Stochastic Calculus, Courant Institute, Fall 2021
http://www.math.nyu.edu/faculty/goodman/teaching/StochCalc2021/index.html

Week 1

Jonathan Goodman, October, 2021

1 About this Stochastic Calculus course

This short course is about building and understanding models of random processes. The models are *stochastic differential equations* (usually called *SDEs*), which are a random process analogue of ordinary differential equations (ODEs). Both ODE and SDE models involve a state variable X_t that describes the state of the "system" (what is being modelled) at time t. Both describe the *increment* $dX_t = X_{t+dt} - X_t$, which depends on X_t in some way. In an SDE model, dX_t also depends on some random "noise" that "arrives" in the interval $[t, t + dt]$. In an SDE model, dt must be positive. We often think of dt as positive in an ODE model, but that is not necessary. An ODE model takes the form

$$dX_t = a(X_t)\, dt \; . \tag{1}$$

An SDE model takes dX_t to be a random variable and specifies its mean and variance

$$\mathrm{E}[\, dX_t] = a(X_t)\, dt \tag{2}$$

$$\mathrm{var}(\, dX_t) = \mu(X_t)\, dt \; . \tag{3}$$

This course is about creating models like this and working with them. Be aware that these formulas are written in a simplified and somewhat incorrect form. More correct versions are (**??**) and (**??**) below.

The above paragraph illustrates the mathematical level of the course. We will often use informal or heuristic reasoning, as is done by most people using stochastic calculus or any other part of applied mathematics. Sometimes we will examine the concepts more carefully. You can think of the definition of the Riemann integral (area defined by boxes with side Δx in the limit $\Delta x \to 0$) as is done in typical calculus classes, but with complete mathematical proofs. The full mathematical framework for stochastic calculus starts with abstract probability measure theory.

Stochastic calculus is a mathematical framework for SDEs, just as ordinary calculus is a framework for ODEs. There is a stochastic integral, the *Ito integral* for integrating noise. This is a complement to the "ordinary" integral $\int \cdots dt$. There is a stochastic chain rule, called *Ito's lemma*, which is used to differentiate functions of the random path X_t.

In stochastic calculus, we call the function X_t (a function of t), a *path*. Since the path is random, stochastic calculus has elements of probability theory

that have no analogue in ordinary differential equations (1). The location of the random path at time t is a random variable that has a probability density $p(x,t)$. This density satisfies a partial differential equation, called the *forward equation* (also *Fokker Planck* equation) that describes how the probability density changes as t increases.

Conditional probability and conditional expectation are central to stochastic calculus. A common situation involves a random path on a time interval $0 \leq t \leq T$ and an intermediate time $t_1 > 0$, $t_1 < T$. You can look at a "sub path" $X_{[0,t_1]}$ (this is all values X_t for $0 \leq t \leq T$) and ask about the path in the future of t_1, which is $X_{[t_1,T]}$. An example in this direction is the *value function* defined by the conditional expectation

$$f(x,t_1) = \mathrm{E}[\, V(X_T) \mid X_{t_1} = x] \ . \tag{4}$$

In an application in financial economics, X_t could represent the "market" at time t and $V(x)$ some measure of value of X, such as $V(x) = x$ (if X represents a price) or $V(x) = x^2$ if we like larger values of x even more. Then the value function is the expectation of this time T value, given information up to some earlier time $t_1 < T$. This value function also satisfies a partial differential equation (PDE) called the *backward equation* (or *Kolmogorov* backward equation, or *Chapman* Kolmogorov equation). The PDE/SDE connection is used both ways. The PDE provides a way learn about SDE models. SDE models provide a way to find solutions to certain PDEs using simulations of the stochastic process. Both directions are often (and often incorrectly) called the *Feynman Kac formula*.

Change of measure formulas are another part of SDE theory that has no analogue in ODE. These formulas, often called *Girsanov* formulas, relate the probability distributions determined by different SDEs. Change of measure is a central concept in the theory of stock options – *Black Scholes* theory and developments from that. The "real world measure" and the *risk neutral* measure (two SDE "worlds") are related using Girsanov's theorem. Change of measure has more practical uses too. It allows you to learn about solutions of one SDE by simulating paths from a different SDE, and then *re-weighting* to compensate for changing the SDE.

