Simulation for Kool Kats

David Goldsman Paul Goldsman

January 25, 2019

Contents

Chapter 1

A Whirlwind Tour of Computer Simulation

1.1 Organization of the Text

Computer simulation is one of the most-widely used tools in operations research, industrial engineering, management science, and, in fact, in many other engineering, mathematical, and science disciplines. In everyday life, you run across systems in which simulation has played a role:

- How many servers are available at the post office when you show up?
- Should we keep a certain number of spare parts in stock in case an important component fails?
- How are issues in the supply chain likely to affect the delivery time of a product?
- What is the best retirement portfolio for a particular person?

The purpose of this text is to present a self-contained course on computer simulation that will help you model complicated systems, write computer programs to allow you to simulate those systems, analyze the outputs of those programs, and then maybe even allow you to improve or even optimize your system.

The course is presented in 11 modules (chapters).

1. This *Whirlwind Tour*, where we give a brief overview of simulation, some motivational demos, and a taste of things to come.
2. *Calculus, Probability, and Statistics* boot camps, intended to make the text completely self-contained. Although you would think this is a programming course, there is also a great deal of theory that will be covered, so that's why it will be necessary to recall our old prerequisite friends from days of yore.

3. *Hand Simulations*, where we introduce various simulation concepts and problems at a low-enough level that will allow us to carry out the simulation computations by hand. This will give us an idea of the types of manipulations that will be required to deal with more-substantive models.

4. *General Simulation Principles*, where we discuss common concepts that underlie most simulation languages. We'll pay particular attention to how the simulation progresses through time by updating what is known as the *future events list*.

5. *Queueing Theory*. Here we step back and review what classical queueing methodology can and cannot do for you — and why we therefore need simulation.

6. *Uniform Number Generation* concerns what you might regard as a simple little problem — namely, how do you generate independent random real numbers between zero and one? It turns out that this is not as easy a you would think, but we show how to get pretty close to achieving that goal. But why spend an entire chapter on such a seemingly minor little problem? Because uniforms are the building blocks for general...

7. ... *Random Variate Generation*! This is where we'll learn how to take those aforementioned uniforms and use them to generate just about anything else — exponentials, Weibulls, arbitrary distributions and even entire stochastic processes like Markov chains, nonhomogeneous Poisson processes, and Brownian motion.

8. *Input Analysis*, which concerns the random variables that are used to drive the simulation. For instance, how do we model the distribution that customer interarrival times follow? Service times? Breakdown times? How do certain decision processes work?

9. *Output Analysis*. Just because a simulation ends doesn't quite mean that you get "the answer." Just like for any other experiment, you have to conduct a formal statistical analysis of the output before you can make sound conclusions. The complicating issue that arises here is that simulation output (e.g., consecutive customers waiting times) is never, ever independent and normally distributed — which means that you can't use the standard analysis that you learned in your first statistics course. Whatever shall you do? That's a raison d'être of the chapter.

10. *Comparing Systems*, where we'll learn how to tell which of a number of alternatives or systems is the "best," where the term "best" can mean several things, for instance, the server configuration that produced the shortest expected customer waiting time, or the inventory policy that maximizes the probability that an order will be delivered on time.

11. *Arena* is a graphics-based simulation language that we will learn during the course. It allows for easy construction, execution, and analysis of a wide variety of simulation processes, most notably, models involving some sort of queueing network. There are numerous fine simulations available, and one could substitute many of these for Arena just as effectively.

Each chapter is itself divided into several sections that can be thought of as lessons that are sometimes self-contained but are usually built on the foundations that have been set so far. After most of the sections, we give you some self-assessment problems that you should be able to do pretty easily; and at the end of the chapter, we present various homework problems, many of which are not too tough, but some of which may need a bit more thought. We also challenge you with several sample exams along the way (with solutions thoughtfully provided) so that you can test your skills as you go. Additional materials will occasionally be posted on our website,

`www.isye.gatech.edu/~sman/courses/6644.`

1.2 Chapter Outline

The goal of the current Chapter 1 is to acquaint the reader with simulation via a short tour of certain aspects of the subject. In particular, we'll talk about the following high-level topics just to get the simulation journey started.

§1.3 Introduction to simulation, where we'll discuss general models, some relevant definitions, and reasons for using simulation.

§1.4 Some historical (hysterical?) facts and...

§1.5 ... typical application areas.

§1.6 A bunch of easy examples for motivational purposes.

§1.7 Generating randomness — we'll need a source of randomness to generate the events that take place during the simulation. How does one produce random variates on a computer?

§1.8 Analyzing randomness — how do we deal with simulations outputs, e.g., consecutive customer waiting times, that can not be handled by traditional statistical methods?

1.3 Introduction to Simulation

This course shows how to create and evaluate a variety of practical models using computer simulation methods. A *model* is a high-level representation of the operation of a real-world process or system. Our concern will be with models that are:

- *Discrete* (vs. continuous) — that is, the state of the model can on change at discrete points in time. The arrivals and departures of customers at a store occur every once in a while. The weather changes continuously.

- *Stochastic* (vs. deterministic) — the systems that we will look at contain randomness.

- *Dynamic* (vs. static). In dynamic models, customers move around and make decisions, servers come on and off duty, and things change. In a static model, the same stuff general happens repeatedly, e.g., playing blackjack over and over and over.

We have actually been working with models since grade school, at least fairly simple ones. The questions arises as to how can we “solve” a model?

- *Analytical* methods. Think of a problem that is solved exactly by solving an equation (or a series of equations).

  Examples: $f'(x) = x$ with the initial condition $f(0) = 0$ solves out to $f(x) = x^2/2$.

  Toss a stone off of a cliff. You can model its position via the usual physics equations, e.g., the vertical position at time t is the sort of famous and obvious $y(t) = y(0) + v(0)t - gt^2/2$.

- *Numerical* methods. If things get a little too complicated to solve by exact analytical methods, you might have to resort to numerical approximations.

  Examples: Try solving $e^x = 3x$ exactly. In fact, you can't do it, but you can use numerical methods (to be discussed later) to find an *approximate* answer, $x \doteq 0.619$.

  Model the weather. This is way too tough for exact analytical models — and may require thousands of differential equations — so you might use numerical methods here.

- *Simulation* methods. Add a little randomness, and you'll need to resort to a simulation model (plenty of examples coming up).

Definition: *Simulation* is the imitation of a real-world process or system over time.

Simulation typically involves the generation of an artificial history for the imitated process or system. This is used to draw inferences concerning the operating characteristics of the real system that is being represented by our simulation model. If our simulation is a good one, then the inferences and conclusions will also be valid for the real system.

Simulation is...

- One of the top industrial engineering / operations research / management sciences technologies.
- Used by academics and practitioners on myriad theoretical and applied problems.
- An indispensable problem-solving methodology.

We use simulation to

- Describe and analyze the behavior of a system.
- Ask “what if” questions about the system. What if customers start showing up more often? What if we add a server?

- Aid in the design and optimization of systems. What is the optimal number of servers so as to maximize profit (revenue – operating costs)?

Pretty much anything can be simulated, for instance,

- *Static* (fixed) systems with no "customer" arrivals. For example, simulate a blackjack strategy, or estimate stock option prices.

- *Dynamic* customer-based systems like Queueing Centers, Manufacturing Processes, Supply Chains, Health Systems, etc.

Remark: *Systems can be real or conceptual.*

Advantages of Simulation. So why would we ever want to simulate? Here are a bunch of reasons to use the methodology.

- Can study models that are too complicated for analytical or numerical treatment.

- Can study detailed relations that might be lost in the analytical or numerical treatment. For example, exactly what happens to a particular customer as he progresses through the system? Or when, precisely, did a queue size exceed its allowable buffer?

- Solve a specific problem. For example, what do you have to do to fix a bottleneck in a production line? If you fix it, does it simply move somewhere else?

- Investigate if a (proposed or existing) system will accomplish its goals.

- Suppose the current system won't accomplish its goals. Now you need to investigate alternatives.

- Can be used to check results and give credibility to conclusions obtained by other methods.

- It's a really nice demo method and can be used to give insight or sell an idea.

- Sometimes it's very easy to construct, run, and analyze a simulation model.

Disadvantages. But into this nice day some rain must fall. There are also some notable reasons to be a little careful when using simulation.

- Sometimes not so easy.

- Simulation can sometimes be very time-consuming and costly, especially if you're retentive about over-modeling and making ultra-pretty graphics.

- Simulations give "random" output (and results can be misinterpreted in many ways).

- For specific problems, better methods than simulation may exist.

But we obviously feel that the good easily outweighs the bad. On balance, simulation is a wonderful methodology that has proven to be extremely useful in a wide variety of applications.

---

§1.3 Assessment

1. We are interested in modeling the arrival and service process at the local McBurger Queen burger joint. Customers come in every once in a while, stand in line, eventually get served, and off they go. Generally speaking, what kind of model are we talking about here? (More than one answer below may be right.)
   (a) Discrete
   (b) Continuous
   (c) Stochastic
   (d) Deterministic
2. Which of the following can be regarded as advantages of simulation? (More than one answer below may be right.)
   (a) Simulation enables you to study models too complicated for analytical or numerical treatment.
   (b) Simulations can serve as very pretty demos that even University of Georgia graduates can understand.
   (c) Simulation can be used to study detailed relations that might be lost in an analytical or numerical treatment.
   (d) Simulations are often tedious and time-consuming to produce.

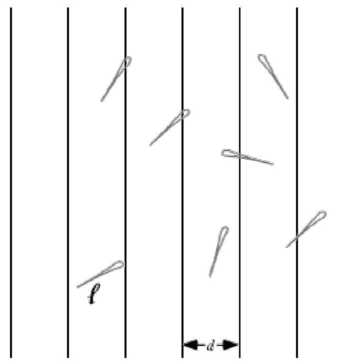
1.4 A History Lesson

In this section, we'll give a little of the history that underlies simulation. Some of the discussion is borrowed from Goldsman et al. (2009).

1.4.1 Buffon's Needle

Georges-Louis Leclerc, Comte de Buffon (1707–1788), was a celebrated French naturalist who anticipated many of the ideas of Darwin and Lamarck on evolution. He is remembered in the history of probability for his famous *needle problem* (Buffon, 1777) — perhaps the first example of a simulation experiment.

The problem is described as follows. If a floor has equally spaced parallel lines a distance d apart and if a needle of length ℓ is tossed at random on the floor where $\ell \leq d$, then what is the probability that the needle will intersect a line?



Buffon calculated $P\big(\text{Needle intersects a line}\big) = 2\ell/(\pi d)$. In particular, if $\ell = d/2$, then the probability is $1/\pi$. Buffon reasoned that by making a significantly large number of tosses, one could estimate π by the reciprocal of the proportion of needles that intersect a line. This is the earliest example of using independent replications of a simulation (i.e., the tosses) to approximate an important physical constant. We will revisit the estimation of π using a similar approach in §1.6.

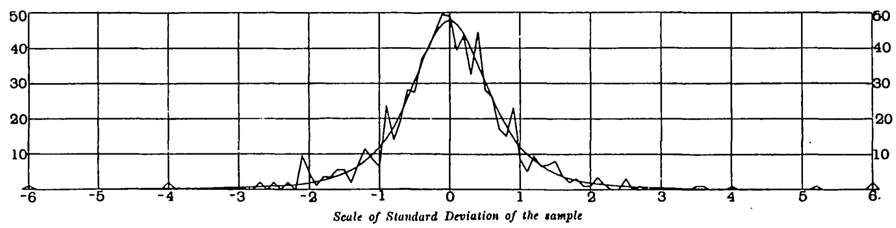
1.4.2 Beer and Student's t distribution

William Sealy Gosset (1876–1937), trained in mathematics and chemistry, became a brewer with Arthur Guinness, Son & Co. Ltd., in 1899. Gosset was faced with the problem of maintaining consistent quality of Guinness's ale and stout, and had to do so in the face of very limited data (small sample sizes). Unfortunately, the state-of-the-art in statistics at the time required large samples so that approximations based on the normal distribution could be used in the analysis. Gosset made his mark by developing appropriate small-sample statistical methodology using Student's t distribution instead of the normal.

To validate his results, Gosset conducted a precomputer simulation experiment by randomly sampling from a population of left middle finger lengths of 3,000 habitual British criminals obtained from New Scotland Yard.

- These measurements were written on 3,000 pieces of cardboard, thoroughly shuffled, and drawn at random to yield a randomly ordered list of the entire population.
- Each consecutive set of 4 measurements from this list was taken as a sample of size $n = 4$, so that there were 750 such samples.
- A standardized statistic from each of the 750 samples was calculated and used to form a histogram. This histogram is superimposed on what amounts to the probability density function of a t distribution with 3 degrees of freedom as shown below; and the achieved fit is quite good.

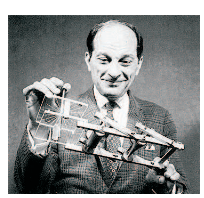
Diagram IV. Comparison of the theoretical frequency curve $y = \frac{1500}{\pi}\left(1+\frac{x^2}{s^2}\right)^{-2}$, with an actual sample of 750 cases.



Guinness allowed Gosset to publish his results, provided he used a pseudonym and no proprietary data was used. These results appeared under the pseudonym "Student" in 1908. This inaugural application of simulation to industrial process control is a remarkable example of the synergy of simulation-based experimentation and analytic techniques in the discovery of the exact solution of what is arguably a classical industrial-engineering problem.

1.4.3 Ulam, Monte Carlo, and the H-Bomb

Stanisław Ulam (1909–1984) was a Polish mathematician who worked on the Manhattan Project and originated the Teller-Ulam design of thermonuclear weapons. Ulam had a great fondness for card games; and in 1946 he was led to a "Monte Carlo" (static simulation) approach for computing the probability that a Canfield solitaire laid out with 52 cards will come out successfully.



In 1946 Ulam also performed detailed calculations showing that Edward Teller's initial design for the hydrogen bomb was inadequate. With the availability in 1946 of ENIAC, one of the first electronic computers, Ulam realized that the Monte Carlo method could be used effectively to estimate the intractable mathematical integrals arising in the design of a workable hydrogen bomb. He proposed an alternative design for the hydrogen bomb that he and Teller ultimately perfected and patented.

Ulam's colleague Nicholas Metropolis coined the term "Monte Carlo," explaining that (Metropolis, 1989)

> Stan had an uncle who would borrow money from relatives because he "just had to go to Monte Carlo." The name seems to have endured.

1.4.4 Industrial Applications and Beyond

Needles, beer, and H-bombs are all nice, but what about the development of real-life industrial applications of simulation? To this end, the 1950s and 1960s saw interest in simulation extend to manufacturing, material handling, and queueing models. Certainly, simulation is the technique of choice when you need to:

- Calculate the movement of parts and the interaction of system components.
- Evaluate the flow of parts through a system.
- Examine conflicting or competing demands for resources.
- Examine contemplated changes prior to their introduction.
- Eliminate (or avoid) major design blunders.

Typical questions that can be addressed by simulation experiments include:

- What will be the throughput? How can we change it?
- Where are the bottlenecks?
- Which is the best design?
- What is the reliability of the system?
- What is the impact of machine breakdowns?

Concurrently, a proliferation of simulation languages arose in the United States and Europe. Notable developments included:

- New "world views" to facilitate easy modeling (see Chapter 4).
- High-level programming commands, including the abilities to model customer arrival patterns, service resources, and customer routings to various locations.
- Automatic output reports with data such as average customer wait times, server utilizations, etc.

- Remark: For many years, the general-purpose simulation language SIMSCRIPT was among the most-popular such applications, particularly in military applications. Readers may recognize its main developer, Harry Markowitz, who — in another life — won the 1990 Nobel Prize in Economics for his work in the field of portfolio optimization.

In the meantime, simulation has come into its own as a serious discipline at the interface of mathematics, probability, statistics, and computer science. Great advances have been made in:

- Mathematical rigor.
- The development of efficient computational algorithms.
- Probabilistic and statistical methods.
- Modeling paradigms.
- Simulation language evolution, including modeling, analysis, and graphics capabilities. We'll talk about one of these languages — Arena — in Chapter 11.
- Use of simulation in a wide variety of applications. In fact, §1.5 gives a small taste of these applications in order to whet your appetite.

---

§1.4 Assessment

1. Who is William Gosset?
   (a) He invented the t distribution that is used ubiquitously in statistics.
   (b) He invented the s distribution that is used ubiquitously in statistics.
   (c) He invented tea.
   (d) He invented the word "ubiquitous".
   (e) He is the brother of Louis Gossett Jr., best known for his fine acting in many films, including *An Officer and a Gentleman*.
2. YES or NO? Has anyone closely related to the field of computer simulation ever won a Nobel Prize?

1.5 Simulation is Everywhere!

This section lists several applications where simulation has found practical use.

- Manufacturing
- Queueing Problems
- Financial Portfolio Analysis
- Traffic Simulation
- Aerospace Simulation
- Retail and Service Sectors
- Supply Chains
- Health Systems

Let's drill down into Health Systems applications of simulation with some additional examples:

- Patient Flow in a Hospital
- Hospital Room Allocation
- Optimization of Doctor / Nurse Scheduling
- Procurement of Supplies
- Disease Surveillance, Disease Propagation
- Humanitarian Logistics

At this point, we present some slightly more-substantive discussions on various real-life projects that your gentle authors have worked on over the years.

Simulating the Spread of Guinea Worm Disease

Guinea worm is a parasite with an interesting life cycle. Guinea worm larvae are ingested by copepods (very small crustaceans) and subsequently infect humans who are unfortunate enough to drink water containing those copepods. The larvae develop into large worms inside a human, eventually eating their way out from some random place in the poor guy's body (*Alien* style) so as to lay new eggs in a water source — thus beginning the cycle again. Guinea worms are gross and often cause infection and disfigurement of their human hosts. In fact, the worms used to be the scourge of parts of Africa and India, but have now almost been eradicated due to the heroic efforts of agencies such as The Carter Center, `www.cartercenter.org/health/guinea_worm/index.html`.

A number of strategies have been used to fight Guinea worms, e.g., education, provisioning of water filters, policing of potential infected water supplies, etc.; but each carries a cost. Simulation involving human behavior can be used to determine which of these strategies (subject to a budget constraint) best usher the worms to extinction.

We note that even though Guinea worms are disgusting, some environmental groups want to stave off their extinction. To this end, see

`www.angelfire.com/ms/guineaworm/`,

especially if you would like to courageously host some of the little fellas in your body.

