Statistical Inference MATH-562

Anthony Davison

Problem 1 Investigate exponential tilting of a baseline density function $f_0(y)$ that is uniform on the interval $\mathcal{Y} = (0, 1)$, when the tilting functions are (a) $s(y) = y$ and (b) $s(y)^{\mathrm{T}} = (\log y, \log(1 - y))$.

Problem 2 If $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, find the approximate distribution of $Y = 1/\overline{X}$ for large n; take care when $\mu = 0$.

Problem 3 In a simple model for the spread of an epidemic in a large closed population of n identical individuals, it can be shown that the the expression $1 - \tau = e^{-R_0\tau}$ relates the basic reproduction number R_0 (the number of susceptible persons infected by a single infective person at the start of the epidemic) to the ultimate fraction τ of the population who are infected.

(a) Show that if $R_0 \le 1$ then there is only one possible value for τ, but that if $R_0 > 1$ then there are two possible values, and explain this heuristically.

(b) If it is positive, the final proportion infected can be estimated by $\widehat{\tau}$, which has an approximate normal distribution with mean τ and variance σ^2/n, where

$$\sigma^2 = \frac{\tau(1-\tau)}{\{1-(1-\tau)R_0\}^2}\{1 + c^2(1-\tau)R_0^2\},$$

where $c = \mathrm{var}(T)^{1/2}/\mathrm{E}(T)$ is the coefficient of variation of the infectious period for an individual, T. Hence obtain an estimator of R_0, and show that this is approximately normal with mean R_0 and variance σ_R^2/n, where $\sigma_R^2 = \{1 + c^2(1-\tau)R_0^2\}/\{\tau(1-\tau)\}$.

(c) Find c^2 when T is constant, uniform on some interval, and exponential. Hence suggest why an upper bound for the variance might be obtained by setting $c = 1$.

Problem 4

(a) The random variable Y follows a Lomax distribution, i.e.,

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

where $\alpha, \theta > 0$ are unknown. Is this an exponential family distribution?

From here on $Y_1, \ldots, Y_n$ represent independent identically distributed variables from the Lomax distribution with unknown θ and known $\alpha > 2$.

(b) Given that

$$\mathrm{E}(Y) = \frac{\theta}{\alpha - 1}, \quad \mathrm{var}(Y) = \frac{\alpha\theta^2}{(\alpha-1)^2(\alpha-2)},$$

use $Y_1, \ldots, Y_n$ to obtain a method-of-moments estimator $\tilde{\theta}$ of θ, and compute $\mathrm{var}(\tilde{\theta})$. Is $\tilde{\theta}$ biased?

(c) Another estimator of θ is $\tilde{\theta}_c = c\overline{Y}$, where $\overline{Y} = n^{-1}(Y_1 + \cdots + Y_n)$ and $c > 0$. Compute the bias and variance of $\tilde{\theta}_c$. What value of c minimises the mean square error of $\tilde{\theta}_c$?