These are the important topics for this course. Hopefully you can re-read this section to review for the final exam and it will all make sense.

2 Probability theory for SDE

The first theoretical issues in probability theory are: How do you describe a random object, and how do you specify its probability distribution? In univeriate probability, the random object is a number. Its distribution is described using a probability density function (PDF) or its cumulative distribution function (CDF). In discrete probability, there is a finite or infinite list of possible objects (random integer, random graph, etc.) and you give the probability of each object. In multi-variate probability with n components, the random object

is n numbers, the components, and a probability density function describes their joint distribution. Notations for these situations are given below.

The probability theory of SDEs cannot be described in any of these ways. The random object is a function of the continuous time variable t. The value of this function at time t is X_t. This is given for some range of t, often $0 \leq t \leq T$, but other ranges are possible. The function is called a *path*. For SDE, the paths usually are continuous functions of t. Thus, the random object is a continuous path defined on some time interval. The probability distribution is given by a *probability measure*. As already stated, this short course of applied aspects of SDEs will not describe probability measures completely.

In principle the probability distribution of a random path is completely determined by the distribution of its starting point, typically X_0, and the dynamical relations (2) (the *drift* or *infinitesimal mean*) and (3) (the *quadratic variation* or *infinitesimal variance*). This is the SDE analogue of the fact that an ordinary differential equation path is determined by its *initial condition* X_0 (typically not random) and the dynamical relation (1). We try to learn about random paths directly from the dynamics without writing expressions for the probability measure. For example, the forward and backward partial differential equations mentioned above are found in this way.

Operations in calculus may be defined using a small time interval Δt, which goes to zero. Stochastic calculus takes this one step further and uses a limit $\Delta t \to 0$ to describe the probability distribution of a random path. To put this in a systematic notation, suppose the time interval of the path is $[0, T]$, choose a number of time steps n and a time interval $\Delta t = T/n$. We will use the following clumsy notation (feel free to suggest or use a less clumsy one) in which $X^{\Delta t} \in \mathbb{R}^n$ is the *discrete time path* consisting of *observations* of the continuous time path at times $t_k = k\Delta t$. The discrete time observations are

$$X_k^{\Delta t} = X_{t_k} \ , \ \ t_k = \Delta t \ .$$

These form the components of the vector $X^{\Delta t}$

$$X^{\Delta t} = \begin{pmatrix} X_{\Delta t} \\ X_{2\Delta t} \\ \vdots \\ X_T \end{pmatrix} = \begin{pmatrix} X_{t_1} \\ X_{t_2} \\ \vdots \\ X_{t_n} \end{pmatrix} \in \mathbb{R}^n \ \ .$$

The dimension of the space of observation vectors, which is n, goes to infinity as $\Delta t \to 0$. In this sense the space of paths – *path space* – is infinite dimensional.

The dimension of the observation vector space is finite, though large. Therefore, there may be a joint probability density for the observation components. The probability density will be written

$$p^{\Delta t}(x_1, x_2, \cdots, x_n) \ .$$

This sequence of probability densities defines the probability distribution of the path. It is possible to write formulas for $p^{\Delta t}$ in some simple cases. The formula

for *Brownian motion* is easy but lengthy to write, but part of it is in Exercise 1. In harder cases there are approximate formulas for $p^{\Delta t}$. In ordinary calculus, we try to avoid going all the way back to $\Delta t \to 0$. Instead we use theorems like the chain rule. In stochastic calculus, we try to reason about path space probabilities without going to $p^{\Delta t}$. Instead, we build up tools such as Ito's lemma and backward equations. Of course, these depend on $\Delta t \to 0$ limits.

SDE models (2)(3) determine probability distributions in path space that have the *Markov property*. This is the property that the future of time t, the probability distribution of the path X_s for $s > t$, depends on X_t only. For example, (2) says that the expected change in X just after time t depends on X_t only. To say this properly, we give corrected versions of (2) and (3) that use conditional expectations.

