\varepsilon)$, differentiate implicitly and solve for $\dot{h}(F_0, F_1 = \delta_x)$ to obtain the influence function

$$IF(x; \psi, F_0) = \frac{\psi(x - \theta)}{E_{F_0}[\psi'(X - \theta)]}.$$

Verify that $E_{F_0}[IF(X)] = 0$.

(iii) Write down the asymptotic normal distribution of the M-estimate. Although this will not apply directly to the median (since then ψ is not differentiable), one can evaluate the denominator of IF using integration by parts; the result then relies only on the differentiability of F_0. Show that the result of this gives the asymptotic distribution of the sample median, as derived in a completely different manner as an application of the Berry-Esséen Theorem.