QUEUEING THEORY AND EPIDEMIC MODELING(COMING SOON!)

ANDREW LEE

Abstract. This is an incomplete draft of the paper. In trying to make this document (relatively) self-contained, and also teach myself the background knowledge for this topic, I ended up being rather ambitious with my goals. The scope of the paper was roughly outlined as follows: 1) preliminaries in probability, 2) preliminaries in stochastic processes and Poisson processes, 3) the basics of queue models, 4) "case studies" with $M/G/1$, $M/M/1$, $M/M/c$, and $M/G/c$ queue models, and finally 5) applications to epidemic modeling. The current document only contains the work done so far on parts 1 and 3, along with some appendices as necessary. The appendix on measure theory is not complete, and will contain a brief section on integration (Lebesgue and Riemann-Stieltjes).

Contents

Date: Fall 2020.

1. Probability

1.1. Probability Spaces and Random Variables.

In this section, we will only deal with discrete probability spaces for the sake of simplicity. However, much of the remainder of this paper will rely on probability spaces in a more general fashion, requiring the use of measure theory. For those unfamiliar with measure theory, Appendix A provides a bare-bones barrage of definitions to introduce the notion of a probability space in terms of measure theory.

Definition 1.1.1. A *discrete probability space* is a pair $(\Omega, \mathbb{P})$, where Ω is a set and $\mathbb{P}\colon \Omega \to [0,1]$ is a function such that

$$\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1.$$

The set Ω is called the *state space*, and the function $\mathbb{P}$ is called the *probability mass function*.

Definition 1.1.2. Let $(\Omega, \mathbb{P})$ be a probability space. An *event* $E \subseteq \Omega$ is a subset of the state space. The probability that an event occurs is given by

$$\mathbb{P}(E) = \sum_{\omega \in E} \mathbb{P}(\omega).$$

In general, a probability space also consists of a σ-algebra $\mathcal{F}$ over Ω, with $\mathbb{P}$ being a measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}(\Omega) = 1$. In this context, an event is an element $E \in \mathcal{F}$ of the σ-algebra. Here, $\mathcal{F}$ is called the *event space*. Note that in the discrete case, the event space $\mathcal{F} = \mathcal{P}(\Omega)$ is the power set of the state space, since any subset of Ω can be a well-defined event. In the remainder of this paper, we will refer to a probability space including the σ-algebra $\mathcal{F}$. That being said, the intuition from the discrete version still carries through.

Given our probability space, it is often helpful to assign each element of the state space a value; for example, H for heads and T for tails, or the numbers 1 through 6 for the six possible results of a dice roll. We call this a random variable.

Definition 1.1.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and $(E, \mathcal{E})$ a measurable space. A $(E, \mathcal{E})$*-valued random variable* is a measurable function[1] $X\colon \Omega \to E$. Then, for some subset $S \in \mathcal{E}$, the probability of X being in S, denoted $\mathbb{P}(X \in S)$, is defined as the probability of the event $X^{-1}(S)$ occurring:

$$\mathbb{P}(X \in S) \coloneqq \mathbb{P}(X^{-1}(S)) = \sum_{\substack{\omega \\ X(\omega) \in S}} \mathbb{P}(\omega).$$

Of course, when $S = \{s\}$ is a singleton set, we write $\mathbb{P}(X = s) \coloneqq \mathbb{P}(X \in \{s\})$.

Intuitively, a random variable is a way to label each element of our state space Ω with elements of E. However, unintuitively, a random variable is neither random nor a variable.

In our everyday mathematical lexicon, a variable is a deterministic quantity. If we let x be the solution to some equation, it can only take on one value, and once we know what it is, *we know what it is.* This may seem obvious, but if we consider a random variable X,

[1]The term *measurable function* is a formality to ensure that the definition is well-behaved with respect to a general not-necessarily-discrete probability space. For a discrete probability space, every function $X\colon \Omega \to E$ is measurable regardless of the nature of E. In the words of measure-theoretic jargon, this is true because the σ-algebra of the discrete probability space is the power set of the state space, meaning that every preimage is necessarily measurable. For more information, see Appendix A.

observing that $X = 2$ in one instance doesn't rule out the possibility that $X = 19$ in another. Algebraically speaking, having a variable that changes its value according to its own whims is rather inconvenient. Fortunately, mathematicians have a wonderful name for things that change value depending on how you evaluate them: *functions*.

Of course, this is *very* closely related to how we define random variables (as functions). But, even if we called them random functions, that would still be inaccurate, because they aren't random. Despite colloquial usage, randomness is not synonymous with being unpredictable; in fact, randomness is anything but unpredictable.

Mathematicians often use the word *random* to describe processes that have an aspect of chance, but ultimately abide by some sense of order. For example, I have no idea what the result of my next coin flip would be, but I would expect to eventually get a roughly equal number of heads and tails if I continue to flip coins for the rest of my life, based on my knowledge that coin flips follow a 50/50 distribution.[2] However, random variables do not fit this moniker: given an element of my state space $\omega \in \Omega$, we know *exactly* what value $X(\omega)$ will take. In this sense, random variables are, in fact, deterministic functions![3]

Random variables are called "random" not because of their own nature, but rather because of the underlying randomness of the choice of $\omega \in \Omega$ from the state space. In the remainder of this section, we will start to see why we speak of them as variables.

1.2. **Probability and Distribution Functions.**

As we continue, we will primarily be concerned with real-valued random variables. That is, random variables where $E = \mathbb{R}$ and $\mathcal{E} = \mathcal{B}$, the Borel σ-algebra.

In the context of a random variable $X \colon \Omega \to E$ over a discrete probability space $(\Omega, \mathbb{P})$, the probability mass function $\mathbb{P}$, or PMF for short, is an apt way to describe the system. Given an element $x \in E$, we can extract meaningful information from $\mathbb{P}(X = x)$ simply as the probability that X takes on the value x. For example, consider a fair dice. Given the discrete probability space $(\Omega, \mathbb{P})$ where $\Omega = \{1, 2, \ldots, 6\}$ and $\mathbb{P} = \frac{1}{6}$ for all $\omega \in \Omega$, the probability that $X = 1$ gives us a very good sense of how often we will roll a 1: around one-sixth of the time.

However, this breaks down if we consider the probability of picking 0.5 out of the interval $[0, 1]$. Obviously, that event has probability zero, and in fact, the probability of picking any $x \in [0, 1]$ is zero. So, we define some new functions to better understand general probability spaces.

Definition 1.2.1. Let X be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *cumulative distribution function* $F_X \colon \mathbb{R} \to \mathbb{R}$ of a random variable X, or CDF for short, is defined as

$$F_X(x) = \mathbb{P}(X \leq x).$$

Moreover, the *complementary cumulative distribution function*, or CCDF[4] for short, is defined as

$$\overline{F}_X(x) = 1 - F_X(x).$$

It turns out that CDFs have some nice properties.

[2]Of course, this would be a colossal waste of time, so the author would like to invite the reader to test this hypothesis as an exercise.

[3]Given the definition, this should be utterly unsurprising.

[4]In other disciplines, this may also be called the *tail distribution*, the *exceedance*, the *reliability function*, or the *survival function*, denoted $S(x)$.

Proposition 1.2.2. *Let F_X be a CDF. Then, F_X has the following properties.*

(1) *(Monotonicity) $a \leq b$ implies $F_X(a) \leq F_X(b)$,*
(2) *(Right Continuity) For all $a \in \mathbb{R}$, $\lim_{x \to a^+} F_X(x) = F_X(a)$,*
(3) *(Limits) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.*

Moreover,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a).$$

These CDFs allow us to classify our random variables.

