Sta 532: Homework #4

1. Find the Kullback-Leibler divergence $\mathrm{KL}(\theta : \theta')$ for the binomial distribution $\mathsf{Bi}(n, \theta)$ with fixed n, for $\theta, \theta' \in (0,1)$.

2. Find the Fisher Information $I_n(\theta)$ for the binomial distribution $\mathsf{Bi}(n, \theta)$ with fixed n, and verify that[1]
$$\mathrm{KL}(\theta : \theta') = \tfrac{1}{2} I_n(\theta)(\theta - \theta')^2 + o(\theta - \theta')^2$$
Suggestion: Write $\theta' = \theta + \epsilon$, and use the Taylor series $\log(1 + x) = x - x^2/2 + \xi^3/3$ for some $\xi \in [0, x]$.
This *always* happens (under the usual regularity conditions)— so, within parametric families, closeness in the Kullback-Leibler is the same as closeness in the *Information metric*, and the Information distance is approximately $\sqrt{2\mathrm{KL}(\theta : \theta')}$ for $\theta' \approx \theta$.

3. Let $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Un}(0, \theta)$ be uniformly distributed on the interval $[0, \theta]$ for some uncertain $\theta > 0$. The MLE is $\hat{\theta}_n = X_n^* := \max\{X_i : \ 1 \le i \le n\}$, the sample maximum. Verify that $\hat{\theta}_n$ is consistent, *i.e.*, that
$$\mathsf{P}_\theta\left[|\hat{\theta}_n - \theta| > \epsilon\right] \to 0$$
for any $\epsilon > 0$, by giving an explicit bound for the indicated probability in terms of n, ϵ, and θ.

4. Again let $\{X_i\} \overset{\text{iid}}{\sim} \mathsf{Un}(0, \theta)$ be uniformly distributed, and let θ have the improper scale-invariant prior density function $\pi(\theta) = \theta^{-1} \mathbf{1}_{\{\theta > 0\}}$. Find the Bayesian posterior mean $\bar{\theta}_n := \mathsf{E}[\theta \mid X_1 \cdots X_n]$ and verify that it too is consistent.

Last edited: January 29, 2015

[1] A note on "little oh" and "big oh" notation: The notation "$f(x) = o\big(g(x)\big)$ at x_0" or "$f(x) \in o\big(g(x)\big)$ at x_0", pronounced "$f(x)$ is little oh of $g(x)$ at x_0", means that $\lim_{x \to x_0} |f(x)/g(x)| = 0$, *i.e.*, that $f(x)$ is negligible compared to $g(x)$ near x_0. Often (as in problem 2) x_0 is zero or infinity and is implicit. The notation "$f(x) = O\big(g(x)\big)$ at x_0" or "$f(x) \in O\big(g(x)\big)$ at x_0", pronounced "$f(x)$ is big oh of $g(x)$ at x_0", means that $|f(x)/g(x)|$ is bounded in a neighborhood of x_0.