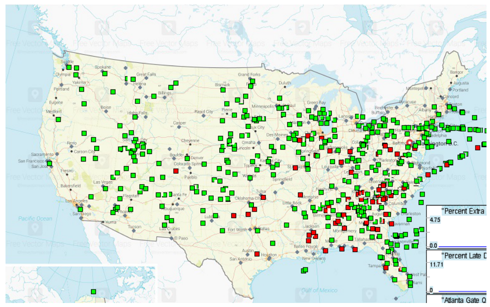
Simulating the Operation of Immunization Clinics

We studied the operations of a number of clinics in San Diego, CA that primarily serve the poor. One of the most-important needs that these clinics fill is that of providing childhood immunizations. Unfortunately, these clinics suffered from certain structural and operational inefficiencies that resulted in long patient waiting times, frustration, and some patients ultimately leaving the clinics before receiving their proper immunizations.

We conducted data analysis on patient arrival and service times; and then we used the results of that analysis as inputs to various clinic simulations. In fact, we simulated several alternative clinic configurations and provider (doctors, nurses, etc.) schedules in order to determine the most-efficacious strategy for each clinic. The exercise resulted in more children getting through the system more quickly, and with a commensurate increase in vaccination rates.

Simulating an Airline Routing Structure

A large airline with several hubs is interested in running a complicated flight schedule involving scores of cities, hundreds of aircraft, and thousands of pilots. A team of Operations Research analysts has proposed such a baseline schedule, but things never work out quite as planned — weather and mechanical-related delays can mess up the best intentions. What if there's a snowstorm in New York City, or a security issue in Chicago? How can these types of issues be mitigated so as to reduce passenger delays, get the flights back on schedule, and reduce lost revenue for the airline?



A potential strategy is to input the existing schedule to a simulation, which is then used to evaluate the schedule's robustness in the face of a problem. Management can then try out via simulation various alternative mitigation strategies; and the results from the respective runs can be used to determine which strategy fares the best.

Simulating a Call Center for Fast-Food Drive-Thru Customers

McBurger Queen is a chain of fast-food establishments serving delicious burgers and ice cream products. Customers can park and then dine inside the restaurant, or (since we live in today's hectic, fast-paced world) they may simply prefer to get their victuals via the drive-thru window. If you use the drive-thru, you typically drive up to the microphone, tell the attendant your order, and then drive to the next station pick up and pay for your taste-tempting order.

Of course, the person taking your drive-thru order over the microphone is not necessarily the same person who prepares and/or eventually hands over your food. So instead of allocating one person per store merely to sit around taking drive-thru orders, might it be possible to instead route all drive-thru orders to a central call center (perhaps even in a different city) serving the entire chain's network? The number of operators at the call center could vary throughout the day to correspond to customer demand patterns, and could be chosen so as to give the customers excellent service — e.g., 98% of all arrivals to the microphones would be answered within two seconds. We used simulation to determine how many servers would be needed to meet this excellent service requirement, even in the face of variable and poorly estimated customer arrival rates. We also determined how many stores would be needed to participate in this program before it could be declared to be cost-effective.

Simulating a Manufacturing Center

A popular car manufacturer produces 300,000 vehicles each year at its large plant in the Southeast. The facility has an enormous footprint, yet there are possible bottleneck points here and there that have been known to cause problems. For instance, moving goods between the plant's warehouse and main production line is often challenging due to various layout and staffing issues at both locations.

So what can be done? Management looked at a number of different potential solution strategies, including: installation of an enhanced transportation system between the warehouse and the line; an improved warehouse layout in which more-popular items were typically housed closer to the transportation system; and an improved production line layout so as to avoid arriving goods from the transportation system piling up in front of the line. Simulations were performed to determine which solution strategies best suited management's needs under various levels of traffic from the warehouse to the line.

---

§1.5 Assessment

1. Which of the following are areas where simulation has found substantial application? (More than one answer below may be correct.)

   (a) Inventory and Supply Chain Analysis

   (b) Financial Analysis

   (c) Manufacturing

   (d) Health Systems

   (e) Transportation Systems

2. Why might simulation be a good tool to analyze supply chains? (More than one answer below may be correct.)

   (a) Supply chains are always deterministic systems.

   (b) Supply chains often have complicated network structures, making exact analysis difficult.

   (c) Supply chains are stochastic systems, with random travel times, lead times, and order patterns.

   (d) Supply chain simulations can be programmed in a matter of minutes.

1.6 Some Easy Examples

This section will discuss various fun little simulation examples to get the creative juices flowing. You can even carry some of these out by yourself.

1. Happy Birthday — we'll talk about a famous result from your Baby Probability class that many people find to be surprising.

2. Let's Make Some π — here we discuss a simple procedure that can be used to estimate the value of π.

3. Fun With Calculus — why bother memorizing all of those awful calculus integration formulas when you can used simulation to do the job?

4. Evil Random Numbers – all stochastic simulations rely on the use of certain random numbers. For that reason, you have to trust that the package's random number generator works as advertised. What could possibly go wrong?

5. Queues 'R Us — everyone stands in line/queue sometime. We'll show how to simulate the simplest possible line.

6. Stock Market Follies — as any investor knows, the stock market is vexingly random. An easy spreadsheet example illustrates this annoying phenomenon.

1.6.1 Happy Birthday

How many people do you need in a room in order to have a 50% chance that at least two will have the same birthday? To keep things easy, let's assume that everyone has a 1/365 chance of being born on any day of the year — so we're not considering freaks that are born on February 29th (like you, Tony Robbins). In addition, we'll assume that everyone in the room is independent — so no twins allowed.

Let A denote the event that there will be a match if there are n people in the room. Then an elementary probability calculation reveals that

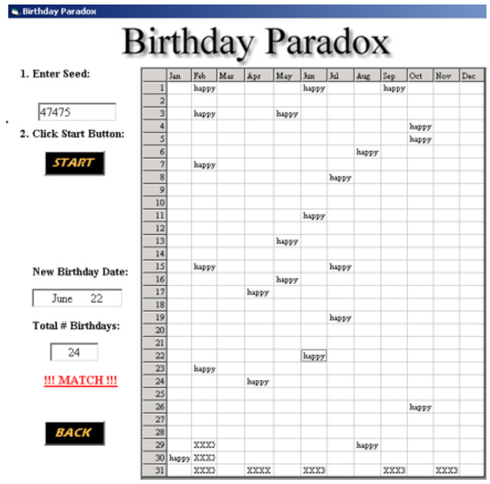
$$P(A) \equiv 1 - \left(\frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - n + 1}{365}\right).$$

Solving (numerically) for the smallest n such that $P(A) \geq 0.5$ gives $n = 23$, which is probably smaller than most people would initially guess.

We could also have arrived at this answer by carrying out a series of simulation experiments. The software application provided at

`www.isye.gatech.edu/~sman/courses/6644/waskos.exe.`

allows one to simulate the birthday problem. The user supplies a "seed" integer to initialize the program and simply clicks and clicks to sequentially generating a people / birthday s one-at-a-time until a match is obtained. At that point, the program stops and outputs the number of people that were needed. That number (say N) almost certainly will not be 23 — it could be much higher or much lower — because the simulation is generating random realizations of the experiments.



In fact, if you run the application many times, you can put together a histogram for N that will converge to the actual probability distribution of N. This provides a bit more information than just the *median* of N that we originally asked for.

You may also have noticed that if you run the experiment twice with the same seed value, then you'll get the *same* answer both times. This phenomenon is explained by the fact that the "random" numbers used in typical simulations aren't really random at all — they are numbers completely determined from an algorithm that merely *appear* to be random to humans. That's not necessarily a bad thing, but more on this later in §1.7.1 and Chapter 6.

1.6.2 Let's Make Some π

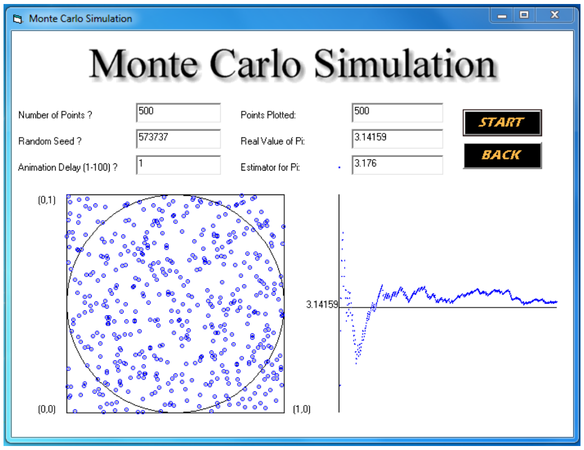
Monte Carlo simulation can be used to estimate π. Here's the basic idea:

- Assume that we begin with a dartboard in the shape of a unit square, with an inscribed circle. Then the area of the entire board is 1, and the area of the circle is $\pi/4$.

- Assuming that randomly tossed darts have an equal chance of landing anywhere on the dartboard, then the probability that a dart will land in the circle is $\pi/4$ (i.e., the ratio of the areas of the circle and square).

- Now throw lots of darts, say n. The Law of Large Numbers (explained in Chapter 2) guarantees that the proportion $\hat{p}_n$ that land in the circle will approach $\pi/4$ as $n \to \infty$.

- Multiply this proportion by 4 to get an estimate for π, i.e., $\hat{\pi}_n \equiv 4\hat{p}_n \to \pi$.



You can play around with the application available in `waskos.exe` to see for yourself how the estimate converges to π. The screenshot shows that after $n = 500$ darts, we achieve an estimate of $\hat{\pi}_{500} = 3.175$ (which, at least, is in the ballpark); and we see that the estimates are more-or-less getting closer to $\pi = 3.14159\ldots$ as n increases.

1.6.3 Fun With Calculus

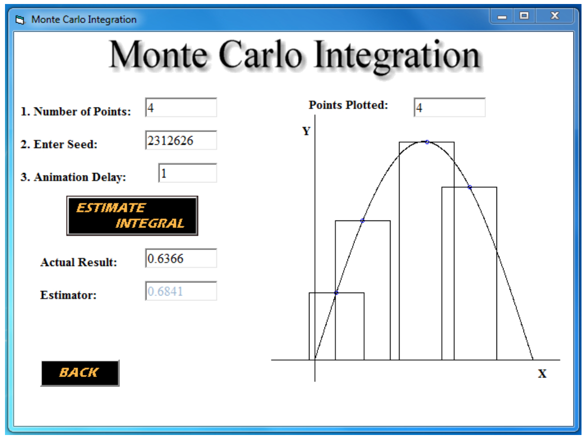
We can approximate definite integrals via simulation. We'll give more details in Chapter 3, but let's at least give the essentials during the present tour. To this end, we'll use simulation to evaluate $f(x) = \int_0^1 \sin(\pi x)\, dx$. You may happen to know that the actual answer is $2/\pi \doteq 0.6366$, but let's pretend for a moment that we don't know this fact.

The approximation that one can use in conjunction with simulation is simply

$$I \;\equiv\; \int_0^1 f(x)\, dx \;\doteq\; \frac{1}{n}\sum_{i=1}^{n} f(U_i),$$

where $U_1, U_2, \ldots, U_n$ are independent random numbers sampled on the interval [0,1]. This may bring back memories of the Riemann sums that you used when you were just learning about integrals. The only difference is our use of the uniform random numbers. Here's the path to glory in plain English:

- Sample n rectangles, such that each is centered randomly on $[0, 1]$, with width $1/n$ and height $f(x)$.
- Add up the areas.
- Make n really, really big.
- Thanks to the Law of Large Numbers, the sum approaches I as n gets large.



The illustration using the `waslos.exe` application is for the trivial case of $n = 4$, and it just so happens that the resulting approximate value of 0.6841 isn't too bad at all (actually, it got very lucky). In any case, the approximation tend to approach $2/\pi$ as n increases.

1.6.4 Evil Random Numbers

In the previous examples, we made use of random number generators that gave us a sequence of numbers $U_1, U_2, \ldots$ that were supposedly independent and uniformly distributed on the interval [0,1]. Now we'll see what happens when you use a bad random number generator. Here's some background.

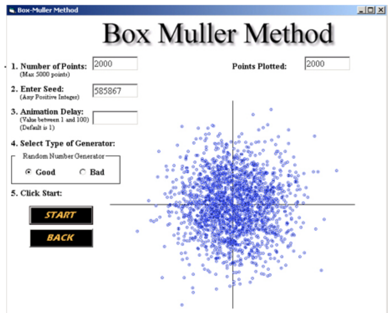
- Let's simulate peoples' heights versus their IQs, both of which we'll assume to be normally distributed and independent of each other.

- The results should look like a two-dimensional version of the famous "bell curve" (i.e., a normal distribution) with most observations in the middle and some on the outside.

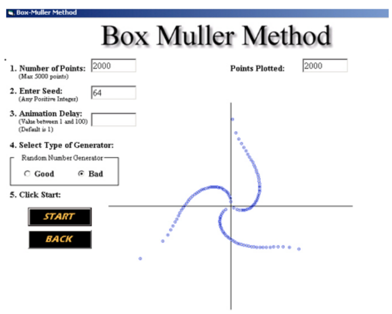
- If we use a "good" random number generator (to be explained in Chapter 6) as the basis to generate the normal observations, you'll see that the observations indeed look pretty normal.

- But a "bad" generator produces a picture that doesn't look normal at all!

- Moral: Better be careful and pay attention to the material in Chapter 6!



Two-dimensional normals constructed with a good random number generator.



Two-dimensional normals constructed with a bad random number generator.

1.6.5 Queues ‘R Us

Simulation can be a very useful tool for the analysis of queues. For instance, suppose there is a single-server queue at your favorite fast-food restaurant, in which:

- Customers show up,
- Wait in line if the server is busy,
- Get served in FIFO (first-in-first-out) order, and
- Leave the system all happy and full.

Question: What happens as the customer arrival rate approaches the service rate? Possible outcomes:

(a) Nothing much changes?

(b) Hamburgers start to taste better?

(c) Dollars begin spontaneously materializing out of the deep fryer?

(d) The line gets pretty long?

If you have ever taken a course in Stochastic Processes (or at least see Chapter 5 for a quick tutorial), then you know that certain simple queueing systems can be analyzed via numerical or exact methods.

Simulation can also be used to evaluate the performance of a queueing system. For instance, as in Chapters 3 and 5, we can simulate customer arrival and service times to estimate the expected customer waiting time, which we illustrate with a simple example. First, some standard notation.

- Customer i's arrival time is denoted by A_i.
- Customer i starts service at time $T_i = \max(A_i, D_{i-1})$ (the maximum of the customer's arrival time and the previous guy's depart time).
- Customer i's waiting time is $W_i^Q = T_i - A_i$.
- Customer i's service time is S_i.
- Customer i's departure time is $D_i = T_i + S_i$.

Example: Suppose we have the sequence of four customer arrival times and service times given in the table below. These force the following values for T_i, W_i^Q, and D_i...

i	A_i	T_i	W_i^Q	S_i	D_i
1	3	3	0	7	10
2	4	10	6	6	16
3	6	16	10	4	20
4	10	20	10	6	26

And so the average waiting time for the four customers is $\sum_{i=1}^{4} W_i^Q/4 = 6.5$. $\Box$

We will show in Chapter 3 how to compute other performance measures such as the percent of time that the server is busy, the average number of people in line, etc. You can also mess around with the queueing application in `waskos.exe` if you'd like more insight.

Remark: Meanwhile, it turns out that the answer to our original question regarding what happens when the arrival rate approaches the service rate is, perhaps surprisingly, (d) — the line will tend to get pretty long.

Remark: Notice anything interesting about the word "queueing"? Or "queueoid"? Hint: Look at all of those vowels!

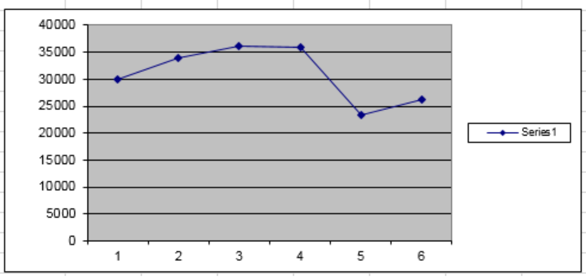
1.6.6 Stock Market Follies

Suppose that we are simulating a portfolio of stocks from different financial sectors, whose values change randomly from year to year with various average rates of increase and volatilities. Simulation allows us to consider different sector mixes for our portfolio, and then watch as the portfolio value changes over time. You can download a simple spreadsheet application demonstrating stock market simulation from

`www.isye.gatech.edu/~sman/courses/6644/Module03-Spreadsheet-Stock-Portfolio.xls.`

The spreadsheet — which will be discussed more-thoroughly in Chapter 3 — follows the evolution of a portfolio of six stock sectors over a five-year period.

	mean	sd	year0	year1	year2	year3	year4	year5
general year performance (so stocks are correlated)				1.12	1.08	1.00	0.73	1.02
energy	0.05	0.30	5000	4094	4507	3607	2991	3156
pharmaceuticals	0.06	0.20	5000	4664	3680	3988	2737	3130
entertainment	0.04	0.10	5000	5877	6974	6548	4670	4501
insurance	0.07	0.05	5000	5729	6454	6797	5597	6141
banking	0.06	0.30	5000	6922	8870	7611	4758	5729
computer technology	0.18	0.30	5000	6557	5761	7340	2657	3568
Totals			30000	33841.95	36246.22	35890.87	23409.34	26224.46

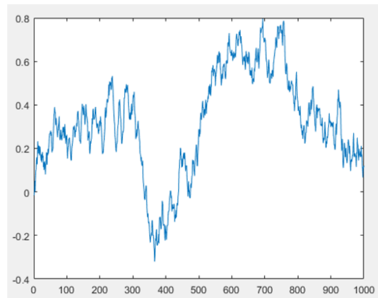


Sadly, in the preceding realization of the simulation, we lost some money by year five; but you can easily re-run the simulation and maybe make some money next time!

1.6.7 Taking a Random Walk

Suppose that every minute, you either earn a dollar or lose a dollar, each with probability 1/2. If you start at 0, what will your fortune be at time n. This is an example of a simple *random walk*, which you may have learned about in a Stochastic Processes course from days of yore. If you let n become large and standardize the fortune properly, we will see in Chapter 7 that the random walk converges to what is known as *Brownian motion*. This can be simulated quite easily and has tremendous applications in many fields, including finance. Indeed the example sample path of Brownian motion below looks quite a bit like a stock price as it evolves over time, eh?

Fun Facts: Brownian motion was discovered by the English botanist Robert Brown in 1827. Its properties were analyzed in some detail by Einstein in 1905, and work on Brownian motion has led to a number of Nobel Prizes, notably, for Black and Scholes for their work on option pricing. Brownian motion has become so important that the color brown was named in its honor on April 1, 2018.



A sample path of Brownian motion.

---

§1.6 Assessment

1. Suppose there are 40 random people in a room. What is the probability that at least two of them will have the same birthday?

