1. Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with probability density function (pdf)
$$f(x) = \frac{1}{\beta} e^{-x/\beta}$$
and cumulative density function (cdf)
$$F(x) = 1 - e^{-x/\beta}, \quad 0 < x < \infty, \quad 0 < \beta < \infty.$$

   (a) Show that $F(X_1), \ldots, F(X_n)$ can be considered as a random sample from a uniform distribution between 0 and 1 by showing that
$$P(F(X_i) \leq x) = x,$$
for $i = 1, \ldots, n$.

   (b) Let $X_{(i)}$ be the order statistics from the random sample $X_1, \ldots, X_n$ and let $Z_i = F(X_{(i)})$. Show that the joint distribution of Z_i and Z_j is
$$f_{Z_i,Z_j}(z_i, z_j) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} z_i^{i-1}(z_j - z_i)^{j-i-1}(1-z_j)^{n-j},$$
where $i < j$ and $0 < z_i < z_j < 1$.

   (c) Let $U = Z_j - Z_i$ and $V = Z_i$. Show that the joint distribution of (U, V) is
$$f_{U,V}(u, v) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} v^{i-1} u^{j-i-1}(1-u-v)^{n-j}.$$
You need to demonstrate the domain of U and V is $u, v > 0$ and $0 < u + v < 1$.

   (d) Show that the marginal distribution of U is
$$f_U(u) = \frac{\Gamma(n+1)}{\Gamma(j-i)\Gamma(n-j+i+1)} u^{j-i-1}(1-u)^{n-j+i},$$
which is pdf of Beta distribution with $\alpha = j - i$ and $\beta = n - j + i + 1$. [Hint: letting $y = v/(1-u)$ will help on solving the complicated integral.]

   (e) A researcher is eager to find a so-called "tolerance interval" $(X_{(i)}, X_{(j)}]$ that covers at least $(100 \times p)$ percent of the distribution at $(100 \times \gamma)$ level. That is, the interval $(X_{(i)}, X_{(j)}]$ satisfies
$$P(F(X_{(j)}) - F(X_{(i)}) \geq p) = \gamma.$$

Given that $i = 1$ and $j = n$, comment on how one can find the probability γ to show the tolerance level of using range $X_{(n)} - X_{(1)}$ to cover at least 80 percent of the distribution.

2. For a women in a certain high-risk population, suppose that the number of lifetime events of domestic violence involving emergency room treatment is assumed to have the Poisson distribution

$$f_X(x|\lambda) = \lambda^x e^{-\lambda}/x!, \quad x = 0, 1, \ldots, \quad \lambda > 0.$$

Let $X_1, \ldots, X_n$ be iid sample randomly chosen for the high-risk population, and each woman in the random sample is asked to recall the number of lifetime events of domestic violence involving emergency room treatment that she has experienced.

(a) Show that the distribution belongs to an exponential family by identifying $h(x)$, $c(\lambda)$, $w(\lambda)$ and $t(x)$, and show that $Y = \sum_{i=1}^n X_i$ is a complete sufficient statistic for λ

(b) Let θ be the probability of a woman ever suffering domestic violence in the past, i.e., $\theta = P(X > 0)$. Show that $\theta = 1 - e^{-\lambda}$ and that $\hat{\theta} = 1 - (1 - 1/n)^Y$ is an unbiased estimator of θ using the fact that Y follows Poisson$(n\lambda)$.

(c) Due to possible recall bias, a researcher decide to dichotomize X_i into

$$Z_i = \begin{cases} 1 & \text{if } X_i > 0 \\ 0 & \text{if } X_i = 0. \end{cases}$$

Let $\bar{Z} = n^{-1} \sum_{i=1}^n Z_i$. Show that $\bar{Z}$ converges in probability to θ and that

$$\sqrt{n}(\bar{Z} - \theta) \to_d N(0, \theta(1 - \theta)).$$

(d) In order to estimate λ, the researcher suggests transformation on $\bar{Z}$. Find a function g such that $g(\bar{Z})$ converges in probability to λ and show that

$$\sqrt{n}(g(\bar{Z}) - \lambda) \to_d N(0, e^{\lambda} - 1).$$

(e) If one would be able to use the original sample $X_1, \ldots, X_n$ to estimate λ, one can show that $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ also converges in probability to λ and that

$$\sqrt{n}(\bar{X} - \lambda) \to_d N(0, \lambda).$$

We now have two consistent estimators $g(\bar{Z})$ and $\bar{X}$ for λ. Which one is preferable? That is, which one has a smaller variance when the sample size is large? Give a heuristic reason why the estimator has a smaller variance.