Assignment Report

Jayashree P(cs16resch11002)
Rashmi HTI(cs16mtech11013)
Sherin Thomas(cs16mtech11016)

Dogs: loglinear model for binary data

Lindley analyses data from Kalbfleisch on the Solomon-Wynne experiment on dogs, whereby they learn to avoid an electric shock. A dog is put in a compartment, the lights are turned out and a barrier is raised, and 10 seconds later an electric shock is applied. The results are recorded as success (Y=1) if the dog jumps the barrier before the shock occurs,or failure (Y=0) otherwise.

Thirty dogs were each subjected to 25 such trials. A plausible model is to suppose that a dog learns from previous trials, with the probability of success depending on the number of previous shocks and the number of previous avoidances. Lindley thus uses the following model

$p_j = A^{x_j} B^{j-x_j}$
for the probability of a shock (failure) at trial j, where
x_j = number of success (avoidances) before trial j and
$j - x_j$ = number of previous failures (shocks).
This is equivalent to the following log linear model $\log p_j = a x_j + b(j-x_j)$

Implementation details:

1. Using Markov Chain Monte Carlo(MCMC):

The two parameter of the model α and β are initialized as α=-1 and β=-1.

From the given information,the probability of each dog getting a shock in a trial is assumed to be similar to a bernoulli distribution with 2 outcomes ie 1 = dog gets a shock and 0 = dog avoids shock.Here the probability of getting a shock associated with each dog in each trial is different and it depends upon all previous histories of shock and avoidance associated with a dog under consideration.

Therefore ,
likelihood $P(D \mid \alpha,\beta) = (p_{1,1}^{X_{1,1}} * (1-p_{1,1})^{1-X_{1,1}}) * (p_{1,2}^{X_{1,2}} * (1-p_{1,2})^{1-X_{1,2}}) * \ldots * (p_{30,25}^{X_{1,2}} * (1-p_{30,25})^{1-X_{1,2}})$

The prior for α and β are assumed to be normal distributions with 0 mean and unit variance.
α~Normal(0,1) and β~Normal(0,1).

Here the prior is not conjugate to the likelihood.Hence we cannot compute the posterior analytically.

So MCMC sampling is used to approximate the posterior.With MCMC we draw samples from a simple proposal distribution where each draw depends only upon the state of the previous draw.

Here we are using a normal distribution as the proposal distribution and samples values for α and β which are less than 0.00001.

Under certain condiitons, the Markov chain will have a unique stationary distribution. In addition, not all samples are used, instead we set up acceptance criteria for each draw based on comparing successive states with respect to a target distribution that ensure that the stationary distribution is the posterior distribution of interest.

We find a target distribution where , posterior ∝ likelihood × prior.

After some time, the Markov chain of accepted draws will converge to the stationary distribution, and we can use those samples as draws from the posterior distribution, and find functions of the posterior distribution.

The acceptance ratio is calculated as
$\rho = p(X|\alpha_{new},\beta_{new})p(\alpha_{new},\beta_{new}) * q(\alpha_{old},\beta_{old}) / p(X|\alpha_{old},\beta_{old})p(\alpha_{old},\beta_{old}) * q(\alpha_{new},\beta_{new})$
where α_{new} and β_{new} are the proposed values for α and β.
Then we generate a random number in the range (0,1).
If the generated random number is less than ρ,the proposed values for α and β are accepted.
A burn-in period is specified ,upto which the samples are rejected even if the generated random number is less than ρ.
Here the burn-in period is used as 10 percent of the number of iterations.
The samples from the post burn-in portion of the iterations are used for predicting the probability of a dog getting shock in the subsequent trials.
Predictions:

Number of shocks in previous trials	**Number of avoidances in previous trials**	**Probability of getting a shock in the next trial**
1	0	0.779
2	0	0.607
2	1	0.473
4	1	0.289
4	4	0.138
4	7	0.066
4	11	0.025
6	11	0.0158
8	14	0.004
8	17	0.002

2. Variational Inference

Posterior ∝ likelihood × prior

Here likelihood is Bernoulli like distribution and the prior is assumed to be Gaussian distribution. Hence the posterior is not in closed form.
So we go for approximation of the posterior using Variational inference, and the approximate distribution is assumed to be Gaussian.
Approximate distribution for α and β has the free parameters m1,v1 and m2,v2 respectively, which we want to optimise so as to make the approximate distribution similar to the actual posterior.

So we try to minimize the KL-divergence between the two distributions, $KL(q(w)||p(w|Y,X))$.

Alternatively we try to maximize the lower bound on the log likelihood of the data.

Lower Bound : $-KL(q(w)||p(w)) + E_q[\log p(Y|X,w)]$

We define an optimization function into which we pass a function handler for the Lower bound calculation and the free parameters with initial values.

In the lower bound function, we find the KL-divergence between the approximate distribution for each of the sampled values of m1,v1,m2,v2 and the Prior distribution first. And then we add it to the expectation of the log likelihood wrt to the approximate distribution.

We iterate through the optimization function until convergence