MATH562 – Fall 2025

Problem Set: Week 9

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 misspecified exponential family models?

   (b) Compute $\imath_1(\theta_g) = \mathrm{Var}_g\{\frac{\partial}{\partial\theta_g}\log f(Y;\theta_g)\}$ and $\jmath_1(\theta_g) = -\,\mathrm{E}_g\{\frac{\partial^2}{\partial\theta_g^2}\log f(Y;\theta_g)\}$ and use them to determine the asymptotic variance $\mathrm{Var}(\hat\theta_g) \doteq \theta_g^3(1-\theta_g)/n$.

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

2. When the generalized Pareto distribution is written as

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

   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 $\Pr(Z_j > z)$ 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?

*3. The generalized Pareto distribution was given in an earlier question. $\lambda > 0$.

   (a) Show that the derivatives with respect to ψ satisfy the first two Bartlett identities only if $\lambda > 2$.
   Hint: Besides appearing in the MLE theory, the Bartlett identities were explicitly listed in the first set of slides. Use that $\int f(y;\psi)\,dy = 1$, for a density f independent of the parameter ψ, and the Leibnitz integration rule.

   (b) If M_n denotes the sample maximum, show that $n^{1/\lambda}(\psi - M_n) \xrightarrow{d} W$ as $n \to \infty$, where $\Pr(W > w) = \exp\{-(w/\psi)^\lambda\}$ for $w > 0$. Deduce that when $\lambda \le 2$ convergence to a limiting distribution for inference on ψ occurs more rapidly than with maximum likelihood estimation.

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 $\hat\theta$ has approximate distribution

$$\hat\theta = \begin{pmatrix}\hat\psi \\ \hat\lambda\end{pmatrix} \sim \mathbb{N}_{p+q}\left\{\begin{pmatrix}\psi \\ \lambda\end{pmatrix}, \begin{pmatrix}\hat\jmath_{\psi\psi} & \hat\jmath_{\psi\lambda} \\ \hat\jmath_{\lambda\psi} & \hat\jmath_{\lambda\lambda}\end{pmatrix}^{-1}\right\},$$

   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 $\mathrm{Var}(\hat{\psi}) \doteq (\hat{\jmath}_{\psi\psi} - \hat{\jmath}_{\psi\lambda}\hat{\jmath}_{\lambda\lambda}^{-1}\hat{\jmath}_{\lambda\psi})^{-1}$.

(b) Show that the profile log likelihood $\ell_p(\psi) = \ell(\psi, \hat{\lambda}_\psi)$ satisfies

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

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