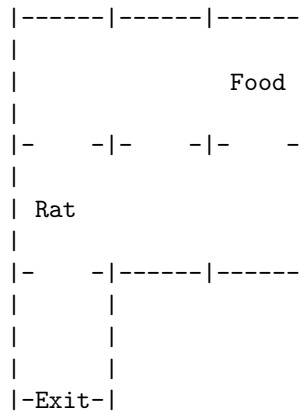


Statistics 831: Probability and Statistical Modeling

Homework #8
Due 3/30/2020

Problems:

1. Solve the following by defining an appropriate Markov chain and using first-step analysis:
 - (a) In repeated (fair) coin tosses:
 - i. Calculate the expected number of tosses until 2 successive heads.
 - ii. Of the two patterns HHT and HTH, which do you expect to observe first and why.
 - (b) Rat maze: for a rat located as shown performing a random walk calculate the probability of finding food before exiting.



Note that the exit is on the wall, ie the rat can move down 1 step without exiting the maze.

2. Define a Markov chain on $0, \dots, N$ by:

$$P_{i,j} = \begin{cases} 1 & i = j = 0 \\ 1/i & 0 \leq j < i \leq N \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

i.e. from state j the chain is equally likely to go to $0, \dots, j-1$.

- (a) Determine the fundamental matrix for the transient states.
 - (b) Determine the distribution of the last positive integer visited.
3. Choose one of the above problems (either 1(a)ii, 1b, or 2b) and validate your results by:
 - (a) Computing large powers of the transition matrix.
 - (b) Simulating realizations of the Markov chain.

4. The Gibbs sampler defined in class cycles through coordinates in a fixed order. Alternatively, we may define a *random-scan* Gibbs sampler, which iteratively chooses $i \in \{1, \dots, d\}$ at random (according to probabilities p_i say), and sets

$$\theta^{(n+1)} = (\theta_1^{(n)}, \dots, \theta_{i-1}^{(n)}, \theta_i^*, \theta_{i+1}^{(n)}, \dots, \theta_d^{(n)})$$

with

$$\theta_i^* \sim \pi(\theta_i \mid \theta_1^{(n)}, \dots, \theta_{i-1}^{(n)}, \theta_{i+1}^{(n)}, \dots, \theta_d^{(n)})$$

- (a) Show that this also produces a Markov chain with stationary distribution π .
 - (b) Give sufficient conditions for this chain to have limiting distribution π .
5. The famous “braking data” of Tukey (1977) is available in R under the dataset name **cars**. It gives the speeds traveled (mph) and braking distances (feet) for 50 cars. It is thought that a good model for this dataset is a quadratic model:

$$y_{ij} = a + bx_i + cx_i^2 + \epsilon_{ij} \quad \text{for } i = 1, \dots, k; \quad j = 1 \dots, n_i$$

- (a) Write down the likelihood function assuming $\epsilon_{ij} \sim N(0, \sigma^2)$.
- (b) Obtain estimates of a, b, c and σ^2 from a standard linear regression.
- (c) View the likelihood in part (a) as a posterior distribution under flat priors, and construct a Metropolis-Hastings algorithm to sample from it using a *Metropolized independence sampler* with proposal distributions selected based on your estimates in (b). Use normals for a, b, c and inverse-gamma for σ^2 .
- (d) Make histograms of the posterior distributions of the parameters. Show any plots or diagnostics used to monitor convergence.
- (e) Consider robustness by modifying the error distribution to $\epsilon_{ij} \sim t_4(0, \sigma^2)$ and re-running your analysis. Do you need to modify your proposal distributions?