# Pre-class work

Do the following exercises from Shonkwiler & Mendivil, Chapter 2.

- **Exercise 14, page 96: Sampling bias for bus waiting times**

  14. (4) (Sampling bias for bus waiting times) Suppose the interarrival time for a city bus has an exponential distribution with parameter $1/\lambda$. A passenger arrives at a uniformly random time and records the time until the next bus arrives. What is the expected waiting time? Use a simulation to get an answer. Is the answer surprising? Now suppose instead that the interarrival time is $U(0, 2\lambda)$. How does this change the situation? (Notice that the expected interarrival time is $\lambda$ in both cases.)

  Note that there are 2 common, equivalent parameterizations of the exponential distribution.

  - $\text{Exponential}(x \mid \lambda) = \lambda\, e^{-\lambda x}$. This is used in Shonkwiler & Mendivil.
  - $\text{Exponential}(x \mid \beta) = \beta^{-1}\, e^{-x/\beta}$. This is used in Scipy. So if you import scipy and generate exponentially distributed random values using `scipy.random.exponential(beta)`, you should use $\beta = \lambda^{-1}$.

  This is an example of a difficult to compute value (the expected waiting time under two different distributions) with a counterintuitive result that be can simulated fairly easily.

- **Exercise 24, page 98: Retirement benefit projection**

  24. (5) (Retirement benefit projection) At age 50 Fannie Mae has \$150,000 invested and will be investing another \$10,000 per year until age 70. Each year the investment grows according to an interest rate that is normally distributed with mean 8% and standard deviation 9%. At age 70, Fannie Mae then retires and withdraws \$65,000 per year until death. Below is given a conditional death probability table. Thus if Fannie Mae lives until age 70, then the probability of dying before age 71 is 0.04979. Simulate this process 1000 times and histogram the amount of money Fannie Mae has at death.

Mortality table, probability of dying during the year by age*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 50 | 0.00832 | 64 | 0.02904 | 78 | 0.09306 | 92 | 0.26593 |
| 51 | 0.00911 | 65 | 0.03175 | 79 | 0.10119 | 93 | 0.28930 |
| 52 | 0.00996 | 66 | 0.03474 | 80 | 0.10998 | 94 | 0.31666 |
| 53 | 0.01089 | 67 | 0.03804 | 81 | 0.11935 | 95 | 0.35124 |
| 54 | 0.01190 | 68 | 0.04168 | 82 | 0.12917 | 96 | 0.40056 |
| 55 | 0.01300 | 69 | 0.04561 | 83 | 0.13938 | 97 | 0.48842 |
| 56 | 0.01421 | 70 | 0.04979 | 84 | 0.15001 | 98 | 0.66815 |
| 57 | 0.01554 | 71 | 0.05415 | 85 | 0.16114 | 99 | 0.72000 |
| 58 | 0.01700 | 72 | 0.05865 | 86 | 0.17282 | 100 | 0.76000 |
| 59 | 0.01859 | 73 | 0.06326 | 87 | 0.18513 | 101 | 0.80000 |
| 60 | 0.02034 | 74 | 0.06812 | 88 | 0.19825 | 102 | 0.85000 |
| 61 | 0.02224 | 75 | 0.07337 | 89 | 0.21246 | 103 | 0.90000 |
| 62 | 0.02431 | 76 | 0.07918 | 90 | 0.22814 | 104 | 0.96000 |
| 63 | 0.02657 | 77 | 0.08570 | 91 | 0.24577 | 105 | 1.0000 |

* Source: Society of Actuaries, Life Contingencies.

You can get the data for this problem here so you don't have to retype the whole table.

# Study guide

The focus of this session is the converse of the previous session.

- In the previous session, we used simple, uniformly distributed random numbers to power our simulations. In this session, we use simulations to generate more complex (and useful) random numbers.
- Some of these simulations lead to well-known probability distributions, while others lead to distributions that are difficult to characterize analytically.

- The words “simulation” and “sampling” are often conflated in the literature since generating a sample from a probability distribution is often effectively a small simulation.

## Shonkwiler & Mendivil, Section 2.1

This section covers some cases you have seen before in this course.

- There are non-simulation techniques for drawing samples from some probability distributions. CDF inversion is the most straightforward, but you need to be able to compute the inverse of the CDF efficiently.

