SM-4331 Exercise 1

1. (a) Let X be a random variable with mean μ and variance σ^2. Show that $\mathrm{P}(|X-\mu| > 2\sigma) \leq 0.25$. What does this inequality tell us about the distribution of X?

   (b) Let $X_1, \ldots, X_n$ be an iid sample from a population with mean μ and variance σ^2. Show that for any $\epsilon > 0$, $\mathrm{P}(|\bar{X}_n - \mu| > \epsilon\sigma) \leq \frac{1}{n\epsilon^2}$. Compare this bound with the approximation implied by the CLT when n is large.

   [Note: The conditions required for these inequalities are minimum, and you may assume that they are met.]

2. Let X and Y be two r.v. with positive and finite variances. The correlation coefficient of X and Y is defined as
$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}.$$
   (If $\rho = 0$, X and Y are said to be uncorrelated or linearly independent).

   (a) Show that $|\rho| < 1$. *Hint: Use the Cauchy-Schwarz inequality.*

   (b) If $Y = aX + b$ for some constant $a \neq 0$ and b, show that $|\rho| = 1$.

   [Note: In fact, $|\rho| = 1$ if and only if $Y = aX + b$ for some constants $a \neq 0$ and b.]

3. Let $X_1, \ldots, X_n$ be a sample from a distribution with mean μ and variance $\sigma^2 \in (0, 1)$. Let $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$, where $\bar{X}_n$ is the sample mean.

   (a) Show that $S_n^2 = \frac{1}{n-1}\sum_{i=1}^n X_i^2 - \frac{n}{n-1}\bar{X}_n^2$.

   (b) Using Slutzky's theorem, show that $S_n^2 \xrightarrow{\mathrm{P}} \sigma^2$.

4. Let $X_1, \ldots, X_n$ be sample from Bern(p), and $Y_1, \ldots, Y_m$ be a sample from Bern(q), and the two samples are independent of one another.

   (a) Find a resonable estimator for $p - q$ and its standard error.

   (b) Find an approximate 95% confidence interval for $p - q$ when both n and m are large.

5. 100 people are given a standard antibiotic to treat an infection and another 100 are given a new antibiotic. In the first group, 90 people recover, while in the second group, 85 people recover. Let p_1 be the probability of recovery with the standard antibiotic, and p be the probability of recovery with the new antibiotic. We are interested in estimating $\theta = p_1 - p_2$. Provide an estimate, standard error, an 80% confidence interval, and a 95% confidence interval for θ.

6. Let $Y_1, \ldots, Y_n$ be a sample from a Poisson distribution with mean $\theta > 0$ unknown.

   (a) Let $Y = Y_1 + \cdots + Y_n$. Find the mean, variance, and the distribution of Y.
   Hint 1: Use the moment generating function (MGF) to solve this. The MGF of a random variable Y_i is given by $M_{Y_i}(t) = \mathrm{E}(e^{tY_i})$. Furthermore, use the fact that $M_Y(t) = \prod_{i=1}^n M_{Y_i}(t)$.
   Hint 2: If X and Z are two random variables such that $M_X(t) = M_Z(t)$, then X and Z have the same probability distribution.

(b) Obtain the MLE for θ and its standard error.

(c) Suppose now that only the first m $(m < n)$ observations of the sample are known explicitly, while for the other $n-m$ only their sum is known, determine the MLE for θ.

7. Find the MLE for λ given a random sample from the gamma distribution with pdf

$$f(x|\lambda, r) = \frac{\lambda^r}{\Gamma(r)} e^{-\lambda x} x^{r-1},$$

where r is a known constant.

8. Find the MLE for θ from a random sample from the population with density function

$$f(y|\theta) = \frac{2y}{\theta^2}$$

where $0 < y \leq \theta$, and $\theta > 0$. **Do not use calculus.** Draw a picture of the likelihood function.

9. Let $X_1, \ldots, X_n$ be a sample from $\mathrm{Unif}(0, \theta)$, where $\theta > 0$ is an unknown parameter. Find the MLE for θ. Derive the distribution for $\hat{\theta}$, and therefore, show that $\hat{\theta}$ is a consistent estimator in the sense that $\hat{\theta} \xrightarrow{\mathrm{P}} \theta$ as $n \to \infty$. *Hint:* $\mathrm{P}(\max_{1 \leq i \leq n} X_i \leq y) = \prod_{i=1}^{n} \mathrm{P}(X_i \leq y)$.

10. Let $X_1, \ldots, X_n$ be a random sample from a Bernoulli distribution, i.e.

$$X_i = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } 1-p \end{cases}$$

for $i = 1, \ldots, n$, where $p \in (0, 1)$ is unknown. Let $\theta = p^2$.

(a) Find the Cramr-Rao lower bound for the variance of unbiased estimators of θ.

(b) Find the MLE $\hat{\theta}$ for θ.

(c) Show that $\mathrm{E}(\hat{\theta}) \neq \theta$.

11. Let $\mathbf{X} = (X_1, \ldots, X_n)^\top$ be a sample from $\mathrm{N}(\mu, \sigma^2)$. Let $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$. Find the Fisher information matrix $\mathcal{I}_{\mathbf{X}}(\boldsymbol{\theta})$, i.e. the Fisher information using all n data points. *Hint: Use $\theta_2 = \sigma^2$ in your calculations.*