Definition 1.2.3. A random variable $X \colon \Omega \to E$ is said to be *discrete* if E is countable. Moreover,

$$F_X(x) = \mathbb{P}(X \leq x) = \sum_{y \leq x} \mathbb{P}(X = y).$$

The equation $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$ should reek of the fundamental theorem of calculus, and for good reason.

Definition 1.2.4. A random variable X is said to be *continuous* if there exists a function $f_X(x)$, called the *probability density function*, or PDF, such that

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx.$$

It follows, then, from the fundamental theorem of calculus, that if X is continuous, its PDF can be determined as the derivative of the CDF; that is,

$$f_X(x) = \frac{d}{dx} F_X(x) \qquad \text{and} \qquad F_X(x) = \int_{-\infty}^x f_X(t)\, dt.$$

Of course, if the PDF exists (and is integrable), it determines the CDF, so we often describe random variables in terms of their PDFs.

The use of derivatives and integrals turns out to be surprisingly common within this field, as we will see in the remainder of this section.

1.3. **Expected Value, Variance, and Moments.** Given a random variable X, it is often very helpful to look at its average behavior. To this effect, we will look at the expected value.

Definition 1.3.1. Let $X \colon \Omega \to \mathbb{R}$ be a real-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The *expected value* of X, denoted $\mathbb{E}[X]$, is

$$\mathbb{E}[X] \coloneqq \int_\Omega X(\omega)\, d\mathbb{P}(\omega).^5$$

If X is discrete with probability mass function $\mathbb{P}_X(x) = \mathbb{P}(X = x)$,

$$\mathbb{E}[X] = \sum_x x\, \mathbb{P}_X(x),$$

[5]This integral is something called a Riemann-Stieltjes integral over the state space Ω. This is the rigorous, measure-theoretic definition of the expected value, but slightly overkill for our purposes, so we'll stick to the definitions given for the discrete and continuous cases.

and if X is continuous with PDF $f_X(x)$, then the expected value can be calculated as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx.$$

In general, we sometimes will denote the *mean* as $\mu = \mathbb{E}[X]$.

A convenient property of the expected value is that it is linear, which follows from the linearity of the integral.

Proposition 1.3.2 (Linearity of Expectation)**.** *Let X and Y be two random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then,*

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \qquad \textit{and} \qquad \mathbb{E}[cX] = c\mathbb{E}[X].^{6}$$

Another important quantity used to study a random variable is the *variance*, which gives a measure of how much a set of numbers spread out from their average value.

Definition 1.3.3. The *variance* of a random variable X with mean $\mu = \mathbb{E}[X]$ is defined as

$$\mathrm{Var}(X) \coloneqq \mathbb{E}[(X-\mu)^2].$$

The variance is also denoted σ^2, since the variance yields the square of the standard deviation.

By linearity of expectation, we can expand the above expression:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[(X-\mu)^2] \\ &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2\end{aligned}$$

Therefore, the variance is equal to the expected value of the square of X minus the square of the mean. In the discrete and continuous cases, we can simplify the definition of variance into the sums/integrals, as follows.

Proposition 1.3.4. *Let X be a discrete random variable on a discrete probability space $(\Omega, \mathbb{P})$. Then, the variance can be defined two equivalent ways:*

$$\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2] = \sum_x \mathbb{P}_X(x)(x-\mu)^2 = \left(\sum_x \mathbb{P}_X(x)x^2\right) - \mu^2 = \mathbb{E}[X^2] - \mu^2.$$

Otherwise, if X is a continuous random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the variance can be expressed as follows.

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2.$$

In general, the mean and variance are two examples of moments, which are different values defined in terms of X and the expectation operator $\mathbb{E}[-]$. Before we consider moments, however, let us first prove a quick fact.

Lemma 1.3.5. *Let X be a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose $g\colon \mathbb{R} \to \mathbb{R}$ is a measurable function. Then, $g(X)$ is also a random variable.*

[6]Considering that random variables are simply deterministic functions, it should be clear that we can add them and multiply them by scalars without trouble. Moreover, we can multiply them (pointwise) without trouble, as well.

Proof. First recall that a random variable is simply a measurable function $X\colon \Omega \to \mathbb{R}$. Since g is a measurable function, it follows that their composition $g \circ X$ is also a measurable function $g \circ X\colon \Omega \to \mathbb{R}$. This can be shown quickly; recall that a function is measurable if the preimage of a measurable set is a measurable set. Let A be an arbitrary measurable set in the codomain of g. Then, its preimage $g^{-1}(A)$ is also measurable, due to the measurability of g. Then, via the measurability of X, $X^{-1}(g^{-1}(A))$ is also measurable. Since $g \circ X$ is a measurable function, it follows that $g(X)$ is a random variable. ■

This lemma tells us that $g(X) = X + Y$, $g(X) = X + c$, and $g(X) = XY$ are all random variables. In this sense, a random variable can be treated like a usual variable, since we can perform all these kinds of manipulations on X.

Of course, since $g(X)$ is a random variable, we can take its expectation. The formal treatment of this following result is a little advanced and relies on measure theory, but the end result is that we can evaluate $\mathbb{E}[g(X)]$ in terms of regular Riemann integration.

Theorem 1.3.6 (Law of the Unconscious Statistician)**.** *Let $X\colon \Omega \to \mathbb{R}$ be a random variable with CDF $F_X(x)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $g\colon \mathbb{R} \to \mathbb{R}$ be a measurable function with respect to the Borel σ-algebra on $\mathbb{R}$. Then*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\, dF_X(x).$$

Recall that in the general case, the expected value is defined as

$$\mathbb{E}[g(X)] = \int_{\Omega} g(X(\omega))\, d\mathbb{P}(\omega).$$

This theorem tells us that we can change from integrating over a measure $\mathbb{P}_(\omega)$ over Ω and instead integrate over the real line. Moreover, in the continuous case, this theorem tells us that

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx,$$

and our "definition"[7] of the expected value in the continuous case comes out as the special case where $g(x) = x$. Of course, this gives us the freedom to consider other functions g, such as $g(X) = X^n$ or $g(X) = (X - \mu)^n$. These are called the moments of X.

Definition 1.3.7. Let X be a continuous random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with PDF f_X. The *nth moment of X* is given by

$$\mathbb{E}[X^n] = \int_{-\infty}^{\infty} x^n f_X(x)\, dx.$$

Similarly, the *nth central moment of X*, where $\mu = \mathbb{E}[X]$ is the first moment, is given by

$$\mathbb{E}[(X - \mu)^n] = \int_{-\infty}^{\infty} (x - \mu)^n f_X(x)\, dx.$$

[7]Air quotes because the definition of expected value does not take this form for the continuous case; it's actually a result of the not-actually-trivial-to-prove law of the unconscious statistician. In fact, the law is named as such because statisticians are often unconscious that what they consider as definition actually is a theorem that requires rigorous proof.

Of course, the mean is the first moment, and the variance is the second central moment. Some quick integration should prove some more results about other moments. The zeroth moment is 1, by definition of the integral of a PDF being a CDF. For the same reason, the zeroth central moment is 1. The first central moment is zero, since of course the mean will not vary with respect to the mean. Going further, the third and fourth central moments can be used to define skewness and kurtosis, respectively. These are beyond the scope of this paper, but of course I will not stop an interested reader from going to learn more.