   (a) Close to 0

   (b) A bit less than 1/2

   (c) Almost exactly 1/2

   (d) Somewhat greater than 1/2

2. Inscribe a circle in a unit square and toss 1000 random darts at the square. Suppose that 800 of those darts land in the circle. Using the technology developed in this lesson, what is the resulting estimate for π?

   (a) -3.2

   (b) 2.8

   (c) 3.0

   (d) 3.2

   (e) 4.0

3. TRUE or FALSE? All random number generators perform pretty much the same.

4. Suppose customers to a barber shop show up at times 4 and 11. Moreover, suppose that it takes the barber 12 minutes to serve customer 1 and then 14 minutes to serve customer 2. When does customer 2 leave the barber?

   (a) 18

   (b) 25

   (c) 30

   (d) 40

1.7 Generating Randomness

We have seen from the previous examples that in order to run simulations, we need a steady supply of random variates to represent quantities such as customer interarrival times, service times, daily demands, and so on. How do we provide these guys to the simulation?

In §1.7.1, we'll start off easy by addressing the generation of a sequence of independent numbers distributed uniformly on the interval [0,1]. These building block uniforms will subsequently be used in §1.7.2 to generate several other more-sophisticated random variables.

If you happen to be a little rusty on your probability theory, don't be a Fretting Freddie about it for now, since we will keep the discussion pretty much self-contained. But if you happen to be a Worry Wanda, you can take a quick look at the probability boot camp in Chapter 2.

1.7.1 Generating Unif(0,1) Random Numbers

We discuss at a high level the goal of generating a sequence of independent Unif(0,1) random numbers, i.e., numbers uniformly distributed on [0,1]. Details are more-fully developed in Chapter 6. To be completely honest, we'll admit at the onset that generating numbers that are truly independent Unif(0,1)'s is actually a difficult, inefficient business, and so we will relax this requirement a bit and instead generate what are known as *pseudo-random numbers (PRNs)*. Such PRNs are generated via a *deterministic* algorithm, and thus are not really random at all. Our goal will therefore switch to producing PRNs that *appear* to be Unif(0,1) to humans and to statistical tests. It turns out that the technology is indeed able to supply simulation users with some extremely good PRNs, so the true randomness issue will be of no practical concern.

In any case, here are some nice points about PRNs (at least the good ones):

- They have well-understood properties, and can be generated in a reproducible way.
- PRN algorithms are simple, efficient, and quick.
- PRNs are virtually indistinguishable from "real" Unif(0,1) random numbers.
- By applying certain mathematical transformations, it is easy to go from Unif(0,1)'s to most other random variables. (Stay tuned.)

As we have mentioned, although PRNs appear to be random, they are actually generated by a deterministic algorithm. The most-common type of algorithm for generating Unif(0,1) PRNs is called the *linear congruential generator (LCG):*

1. Choose an integer "seed," X_0, to initialize the algorithm.

2. Set the next integer $X_i = (aX_{i-1}+c) \bmod (m)$, where a, c, and m are carefully chosen integer constants, and mod is the modulus (or remainder) function, e.g., $17 \bmod (6) = 5$.

3. Set the ith PRN to $R_i = X_i/m$. This transforms the integer X_i to a PRN $R_i \in [0, 1]$.

Trivial Example: For purposes of illustration, consider the LCG

$$X_i = (5X_{i-1} + 3) \bmod (8)$$

If $X_0 = 0$, we have $X_1 = (5X_0 + 3) \bmod 8 = 3$; and continuing,

i	0	1	2	3	4	5	6	7	8	9
X_i	0	3	2	5	4	7	6	1	0	3
R_i	0	$\frac{3}{8}$	$\frac{2}{8}$	$\frac{5}{8}$	$\frac{4}{8}$	$\frac{7}{8}$	$\frac{6}{8}$	$\frac{1}{8}$	0	$\frac{3}{8}$

so that the sequence starts repeating with $X_8 = 0$. $\Box$

This is a *full-period generator*, since it has cycle length $m = 8$ before the repetition starts. Generally speaking, full-period is a good thing. Of course, one would never use a PRN generator that had a cycle length of only 8, but that problem will be resolved in the next example...

Better Example: Here's an LCG that works OK if you're on a desert island — it's not nearly the best ever, but at least you can take it home to your parents once you get off the island.

$$X_i = 16807X_{i-1} \bmod (2^{31} - 1) \quad \text{and} \quad R_i = X_i/m.$$

This LCG produces pretty decent PRNs in the sense that the cycle length is over two billion, and it passes most reasonable statistical tests for independent Unif(0,1)'s. But there are way better generators are out there, which we'll save for later. $\Box$

1.7.2 Generating Other Random Variates

PRNs are the basic building blocks for everything else. Starting from PRNs that are supposedly independent Unif(0,1)'s, it is possible generate random variates from many other probability distributions by applying appropriate transformations. We present a couple of archetypal examples from Chapter 7.

Example: If U is Unif(0,1), then the *inverse transform method* reveals that $(-1/\lambda)\ln(U)$ gives us an Exponential(λ) random variate. $\Box$

Example: If U_1 and U_2 are independent Unif(0,1) random variables, then (remarkably) the *Box-Muller method* establishes the fact that the pair

$$Z_1 = \sqrt{-\ell n(2\pi U_1)}\cos(2\pi U_2) \quad \text{and} \quad Z_2 = \sqrt{-\ell n(2\pi U_1)}\sin(2\pi U_2)$$

are independent standard normal random variates. □

Chapter 7 will discuss a library of random variate generate techniques — even the Brownian motion stochastic process discussed in §1.6.7.

---

§1.7 Assessment

1. Suppose we are using the (terrible) pseudo-random number generator

$$X_i \ = \ (5X_{i-1} + 3)\,\mathrm{mod}(8),$$

with starting value ("seed") $X_0 = 1$. Find the second PRN, $U_2 = X_2/m = X_2/8$.

   (a) 0
   (b) 1/8
   (c) 3/8
   (d) 3

2. Suppose that we generate a pseudo-random number $U = 0.728$. Use this to generate an Exponential($\lambda = 3$) random variate.

   (a) -0.106
   (b) 0.106
   (c) -0.952
   (d) 0.952

1.8 Analyzing Randomness

Simulation output can be badly behaved and difficult or impossible to analyze using the "usual" statistical methods. The issue is that the observations we encounter in simulations often violate the assumptions that you learned in your Baby Stats class. For example, consider consecutive customer waiting times in the line at the post office. Such waits are:

- Not normally distributed — they are usually skewed.
- Not identically distributed — patterns change throughout the day.
- Not independent — they are usually somewhat correlated. If you wait a long time in line, then the guy next to you is likely to wait a long time in line as well.

If you use standard statistical methods to analyze such naughty data, then the conclusions you draw might be misleading. For instance, what you think is a 95% confidence interval for the mean waiting time might actually only be a 30% confidence interval (unknown to you)!

In order to carry out proper simulation output analysis, we usually divide the output world into two general cases: *terminating simulations* and *steady-state simulations*. The methods that we use to conduct the analysis differ somewhat depending on the category.

1.8.1 Terminating Simulations

In a terminating simulation, we are interested in the short-term behavior of a system. These sorts of simulations characteristically have clear beginnings and endings. Typical questions that terminating simulations may address are:

- What is the average customer waiting time in a bank over the course of a day?
- What is the average number of infected people during an influenza pandemic?

Terminating simulations are usually analyzed via the method of *independent replications*, which consists of the following:

- Make multiple, independent replications (runs) of the simulation. Each run is made under identical conditions — i.e., with the same system parameters and setup — but using different random numbers to drive the simulation.
- The *sample means* resulting from the independent replications are assumed to be approximately independent and identically distributed (i.i.d.) normal random variables. At this point, we can use classical statistical techniques on the i.i.d. sample means. (We repeat, however, that the original observations from each replication, e.g., the consecutive waiting times, may be correlated and highly non-normal.)

1.8.2 Steady-State Simulations

In steady-state simulation, we are interested in the long-term behavior of a system. A continuously running assembly line, for instance, might be a good candidate for analysis via steady-state simulation analysis, since we would likely be interested in long-term behavior of the line.

Most simulations do not begin in steady state, and usually require a warm-up period before collecting representative data, in order to avoid possible (likely!) *initialization (or start-up) bias*. For example, the waiting times of the first few customers in a queueing system that starts out empty are likely to be shorter, on average, than steady-state waiting times. In any case, failure to warm up the simulation can ruin statistical analysis, since the behavior of the system in the initial period is likely to be very different than its behavior in steady state.

There are many methods for dealing with steady-state simulation data, some of which we'll talk about in more detail in Chapter 9:

- Batch Means
- Overlapping Batch Means / Spectral Analysis
- Standardized Time Series
- Regeneration

The bellwether *method of batch means* is a powerful, yet relatively simple technique for taking steady-state simulation output data and rendering it into a form suitable for analysis using standard statistical methods. The method consists of the following steps:

- Make one long run (instead of many shorter replications).
- Warm up the simulation before collecting any data.
- Chop the remaining (i.e., the "steady-state") observations into contiguous batches.
- The *sample means* from each batch are assumed to be approximately i.i.d. normal.
- Now you can use classical statistical techniques on the i.i.d. batch means (not on the original observations!).

---

§1.8 Assessment

1. TRUE or FALSE? Simulation outputs such as consecutive customer waiting times are almost always independent and identically distributed normal random variables.

2. Let's simulate a bank that closes at 4:30 p.m. What kind of simulation approach would you take?

   (a) Steady-state simulation

   (b) Terminating simulation

   (c) Arnold Schwarzenegger simulation

   (d) I'm from The University of Georgia. What is simulation? And what is bank?

1.9 Exercises

1. (§1.3) Suppose you throw a rock off a cliff having height $h_0 = 1000$ feet. You're a strong bloke, so the initial downward velocity is $v_0 = -100$ feet/sec (slightly under 70 miles/hr). Further, in this neck of the woods, it turns out there is no friction in the atmosphere — amazing! Now you remember from your Baby Physics class that the height after time t is

$$h(t) \;=\; h_0 + v_0 t - 16t^2.$$

When does the rock hit the ground?

2. (§1.3) Consider a single-server queueing system where the times between customer arrivals are independent, identically distributed Exp(λ = 2/hr) random variables; and the service times are i.i.d. Exp(μ = 3/hr). Unfortunately, if a potential arriving customer sees that the server is occupied, he gets mad and leaves the system. Thus, the system can have either 0 or 1 customer in it at any time. This is what's known as an M/M/1/1 queue. If $P(t)$ denotes the probability that a customer is being served at time t, trust me that it can be shown that

$$P(t) \;=\; \frac{\lambda}{\lambda+\mu} + \left[P(0) - \frac{\lambda}{\lambda+\mu}\right] e^{-(\lambda+\mu)t}.$$

If the system is empty at time 0, i.e., $P(0) = 0$, what is the probability that there will be no people in the system at time 1 hr?

3. (§1.4) Harry Markowitz (one of the big wheels in simulation language development) won his Nobel Prize for portfolio theory in 1990, though the work that earned him the award was conducted much earlier in the 1950s. Who won the 1990 Prize with him?

4. (§1.5) Which of the following situations might be good candidates to use simulation? (There may be more than one correct answer.)

   (a) We are interested in investing our entire portfolio in fixed-interest U.S. bonds, and we are interested in determining the portfolio's value in 5 years.

   (b) We are interested in investing one half of our portfolio in fixed-interest U.S. bonds and the remaining half in a stock market equity index. We have some information concerning the distribution of stock market returns, but we do not really know what will happen in the market with certainty.

   (c) We have a new strategy for baseball batting orders, and we would like to know if this strategy beats other commonly used batting orders (e.g., a fast guy bats first, a big, strong guy bats fourth, etc.). We have information on the performance of the various team members, but there's a lot of randomness in baseball.

(d) We have an assembly station in which "customers" (for instance, parts to be manufactured) arrive every 5 minutes exactly and are processed in precisely 4 minutes by a single server. We would like to know how many parts the server can produce in a hour.

(e) Consider an assembly station in which parts arrive randomly, with independent exponential interarrival times. There is a single server who can process the parts in a random amount of time that is normally distributed. Moreover, the server takes random breaks every once in a while. We would like to know how big any line is likely to get.

(f) Suppose we are interested in determining the number of doctors needed on Friday night at a local emergency room. We need to insure that 90% of patients get treatment within one hour.

5. (§1.6) The planet Glubnor has 50-day years.

   (a) Suppose there are 2 Glubnorians in the room. What's the probability that they'll have the same birthday?

   (b) Now suppose there are 3 Glubnorians in the room. (They're big, so the room is getting crowded.) What's the probability that at least two of them have the same birthday?

6. (§1.6) Inscribe a circle in a unit square and toss $n = 500$ random darts at the square.

   (a) Suppose that 380 of those darts land in the circle. Using the technology developed in this lesson, what is the resulting estimate for π?

   (b) What would our estimate be if we let $n \to \infty$ and we applied the same ratio strategy?

   (c) With the previous question in mind, suppose that we can somehow toss n random darts into a unit cube. Further, suppose that we've inscribed a *sphere* with radius 1/2 inside the cube. Let $\hat{p}_n$ be the proportion of the n darts that actually fall within the sphere. Give a Monte Carlo scheme to estimate π.

7. (§1.6) Suppose customers arrive at a single-server ice cream parlor times 3, 6, 15, and 17. Further suppose that it takes the server 7, 9, 6, and 8 minutes, respectively, to serve the four customers. When does customer 4 leave the shoppe?

8. (§1.7) Suppose we are using the (awful) pseudo-random number generator

$$X_i \;=\; (5X_{i-1} + 1)\,\mathrm{mod}(8),$$

   with starting value ("seed") $X_0 = 1$. Find the second PRN, $U_2 = X_2/m = X_2/8$.

9. (§1.7) Suppose we are using the "decent" pseudo-random number generator

$$X_i \;=\; 16807\, X_{i-1}\, \mathrm{mod}(2^{31}-1),$$

with seed $X_0 = 12345678$. Find the resulting integer X_1. Feel free to use something like Excel if you need to.

10. (§1.7) Suppose that we generate a pseudo-random number $U = 0.128$. Use this to generate an Exponential($\lambda = 1/3$) random variate.

11. (§1.8) Which scenarios are most apt for a steady-state analysis? (More than one answer may be right.)

    (a) We simulate a bank from noon till 1:00 pm.

    (b) We investigate a production line that runs 24/7.

    (c) We are interested in seeing what our portfolio is likely to be 3 months from now.

    (d) We try to estimate the unemployment rate 30 years from now.

Chapter 2

Calculus, Probability, and Statistics Boot Camps

Simulation has a certain prerequisites, including calculus, probability, and statistics. It's been our experience that some of our readers experience minor memory issues and therefore want a wee bit of review on these topics; so this chapter offers corresponding boot camps. Now, this material is sort of *optional* if you're already comfortable with it, but if not, you really ought to take a look.

Ah, the memories are flooding back! We start out with a calculus primer in §§2.1–2.4, including some old friends: derivatives, solving equations for zeroes, integration, and numerical integration. Then it's off to probability camp in §§2.5–2.13, where we'll review probability basics and the concept of random variables (RVs). Since this is a simulation text, the next task is to do a bit of an aside and learn about a very easy way to simulate certain RVs. Then it's time to read about Great Expectations (you'll get the joke when you see it), after which we'll discuss functions of RVs and bivariate RVs. If all of these terms are familiar to you, great! If not, well, that's why we're going to camp!

We continue our probability review with a tough lesson on conditional expectation (which you may not have seen when you first took probability / statistics), but then easier lessons on covariance, probability distributions, and the normal distribution. The normal distribution is especially important because of its link to the Central Limit Theorem, which is of course the most-important result in the history of The Universe.

Finally, §§2.14–2.16 go over a smattering of statistics, including point and confidence interval estimation, after which we can come home from camp.

I. Calculus Primer

Goal: This section provides a brief review of various calculus tidbits that we'll be using later on.

2.1 Basics and Derivatives

First of all, let's suppose that $f(x)$ is a *function* that maps values of x from a certain *domain* X to a certain *range* Y, which we can denote by the shorthand $f : X \to Y$.

Example: If $f(x) = x^2$, then the function takes x-values from the real line $\mathbb{R}$ to the nonnegative portion of the real line $\mathbb{R}^+$. $\Box$

Definition: We say that $f(x)$ is a *continuous* function if, for any x_0 and $x \in X$, we have $\lim_{x \to x_0} f(x) = f(x_0)$, where "lim" denotes a *limit* and $f(x)$ is assumed to exist for all $x \in X$.

Examples: The function $f(x) = 3x^2$ is continuous for all x. The function $f(x) = \lfloor x \rfloor$ (round down to the nearest integer, e.g., $\lfloor 3.4 \rfloor = 3$) has a "jump" discontinuity at any integer x. $\Box$

Intermediate Value Theorem: If $f(x)$ is continuous on the interval $x \in [a, b]$, then it also takes any value between $f(a)$ and $f(b)$ at some point in the interval.

Definition: If $f(x)$ is continuous, then it is *differentiable* (has a *derivative*) if

$$\frac{d}{dx} f(x) \;\equiv\; f'(x) \;\equiv\; \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists and is well-defined for any given x. Think of the derivative as the slope of the function.

Examples: Some well-known derivatives are:

$$\begin{aligned}
[x^k]' &= kx^{k-1}, \\
[e^x]' &= e^x, \\
[\sin(x)]' &= \cos(x), \\
[\cos(x)]' &= -\sin(x), \\
[\ell\text{n}(x)]' &= \frac{1}{x},
\end{aligned}$$

$$[\arctan(x)]' = \frac{1}{1+x^2}. \quad \Box$$

Theorem: Some well-known properties of derivatives are:

$$[af(x)+b]' = af'(x),$$

$$[f(x)+g(x)]' = f'(x)+g'(x),$$

$$[f(x)g(x)]' = f'(x)g(x)+f(x)g'(x) \quad \textit{(product rule)},$$

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{g(x)f'(x)-f(x)g'(x)}{g^2(x)} \quad \textit{(quotient rule)}^1,$$

$$[f(g(x))]' = f'(g(x))g'(x) \quad \textit{(chain rule)}^2.$$

Example: Suppose that $f(x) = x^2$ and $g(x) = \ell\mathrm{n}(x)$. Then

$$[f(x)g(x)]' = \frac{d}{dx}x^2\ell\mathrm{n}(x) = 2x\ell\mathrm{n}(x)+x,$$

$$\left[\frac{f(x)}{g(x)}\right]' = \frac{d}{dx}\frac{x^2}{\ell\mathrm{n}(x)} = \frac{2x\ell\mathrm{n}(x)-x}{\ell\mathrm{n}^2(x)},$$

$$[f(g(x))]' = 2g(x)g'(x) = \frac{2\ell\mathrm{n}(x)}{x}. \quad \Box$$

Remark: The second derivative $f''(x) \equiv \frac{d}{dx}f'(x)$ and is the "slope of the slope." If $f(x)$ is "position," then $f'(x)$ can be regarded as "velocity," and as $f''(x)$ as "acceleration."

