# CSCE 619-600: Computer Networks
# Homework 1 due January 31, 2017

## 1 Purpose

The goal of this homework is to understand random variable generation, verify examples shown in class, and learn Latex.

## 2 Description

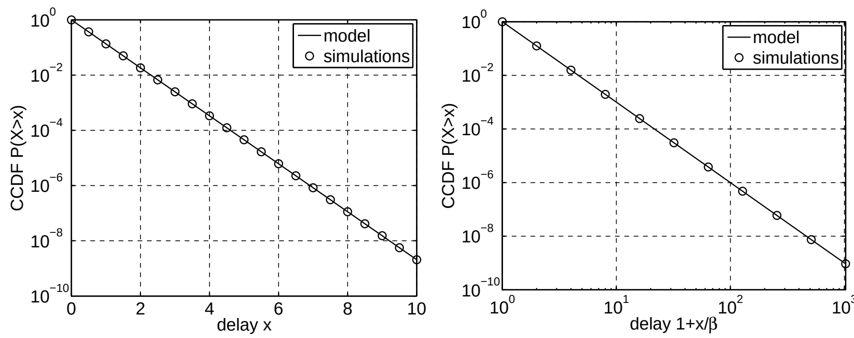
### 2.1 Random Variable Generation

In this section, you will prove and validate a simple method of generating arbitrary random variables using Mersenne Twister (or other suitable random-number generators). Assume the objective is to obtain samples of a random variable $X$ with some CDF $F(x)$, but you have access only to a number generator that produces realizations of a random variable $U$ uniformly distributed in $(0, 1)$. If your language (e.g., C++11, Matlab) offers shortcuts to this process, refrain from using them.

1. Prove that random variable $Y = F^{-1}(U)$, where $F^{-1}$ is the inverse of $F(x)$, has the same distribution as $X$. Show detailed derivations and motivate each of the steps.

2. After the proof, use simulations to generate exponential and Pareto variables with respective distributions $F_1(x) = 1 - e^{-2x}$ and $F_2(x) = 1 - (1 + x)^{-3}$. For verification, plot the empirical (i.e., measured in simulations) tail of the generated distribution and that of the corresponding model on the same graph. Make sure to plot the model as a continuous curve and simulations as isolated markers (see Figure 1). Discuss how likely each distribution is to reach extreme values (e.g., larger than 10 hours) and explain what "heavy-tailed" means in this context.

3. Note that the Pareto case is best plotted on a *log-log scale*, i.e., the $x$-axis containing $\log(1 + x/\beta)$ and the $y$-axis being $\log P(X > x)$. For the exponential case, only the $y$ scale should be logarithmic. Prove that both model tails are straight lines in these graphs and confirm whether your simulations satisfy this property.




(a) exponential (b) Pareto

Figure 1: Illustration of proper representation of CCDFs. Note the usage of subfigures, descriptive labeling of all axes and legend, and readable font size.

4. Finally, examine the mean of each empirical distribution and compare it to the theoretical one.

## 2.2 Bus Example (Random Arrival)

Simulate the model of bus arrival explained in class and verify some of the analytical results shown in the first lecture. Assume that the inter-bus delays are given by independent variables $X_1, X_2, \ldots$ with CDF $F(x)$ and that you arrive to the bus stop at a random time $t$ during the day. The objective is to obtain the distribution of *wait time* to the next bus and find out its shape through curve fitting.

To accomplish this, assume in your simulation that buses start arriving to the bus stop at time 0, while your arrival time $t$ is uniformly random in $[100, 300]$ hours. Each iteration $i$ of your simulation should generate a bus schedule in the interval $[0, t]$ and compute one sample of wait time $W_i(t)$. Repeating this $N = 100{,}000$ times, you can build an empirical distribution $G(x)$ of wait times $\{W_1(t), \ldots, W_N(t)\}$.

1. Examine two separate cases of inter-bus delay distributions: exponential $F_1(x) = 1 - e^{-\lambda x}$ and Pareto $F_2(x) = 1 - (1 + x/\beta)^{-3}$. Select parameters $\lambda$ and $\beta$ such that the average inter-bus delay is $E[X_i] = 1/3$ of an hour (i.e., 20 minutes). Show full derivations for $\lambda$ and $\beta$. Note: your units of time $t$ are *hours*, not minutes or seconds.

2. Plot the *tails* of $G_1(x)$ and $G_2(x)$.

3. Using curve fitting, demonstrate that $G_1(x) = 1 - e^{-\lambda x}$ and $G_2(x) = 1 - (1 + x/\beta)^{-2}$.

4. Observe and discuss the mean wait time under both distributions.

### 2.3 Bus Example (Conditional Wait)

Create a new simulation to compute the mean wait time $V(t)$ under the condition that your arrival is exactly $t = 1/12$ hour (i.e., 5 minutes) after the last bus has departed for both exponential and Pareto $F(x)$.

1. Generate a stream of random variables $X_i$ from distribution $F(x)$ and compute each wait delay as

$$V_i(t) = \begin{cases} X_i - t & X_i \geq t \\ \text{undefined} & \text{otherwise} \end{cases}. \qquad (1)$$

   Here, "undefined" means that you simply skip over that inter-bus delay. After generating 100,000 random variables $X_i$, you will have enough samples $V_i(t)$ to build a decently accurate distribution $H(x) = P(V_i(t) < x)$. Plot its tails using the parameters of the first problem (i.e., exponential with $\lambda = 2$ and Pareto with $\alpha = 3, \beta = 1$) and make observation to their shape.

2. Derive the tail distribution $P(V_i(t) > x)$ of wait times. Compare to simulations in part 1 above. Hint: see class slides that proved the memoryless property of exponential distributions.

3. Using the expectation of a non-negative random variable shown in class, derive a formula for the average wait time $E[V_i(t)]$ experienced by the passengers for a general $F(x)$ and any $t$. Simplify it for exponential and Pareto cases with generic parameters $\lambda$ and $(\alpha, \beta)$. Now recompute the examples shown on the first day of class.

4. Vary the mean inter-bus delay $E[X_i]$ (i.e., by changing $\lambda$ and $\beta$) and plot both the model and simulations of $E[V_i(t)] = E[X_i - t | X_i > t]$ to show how well they match over a range of values $E[X_i]$. To avoid confusion, make sure that $t$ is measured in the same time-units as $E[X_i]$, i.e., $t = 1/12$ represents 5 minutes.