Though the above examples of $g(X)$ are definitely useful ways to extract information about X, it turns out that there are other functions g that give very powerful tools, some of which can be used to compute these moments.

1.4. **Integral Transforms.**

Let us first begin with the motivation for using integral transforms, in general. At its core, problem-solving in mathematics depends upon seeing the problem from different angles. Integral transforms achieve this very task.

Roughly speaking, an integral transform takes a function in one "space" and transforms it into another function in a second "space," which is easier to work with. Then, by manipulating and solving the equation in the second "space," we can take the inverse transform to get the correct answer in the original space. The archetypal example is the Laplace transform $\mathcal{L}$ on differential equations, which takes a differential equation from the "time space" to the "frequency space." This has the effect of, roughly speaking, turning differentiation and integration into simple multiplication and division by an operator variable s. Of course, the algebra in the frequency space turns out to be much simpler, making the Laplace transform very useful for solving differential equations, as shown in Figure 1.

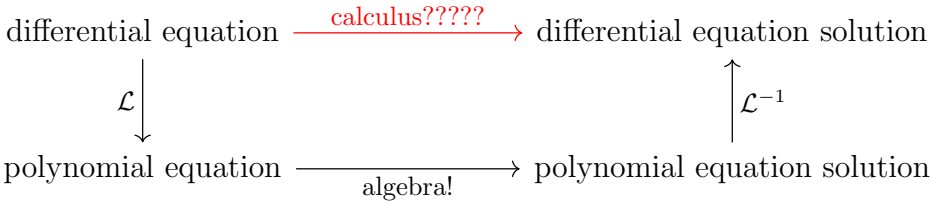

Figure 1. A schematic for using the Laplace transform to solve a differential equation.

Now, we shall continue with transforms on random variables. First, we will consider a simpler case, where X is discrete (and, for simplicity, non-negative).

Definition 1.4.1. Let $X\colon \Omega \to \mathbb{Z}_{\geq 0}$ be a nonnegative discrete random variable on $(\Omega, \mathbb{P})$, where $\mathbb{P}_X(k) = \mathbb{P}(X = k)$ is the probability mass function. Then, the *z-transform*[8] of X is defined by

$$\mathcal{Z}\{\mathbb{P}_X\}(z) = \mathbb{E}[z^X] = \sum_{k=0}^{\infty} \mathbb{P}_X(k) z^k.$$

In this case, our function is $g(X) = z^X$. Here, the values z^k are the values that our random variable z^X is taking, with probability $\mathbb{P}_X(k)$.

[8]This may also be called the probability generating function of X, in which case it may be written as $G(z)$. But, $\mathcal{Z}$ looks really cool, and z-transform sounds infinitely cooler than probability generating function, so we're going to stick to that.

At this point, you may be wondering, well, what's that z doing there? Well, it turns out that we can let z be any complex number, which means the z-transform is defined over the whole complex plane. We will not be dealing with the intricacies of this fact here, but there is something convenient about z being able to take different values in $\mathbb{C}$, as we will see.

Note that $\mathcal{Z}\{\mathbb{P}_X\}(0) = \mathbb{P}_X(0)$, and $\mathcal{Z}\{\mathbb{P}_X\}(1) = 1$. Of course, this isn't particularly interesting, but let us consider taking the derivative:

$$\mathcal{Z}\{\mathbb{P}_X\}'(z) = \sum_{k=0}^{\infty} k\mathbb{P}_X(k)z^{k-1}.$$

Letting $z = 1$, we see that

$$\mathcal{Z}\{\mathbb{P}_X\}'(1) = \sum_{k=0}^{\infty} k\mathbb{P}_X(k) = \mathbb{E}[X]$$

is the expected value!

We can do a similar process for the nth moment, simply taking the second derivative. Note that in order to get a term of k^2 instead of $k(k-1)$, we have to first multiply by z on both sides:

$$\begin{aligned} z\mathcal{Z}\{\mathbb{P}_X\}'(z) &= \sum_{k=0}^{\infty} k\mathbb{P}_X(k)z^{k} \\ (z\mathcal{Z}\{\mathbb{P}_X\}'(z))' &= \sum_{k=0}^{\infty} k^2\mathbb{P}_X(k)z^{k-1} \\ z\mathcal{Z}\{\mathbb{P}_X\}''(z) + \mathcal{Z}\{\mathbb{P}_X\}'(z) &= \sum_{k=0}^{\infty} k^2\mathbb{P}_X(k)z^{k-1} \end{aligned}$$

Of course, this means that we can plug in $z = 1$ to get

$$\mathbb{E}[X^2] = \mathcal{Z}\{\mathbb{P}_X\}''(1) + \mathcal{Z}\{\mathbb{P}_X\}'(1)$$

and therefore

$$\mathrm{Var}(X) = \mathcal{Z}\{\mathbb{P}_X\}''(1) + \mathcal{Z}\{\mathbb{P}_X\}'(1) - \left(\mathcal{Z}\{\mathbb{P}_X\}'(1)\right)^2.$$

Now, using the z-transform to get the moments is nice and all, but an astute (read: lazy) reader might wish for a simpler formulation. And, in fact, for many random variables, there is a nice function to generate the moments of X, aptly called a *moment-generating function*.

Definition 1.4.2. Let X be a nonnegative real-valued random variable. The *moment-generating function*, denoted $M_X(t)$, is

$$M_X(t) \coloneqq \mathbb{E}[e^{tX}].$$

A very nice property of the moment-generating function is that $M_X^{(n)}(0)$, where $f^{(n)}$ is the nth derivative, is equal to the nth moment. This is not too surprising, given some quick

back-of-the-napkin algebra:

$$\begin{aligned} M_X^{(n)}(t) = (\mathbb{E}[e^{tX}])^{(n)} &= \left(\mathbb{E}\left[1 + tX + \frac{t^2X^2}{2!} + \cdots\right]\right)^{(n)} \\ &= \left(1 + t\mathbb{E}[X] + \frac{t^2}{2!}\mathbb{E}[X^2] + \cdots\right)^{(n)} \\ &= \mathbb{E}[X^n] + t\mathbb{E}[X^{n+1}] + \frac{t^2}{2!}\mathbb{E}[X^{n+2}] + \cdots \end{aligned}$$

and therefore that

$$M_X^{(n)}(t) = \mathbb{E}[X^n].$$

The moment-generating function is nice in that it works for any sort of random variable. However, the integral corresponding to the moment-generating function does not always converge (owing to the e^t term), so we often also use something called the Laplace-Stieltjes transform.[9]

Definition 1.4.3. Let X be a nonnegative continuous real-valued random variable with CDF F_X. The *Laplace-Stieltjes transform* of X, denoted $\mathcal{L}_X(s)$, is

$$\mathcal{L}_X^*(s) = \mathbb{E}[e^{-sX}] = \int_0^\infty e^{-sx} f_X(x)\, dx.$$

Note that $\mathcal{L}_X^*(-t) = M_X(t)$. Moreover, we can see the Laplace-Stieltjes transform as the continuous version of the z-transform, substituting $z = e^{-s}$. Of course, we can use the Laplace-Stieltjes transform to get the moments:

$$\mathbb{E}[X^n] = (-1)^n \mathcal{L}_X^{(n)}(0).$$

1.5. **Discrete Random Variables and Their Distributions.**

After that flurry of abstract constructions, let us define some specific random variables. Each of these random variables will have some number of *parameters* and a probability mass function $\mathbb{P}$. Then, we can state their means and variances, followed by some examples and properties of interest. Though it will not be done here, we highly encourage our readers to use the methods of z-transforms or moment-generating functions to derive these quantities for themselves.

