Bayesian Inference for Surveys
Homework # 3

1. Let $Y = \{Y_1, Y_2, Y_3, Y_4\}$ be a population of size $N = 4$ and let a simple random sample of size $n = 3$ be drawn without replacement. The goal is to estimate the population mean, $(Y_1+Y_2+Y_3+Y_4)/4$. The following table shows an estimate (t) of the population mean for each of the 4 possible samples.

Sample Number	Sample	Estimate (t)
1	(Y_1, Y_2, Y_3)	$(1/3 + c)Y_1 + (1/3 - c)Y_2 + Y_3/3$
2	(Y_1, Y_2, Y_4)	$(1/3 - c)Y_1 + (1/3 + c)Y_2 + Y_4/3$
3	(Y_1, Y_3, Y_4)	$(Y_1 + Y_3 + Y_4)/3$
4	(Y_2, Y_3, Y_4)	$(Y_2 + Y_3 + Y_4)/3$

   (a) Assume that Y is fixed and c in the definition of t is known (only random variable is the inclusion indicator, I). Show that t is unbiased.

   (b) Compute the sampling variance of t.

   (c) Show that the sampling variance of t can be made arbitrarily smaller than the sampling variance of the sample mean $\bar{y}$, for an appropriate choice of c and thus proving that the sample mean is not the uniformly minimum variance unbiased estimator of the population mean, for all values of Y. [Note: A general result that uniformly minimum variance unbiased (linear) estimate of the population mean does not exist for almost all designs was proved by Godambe (1955, JRSS, Series B)].

2. Let $y_1, y_2, \ldots, y_n$ be a random sample of size n from a normal distribution with mean θ and variance $c^2\theta^2$ where c is known and $\theta > 0$. Find the maximum likelihood estimate of θ and its asymptotic variance.

3. A firm has 120 employees and wants to decide between two firms, A and B, for training their employees. A sample 20 employees are chosen with ten employees each sent to the two firms for training. Upon their return, all 20 employees were assessed using a proficiency test. The scores of the 20 employees are given in the table below. Assume that the scores are normally distributed and the non-informative prior, $\pi(\mu, \sigma) \propto \sigma^{-1}$ for the mean and standard deviation.

Firm A	Firm B
60	68
72	90
78	88
92	88
78	68
57	72
62	72
72	92
88	50
71	99

   (a) What is posterior distribution of the population mean of the scores, if all 120 employees were trained by firm A (firm B)?

   (b) A score of 75 or more on the test is classified as proficient. What is the posterior distribution of the proportion of proficient employees if all 120 employees were trained by firm A (firm B)?

   (c) Based on (a) and (b), do you have any reasons to prefer one firm over the other?