The conditional expectations and probabilities we need depend on pieces of the path X. We write $X_{[s_0,s_1]}$ to denote the path from $t = s_0$ to $t = s_1$. In particular, we need the path up to a time s_1, which is $X_{[0,s_1]}$. Naturally, this assumes $0 \leq s_0 \leq s_1$. We write $X^{\Delta t}_{[0,s_1]}$ for the set of discrete time observations that happen up to time s_1. The components are the numbers X_{t_k} for all k with $t_k \leq s_1$. The dimension of the space of vectors $X^{\Delta t}_{[0,s_1]}$ depends on Δt and s_1 and is given by

$$\left\lfloor \frac{s_1}{\Delta t} \right\rfloor = \text{ greatest integer } \leq \frac{s_1}{\Delta t} \, .$$

Here, $\lfloor x \rfloor$, the "floor" of x, is what you get by rounding down to the nearest integer. The *past* of a time t is determined by the partial observation paths $X^{\Delta t}_{[0,s_1]}$, in the limit $\Delta t \to 0$.

The conditional probabilities we need are distributions of X_{s_2} conditioned on knowing the path up to time s_1. If $\Delta t > 0$ then this can be expressed in the language of simple conditional probability. We seek the distribution of X_{s_2} conditioned on values of the discrete time observations before time s_1. Denote these values by x_1, x_2, etc. The conditional density is

$$X_{s_2} \sim p(\cdot \mid X_{\Delta t} = x_1, X_{t_2} = x_2, \cdots, X_{t_k} = x_k) \, . \tag{5}$$

Here, $k = \lfloor s_1 / \Delta t \rfloor$ is the number of observation times up to s_1. The conditional probability is defined in the usual finite dimensional way. If Y and Z are random variables with some joint PDF, and if z is a possible value of Z, then the conditional probability density is

$$p(y \mid Z = z) = \frac{p(y,z)}{\int p(y',z) \, dy'} \, .$$

The conditional density of (5) is this, with Y given by X_{s_2} and Z being the observation values $X_{t_1}, \cdots, X_{t_k}$. The point here is not to write a lot of formulas, but to make it clear what these conditional probabilities and densities are.

In this notation, the *Markov property* is the property that $p(\cdot | \cdots)$ of (5) depends only on x_k and not on earlier values:

$$p(\cdot \mid X_{\Delta t} = x_1, X_{t_2} = x_2, \cdots, X_{t_k} = x_k) = p(\cdot \mid X_{t_k} = x_k) \, . \tag{6}$$

The limit $\Delta t \to 0$ defines a conditional probability conditional on the path $X_{[0,s_1]}$. These probabilities conditional on the whole partial path (the whole of $X_{[0,s_1]}$) require measure theory to define completely, but an informal statement of the Markov property is that the PDF of X_{s_2} conditional on the whole path $X_{[0,s_1]}$ is the same as the probability conditional on the las value, X_{s_1}.

$$p(\,\cdot \mid X_{[0,s_1]}) = p(\,\cdot \mid X_{s_1}) \tag{7}$$

A future class will discuss a modern form of conditional probability.

The correction to (2) and (3) is only to say that X_t has the Markov property the infinitesimal mean and variances as conditional

$$\mathrm{E}[\,dX \mid X_t = x] = a(x)\,dt\;, \tag{8}$$

$$\mathrm{var}[\,dX \mid X_t = x] = \mu(x)\,dt\;. \tag{9}$$

We write Δt and $\Delta x = X_{t+\Delta t} - X_t$ instead of dt and dX to describe changes that are small but not infinitely small. It is important to keep in mind the constraint $\Delta t > 0$. In the informal language of differentials, we write $dt > 0$, but pure literal minded mathematicians ask how it's possible to be both infinitely small and positive.

$$\mathrm{E}[\,\Delta X \mid X_t = x] = a(x)\,\Delta t + O(\Delta t^2)\;, \tag{10}$$

$$\mathrm{var}[\,\Delta X \mid X_t = x] = \mu(x)\,\Delta t + O(\Delta t^2)\;. \tag{11}$$

You can look up the "big Oh" notation $O(\Delta t^2)$ in Wikipedia if you're not familiar with it. There will be more material on it in future classes.

