1. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $U[0,1]$ (uniform on the interval $[0,1]$) distribution, and let $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ be the order statistics. In particular, $X_{(1)} = \min(X_1, \ldots, X_n)$ and $X_{(n)} = \max(X_1, \ldots, X_n)$.

   (a) Let $\{a_n\}_{n\geq 1}$, $\{b_n\}_{n\geq 1}$ be positive sequences. Express the probability

   $$\mathrm{P}\left(X_{(1)} > \frac{x}{a_n},\ X_{(n)} < 1 - \frac{y}{b_n}\right)$$

   as a function of x, y.

   (b) Set $a_n = b_n := n$, and use part (a) to determine the joint asymptotic distribution of $(nX_{(1)}, n(1 - X_{(n)}))$. Conclude that $nX_{(1)}$ and $n(1 - X_{(n)})$ are asymptotically independent.

   (c) Use part (b) to find the limiting distribution of $n(1 - X_{(n)} + X_{(1)})$. Please write down the density of this distribution explicitly.

   (d) For $Y_1, \ldots, Y_n$ i.i.d. $U[a, b]$ with $a < b$, both unknown, use part (c) to find an asymptotic confidence interval at level $1-\alpha \in (0,1)$ for the range $b-a$. (Hint: Consider the transformation $Y = (b-a)X + a$.)

2. A population is made up of two subpopulations 1 and 2 in known proportions $p \in (0,1)$ and $1-p$, respectively. For $i = 1, 2$ let μ_i and σ_i^2 denote the unknown mean and known variance, respectively, of subpopulation i.

   (a) For $n, m \geq 1$ let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ denote i.i.d. samples drawn independently from subpopulations 1 and 2, respectively, and let $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$ and $\overline{Y}_m = (1/m)\sum_{i=1}^m Y_i$. Show that, for any $n, m \geq 1$,

   $$\widehat{\theta}_{n,m} := p\overline{X}_n + (1-p)\overline{Y}_m$$

   is unbiased for the mean of the overall population.

   (b) Compute $\mathrm{Var}(\widehat{\theta}_{n,m})$.

   (c) Now suppose that each data point drawn from subpopulation 1 costs \$1, and each drawn from subpopulation 2 costs \$$c$, for some known quantity $c > 0$. How should n and m be chosen to achieve the minimal variance $\mathrm{Var}(\widehat{\theta}_{n,m})$ subject to a total sampling budget of \$$b$, where b is a known, large quantity? For this part and the next you can ignore the discreteness of n, m and treat these quantities as continuous.

   (d) What is the minimal value of this variance in part 2c?