Moreover, in each definition, the codomain of the random variable is specified to be the *support*; that is, the set on which the probability mass function is not zero.

The (discrete) uniform distribution is fairly simple to understand, having equal probability of returning one of n consecutive values.

Definition 1.5.1. The *discrete uniform distribution*, denoted $\text{Unif}_D(a, b)$, has two parameters, $a, b \in \mathbb{Z}$ with $a < b$. A random variable $X \colon \Omega \to \{a, a+1, \ldots, b\}$ where $X \sim$

[9]Another solution to this problem would be to substitute *it* for t, turning into something called the characteristic function. The characteristic function of a (continuous) random variable would then be the Fourier transform of its PDF. This is, of course, incredibly interesting, but we will not discuss this in this paper.

$\mathrm{Unif}_D(a,b)$[10] has PMF

$$\mathbb{P}_X(k) = \begin{cases} \dfrac{1}{b-a+1} & \text{if } k = a, a+1, \ldots, b, \\ 0 & \text{otherwise.} \end{cases}$$

We often define a new value $n \coloneqq b - a + 1$, since there are n equally likely outcomes. For this distribution, the mean is $\mu = \frac{a+b}{2}$ and the variance is $\sigma^2 = \frac{n^2-1}{12}$.

For example, the distribution of a dice roll would be $\mathrm{Unif}_D(1,6)$.

The Bernoulli distribution encodes the result of a weighted coin toss, where the coin turns up heads with probability p and tails with probability $1-p$ (encoded numerically as 1 and 0, respectively). We call such a coin a p-coin.

Definition 1.5.2. The *Bernoulli distribution*, denoted $\mathrm{Bern}(p)$, has one parameter $0 \le p \le 1$. A random variable $X \colon \Omega \to \{0,1\}$ where $X \sim \mathrm{Bern}(p)$ has PMF

$$\mathbb{P}_X(1) = p \qquad \text{and} \qquad \mathbb{P}_X(0) = 1 - p.$$

For this distribution, the mean is $\mu = p$ and the variance is $\sigma^2 = p(1-p)$.

The binomial distribution is the distribution of the number of heads when flipping a p-coin n times; that is, the distribution of n independent Bernoulli trials.

Definition 1.5.3. The *binomial distribution*, denoted $B(n,p)$ has two parameters $n \in \mathbb{Z}_{\ge 0}$ and $0 \le p \le 1$. A random variable $X \colon \Omega \to \{0,1,\ldots,n\}$ where $X \sim B(n,p)$ has PMF

$$\mathbb{P}_X(k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

For this distribution, the mean is np and the variance is $np(1-p)$.

It follows that a $B(1,p)$ random variable is a $\mathrm{Bern}(p)$ random variable. In this sense, the Bernoulli distribution is a special case of the binomial distribution.

The geometric distribution, like the binomial distribution, is also based on Bernoulli trials. But this time, we stop at the first success.

Definition 1.5.4. The *geometric distribution*[11], denoted $\mathrm{Geom}(p)$, has one parameter $0 \le p \le 1$. A random variable $X \colon \Omega \to \mathbb{Z}_{>0}$ where $X \sim \mathrm{Geom}(p)$ has PMF

$$\mathbb{P}_X(k) = (1-p)^{k-1} p.$$

The geometric distribution has mean $\frac{1}{p}$ and variance $\frac{1-p}{p^2}$.

The geometric distribution exhibits a property that is very important in queueing theory, the property of being *memoryless*.

Definition 1.5.5. A discrete positive random variable X is said to be *memoryless* if its future behavior does not depend on its past behavior. Stated formally, that means that the

[10]The notation $X \sim \mathrm{Unif}_D(a,b)$ means that X is distributed according to the distribution $\mathrm{Unif}_D(a,b)$.

[11]The geometric distribution can also be defined as the number of failures before the first success, rather than the first success. This definition gives a random variable $Y = X - 1$. We will use the first convention.

probability that X takes on a value greater than $m+n$ given that X is already greater than n is equal to the probability that X is greater than m; that is,

$$\mathbb{P}(X > m+n \mid X > n) = \mathbb{P}(X > m).^{12}$$

It should be clear that a geometric random variable is memoryless. After all, the probability that the next m trials will be failures should not depend on my previous n failures. It turns out that the geometric random variable is the only discrete random variable with this property. For a proof of this fact, refer to Appendix 1.7.

The Poisson distribution is a very important distribution in probability, and it counts the probability of a given number of events occurring in a fixed interval of time, given that these events occur independent of each other and with a constant mean rate.

Definition 1.5.6. The *Poisson distribution*, denoted $\text{Pois}(\lambda)$, has one parameter $\lambda > 0$. A random variable $X\colon \Omega \to \mathbb{Z}_{\geq 0}$ where $X \sim \text{Pois}(\lambda)$ has PMF

$$\mathbb{P}_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Interestingly, the Poisson distribution has both mean and variance equal to $\mu = \sigma^2 = \lambda$.

For example, if I receive an average of 100 emails a day, then the distribution of the number of emails I get should follow a Poisson distribution, since the number of emails I receive on Tuesday should be independent of the number of emails I received last Thursday.

The Poisson distribution is also fairly closely related to the binomial distribution, in fact, by fixing $\lambda = np$ and letting $n \to \infty$ and $p \to 0$, the Poisson distribution is a good approximation for the binomial distribution. We state this result as follows.

Theorem 1.5.7 (Poisson limit theorem)**.** *Let $\mathbb{P}_{B(n,p)}(k)$ and $\mathbb{P}_{\text{Pois}(\lambda)}(k)$ be the probability mass functions for the binomial and Poisson distributions with their respective parameters, respectively:*

$$\mathbb{P}_{B(n,p)}(k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \textit{and} \qquad \mathbb{P}_{\text{Pois}(\lambda)}(k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Then, fixing $\lambda = np$ and taking the limit $n \to \infty$,

$$\mathbb{P}_{\text{Pois}(\lambda)}(k) = \frac{\lambda^k e^{-\lambda}}{k!} = \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} = \lim_{n\to\infty} \mathbb{P}_{B(n,\lambda/n)}(k).$$

Therefore, for large n, the binomial distribution resembles a Poisson distribution with mean $\lambda = np$.

This is sometimes called the law of rare events, since fixing $\lambda = np$ while taking $n \to \infty$ makes $p \to 0$.

Proof. We wish to show that

$$\frac{\lambda^k e^{-\lambda}}{k!} = \lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k}.$$

[12]The notation $\mathbb{P}(A \mid B)$ for two events A and B is the *conditional probability of A given B*. The equation for this is given by Bayes' Law:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)\,\mathbb{P}(B \mid A)}{\mathbb{P}(B)}.$$

We do so by manipulating the right-hand side, then taking the limit. First, let us fix $\lambda = np$, and substitute $p = \frac{\lambda}{n}$.

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} &= \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-k+1)}{n^k}\right) \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}\end{aligned}$$

Taking limits over each portion in parentheses, we see that

$$\lim_{n\to\infty} \left(\frac{n(n-1)\cdots(n-k+1)}{n^k}\right) = \lim_{n\to\infty} \left(\frac{n^k + O(n^{k-1})}{n^k}\right) = 1$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^{-k} = 1.$$

Therefore,

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{k} p^k (1-p)^{n-k} &= \frac{\lambda^k}{k!}(1)(e^{-\lambda})(1) \\ &= \frac{\lambda^k e^{-\lambda}}{k!}.\end{aligned}$$

And the result holds.

