7 Foundations of Data Science (djw1005)

(a) Let $X_1, \ldots, X_n$ be independent binary random variables, $\P(X_i = 1) = \theta$, $\P(X_i = 0) = 1 - \theta$, for some unknown parameter $\theta$. Using Uniform[0,1] as the prior distribution for $\theta$, find the posterior distribution. [Note: For your answer, and in answer to parts (b) and (d), give either a named distribution with its parameters, or a normalised density function.] [3 marks]

I have collected a dataset of images, and employed an Amazon Mechanical Turk worker to label them. The labels are binary, nice or nasty. To assess how accurate the worker is, I first picked 30 validation images at random, found the true label myself, and compared the worker’s label. The worker was correct on 25 and incorrect on 5.

(b) Let $\theta$ be the probability that the worker labels an image incorrectly. Using Beta(0.1, 0.5) as the prior distribution for $\theta$, find the posterior. [3 marks]

I next ask the worker to label a new test image, and they tell me the image is nice. Let $z \in \{\text{nice, nasty}\}$ be the true label, and let the prior distribution for $z$ be $\Pr(\text{nice}) = 0.1$, $\Pr(\text{nasty}) = 0.9$.

(c) For both $z = \text{nice}$ and $z = \text{nasty}$, find

$$\P(\text{worker says nice} \mid z, \theta).$$

Hence find the posterior distribution of $(z, \theta)$. Your answer may be left as an un-normalised density function. [5 marks]

(d) Find the posterior distribution of $z$. [5 marks]

My colleague has more grant money and she can employ 3 workers to rate each image. On a test set of 30 images, she found that they all agreed on 15 images, worker 1 was the odd one out on 8 of the images, worker 2 was the odd one out on 4, and worker 3 was the odd one out on 3.

(e) Let $\theta_i$ be the probability that worker $i$ labels an image incorrectly. Find the posterior distribution of $(\theta_1, \theta_2, \theta_3)$. Your answer may be left as an un-normalised density function. [4 marks]

Hint. The Beta($\alpha, \beta$) distribution has mean $\alpha / (\alpha + \beta)$ and density

$$\Pr(x) = \binom{\alpha + \beta - 1}{\alpha - 1} x^{\alpha - 1} (1-x)^{\beta - 1}, \quad x \in [0,1].$$