Homework 2
Statistics 201B
Due Nov. 1, 2018

When not specified, please do not use R to get answers.

1. A manufacturer of booklets packages them in boxes of 100. It is known that, on average, the booklets weigh one ounce, with a standard deviation of 0.05 ounce. The manufacturer is interested in calculating

$$P(100 \text{ booklets weigh more than } 100.4 \text{ ounces}),$$

a number that would help detect whether too many booklets are being put into a box. Explain how you would calculate an approximate value of this probability. Mention any relevant theorems or assumptions needed.

2. Let $X_1, \dots, X_n$ be distinct observations (no ties). Let $X_1^*, \dots, X_n^*$ denote a bootstrap sample (a sample from the empirical CDF), and let $\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*$. Find: $E(\bar{X}_n^*|X_1, \dots, X_n)$, $V(\bar{X}_n^*|X_1, \dots, X_n)$, $E(\bar{X}_n^*)$, and $V(\bar{X}_n^*)$.

3. Let $X_1, \dots, X_n$ be *iid* with density

$$P_\theta(X = x) = \theta^x(1-\theta)^{1-x} \quad \text{for } x = 0, 1 \text{ and } 0 \le \theta \le 1/2$$

Find the method of moments estimator and the MLE of θ.

4. One observation is taken on a discrete random variable X with density $f(x;\theta)$, where $\theta \in \{1, 2, 3\}$. Find the MLE of θ.

x	$f(x;1)$	$f(x;2)$	$f(x;3)$
0	1/3	1/4	0
1	1/3	1/4	0
2	0	1/4	1/4
3	1/6	1/4	1/2
4	1/6	0	1/4

5. Verify the statements made in class about the Fisher information matrix $I_n(\mu, \sigma)$ and its inverse when $X_1, \dots, X_n \stackrel{iid}{\sim} N(\mu, \sigma^2)$.

6. Let $X_1, \dots, X_n \stackrel{iid}{\sim} Unif(0, \theta)$. Show that the MLE is consistent. Hint: Let $Y = \max\{X_1, \dots, X_n\}$. For any c, $P(Y < c) = P(X_1 < c, X_2 < c, \dots, X_n < c) = P(X_1 < c)P(X_2 < c) \cdots P(X_n < c)$.

7. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} N(\theta, 1)$. Define $Y_i = I\{X_i > 0\}$. Let $\psi = P(Y_1 = 1)$.

   (a) Find the MLE of ψ.

   (b) Find an approximate 95% confidence interval for ψ.

   (c) Define $\tilde{\psi} = \frac{1}{n}\sum_{i=1}^{n} Y_i$. Show that $\tilde{\psi}$ is a consistent estimator of ψ.

   (d) Compute the asymptotic relative efficiency of $\tilde{\psi}$ to $\hat{\psi}$. Hint: Use the delta method to get the standard error of the MLE. Then compute the standard error (i.e., the standard deviation) of $\tilde{\psi}$.

   (e) Suppose that the data are not really normal. Show that $\hat{\psi}$ is not consistent. What, if anything, does $\hat{\psi}$ converge to?

8. Let $X_1, \ldots, X_n \stackrel{iid}{\sim} Gamma(\alpha, \beta)$.

   (a) Find the MLE of β assuming α is known.

   (b) Find the Fisher information and construct an approximate 95% normal-based confidence interval for β.

   (c) When both α and β are unknown, there is no closed-form expression for the MLE. The file `berkeleyprecip.csv` on bCourse contains total monthly precipitation data for Berkeley, CA, going back to 1919. In `R`, calculate the total winter precipitation for each year (removing missing values) using

```
precip <- read.csv("berkeleyprecip.csv", header = TRUE)
precip[precip==-99999] <- NA # Missing values
winter.precip <- precip$DEC + precip$JAN + precip$FEB
winter.precip <- winter.precip[!is.na(winter.precip)]
```

   Numerically find the MLEs for α and β under the model that the values for each year are *iid* with distribution $Gamma(\alpha, \beta)$. Approximate the observed Fisher information matrix and use it to construct 95% normal-based confidence intervals for α and β. Hint: Look at the steps in `betaexample.R`. Turn in your MLEs and confidence intervals along with a comment about the numerical optimization: what evidence do you have about whether the algorithm found a global optimum?