Problem 1 $Y_1, \ldots, Y_n$ is a random sample from the Lomax distribution with unknown θ and known $\alpha > 0$,

$$\mathrm{P}(Y \le y) = \begin{cases} 1 - \dfrac{\theta^\alpha}{(\theta + y)^\alpha}, & y > 0, \\ 0, & y \le 0, \end{cases}$$

where $\alpha, \theta > 0$. Find the expected information for θ and hence compare the maximum likelihood estimator with the moments estimator found on Sheet 2.

Problem 2 Suppose that a random sample $Y_1, \ldots, Y_n$ from the exponential density is rounded down to the nearest δ, giving δZ_j, where $Z_j = \lfloor Y_j/\delta \rfloor$. Then the loss of information due to rounding is the ratio of the Fisher information based on the rounded data to that based on the original data.

(a) Show that the likelihood contribution from a rounded observation can be written as $(1-e^{-\lambda\delta})e^{-Z\lambda\delta}$, and deduce that the Fisher information for λ based on the rounded sample is

$$\imath(\delta) = n\delta^2 \exp(-\lambda\delta)\{1 - \exp(-\lambda\delta)\}^{-2}, \quad \delta, \lambda > 0.$$

(b) Show that $\imath(\delta)$ has limit n/λ^2 as $\delta \to 0$, and deduce that if $\lambda = 1$ the loss of information when data are rounded down to the nearest integer rather than recorded exactly is less than 10%.

(c) Find the loss of information when $\lambda\delta = 0.1$, and comment briefly.

(d) By considering the resulting average log likelihood as $n \to \infty$, show that replacing Y_j by δZ_j in the usual likelihood leads to slight overestimation of λ in large samples.

Problem 3 A sample of $n = 16$ Vaudois number plates has maximum 523308 and average 320869. Suppose that these were sampled uniformly and independently on the interval $(0, \theta)$, where θ is the highest number plate in the canton.

(a) A random sample $Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} U(0, \theta)$ is available. Find the mean and variance of $\overline{Y}$ and hence suggest an unbiased estimator of θ. Show that $Q = \overline{Y}/\theta$ is a pivot and find its approximate distribution. How could you find its exact distribution?

(b) Use the calculation in (a) to obtain an approximate 95% confidence interval for θ.

(c) Compare your interval from (b) with that computed using the maximum as suggested in the lectures. Which do you prefer, and why?

Problem 4 Consider a random sample of bivariate normal pairs (Y_j, X_j) $(j = 1, \ldots, n)$ with unknown mean vector (ψ, λ), and known values of $\mathrm{var}(Y_j) = \sigma_1^2$, $\mathrm{var}(X_j) = \sigma_2^2$ and $\mathrm{corr}(X_j, Y_j) = \rho$. Suppose that independent observations $X_{n+1}, \ldots, X_{n+m} \overset{\text{iid}}{\sim} \mathcal{N}(\lambda, \sigma_2^2)$ are also available. Such data might arise when measurements Y that are expensive or difficult to obtain are correlated with others, X, that are much cheaper or easier to obtain, so that it is only affordable to make n joint measurements but $m \gg n$ auxiliary measurements on X alone can also be provided. Under what circumstances do the auxiliary measurements aid in estimating ψ, and by how much?

(a) Write down the log likelihood function for ψ and λ based on the data, and show that $\widehat{\psi} = \overline{Y}_n + \rho\sigma_1(\overline{X}_{n+m} - \overline{X}_n)/\sigma_2$, where $n\overline{Y}_n = \sum_{j=1}^n Y_j$, $n\overline{X}_n = \sum_{j=1}^n X_j$ and $(n+m)\overline{X}_{n+m} = \sum_{j=1}^{n+m} X_j$.

(b) Check that both $\overline{Y}_n$ and $\widehat{\psi}$ are unbiased estimators of ψ, find $\mathrm{var}(\overline{Y}_n)$, show that $\widehat{\psi}$ has asymptotic variance $\sigma_1^2\{1 - m\rho^2/(n+m)\}/n$, and hence give the gain in relative efficiency due to the auxiliary data. Discuss how and why this changes when (i) $\rho = 0$, (ii) $m \to \infty$, and (iii) $\rho \to \pm 1$.