Assignment 4
CS 230 /CS 561: Probability and Statistics for CS
Course Instructor : Dr. Satyanath Bhat
Teaching Assistant : Vipin Gautam, Divya Mishra
Teaching Fellow : Prajyot Patil

INSTRUCTIONS: Please submit your solutions for assignment in google classroom.

I Bernoulli Process

1. Simulate a Bernoulli process and empirically estimate the probability mass function for the number of trials till k^{th} arrival. Use $p = 0.5$. Optionally, compare the empirical pmf with the theoretical pmf (Pascal's or Negative Binomial distribution).

2. Using a simulation of Bernoulli process, estimate the probability mass function of the length of the first continuous strings of zeros. For example, for the sequence 1,1,0,0,0,1,1,0,1 ... the length of first continous string of zeros is 3. Use $p = 0.5$. Optionally compare it with the theoretical pmf (Geometric).

II Poisson Process

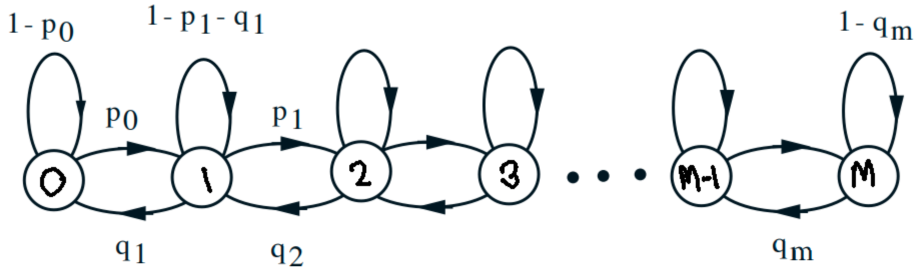
1. Imagine that there are two queues in a hospital, one for senior citizens and another for general citizens. Assume both these are poisson processe with rates 1 per hour and 2 per hour. These queues are combined at some point to generate a new single queue. Estimate (and plot) the pmf of the combined queue using simulation. For simulation, use samples (arrivals/hr) from appropriate Poisson distributions. Optionally, compare the empirical pmf you generated with theoretical pmf (Poisson with rate 3/hr)

2. Generate a Poisson process with rate 2 per hour. Record the time of the k^{th} arrival in the Poisson process. Generate several samples and plot the histogram. Optionally compare with the pdf of Erlang's distribution. [Hint: For Poisson process simulation, there are two possible ways. One of which is to choose a very small delta and look at poisson arrivals in that period delta. Another approach is to simulate interarrival time as exponential distribution.]

3. A certain bus company operates buses according to a Poisson process with rate 1 bus every 10 minutes. Simulate a day of bus operations and record every bus arrivals during the day at a certain bus stop. The transport inspector randomly arrives at a bus stop and notes the interarrival times between the last bus (which he missed) and the next bus. Using several simulations, estimate the mean interarrival time in the process described above, compare your estimate with the one predicted by theory.

III Markov Chain

Consider a Markov Chain with state transistion shown below. Use $p_i = q_i = 1/4$ and $M = 4$. Simulate a run of markov chain begining at state 0 at n=0 and estimate the n=100 step transistion probabilities. Assuming that the markov chain has achieved steady state, compare your estimate with the theoretical steady state probabilities.



IV Statistics

Consider an election between two candidates A and B. Consider an exit poll problem where the statistician wishes to estimate the support μ_A for candidate A. The statistician wishes to estimate the support $\hat{\mu}_A$, such that with probability 0.05 the estimate is NOT within the interval $[\mu_A - 0.03, \mu_A + 0.03]$. Mathematically, it is required that,

$$P(|\hat{\mu}_A - \mu_A| \geq 0.03) \leq 0.05$$

1. Using Chebyshev's inequality, check that the number of individual whose opinion should be sought is 5556. You might use the upper bound of 0.25 for the variance term in your calculations. Please DO NOT submit this.

2. Using Central Limit theorem, check that the number of individual whose opinion should be sought is 1067. You might use the upper bound of 0.25 for the variance term in your calculations. Please DO NOT submit this.

Now, observe that the population of the country is not required and the actual support that candidate A receives is also immaterial in the calculations. Isn't this fantastic! Now, using simulations, we would like to verify how well the statistician performs in practice using these tools. Towards this assume that candidate A has support $\mu_A = 0.55$, simulate several runs of exit polls using both strategies (n=5556 and n = 1067). In each run, check if the empirical estimate $\hat{\mu}_A$ is within $[\mu_A - 0.03, \mu_A + 0.03]$. Using this method, you can therefore estimate the probability $P(|\hat{\mu}_A - \mu_A| \geq 0.03)$ for both strategies. Report how well were each of the strategies performed in practice and which one would you use.