If $f(x)$ is continuous and has a nice, beautiful derivative, then a *(local) minimum* or *local maximum* of $f(x)$ can only occur when the slope of $f(x)$ is zero, i.e., only when $f'(x) = 0$, say at $x = x_0$. Exception: A min or max can occur at the *endpoints* of the interval of interest.

Then if $f''(x_0) < 0$, you get a max; if $f''(x_0) > 0$, you get a minimum; and if $f''(x_0) = 0$, you get a *point of inflection*.

Example: Find the value of x that minimizes $f(x) = e^{2x} + e^{-x}$. The minimum can only occur when $f'(x) = 2e^{2x} - e^{-x} = 0$. After a little algebra, we find that this occurs at $x_0 = -(1/3)\ell\mathrm{n}(2) \approx -0.231$. It's also easy to show that $f''(x) > 0$ for all x; and so x_0 yields a minimum. $\Box$

---

[1] Ho dee Hi minus Hi dee Ho over Ho Ho.

[2] www.youtube.com/watch?v=gGAiW5dOnKo

§2.1 Assessment

1. If $f(x) = \ell\text{n}(2x)$, find the derivative $f'(x)$.
   (a) $2x$
   (b) $\frac{1}{2}\ell\text{n}(x)$
   (c) $1/x$
   (d) $x/2$
2. If $f(x) = \sin(\ell\text{n}(x))$, find the derivative $f'(x)$.
   (a) $\cos(\ell\text{n}(x))$
   (b) $\cos(\ell\text{n}(x))/x$
   (c) $\sin(\ell\text{n}(x))/x$
   (d) $\sin(\ell\text{n}(x))$
   (e) I'm from UGA — I'm scared of math, and I don't know!

2.2 Saved By Zero! Solving Equations

Finding Zeroes: Speaking of solving for a 0, how might you do it if a continuous function $g(x)$ is a complicated nonlinear fellow?

- Trial-and-error (not so great).
- Bisection (divide-and-conquer).
- Newton's method (or some variation)
- Fixed-point method (perhaps we'll get around to this later).

Bisection: Suppose you can find x_1 and x_2 such that $g(x_1) < 0$ and $g(x_2) > 0$. (We'll follow similar logic if the inequalities are both reversed.) By the Intermediate Value Theorem (IVT), there must be a zero in $[x_1, x_2]$, that is, $x^\star \in [x_1, x_2]$ such that $g(x^\star) = 0$.

We'll use a divide-and-conquer strategy to conduct our search for $x^\star$. To this end, take $x_3 = (x_1 + x_2)/2$. If $g(x_3) < 0$, then there must be a zero in $[x_3, x_2]$. Otherwise, if $g(x_3) > 0$, then there must be a zero in $[x_1, x_3]$. In either case, you've reduced the length of the search interval.

Continue in this same manner until the length of the search interval is as small as desired.

Exercise: Try this out for $g(x) = x^2 - 2$, and come up with an approximation for $\sqrt{2}$.

Solution: Let's start out with $x_1 = 1$ and $x_2 = 2$, so that $g(x_1) = -1 < 0$ and $g(x_2) = 2 > 0$; and then the IVT implies that the zero $x^\star \in [1, 2]$.

Then $x_3 = (x_1 + x_2)/2 = 1.5$ so that $g(x_3) = 0.25 > 0$; and now the IVT implies $x^\star \in [1, 1.5]$. Continuing, we have the following iterations.

i	x_i	$g(x_i)$	comments
1	1	-1	
2	2	2	look in [1,2]
3	1.5	0.25	look in [1,1.5]
4	1.25	-0.438	look in [1.25,1.5]
5	1.375	< 0	look in [1.375,1.5]
6	1.4375	> 0	look in [1.375,1.4375]

Thus, at this point, we know that $x^\star \in [1.375, 1.4375]$; and we see that the x_i's do seem to be converging to the true answer of $\sqrt{2} = 1.414\ldots$. $\Box$

Newton's Method: This technique is a bit faster that bisection, since it takes advantage of information about the derivative of $g(x)$.

Suppose you can find a reasonable first guess for the zero, say, x_i, where we start off at iteration $i = 0$. If $g(x)$ has a nice, well-behaved derivative (which doesn't happen to be too flat near the zero of $g(x)$), then iterate your guess as follows:

$$x_{i+1} = x_i - \frac{g(x_i)}{g'(x_i)}.$$

Keep going until things appear to converge.

This makes sense since for x_i and x_{i+1} close to each other and the zero $x^\star$, we have

$$g'(x_i) \approx \frac{g(x^\star) - g(x_i)}{x^\star - x_i}.$$

Exercise: Try Newton out for $g(x) = x^2 - 2$, noting that the iteration step is to set

$$x_{i+1} = x_i - \frac{x_i^2 - 2}{2x_i} = \frac{x_i}{2} + \frac{1}{x_i}.$$

Solution: Let's start with a bad guess of $x_1 = 1$. Then

$$\begin{aligned}
x_2 &= \frac{x_1}{2} + \frac{1}{x_1} = \frac{1}{2} + 1 = 1.5 \\
x_3 &= \frac{x_2}{2} + \frac{1}{x_2} \approx \frac{1.5}{2} + \frac{1}{1.5} = 1.4167 \\
x_4 &= \frac{x_3}{2} + \frac{1}{x_3} \approx 1.4142 \quad \text{Wow!} \quad \Box
\end{aligned}$$

---

§2.2 Assessment

1. Which of the following methods *cannot* be used to find the zeroes of a complicated function?

   (a) trial-and-error

   (b) bisection

   (c) Newton's method

   (d) Newman's method acting

2. Use your favorite numerical method to solve $g(x) = x^2 - 3 = 0$ for $x \in [0, 1]$.

   (a) $x = 1.414$

   (b) $x = 1.732$

   (c) $x = 2$

   (d) x is an imaginary number

2.3 Integrals and Beyond

Definition: The function $F(x)$ having derivative $f(x)$ is called the *antiderivative* (or *indefinite integral*). It is denoted by $F(x) = \int f(x)\,dx$.

Fundamental Theorem of Calculus: If $f(x)$ is continuous, then the area under the curve for $x \in [a, b]$ is denoted and given by the *definite integral* [3]

$$\int_a^b f(x)\,dx \;\equiv\; F(x)\Big|_a^b \;\equiv\; F(b) - F(a).$$

Example: Some well-known indefinite integrals are:

$$\begin{aligned}
\int x^k\,dx &= \frac{x^{k+1}}{k+1} + C \quad \text{for } k \neq -1 \\
\int \frac{dx}{x} &= \ell\mathrm{n}|x| + C, \\
\int e^x\,dx &= e^x + C, \\
\int \cos(x)\,dx &= \sin(x) + C,
\end{aligned}$$

[3] "I'm *really* an integral!"

$$\int \frac{dx}{1+x^2} \;=\; \arctan(x) + C,$$

where C is an arbitrary constant. $\square$

Example: It is easy to see that

$$\int \frac{d\,\text{cabin}}{\text{cabin}} \;=\; \ell\text{n}|\text{cabin}| + C \;=\; \text{houseboat}. \quad \square$$

Theorem: Some well-known properties of definite integrals are:

$$\int_a^a f(x)\,dx \;=\; 0,$$

$$\int_a^b f(x)\,dx \;=\; -\int_b^a f(x)\,dx,$$

$$\int_a^b f(x)\,dx \;=\; \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Theorem: Some other properties of general integrals are:

$$\int [f(x)+g(x)]\,dx \;=\; \int f(x)\,dx + \int g(x)\,dx,$$

$$\int f(x)g'(x)\,dx \;=\; f(x)g(x) - \int g(x)f'(x)\,dx \quad \textit{(integration by parts)}^4,$$

$$\int f(g(x))g'(x)\,dx \;=\; \int f(u)\,du \quad \textit{(substitution rule)}^5.$$

Example: Using integration by parts with $f(x) = x$ and $g'(x) = e^{2x}$ and the chain rule, we have

$$\int_0^1 xe^{2x}\,dx \;=\; \frac{xe^{2x}}{2}\Big|_0^1 - \int_0^1 \frac{e^{2x}}{2}\,dx \;=\; \frac{e^2}{2} - \frac{e^{2x}}{4}\Big|_0^1 \;=\; \frac{e^2+1}{4}. \quad \square$$

Definition: Derivatives of arbitrary order k can be written as $f^{(k)}(x)$ or $\frac{d^k}{dx^k}f(x)$. By convention, $f^{(0)}(x) = f(x)$.

The *Taylor series expansion* of $f(x)$ about a point a is given by

$$f(x) \;=\; \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(x-a)^k}{k!}.$$

[4] www.youtube.com/watch?v=OTzLVIc-O5E
[5] www.youtube.com/watch?v=eswQl-hcvU0

The *Maclaurin series* is simply Taylor expanded around $a = 0$.

Example: Here are some famous Maclaurin series.

$$\begin{aligned}
\sin(x) &= \sum_{k=0}^{\infty} \frac{(-1)^{k+1} x^{2k+1}}{(2k+1)!}, \\
\cos(x) &= \sum_{k=0}^{\infty} \frac{(-1)^{k} x^{2k}}{(2k)!}, \\
e^x &= \sum_{k=0}^{\infty} \frac{x^k}{k!}.
\end{aligned}$$

Example: And while we're at it, here are some miscellaneous sums that you should know.

$$\begin{aligned}
\sum_{k=1}^{n} k &= \frac{n(n+1)}{2}, \\
\sum_{k=1}^{n} k^2 &= \frac{n(n+1)(2n+1)}{6}, \\
\sum_{k=0}^{\infty} p^k &= \frac{1}{1-p} \quad \text{(for } -1 < p < 1\text{)}.
\end{aligned}$$

Theorem: Occasionally, we run into trouble when taking indeterminate ratios of the form $0/0$ or ∞/∞. In such cases, *L'Hôspital's Rule*[6] is useful: If the limits $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both go to 0 or both go to ∞, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}.$$

Example: L'Hôspital shows that

$$\lim_{x\to 0} \frac{\sin(x)}{x} = \lim_{x\to 0} \frac{\cos(x)}{1} = 1. \quad \Box$$

---

§2.3 Assessment

1. Find $\int_0^1 (x+1)^2\, dx$.

   (a) 1/2

[6] This rule makes me sick.

   (b) 7/2
   (c) 8/3
   (d) 7/3

2. Find $\int_1^2 \ell\text{n}(x)\,dx$.
   (a) 1
   (b) $2\ell\text{n}(2) - 1$
   (c) $\ell\text{n}(2)$
   (d) $3\ell\text{n}(2) - 1$

3. Find $\lim_{x\to 0}(e^x - 1)/\sin(x)$. (Hint: This problem will make you so sick, you'll have to go to the ...?)
   (a) 1
   (b) 0
   (c) ∞
   (d) undetermined

2.4 Numerical Integration Methods

Riemann Sums: Sometimes you just can't solve an integral in closed form. In those cases, we can use a variety of methods to numerically *approximate* the integral. Perhaps you still have nightmares about the Riemann sum approximations that you learned about before you actually started doing integrals; luckily, the passage of time makes all things easier, and this is certainly the case here.

The trick is simply to approximate the area under the nice, continuous function $f(x)$ from $x = a$ to b by adding up the areas of n adjacent rectangles of width $\Delta x = (b-a)/n$ and height $f(x_i)$, where $x_i = a + i\Delta x$ is the right-hand endpoint of the ith rectangle. Thus,

$$\int_a^b f(x)\,dx \quad \approx \quad \sum_{i=1}^{n} f(x_i)\Delta x \;=\; \frac{b-a}{n}\sum_{i=1}^{n} f\left(a + \frac{i(b-a)}{n}\right).$$

In fact, as $n \to \infty$, this result becomes an equality.

Exercise: Try it out on $\int_0^1 \sin(\pi x/2)\,dx$ (which secretly equals $2/\pi$) for different values of n, and see for yourself.

Solution: Since we're so nice (and since we don't want you to have nightmares again), we've made things easy. In this problem, we've thoughtfully taken $a = 0$ and $b = 1$, so that $\Delta x = 1/n$ and $x_i = i/n$, which simplifies the notation a bit. Then

$$\int_a^b f(x)\,dx \;=\; \int_0^1 f(x)\,dx \;\approx\; \sum_{i=1}^{n} f(x_i)\Delta x \;=\; \frac{1}{n}\sum_{i=1}^{n} \sin\left(\frac{\pi i}{2n}\right).$$

For $n = 100$, this calculates out to a value of 0.6416, which is pretty close to the true answer of $2/\pi \approx 0.6366$. $\square$

Trapezoid Rule: Here's a well-known method that works a little more quickly (i.e., requires smaller n) than Riemann. Now we have

$$\begin{aligned}\int_a^b f(x)\,dx \;&\approx\; \left[\frac{f(x_0)}{2} + \sum_{i=1}^{n-1} f(x_i) + \frac{f(x_n)}{2}\right]\Delta x \\ &=\; \frac{b-a}{n}\left[\frac{f(a)}{2} + \sum_{i=1}^{n-1} f\left(a + \frac{i(b-a)}{n}\right) + \frac{f(b)}{2}\right].\end{aligned}$$

The messy term in brackets arises from the use of adjacent trapezoids instead of rectangles to approximate the area.

You may want to try it out on $\int_0^1 \sin(\pi x/2)\,dx$.

Monte Carlo Integration: We discussed MC integration a bit in §1.6; and you will soon learn more details in Chapter 3. But for now, let $U_1, U_2, \ldots, U_n$ denote a sequence of independent Unif(0,1) random numbers. You can easily obtain such a sequence from, e.g., Excel using the function `RAND()`. In any case, it can be shown that

$$\int_a^b f(x)\,dx \;\approx\; \frac{b-a}{n}\sum_{i=1}^{n} f(a + (b-a)U_i),$$

with the result becoming an equality as $n \to \infty$.

Yet again, feel free to try it out on $\int_0^1 \sin(\pi x/2)\,dx$.

---

§2.4 Assessment

1. Which of the following is *not* an integration method discussed in this lesson?

   (a) Riemann sums

   (b) Newmann sums

   (c) Trapezoid Rule

   (d) The Monte Carlo method

2. How does a mathematician capture a wild man-eating zoid?

(a) You catch a zoid.

(b) You capture a zoid.

(c) You trap a zoid.

(d) Trick question! It's always best to avoid a zoid altogether!

3. Find the approximate value of the integral $\int_0^1 x^2\,dx$ using the lesson's form of the Riemann sum with $n = 4$, specifically,

$$\int_0^1 x^2\,dx \;\approx\; \frac{1}{n}\sum_{i=1}^{n}(i/n)^2.$$

(a) -3

(b) 15/32

(c) 0

(d) 11

II. Probability Primer

We now present a crash course that will give the reader access to the fundamentals of applied probability.

2.5 Probability Basics

We assume that the reader is familiar with the basic concepts of sample spaces and events, and the definition of probability. But if you feel that you need a bit more of a review, you can see

`www.isye.gatech.edu/~sman/courses/6739.`

Definition: $P(A|B) \equiv P(A \cap B)/P(B)$ is the *conditional probability of A given B*.

Example: Toss a fair die. Let $A = \{1, 2, 3\}$ and $B = \{3, 4, 5, 6\}$. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{4/6} = 1/4. \quad \square$$

Definition: If $P(A \cap B) = P(A)P(B)$, then A and B are *independent* events.

Theorem: If A and B are independent, then $P(A|B) = P(A)$.

Proof: By the definition of conditional probability and then independence, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A). \quad \square$$

Example: Toss two dice. Let $A =$ "Sum is 7" and $B =$ "First die is 4". Then

$$P(A) = 1/6, \quad P(B) = 1/6, \quad \text{and}$$

$$P(A \cap B) = P((4, 3)) = 1/36 = P(A)P(B).$$

So A and B are independent. $\square$

Definition: A *random variable (RV)* X is a function from the sample space Ω to the real line $\mathbb{R}$, i.e., $X : \Omega \to \mathbb{R}$.

Example: Let X be the sum of two dice rolls. Then $X((4,6)) = 10$. In addition,

$$P(X = x) \;=\; \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \quad\vdots & \\ 1/36 & \text{if } x = 12 \\ 0 & \text{otherwise} \end{cases} \qquad \Box$$

Definition: If the number of possible values of a RV X is finite or countably infinite, then X is a *discrete* RV. Its *probability mass function (pmf)* is $f(x) \equiv P(X = x)$. Note that $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) \;=\; \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad \Box$$

Example: Here are some well-known discrete RVs: Bernoulli(p), Binomial(n,p), Geometric(p), Negative Binomial, Poisson(λ), etc. (If you don't remember them, don't worry… we'll pick them up as we need them.)

Notation: "$\sim$" means "is distributed as." For instance, $X \sim \text{Bin}(n,p)$ means that X has the Binomial distribution with parameters n and p.

Definition: A *continuous* RV is one with probability zero at every individual point. A RV is continuous if there exists a *probability density function (pdf)* $f(x)$ such that $P(X \in A) = \int_A f(x)\,dx$ for every set A. Note that $\int_x f(x)\,dx = 1$.

Example: Pick a random number between 3 and 7. Then

$$f(x) \;=\; \begin{cases} 1/4 & \text{if } 3 \le x \le 7 \\ 0 & \text{otherwise} \end{cases} \qquad \Box$$

Example: Here are some well-known continuous RVs: Uniform(a,b), Exponential(λ), Normal(μ, σ^2), etc.

Definition: For any RV X (discrete or continuous), the *cumulative distribution function (cdf)* is

$$F(x) \;\equiv\; P(X \le x) \;=\; \begin{cases} \sum_{y \le x} f(y) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{x} f(y)\,dy & \text{if } X \text{ is continuous} \end{cases}$$

Note that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. In addition, if X is continuous, then $\frac{d}{dx}F(x) = f(x)$.

Example: Flip 2 coins. Let X be the number of heads.

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \le x < 1 \\ 3/4 & \text{if } 1 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases} \qquad \Box$$

Example: Suppose $X \sim \mathrm{Exp}(\lambda)$ (i.e., X has the exponential distribution with parameter λ). Then $f(x) = \lambda e^{-\lambda x}$, $x \ge 0$, and the cdf is $F(x) = 1 - e^{-\lambda x}$, $x \ge 0$. $\Box$

---

§2.5 Assessment

1. Toss a 4-side die twice (you know, one of those goofy Dungeons and Dragons pyramid dice things). Assuming the die is numbered 1,2,3,4, what's the probability that the sum will equal 3?
   (a) 0
   (b) 1/2
   (c) 13/16
   (d) 1/8
2. TRUE or FALSE? $f(x) = 3e^{-x}$ for $x > 0$ is a legitimate probability density function.

