HW5, Solution to HW5 Problems 1 & 2

#(1) This problem is really intended to show you that the standard \( p \)'th quantile estimator approximately solving the equation

\[
F_n(\hat{\tau}) = n^{-1} \sum_{i=1}^{n} I_{[X_i \leq \hat{\tau}]} = p
\]

is actually representable as a minimum contrast estimator. The function

\[
\rho(x, \tau) = (x - \tau) (I_{[x > \tau]} - 1 + p)
\]

is a contrast function whenever the quantile is unique (which does not even require continuity of the distribution function). But for differentiability, we do need continuity of the distribution function, in which case we check, with respect to the true distribution of the samples of random variables \( X_i, \ i = 1 \ldots, n \),

\[
E(\rho(X, t)) = -\int_{t}^{\infty} (x-t) d(1-F(x)) - (1-p)(E(X) - t)
\]

\[
= \int_{t}^{\infty} (1-F(x)) dx + 1 - p
\]

This function is differentiable, and its derivative \( F'(t) - p \) is zero (by assumption) if and only if \( t = \tau = F^{-1}(p) \). This completes the contrast-function verification. As remarked in class, the second derivative of the contrast function is a.e. 0, so the asymptotic normality cannot be made to follow from the Taylor series expansions as in the main asymptotic-normality theorems for minimum-contrast estimators, but the asymptotic normality does hold if the density \( f(\tau) \) of the random variables \( X_i \) exists and is positive at \( \tau = F^{-1}(p) \), in which case

\[
\sqrt{n}(\hat{\tau} - \tau) \overset{D}{\to} N(0, p(1-p)/f^2(\tau))
\]

#(2) (a) Here the main work is done already in the book’s equation (5.49) with \( \mu(w) = e^w/(1+e^w) \), \( g(r) = \mu^{-1}(r) = \log(r/(1-r)) \), and with \( \phi_i = \phi/t_i \equiv 1 \) for all \( i \). This gives the likelihood equation (for the parameter \( \theta = (a, b) \)) as being of the form

\[
\sum_{i=1}^{n} g(X_i) (Y_i - \mu(\theta' m(X_i))) = 0 , \quad g(x) = m(x) \equiv \frac{1}{x}
\]

(b) When \( E(g(X)^{\otimes 2}) \) exists and is assumed nonsingular as a \( 2 \times 2 \) matrix, the root of the above estimating equation for general \( g \) exists and is unique by convexity for all sufficiently large \( n \), and the general contrast-estimation or
estimating-equation theorems imply that as $n \to \infty$ the solution has $\sqrt{n}(\hat{\theta} - \theta)$ asymptotically normal with mean $0$ and variance $A^{-1}B A^{-1 \text{tr}}$, where

$$A = E(\mu(X)(1 - \mu(X)) g(X) m(X)^{\text{tr}}), \quad B = E(\mu(X)(1 - \mu(X)) g(X)^{\otimes 2})$$

All of this reasoning holds just the same whether the distribution of the r.v.’s $X_i$ (which means the probability $p = E(X_i)$ in the binary case) is assumed known or unknown.

For part (c), the issue is to show that the asymptotic variance matrix $A^{-1}B A^{-1 \text{tr}}$ is made smallest in the sense of positive-definite ordering (only) by the choice $g(x) = c m(x)$, where $c > 0$ is an arbitrary constant. One way to argue this is by general principles, saying that (whether $p$ is assumed known or unknown), the asymptotically optimal MLE corresponds to the estimating equation with $g = m$, and therefore that this is the choice uniquely minimizing the asymptotic variance, making it equal to the inverse of $C = E(\mu(X)(1 - \mu(X)) m(X)^{\otimes 2})$.

But there is a more direct way to argue that $A^{-1}B A^{-1 \text{tr}} - C$ is positive definite for all choices of $g$ not proportional to $m$. (We made an argument in class via Cauchy-Schwarz for the case of scalar $g$ and $m$, but that argument did not cover the general case.) Indeed, the matrix obtained by entrywise integration is evidently nonnegative definite, where $dG(x) \equiv \mu(x)(1 - \mu(x)) dF(x)$:

$$\int (g(x) - AC^{-1}m) dG(x) = B - AC^{-1}A^{\text{tr}}$$

Thus in positive-definite matrix ordering

$$B \geq AC^{-1}A^{\text{tr}} \Rightarrow A^{-1}B A^{-1 \text{tr}} \geq C^{-1}$$

which is the desired result.