Math 697: Homework 1

Exercise 1 (a) For positive numbers a and b, the $Pareto(a,b)$ distribution has p.d.f $f(x) = ab^a x^{-a-1}$ for $x \geq b$ and $f(x) = 0$ for $x < b$. Apply the inversion method to generate $Pareto(a,b)$.
(b) The standardized logistic distribution has the p.d.f $f(x) = \frac{e^{-x}}{(1+e^{-x})^2}$. Use the inversion method to generate a random variable having this distribution.

Exercise 2 In class we have formulated the rejection method for continuous random variables but it can be extended to discrete random variables too.
(a) Suppose now X and Y are discrete random variables, both taking values in the same finite or countable set S. Formulate and prove the rejection method in this case.
(b) Set-up an algorithm to simulate a Poisson random variable with parameter λ using a geometric random variable with parameter p. Discuss the choice of p if λ is fixed.

Exercise 3 Consider the technique of generating a $\Gamma_{n,\lambda}$ random variable by using the rejection method with $g(x)$ being the p.d.f of an exponential with parameter λ/n.

1. Show that the average number of iterations of the algorithm is $n^n e^{1-n}/(n-1)!$.
2. Use Stirling formula to show that for large n the answer in 1. is approximately $e\sqrt{(n-1)/2\pi}$.
3. Show that the rejection method is equivalent to the following
   - **Step 1:** Generate Y_1 and Y_2 independent exponentials with parameters 1.
   - **Step 2:** If $Y_1 < (n-1)[Y_2 - \log(Y_2) - 1]$ return to step 1.
   - **Step 3:** Set $X = nY_2/\lambda$.

Exercise 4 (Generating a uniform distribution on the permutations)

In this problem we will use the following notation. If x is positive real number we denote by $[x]$ the integer part of x, i.e. $[x]$ is the greatest integer less than or equal x. For example $[2.37] = 2$.

Consider a permutation of $(1, 2, 3, \cdots n)$. We denote by S(i) the element in position i. For example for the permutation $(2, 4, 3, 1, 5)$ of 5 elements we have $S(1) = 2$, $S(2) = 4$, and so on.

Consider the following algorithm

1. Set $k = 1$
2. Set $S(1) = 1$
3. If $k = n$ stop. Otherwise let $k = k + 1$.
4. Generate a random number U, and let

$$S(k) = S([kU] + 1)\,,$$

$$S([kU] + 1) = k\,.$$

   Go to step 3.

Show that at iteration k, – i.e. when the value of $S(k)$ is initially set– $S(1), S(2), S(k)$ is a random permutation of $1, 2, \cdots, k$, i.e., all permuation are equally likely and occur with probability $1/k!$.

Hint: Relate the probability P_k on the set of permutation of obtained at iteration k with the probability P_{k-1} obtained at iteration $k-1$.

Exercise 5 On a friday night you enter a BBQ restaurant which promises that every customer is served within a minute. Unfortunately there are 30 customers in line and you an appointment will force you to leave in 40 minutes. Being a probabilist you assume that the waiting time of each customer is exponential is mean 1. Estimate the probability that you will miss your appointment if you wait in line until you are served using (a) Chebyshev inequality, (b) The central limit theorem, (c) Chernov bounds.

Exercise 6 (Hit-or-miss method)

1. Suppose that you wish to estimate the volume of a set B contained in the Euclidean space $\mathbf{R}^k$. You know that B is a subset of A and you know the volume of A. The "hit-or-miss" method consists in choosing n independent points uniformly at random in A and use the fraction of points which lands in B to get an estimate of the volume of B. (We used this method to compute the number π in class.) Write down the estimate I_n obtained with this method and compute $\text{var}(I_n)$. (This will be expressed in terms of the volume of A and B.)

2. Suppose now that D is a subset of A and that we know the volume of D and the volume of $D \cap B$. You decide to estimate the volume of B by choosing n points at random from $A \setminus D$ and counting how many land in B. What is the corresponding estimator I'_n of the volume of B for this second method? Show that this second method is better than the first one in the sense that $\text{var}(I'_n) \leq \text{var}(I_n)$.

3. How would you use this method concretely to improve the estimation of the number π? Compute the corresponding variances.

Exercise 7

Suppose f is a function on the interval $[0, 1]$ with $0 < f(x) < 1$. Here are two ways to estimate $I = \int_0^1 f(x)dx$.
(a) Use the "hit-or-miss" from the previous problem with $A = [0, 1] \times [0, 1]$ and $B = \{(x, y) : 0 \leq x \leq 1\,,\, 0 \leq y \leq f(x)\}\,.$
(b) Use the simple sampling algorithm with $U_1, U_2, \cdots$ be i.i.d. uniform random variables on $[0, 1]$ and

$$\hat{I}_n \;=\; \frac{1}{n}\sum_{i=1}^{n} f(U_i)\,.$$

Find which one of theses two methods is the most efficient.

Exercise 8 (Antithetic variables) In this problem we describe an example of a method to reduce the variance of the simple sampling method.

1. Suppose that k and h are both nondecreasing (or both nonincreasing) functions then show that
$$\mathrm{cov}(k(X), h(X)) \geq 0\,.$$
Hint: Let Y be a random variable which is independent of X and has the same distribution as X. Then by our assumption on h, k we have $(k(X) - k(Y))(h(X) - h(Y)) \geq 0$. Take then expectations.

2. Consider the integral $I = \int_0^1 k(x)dx$ and assume that k is nondecreasing (or nonincreasing). The simple sampling estimator is
$$I_n = \frac{1}{n}\sum_{i=1}^{n} k(U_i)\,.$$
where U_i are independent $U([0,1])$ random variables. Consider now the alternative estimator: for n even set
$$\hat{I}_n = \frac{1}{n}\sum_{i=1}^{n/2} k(U_i) + k(1-U_i)\,.$$
where U_i are independent $U([0,1])$ random variables. Show that I_n is an estimator for I and that $\mathrm{var}(\hat{I}_n) \leq \mathrm{var}(I_n)$.
Hint: Use part 1. to show $\frac{1}{2}\mathrm{var}(k(U_1) + k(1-U_1)) \leq \mathrm{var}(k(U_1))$.