Statistical Simulations

Evana Sorfina Mohd Nazri

Mentor: Medha Agarwal

During the course of the Statistics and Probability Association's Directed Reading Program, I studied methods of obtaining independent and identically distributed random samples for both continuous and discrete random variables. I learned both the mathematical theory behind generating IID samples, as well as methods in coding them in R. This project was mostly theory heavy and so my slides are, *not slides*, as the math and graphs can be more clearly seen as a pdf file! This report will be a summary of the slides and a wrap up of what I have learned. A more detailed explanation can be viewed in the slides.

The motivation behind sampling is that often times we do not have a closed form solution for distributions. The lack of closed form solutions limits us from calculating the expectation for these distributions directly. Thus, we need to generate samples because it enables us to get the expectations by calculating from the samples. In the big picture sense, if we have some kind of computer program, we can ask it to generate any type of distribution we want it to generate, such as the exponential distribution, normal distribution, gamma distribution. But behind the scenes the computer is doing something to generate numbers from that distribution, and these are a few of the methods of how the computer generates it.

The building block of computational simulation is the generation of uniform random numbers. If we can draw from $U(0, 1)$, then we can draw from most other distributions. Computers can generate numbers between $(0, 1)$, which although are not exactly random (and in fact deterministic), but have the appearance of being $U(0, 1)$ random variables. These draws from $U(0, 1)$ are pseudorandom (fake random!) draws. A few methods of generating these random numbers include the multiplicative congruential method, and the mixed congruential method. We can see that the resulting graph from generating these random numbers indeed gives us the uniform distribution, and we are now set to move on to generating other distributions that we want.

In the following weeks, we discussed methods of generating discrete and continuous random variables. Discrete distributions include Bernoulli, Poisson, Geometric, etc. In generating discrete random samples, we studied the inverse transform, and accept reject method. The inverse transform method works by generating a random number U, and finding the interval which U lies on based on the probabilities of p (of our target distribution). The next method in generating discrete random samples is the accept reject method. For accept-reject, suppose we have an efficient method for generating a random variable having pmf q, we can use it as a basis for simulating p. By first simulating a random variable Y having mass function q and accepting this simulated value with a probability proportional to py/qy. Other miscellaneous methods of generating discrete random samples include taking advantage of the relationship between two (or more) distributions.

Next, we discussed methods of generating continuous random samples for distributions such as the exponential, Gamma, and Beta distribution. Both of the methods above, inverse transform, and accept reject hold for such distributions (with a little modifications), and we also discussed an additional method which is importance sampling.

In importance sampling, we have our target distribution p, which we cannot sample from, and another distribution q which we can sample from. Importance sampling does not seem intuitive at first, as we are sampling from q to get the expectations of p, but with mathematical manipulation, we get that sampling from q often gets us accurate results for expectations, and in certain cases, sampling from q gives us a better estimation of the expectation than directly sampling from p! Such certain cases could be estimating the expectations of something happening at the tails of the distribution as for example in the insurance sector where we want to estimate the expectations of rare events happening. Lastly, the final mini example about optimal proposals show how we choose the importance distribution q that minimizes the variance estimators.

All in all, I really enjoyed the time I spent in the SPA DRP and I'm grateful for the opportunity it gave me to learn about simulations on a deeper level. I thank my mentor Medha Agarwal for being so patient with me throughout this quarter and helping me until the end with slides and presentation preparation!

References

Ross, Sheldon M. Simulation. Academic Press, 2013.