Stat 701 Handout on HW3 Misspecified Model Problem, (C)

In this problem, you are asked to calculate relative efficiency based on the maximum likelihood estimator in the double-exponential distribution, $g(x,\mu) = (1/2)e^{-|x-\mu|}$. The resulting estimator, the sample median $\tilde{\mu}$, can be viewed as an estimating equation estimator based on the estimating function

$$\psi(x,\mu) \; = \; 2\, I_{[x \le \mu]} \; - \; 1 \qquad \text{or} \qquad 2\, I_{[x \ge \mu]} - 1$$

You are asked to do theoretical calculations involving asymptotic normality of $\sqrt{n}\,(\tilde{\mu} - \mu)$ with mean 0 and an asymptotic variance that you must calculate for true distribution either $g(x,\mu_0)$ or $\mathcal{N}(\mu_0, 1)$. But this estimating function ψ and associated estimator do not satisfy the assumptions of our Theorem establishing asymptotic normality. Although there are extended estimating-equation theoretical results in the statistical literature, the idea for an alternative estimating equation that I sketched in class does not actually produce the same sample-median estimator.

So what I give here is an explanation, as brief as I can make it, of how to get the theoretical asymptotic normality of the sample median and compute the variance, so that you can do problem (C). You will see that there are three steps. **To complete Problem (C) for HW3, I will ask you to take the first two steps as given and complete only the third one.**

Step 1. The sample median is a consistent estimator of the population median, assuming that the *iid* observations X_i come from a density f with unique median, i.e., one for which the distribution function $F(t) = \int_{-\infty}^{t} f(x)dx$ satisfies $F(\mu) = 1/2$ and for arbitrarily small $\delta > 0$, $F(\mu - \delta) < 1/2 < F(\mu + \delta)$.

Since we know that the sample median $\tilde{\mu}$ is either the order statistic $X_{((n+1)/2)}$ if n is odd, and $(X_{(n/2)} + X_{(n/2+1)})/2$ if n is even, if follows that $|F_n(\tilde{\mu}) - 1/2| \le 1/n$, where $F_n(t) = n^{-1}\sum_{i=1}^{n} I_{[X_i \le t]}$ is the *empirical distribution function*. Therefore, using $[y]$ to denote the greatest integer $\le y$, as $n \to \infty$

$$P(\tilde{\mu} \ge 1/2 + \delta) \; \le \; P(X_{[(n+1)/2]} \ge 1/2 + \delta) \; \le \; \texttt{pbinom}([\frac{n+1}{2}],\, n,\, 1/2 + \delta) \; \to \; 0$$

and a similar argument shows $P(\tilde{\mu} \le \mu - \delta) \to 0$. Thus $\tilde{\mu}$ is consistent.

Step 2. We know that $F(\mu) = 1/2$ and $|F_n(\mu) - 1/2| \le 1/n$. So we look at the difference between $F_n(t)$ and $F(t)$, which we know is small for large n and fixed t by the

Law of Large Numbers. In fact, using the central limit theorem, we know for each t that as $n \to \infty$

$$\sqrt{n}\,(F_n(t) - F(t)) \quad \xrightarrow{\mathcal{D}} \quad \text{a random continuous function}$$

This kind of result is proved in STAT 601 using weak-convergence theory on the space of continuous functions (or empirical process theory). The implication which we use below is: for all sequences $\epsilon_n \searrow 0$ as $n \to \infty$,

$$\sup_{x:\,|x-\mu|\le\epsilon_n} |\sqrt{n}\,(F_n(x) - F(x)) \;-\; \sqrt{n}\,(F_n(\mu) - F(\mu))| \;\xrightarrow{P}\; 0 \tag{1}$$

Step 3. This is the part that you will fill in to complete your asymptotic variance calculation in Problem (C). We know that $F(\mu) = 1/2$ and $|F_n(\mu) - 1/2| \le 1/n$, and by (1),

$$\sqrt{n}\,(F(\tilde{\mu}) \;-\; 1/2) \;-\; \sqrt{n}\,(1/2 \;-\; F_n(\mu)) \;\xrightarrow{P}\; 0 \tag{2}$$

and also, by the ordinary CLT

$$\sqrt{n}\;(F_n(\mu) - 1/2) \;=\; \frac{1}{\sqrt{n}}\sum_{i=1}^{n} (I_{[X_i \le \mu]} \;-\; 1/2) \;\xrightarrow{\mathcal{D}}\; \mathcal{N}(0,\,1/4) \tag{3}$$

Use these facts to prove the asymptotic normality of $\sqrt{n}(F(\tilde{\mu}) \;-\; F(\mu))$, and then use continuous differentiability of F at μ together with the Delta Method to prove asymptotic normality of $\sqrt{n}\,(\tilde{\mu} - \mu)$.

The asymptotic variance of $\sqrt{n}\,(\tilde{\mu} - \mu)$ in Step 3 is what you will use to complete the Asymptotic Relative Efficiency calculations (separately for f the standard normal and double-exponential density) in Problem (C).