1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with pdf

$$f_X(x) = \frac{1}{\theta} e^{-x/\theta}, \quad 0 < x < \infty, \quad 0 < \theta < \infty.$$

A researcher is interested in deriving the distribution of a random variable $U = X_1 / \sum_{i=1}^n X_i$, which is a ratio of any given variable to the summation of n variables.

(a) To derive the distribution, one statistician suggests to create another random variable $V = \sum_{i=1}^n X_i$ and write $V = X_1 + Y_1$, where $Y_1 = \sum_{i=2}^n X_i$. Find the inverse function for X_1 and Y_1 as a function of U and V and derive the determinant of the Jacobian matrix.

(b) Show that the joint pdf of U and V is

$$f_{U,V}(u, v) = \frac{1}{\Gamma(n-1)\theta^n} (1-u)^{n-2} v^{n-1} e^{-v/\theta},$$

using the fact that Y_1 follows a Gamma distribution with pdf

$$f_{Y_1}(y) = \frac{1}{\Gamma(n-1)\theta^{n-1}} y^{n-2} e^{-y/\theta}, \quad 0 < y < \infty.$$

(c) Make an argument that U and V are independent and derive the marginal distributions of U and V.

(d) Show that $V = \sum_{i=1}^n X_i$ is a complete and sufficient statistic and that $U = X_1 / \sum_{i=1}^n X_i$ is an ancillary statistic of θ.

(e) Let an indicator function $\delta(X_1)$ be defined by

$$\delta(X_1) = \begin{cases} 1 & \text{if } X_1 > c, \\ 0 & \text{otherwise,} \end{cases}$$

where c is a constant. Show that

$$E\{\delta(X_1) | \sum_{i=1}^n X_i = t\} = (1 - c/t)^{n-1},$$

and that

$$E(X_1 | \sum_{i=1}^n X_i = t) = t/n,$$

using Basu's Theorem and $E\{\delta(X_1)\} = P(X_1 > c)$.

2. An event occurrence, e.g., mortality or hospitalization, can be considered as an *end point* in a clinical trial. A biostatistician uses $X_1, \ldots, X_n$ to represent a random sample of size n for the *control* group and likewise $Y_1, \ldots, Y_n$ for the *treatment* group. One common quantity a biomedical researcher is interested for the comparison between control and treatment groups is called *odds ratio*, which is a ratio of two odds.

(a) Given that $X_1, \ldots, X_n$ follow Bernoulli(θ_1), $0 < \theta_1 < 1$, and that $Y_1, \ldots, Y_n$ follow Bernoulli(θ_2), $0 < \theta_2 < 1$, derive the limiting distribution of $\bar{X} = n^{-1} \sum_{i=1}^{n} X_i$ and $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ using Central Limit Theorem (CLT).

(b) The odds, defined by $\gamma_1 = \theta_1/(1-\theta_1)$ and $\gamma_2 = \theta_2/(1-\theta_2)$, can be used to describe how large θ_1 and θ_2 are. However, due to a restricted range of γ_1 and γ_2, a biostatistician tends to work on log-odds, which is defined by $\log(\gamma_1)$ and $\log(\gamma_2)$, respectively, for control and treatment groups. If one uses $\log(\hat{\gamma}_1) = \log(\bar{X}/(1-\bar{X}))$ and $\log(\hat{\gamma}_2) = \log(\bar{Y}/(1-\bar{Y}))$ to estimate $\log(\gamma_1)$ and $\log(\gamma_2)$, derive the limiting distributions of the two log-odds estimators.

(c) The logarithm of the odds ratio, which is defined by $\log(\gamma_1/\gamma_2)$, can then be estimated by the difference between two log-odds estimators, i.e., $\log(\hat{\gamma}_1) - \log(\hat{\gamma}_2)$. Assuming X and Y are independent, show that the limiting (asymptotic) distribution of $\log(\hat{\gamma}_1/\hat{\gamma}_2)$ is

$$\sqrt{n}\{\log(\hat{\gamma}_1/\hat{\gamma}_2) - \log(\gamma_1/\gamma_2)\} \to_d N(0, \sigma^2),$$

with σ^2 as a function of θ_1 and θ_2.

[Hint: If $X_n \to_d X$ and $Y_n \to_d Y$, then $X_n + Y_n \to_d X + Y$ when X_n and Y_n are independent for each n.]

3. Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution with pdf

$$f_X(x) = \frac{1}{\theta}, \quad 0 < x < \theta.$$

(a) Show that

$$P(X_{(n)} \leq x) = \left(\frac{x}{\theta}\right)^n,$$

where $X_{(n)}$ is the maximum order statistic, and that, for any $\epsilon \in (0, \theta)$,

$$P(|X_{(n)} - \theta| \leq \epsilon) = 1 - \left(1 - \frac{\epsilon}{\theta}\right)^n,$$

and that $X_{(n)}$ converges in probability to θ.

(b) Show that $Z_n = n(\theta - X_{(n)})$ converges in distribution to an exponential distribution with mean θ, using the fact that $\lim_{n\to\infty}(1 - x/n)^n = e^{-x}$ for some $x \in (0, n)$.

(c) [Bonus] Show that $Y_n = n\{1 - F_X(X_{(n)})\}$ converges in distribution to an exponential distribution with mean 1, where $F_X(x)$ is the cdf of X.