STOR 655 problems for CWE August 2010

Let $X_1, X_2, \ldots$ be iid samples from a mixture density

$$f(x|\theta) = \theta\ g(x) + (1-\theta)\ h(x)$$

with an unknown parameter $\theta \in (0,1)$ and known densities g and h. Write $X^n = (X_1, \ldots, X_n)$ for a sample of size n. Except for part (c), we assume the curves g and h have at most a finite number of intersections.

(a) Find a MLE of θ based on X^n.

(b) Show that the Fisher information with respect to $f(\cdot|\theta)$ is given by

$$I(\theta) = \frac{1}{\theta(1-\theta)} \left[1 - \int_{-\infty}^{\infty} \frac{g(x)h(x)}{\theta\ g(x) + (1-\theta)\ h(x)}\ dx\right]. \qquad (1)$$

[**Hint:** Use $g - h = (f-h)/\theta$ and a similar expression for $h - g$.]

(c) In the extreme case with $g \equiv h$, show that θ becomes unidentifiable, i.e. the observations X^n tell us nothing about θ.

(d) At the other extreme, suppose the sample space has a subset A (known to us) to which g assigns probability one while h assigns probability zero. Find a MLE $\widehat{\theta}_n$ for θ based on X^n.

(e) Is the MLE $\widehat{\theta}_n$ derived in (d) asymptotically efficient? Explain why.

(f) Again, assume the curves g and h only intersect at a finite number of points. Find a minimal sufficient statistic T for θ based on X^n. [**Hint:** Recall the argument that leads to the minimal sufficient statistics for Cauchy population with location parameter θ.]

(g) Consider the special situation where g and h are densities for $N(-\mu, 1)$ and $N(\mu, 1)$ respectively with a know constant $\mu > 0$. For testing $H_0 : \theta = 1/2$ vs $H_a : \theta \neq 1/2$ based on X^n, which of two cases is easier: small μ or large μ? Explain why.