Problem Set 4
Random Variables Generation & Probability Bounds

Date: Wednesday, June 26^{th}, 2019.

Problem 1:

Let X be a continuous RV with a strictly increasing CDF $F_X(\cdot)$. Show that the RV $U = F_X(X)$ has a uniform distribution $\mathcal{U}([0,1])$.

Problem 2:

Using the inverse transform method to to generate a RV from the following distributions:

a) Exponential distribution with PDF $f(x) = \lambda e^{-\lambda x},\ x \geq 0$

b) Arcsine distribution with PDF $f(x) = \frac{1}{\pi}\left(x(1-x)\right)^{-\frac{1}{2}},\ 0 < x < 1$

c) Weibull distribution with PDF $f(x) = \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta},\ x \geq 0$

Problem 3:

Let X be a RV with distribution F and assume we can sample from Y, a RV with distribution G. Provided that $\sup\limits_x \frac{f(x)}{g(x)} \leq c$, we can sample from F using the Acceptance-Rejection Algorithm:

1. Generate a RV Y distributed as G.
2. Generate $U \sim \mathcal{U}([0,1])$ (independent from Y)
3. If $U \leq \frac{f(Y)}{cg(Y)}$, then set $X = Y$ ("accept") ; otherwise go back to step 1 ("reject").

a) The number of times N that steps 1 and 2 need to be called (i.e., the number of iterations needed to successfully generate X) is RV. Give its distribution and express its parameters as a function of c. Deduce the average the number of iterations required until an X is successfully generated.

b) Show that the accept/rejection algorithm produces a variable Y distributed according to F, i.e.

$$\mathbb{P}\left(Y \leq y | U \leq \frac{f(Y)}{cg(Y)}\right) = F(y).$$

Problem 4:

The PDF for the Cauchy distribution centered at $x_0 \in \mathbb{R}$ and with scale parameter $\gamma > 0$ is given by

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left(1 + \left(\frac{x - x_0}{\gamma}\right)^2\right)}$$

(a) Show by integration of the PDF that the CDF of the Cauchy distribution with $x_0 = 0$ and $\gamma = 1$ is given by $F(x; 0, 1) = tan^{-1}(x)/\pi + 1/2$.

(b) Describe and implement an algorithm for sampling $X \sim F(\cdot; 0, 1)$ in MATLAB.

(c) It is possible to extend the preceding method to sample from $X \sim F(\cdot; x_0, \gamma)$ for any $x_0 \in \mathbb{R}$ and $\gamma > 0$. How?

Problem 5:

Let X be a positive random variable with $E[X] = 10$. Using Jensen's inequality, what can you say about

the following quantities?
(a) $E\left[\frac{1}{X+1}\right]$
(b) $E\left[e^{\frac{1}{X+1}}\right]$
(c) $E\left[\ln\sqrt{X}\right]$

Problem 6:

Let X be an Binomial random variable with success probability in each trial $p = 1/2$ and number of trials n. Let $\alpha = \frac{3}{4}$. Please verify the following inequalities,
(a) $P(X \geq \alpha n) \leq \frac{2}{3}$ (Markov Inequality).
(b) $P(X \geq \alpha n) \leq \frac{4}{n}$ (Chebyshev Inequality)
(c) $P(X \geq \alpha n) \leq \left(\frac{16}{27}\right)^{\frac{n}{4}}$ (Chernoff Inequality)
Comments: You will find the bound given by Markov is the "weakest". The bound given by Chebyshev is "stronger" than the one given by Markov, when n is big. The "strongest" bound is the chernoff bound, which goes to zero exponentially fast as n increases.

Problem 7:

We consider the RV X with PDF

$$f(x) = Cxe^{-\frac{2}{3}x^2}, x \geq 0$$

(a) Determine the constant C so that $f(\cdot)$ is a valid PDF.
(b) Write a MATLAB function AR(n) that allows to simulate realizations of X by acceptance-rejection and returns a n sample $\{X_i\}_{i=1}^n$ and the acceptance rate. (We can for example upper bound $xe^{-x^2/3}$ on $\mathbb{R}_+$ and use an appropriate Gaussian distribution and restricted to $\mathbb{R}_+$ or use an exponential distribution.)
(c) Determine the distribution function F of X and write a second function IT(n) which makes it possible to simulate realizations of $X \sim F$ using the inverse transform method.
(d) Plot the histograms of the realizations obtained by AR(n) and IT(n) and compare them theoretical density f. Compare the two methods in terms of complexity and computational cost.

Problem 8:

Let X be an exponential random variable with parameter $\lambda = 1$. Compute the Markov inequality, the Chebyshev inequality, and the Chernoff bound to obtain bounds on $P(X \geq a)$ as a function of a. Also compute $P(X \geq a)$.
(a) For what values of a is the Markov inequality smaller than the Chebyshev inequality?
(b) Using MATLAB, plot the Markov bound, the Chebyshev bound, the Chernoff bound, and $P(X \geq a)$ for $2 \leq a \leq 6$ on the same graph. For what range of a is the Markov bound the smallest? the Chebyshev? Now use MATLAB command 'semilogy' to draw the same four curves for $6 \leq a \leq 20$. Which bound is the smallest?

Problem 9:

Your friend tells you that he had four job interviews last week. He says that based on how the interviews went, he thinks he has a 20% chance of receiving an offer from each of the companies he interviewed with. Nevertheless, since he interviewed with four companies, he is 90% sure that he will receive at least one offer. Is he right?