■

1.6. **Continuous Random Variables and Their Distributions.**

Since all of these random variables are continuous, they have both a CDF and a PDF. We will list both, but a continuous random variable is often thought of more in terms of the PDF rather than the CDF. The first (perhaps obvious) example is the continuous uniform distribution, the continuous analogue of the discrete uniform distribution.

Definition 1.6.1. The *continuous uniform distribution*, denoted $\text{Unif}_C(a,b)$, has two parameters $a, b \in \mathbb{R}$ with $a < b$. A random variable $X\colon \Omega \to [a,b]$ where $X \sim \text{Unif}_C(a,b)$ has PDF

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \le x \le b, \\ 0 & \text{otherwise} \end{cases}$$

and CDF

$$F_X(x) = \begin{cases} 0 & \text{for } x < a, \\ \dfrac{1}{b-a}(x-a) & \text{for } a \le x \le b, \\ 1 & \text{for } x > b. \end{cases}$$

Just as in the discrete case, the distribution has mean $\mu = \frac{a+b}{2}$ and variance $\sigma^2 = \frac{(b-a)^2}{12}$.

The exponential distribution is named for its distribution functions, which have exponential components. It has the key property of being memoryless, making it a continuous analogue of the geometric distribution.

Definition 1.6.2. The *exponential distribution*, denoted $\text{Exp}(\lambda)$, has one parameter $\lambda > 0$. A random variable $X : \Omega \to [0, \infty)$ where $X \sim \text{Exp}(\lambda)$ has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0 \end{cases}$$

and CDF

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

For the exponential distribution, it is sometimes easier to work with the CCDF

$$\overline{F}_X(x) = \begin{cases} e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

The mean is $\frac{1}{\lambda}$, and the variance is $\frac{1}{\lambda^2}$.

The parameter λ is often called the *rate parameter*. A *scale* parameter μ is sometimes used instead, where $\mu = \frac{1}{\lambda}$. The fact that the exponential distribution is memoryless follows from its CCDF. In fact, it turns out that the exponential random variable is the only memoryless random variable. For a proof, refer to Appendix 1.7.

The Erlang distribution is another random variable used quite often in queueing theory. Intuitively, the Erlang distribution is the distribution of some number of iid exponential random variables. The distribution is named for A. K. Erlang, who worked on modeling telephone traffic via queueing theory.

Definition 1.6.3. The *Erlang distribution*, denoted $\text{Er}(k, \lambda)$, has two parameters, a shape parameter k and a rate parameter $\lambda > 0$. A random variable $X : \Omega \to [0, \infty)$ where $X \sim \text{Er}(k, \lambda)$ has PDF

$$f_X(x) = \begin{cases} \dfrac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0 \end{cases}$$

and CDF

$$F_X(x) = \begin{cases} 1 - \displaystyle\sum_{n=0}^{k-1} \frac{(\lambda x)^n e^{-\lambda x}}{n!} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

Again, the CCDF takes the slightly nicer form

$$F_X(x) = \begin{cases} \displaystyle\sum_{n=0}^{k-1} \frac{(\lambda x)^n e^{-\lambda x}}{n!} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0. \end{cases}$$

The mean is $\frac{k}{\lambda}$, and the variance is $\frac{k}{\lambda^2}$.

The Erlang distribution is a generalization of the exponential distribution, as the distributions $\text{Er}(1, \lambda)$ and $\text{Exp}(\lambda)$ are equivalent. That is, the exponential distribution is the Erlang distribution with shape parameter $k = 1$.

1.7. **Memoryless Random Variables.**

Memoryless random variables are extremely important in queueing theory. In this section, we will discuss the geometric and exponential random variables. We begin by stating the two main propositions.

Proposition 1.7.1. *Let X be a discrete random variable. Then, X is memoryless if and only if $X \sim \text{Geom}(p)$.*

Proposition 1.7.2. *Let X be a continuous random variable. Then, X is memoryless if and only if $X \sim \text{Exp}(\lambda)$.*

It follows that the geometric (resp. exponential) distributions is the only memoryless discrete (resp. continuous) random variable. Let us first restate the definition of memorylessness in the discrete case.

Definition 1.7.3. A discrete positive[13] random variable X is said to be *memoryless* if

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n)$$

for $m, n \in \mathbb{Z}_{\geq 0}$.

Proof of Proposition 1.7.1. First we show the forward direction. Suppose X is memoryless; that is,

$$\mathbb{P}(X > m + n \mid X > m) = \mathbb{P}(X > n). \tag{1.7.1}$$

Then, we can define the CCDF of X:

$$\overline{F}_X(x) = \mathbb{P}(X > x).$$

Using Bayes' Law, we can simplify the LHS of (1.7.1), resulting in the expression

$$\frac{\mathbb{P}(X > m + n)\,\mathbb{P}(X > m \mid X > m + n)}{\mathbb{P}(X > m)} = \mathbb{P}(X > n).$$

Of course, $\mathbb{P}(X > m \mid X > m + n) = 1$, so we have the relation

$$\mathbb{P}(X > m + n) = \mathbb{P}(X > n)\mathbb{P}(X > m).$$

Stated in terms of the CCDF, the CCDF has to follow the relation

$$\overline{F}_X(m + n) = \overline{F}_X(m)\overline{F}_X(n) \tag{1.7.2}$$

Letting $m = n = 1$, it follows that

$$\overline{F}_X(2) = \overline{F}_X(1)^2.$$

Moreover, we can use the recursive relation

$$\overline{F}_X(n + 1) = \overline{F}_X(n)\overline{F}_X(1)$$

to prove the explicit expression

$$\overline{F}_X(n) = \overline{F}_X(1)^n \tag{1.7.3}$$

using induction. From the definition of the CDF, we can see that

$$\mathbb{P}(X = n) = \mathbb{P}(n - 1 < X \leq n) = F_X(n) - F_X(n - 1) = \overline{F}_X(n - 1) - \overline{F}_X(n).$$

[13]The derivation is slightly neater with positive random variables, but a simple shift by 1 will let us derive this for nonnegative random variables.

In terms of $\overline{F}_X(1)$,

$$\mathbb{P}(X = n) = \overline{F}_X(1)^{n-1} - \overline{F}_X(1)^n = (\overline{F}_X(1))^{n-1}(1 - \overline{F}_X(1)).$$

Of course, substituting $1 - p = \overline{F}_X(1)$, we get that

$$\mathbb{P}(X = n) = (1 - p)^{n-1}p$$

and thus X is geometrically distributed.

Going in the other direction, it is not difficult to show that the geometric distribution follows the relations (1.7.2) and (1.7.3), with $p = 1 - \overline{F}_X(1)$. Thus the proposition is true. ■

The proof of the continuous case is similar, so we will present a slightly condensed version here.

Proof of Proposition 1.7.2. The definition for a memoryless continuous positive random variable is practically identical to the discrete case, except that now our random variables are (positive) real-valued.[14] From (1.7.2), we can conclude that for two nonnegative reals $s, t \in \mathbb{R}_{\geq 0}$,

$$\overline{F}_X(s + t) = \overline{F}_X(s)\overline{F}_X(t). \tag{1.7.4}$$

This implies that, just as before,

$$\overline{F}_X(t) = \overline{F}_X(1)^t. \tag{1.7.5}$$

Then, substituting $e^{-\lambda} = \overline{F}_X(1)$ for $\lambda > 0$ (which implies that $0 < \overline{F}_X(1) < 1$), we get that

$$\overline{F}_X(t) = e^{-\lambda t}, \tag{1.7.6}$$

which is exactly the CCDF of the exponential distribution.