2.6 Simulating Random Variables

We'll make a brief aside here to show how to simulate some very simple random variables.

Example: Consider a *discrete uniform (D.U.)* distribution on $\{1, 2, \ldots, n\}$, i.e., $X = i$ with probability $1/n$ for $i = 1, 2, \ldots, n$. (Think of this as an n-sided dice toss for you Dungeons and Dragons fans.)

If $U \sim \mathrm{Unif}(0, 1)$, we can obtain a D.U. random variate simply by setting $X = \lceil nU \rceil$, where $\lceil \cdot \rceil$ is the "ceiling" (or "round up") function.

For example, if $n = 10$ and we sample a Unif(0,1) random variable $U = 0.73$, then $X = \lceil 7.3 \rceil = 8$. $\Box$

Example: Here's how to simulate another discrete RV. Suppose that the pmf is

$$P(X=x) \;=\; \begin{cases} 0.25 & \text{if } x=-2 \\ 0.10 & \text{if } x=3 \\ 0.65 & \text{if } x=4.2 \\ 0 & \text{otherwise} \end{cases}$$

We can't use a die toss to simulate this random variable. Instead, we'll use what's called the *inverse transform method*. To do so, construct the following table.

x	$f(x)$	$P(X \le x)$	Unif(0,1)'s
-2	0.25	0.25	$[0.00, 0.25]$
3	0.10	0.35	$(0.25, 0.35]$
4.2	0.65	1.00	$(0.35, 1.00)$

Sample $U \sim \text{Unif}(0,1)$, and choose the corresponding x-value, i.e., $X = F^{-1}(U)$, where $F^{-1}(\cdot)$ denotes the inverse of the cdf. For example, $U = 0.46$ means that $X = 4.2$. $\Box$

Now we'll use the inverse transform method to generate a continuous random variable. We'll talk about and prove the following result a little later....

Inverse Transform Theorem: If X is a continuous random variable with cdf $F(x)$, then the random variable $F(X) \sim \text{Unif}(0,1)$.

This suggests a way to generate realizations of the RV X. Simply set $F(X) = U \sim \text{Unif}(0,1)$ and solve for $X = F^{-1}(U)$.

Example: Suppose $X \sim \text{Exp}(\lambda)$. Then $F(x) = 1 - e^{-\lambda x}$ for $x > 0$. Set $F(X) = 1 - e^{-\lambda X} = U$. Now solve for X,

$$X \;=\; \frac{-1}{\lambda}\,\ell\text{n}(1-U) \;\sim\; \text{Exp}(\lambda).$$

(Don't worry that we used $\ell\text{n}(1-U)$ here but $\ell\text{n}(U)$ in §1.7. Why?) $\Box$

Example (Generating Uniforms): All of the above RV generation examples relied on our ability to generate *independent and identically distributed (iid)* Unif(0,1) RVs. For now, let's assume that we can at least generate *deterministic* *pseudo-random numbers (PRNs)* that *appear* to be iid Unif(0,1); see Chapter 6.

If you don't like programming, you can be a little lazy and use the Excel function `RAND()` or something similar to generate PRNs.

If you do like programming, here's an algorithm to generate reasonable PRNs, $U_1, U_2, \ldots$, from scratch. Start by picking a *seed* integer X_0, and iteratively calculate

$$X_i \;=\; 16807X_{i-1}\text{mod}(2^{31}-1), \quad i = 1, 2, \ldots.$$

Then set $U_i = X_i/(2^{31}-1)$, $i = 1, 2, \ldots$.

And here's an easy, old-fashioned FORTRAN implementation of the above algorithm from Bratley et al. (1987).

```
FUNCTION UNIF(IX)
K1 = IX/127773     (this division truncates, e.g., 5/3 = 1.)
IX = 16807*(IX - K1*127773) - K1*2836     (update seed)
IF(IX.LT.0)IX = IX + 2147483647
UNIF = IX * 4.656612875E-10
RETURN
END
```

In the above function, we input a positive integer `IX` and the function returns the PRN `UNIF`, as well as an updated `IX` that we can use again. □

Some Exercises: In the following, we'll assume that you can use Excel (or whatever) to simulate independent Unif(0,1) RVs. (We'll review independence in a little while.)

1. Make a histogram of $X_i = -\ell\text{n}(U_i)$, for $i = 1, 2, \ldots, 10000$, where the U_i's are independent Unif(0,1) RVs. What kind of distribution does it look like?

2. Suppose X_i and Y_i are independent Unif(0,1) RVs, $i = 1, 2, \ldots, 10000$. Let $Z_i = \sqrt{-2\ell\text{n}(X_i)}\,\sin(2\pi Y_i)$, and make a histogram of the Z_i's based on the 10000 replications.

3. Suppose X_i and Y_i are independent Unif(0,1) RVs, $i = 1, 2, \ldots, 10000$. Let $Z_i = X_i/(X_i - Y_i)$, and make a histogram of the Z_i's based on the 10000 replications. This may be somewhat interesting. It's possible to derive the distribution analytically, but it takes a lot of work.

---

§2.6 Assessment

1. Suppose X is a continuous random variable with cumulative distribution function $F(x)$. What is the distribution of the nasty random variable $F(X)$?

   (a) Normal

   (b) Unif(0,1)

   (c) Exponential

   (d) Weibull

2. Suppose that U is a Unif(0,1) random variable. Name the distribution of $X = -\ell\mathrm{n}(1-U)$.

   (a) Normal

   (b) Unif(0,1)

   (c) Exponential

   (d) Weibull

3. BONUS: TRUE or FALSE? $2^{31}-1$ is a prime number.

2.7 Great Expectations

Definition: The *expected value* (or *mean*) of a RV X is

$$\mathrm{E}[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x)\,dx & \text{if } X \text{ is continuous} \end{cases} = \int_{\mathbb{R}} x\,dF(x),$$

where the last expression is standard notation often used to save space.

Example: Suppose that $X \sim \mathrm{Bernoulli}(p)$. Then

$$X = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p\ (=q) \end{cases}$$

and we have $\mathrm{E}[X] = \sum_x x f(x) = p$. $\Box$

Example: Suppose that $X \sim \mathrm{Uniform}(a,b)$. Then

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

and we have $\mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\,dx = (a+b)/2$. $\Box$

Example: Suppose that $X \sim \mathrm{Exponential}(\lambda)$. Then

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

and we have (after integration by parts and L'Hôspital's Rule)

$$\mathrm{E}[X] = \int_{\mathbb{R}} x f(x)\,dx = \int_0^\infty x\lambda e^{-\lambda x}\,dx = \frac{1}{\lambda}. \quad \Box$$

Definition / Theorem: Here is a nice tool that is informally known as the *Law of the Unconscious Statistician (LOTUS)*. It's actually a theorem though some foolhardy folks think it's a definition. In any case, suppose that $h(X)$ is some function of the RV X. Then

$$\mathrm{E}[h(X)] = \begin{cases} \sum_x h(x)f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} h(x)f(x)\,dx & \text{if } X \text{ is continuous} \end{cases}.$$

The function $h(X)$ can be anything "nice", e.g., $h(X) = X^2$ or $1/X$ or $\sin(X)$ or $\ell\mathrm{n}(X)$.

Example: Suppose X is the following discrete RV:

x	2	3	4
$f(x)$	0.3	0.6	0.1

Then $\mathrm{E}[X^3] = \sum_x x^3 f(x) = 8(0.3) + 27(0.6) + 64(0.1) = 25.$ $\Box$

Example: Suppose $X \sim \mathrm{Unif}(0, 2)$. Then

$$\mathrm{E}[X^n] = \int_{\mathbb{R}} x^n f(x)\,dx = 2^n/(n+1). \quad \Box$$

Definition: $\mathrm{E}[X^n]$ is the nth *moment* of X.

$\mathrm{E}[(X - \mathrm{E}[X])^n]$ is the nth *central moment* of X.

$\mathrm{Var}(X) \equiv \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$ is the *variance* of X. (That second equality is a theorem.)

The *standard deviation* of X is $\sqrt{\mathrm{Var}(X)}$.

Example: Suppose $X \sim \mathrm{Bern}(p)$. Recall that $\mathrm{E}[X] = p$. Then

$$\mathrm{E}[X^2] = \sum_x x^2 f(x) = p \quad \text{and}$$

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = p(1-p). \quad \Box$$

Example: Suppose $X \sim \mathrm{Exp}(\lambda)$. By LOTUS,

$$\mathrm{E}[X^n] = \int_0^\infty x^n \lambda e^{-\lambda x}\,dx = n!/\lambda^n.$$

$$\text{Var}(X) \;=\; \text{E}[X^2] - (\text{E}[X])^2 \;=\; \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \;=\; 1/\lambda^2. \quad \Box$$

Theorem: $\text{E}[aX+b] = a\text{E}[X]+b$ and $\text{Var}(aX+b) = a^2\text{Var}(X)$.

Example: If $X \sim \text{Exp}(3)$, then

$$\text{E}[-2X+7] \;=\; -2\text{E}[X]+7 \;=\; -\frac{2}{3}+7.$$

$$\text{Var}(-2X+7) \;=\; (-2)^2\text{Var}(X) \;=\; \frac{4}{9}. \quad \Box$$

Definition: $M_X(t) \equiv \text{E}[e^{tX}]$ is the *moment generating function (mgf)* of the RV X. ($M_X(t)$ is a function of t, *not* of X!)

Example: $X \sim \text{Bern}(p)$. Then

$$M_X(t) \;=\; \text{E}[e^{tX}] \;=\; \sum_x e^{tx} f(x) \;=\; e^{t\cdot 1}p + e^{t\cdot 0}q \;=\; pe^t + q. \quad \Box$$

Example: $X \sim \text{Exp}(\lambda)$. Then

$$M_X(t) \;=\; \int_{\Re} e^{tx} f(x)\,dx \;=\; \lambda \int_0^\infty e^{(t-\lambda)x}\,dx \;=\; \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \Box$$

Theorem: So why is it called the *moment* generating function? Under certain technical conditions,

$$\text{E}[X^k] \;=\; \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}, \quad k = 1, 2, \ldots.$$

Thus, you can *generate* the moments of X from the mgf — hence the moniker!

Example: $X \sim \text{Exp}(\lambda)$. Then $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. So

$$\text{E}[X] \;=\; \frac{d}{dt} M_X(t)\bigg|_{t=0} \;=\; \frac{\lambda}{(\lambda - t)^2}\bigg|_{t=0} \;=\; 1/\lambda.$$

Further,

$$\text{E}[X^2] \;=\; \frac{d^2}{dt^2} M_X(t)\bigg|_{t=0} \;=\; \frac{2\lambda}{(\lambda - t)^3}\bigg|_{t=0} \;=\; 2/\lambda^2.$$

Thus,

$$\text{Var}(X) \;=\; \text{E}[X^2] - (\text{E}[X])^2 \;=\; \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \;=\; 1/\lambda^2. \quad \Box$$

Moment generating functions have many other important uses, some of which we'll talk about later in this course.

---

§2.7 Assessment

1. Suppose X is a continuous random variable with p.d.f. $f(x) = 3x^2$ for $0 \le x \le 1$. Find $\mathrm{E}[X]$.
   (a) 2/3
   (b) 3/4
   (c) 1
   (d) 3/2
2. Suppose X is a continuous random variable with p.d.f. $f(x) = 3x^2$ for $0 \le x \le 1$. Find $\mathrm{E}[1/X]$.
   (a) 2/3
   (b) 3/4
   (c) 1
   (d) 3/2
3. The abbreviation "mgf" stands for...
   (a) Mom's generating function
   (b) Mega-gigundo function
   (c) Most-glorious function
   (d) Moment generating function

2.8 Functions of a Random Variable

Problem: Suppose we have a RV X with pmf / pdf $f(x)$. Let $Y = h(X)$. Find $g(y)$, the pmf / pdf of Y.

Examples (take my word for it for now):

If $X \sim \mathrm{Nor}(0, 1)$, then $Y = X^2 \sim \chi^2(1)$.

If $U \sim \mathrm{Unif}(0, 1)$, then $Y = -\frac{1}{\lambda}\ell\mathrm{n}(U) \sim \mathrm{Exp}(\lambda)$.

Discrete Example: Let X denote the number of H's from two coin tosses. We want the pmf for $Y = X^3 - X$.

x	0	1	2
$f(x)$	1/4	1/2	1/4
$y = x^3 - x$	0	0	6

This implies that

$$g(0) \;=\; P(Y=0) \;=\; P(X=0 \text{ or } 1) \;=\; 3/4 \quad \text{and} \quad g(6) \;=\; P(Y=6) \;=\; 1/4.$$

In other words,

$$g(y) \;=\; \begin{cases} 3/4 & \text{if } y=0 \\ 1/4 & \text{if } y=6 \end{cases}. \quad \Box$$

Continuous Example: Suppose X has pdf $f(x) = |x|$, $-1 \le x \le 1$. Find the pdf of $Y = X^2$. The game plan will be to find the cdf of Y and then differentiate to get its pdf. First of all, the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} |x|\,dx \;=\; y, \quad 0 < y < 1. \end{aligned}$$

Thus, the pdf of Y is $g(y) = G'(y) = 1$, $0 < y < 1$, indicating that $Y \sim \text{Unif}(0,1)$. $\Box$

Inverse Transform Theorem (ITT): Suppose X is a continuous random variable having cdf $F(x)$. Then, amazingly, $F(X) \sim \text{Unif}(0,1)$.

Proof: Let $Y = F(X)$. Then the cdf of Y is

$$P(Y \le y) \;=\; P(F(X) \le y) \;=\; P(X \le F^{-1}(y)) \;=\; F(F^{-1}(y)) \;=\; y,$$

which is the cdf of the Unif(0,1). $\Box$

This result is of fundamental importance when it comes to generating random variates during a simulation.

Example: Here is how we generate exponential RVs. Suppose $X \sim \text{Exp}(\lambda)$, with cdf $F(x) = 1 - e^{-\lambda x}$ for $x \ge 0$. So the ITT implies that

$$F(X) \;=\; 1 - e^{-\lambda X} \;\sim\; \text{Unif}(0,1).$$

Now let $U \sim \text{Unif}(0,1)$ and set $F(X) = U$. Then we have

$$X \;=\; \frac{-1}{\lambda}\,\ell\text{n}(1-U) \;\sim\; \text{Exp}(\lambda).$$

For instance, if $\lambda = 2$ and $U = 0.27$, then $X = 0.157$ is an Exp(2) realization. $\Box$

Exercise: Suppose that X has the Weibull distribution with cdf

$$F(x) \; = \; 1 - e^{-(\lambda x)^\beta}, \quad x > 0.$$

If you set $F(X) = U$ and solve for X, show that you get

$$X \; = \; \frac{1}{\lambda}\left[-\ell\text{n}(1-U)\right]^{1/\beta}.$$

Now pick your favorite λ and β, and use this result to generate values of X. In fact, make a histogram of your X values. Are there any interesting values of λ and β you could've chosen?

Bonus Theorem: Here's another way to get the pdf of $Y = h(X)$ for some nice continuous function $h(\cdot)$. The cdf of Y is

$$F_Y(y) \; = \; P(Y \le y) \; = \; P(h(X) \le y) \; = \; P(X \le h^{-1}(y)).$$

By the chain rule (and since a pdf must be ≥ 0), the pdf of Y is

$$f_Y(y) \; = \; \frac{d}{dy}F_Y(y) \; = \; f_X(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right|.$$

And now, here's how to prove LOTUS!

$$\begin{aligned} \mathrm{E}[Y] \quad &= \quad \int_{\mathbb{R}} y f_Y(y)\,dy \; = \; \int_{\mathbb{R}} y f_X(h^{-1}(y))\left|\frac{d}{dy}h^{-1}(y)\right|\,dy \\ \text{``=''} \quad & \quad \int_{\mathbb{R}} y f_X(h^{-1}(y))\,dh^{-1}(y) \; = \; \int_{\mathbb{R}} h(x) f_X(x)\,dx. \quad \Box \end{aligned}$$

---

§2.8 Assessment

1. Suppose X is the result of a 4-sided die toss having sides numbered $-2, -1, 1, 2$. Find the probability mass function of $Y = X^2$.

   (a) $P(Y = 1) = P(Y = 4) = 1/2$
   (b) $P(Y = 1) = P(Y = 2) = 1/2$
   (c) $P(Y = -2) = P(Y = -1) = P(Y = 1) = P(Y = 2) = 1/4$
   (d) $P(Y = -4) = P(Y = -1) = P(Y = 1) = P(Y = 4) = 1/4$

2. Suppose X is a continuous random variable with pdf $f(x) = 2x$ for $0 < x < 1$. Find the pdf of $Y = \sqrt{X}$.

   (a) $3y^2$, for $0 < y < 1$
   (b) x^2, for $0 < x < 1$
   (c) $4y^3$, for $-1 < y < 1$
   (d) $4y^3$, for $0 < y < 1$

2.9 Jointly Distributed Random Variables

Simulations frequently make use of and/or generate *correlated* data. So we'll need to study how random variables interact together — as opposed to operating separately in isolation. To keep things as simple as possible for this review, let's merely consider two interacting RVs — think height and weight.

Definition: The *joint cdf* of X and Y is

$$F(x,y) \;\equiv\; P(X \le x, Y \le y), \quad \text{for all } x, y.$$

Remark: The *marginal cdf* of X is $F_X(x) = F(x,\infty)$. (We use the X subscript to remind us that it's just the cdf of X all by itself.) Similarly, the marginal cdf of Y is $F_Y(y) = F(\infty, y)$.

Definition: If X and Y are discrete, then the *joint pmf* of X and Y is $f(x,y) \equiv P(X = x, Y = y)$. Note that $\sum_x \sum_y f(x,y) = 1$.

Remark: The *marginal pmf* of X is

$$f_X(x) \;=\; P(X = x) \;=\; \sum_y f(x,y).$$

The marginal pmf of Y is

$$f_Y(y) \;=\; P(Y = y) \;=\; \sum_x f(x,y).$$

Example: The following table gives the joint pmf $f(x,y)$, along with the accompanying marginals. In this example, X is the distribution of a University of Georgia student's grade point average, and Y is the student's IQ.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.3	0.2	0.1	0.6
$Y=60$	0.1	0.2	0.1	0.4
$f_X(x)$	0.4	0.4	0.2	1

Definition: If X and Y are continuous, then the *joint pdf* of X and Y is $f(x,y) \equiv \frac{\partial^2}{\partial x \partial y} F(x,y)$. Note that $\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\,dx\,dy = 1$.

