6 Data Science (djw1005)

This question concerns a machine for making widgets. The machine is unreliable, and the proportion of defective widgets varies from day to day. For a given day, let Θ be the probability that a given widget is defective, and assume that widgets that day are conditionally independent given Θ.

You may give either analytical or computational answers. If you give computational answers, you must provide clearly-commented pseudocode.

(*a*) You sample n widgets, and find that x are defective. Find a posterior confidence interval for Θ, given the prior belief that $\Theta \sim \text{Beta}(\alpha, \beta)$. [4 marks]

(*b*) An alternative sampling strategy is to keep sampling widgets until you find the first defective widget. Let x be the number of widgets you end up sampling. Find a posterior confidence interval for Θ, given the prior belief that $\Theta \sim \text{Beta}(\alpha, \beta)$. [3 marks]

(*c*) For the sampling strategy in part (*b*), find the expected number of samples needed. Assume that α and β are positive integers with $\alpha > 1$; your answer should be a function of α and β. [4 marks]

(*d*) An engineer tells you that on a given day, the machine is either healthy or cranky, each equally likely, and that $\Theta \sim \text{Beta}(\alpha_m, \beta_m)$ where $m \in \{h, c\}$ indicates whether the machine is healthy or cranky. Find the posterior probability that the machine is cranky, for each of the two sampling strategies above. [9 marks]

[*Note:* The Beta distribution has density $\Pr(x; \alpha, \beta) = x^{\alpha-1}(1-x)^{\beta-1}/B(\alpha, \beta)$ where, for integer α and β, $B(\alpha, \beta) = (\alpha-1)!(\beta-1)!/(\alpha+\beta-1)!$]