A *huge* caveat to this proof is the step from (1.7.4) to (1.7.5) and consequently (1.7.6). The equation in (1.7.4) is closely related to something called the *Cauchy functional equation*, which has many famously pathological solutions. We will not cover the rigorous basis for this step in this paper, but its validity comes from the fact that a CCDF is, by definition, monotonic, continuous, bounded, and positive. These conditions are enough to restrict the solutions of (1.7.4) to just the exponential.

Of course, just as before, it is trivial to check the reverse direction. ■

[14]Technically, since the inequalities are strict, the domain of the CCDF includes 0, even though 0 is not in the support of the random variable; i.e. $f_X(0) = 0$.

2. Queue Models

Queueing theory has become an extremely useful framework to study the formation, function, and congestion of queues. Of course, queues are called lines in the American lexicon, but it is a truth universally acknowledged that "queueing theory" sounds more official than "lining theory."

The most simple kind of queue consists of two primary sections: the server, and the queue of customers awaiting service. For example, the server could be a bank teller, and the queue would be a line of people asking the bank for loans. Or, the server could be a computer, processing packets of data. In fact, the interpretation for a queueing model is usefully vague, allowing us to apply the same framework for different systems.

A useful motivating question in queueing theory is asking the question of when queues actually form.

Let us consider a fast food restaurant that is able to serve 600 customers per hour. (It's not called fast food for no reason.) Of course, if there are more than 600 customers arriving per hour, then the queue will eventually build up and get arbitrarily long. By the nature of these things, everyone who stands in line will eventually be served, but obviously this doesn't make for an interesting experience, both on the part of the people waiting in line and the mathematician eagerly watching from the side.

Now, suppose that there are less than 600 customers arriving per hour. Then, the service rate is faster than the arrival rate, so what's the problem? Well, the answer lies in the variability of these arrivals. Of course, our restaurant could easily handle 300 customers arriving at equally spaced intervals over an hour, but they would be hard-pressed to provide timely service if a nearby math conference had a lunch break and 300 hungry mathematicians decided to test their Chicken McNugget conjectures and arrived in one big pack.

Of course, the distributions of arrival and serving can be whatever they want, so we use a form of notation called Kendall notation as a shorthand for a particular queueing process. Traditionally, this notation contains three slots, but there is also a six-part versions with more information. We will most often use the three-part notation in this paper, but we will begin with a full treatment of the full version.

2.1. **Kendall Notation.**

Kendall notation specifies six different factors that govern a queueing process: time between arrivals A, service time distribution S, service capacity c, queue capacity K, population of jobs N, and queueing discipline D. All together, a specific queue can be characterized by these six factors, and will be notated $A/S/c/K/N/D$. This section will define all of these terms.

The first factor is the *time between arrivals*. This factor describes the distribution of the inter-arrival times. Of the many ways this can be distributed, we are primarily interested in the D, M, GI, and G cases. D stands for a degenerate[15] process, where the inter-arrival times are constant and therefore utterly uninteresting. M stands for Markovian (or memoryless), denoting exponentially distributed inter-arrival times. Of course, this means that the arrivals follow a Poisson process. More specifically, we can also have an M_t process, where the inter-arrival times are exponentially distributed according to a time-varying parameter $\lambda(t)$. GI stands for general independent, which means that the time between arrivals can be iid

[15]Appropriate synonyms with the same first letter (written in decreasing order of seriousness) include: deterministic, defective, deject, depraved, disappointing, dreadful, and of course, dumb.

random variables from any probability distribution. Finally, G stands for general, where the inter-arrival times can be dependent.

The second factor is the *service time distribution*. This factor describes the distribution of the service times. Similar to the inter-arrival times, we will only consider M and G.

The third factor is the *service capacity*. Simply put, this denotes the number of servers. In terms of mathematical significance, this is often either 1, some general variable c, or ∞.

The fourth factor is the *queue capacity*. This, of course, denotes the capacity of the queue; if the queue is at capacity, new arrivals are turned down. If left unspecified, the queue capacity is often taken to be unlimited, making $K = \infty$.

The fifth factor is the *population of jobs*. This can be considered as the total population of customers potentially in need of service. The effect of this factor becomes more apparent for smaller populations, since a long queue will reduce the outside population, reducing the effective rate of arrival. If left unspecified, the population is generally taken to be infinite, with $N = \infty$.

The sixth, and final, factor is the *service discipline*. This refers to the order in which the servers provide services for the customers. For example, the queue may have a FIFO (first in, first out) discipline, or a LIFO (last in, first out) discipline.[16] Alternatively, we may have a PS (processor sharing) discipline, where each customer in the queue receives an equal fraction of service. If left unspecified, the service discipline will be FIFO.

Therefore, a queue $A/S/c/K/N/D$ can be determined by specifying each of these variables. If the final three are left unspecified, they are assumed to be $\infty/\infty/$ FIFO.

2.2. **Utilization Level.**

We begin our discussion of queues by discussing the mean behavior of a system, and defining an important metric called the utilization level. Intuitively, the utilization level, denoted ρ, is the percent of time that the server is busy.

Suppose that we have a general $G/G/1$ single-server queue. For this specific queue, we are assuming that we have a steady state; that is, the queue would still behave the same way if we had been running it for infinite time. Then, we will have a well-defined mean arrival rate λ. Denoting S as the random variable for service times, we can denote $\mathbb{E}(S)$ as the mean service time. First, let us restrict ourselves to the case where the mean arrival rate is less than the mean service *rate*; that is,

$$\lambda < \frac{1}{\mathbb{E}(S)}.$$

This can also be expressed as the relation

$$\lambda \mathbb{E}(S) < 1.$$

But, this quantity $\lambda \mathbb{E}(S)$ has much more significance. To see this, let us run this queue for an arbitrarily long time, say L. Then, since we're in a steady state, the number of people that have come in during that time should be equal to the number of people that have come out during that time. Of course, the number of people that have come in is equal to λL.

However, since the servers are not active for the entire time period L, we define ρ as the fraction of time in which the servers actually are active; the servers are active for ρL units

[16]Technically, these terms are only accurate for a single-server queue. In general, for a multi-server queueing process, the analogous disciplines would be FCFS (first come, first served) and LCFS (last come, first served).

of time. Then, we can express the above sentence as the relation

$$\lambda L = \frac{1}{\mathbb{E}(S)}\rho L.$$

Therefore,

$$\rho = \lambda \mathbb{E}(S).$$

Thus, the quantity $\lambda\mathbb{E}(S)$ is precisely the fraction of time that the server is active! This is formally stated in the following definition.

Definition 2.2.1. Let λ be the mean arrival rate and S the random variable associated with service times for a general single-server $G/G/1$ queue. Then, we define the *utilization level*[17] ρ as

$$\rho = \lambda \mathbb{E}(S).$$

Intuitively, the utilization level is the fraction of time that the server is busy. Of course, $1 - \rho$ is the fraction of time that the server is not busy.

Of course, there are other metrics besides the utilization level. For example, one may want to consider the distribution of the waiting time or the sojourn[18] time (the waiting time plus the service time). In general, however, we often like to study *stationary mean performance measures*, such as the mean waiting time or the mean sojourn time in a steady state.

2.3. **Little's Law.**

Little's Law gives us a very important relation between the mean number of customers in the system $\mathbb{E}(L)$, the mean sojourn time $\mathbb{E}(J)$, and the mean rate of arrival.