3 Brownian motion

Stochastic calculus uses Brownian motion in several ways. A *diffusion process*, or a *diffusion* is a random process described by a stochastic differential equation. Brownian motion is the simplest diffusion process. Other diffusions are described by what they have in common with Brownian motion and the differences between them. Moreover, Brownian motion is used as a model of *noise* for other diffusions. Therefore, we use X_t to denote Brownian motion if it is the only process being discussed, but W_t if Brownian motion is being used as a noise source. This will start here with the X_t notation then switch to W_t when X_t becomes a different process.

A "standard" Brownian motion starts with $X_0 = 0$. The path is a continuous function of t. The *increments* $X_{s_2} - X_{s_1}$ are Gaussian with mean zero and variance equal to the size of the interval:

$$\mathrm{E}[\,X_{s_2} - X_{s_1}] = 0 \tag{12}$$

$$\mathrm{var}[\,X_{s_2} - X_{s_1}] = s_2 - s_1\;. \tag{13}$$

Moreover, the increments over disjoint time intervals are independent. If $s_1 < s_2 < \cdots < s_k$, then the random increments $X_{s_{j+1}} - X_{s_j}$ are all independent.

Brownian motion has the Markov property because nothing in one time interval has any influence on what happens in a different interval. The distribution of X_{s_2}, conditional on the path $X_{[0,s_1]}$, is $X_{s_1} + \mathcal{N}(0, s_2 - s_1)$. The expression $\mathcal{N}(\mu, \sigma^2)$ refers to a Gaussian random variable with mean μ and variance σ^2. Ihe random variable $\mathcal{N}(\mu, \sigma^2)$ is implicitly understood to be independent of other random variables in the same discussion. In this case, it is independent of X_{s_1}.

Brownian motion may be "derived" as a large scale, or *coarse grained* model for the steady arrival of random "information". The random information consists of an i.i.d. sequence of random variables U_i with mean zero and variance σ^2. The *central limit theorem* concerns the distribution of

$$S_M = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} U_i \ .$$

It says that S_M is approximately Gaussian with mean zero and variance σ^2, if M is large. This applies to any M "samples" from the sequence U_i, not necessarily the first M elements of the sequence. In particular, the next M samples give an independent random sum with the same distribution

$$S_M^{(1)} = \frac{1}{\sqrt{M}} \sum_{i=1}^{M} U_i$$

$$S_M^{(2)} = \frac{1}{\sqrt{M}} \sum_{i=M+1}^{2M} U_i$$

etc.

The sums $S_M^{(1)}$ and $S_M^{(2)}$ are both approximately normal for large M, and they are exactly independent. Therefore, their joint distribution is approximately a two component Gaussian with independent Gaussian components. The individual random numbers U_i are sometimes called *shocks*, particularly in finance and economics. This terminology seems funny to me, as you would imagine that economists would eventually stop being "shocked" by new information as each new U_i arrives.

These sums may be *normalized*, or *scaled*, or *re-scaled* by a scaling parameter C. "Clearly",

$$Y_M = \frac{C}{\sqrt{M}} \sum_{i=1}^{M} U_i$$

is approximately Gaussian with mean 0 and variance $C^2\sigma^2$.

By *coarse-graining*, we mean describing changes in the running sum $\sum U_i$, in an approximate but useful way, when there are many terms in any increment of time. For example, a stock price can "tick" (change) up or down by one cent at a time, which is not Gaussian because a tick has only two possible values. However, we may model the change over an hour as approximately Gaussian because there are so many independent ticks in an hour. There are *fine-grained*

models that describe individual or short sequences of ticks, called *market microstructure models*. The coarse-grained model is simpler.

In technical notation, choose T and $\Delta t = T/n$. Suppose the running sum is coarse-grained so that there are M "shocks" in any time interval $[t_k, t_{k+1}]$. Overall there are nM shocks. The time between shocks is T/nM. Shock i arrives at time i/nM (so shock with $i = nM$ arrives at time T). The overall sum is

$$\sum_{i=1}^{nM} U_i \ .$$

We start by "scaling", or "normalizing", which means choosing a scaling factor C so that the variance of the overall sum is the variance of X_T, which is T. The result (you easily check) is

$$C = \frac{\sqrt{T}}{\sqrt{nM\sigma^2}} \ .$$

The scaled running sum is (get the largest i in the sum by rounding t/nM down to the nearest integer)

$$Y_t = \frac{\sqrt{T}}{\sqrt{nM\sigma^2}} \sum_{i=1}^{\lfloor \frac{t}{nM} \rfloor} U_i \ .$$

This Y is a scaled running sum process that is approximately described by Brownian motion.

