6th problem set,

1. Suppose $X_1, X_2 \sim \text{i.i.d. Uniform}(\theta - 1/2,\ \theta + 1/2)$.

   (a) Show that $(\min\{X_1, X_2\},\ \max\{X_1, X_2\})$ is a 50% confidence interval for θ.

   (b) Suppose you observe that min$=84.03$ and max$=84.04$. Would you be confident that θ is within the 50% confidence interval? Suppose you observe that min$=64.52$ and max$=65.48$? Then would you be confident that θ is within the 50% confidence interval? Carefully explain your reasoning.

   (c) Suppose we assign an improper prior distribution to θ, as follows: $f_\theta(u) = c > 0$ for all $u \in \mathbb{R}$. (We fail to notice that $\int_{-\infty}^{\infty} c\,du \not< \infty$.) Let A and B be respectively the first and third quartiles of the posterior distribution. (The first quartile of a distribution is a number a such that the distribution assigns probability $1/4$ to the interval $(-\infty,\ a)$. The third quartile of a distribution is a number b such that the distribution assigns probability $3/4$ to the interval $(-\infty,\ b)$ (and probability $1/4$ to the interval $(b,\ \infty)$).) Express A and B as functions of $\min\{X_1, X_2\}$ and $\max\{X_1, X_2\}$). Find A and B in particular in case min$=84.03$ and max$=84.04$. Find A and B in case min$=64.52$ and max$=65.48$.

   (d) Show that (A, B) is a 50% confidence interval for θ, i.e., when the value of θ is fixed, we have $\mathbf{Pr}(A < \theta < B) = 0.5$.

   (e) Briefly discuss ways in which the 50% confidence interval of part (d) is better than the 50% confidence interval of part (a).

2. (a) In the regression problem of the 5th problem set, show that there is an $(n-2)$-dimensional space of vectors $c \in \mathbb{R}^n$ such that $\mathbf{E}(c'Y) = \mathbf{E}(c_1Y_1 + \cdots + c_nY_n) = 0$. (The random variable $c'Y$ is called a "linear unbiased estimator of zero.")

   (b) Suppose we have $n = 6$ and $x_1 = 2$, $x_2 = 3$, $x_3 = 5$, $x_4 = 5$, $x_5 = 7$, $x_6 = 9$. Find four linear unbiased estimators of zero that are linearly independent of each other, i.e., four vectors that can be put in the role of c in part (a), none of which is a linear combination of the others.

   (c) Let $\widehat{\beta_1} = d'Y$ be the least-squares estimator of of the slope in the regression problem of the 5th problem set (in #4 on the 5th problem set, you found the value of d). Show that $\widehat{\beta_1}$ is uncorrelated with every linear unbiased estimator of zero.

   (d) Let d and c be as above. Show that $(d+c)'Y$ is an unbiased estimator of β_1. Show that $\mathbf{var}((d+c)'Y) = \mathbf{var}(d'Y) + \mathbf{var}(c'Y) \geq \mathbf{var}(d'Y)$. Use the result to show that $d'Y$ is the best linear unbiased estimator of β_1. ("Linear" means linear in Y, i.e., of the form $d'Y$ for some vector d.) (This is not yet the Gauss-Markov theorem mentioned in the summary of March 6th, because that theorem has much weaker assumptions (and so, is a much stronger theorem) than those on which we are relying here.)

3. Let $X_1, \ldots, X_{15} \sim \text{i.i.d.}\, N(\mu, \sigma^2)$. Use what you know about the probability distribution of $\sum_{i=1}^{15}(X_i - \overline{X})^2$ to find a 95% confidence interval for σ^2, i.e., find two statistics A and B such that $\mathbf{Pr}(A < \sigma^2 < B) = 0.95$, and $\mathbf{Pr}(\sigma^2 < A) = 0.05/2$ and $\mathbf{Pr}(\sigma^2 > B) = 0.05/2$.