Problem 10:
Let X and Y be two random variables with $E[X] = 1$, $Var(X) = 4$, and $E[Y] = 2$, $Var(Y) = 1$. Find the maximum possible value for $E[XY]$.

Problem 11:
A system consists of 4 components in series, so the system works properly if all of the components are functional. In other words, the system fails if and only if at least one of its component fails. Suppose that we know that the probability that the component i fails is less than or equal to $p_f = \frac{1}{100}$, for $i = 1, 2, 3, 4$. Find an upper bound on the probability that the system fails.

Problem 12:
Let $X \sim Geometric(p)$. Using Markov's inequality find an upper bound for $P(X \geq a)$, for a positive integer a. Compare the upper bound with the real value of $P(X \geq a)$.

Problem 13:
The number of customers visiting a store during a day is a random variable with mean $E[X] = 100$ and variance $Var(X) = 225$. Using the Chebyshev's inequality, find an upper bound for having more than 120 or less than 80 customers in a day. That is, find an upper bound on

$$P(X \leq 80 \text{ or } X \geq 120).$$

Problem 14:
Use the inverse transformation method to generate a random variable having distribution function

$$F(x) = \frac{x^2 + x}{2}, 0 \leq x \leq 1.$$

Problem 15:
Let X have a standard Cauchy distribution.

$$F_X(x) = \frac{1}{\pi} \arctan(x) + \frac{1}{2}.$$

Assuming you have $U \sim Uniform(0, 1)$, explain how to generate X. Then, use this result to produce 1000 samples of X and compute the sample mean. Repeat the experiment 100 times. What do you observe and why?

Problem 16:
Use the rejection method to generate a random variable having PDF $f(x) = 20x(1 - x)^3,\ 0 < x < 1$.
Hint: Assume $g(x) = 1$ for $0 < x < 1$.

Problem 17:
If we desire an $X \sim N(\mu, \sigma^2)$, then we can express it as $X = \sigma Z + \mu$, where Z denotes a RV with the $N(0, 1)$ distribution. Thus, it suffices to find an algorithm for generating $Z \sim N(0, 1)$.
Moreover, if we can generate from the absolute value, $|Z|$, then by symmetry we can obtain our Z by independently generating a RV S (for sign) that is ± 1 with probability $1/2$ and setting $Z = S|Z|$. In

other words, we generate a uniform RV U, and set $Z = |Z|$ if $U \leq 0.5$ and set $Z = |Z|$ if $U > 0.5$. Therefore, we only need to sample from $|Z|$. Note that the RV $|Z|$ is non-negative and has density

$$f(x) = \frac{2}{\sqrt{2\pi}} e^{\frac{x^2}{2}}, x \geq 0.$$

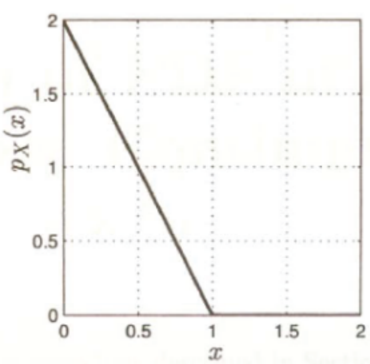
Use the rejection method to generate a standard normal random variable.
Hint: Assume $g(x)$ is the pdf of the exponential distribution with $\lambda = 1$.

Problem 18:

Let X be the number of successes in n Bernoulli trials where the probability of success is p. Let $Y = \frac{X}{n}$ be the average number of successes per trial. Apply the Chebyshev inequality to the event $|Y - p| > a$. What happens as $n \to \infty$.

Problem 19:

Find a transformation so that $X = g(U)$, where $U \sim Uniform(0, 1)$ has the PDF shown in the below figure.



Problem 20:

A RV Y has the PDF

$$f_Y(y) = \frac{4a^4 y}{(a^2 + y^2)^3}, \; y \geq 0,$$

where a is a real positive constant. Find the transformation $g(\cdot)$ needed to generate $Y = g(X)$ from a RV X that is uniformity distributed on $(0, 1)$.

Problem 21:

Suppose that we roll a standard fair die 100 times. Let X be the sum of the numbers that appear over the 100 rolls. Use Chebyshev's inequality to bound $P(|X - 350| \geq 50)$.

Problem 22:

A random variable X is always strictly larger than -100. You know that $E[X] = 60$. Give the best upper bound you can on $P(X \geq 20)$.

Problem 23:

Let K be a Poisson random variable with expected value α. Use the Chernoff bound to find an upper bound on $P(K \geq c)$.

Problem 24:

An isolated edge in a network is an edge that connects two nodes in the network such that neither of the two nodes is connected to any other nodes in the network. Let C_n be the event that a graph randomly generated according to $G(n, p)$ model has at least one isolated edge. $G(n, p)$ is the model that there are n nodes and each pair of nodes are connected by one edge with probability p.

(a) Show that

$$P(C_n) \leq \binom{n}{2} p(1-p)^{2(n-2)}$$

(b) Show that, for any constant $b > \frac{1}{2}$ if $p = p_n = b\frac{\ln(n)}{n}$ then

$$\lim_{n\to\infty} P(C_n) = 0.$$

Hint: It is convenient to use the following inequality: $1 - x \leq e^{-x}$ for all $x \in \mathbb{R}$.

Problem 25:

Use the rejection method to generate a random variable having the PDF $f(x) = \frac{4}{3\sqrt{\pi}} x^{\frac{3}{2}} e^{x}, \ x > 0$.

Hint: Assume $g(x)$ is the pdf of an exponential distribution with parameter $\lambda = \frac{2}{5}$.