The increment of Y over the time interval $[t_k, t_{k+1}]$ is determined by the M shocks that in that time interval

$$\Delta Y_k = Y_{t_{k+1}} - Y_{t_k} = \frac{\sqrt{T}}{\sqrt{nM\sigma^2}} \sum_{i=1+kM}^{(k+1)M} U_i \ . \tag{14}$$

The mean is zero and the variance is (after some calculation)

$$\mathrm{var}(\,\Delta Y_k) = \frac{T}{nM\sigma^2}\, M\sigma^2 = \Delta t \ .$$

Furthermore, the increments ΔY_k are approximately Gaussian when M is large. Distinct increments always are independent. In the limit $M \to \infty$, we get paths whose increments are independent and exactly Gaussian with the right mean and variance. Brownian motion paths satisfy an SDE model with infinitesimal mean $a = 0$ and infinitesimal variance $\mu = 1$.

The *Donsker invariance principle* is the theorem that the distribution Y_t converges to the distribution of Brownian motion (described above) in the limit $M \to \infty$. It is named for Professor Monroe (Monie) Donsker, who was a Courant Institute faculty member for several decades. Keep in mind that the numbers Y_t do not converge to numbers X_t as $M \to \infty$. Only the distribution converges. In fact, if M is replaced by $2M$ (for example) all of the sums (except $k = 0$) in (14) change "completely", in that they sums for $2M$ are independent of the sums for M.

4 Geometric Brownian motion

Geometric Brownian motion is a diffusion process in which dX is proportional to X. This may be appropriate for modeling a population where dX specified as percentages of the population. It may be used to model a population where the number of birth and death rate is proportional to the population. It is used to model the random changes in the price of a "stock" (financial asset). Here too, prices go up and down in percentages. The infinitesimal mean being proportional to x means

$$a(x) = rx \ . \tag{15}$$

The parameter r is the expected growth rate (for populations) or the expected rate of return (for prices). The infinitesimal standard deviation of dX also should be proportional to x. This implies that the infinitesimal variance is proportional to x^2. The coefficient is written as σ^2 (it's positive) where σ is the *volatility*.

$$\mu(x) = \sigma^2 x^2 \ . \tag{16}$$

The standard deviation of the increment is the square root of the variance. Here, this implies over an interval of length Δt

$$\text{std dev } (\Delta X_t \mid X_t = x) \approx \sigma\sqrt{\Delta t} \ .$$

We think of $\sqrt{\Delta t}$ as being much larger than Δt. For example, if $\Delta t = 1/256$ ($256 = 2^8$ being the approximate number of trading days in a year) then $\sqrt{\Delta t} = 1/16 = 2^{-4}$, which is larger by a factor of 16.

The term *geometric* comes from the fact that the "log process" $Y_t = \log(X_t)$ is closely related to ordinary Brownian motion. After Week 2, you will be able to verify that it has constant infinitesimal variance of the log process is $\mu = \sigma^2$ and constant infinitesimal mean

$$a = r - \frac{1}{2}\sigma^2 \ . \tag{17}$$

The $\frac{1}{2}\sigma^2$ is the "Ito term" that we will find in Week 2 using *Ito's lemma*. Let W_t be a Brownian motion as above with infinitesimal mean $a = 0$ and infinitesimal variance $\mu = 1$. Then Y_t may be written

$$Y_t = y_0 + \sigma W_t + (r - \frac{1}{2}\sigma^2)t \ . \tag{18}$$

You can check this, since $\text{var}(dY) = \sigma^2 \text{var}(dW) = \sigma^2 dt$, and $\text{E}[dY] = \text{E}[dW](r - \frac{1}{2}\sigma^2)dt$. A geometric series is the exponential of an arithmetic series. If we think of W_t as "arithmetic" Brownian motion, then *geometric Brownian motion* is a fitting way to describe the process

$$X_t = X_0 e^{\sigma W_t + \left(r - \frac{1}{2}\sigma^2\right)t} \ . \tag{19}$$

This formula expresses a diffusion process X_t in terms of a Brownian motion. The classes of weeks 3 and 4 describe the *stochastic differential equation* approach to stochastic modeling that relates any diffusion to a corresponding Brownian motion.