Theorem 2.3.1 (Little's Law)**.** *Let L, J be random variables corresponding to the number of customers in the system and the sojourn time, respectively. Let λ be the mean rate of arrival. If the limits for λ and $\mathbb{E}(J)$ exist (i.e., $\lambda, \mathbb{E}(J) < \infty$), then the limit for $\mathbb{E}(L)$ exists (i.e., $\mathbb{E}(L) < \infty$) and*

$$\mathbb{E}(L) = \lambda \mathbb{E}(J).$$

As always, we assume that the queue remains in steady-state; i.e. that $\rho = \lambda\mathbb{E}(S) < 1$.

Intuitively, this can be understood as follows. Suppose that each person in the queue pays a dollar per unit of time they spend in the system. Then, there are two ways that they could pay this sum. First, they could pay "continuously" in time, which means that the system would net $\mathbb{E}(L)$ dollars per unit time. Alternatively, they could pay at the end of their tenure in the system. Then, each person would pay $\mathbb{E}(J)$ as they leave the system. In steady-state, the mean rate of arrival is equal to the mean rate of departure, so the mean revenue would be equal to $\lambda\mathbb{E}(J)$. Thus the two quantities are equal.

Little's law is rather simple to state, but surprisingly difficult to prove, and therefore we will leave it as an exercise to the reader...to go to Appendix B and read the proof!

Now, Little's law is rather unsurprising on its own. However, there is a slight nuance that makes it extremely powerful, and that is the flexibility we are afforded with when defining "the system." Of course, we can apply Little's law to the entire queue, including the service portion. Or, we could also consider a *fraction* of the arrivals. For example, if we saw fit, we could divide the arrivals based on the ownership of glasses. Then, regardless of the service

[17]This is also called the utilization rate, traffic intensity, offered load, occupation rate, server utilization, or simply utilization.

[18]This is also called the time in the system, or the delay.

discipline, (which may treat our bespectacled friends with more bountiful generosity; i.e., more slowly) Little's law still applies! That is, the average number of bespectacled customers is equal to *their* average arrival rate multiplied by *their* average sojourn time. This way, we can invoke Little's law for parts of the queue.

Alternatively, we can also invoke Little's law for a part of the physical system. For example, given a single-server queue, we can define the "system" as only the server. Then, the arrival rate is still λ, but this time the random variable J corresponds to the service times S, and the random variable L corresponds to the mean number of servers in the system. Of course, this second part is confusing: what does it mean to consider the mean number of servers in a single-server system? Well, if we expand our perception of what it means to be "in the system," we can say that the server is in the system if they are, well, currently serving someone. Then, $\mathbb{E}(L)$ admits a very simple interpretation: the utilization level ρ! Thus, we find that the equation

$$\rho = \lambda \mathbb{E}(S)$$

is simply a special case of Little's law applied to the server!

2.4. **The PASTA Property.**

It is often said that the intellectual prowess of a field of study can be judged by the quality of its acronyms.[19] To this effect, queueing theory is emerging as *the* powerhouse of academia in the twenty-first century, owing this title purely to the marvelously named PASTA property. Disappointingly, this has nothing to do with the delectable cultural delicacies of the Italian peninsula, but we shall discuss it nonetheless.

Roughly speaking, the PASTA property states that on average, for a Poisson process, a customer entering the queue will find the same situation as an outside observer looking at the queue. This can be summed up by saying that Poisson Arrivals See Time Averages, or of course, PASTA for short.[20]

[19]Nobody says this; I just made it up.

[20]This is incomplete. There will be more PASTA.

Appendix A. Measure Theory

Let us first recall the definition of a measure space, starting from the notions of a σ-algebra and a measure.

First, we define a σ-algebra over a set. Roughly speaking, a σ-algebra is the collection of subsets of our set which are of interest. In general, measure spaces tend not to be well-behaved if every subset of the underlying set is "measurable," so the σ-algebra restricts us to consider only some of those sets, which we know will be well-behaved. Then, a set equipped with a σ-algebra can be further equipped by a measure, to make a measure space.

Definition A.1. A *σ-algebra* over X is a collection $\mathcal{A}$ of subsets of X which obeys the following conditions:[21]

(1) $X \in \mathcal{A}$,
(2) $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$, and
(3) $A_1, A_2, \ldots \in \mathcal{A}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The pair $(X, \mathcal{A})$ is called a measurable space, and elements $A \in \mathcal{A}$ are called measurable sets.

Definition A.2. Let X be a set and $\mathcal{A}$ a σ-algebra on X. A function $\mu \colon \mathcal{A} \to [0, \infty]$ is called a *measure* on $(X, \mathcal{A})$ if it satisfies the following two properties:

(1) $\mu(\varnothing) = 0$,
(2) (Countable Additivity) For a countable collection $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}$ of pairwise disjoint sets in $\mathcal{A}$,

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

Any measurable space $(X, \mathcal{A})$ can be "equipped" with any measure μ, hence the name. Then, by carrying through with this procedure, we obtain a measure space.

Definition A.3. A measure space is a triple $(X, \mathcal{A}, \mu)$ where X is a set, $\mathcal{A}$ is a σ-algebra on X, and μ is a measure on $(X, \mathcal{A})$.

Definition A.4. A probability space is a measure space $(X, \mathcal{A}, \mu)$ where $\mu(X) = 1$. We will often use the notation $(\Omega, \mathcal{F}, \mathbb{P})$ for a probability space.

The notation $(\Omega, \mathcal{F}, \mathbb{P})$ is simply to concretely fix ourselves into the realm of probability. Our set X becomes our state space Ω, our σ-algebra $\mathcal{A}$ becomes an event space $\mathcal{F}$, and of course our measure μ becomes a probability mass function $\mathbb{P}$.

Now, let us consider maps between measurable spaces.

Definition A.5. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces. A function $f \colon X \to Y$ is said to be *measurable* if the preimage of every $\mathcal{B}$-measurable set is $\mathcal{A}$-measurable; that is, if $B \in \mathcal{B}$,

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\} \in \mathcal{A}.$$

[21] These conditions imply that $\mathcal{A}$ contains the empty set and is closed under countable intersections, which is an equivalent definition.

Appendix B. Proof of Little's Law

Proof of Little's Law. This proof is due to [Sti74].

Suppose that customers arrive at a queue system at time points $(t_i) \coloneqq \{t_1, t_2, \ldots\}$, where $0 \le t_1 \le t_2 \le \cdots$. Let $I_n(t)$ be an indicator function for the nth customer; that is,

$$I_n(t) = \begin{cases} 1 & \text{if the } n\text{th customer is in the system at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

Then, we can easily define the sojourn time of the nth customer j_n as the integral of I_n; that is,

$$j_n \coloneqq \int_0^\infty I_n(t)\,dt.$$

Similarly, we can easily define the number of customers in the system at time t, $L(t)$:

$$L(t) \coloneqq \sum_{n=1}^{\infty} I_n(t).$$

Now, suppose we have some $T \ge 0$. Then, it follows that $\int_0^T L(t)\,dt$ is the total amount of time spent in the system in the interval $[0, T]$. We can bound this quantity with two other quantities: $A(t)$ and $D(t)$, denoting the sum of the sojourn times of the customers who have arrived or departed in $[0, t]$, respectively. Let us prove this bound.

