**COMPUTER SCIENCE TRIPOS Part IB – mock – Paper 6**

## 2 Data Science (DJW)

Let $x_1, \ldots, x_n$ be observed values, which we believe are sampled independently from the distribution Uniform$[\mu - \theta, \mu + \theta]$, for some parameters $\mu \in \mathbb{R}$ and $\theta > 0$.

(*a*) Suppose $\mu$ is known and $\theta$ is unknown. Use $\Theta \sim \text{Pareto}(b_0, \alpha_0)$ as the prior for $\theta$, where $b_0$ and $\alpha_0$ are constants. (The Pareto distribution is described below.)

(*i*) What is the prior density of $\Theta$? [1 mark]

(*ii*) Show that the posterior distribution of $\Theta$ is Pareto, and give its parameters. [5 marks]

(*iii*) Calculate a 95% posterior confidence interval for $\Theta$. [4 marks]

(*b*) Suppose $\mu$ and $\theta$ are both unknown. Use Normal$(c_0, \sigma_0^2)$ as the prior for $\mu$, and Pareto$(b_0, \alpha_0)$ as the prior for $\theta$. Here $c_0$, $\sigma_0$, $b_0$, and $\alpha_0$ are all constants.

(*i*) Find the joint posterior density of the two parameters. [*Note:* Leave your answer as an unnormalized density function.] [3 marks]

(*ii*) Give pseudocode to generate a weighted sample from this density. Your code should produce a list of $m$ sampled pairs $\big[(\mu_1, \theta_1), \ldots, (\mu_m, \theta_m)\big]$ together with weights $[w_1, \ldots, w_m]$. [3 marks]

(*iii*) Give pseudocode to find a 95% posterior confidence interval for $\Theta$. [4 marks]

*Note:* If $X \sim \text{Pareto}(b, \alpha)$ then it has cumulative distribution function

$$\mathbb{P}(X \leq x) = \left[1 - \left(\frac{b}{x}\right)^{\alpha}\right] 1_{x \geq b}$$

and it may be sampled using

```
b * (1 + numpy.random.pareto(a=α))
```