5 Exercises

1. (formulas and properties of Brownian motion)

   (a) Use the reasoning of Section 3 to show that if $0 \leq t_0 \leq \cdots \leq t_k$ then the increments $\Delta Y_k = Y_{t_{k+1}} - Y_{t_k}$ converge to independent Gaussians with mean zero and variance $t_{k+1} - t_k$. For the rest of this exercise, assume that X_t has this distribution.

   (b) Suppose that X_t is Brownian motion as defined in Section 3. Suppose $t_1 < t_2 < t_3$ and consider the increments $Z_1 = X_{t_2} - X_{t_1}$, $Z_2 = X_{t_3} - X_{t_2}$, and $Z_3 = X_{t_3} - X_{t_1}$. Show that the distribution of Z_3 as an increment of Brownian motion (a Gaussian with mean zero and variance $t_3 - t_1$) agrees with the distribution of Z_3 as the sum of Z_1 and Z_2 and depending on the joint distribution of Z_1 and Z_2.

   (c) Suppose $t_1 < t < t_2$. This means that t may be written as a *convex combination* of t_1 and t_2 as

   $$t = \lambda t_1 + (1 - \lambda) t_2 \ , \ \ 0 \leq \lambda \leq 1 \ .$$

   (t is a weighted average of t_1 and t_2 with t_1 getting weight λ.) Use notation $(U, V, W) = (X_{t_1}, X_{t_2}, X_{t_3})$. Write a formula for the joint PDF: $(U, V, W) \sim p(u, v, w)$. Get this from the distribution of U and then the conditional distribution of V given U, then the conditional distribution of W given U and V. Why does the latter depend only on V?

   (d) Use the result of part (c) to find the conditional distribution of the interior value V, conditional on knowing the endpoint values $U = u$ and $W = w$. Find the conditional mean and variance of V in terms of u, w, λ, and $t_3 - t_1$. In particular, show that the conditional mean is given by interpolation and the conditional variance is less than the conditional variance of V given only $U = u$. Why?

2. A linear combination of Gaussians is Gaussian if the terms in the sum are jointly Gaussian. Suppose X_t is Brownian motion and consider the integral

   $$Z = \int_0^T X_t \, dt \ .$$

   Show that Z is Gaussian and find its mean and variance. For this, make a Riemann sum approximation to the integral and use the properties of the numbers X_{t_k}.

3. (Brownian motion has "rough paths")

   (a) Suppost $Z \sim \mathcal{N}(0, \sigma^2)$. Find a formula for $\mathrm{E}[\,|Z|\,]$.

(b) Suppose we add the sizes of the increments of Brownian motion, giving the *rectified* sum

$$R = \sum_{k=1}^{n} |\Delta X_k| \ .$$

For this formula, take $\Delta X_k = X_{t_{k+1}} - X_{t_k}$, with $t_k = k\Delta t$ and $\Delta t = T/n$. Find $E[R]$ and show $E[R] \to \infty$ as $\Delta t \to 0$. The average speed, if it could be defined, would be

$$\text{average speed} = \lim_{\Delta t \to 0} \frac{1}{T} \sum_{k=1}^{n} |\Delta X_k| \ .$$

Show that the average speed of a Brownian motion path is infinite. (Note: this is one question being asked several ways for emphasis.)

4. Let $p(x,t)$ be the PDF of Brownian motion, X_t. Use the properties of Brownian motion to write a formula for p. Show by direct calculation that this satisfies the *forward equation*

$$\partial_t p(x,t) = \frac{1}{2} \partial_x^2 p(x,t) \ .$$

5. The *exponential* random variable is a positive random number with PDF $u(t) = \lambda e^{-\lambda t}$ if $t > 0$ and $u(t) = 0$ if $t < 0$. Call this distribution $\text{Exp}(\lambda)$. Download and run the code `ExponentialSampler.py`. This is Python 3. It uses Numpy version 18 or later. If you get an error message because of the random number generator, you may need to update your Numpy. I strongly advise using Python 3 in command line mode rather than in an IDE. That's what most serious developers do. You should get some output at the terminal, a plot should pop up, and it should write a plotfile `ExponentialHistogram.pdf`. You have to close the popup file manually each time you run the code.

