Possible points: 100

1. **Problem 1 [30 pts]**

   A startup company A has a cool new product that they supply to company B, but it's a bit unreliable.

   Company A has in the past sent two large shipments to company B. Among those, 10% were defective. This is considered the prior information.

   Now, company B wants to give company A one more chance, and orders a shipment of 50 units. The shipment arrives, and out of that shipment, only 1 unit is defective. This is considered the data at hand.

   Encouraged, the company B is considering ordering a large shipment of size 500 in the future.

   (a) What would be the best frequentist estimate of the proportion of defective items for this new shipment of size 500?

   (b) What is the variance of the estimator you used in part a?

   (c) Let's do a Bayesian analysis of this problem now. Propose a prior distribution based on the available prior information.

   (d) Derive the posterior density based on the data (1 out of 50 defective items) and the above prior

   (e) Derive the posterior predictive distribution for the number of defective items for a future sample of 500

   (f) Plot the probability mass function derived above

   (g) Under the square error loss, find the Bayesian estimator of the number of defective items for this new shipment of size 500

   (h) Provide the actual estimate of the number of defective items for this new shipment of size 500 using the estimator above

   (i) Provide the central 95% credible interval for the number of defective items for this new shipment of size 500

   (j) Provide the 95% highest probability density (HPD) credible interval for the number of defective items for this new shipment of size 500

   (k) Report the estimate and the 95% HPD for the analyses with 2 alternative priors; one pessimistic and one optimistic.

2. **Problem 2 [50 pts]**

   197 animals are distributed into 4 categories: $Y = (y_1, y_2, y_3, y_4)$ according to the genetic linkage model, where the probabilities of falling into the 4 categories are given as:

   $$((2+\theta)/4,\ (1-\theta)/4,\ (1-\theta)/4,\ \theta/4).$$

   (a) What is the likelihood for the data $Y = (125, 18, 20, 34)$

   (b) What is the likelihood for the data $Y = (14, 0, 1, 5)$

   (c) Use the Newton-Raphson algorithm to obtain the MLE $\widehat{\theta}$ of θ for the data $Y = (125, 18, 20, 34)$. Try starting your algorithm at $\theta = 0.1$, 0.3, 0.6, 0.9. How did you assess the convergence of this algorithm?

(d) Repeat (c) for the data $Y = (14, 0, 1, 5)$. Did the algorithm converge for all the starting points?

(e) Plot the normalized likelihood and the associated normal approximation in the same figure for the data $Y = (125, 18, 20, 34)$. Discuss the adequacy of the normal approximation.

(f) Repeat (e) for the data $Y = (14, 0, 1, 5)$

(g) Using a Beta $Be(7, 3)$ prior on θ, derive the posterior density for θ for the data $Y = (14, 0, 1, 5)$, and repeat part (d). Did the algorithm converge for all starting values now?

3. **Problem 3 [10 pts]**

Consider a generalized absolute error loss function given as follows:

$L(\theta, \delta) = k_2(\theta - \delta)$ if $\theta > \delta$ and $L(\theta, \delta) = k_1(\delta - \theta)$ if $\theta \leq \delta$

Find a Bayes estimator associated with this loss, for an arbitrary prior π. Show all work.