Remark: The *marginal pdf's* of X and Y are

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \quad \text{and} \quad f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx.$$

Example: Suppose the joint pdf is

$$f(x, y) = \frac{21}{4}x^2 y, \quad x^2 \le y \le 1.$$

Then the marginal pdf's are:

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2 y\,dy = \frac{21}{8}x^2(1 - x^4), \quad -1 \le x \le 1$$

and

$$f_Y(y) = \int_{\mathbb{R}} f(x, y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2 y\,dx = \frac{7}{2}y^{5/2}, \quad 0 \le y \le 1. \quad \Box$$

Definition: X and Y are *independent* RVs if

$$f(x, y) = f_X(x) f_Y(y) \quad \text{for all } x, y.$$

Theorem: X and Y are independent if you can write their joint pdf as $f(x, y) = a(x)b(y)$ for some functions $a(x)$ and $b(y)$, and x and y don't have *"funny limits"* (their domains do not depend on each other).

This is a relatively easy way of determining whether or not X and Y are independent, without having to explicitly calculate $f_X(x)$ and $f_Y(y)$.

Examples: If $f(x, y) = cxy$ for $0 \le x \le 2$, $0 \le y \le 3$, then X and Y are independent.

If $f(x, y) = \frac{21}{4}x^2 y$ for $x^2 \le y \le 1$, then X and Y are *not* independent.

If $f(x, y) = c/(x + y)$ for $1 \le x \le 2$, $1 \le y \le 3$, then X and Y are *not* independent. $\Box$

Definition: The *conditional pmf / pdf* of Y given $X = x$ is $f(y|x) \equiv f(x, y)/f_X(x)$ (assuming $f_X(x) > 0$).

This is a legit pmf / pdf. For example, in the continuous case, $\int_{\mathbb{R}} f(y|x)\,dy = 1$, for any x.

Example: Suppose $f(x, y) = \frac{21}{4}x^2 y$ for $x^2 \le y \le 1$. Then

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1 - x^4)} = \frac{2y}{1 - x^4}, \quad x^2 \le y \le 1. \quad \Box$$

Theorem: If X and Y are independent, then $f(y|x) = f_Y(y)$ for all x, y.

Proof: By the definition of conditional probability and independence,

$$f(y|x) \;=\; \frac{f(x,y)}{f_X(x)} \;=\; \frac{f_X(x)f_Y(y)}{f_X(x)}. \quad \square$$

---

§2.9 Assessment

1. The following table gives the joint pmf $f(x,y)$ of two random variables X (the GPA of a University of Georgia student) and Y (his IQ).

$f(x,y)$	$X=2$	$X=3$	$X=4$
$Y=40$	0.4	0.1	0.1
$Y=50$	0.1	0.2	0.1

   What's the probability that a random UGA student has an IQ of 50?

   (a) 0
   (b) 0.1
   (c) 0.4
   (d) 0.6

2. YES or NO? Suppose X and Y have joint p.d.f. $f(x,y) = 6xy^2$ for $0 < x < 1$ and $0 < y < 1$. Are X and Y independent?

2.10 Conditional Expectation

Definition: The *conditional expectation* of Y given $X = x$ is

$$\mathrm{E}[Y|X=x] \;\equiv\; \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x)\, dy & \text{continuous} \end{cases}$$

Example: The expected weight of a person who is 7 feet tall ($\mathrm{E}[Y|X=7]$) will probably be greater than that of a random person from the entire population ($\mathrm{E}[Y]$).

Old Continuous Example: $f(x,y) = \frac{21}{4}x^2 y$, if $x^2 \le y \le 1$. Then

$$\mathrm{E}[Y|x] \;=\; \int_{\mathbb{R}} y f(y|x)\, dy \;=\; \int_{x^2}^{1} \frac{2y^2}{1-x^4}\, dy \;=\; \frac{2}{3}\cdot\frac{1-x^6}{1-x^4}. \quad \square$$

Double Expectations Theorem: $\mathrm{E}[\mathrm{E}(Y|X)] = \mathrm{E}[Y]$.

Proof: Let's just do the continuous case. By the Unconscious Statistician,

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(Y|X)] &= \int_{\mathbb{R}} \mathrm{E}(Y|x) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y f(y|x)\,dy \right) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y|x) f_X(x)\,dx\,dy \\
&= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}} y f_Y(y)\,dy \;=\; \mathrm{E}[Y]. \quad \Box
\end{aligned}$$

Old Example: Suppose (yet again) $f(x,y) = \frac{21}{4}x^2 y$, if $x^2 \le y \le 1$. By previous examples, we know $f_X(x)$, $f_Y(y)$, and $\mathrm{E}[Y|x]$. Find $\mathrm{E}[Y]$.

Solution #1 (old, boring way):

$$\mathrm{E}[Y] = \int_{\mathbb{R}} y f_Y(y)\,dy = \int_0^1 \frac{7}{2} y^{7/2}\,dy = \frac{7}{9}.$$

Solution #2 (new, exciting way):

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y|X)] = \int_{\mathbb{R}} \mathrm{E}(Y|x) f_X(x)\,dx \\
&= \int_{-1}^{1} \left(\frac{2}{3} \cdot \frac{1-x^6}{1-x^4} \right) \left(\frac{21}{8} x^2 (1-x^4) \right) dx = \frac{7}{9}.
\end{aligned}$$

Notice that both answers are the same (good)! $\Box$

Example: We'll discuss a cutesy way to calculate the mean of the Geometric distribution. Let $Y \sim \mathrm{Geom}(p)$, e.g., Y could be the number of coin flips before H appears, where $P(\mathrm{H}) = p$. From Baby Probability class, we know that the pmf of Y is $f_Y(y) = P(Y = y) = q^{y-1}p$, for $y = 1, 2, \ldots$, where $q = 1 - p$. The old-fashioned way to calculate the mean is:

$$\mathrm{E}[Y] = \sum_y y f_Y(y) = \sum_{y=1}^{\infty} y q^{y-1} p = 1/p,$$

where the last step follows because I tell you so. $\Box$

But if you are not quite willing to believe me, let's use double expectation to do what's called a *"standard one-step conditioning argument."* To this end, define $X = 1$ if the first flip is H; and $X = 0$ otherwise. Based on the result X of the first step, we have

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y|X)] \;=\; \sum_x \mathrm{E}(Y|x) f_X(x) \\ &= \mathrm{E}(Y|X=0)P(X=0) + \mathrm{E}(Y|X=1)P(X=1) \\ &= (1+\mathrm{E}[Y])(1-p) + 1(p). \quad \text{(why?)} \end{aligned}$$

Solving, we get $\mathrm{E}[Y] = 1/p$ again! $\Box$

Computing Probabilities by Conditioning

You can use conditioning arguments to calculate all sorts of good stuff. We'll look at a couple of well-known applications in what follows. In general, let A be some event, and define the RV $Y = 1$ if A occurs; and $Y = 0$ otherwise. Then

$$\mathrm{E}[Y] \;=\; \sum_y y f_Y(y) \;=\; P(Y=1) \;=\; P(A).$$

Similarly, for any RV X, we have

$$\mathrm{E}[Y|X=x] \;=\; \sum_y y f_Y(y|x) \;=\; P(Y=1|X=x) \;=\; P(A|X=x).$$

Thus,

$$\begin{aligned} P(A) &= \mathrm{E}[Y] \;=\; \mathrm{E}[\mathrm{E}(Y|X)] \\ &= \int_{\mathbb{R}} \mathrm{E}[Y|X=x]\, dF_X(x) \\ &= \int_{\mathbb{R}} P(A|X=x)\, dF_X(x). \end{aligned}$$

Example/Theorem: If X and Y are independent cts RVs, then

$$P(Y < X) \;=\; \int_{\mathbb{R}} P(Y < x) f_X(x)\, dx.$$

Proof: Follows from above result if we let the event $A = \{Y < X\}$. $\Box$

Example: If $X \sim \mathrm{Exp}(\mu)$ and $Y \sim \mathrm{Exp}(\lambda)$ are indep RVs, then

$$\begin{aligned} P(Y < X) &= \int_{\mathbb{R}} P(Y < x) f_X(x)\, dx \\ &= \int_0^\infty (1 - e^{-\lambda x}) \mu e^{-\mu x}\, dx \\ &= \frac{\lambda}{\lambda + \mu}. \quad \Box \end{aligned}$$

Variance Decomposition Theorem: The following result is sometimes used in what are called *variance reduction techniques*. (See Chapter 10 for some insight.)

$$\mathrm{Var}(Y) \;=\; \mathrm{E}\left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathrm{E}(Y|X)\right]$$

Proof (from Ross): By definition of variance and double expectation,

$$\begin{aligned}\mathrm{E}\left[\mathrm{Var}(Y|X)\right] &= \mathrm{E}\left[\mathrm{E}(Y^2|X) - \{\mathrm{E}(Y|X)\}^2\right] \\ &= \mathrm{E}(Y^2) - \mathrm{E}\left[\{\mathrm{E}(Y|X)\}^2\right].\end{aligned}$$

Similarly,

$$\begin{aligned}\mathrm{Var}\left[\mathrm{E}(Y|X)\right] &= \mathrm{E}\left[\{\mathrm{E}(Y|X)\}^2\right] - \{\mathrm{E}\left[\mathrm{E}(Y|X)\right]\}^2 \\ &= \mathrm{E}\left[\{\mathrm{E}(Y|X)\}^2\right] - \{\mathrm{E}(Y)\}^2.\end{aligned}$$

Thus,

$$\mathrm{E}\left[\mathrm{Var}(Y|X)\right] + \mathrm{Var}\left[\mathrm{E}(Y|X)\right] \;=\; \mathrm{E}(Y^2) - \{\mathrm{E}(Y)\}^2 \;=\; \mathrm{Var}(Y). \quad \Box$$

---

§2.10 Assessment

This is a difficult lesson, but don't panic. Let's at least see how well you understand the basics.

1. The following table gives the joint pmf $f(x, y)$ of two random variables X (the GPA of a University of Georgia student) and Y (his IQ).

$f(x,y)$	$X = 2$	$X = 3$	$X = 4$	$f_Y(y)$
$Y = 40$	0.4	0.1	0.1	0.6
$Y = 50$	0.1	0.2	0.1	0.4
$f_X(x)$	0.5	0.3	0.2	1

What's the conditional probability that a random UGA student has an IQ of 50 given that his GPA = 2?

(a) 0

(b) 0.2

(c) 0.4

(d) 0.6

2. Which are legitimate expressions for $\mathrm{E}[Y|x]$? (There may be more than one correct answer.)

(a) $\mathrm{E}[Y]$
(b) $\int_{\mathbb{R}} y f(y|x)\, dy$
(c) $\int_{\mathbb{R}} y f(y|x)\, dx$
(d) $\frac{1}{f_X(x)} \int_{\mathbb{R}} y f(x,y)\, dy$

2.11 Covariance and Correlation

We start with 2-D LOTUS, which will be useful for subsequently calculating quantities such as covariance and correlation, as well as for proving some of the following results (which we will not actually do here).

"Definition" (two-dimensional LOTUS): Suppose that $h(X,Y)$ is some function of the RVs X and Y. Then

$$\mathrm{E}[h(X,Y)] \;=\; \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & \text{if } (X,Y) \text{ is discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f(x,y)\, dx\, dy & \text{if } (X,Y) \text{ is continuous} \end{cases}$$

Now, some preliminary results on what happens when we add up random variables.

Theorem: Whether or not X and Y are independent, we have $\mathrm{E}[X+Y] = \mathrm{E}[X]+\mathrm{E}[Y]$.

Theorem: If X and Y are *independent*, then $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. (Stay tuned for dependent case.)

Definition: $X_1, X_2, \ldots, X_n$ form a *random sample* from $f(x)$ if (i) $X_1, X_2, \ldots, X_n$ are independent, and (ii) each X_i has the same pmf / pdf $f(x)$.

Notation: $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$. (The term "$\overset{\text{iid}}{\sim}$" reads *independent and identically distributed.*)

Example: Suppose $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} f(x)$ and define the *sample mean* $\bar{X}_n \equiv \sum_{i=1}^n X_i/n$. Since the X_i's are iid, we have

$$\mathrm{E}[\bar{X}_n] \;=\; \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] \;=\; \frac{1}{n}\sum_{i=1}^n \mathrm{E}[X_i] \;=\; \mathrm{E}[X_1]$$

and

$$\mathrm{Var}(\bar{X}_n) \;=\; \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^n X_i\right) \;=\; \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}(X_i) \;=\; \frac{\mathrm{Var}(X_1)}{n}.$$

Thus, the expected value of the sample mean doesn't change, but the variance *decreases* as n increases. (More consequences from this later.) $\square$

However, not all RVs are independent...

Definition and Theorem: The *covariance* between X and Y is

$$\text{Cov}(X,Y) \;\equiv\; \text{E}[(X-\text{E}[X])(Y-\text{E}[Y])] \;=\; \text{E}[XY]-\text{E}[X]\text{E}[Y].$$

Note that $\text{Var}(X) = \text{Cov}(X,X)$.

Covariance measures the direction of the relationship between X and Y. So IQ and GPA would likely have a positive covariance, while bonds and stock prices (which tend to move in opposite directions) would have a negative covariance.

Theorem: If X and Y are independent RVs, then $\text{Cov}(X,Y) = 0$.

Remark: But $\text{Cov}(X,Y) = 0$ doesn't mean X and Y are independent!

Example: Suppose $X \sim \text{Unif}(-1,1)$ and $Y = X^2$. Then X and Y are clearly dependent. However,

$$\text{Cov}(X,Y) \;=\; \text{E}[X^3]-\text{E}[X]\text{E}[X^2] \;=\; \text{E}[X^3] \;=\; \int_{-1}^{1}\frac{x^3}{2}\,dx \;=\; 0. \quad \Box$$

Theorem: $\text{Cov}(aX,bY) = ab\,\text{Cov}(X,Y)$.

Theorem: Whether or not X and Y are independent,

$$\text{Var}(X+Y) \;=\; \text{Var}(X)+\text{Var}(Y)+2\text{Cov}(X,Y)$$

and

$$\text{Var}(X-Y) \;=\; \text{Var}(X)+\text{Var}(Y)-2\text{Cov}(X,Y).$$

Definition: The *correlation* between X and Y is

$$\rho \;\equiv\; \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Theorem: $-1 \le \rho \le 1$.

A correlation $\rho \doteq 1$ is regarded as "high" (e.g., GRE and SAT scores for the same person); $\rho \doteq -1$ is "high negative" (e.g., snowfall vs. temperature); and $\rho \doteq 0$ is "low" (e.g., IBM stock price and temperature on Mars — or so they want you to believe!)

Example: Consider the following joint pmf.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.00	0.20	0.10	0.3
$Y=50$	0.15	0.10	0.05	0.3
$Y=60$	0.30	0.00	0.10	0.4
$f_X(x)$	0.45	0.30	0.25	1

$\mathrm{E}[X]=2.8$, $\mathrm{Var}(X)=0.66$, $\mathrm{E}[Y]=51$, $\mathrm{Var}(Y)=69$,

$$\mathrm{E}[XY] \;=\; \sum_x \sum_y xyf(x,y) \;=\; 140,$$

and

$$\rho \;=\; \frac{\mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y]}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \;=\; -0.415. \quad \Box$$

Portfolio Example: Consider two assets, S_1 and S_2, with expected returns $\mathrm{E}[S_1]=\mu_1$ and $\mathrm{E}[S_2]=\mu_2$, and variabilities $\mathrm{Var}(S_1)=\sigma_1^2$, $\mathrm{Var}(S_2)=\sigma_2^2$, and $\mathrm{Cov}(S_1,S_2)=\sigma_{12}$. Define a *portfolio* $P=wS_1+(1-w)S_2$, where $w\in[0,1]$. Then

$$\mathrm{E}[P] \;=\; w\mu_1+(1-w)\mu_2$$

$$\mathrm{Var}(P) \;=\; w^2\sigma_1^2+(1-w)^2\sigma_2^2+2w(1-w)\sigma_{12}.$$

Setting $\frac{d}{dw}\mathrm{Var}(P)=0$, we obtain the critical point that (hopefully) minimizes the variance of the portfolio,

$$w \;=\; \frac{\sigma_2^2-\sigma_{12}}{\sigma_1^2+\sigma_2^2-2\sigma_{12}}. \quad \Box$$

Portfolio Exercise: Suppose $\mathrm{E}[S_1]=0.2$, $\mathrm{E}[S_2]=0.1$, $\mathrm{Var}(S_1)=0.2$, $\mathrm{Var}(S_2)=0.4$, and $\mathrm{Cov}(S_1,S_2)=-0.1$.

- What value of w maximizes the expected return of the portfolio?
- What value of w minimizes the variance? (Note the negative covariance I've introduced into the picture.)
- Let's talk trade-offs.

---

§2.11 Assessment

1. If X and Y both have mean 0 and variance 4, and $\mathrm{Cov}(X,Y)=1$, find the correlation between X and Y.

   (a) 0

(b) 1

(c) 1/2

(d) 1/4

2. If X and Y both have mean 0 and variance 4, and $\text{Cov}(X, Y) = 1$, find $\text{Var}(X + Y)$.

   (a) 8

   (b) 9

   (c) 10

   (d) 11

3. TRUE or FALSE? If X and Y are uncorrelated, then they're independent.

2.12 Some Probability Distributions

In this section, we'll meet some old distribution friends and list out some of their properties. First of all, we'll start out with some *discrete distributions*.

$X \sim$ Bernoulli(p). The Bernoulli takes on two values — 1 and 0 — informally representing "success" (S) and "failure" (F). In a simulation project, for instance, a "success" might represent the event that a simulated order was delivered on time. In any case, the pmf, expected value, variance, and mgf are as follows.

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p \ (= q) & \text{if } x = 0 \end{cases},$$

$\text{E}[X] = p$, $\text{Var}(X) = pq$, and $M_X(t) = pe^t + q$.

$Y \sim$ Binomial(n, p). If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$ (i.e., *Bernoulli(p) trials*), then $Y = \sum_{i=1}^n X_i \sim \text{Bin}(n, p)$. This RV can be regarded as the number of successes out of n Bern(p) trials. For instance, how many times does a "5" come up if we toss a standard die 20 times?

$$f(y) = \binom{n}{y} p^y q^{n-y}, \quad y = 0, 1, \ldots, n,$$

$\text{E}[Y] = np$, $\text{Var}(Y) = npq$, and $M_Y(t) = (pe^t + q)^n$.

