Solution 1

(a) Clearly $Y_{(n)} \leq y$ iff all the $Y_j \leq y$, and since the Y_j are independent, $\mathrm{P}(Y_{(n)} \leq y) = F(y)^n$, so the density function is $nf(y)F(y)^{n-1}$.

Likewise $Y_{(1)} > y$ iff all the $Y_j > y$, so $\mathrm{P}(Y_{(1)} > y) = \{1 - F(y)\}^n$, giving $\mathrm{P}(Y_{(1)} \leq y) = 1 - \{1 - F(y)\}^n$ and density function $nf(y)\{1 - F(y)\}^{n-1}$.

(b) The probability density for a particular permutation of $Y_1, \ldots, Y_n$ to equal $y_1, \ldots, y_n$ is $\prod_{j=1}^n f(y_j)$. But since there are $n!$ such permutations, all with the same density, the density values for all $n!$ permutations must be added to give the joint density for the order statistics, i.e.,

$$f_{Y_{(1)},\ldots,Y_{(n)}}(y_1, \ldots, y_n) = n! \prod_{j=1}^n f(y_j), \quad y_1 < \cdots < y_n.$$

A similar argument can be applied to the c.d.f. Indeed, $F_{(1),\ldots,(n)}(\boldsymbol{y}) = \mathrm{P}(\{Y_{(i)} \leq y_i,\ i = 1, \ldots, n\}) \overset{*}{=} n!\mathrm{P}(\{Y_i \leq y_i,\ i = 1, \ldots, n\})$, where the latter c.d.f. factorizes, and taking partial derivatives leads to the previous result on the density scale. The equality $*$ follows again from a permutation argument, as the domain restriction on $\boldsymbol{y}$ keeps the ordered sample identical.

We obtain the marginal density of $Y_{(n)}$ by integration over y_1, then y_2, $\ldots$, up to y_{n-1}. To do so, one just has to keep in mind the identity

$$\int_{-\infty}^{y_{k+1}} \frac{F(y_k)^{k-1}}{(k-1)!} f(y_k) \mathrm{d}y_k = \frac{F(y_{k+1})^k}{k!},$$

as the integral essentially recurses this formula.

The computation for the minimum is similar, integrating over $y_n \in (y_{n-1}, \infty)$, then over $y_{n-1} \in (y_{n-2}, \infty)$, etc. and resulting in $nf(y_1)\{1 - F(y_1)\}^{n-1}$. This time the key recursion is

$$\int_{y_{k-1}}^{\infty} \frac{(1 - F(y_k))^{n-k}}{(n-k)!} f(y_k) \mathrm{d}y_k = \frac{(1 - F(y_{k-1}))^{n-(k-1)}}{(n-(k-1))!}.$$

(c) The uniform density on $(0, a)$ is $1/a$ for $y \in (0, a)$, which gives $n!/a^n$ for $0 < y_1 < \cdots < y_n < a$ when inserted into (b).

Solution 2

(a) As $\min(Y_1, \ldots, Y_r) > x$ if and only if $Y_1 > x, \ldots, Y_r > x$, we have

$$\mathrm{P}\{\min(Y_1, \ldots, Y_r) \leq x\} = 1 - \mathrm{P}\{\min(Y_1, \ldots, Y_r) > x\} = 1 - \mathrm{P}(Y_1 > x)^r = 1 - \exp(-r\lambda x), \quad x > 0,$$

and for $x, y > 0$, $\mathrm{P}(Y - x > y \mid Y > x)$ equals

$$\frac{\mathrm{P}(Y - x > y, Y > x)}{\mathrm{P}(Y > x)} = \frac{\mathrm{P}(Y > y + x)}{\mathrm{P}(Y > x)} = \exp\{-\lambda(x + y)\}/\exp(-\lambda x) = \exp(-\lambda y),$$

as required.

