Homework 9
Due Thursday Nov 13 by 3:00

1. Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$. Let $\widehat{p} = \frac{1}{n}\sum_{i=1}^n X_i$. Let $X_1^*, \ldots, X_n^*$ denote a bootstrap sample and let $\widehat{p}^* = \frac{1}{n}\sum_{i=1}^n X_i^*$.

   (a) What is the exact distribution of $n\widehat{p}^*$, conditional on $X_1, \ldots, X_n$?

   (b) Find an explicit expression for the bootstrap variance. That is, find $\mathsf{Var}(\widehat{p}^*|X_1, \ldots, X_n)$.

   (c) What is the asymptotic distribution of $\sqrt{n}(\widehat{p} - p)$? What is the asymptotic distribution of $\sqrt{n}(\widehat{p}^* - \widehat{p}) \,|\, X_1, \ldots, X_n$?

2. Let $X_1, \ldots, X_n \sim \text{Uniform}(0, \theta)$. Let $\widehat{\theta}_n = X_{(n)} = \max\{X_1, \ldots, X_n\}$. Let $X_1^*, \ldots, X_n^*$ denote a (nonparametric) bootstrap sample and let $\widehat{\theta}_n^* = \max\{X_1^*, \ldots, X_n^*\}$. Show that

$$P(\widehat{\theta}_n^* = \widehat{\theta}) = 1 - \left(1 - \frac{1}{n}\right)^n \approx 1 - \frac{1}{e} = 0.63.$$

   Explain why this suggests that the bootstrap will not work well in this case. (Compare the properties of the distribution of $\widehat{\theta}^*$ and the distribution of $\widehat{\theta}$.)

3. Let $X_1, \ldots, X_n \sim N(\theta, 1)$ and let $\theta \sim N(a, b^2)$.

   (a) Find the posterior distribution $\pi(\theta|X_1, \ldots, X_n)$.

   (b) Find c_n such that

$$P(\theta \in C_n|X_1, \ldots, X_n) = 1 - \alpha$$

   where $C_n = [\overline{\theta}_n - c_n, \overline{\theta}_n + c_n]$ and $\overline{\theta}_n$ is the posterior mean.

   (c) Find

$$\mathsf{Cov}_{C_n}(\theta) = P_\theta(\theta \in C_n),$$

   the frequentist coverage of C_n. Plot $\mathsf{Cov}_{C_n}(\theta)$ as a function of θ. Also plot $\mathsf{Cov}_{D_n}(\theta)$ where

$$D_n = \left[\overline{X}_n - \frac{z_{\alpha/2}}{\sqrt{n}},\ \overline{X}_n + \frac{z_{\alpha/2}}{\sqrt{n}}\right].$$

   How does $\mathsf{Cov}_{C_n}(\theta)$ compare to $\mathsf{Cov}_{D_n}(\theta)$?

4. Let $X_i \sim N(\mu_i, 1)$ for $i = 1, \ldots, n$. Let

$$\pi(\mu_1, \ldots, \mu_n) \propto 1.$$

   (a) Find the posterior for $\mu = (\mu_1, \ldots, \mu_n)$.

   (b) Suppose we want to estimate $\theta = \sum_{i=1}^n \mu_i^2$. Find the posterior distribution $\pi(\theta|X_1, \ldots, X_n)$ for θ.

Hint: The following information will be helpful. A random variable V has a χ^2_k distribution if $V = \sum_{j=1}^k Z_j^2$ where $Z_1, \ldots, Z_k \sim N(0,1)$. Note that $\mathbb{E}(V) = k$ and $\mathsf{Var}(V) = 2k$. A random variable V has a non-central-$\chi^2_k(\nu)$ distribution with non-centrality parameter $\nu = \sum_j a_j^2$ if $V = \sum_{j=1}^k (Z_j + a_j)^2$ where $Z_1, \ldots, Z_k \sim N(0,1)$. In this case, $\mathbb{E}(V) = 2k + \nu$ and $\mathsf{Var}(V) = 2k + 4\nu$.

(c) Find the posterior mean $\overline{\theta}_n = \mathbb{E}(\theta | X_1, \ldots, X_n)$.

(d) Find the bias and variance of $\overline{\theta}_n$. Is $\overline{\theta}_n$ consistent?

The next two questions are optional.

(e) Find c_n such that $P(\theta \in C_n | X_1, \ldots, X_n) = 1 - \alpha$ where $C_n = [0, c_n]$. (Hint: c_n is just the quantile of a non-central χ^2.)

(f) Find (approximately) the frequentist coverage of C_n.

Hint: If V has a non-central-$\chi^2_k(\nu)$ distribution and if k is large then $V \approx N(2k + \nu, 2k + 4\nu)$.