University of Technology Sydney
School of Mathematical and Physical Sciences

Mathematical Statistics (37262) –
Class 11 Preparation Work

1. The number of calls a helpdesk receives each hour is modelled as an independent realisation of a $Poi(\lambda)$ variable for some unknown $\lambda > 0$.

   Prior to the gathering of any observations, it is believed that the uncertainty around the value of λ can be described by a gamma distribution, $\lambda \sim Gamma(\alpha, \beta)$ for some $\alpha, \beta > 1$.

   That is, $f(\lambda) = \dfrac{\beta^{\alpha} \lambda^{(\alpha-1)} e^{-\beta\lambda}}{\Gamma(\alpha)}$.

   Observations $x_1, x_2, ..., x_n$ of the number of calls in an hour are made for each of the first *n* hours of a day.

   i) Find the likelihood $L(x_1, x_2, ..., x_n | \lambda)$ of observing the dataset $x_1, x_2, ..., x_n$ for a given value of λ.

   ii) Find the posterior distribution for λ, describing the uncertainty around the parameter's value after the dataset has been observed.
   Show that the gamma distribution is a conjugate prior for observations from a Poisson distribution.

   iii) What is the mean of the posterior distribution?

Hint: You may assume without proof that, for $\lambda \sim Gamma(\alpha, \beta)$, $E(\lambda) = \dfrac{\alpha}{\beta}$.

2. Samples are drawn from a normal random variable $N(\theta,1)$ with known variance 1 but where the value of the mean $\theta > 0$ is not known.

i) Find the likelihood $L(y_1,y_2,...,y_n|\theta)$ of observing the dataset $y_1,y_2,...,y_n$ for a given value of θ.

ii) The prior belief about θ is that $\theta \sim \exp(10)$.

Show that the posterior distribution of θ given observations $y_1,y_2,...,y_n$ is given by

$$f(\theta|y_1,y_2,...,y_n) = \frac{e^{\frac{-\sum_{i=1}^{n}(y_i-\theta)^2-20\theta}{2}}}{\int_0^\infty e^{\frac{-\sum_{i=1}^{n}(y_i-\theta)^2-20\theta}{2}}\,d\theta}$$

iii) Explain how, given a symmetric proposal distribution and a fixed starting value θ_0, the Metropolis-Hastings algorithm could be applied to simulate samples from the posterior distribution of θ.

Clearly describe each step of the algorithm.

3. Independent samples $x_1,x_2,...,x_n$ are drawn from a uniform random variable $X \sim U[0,\theta]$ where the value of $\theta > 0$ is not known.

i) Write down the probability density function of $X \sim U[0,\theta]$.

ii) Hence find the likelihood of the sample $x_1,x_2,...,x_n$, $L(x_1,x_2,...,x_n|\theta)$.

Prior to any observations being made, the belief around the value of θ is described a Pareto variable, $\theta \sim Pareto(m,\alpha)$.

That is, $f(\theta) = \begin{cases} \dfrac{\alpha m^\alpha}{\theta^{\alpha+1}} & \theta \in [m,\infty) \\ 0 & \text{otherwise} \end{cases}$ for $m > 0, \alpha > 0$.

iii) Show that the posterior distribution describing the belief around the value of θ is also described by a Pareto variable.

iv) Is the Pareto distribution a conjugate prior for this likelihood function? Justify your answer.

Hint: Think about the range of $X \sim U[0,\theta]$ when considering the likelihood function.