(b) As $\mathrm{P}(E_j/\lambda \leq x) = \mathrm{P}(E_j \leq \lambda x) = 1 - \exp(-\lambda x) = \mathrm{P}(Y_j \leq x)$, we have $Y_j \overset{\mathrm{D}}{=} E_j/\lambda$. We argue as follows:

- $Y_{(1)}$ is the smallest of n independent exponential variables, so it is exponential with parameter $n\lambda$ and therefore we can write $Y_{(1)} \overset{\mathrm{D}}{=} E_1/(n\lambda)$;

- the remaining $n-1$ variables have the lack of memory property, so given that $Y_{(1)} = x$ the remaining $Y_j - x$ have exponential distributions with parameter λ. Thus $Y_{(2)} - Y_{(1)}$ is the minimum of $n-1$ exponential variables, i.e., $Y_{(2)} - Y_{(1)} \stackrel{\mathrm{D}}{=} E_2/\{(n-1)\lambda\}$;
- iterating the argument by successively conditioning on $Y_{(2)}, \ldots, Y_{(n-1)}$ and obtaining the distributions of $Y_{(3)} - Y_{(2)}, \ldots, Y_{(n)} - Y_{(n-1)}$ gives the stated representation.

An alternative solution is to work with increments. Define $D_1 = Y_{(1)}$ and $D_k = Y_{(k)} - Y_{(k-1)}$ for all $k = 2, \ldots, n$. The joint density of the ordered sample is

$$f_{Y_{(1)},\ldots,Y_{(n)}}(y_1, \ldots, y_n) = n!\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} y_i\right).$$

Now note that the inverse transformation $y_j(\boldsymbol{d}) = \sum_{i=1}^{j} d_i$ for all j has as jacobian $(J(\boldsymbol{d}))_{i,j} = \delta_{i\geq j}$ which has determinant one. Therefore the joint density of $(D_1, \ldots, D_n)$ is

$$\begin{aligned}
f_{\boldsymbol{D}}(\boldsymbol{d}) = f_{Y_{(1)},\ldots,Y_{(n)}}(\boldsymbol{y}(\boldsymbol{d})) &= n!\lambda^n \exp\left(-\lambda \sum_{i=1}^{n}\sum_{j=1}^{i} d_j\right) \\
&= \prod_{j=1}^{n} ((n-j+1)\lambda) \cdot \exp\left(-\lambda \sum_{j=1}^{n} (n-j+1) d_j\right) \\
&= \prod_{j=1}^{n} ((n-j+1)\lambda \exp\left(-\lambda(n-j+1)d_j\right)),
\end{aligned}$$

which we can see factorized in n exponential densities. Therefore we conclude that $D_i \sim \exp((n-j+1)\lambda)$ and jointly independent. This means that we can represent $D_j \stackrel{d}{=} E_j/(\lambda(n-j+1))$, with $E_1, \ldots, E_n \stackrel{\text{iid}}{\sim} \exp(1)$. Taking into account that $Y_{(k)} = \sum_{j=1}^{k} D_j$, we conclude the desired representation.

(c) A standard exponential variable has mean and variance both equal to 1, so for the mean

$$\mathrm{E}(Y_{(r)}) = \frac{1}{\lambda}\sum_{j=1}^{r} \frac{1}{n+1-j}$$

and for the covariance

$$\begin{aligned}
\operatorname{cov}(Y_{(r)}, Y_{(s)}) &= \operatorname{cov}\left(\frac{1}{\lambda}\sum_{i=1}^{r} \frac{E_i}{n+1-i}, \frac{1}{\lambda}\sum_{j=1}^{s} \frac{E_j}{n+1-j}\right) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{1}{\lambda^2(n-j+1)(n-i+1)} \operatorname{cov}(E_i, E_j) \\
&= \sum_{i=1}^{r}\sum_{j=1}^{s} \frac{\delta_{i,j}}{\lambda^2(n-j+1)(n-i+1)} \\
&= \sum_{j=1}^{m} \frac{1}{\lambda^2(n-j+1)^2}, \quad r, s \in \{1, \ldots, n\},
\end{aligned}$$

with $m = \min(s, r)$ and the second formula giving the variance when $r = s$. These are useful in assessing QQplots, since they give the expectation and variance of (individual) order statistics.