   (a) Show that $\text{E}[T] = \frac{1}{\lambda}$.

   (b) Show that if U is a standard uniform ($U \in [0,1]$ uniformly distributed), then $T = -\frac{1}{\lambda}\log(U)$ has $T \sim \text{Exp}(\lambda)$. The random number generator `rg.random()` returns a standard uniform. This explains line 32 in `ExponentialSampler.py`.

   (c) Consider the conditional distribution of $T \sim \text{Exp}(\lambda)$ conditional on $T > t_s$. Show that this conditional distribution is $t_s + \text{Exp}(\lambda)$. This means that conditional on $T > T_s$, the extra time $T - t_s$ is also exponential with the same rate. The exponential distribution is used to model how long it takes a new lightbulb (or any device) to break. Explain that in the exponential model, a lightbulb that has not broken yet is as good as new.

(d) Explain why the procedure `sim(..)` produces T that with the conditional distribution of part (c).

(e) Let T_k be an independent sequence $T_k \sim \text{Exp}(\lambda)$. Let $N = \min\{k | T_k > t_s\}$. Calculate $\text{E}[N]$ as a function of λ and t_s. Show that the number produced by the code `ExponentialSampler.py` agrees with this.

6. Modify the code from Exercise 5 to study the hitting time related to exponential random variables. Define

$$R_n = \sum_{k=1}^{n} T_k \ , \quad T_k \sim \text{Exp}(\lambda) \ \text{ i.i.d.}$$

From Exercise 5 part (a), we know $\text{E}[R_n] = \frac{n}{\lambda}$. Consider a positive "gap" g, and consider

$$N = \min\left\{ n \mid R_n > \frac{n}{\lambda} + g \right\} \ .$$

Consider the random variables $Y_k = T_k - \frac{1}{\lambda}$ and $S_n = R_n - \frac{n}{\lambda}$. let X_t be a standard Brownian motion with $X_0 = 0$ and $\text{var}(X_t) = t$. The first *hitting time* (also called *first passage time*) at $M > 0$ is

$$T = \min\{t \mid X_t \geq M\} \ .$$

We will see (Week 3) that T has PDF

$$u(t) = \frac{M}{\sqrt{2\pi t^3}} e^{-\frac{M^2}{2t}} \ .$$

Use this density and the Brownian motion scaling of the random walk S_n to estimate the distribution of N when g is large, with λ fixed. You may assume that the hitting time N of the random walk, properly scaled, is approximately related to the corresponding hitting time for Brownian motion.

Modify `ExponentialSampler.py` from Exercise 5 to test this theoretical prediction. Modify the function `sim` to simulate a first hitting time for the random walk. Make a histogram of the random times N. Make bins of width L, so that the *bin counts* are $B_j = \#\{k \mid jL \leq N_k < (j+1)L\}$. Make a plot showing the empirical bin counts and the theoretical prediction. Make a table of this information. The Brownian motion approximation is valid when g is large, because that forces a large number of steps before $R_n \geq \frac{1}{\lambda} + g$. The histogram will be too noisy to be useful if L is too small. If the bins are too large, then there will not be enough bins to be an interesting test of the theory. Once your code is running, experiment with parameters to get as good a agreement with theory as you can.

Your code must follow style rules followed by `ExponentialSampler.py`. These include: all floating point number output must be formatted. Never use [str(x) if x is a floating point number in the code you upload. You

are free to do that while debugging. Comment a lot. Make the comments useful. Use white space to make things line up vertically as much as possible to make the code easy to read, Use a docstring for any function. Put your name and contact information at the top, along with when and why you wrote the code. Put relevant numbers in the plot title and legends. Do not "hard wire" code parameters. Every code parameter should have a variable name and an assignment statement with a comment. Tabular output should be aligned under table headings.