Math 697: Homework 1

Exercise 1

1. 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 $Pareto(a,b)$.

2. 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 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 3 (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 the following algorithm to generate a random permutation of the elements $1, 2, 3, \cdots, n$. We will 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.

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.

Explain, in words, what the algorithm is doing. Show that at iteration k, – i.e. when the value of $S(k)$ is initially set– $S(1), S(2), \cdots, S(k)$ is a random permutation of $1, 2, \cdots, k$.
Hint: Relate the probability P_k obtained at iteration k with the probability P_{k-1} obtained at iteration $k-1$.

Exercise 4 Compute the Legendre transform u^* of the logarithmic m.g.f u for the random variables $Poisson(\lambda)$ and $Exp(\lambda)$. Discuss in details where u^* is finite or not.

Exercise 5 We have seen in class that if X_i are independent $B(1,p)$ Bernoulli random variable and $S_n = X_1 + \cdots + X_n$ then for $a > p$

$$P\left(\frac{S_n}{n} \geq a\right) \leq e^{-nu_p^*(a)}$$

where

$$u_p^*(z) = \begin{cases} z\log\left(\frac{z}{p}\right) + (1-z)\log\left(\frac{1-z}{1-p}\right) & \text{if } 0 \leq \text{z} \leq 1 \\ +\infty & \text{otherwise} \end{cases} .$$

and a similar bound for $a < p$. In order to make that bound more easy to use in practice

1. Show that for $a > 0$ and $0 < p < 1$ we have

$$u_p^*(z) - 2(z-p)^2 \geq 0 \,.$$

   Hint: Differentiate twice.

2. Show that for any $\epsilon > 0$

$$P\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) \leq 2e^{-2n\epsilon^2} \,.$$

Exercise 6 On a friday night you enter a fast food 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) Cramer's theorem.

Exercise 7 Consider the problem of estimating π that we have considered in class. Estimate the number of trials you should perform to ensure that with probability $1-\delta$ your result is at distance no more than ϵ from the true value of π. Do this in two ways: (1) Use the central limit theorem, (b) Use the estimates of Exercise 5 and Cramer's theorem.

Exercise 8 (Hit-or-missmethod)

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 $\text{var}(I_n)$.

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 use this method concretely to the estimation of the number π? Compute the corresponding variances.

Exercise 9

Suppose f is a function on the interval $[0,1]$ with $0 < f(x) < 1$. Here are two ways to estimate $I = \int_0^1 f(x)dx$.

1. Use the "hit-or-miss" from the previous problem with $A = [0,1] \times [0,1]$ and $B = \{(x,y) : 0 \le x \le 1\,,\, 0 \le y \le f(x)\}$.

2. Let $U_1, U_2, \cdots$ be i.i.d. uniform random variables on $[0,1]$ and use the estimator

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

   Show that that $\hat{I}_n$ has smaller variance than the estimator of (a).

Exercise 10 (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

$$\mathrm{cov}(k(X), h(X)) \ge 0\,.$$

   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 $(k(X) - k(Y))(h(X) - h(Y)) \ge 0$. 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 $\mathrm{var}(\hat{I}_n) \le \mathrm{var}(I_n)$.
   Hint: Use part 1. to show $\frac{1}{2}\mathrm{var}(k(U_1) + k(1 - U_1)) \le \mathrm{var}(k(U_1))$.

3. Let $k(x) = 4\sqrt{1 - x^2}$. Then $I = \pi$. Compute $\mathrm{var}(\hat{I}_n)$ and $\mathrm{var}(I_n)$.