1. (a) The likelihood function is

$$L(\theta; Y_1, \cdots, Y_n) = \prod_{j=1}^{n} \frac{\theta^{Y_j}}{Y_j!} e^{-\theta} = \theta^Y e^{-n\theta} \Big/ \prod_{j=1}^{n} Y_j!.$$

Hence, Y is sufficient for θ. Further, $EY = nEY_1 = n\theta$, and $\text{Var}(Y) = n\text{Var}(Y_1) = n\theta$. Since $E(e^{tY}) = \{E(e^{tY_1})\}^n = [\exp\{\theta(e^t - 1)\}]^n = \exp\{n\theta(e^t - 1)\}$, $Y \sim Poisson(n\theta)$.

(b) $l(\theta) = Y \log \theta - n\theta$. Let $\dot{l}(\theta) = Y/\theta - n = 0$, leading to $\hat{\theta} = Y/n$.

(c) From (a), $Z \sim Poisson((n-m)\theta)$. Therefore the likelihood now is

$$L(\theta; Y_1, \cdots, Y_m, Z) = \Big\{ \prod_{j=1}^{m} \frac{\theta^{Y_j}}{Y_j!} e^{-\theta} \Big\} \frac{\{(n-m)\theta\}^Z}{Z!} e^{-(n-m)\theta} \propto L(\theta; Y_1, \cdots, Y_n).$$

Hence the MLE is the same as in (b).

(d) Since $Z = Y_{m+1} + \cdots + Y_n$ is a sufficient statistic for θ with observations $Y_{m+1}, \cdots, Y_n$, there is no loss of information for knowing Z only as far as the estimation for θ is concerned. So the MLE based on $Y_1, \cdots, Y_m$ and Z is the same as the MLE based on the whole sample $Y_1, \cdots, Y_n$.

2. The likelihood function can be written

$$L(\lambda) = \prod_{i=1}^{n} f(x_i) = \frac{1}{[\Gamma(r)]^n} \exp\left(-\lambda \sum x_i\right) \left[\prod_{i=1}^{n} x_i\right]^{r-1} \lambda^{nr}.$$

The log-likelihood is

$$l(\lambda) = \text{constant} - \lambda \sum x_i + nr \ln \lambda,$$

where the constant depends on the observations and on r but not on λ. Let $\frac{d}{d\lambda} l(\lambda) = 0$, leading to $\hat{\lambda} = r/\bar{X}$. (Notice that the log-likelihood here is of the form in Question 1.)

3. The density function may be written

$$f(y; \theta) = \frac{2y}{\theta^2} I_{(0,\infty)}(y) I_{(-\infty,\theta)}(y).$$

so the joint density for a sample of observations is

$$f(\mathbf{y}; \theta) = \frac{2^n \prod_{i=1}^n y_i}{\theta^{2n}} \left[\prod_{i=1}^{n} I_{(0,\infty)}(y_i)\right] \left[\prod_{i=1}^{n} I_{(-\infty,\theta)}(y_i)\right] = \frac{2^n \prod_{i=1}^n y_i}{\theta^{2n}} \left[\prod_{i=1}^{n} I_{(0,\infty)}(y_i)\right] I_{(-\infty,\theta)}(y_{(n)}),$$

where $y_{(n)}$ is the largest observation. The likelihood is

$$L(\theta; \mathbf{y}) = C\theta^{-2n} I_{(-\infty,\theta)}(y_{(n)}),$$

where the constant C depends on the observations, but not on θ. To make the log-likelihood large, one must make θ as close to zero as possible without taking it below $y_{(n)}$. The maximum likelihood estimator of θ is therefore $Y_{(n)}$, the largest observation.

4. The log-likelihood of one observation is

$$l(\mu; y) = y - \mu - 2 \ln \left[1 + \exp(y - \mu)\right].$$

Differentiating with respect to μ gives the score function

$$-1+2\frac{\exp(y-\mu)}{1+\exp(y-\mu)}=-1+2F(y;\mu),$$

where $F(y;\mu)$ is the distribution function for the logistic distribution

$$F(y;\mu)=\int_{-\infty}^{y}f(y;\mu)dy=\frac{\exp(y-\mu)}{1+\exp(y-\mu)}.$$

Since by the probability integral transformation we know that $F(Y;\mu)$ has a uniform distribution on $(0,1)$, it follows that

$$\mathrm{Var}\{-1+2F(Y;\mu)\}=4\mathrm{Var}\{F(Y;\mu)\}=4/12=1/3,$$

and so the information in a sample of size n is $\mathcal{I}(\mu)=n/3$. The score function for the sample of size n is

$$s(\mu;\mathbf{y})=\sum_{i=1}^{n}[-1+2F(y_i;\mu)].$$

The iteration for the method of scoring is

$$\hat{\mu}_{r+1}=\hat{\mu}_r+[\mathcal{I}(\hat{\mu}_r)]^{-1}s(\hat{\mu}_r;\mathbf{y})=\hat{\mu}_r+3\sum_{i=1}^{n}[-1+2F(y_i;\hat{\mu}_r)]/n.$$

