University of Manitoba
Department of Statistics

**STAT 4600: Computational Statistics**

Assignment 7

Due: Thursday, April 5

## Question 1:

Consider an experiment where one of two coins is selected at random and tossed $m$ times. Coin 1, with both sides showing *Heads*, is selected with probability $w$. Coin 2, which gives *Tails* with probability $p$ when tossed, is chosen with probability $1-w$.

Let $X$ be the number of outcomes of *Tails* observed in the $m$ tosses. It is possible to show that

$$\mathbb{P}(X=x)=\begin{cases} w+(1-w)(1-p)^m & \text{if } x=0,\\ (1-w)\binom{m}{x}p^x(1-p)^{m-x} & \text{if } x=1,\dots,m. \end{cases}$$

This distribution is known as the zero-inflated binomial distribution.

Now, the coin flipping experiment was performed a total of $n=91$ times. For each repetition of the experiment, the selected coin was flipped $m=5$ times and the number $i$ of *Tails* (out of 5 tosses) was recorded. This lead to the data shown in Table 1.

(A) Draw contours of the likelihood of $\theta=(w,p)'$. Then, add the parametrized curve $(p,\hat{w}(p))'$ to your contours. Explain what is the link between this curve and the contours.

Finally, numerically obtain $\hat{\theta}_{MLE}$ through maximizing the (bivariate) likelihood of $\theta$.

(B) Graph the profile log-likelihood of $p$.

Then, obtain $\hat{\theta}_{MLE}$ by making use of this profile log-likelihood.

**Note:** For the above, you may use the following without proof. However, do make sure that you know how to derive these results.

If the coin flipping experiment is repeated $n$ times, leading to observations $\mathbf{x}=(x_1,x_2,\dots,x_n)'$, then

$$L(\theta;\mathbf{x})=\left[w+(1-w)(1-p)^m\right]^{n_0}(1-w)^{n-n_0}p^{n\bar{x}}(1-p)^{n(m-\bar{x})-n_0m}\prod_{i=1}^{n}\binom{m}{x_i},$$

where $\bar{x}$ denotes the usual sample mean, and $n_0$ the number of observations equal to zero.
Also, the MLE of $w$, for a known value of $p$, is given by

$$\hat{w}(p)=\frac{n_0-n(1-p)^m}{n[1-(1-p)^m]}.$$

Table 1: Data for Question 1 – the Frequency row refers to the number of repetitions of the experiment that lead to $i$ *Tails* from flipping the selected coin 5 times.

| Number of *Tails* $i$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Frequency | 39 | 14 | 22 | 12 | 3 | 1 |

**Question 2:**

In the development of drugs, bioassay experiments are often performed on animals. In a typical experiment, various dose levels of a compound are administered to batches of animals and a binary outcome (positive or negative) is recorded for each animal.

Specifically, consider a toxicity study where an experiment was performed and led to data in the form of a dose level (in log g/ml), the number of deaths and the number of animals in each of four groups. Let $Y_i$ denote the number of deaths observed in the $i^{\text{th}}$ group out of $n_i$ animals with dose level $x_i$. The observed data are displayed in Table 2.

Now, assume

$$Y_i \mid \theta_0, \theta_1 \sim \text{Binomial}(n_i, p_i),$$

where the probability $p_i$ of a death in group $i$ follows the logistic model

$$\log\left(\frac{p_i}{1-p_i}\right) = \theta_0 + \theta_1 x_i$$

or,

$$p_i = \frac{e^{\theta_0+\theta_1 x_i}}{1+e^{\theta_0+\theta_1 x_i}}.$$

Taking a Bayesian approach to inference, we assume a priori that

$$\Theta_0 \sim \text{Unif}(-5,5) \qquad \text{and} \qquad \Theta_1 \sim \text{Unif}(0,50).$$

Note that $\theta_1$ is assumed to be positive as it is believed toxicity (and risk of death) increases with an increase in dosage.

(A) Find the posterior mode.

(B) Approximate the Bayes estimators of $\theta_0$ and $\theta_1$ by running an appropriate Monte Carlo experiment.

(C) Construct a 90% equal-tails credible region for $\theta_1$.

(D) Find the posterior median of $\Theta_1$.

(E) An important parameter, referred to as LD-50, is defined as $\lambda = -\theta_0/\theta_1$ and corresponds to the dose $x$ for which the probability of death is equal to 1/2.

Approximate the Bayes estimator of $\lambda$, given by

$$\hat{\lambda}_B = \mathbb{E}[\Lambda|Y=y] = -\mathbb{E}\big[\Theta_0/\Theta_1\big|y\big],$$

where $y = (y_1, y_2, y_3, y_4)'$ denotes the vector of observed $y$ values.

Table 2: Data for Question 2 – outcomes of the bioassay experiment

| Dose ($x_i$) | Deaths ($y_i$) | Sample size ($n_i$) |
|---|---|---|
| -0.86 | 0 | 5 |
| -0.30 | 1 | 6 |
| -0.05 | 3 | 5 |
| 0.73 | 5 | 6 |