- The Bernoulli trial (a biased coin toss) uses the check

  ```
  if uniform(0,1) < p
  ```

  to determine whether the trial results in a success or a failure.

- Choosing one of multiple discrete outcomes (the spinner/roulette wheel from Sayama) uses the cumulative sum of the probabilities of the possible outcomes and a uniformly distributed random number. This is a generalization of the Bernoulli trial
  .

  ```
  probabilities = [0.1, 0.2, 0.3, 0.4]  # four possible outcomes
  cumulative_prob = scipy.cumsum(probabilities)
  sample = cumulative_prob.searchsorted(scipy.random.uniform(0,
  1))
  ```

## Shonkwiler & Mendivil, Section 2.3

- A sample from the binomial distribution can be seen as a simulation of a number of Bernoulli trials – do $n$ flips of a coin with bias $p$ and count the number of heads.

## Shonkwiler & Mendivil, Sections 2.4 and 2.5

- The Poisson and exponential distributions are related.
  - The exponential distribution models waiting times – the duration between two events, for example between one bus arriving and the next bus arriving at a bus stop.
  - The Poisson distribution models rates – the number of events per unit of time, for example, the number of buses arriving at a bus stop per hour.

- Section 2.5.1 shows how to generate samples from the exponential distribution using CDF inversion. The CDF of the exponential distribution is

  $$F(x) = 1 - e^{-\lambda x}$$

  where $\lambda$ is a parameter of the exponential distribution over waiting time $x$. The inverse of the CDF is simply

  $$x(F) = -\lambda^{-1} \log(1 - F)$$

  So if we draw a uniformly distributed random number, $F$, and plug it into the inverse CDF above, $x$ will be exponentially distributed.

- A Poisson sample can be simulated by generating samples from the exponential distribution (waiting times between events) and counting how many events occur per unit interval.

- Section 2.5.2: As you've already seen in this course, there is a whole field called queueing theory. We've seen that queueing problems can be difficult to analyze mathematically in complex situations and simulation is often used to derive results instead. Queueing theory has applications in telecommunications networks (routing and queueing of packets in IP networks), distributed software systems, scheduling, traffic modeling, etc.

## Connection to *CS146: Modern Computational Statistics*

- The simulations you have seen so far in this course are essentially generating samples from some really complex probability distributions.

- There is a deep connection with probabilistic inference here, which is why Monte Carlo methods are covered in some detail in CS146.

- The key idea is that inference is the process of estimating model parameters from data, while simulations generate data when given specific parameter settings.

- This idea is formalized in Bayes' equation

  P(model parameters | simulation results)
  = (1/Z) P(simulation results | model parameters) P(model parameters)

  where

  - P(simulation results | model parameters) is the output from a simulation,

    - P(model parameters | simulation results) is the inference result (known as the posterior distribution),
    - P(model parameters) is the prior distribution,
    - Z is constant with respect to the simulation parameters.

# Frequently asked questions

## Pre-class work, Exercise 14: How to implement the "uniformly distributed passenger arrival times"?

What's difficult about this, is that we don't have an obvious upper bound for the uniform distribution. In other words, we need to pick the reasonable values for $a$ and $b$ in Uniform$(a, b)$.

The trick is to pick the interval $[a, b]$ wide enough so that several buses arrive in that window, and at the same time make sure that the passenger does not miss that last bus (there is always another bus for the passenger to catch). You could use the following settings for the simulation:

- $\lambda = 1$ minute
- $a = 0,\ b = 10$ meaning the passenger arrives before minute 10.
- Generate 100 bus arrival times. If there is 1 bus per minute, on average, then by the time there have been 100 buses, we will be well past minute 10.

## What is the difference between using $\lambda$ and $\beta = 1/\lambda$ in the exponential distribution?

When modeling events and the times at which events occur, $\beta$ is interpreted as the average time between two consecutive events, and $\lambda$ is interpreted as the average rate at which the events happen (number of events per time unit).

Both the exponential distribution and the Poisson distribution can be characterized by either of these two parameters, with the same interpretation of the parameters. For instance, in the distributions Poisson$(k \mid \lambda)$ and Exponential$(x \mid \lambda)$, $\lambda$ means the same thing (average rate). The difference is in the meaning of the random variable or the samples from the distributions. The samples from the Poisson distribution represent the number of events per unit of time (for example, the number of buses per minute) while the samples from the exponential distribution represent the time between two consecutive events (for example, the number of minutes between two buses).