Problem 1 The *order statistics* of a random sample $Y_1, \ldots, Y_n$ from a continuous distribution F with density function f are defined to be the ordered values $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n-1)} < Y_{(n)}$. Thus $Y_{(1)}$ and $Y_{(n)}$ are respectively the sample minimum and maximum and, if n is odd, the sample median is $Y_{((n+1)/2)}$.

(a) Find the distribution functions of $Y_{(1)}$ and $Y_{(n)}$ and hence obtain their density functions.

(b) Use the fact that $n!$ permutations of the $Y_1, \ldots, Y_n$ would yield the same values of $Y_{(1)}, \ldots, Y_{(n)}$ to explain why the joint density of the order statistics is

$$n! f(y_1) \cdots f(y_n), \quad y_1 < \cdots < y_n,$$

and hence recover the density functions in (a).

(c) Find the joint density of the order statistics of a random sample $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} U(0, a)$.

Problem 2 Consider the order statistics $0 < Y_{(1)} < \cdots < Y_{(n)}$ of a random sample $Y_1, \ldots, Y_n \stackrel{\text{iid}}{\sim} \exp(\lambda)$, and write $\stackrel{\text{D}}{=}$ for 'has the same distribution as'.

(a) Show that $\min(Y_1, \ldots, Y_r) \sim \exp(r\lambda)$, and that each Y_j has the lack-of-memory property

$$\mathrm{P}(Y - x > y \mid Y > x) = \mathrm{P}(Y > y), \quad x, y > 0.$$

(b) Show that $Y_j \stackrel{\text{D}}{=} E_j/\lambda$ with $E_1, \ldots, E_n \stackrel{\text{iid}}{\sim} \exp(1)$, and hence obtain the *Renyi representation*

$$Y_{(r)} \stackrel{\text{D}}{=} \frac{1}{\lambda} \sum_{j=1}^{r} \frac{E_j}{n+1-j}, \quad r = 1, \ldots, n.$$

(c) Find the means and covariances of $Y_{(1)}, \ldots, Y_{(n)}$.

Problem 3 Consider the homogeneous Poisson process with rate $\lambda > 0$ observed on $[0, t_0]$, and suppose that events occur at times $0 < t_1 < \cdots < t_n < t_0$.

(a) Show that

$$\mathrm{P}\{N(t_0) = n\} = \int_0^{t_0} \mathrm{d}t_n \int_0^{t_n} \mathrm{d}t_{n-1} \cdots \int_0^{t_2} \mathrm{d}t_1 \, \lambda^n e^{-\lambda t_0}, \quad n = 1, 2, \ldots,$$

and deduce that $N(t_0)$ has the Poisson distribution with mean λt_0.

(b) Show that the log likelihood for λ based on the times $t_1, \ldots, t_n$ is

$$\ell(\lambda) = n \log \lambda - \lambda t_0, \quad \lambda > 0,$$

and deduce that the unique maximum likelihood estimator is $\widehat{\lambda} = n/t_0$ and that the expected information is $\imath(\lambda) = t_0/\lambda$. Are these the same as for the log likelihood based on $N(t_0)$? Explain.

(c) Find the distribution of the event times conditional on the event $N(t_0) = n$.

(d) We can check the homogeneous Poisson process model for the Bengal typhoon data with the following `R` code:

```
load("bengal.dat")  # data available on Moodle page
bengal # look at event times
u <- (bengal-1877)/101 # rescale them to (0,1) and plot their empirical CDF
plot(u,c(1:141)/141,type="s",panel.first=abline(0,1,col="grey"))
ks.test(u,y="dunif") # Kolmogorov-Smirnov test of uniformity
```

Explain the connection to (c). What do you conclude?