

**HW2 STT 465
(MSU, Fall, 2015)**

Due: Wednesday, Sept. 30th, 2015 in class (hard copy) or by e-mail (pdf).

Note: support all your answers with either derivations or computing code (preferable R). For computing code provide an appendix with sub-titles indicating what question does the code applies to.

1. Beta Binomial Model

Gout is a complex form of arthritis, characterized by sudden, severe attacks of pain, redness and tenderness in the joint (Mayo Clinic). The incidence of Gout is known to vary between male and female and also between ethnic groups (Singh, 2013; PMCID: PMC3545402).

The data set gout.txt (<https://github.com/gdlc/STT465/blob/master/gout.txt>) contains data on Gout, sex and ethnicity for a total of 3,211 patients. The objective is to infer the incidence of Gout by sex and ethnic group (i.e., for groups defined as follows: WF=white-female, WM=white-male, BF=black-female, BM=black-male) using the Beta-binomial model discussed in the class.

1.1. (Maximum Likelihood) Using the Gout data set, compute and report the Maximum Likelihood Estimator (MLE) and an approximate 95% CI for the probability of developing Gout by group. Write down the likelihood function, provide a derivation of the MLE, its expected value and variance, and evaluate the estimator for each of the groups.

1.2. (Prior). Consider a Beta prior with shape parameters shape1=1.2 and shape2=2. For this prior: (i) Present plot of the prior density over it's support [0,1], (ii) report the prior mean and (iii) a prior 95% credibility region.

1.3. (Analytics) (i) Write down the Beta-Binomial model (i.e., likelihood and prior) , (ii) derive the Posterior distribution of the probability of Gout given the data (at this point ignore sex/ethnicity, i.e., assume a homogeneous population and present your results in terms of the number of cases and numbers of controls, or functions of it), and (ii) present an expression for the posterior mean and posterior variance of the probability of developing Gout.

1.4. (Stratified analysis) Using the Gout data, the prior distribution of 1.1, and the posterior distribution derived in 1.3, report:

- The posterior mean by group
- A 95% posterior credibility interval for each of the groups
- A single plot with a display of the four posterior densities (i.e., one per group).

Note: for this question carry out stratified analyses, i.e., separate analyses per group.

1.5. Summarize your findings (2-3 sentences) from questions 1.1-1.4.

2. Monte Carlo Methods

The objective is to estimate the posterior distribution of the odds, $\frac{\theta}{1-\theta}$, of developing Gout per group.

2.1. Propose a MC algorithm to draw samples from the posterior distribution of the odds.

2.2. Implement the algorithm proposed in 2.1 and report:

- (1) Estimated posterior mean, posterior standard deviation and 95% posterior credibility region.
- (2) An approximate estimate of the MC standard error of the estimated posterior mean.