Solution 3

(a) We saw in the lectures that the joint density of the times of the n events is

$$e^{-\mu(0,t_0)} \prod_{j=1}^{n} \dot{\mu}(t_j), \quad 0 < t_1 < \cdots < t_n < t_0,$$

and, under the homogeneous Poisson assumption, setting $\dot{\mu}(t) = \lambda$ this reduces to $\lambda^n e^{-\lambda t_0}$. To observe n events in $[0, t_0]$, we need to observe t_n in this interval, then t_{n-1} in $[0, t_n]$, and so on, recursively, meaning integrating the density in this domain we get

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

but now note that for

$$\int_0^{t_{k+1}} \frac{t_k^{k-1}}{(k-1)!} dt_k = \frac{t_{k+1}^k}{k!}$$

and therefore

$$\mathrm{P}\{N(t_0) = n\} = \lambda^n \exp(\lambda t_0) \int_0^{t_0} \frac{t_n^{n-1}}{(n-1)!} dt_n = \frac{(\lambda t_0)^n}{n!} \exp(-\lambda t_0)$$

which is the p.m.f. of a Poisson distribution with mean λt_0.

(b) As before, denote by $N(t_0)$ the number of events in $[0, t_0]$. Due to the homogeneous Poisson assumption, we saw in (a) that the likelihood function based on the event times depends only on λ, t_0, and the counting process $N(\cdot)$. Indeed, from (a) we see that the log-likelihood is

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

First order derivatives $\ell'(\lambda) = N(t_0)/\lambda - t_0$ lead to a MLE of $\widehat{\lambda} = N(t_0)/t_0$, the observed rate of events per unit time in $[0, t_0]$. This is indeed a maximum as $\ell''(\lambda) = -N(t_0)/\lambda^2$, negative for all positive λ as long as $N(t_0) = 0$, leading to a unique MLE. If $N(t_0) = 0$ then the log-likelihood is constant in λ. While every value of λ would solve the score equation, Fisher's information is always zero, meaning our observations (or, rather, the lack of them) are completely uninformative towards λ.

If we were to base inference on the log-likelihood for the number of events we would have, again from (a)

$$\ell(\lambda) = N(t_0) \log \lambda + N(t_0) \log t_0 - \lambda t_0 - \log(N(t_0)!), \quad \lambda > 0,$$

which is the same log-likelihood as before up to additive constants not depending on λ, meaning inference is identical.

It is clear from $\ell(\lambda)$ that under this model the number of events is a sufficient statistic (*statistique exhaustive* — check this if unsure), so under this particular model the times are irrelevant for inference; what matters is the number of events.

(c) This conditional density is

$$f_{T_1,\ldots,T_n|N(t_0)}(t_1, \ldots, t_n \mid n) = \frac{f_{T_1,\ldots,T_n,N(t_0)}(t_1, \ldots, t_n, n)}{\mathrm{P}(N(t_0) = n)} \stackrel{*}{=} \frac{n! \exp(\lambda t_0)}{(\lambda t_0)^n} \cdot \lambda^n \exp(-\lambda t_0) = \frac{n!}{t_0^n},$$

where in $*$ we used the fact that in the density we computed in (a) (which is the one we plugged in for the joint density) the events are already restricted to $[0, t_0]$, so that the joint with the number of events reduced to the marginal of the event times. Hence the conditional distribution of the event times, given that there are n events, is that of the order statistics of a uniform sample on $(0, t_0)$ (see Problem 1(a)).

(d) The plot clearly shows that the data are under the diagonal line that would correspond to a uniform sample, and this is confirmed by the Kolmogorov–Smirnov test, which has a tiny P-value. Hence there is strong evidence against the model. The connection with (c) is that the rescaled data t_j/t_0 should be a uniform sample on $U(0, 1)$ if the model is correct, and clearly this is not the case.