$X \sim$ Geometric(p) is the number of Bern(p) trials until a success occurs. For example, "FFFS" implies that $X = 4$. The Geometric distribution also comes into play when carrying out certain RV generation methods (see Chapter 7). Its pmf and moments are:

$$f(x) = q^{x-1}p, \quad x = 1, 2, \ldots,$$

$\mathrm{E}[X] = 1/p$, $\mathrm{Var}(X) = q/p^2$, and $M_X(t) = pe^t/(1-qe^t)$.

$Y \sim \mathrm{NegBin}(r, p)$ is the sum of r iid Geom(p) RVs, i.e., the time until the rth success occurs. For example, "FFFSSFS" implies that $\mathrm{NegBin}(3, p) = 7$. Here we have

$$f(y) \;=\; \binom{y-1}{r-1} q^{y-r} p^r, \quad y = r, r+1, \ldots,$$

$\mathrm{E}[Y] = r/p$, $\mathrm{Var}(Y) = qr/p^2$, and $M_Y(t) = \left[pe^t/(1-qe^t)\right]^r$.

$X \sim \mathrm{Poisson}(\lambda)$. The Poisson distribution is ubiquitous in simulations, especially when dealing with any type of "arrival" process. For instance, how many customers arrive at a service station during a certain time period? How many accidents occur at the manufacturing plant in a given month? How many raisins will you find in 3 cookies? Before giving details on its pmf and moments, let's see where the distribution comes from.

Definition: A *counting process* $N(t)$ tallies the number of "arrivals" observed in $[0, t]$. A *Poisson process* is a special counting process satisfying the following.

i. Arrivals occur one-at-a-time at rate λ (e.g., $\lambda = 4$ customers / hr). Thus, customer arrivals at McBurger Queen are not quite Poisson because people arrive in *groups*. Perhaps the groups themselves, however, might adhere to this assumption.

ii. *Independent increments*, i.e., the numbers of arrivals in disjoint time intervals are independent. In other words, the number of arrivals between 3:00 am and 5:00 am would be completely independent of the number of arrivals between 11:00 am and 4:00 pm.

iii. *Stationary increments*, i.e., the distribution of the number of arrivals during the time interval $[s, s+t]$ only depends on the length t. The number of neutrinos hitting a detector probably isn't dependent on the time of day and so would satisfy this condition. But the number of arrivals to a restaurant is clearly time dependent (e.g., more arrivals during breakfast, but very few between 3:00 pm and 4:00 pm).

$X \sim \mathrm{Pois}(\lambda)$ is *defined* as the number of arrivals that a Poisson process experiences in one time unit, i.e., $N(1)$. The pmf and various moments are:

$$f(x) \;=\; \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots,$$

$\mathrm{E}[X] = \lambda = \mathrm{Var}(X)$, and $M_X(t) = e^{\lambda(e^t - 1)}$.

At this point, we'll discuss some important *continuous distributions.*

$X \sim \text{Uniform}(a, b)$. The uniform is often used to model random quantities for which we have almost no information besides the lowest and highest possible values. For example, the SAT math scores of University of Georgia students are Unif(200,500). We have

$$f(x) = \frac{1}{b-a}, \quad \text{for } a \le x \le b,$$

with $\text{E}[X] = (a+b)/2$, $\text{Var}(X) = (b-a)^2/12$, and $M_X(t) = (e^{tb} - e^{ta})/(tb - ta)$.

$X \sim \text{Exponential}(\lambda)$. The exponential is useful for modeling customer interarrival times and machine breakdown times, among many other applications. Its pdf and friends are

$$f(x) = \lambda e^{-\lambda x}, \quad \text{for } x \ge 0,$$

$\text{E}[X] = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$, and $M_X(t) = \lambda/(\lambda - t)$ for $t < \lambda$.

Theorem: The exponential distribution has the *memoryless property*, i.e., for $s, t > 0$,

$$P(X > s+t | X > s) = P(X > t).$$

If a lightbulb's lifetime is exponential, then at any time, it is "as good as new" (and its eventual failure comes completely out of the blue).

Proof: Noting that the cdf of the Exp(λ) is $F(x) = 1 - e^{-\lambda x}$ for $x > 0$, we have

$$\begin{aligned} P(X > s+t | X > s) &= \frac{P(X > s+t \cap X > s)}{P(X > s)} \\ &= \frac{P(X > s+t)}{P(X > s)} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \\ &= e^{-\lambda t}. \quad \Box \end{aligned}$$

Example: Suppose $X \sim \text{Exp}(\lambda = 1/100)$. Then

$$P(X > 200 | X > 50) = P(X > 150) = e^{-\lambda t} = e^{-150/100}. \quad \Box$$

Remark: The exponential is the only continuous distribution having the memoryless property. It also turns out that the Geom(p) is the only discrete distribution that is memoryless.

Theorem: If $N(t)$ is a Poisson process with rate λ. Then the times between "arrivals" are iid Exp(λ). This theorem will come up a number of times in the subsequent chapters, and is actually a bit tricky to prove (so we'll be lazy and forgo that corvée here).

$X \sim \text{Gamma}(\alpha, \lambda)$. The gamma is a flexible distribution that generalizes the exponential and is widely used in reliability work. Its pdf is

$$f(x) \;=\; \frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x \ge 0,$$

where the gamma function is

$$\Gamma(\alpha) \;\equiv\; \int_0^\infty t^{\alpha-1} e^{-t}\, dt.$$

Moreover, $\text{E}[X] = \alpha/\lambda$, $\text{Var}(X) = \alpha/\lambda^2$, $M_X(t) = \big[\lambda/(\lambda - t)\big]^\alpha$ for $t < \lambda$.

Theorem: If $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$, then $Y \equiv \sum_{i=1}^n X_i \sim \text{Gamma}(n, \lambda)$.

The $\text{Gamma}(n, \lambda)$ is also called the $\text{Erlang}_n(\lambda)$. It has cdf

$$F_Y(y) \;=\; 1 - e^{-\lambda y} \sum_{j=0}^{n-1} \frac{(\lambda y)^j}{j!}, \quad y \ge 0.$$

$X \sim \text{Triangular}(a, b, c)$. This distribution is good for modeling things with limited data — a is the smallest possible value, b is the "most likely," and c is the largest. Its pdf is

$$f(x) \;=\; \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{if } a < x \le b \\ \frac{2(c-x)}{(c-b)(c-a)} & \text{if } b < x \le c \\ 0 & \text{otherwise} \end{cases}$$

with $\text{E}[X] = (a + b + c)/3$.

$X \sim \text{Beta}(a, b)$. The beta is a often used to model random variables that exist on a fixed interval; and the uniform and certain triangular distributions are special cases. The baseline version of the beta has pdf

$$f(x) \;=\; \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}, \quad \text{for } 0 \le x \le 1 \text{ and } a, b > 0.$$

It has mean and variance

$$\text{E}[X] \;=\; \frac{a}{a+b} \quad \text{and} \quad \text{Var}(X) \;=\; \frac{ab}{(a+b)^2(a+b+1)}.$$

$X \sim \text{Normal}(\mu, \sigma^2)$. The normal is far and away the most-important distribution out there. It's used to model everything from heights and weights to IQs to sizes of manufactured parts to generic sums and averages. It has pdf

$$f(x) \;=\; \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R},$$

as well as $\mathrm{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$, and $M_X(t) = \exp(\mu t + \frac{1}{2}\sigma^2 t^2)$.

Theorem: A linear transform of a normal is still normal with the obvious mean and variance. If $X \sim \mathrm{Nor}(\mu, \sigma^2)$, then $aX + b \sim \mathrm{Nor}(a\mu + b, a^2\sigma^2)$.

Corollary: If $X \sim \mathrm{Nor}(\mu, \sigma^2)$, then $Z \equiv \frac{X-\mu}{\sigma} \sim \mathrm{Nor}(0, 1)$, the *standard normal distribution*, with pdf $\phi(z) \equiv \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ and cdf $\Phi(z)$, which is tabled extensively or available in any reasonable software package. E.g., using Excel,

$$\Phi(1.96) \;=\; \texttt{NORM.S.DIST(1.96,1)} \;\doteq\; 0.975.$$

Theorem: Linear combinations of independent normals are still normal. In particular, if $X_1, X_2, \ldots, X_n$ are *independent* with $X_i \sim \mathrm{Nor}(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, then

$$\sum_{i=1}^n a_i X_i + b \;\sim\; \mathrm{Nor}\left(\sum_{i=1}^n a_i \mu_i + b,\, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

Baby Corollary: If X_1 and X_2 are *independent* with $X_i \sim \mathrm{Nor}(\mu_i, \sigma_i^2)$, $i = 1, 2$, then $X_1 + X_2 \sim \mathrm{Nor}(\mu_1 + \mu_2,\, \sigma_1^2 + \sigma_2^2)$.

Example: Suppose $X \sim \mathrm{Nor}(3, 4)$, $Y \sim \mathrm{Nor}(4, 6)$, and X and Y are independent. Find the distribution of $2X - 3Y + 1$. By the above theorem, $2X - 3Y + 1$ will be normal, so all that remains is to find the mean and variance. To this end,

$$\mathrm{E}[2X - 3Y + 1] \;=\; 2(3) - 3(4) + 1 \;=\; -5,$$

$$\mathrm{Var}(2X - 3Y + 1) \;=\; 2^2(4) + (-3)^2(6) \;=\; 70,$$

and so $2X - 3Y + 1 \sim \mathrm{Nor}(-5, 70)$. $\square$

---

§2.12 Assessment

1. If the number of accidents at a factory this year is a Pois(3) random variable, find the probability that we'll have exactly 2 accidents this year.

   (a) 0.224

   (b) 2/3

   (c) 0.112

   (d) 1/3

2. If X has a normal distribution with mean 2 and variance 9, find the probability that $X \leq 5$.

(a) 0.5

(b) 0.1587

(c) 0.975

(d) 0.8413

2.13 Limit Theorems

Corollary (of normal linear combinations theorem): If $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Nor}(\mu, \sigma^2)$, then the *sample mean*

$$\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^{n} X_i \sim \text{Nor}(\mu, \sigma^2/n).$$

This is a special case of the *Law of Large Numbers*, which says that $\bar{X}_n$ approximates μ well as n becomes large. In fact, the random variable $\bar{X}_n$ actually *converges* to the constant μ as $n \to \infty$.

Thus, we see that a random variable can in some sense converge to something. With this in mind, can a sequence of *cdfs* converge to something?

Definition: The sequence of RVs $Y_1, Y_2, \ldots$ with respective cdfs $F_{Y_1}(y), F_{Y_2}(y), \ldots$ *converges in distribution* to the RV Y having cdf $F_Y(y)$ if $\lim_{n\to\infty} F_{Y_n}(y) = F_Y(y)$ for all y for which $F_Y(y)$ is continuous. Notation: $Y_n \xrightarrow{d} Y$.

Idea: If $Y_n \xrightarrow{d} Y$ and n is large, then you ought to be able to approximate the distribution of Y_n by the limit distribution of Y.

And now what we've all been waiting for — the most-important theorem in the world (at least until we talk about more-general versions later on). The Central Limit Theorem says that sums and averages of almost anything can often be well-approximated by a normal distribution.

Central Limit Theorem (CLT): If $X_1, X_2, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ with mean μ and finite variance σ^2, then

$$Z_n \equiv \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \text{Nor}(0,1).$$

Thus, the cdf of Z_n approaches $\Phi(z)$ as n increases. The CLT usually works well if the pmf/pdf is fairly symmetric and $n \geq 15$.

Example: If $X_1, X_2, \ldots, X_{100} \stackrel{\text{iid}}{\sim} \text{Exp}(1)$ (so $\mu = \sigma^2 = 1$), then

$$\begin{aligned} & P\left(90 \le \sum_{i=1}^{100} X_i \le 110\right) \\ &= P\left(\frac{90-100}{\sqrt{100}} \le Z_{100} \le \frac{110-100}{\sqrt{100}}\right) \\ &\approx P(-1 \le \text{Nor}(0,1) \le 1) \;=\; 0.6827. \end{aligned}$$

By the way, since $\sum_{i=1}^{100} X_i \sim \text{Erlang}_{k=100}(\lambda = 1)$, we can use the cdf (which may be tedious to evaluate) or software such as Minitab to obtain the *exact* value of $P(90 \le \sum_{i=1}^{100} X_i \le 110) = 0.6835$.

Wow! The CLT and exact answers match nicely! $\square$

Exercise: Demonstrate that the CLT actually works.

1. Pick your favorite RV X_1. Simulate it 1000 or so times and make a histogram.
2. Now suppose X_1 and X_2 are iid from your favorite distribution. Make a histogram of $X_1 + X_2$.
3. Now $X_1 + X_2 + X_3$.
4. ... Now $X_1 + X_2 + \cdots + X_n$ for some reasonably large n.
5. Does the CLT work for the Cauchy distribution, i.e., $X = \tan(2\pi U)$, where $U \sim \text{Unif}(0,1)$? Answer: Nope — the Cauchy violates the finite variance assumption.

---

§2.13 Assessment

1. What is the most-important theorem in the universe?
   (a) Eastern Limit Theorem
   (b) Central Limit Theorem
   (c) Mountain Limit Theorem
   (d) Pacific Limit Theorem
2. Let's take a bunch of independent observations from a "well-behaved" distribution. The Central Limit Theorem says that the standardized sample mean of those observations converges to what distribution?
   (a) exponential
   (b) Weibull
   (c) binomial
   (d) normal

III. Statistics Primer

Our last stop at boot camp is Statistics City. Statistical methods are necessary in order for us to (i) estimate parameters that will be used as input to models, (ii) analyze simulation output, and (iii) select the best of a number of potential solution strategies for problems of interest. We'll just cover the basics here, in particular, simple point estimation and confidence interval techniques. Additional material will be presented as needed in Chapters 6 and 8–10.

2.14 Introduction to Estimation

Definition: A *statistic* is a function of the observations $X_1, X_2, \ldots, X_n$, and not explicitly dependent on any *unknown parameters* such as $\mu = \mathrm{E}[X_i]$ or $\sigma^2 = \mathrm{Var}(X_i)$.

Examples: Well-known statistics include the *sample mean* $\bar{X} \equiv \frac{1}{n}\sum_{i=1}^{n} X_i$ and the *sample variance* $S^2 \equiv \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Statistics are *random variables*. If we take two different samples, we'd expect to get two different values of a statistic. A statistic is used to estimate some unknown parameter like μ or σ^2.

Let $X_1, \ldots, X_n$ be iid RVs and let $T(\boldsymbol{X}) \equiv T(X_1, X_2, \ldots, X_n)$ be a statistic based on the X_i's. Suppose we use $T(\boldsymbol{X})$ to estimate some generic unknown parameter θ. Then $T(\boldsymbol{X})$ is called a *point estimator* for θ.

Examples: $\bar{X}$ is usually a point estimator for the mean $\mu = \mathrm{E}[X_i]$, and S^2 is often a point estimator for the variance $\sigma^2 = \mathrm{Var}(X_i)$.

It would be nice if $T(\boldsymbol{X})$ had certain properties:

- Its expected value should equal the parameter it's trying to estimate. (So on average, we're correct.)
- It should have low variance. (So we don't bounce around too much.)

We now discuss issues regarding the expectation of an estimator.

Definition: $T(\boldsymbol{X})$ is *unbiased* for θ if $\mathrm{E}[T(\boldsymbol{X})] = \theta$.

Theorem: Suppose $X_1, X_2, \ldots, X_n$ are iid anything with mean μ. Then $\bar{X}$ is *always unbiased* for μ. That's why $\bar{X}$ is called the *sample mean*.

Proof: Very easy.

$$\mathrm{E}[\bar{X}] \;=\; \mathrm{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] \;=\; \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}[X_i] \;=\; \mathrm{E}[X_i] \;=\; \mu. \quad \Box$$

Example: In particular, suppose $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Then $\bar{X}$ is unbiased for $\mu = \mathrm{E}[X_i] = 1/\lambda$. $\Box$

Remark: Warning! Be careful because it turns out that $1/\bar{X}$ is a bit *biased* for λ in the exponential case, i.e., $\mathrm{E}[1/\bar{X}] \neq 1/\mathrm{E}[\bar{X}] = \lambda$.

Theorem: Suppose $X_1, X_2, \ldots, X_n$ are iid anything with mean μ and variance σ^2. Then $\mathrm{E}[S^2] = \mathrm{Var}(X_i) = \sigma^2$, so that S^2 is *always unbiased* for σ^2. This is why S^2 is called the *sample variance*.

Proof: First, some algebra gives

$$S^2 \;=\; \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \;=\; \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right).$$

Since $\mathrm{E}[X_i] = \mathrm{E}[\bar{X}] = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$, we have

$$\begin{aligned}
\mathrm{E}[S^2] &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathrm{E}[X_i^2] - n\mathrm{E}[\bar{X}^2]\right) \\
&= \frac{n}{n-1}\left(\mathrm{E}[X_1^2] - \mathrm{E}[\bar{X}^2]\right) \\
&= \frac{n}{n-1}\left(\mathrm{Var}(X_1) + (\mathrm{E}[X_1])^2 - \mathrm{Var}(\bar{X}) - (\mathrm{E}[\bar{X}])^2\right) \\
&= \frac{n}{n-1}(\sigma^2 - \sigma^2/n) \;=\; \sigma^2. \quad \Box
\end{aligned}$$

Example: Suppose $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Then S^2 is unbiased for $\mathrm{Var}(X_i) = 1/\lambda^2$. $\Box$

Remark: We're really serious about the necessity of the theorem's *independence* assumption. Things can go horribly wrong with dependent data — if you don't believe us, see Chapter 9!

Remark: Another warning! S is *biased* for the standard deviation σ.

The next example demonstrates that some unbiased estimators are better than others. The take away will be that you can break the tie between two unbiased estimators by looking at their variances.

Example: Suppose that $X_1, X_2 \ldots, X_n \stackrel{\text{iid}}{\sim} \text{Unif}(0, \theta)$, i.e., the pdf is $f(x) = 1/\theta$, $0 < x < \theta$. Here, the parameter θ is the unknown upper bound on the distribution, which we have to guess based on the observations. For instance, if we see $X_1 = 15.9$, $X_2 = 1.3$, $X_3 = 27.6$, and $X_4 = 11.8$, then our guess for θ must be *at least* $\max_i X_i = 27.6$ — and maybe even a little more — else it won't make any sense.

We'll consider two estimators for θ: $Y_1 \equiv 2\bar{X}$ and $Y_2 \equiv \frac{n+1}{n} \max_{1 \le i \le n} X_i$. Which one is better?

First of all, since $\text{E}[Y_1] = 2\text{E}[\bar{X}] = 2\text{E}[X_i] = \theta$, we see that Y_1 is unbiased for θ.

It's also the case that Y_2 is unbiased, but it takes a little more work to show this, so we'll be lazy and defer the details of our toils to Chapter 8.

