Statistics 610 fall 2014
Homework # 5
Due: Thursday 9 October

[5.1] Suppose $x_1, x_2, \dots$ are iid $N(\mu, \sigma^2)$ distributed random variables. Define $T_n = \sum_{i \le n} (x_i - \overline{x}_n)^2$ where $\overline{x}_n = n^{-1} \sum_{i \le n} x_i$. Find the constant c_n that minimizes $\mathbb{E}|c_n T_n - \sigma^2|^2$. Hint: The answer is not $c_n = (n-1)^{-1}$.

[5.2] For a fixed density function f on the real line define

$$f_\theta(z) = \frac{1}{\sigma} f\left(\frac{z-\mu}{\sigma}\right) \qquad \text{where } \theta = (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+.$$

Find the information matrix

$$\mathbb{I}(\mu, \sigma) = \text{var}_\theta \left(\frac{\partial \log f_\theta}{\partial \theta}\right),$$

the 2×2 variance matrix for the random vector $[\partial \log f_\theta / \partial \mu, \partial \log f_\theta / \partial \sigma]$.

[5.3] Let $\mathbb{P}_\theta$ denote the uniform distribution on $[0, \theta]^2$, for $\theta > 0$. That is, the coordinates x_1 and x_2 are independent Uniform$[0, \theta]$ under $\mathbb{P}_\theta$. Let $S := x_1 + x_2$ and $M := \max(x_1, x_2)$. Consider estimation of θ, with the excellence of an estimator $\widehat{\theta}$ being judged by $\mathbb{E}_\theta(\widehat{\theta} - \theta)^2$, the smaller the better.

(i) Define $G(m) := \mathbb{E}_\theta(S \mid M = m)$ for each m in $[0, \theta]$. Explain why $G(M)$ is preferred to S as an estimator for θ.

(ii) Define $H(s) := \mathbb{E}_\theta(3M/2 \mid S = s)$ for each s in $[0, 2\theta]$. Explain why $H(S)$ is not preferred to $3M/2$ as an estimator for θ.

(iii) Explain why $\mathbb{E}_\theta(2x_1 \mid S)$ is preferred to $2x_1$ as an estimator for θ.

[5.4] Under the $\mathbb{P}_\theta$ model, suppose $x_1, \dots, x_n$ are independent with each distributed Unif$(\theta - 1/2, \theta + 1/2)$. (You can think of $\mathbb{P}_\theta$ as the uniform distribution on $(\theta - 1/2, \theta + 1/2)^n$ and the x_i's as the coordinates of a generic x in $\mathbb{R}^n$.) The order statistics $x_{(1)} < x_{(2)} < \dots < x_{(n)}$ are just the values $x_1, \dots, x_n$ rearranged in increasing order.

(i) Show that $T(x) = (x_{(1)}, x_{(n)})$ is a sufficient statistic for $\{\mathbb{P}_\theta : \theta \in \mathbb{R}\}$.

(ii) Describe the conditional distribution of x given $T(x) = (t_1, t_2)$.

(iii) For $\theta - 1/2 < y < z < \theta + 1/2$ and δ small enough, show that

$$\mathbb{P}_\theta\{y < x_{(1)} < y + \delta, z < x_{(n)} < z + \delta\} = \delta^2 n(n-1)(z-y)^{n-2} + o(\delta^2).$$

(iv) Find the distribution of $x_{(n)} - x_{(1)}$. Sketch the density function.

(v) Mimic the argument given in class (for $n = 2$) to find a 90% confidence interval J_x for θ.

(vi) Find the expected length of J_x and the distribution of the length of J_x. (A labelled, hand-drawn sketch of the distribution function would help. It would be useful to show where 0.9 is on the vertical axis. You might prefer to use R to draw pictures for some values of n.)