ECE/CS 541: Computer System Analysis

Homework #4 (**Due 12/07/2018**)

Student Name: Student's name *Netid: netid*

The intent of this homework is to get you some hands on experience with generating random variates and performing statistical tests.

You are to submit this homework electronically on `Compass2g` by the end of day on December 7, 2018 at 11:59 pm.

Problem 1 (25 points): Generating Exponentials

We are interested in a sampling process that generates a pair of random variable (X, Y), such that for any pair of samples (x, y), we always have that $x > y$. Furthermore, we require that $X \sim exp(\lambda)$ and $Y \sim exp(2\lambda)$ (obviously, X and Y are not independent).

a) Describe a simple way to generate the pairs (X, Y).

b) Now suppose instead of requiring that $Y \sim exp(2\lambda)$, we require that $Y \sim exp(\frac{\lambda}{2})$. Show that in this case, it is **impossible** to generate (X, Y) such that for all samples $x > y$.

Problem 2 (25 points): Combining Distributions

Suppose that it is relatively easy to simulate from a set of given distributions F_i for $i = \{1, 2, 3, \dots, n\}$.

a) Assuming that n is small, describe a way to simulate from

$$F(x) = \sum_{i=1}^{n} P_i F_i(x), \text{ where } \qquad P_i \geq 0 \ \forall i, \qquad \sum_{i=1}^{n} P_i = 1$$

b) Using the above results, give a way to simulate from $F(x)$ where

$$F(x) = \begin{cases} \dfrac{1 - e^{-2x} + 2x}{3}, & 0 < x < 1 \\ \dfrac{3 - e^{-2x}}{3}, & 1 < x < \infty \end{cases}$$

Problem 3 (20 points): Generating a Triangular Distribution

Consider a random variable X with its probability density function given by

$$f_X(x) = \begin{cases} \dfrac{2}{z}t, & \text{for } x \leq z, \\ 2\left(1 - \dfrac{x - z}{1 - z}\right), & \text{for } z < x \leq 1 \end{cases}$$

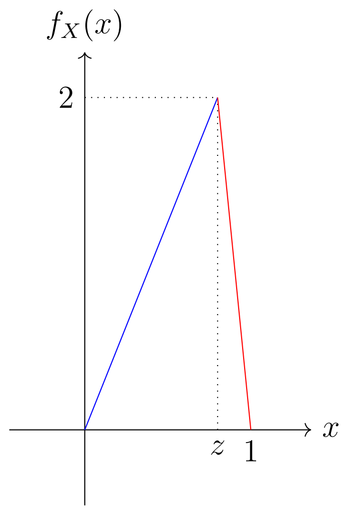

Figure 1: The pdf of X for Problem 3.

for some $0 < z < 1$. We can see $f_X(x)$ for $z = 0.8$ in Figure 1.

Give an algorithm for generating this distribution using the inverse transform method.

Problem 4 (20 points): Validating the Claims of your Fellow Researchers

In a recent paper at SIGCOMM'2017, researchers used data from Facebook's data-centers to the study the behavior of micro-bursts (extremely short-term bursts in traffic) [1]. In that paper, the authors show using the KS test that the inter-arrival times between micro-bursts are indeed *not* exponential, and thus cannot be modeled as a Poisson arrival process. In this exercise, we set out to validate that claim.

Use the provided data-set (`https://github.com/zhangqiaorjc/imc2017-data`), specifically the three files `fig4_interburst_dur_*.csv` to implement the KS test and validate the authors claim. You can use any programming language to read and manipulate the data, however, you are **not** allowed to use any KS test packages. Yes we know that packages exist and are easy to use. The goal of this exercise is to gain a better understanding of what happens behind the scenes when running the KS-test and avoid the pitfalls of p-value hacking (c.f. Figure 2). Show your work and specify your chosen confidence level.

Problem 5 (10 points): LCGs

Consider the LCG pseudo-random number generator given by

$$X_{n+1} = (aX_n + c)\%m$$

where $a = 5$, $c = 1$, $m = 16$, and $X_0 = 1$.

a) Does this RNG have full period? Why or why not?

b) If you answered yes to the above question, would you recommend the use of this RNG in practice? Why or why not? *Hint.* Generate a couple of numbers and look at their binary representation.

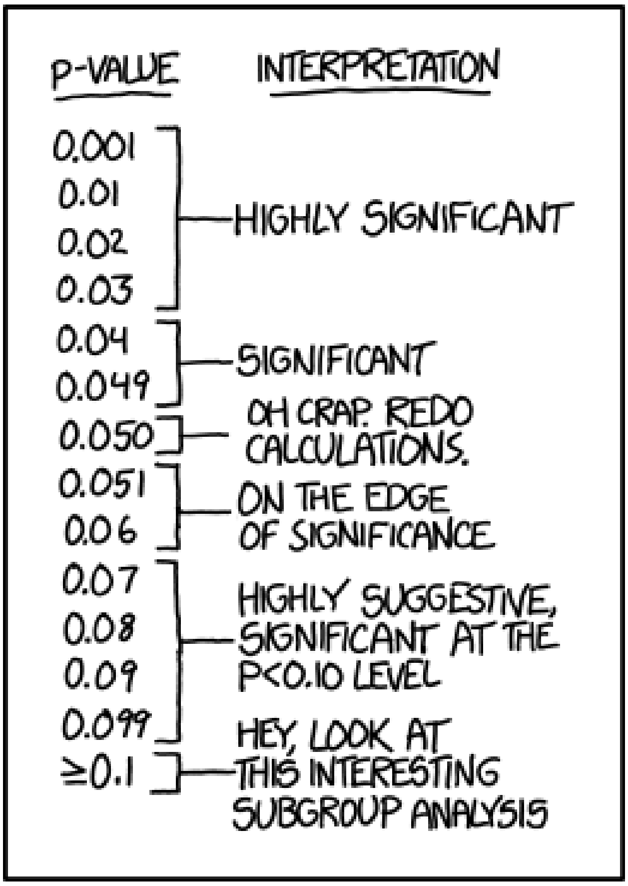

Figure 2: p-value hacking from https://xkcd.com/1478/

References

[1] Qiao Zhang, Vincent Liu, Hongyi Zeng, and Arvind Krishnamurthy. High-resolution measurement of data center microbursts. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, pages 78–85, New York, NY, USA, 2017. ACM.