Since Y_1 and Y_2 are both unbiased for θ, we'll break the tie by picking the estimator with the *lower variance*. After all, if you're taking a shower, which of the following would you prefer?

- A nice, even, warm, low-variance water flow where the temperature never changes?
- A high-variance water flow that's freezing half the time and boiling half the time (so "just right" on average but always awful)?

Anyhow, after performing algebra similar to that employed above, we have

$$\text{Var}(Y_1) = \frac{\theta^2}{3n} \quad \text{and} \quad \text{Var}(Y_2) = \frac{\theta^2}{n(n+2)}.$$

Thus, Y_2 has *much lower variance* than Y_1; and so Y_2 is the better estimator since the use of the "max" effectively pins down in our guess for θ. □

Mean Squared Error

The previous example shows that we really ought to consider the expected value *and* variance when evaluating the efficacy of potential estimators.

Definition: The *bias* of $T(\boldsymbol{X})$ as an estimator of θ is $\text{Bias}(T) \equiv \text{E}[T] - \theta$.

The *mean squared error* of $T(\boldsymbol{X})$ is $\text{MSE}(T) \equiv \text{E}[(T - \theta)^2]$.

Theorem: After some algebra, we get an easier expression for MSE that combines the bias and variance of an estimator

$$\text{MSE}(T) = \text{Var}(T) + (\underbrace{\text{E}[T] - \theta}_{\text{Bias}})^2.$$

Remark: Lower MSE is better — we usually won't get upset by a *little* bias.

Definition: The *relative efficiency (RE)* of T_2 to T_1 is $\text{MSE}(T_1)/\text{MSE}(T_2)$. This is an effective measure for comparing estimators; if RE < 1, then we'd want T_1.

Example: Consider our old friend, $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Unif}(0, \theta)$, for which we evaluated two estimators: $Y_1 = 2\bar{X}$ and $Y_2 = \frac{n+1}{n} \max_i X_i$. We showed showed before that $\text{E}[Y_1] = \text{E}[Y_2] = \theta$ (so both are unbiased); and furthermore, $\text{Var}(Y_1) = \theta^2/(3n)$ and $\text{Var}(Y_2) = \theta^2/[n(n+2)]$. Thus,

$$\text{RE} \;=\; \frac{\text{MSE}(T_1)}{\text{MSE}(T_2)} \;=\; \frac{\theta^2/(3n)}{\theta^2/[n(n+2)]} \;\approx\; n/3.$$

In fact, RE ≥ 1 for all n; and so T_2 is again the better choice. $\Box$

---

§2.14 Assessment

1. Suppose that we are using some estimator T to estimate an unknown parameter θ. Further suppose that T has a bias of 3 and a variance of 5. What is T's mean squared error?

   (a) 8

   (b) 28

   (c) 14

   (d) 34

2. TRUE or FALSE? The sample mean is always unbiased for the true mean. And, while we're at it, the sample variance is always unbiased for the true variance.

2.15 Maximum Likelihood Estimation

Although we love and admire unbiased estimators, we saw in §**??** that they were occasionally funky. For instance, (i) they're not unique, (ii) they can sometimes have goofy variance properties, (iii) they may be difficult to obtain, and (iv) they lack the intuitive "invariance" property in that $T(\boldsymbol{X})$ being unbiased for θ does not necessarily imply $h(T(\boldsymbol{X}))$ is unbiased for $h(\theta)$ for a given function $h(\cdot)$. Maximum likelihood estimators are an alternative class of estimators that often get around some of the above issues.

Definition: Consider an iid random sample $X_1, X_2, \ldots, X_n$, where each X_i has pmf/pdf $f(x)$. Further, suppose that θ is some unknown parameter from X_i. The *likelihood function* is $L(\theta) \equiv \prod_{i=1}^{n} f(x_i)$.

Definition: The *maximum likelihood estimator (MLE)* of θ is the value of θ that maximizes $L(\theta)$. The MLE is a function of the X_i's and is a random variable.

Example: Suppose $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Let's find the MLE for λ. First, we'll get the likelihood function,

$$L(\lambda) \;=\; \prod_{i=1}^n f(x_i) \;=\; \prod_{i=1}^n \lambda e^{-\lambda x_i} \;=\; \lambda^n \exp\Big(-\lambda \sum_{i=1}^n x_i\Big).$$

Now maximize $L(\lambda)$ with respect to λ. We could probably take the derivative and plow through all of the horrible algebra, but this is usually an onerous undertaking. Might there be a useful trick that would hasten our endeavors?

Useful Trick: Since the natural log function is one-to-one, it's easy to see that the λ that maximizes $L(\lambda)$ also maximizes $\ell\text{n}(L(\lambda))$! Thus, we'll maximize the log-likelihood function,

$$\ell\text{n}(L(\lambda)) \;=\; \ell\text{n}\bigg(\lambda^n \exp\Big(-\lambda \sum_{i=1}^n x_i\Big)\bigg) \;=\; n\ell\text{n}(\lambda) - \lambda \sum_{i=1}^n x_i.$$

Use of this simpler expression certainly makes our job less horrible. Now we have

$$\frac{d}{d\lambda}\ell\text{n}(L(\lambda)) \;=\; \frac{d}{d\lambda}\Big(n\ell\text{n}(\lambda) - \lambda \sum_{i=1}^n x_i\Big) \;=\; \frac{n}{\lambda} - \sum_{i=1}^n x_i \;\equiv\; 0,$$

where we set things = 0 in order to obtain the critical point. This implies that the MLE is $\hat{\lambda} = 1/\bar{X}$. $\Box$

Remarks: A panoply of tidbits...

- $\hat{\lambda} = 1/\bar{X}$ makes sense since $\text{E}[X] = 1/\lambda$.
- At the end of our labors, we put a little $\widehat{\text{hat}}$ over λ to indicate that this is the MLE.
- At the end, we make all of the little x_i's into big X_i's to indicate that this is a random variable.
- Just to be careful, you probably ought to perform a second-derivative test, but we won't blame you if you don't.
- We will present a bevy of additional examples — some of which you will not have seen in your Baby Stats class — in Chapter 8.

The following theorem is one of the main reasons why we like MLEs so much. It will allow us to find MLEs for a larger class of parameters tnan you might at first think.

Theorem (Invariance Property of MLEs): If $\hat{\theta}$ is the MLE of some parameter θ and $h(\cdot)$ is a one-to-one function, then $h(\hat{\theta})$ is the MLE of $h(\theta)$.

Example: Suppose $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$. We define the *survival function* as

$$\bar{F}(x) \;=\; P(X > x) \;=\; 1 - F(x) \;=\; e^{-\lambda x}.$$

In addition, recall from the previous example that the MLE for λ is $\hat{\lambda} = 1/\bar{X}$. Then the Invariance Property says that the MLE of $\bar{F}(x)$ is

$$\widehat{\bar{F}(x)} \;=\; e^{-\hat{\lambda}x} \;=\; e^{-x/\bar{X}}.$$

This kind of thing is used all of the time the actuarial sciences; and we will use this methodology ourselves in Chapter 8 to perform goodness-of-fit tests for certain simulation input processes. $\Box$

---

§2.15 Assessment

1. Suppose that we are looking at i.i.d. Exp(λ) customer service times. We observe times of 20, 40, and 90 minutes. What's the maximum likelihood estimator for λ?

   (a) 3

   (b) 50

   (c) 1/50

   (d) 1/3

2. If the MLE for some parameter θ has been calculated as $\hat{\theta} = 3$, then what's the MLE for e^{θ}?

   (a) e^3

   (b) $e^{1/3}$

   (c) 1/3

   (d) I'm from the University of Georgia and I'm confused, cold, and wet.

2.16 Distributional Results and Confidence Intervals

There are a number of distributions (including the normal) that come up in statistical sampling problems. Here are a few:

Definitions: If $Z_1, Z_2, \dots, Z_k$ are iid Nor(0,1), then $Y = \sum_{i=1}^{k} Z_i^2$ has the *χ^2 distribution with k degrees of freedom (df)*. Notation: $Y \sim \chi^2(k)$. Note that $\text{E}[Y] = k$ and

$\text{Var}(Y) = 2k$.

If $Z \sim \text{Nor}(0,1)$, $Y \sim \chi^2(k)$, and Z and Y are independent, then $T = Z/\sqrt{Y/k}$ has the *Student t distribution with k df.* Notation: $T \sim t(k)$. Note that the $t(1)$ is the *Cauchy distribution.* Gesundheit!

If $Y_1 \sim \chi^2(m)$, $Y_2 \sim \chi^2(n)$, and Y_1 and Y_2 are independent, then $F = (Y_1/m)/(Y_2/n)$ has the *F distribution with m and n df.* Notation: $F \sim F(m,n)$.

How (and why) would one use the above facts? Because they can be used to construct *confidence intervals (CIs)* for $\mu = \text{E}[X_i]$ and $\sigma^2 = \text{Var}(X_i)$ under a variety of assumptions.

A *$100(1-\alpha)\%$ two-sided CI* for an unknown parameter θ is a random interval $[L, U]$ such that $P(L \le \theta \le U) = 1 - \alpha$.

Here are some examples / theorems, all of which assume that we observe iid normal observations $X_1, X_2, \ldots, X_n$.

Theorem: If σ^2 is *known*, then a $100(1-\alpha)\%$ two-sided CI for μ is

$$\bar{X}_n - z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + z_{\alpha/2}\sqrt{\frac{\sigma^2}{n}},$$

where z_γ is the $1-\gamma$ quantile of the standard normal distribution, i.e., $z_\gamma \equiv \Phi^{-1}(1-\gamma)$.

Theorem: If σ^2 is *unknown* (which is, by far, the most-useful case), then a $100(1-\alpha)\%$ two-sided CI for μ is

$$\bar{X}_n - t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}} \;\le\; \mu \;\le\; \bar{X}_n + t_{\alpha/2,n-1}\sqrt{\frac{S^2}{n}},$$

where $t_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $t(\nu)$ distribution.

Example: Consider 20 residual flame times (in sec.) of treated specimens of children's nightwear. (Don't worry — children were not in the nightwear when the clothing was set on fire.)

9.85	9.93	9.75	9.77	9.67
9.87	9.67	9.94	9.85	9.75
9.83	9.92	9.74	9.99	9.88
9.95	9.95	9.93	9.92	9.89

Let's get a 95% CI for the mean residual flame time. After a little algebra, we obtain

$$\bar{X} \;=\; 9.8525 \quad \text{and} \quad S \;=\; 0.09646.$$

Further, you can use the Excel function `t.inv(0.975,19)` to get t-quantile $t_{\alpha/2,n-1} = t_{0.025,19} = 2.093$. Then the *half-length* of the CI is

$$H \;=\; t_{\alpha/2,n-1}\sqrt{S^2/n} \;=\; \frac{(2.093)(0.0965)}{\sqrt{20}} \;=\; 0.0451.$$

Thus, the CI is $\mu \in \bar{X} \pm H$, or $9.8074 \le \mu \le 9.8976$. $\Box$

Theorem: A $100(1-\alpha)\%$ two-sided CI for σ^2 is

$$\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2},n-1}} \;\le\; \sigma^2 \;\le\; \frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2},n-1}},$$

where $\chi^2_{\gamma,\nu}$ is the $1-\gamma$ quantile of the $\chi^2(\nu)$ distribution.

---

§2.16 Assessment

1. TRUE or FALSE? The length of a confidence interval increases as you demand higher confidence (larger $1-\alpha$).

2. TRUE or FALSE? A 95% confidence interval means that you are 95% sure that the true parameter lies somewhere in the interior of the interval.

2.17 Exercises

1. (§2.1) If $f(x) = \ell\text{n}(2x-3)$, find the derivative $f'(x)$.

2. (§2.1) If $f(x) = \cos(1/x)$, find the derivative $f'(x)$.

3. (§2.2) Suppose that $f(x) = e^{4x} - 4e^{2x} + 4$. Use any method you want to find a zero of $f(x)$, i.e., x such that $f(x) = 0$.

4. (§2.3) Find $\int_0^1 (2x+1)^2\,dx$.

5. (§2.3) Find $\int_1^2 e^{2x}\,dx$.

6. (§2.3) Find $\lim_{x\to 0} \frac{\sin(x)-x}{x}$. Hint: you may need to go to the L'Hôspital.

7. (§2.4) Find the approximate value of the integral $\int_0^2 (x-1)^2\,dx$ using the lesson's form of the Riemann sum with $f(x) = (x-1)^2$, $a = 0$, $b = 2$, and $n = 4$.

8. (§2.5) If $P(A) = P(B) = P(C) = 0.6$, and A, B, and C are independent, find the probability that *exactly one* of A, B, and C occurs.

9. (§2.5) Toss 3 dice. What's the probability that a "4" will come up exactly twice?

10. (§2.6) Suppose U and V are independent Uniform(0,1) random variables. (You can simulate these using the `RAND()` function in Excel, for instance.) Consider the nasty-looking random variable

$$Z \;=\; \sqrt{-2\ell\text{n}(U)}\,\cos(2\pi V),$$

where the cosine calculation is carried out in *radians* (not degrees). Go ahead and calculate $Z\ldots$ don't be afraid. Now, repeat this task 1000 times (easy to do in Excel) and make a histogram of the 1000 Z's. What distribution does this look like?

11. (§2.7) Suppose that X is a discrete random variable having $X = -1$ with probability 0.2, and $X = 3$ with probability 0.8.

    (a) Find $\text{E}[X]$.

    (b) Find $\text{Var}(X)$.

    (c) Find $\text{E}[3 - \frac{1}{X}]$.

12. (§2.7) Suppose X is a continuous random variable with pdf $f(x) = 4x^3$ for $0 \le x \le 1$. Find $\text{E}[1/X^2]$.

13. (§2.8) Suppose X is the result of a 5-sided die toss having sides numbered $-2, -1, 0, 1, 2$. Find the probability mass function of $Y = X^2$.

14. (§2.8) Suppose X is a continuous random variable with pdf $f(x) = 2x$ for $0 < x < 1$. Find the pdf $g(y)$ of $Y = X^2$. (This may be easier than you think.)

15. (§2.9) Suppose that $f(x, y) = 6x$ for $0 \le x \le y \le 1$.

    (a) Find $P(X < 1/2$ and $Y < 1/2)$.

    (b) Find the marginal pdf $f_X(x)$ of X.

16. (§2.9) YES or NO? Suppose X and Y have joint pdf $f(x, y) = cxy/(1+x^2+y^2)$ for $0 < x < 1$, $0 < y < 1$, and whatever constant c makes the nasty mess integrate to 1. Are X and Y independent?

17. (§2.10) Suppose that $f(x, y) = 6x$ for $0 \le x \le y \le 1$. It turns out (take our word for it) that the marginal pdf of X is

$$f_X(x) \;=\; 6x(1-x) \quad \text{for } 0 \le x \le 1.$$

    (a) Find the conditional pdf of Y given that $X = x$.

    (b) Find $\text{E}[Y|X = x]$.

   (c) Find $\mathrm{E}\big[\mathrm{E}[Y|X]\big]$.

18. (§2.11) Suppose that the correlation between December snowfall and temperature in Siberacuse, NY is -0.5. Further suppose that $\mathrm{Var}(S) = 100$ in^2 and $\mathrm{Var}(T) = 25$ (degrees F)^2. Find $\mathrm{Cov}(S,T)$ (in units of degree inches, whatever those are).

19. (§2.11) If X and Y both have mean -7 and variance 4, and $\mathrm{Cov}(X,Y) = 1$, find $\mathrm{Var}(3X - Y)$.

20. (§2.12) You may recall that the pmf of the Geometric(p) distribution is $f(x) = (1-p)^{x-1}p$, $x = 1, 2, \ldots$. If the number of orders at a production center this month is a Geom(0.7) random variable, find the probability that we'll have at most 3 orders.

21. (§2.12) Suppose the SAT math score of a University of Georgia student can be approximated by a normal distribution with mean 400 and variance 225. Find the probability that the UGA Einstein will score at least a 415.

22. (§2.13) What is the most-important theorem in the universe?

    (a) Eastern Limit Theorem

    (b) Central Limit Theorem

    (c) Central Limit Serum

    (d) Central Simit Theorem (simit is a tasty Turkish bagel)

23. (§2.13) If $X_1, \ldots, X_{400}$ are i.i.d. from some distribution with mean 1 and variance 400, find the approximate probability that the sample mean $\bar{X}$ is between 0 and 2.

24. (§2.14) Suppose we collect the following observations: 7, -2, 1, 6. What is the sample variance?

25. (§2.14) Consider two estimators, T_1 and T_2, for an unknown parameter θ. Suppose that the $\mathrm{Bias}(T_1) = 0$, $\mathrm{Bias}(T_2) = \theta$, $\mathrm{Var}(T_1) = 4\theta^2$, and $\mathrm{Var}(T_2) = \theta^2$. Which estimator might you decide to use and why?

26. (§2.15) Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. $\mathrm{Pois}(\lambda)$. Find $\hat{\lambda}$, the MLE of λ. (Don't panic — it's not that difficult.)

27. (§2.15) Suppose that we are looking at i.i.d. Exp(λ) customer service times. We observe times of 2, 4, and 9 minutes. What's the maximum likelihood estimator of λ^2?

28. (§2.16) Suppose we collect the following observations: 7, -2, 1, 6 (as in a previous question in this homework). Let's assume that these guys are i.i.d. from a normal distribution with *unknown* variance σ^2. Give me a two-sided 95% confidence interval for the mean μ.

Bibliography

T. Aktaran-Kalaycı, C. Alexopoulos, N. T. Argon, D. Goldsman, and J. R. Wilson. (This is an example of a journal paper with a lot of authors) Exact expected values of variance estimators for simulation. *Naval Research Logistics*, 54:397–410, 2007.

P. Billingsley. *(This is an example of a book) Convergence of Probability Measures*. John Wiley & Sons, New York, 1968.

P. Bratley, B. L. Fox, and L. E. Schrage. *A Guide to Simulation* (2nd edition). Springer-Verlag, New York, 1987.

G. Buffon. Essai d'arithmétique morale. *Histoire naturelle, générale er particulière*, *Supplément*, 4:46–123, 1777.

D. Goldsman, R. E. Nance, and J. R. Wilson. A brief history of simulation. In M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, editors, *Proceedings of the 2009 Winter Simulation Conference*, pages 310–313. Institute of Electrical and Electronics Engineers, Piscataway, NJ, 2009.

N. Metropolis. The beginning of the Monte Carlo method. In N. G. Cooper, editor, *Reflections on the Life and Legacy of Stanislaw Ulam*, pages 125–130. Cambridge University Press New York, 1989.

Wolfram Research. (This is an example of software) Mathematica, Version 9.0, Champaign, IL, 2014.