Problem 1 Data $Y_1, \ldots, Y_n$ treated as a random sample from the geometric density f with support $\mathcal{Y} = \{0, 1, \ldots\}$ and parameter $\theta \in (0, 1)$ are in fact from the Poisson density g with mean $\lambda > 0$.

(a) Show that $\mathrm{E}_g\{\log f(Y; \theta)\}$ is maximised by $\theta_g = (1 + \lambda)^{-1}$, and check that this matches the means of the models. Is this a general feature of mis-specified exponential family models?

(b) Find $\imath_1(\theta_g)$ and $\hbar_1(\theta_g)$ and show that the sandwich variance formula gives $\mathrm{var}(\widehat{\theta}_g) \doteq \theta_g^3(1 - \theta_g)/n$.

(c) Show that the maximum likelihood estimator of θ based on $Y_1, \ldots, Y_n$ is $\widehat{\theta} = 1/(1+\overline{Y})$ and use the delta method to find its asymptotic variance. Is this a surprise?

Problem 2 When the generalized Pareto distribution is written as

$$\mathrm{P}(Y > y) = (1 - y/\psi)_+^\lambda, \quad 0 < y < \psi, \quad \psi, \lambda > 0,$$

where $a_+ = \max(a, 0)$, the parameter ψ represents the upper support point for Y.

(a) Find the profile log likelihood for ψ based on a random sample $Y_1, \ldots, Y_n$.

(b) Show that if ψ is regarded as fixed, then the minimal sufficient statistic for λ is $S_\psi = \sum_j Z_j$, where $Z_j = -\log(1 - Y_j/\psi)$. By considering $\mathrm{P}(Z_j > z)$ or otherwise, show that S_ψ has a gamma distribution and deduce that the conditional density of the data given $S_\psi = s_\psi$ is

$$f(y_1, \ldots, y_n \mid s_\psi; \psi) = \frac{\Gamma(n) e^{s_\psi}}{\psi^n s_\psi^{n-1}}, \quad 0 < y_1, \ldots, y_n < \psi, \sum_{j=1}^n \log(1 - y_j/\psi) = s_\psi.$$

(c) Compare the profile log likelihood with the log likelihood obtained from (b). Which is preferable?

Problem 3 The generalized Pareto distribution was given in an earlier question.

(a) Show that the derivatives with respect to ψ satisfy the first two Bartlett identities only if $\lambda > 2$.

(b) If M_n denotes the sample maximum, show that $n^{1/\lambda}(\psi - M_n) \xrightarrow{D} W$ as $n \to \infty$, where $\mathrm{P}(W > w) = \exp\{-(w/\psi)^\lambda\}$ for $w > 0$. Deduce that when $\lambda \leq 2$ convergence to a limiting distribution for inference on ψ occurs more rapidly than with maximum likelihood estimation, but if $\lambda > 2$ then maximum likelihood is the better option, at least asymptotically.

Problem 4 Suppose that the parameter θ consists of a $p \times 1$ parameter of interest ψ and a $q \times 1$ nuisance parameter λ, and that the maximum likelihood estimator $\widehat{\theta}$ has approximate distribution

$$\widehat{\theta} = \begin{pmatrix} \widehat{\psi} \\ \widehat{\lambda} \end{pmatrix} \dot\sim \mathcal{N}_{p+q}\left\{ \begin{pmatrix} \psi \\ \lambda \end{pmatrix}, \begin{pmatrix} \widehat{\jmath}_{\psi\psi} & \widehat{\jmath}_{\psi\lambda} \\ \widehat{\jmath}_{\lambda\psi} & \widehat{\jmath}_{\lambda\lambda} \end{pmatrix}^{-1} \right\},$$

where the circumflex denotes a quantity evaluated at the overall maximum likelihood estimate.

(a) Use the formula for the inverse of a partitioned matrix to show that $\mathrm{var}(\widehat{\psi}) \doteq (\widehat{\jmath}_{\psi\psi} - \widehat{\jmath}_{\psi\lambda}\widehat{\jmath}_{\lambda\lambda}^{-1}\widehat{\jmath}_{\lambda\psi})^{-1}$.

(b) Show that the profile log likelihood $\ell_{\mathrm{p}}(\psi) = \ell(\psi, \widehat{\lambda}_\psi)$ satisfies

$$\widetilde{\jmath}_{\mathrm{p}} = -\frac{\partial^2 \ell_{\mathrm{p}}(\psi)}{\partial\psi\partial\psi^{\mathrm{T}}} = \widetilde{\jmath}_{\psi\psi} - \widetilde{\jmath}_{\psi\lambda}\widetilde{\jmath}_{\lambda\lambda}^{-1}\widetilde{\jmath}_{\lambda\psi},$$

where a tilde denotes a quantity evaluated at $\widehat{\theta}_\psi = (\psi, \widehat{\lambda}_\psi)$. Deduce that $\widehat{\psi} \dot\sim \mathcal{N}_p(\psi, \widehat{\jmath}_{\mathrm{p}}^{-1})$.