Lemma B.1. *For all $T \ge 0$,*

$$D(T) \le \int_0^T L(t)\,dt \le A(T).$$

Proof of Lemma B.1. By definition, $j_n = \int_0^\infty I_n(t)\,dt$, so it follows that

$$j_n \ge \int_0^T I_n(t)\,dt,$$

with equality if $t_n + j_n \le T$. Moreover, we can take the integral of L as follows.

$$\int_0^T L(t)\,dt = \int_0^T \sum_{n=1}^{\infty} I_n(t)\,dt = \sum_{n=1}^{\infty} \int_0^T I_n(t)\,dt$$

We can write $A(t)$ and $D(t)$ as follows:

$$A(t) = \sum_{\substack{n \\ t_n \le T}} j_n \qquad \text{and} \qquad D(t) = \sum_{\substack{n \\ t_n + j_n \le T}} j_n.$$

Since $\int_0^T I_n(t)\,dt = 0$ if $t_n > T$, it follows that

$$\sum_{n=1}^{\infty} \int_0^T I_n(t)\,dt = \sum_{\substack{n \\ t_n \le T}} \int_0^T I_n(t)\,dt.$$

Thus,

$$\int_0^T L(t)\,dt \le \sum_{\substack{n \\ t_n \le T}} j_n = A(t).$$

Similarly, since $j_n = \int_0^T I_n(t)\,dt$ when $t_n - j_n \le T$, it follows that

$$D(t) = \sum_{\substack{n \\ t_n + j_n \le T}} j_n = \sum_{\substack{n \\ t_n + j_n \le T}} \int_0^T I_n(t)\,dt \le \sum_{\substack{n \\ t_n \le T}} \int_0^T I_n(t)\,dt = \int_0^T L(t)\,dt,$$

since $t_n + j_n \le T$ is a more strict condition than $t_n \le T$. Thus the inequality is true. ■

Intuitively, this makes sense: if we think of each consumer paying one dollar per unit of time they spend in the system, then $S(T)$ is the cost if each customer pays the lump sum at their arrival, $D(T)$ is the cost if each customer pays the lump sum at their departure, and $\int_0^T L(t)\,dt$ is the cost if each consumer pays continuously.

Let $t_0 = 0$ and $N(t) = \max\{n \mid t_n \le t\}$ be the cumulative number of arrivals by time t. Moreover, define λ, $\mathbb{E}(J)$, and $\mathbb{E}(L)$ as follows.

$$\lambda = \lim_{T\to\infty} \frac{N(T)}{T}, \qquad \mathbb{E}(J) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} j_i, \qquad \mathbb{E}(L) = \lim_{T\to\infty} \frac{1}{T}\int_0^T L(t)\,dt.$$

By the assumptions of the theorem, we know that $\lambda < \infty$ and $\mathbb{E}(J) < \infty$; both of the limits exist. Moreover, we can say more, with the lemma below.

Lemma B.2. *If $\lambda < \infty$ and $\mathbb{E}(J) < \infty$, then*

$$\lim_{n\to\infty} \frac{j_n}{t_n} = 0.$$

The proof of this lemma is presented as Lemma 4 in [Sti72].

Using this lemma, we can finish the proof. Since t_n is well defined and finite for all n, it follows that $T \to \infty$ as $N(T) \to \infty$. Therefore, we can rewrite $\mathbb{E}(J)$ as

$$\mathbb{E}(J) = \lim_{T\to\infty} \frac{1}{N(T)} \sum_{i=1}^{N(T)} j_i.$$

Then,

$$\lambda\mathbb{E}(J) = \lim_{T\to\infty} \frac{N(T)}{T}\frac{1}{N(T)} \sum_{i=1}^{N(T)} j_i = \lim_{T\to\infty} \frac{1}{T} \sum_{i=1}^{N(T)} j_i = \lim_{T\to\infty} \frac{1}{T} A(T).$$

Therefore, from Lemma B.1, it suffices to show that

$$\lim_{T\to\infty} \frac{1}{T} A(T) = \lim_{T\to\infty} \frac{1}{T} D(T),$$

which in turn implies that both are equal to

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T L(t)\,dt = \mathbb{E}(L)$$

by squeezing.

From Lemma B.2, we know that $\lim_{n\to\infty} \frac{j_n}{t_n} = 0$. Let $\varepsilon > 0$. Then, there exists an N such that $n > N$ implies that $\frac{j_n}{t_n} < \varepsilon$. Pick a large T, such that $N(T) > N$. Then,

$$D(T) = \sum_{\substack{n \\ t_n + j_n \le T}} j_n = \sum_{\substack{n \le N \\ t_n + j_n \le T}} j_n + \sum_{\substack{n > N \\ t_n + j_n \le T}} j_n.$$

Since $j_n < \varepsilon t_n$,

$$D(T) \geq \sum_{\substack{n \leq N \\ t_n + j_n \leq T}} j_n + \sum_{\substack{n > N \\ t_n + \varepsilon t_n \leq T}} j_n = \sum_{\substack{n \leq N \\ t_n + j_n \leq T}} j_n - \sum_{\substack{n \leq N \\ t_n + \varepsilon t_n \leq T}} j_n + \sum_{\substack{n \\ t_n \leq \frac{T}{1+\varepsilon}}} j_n.$$

The first and second term in the RHS are bounded by $\sum_{n \leq N} j_n$. Moreover, note that due to Lemma B.2,

$$\lim_{T \to \infty} \frac{\sum_{n \leq N} j_n}{T} = 0,$$

since $t_n \to \infty$ as $T \to \infty$. Therefore,

$$\lim_{T \to \infty} \frac{A(T)}{T} \geq \lim_{T \to \infty} \frac{D(T)}{T} \geq \frac{1}{1+\varepsilon} \lim_{T \to \infty} \frac{A(T)}{T}.$$

Since ε was arbitrary, it follows that the limits are equal and the result follows. ■

References

[AR15] Ivo Adan and Jacques Resing. "Queueing Systems". Mar. 26, 2015.

[Ber06] Ryan Berry. "Queueing Theory". 2006.

[Hav13] Moshe Haviv. *Queues: A Course in Queueing Theory*. Springer New York, May 20, 2013. ISBN: 978-1-4614-6765-6.

[HS+10] Carlos Hernandez-Suarez et al. "An application of queuing theory to SIS and SEIS epidemic models". In: *Mathematical Biosciences and Engineering* 7.4 (2010), pp. 809–823. DOI: 10.3934/mbe.2010.7.809.

[Pal+20] Sergio Palomo et al. "Flattening the Curve: Insights From Queueing Theory". In: (Apr. 20, 2020). URL: http://arxiv.org/abs/2004.09645.

[Ros95] Sheldon M. Ross. *Stochastic Processes*. 2nd edition. New York: Wiley, Feb. 8, 1995. ISBN: 978-0-471-12062-9.

[Sti72] Shaler Stidham. "$L = \lambda W$: A Discounted Analogue and a New Proof". In: *Operations Research* 20.6 (Dec. 1, 1972), pp. 1115–1126. DOI: 10.1287/opre.20.6.1115.

[Sti74] Shaler Stidham. "Technical Note—A Last Word on $L = \lambda W$". In: *Operations Research* 22.2 (Apr. 1, 1974), pp. 417–421. DOI: 10.1287/opre.22.2.417.

[TB09] Pieter Trapman and Martinus C. J. Bootsma. "A useful relationship between epidemiology and queueing theory: The distribution of the number of infectives at the moment of the first detection". In: *Mathematical Biosciences* 219.1 (May 2009), pp. 15–22. DOI: 10.1016/j.mbs.2009.02.001.

[Zuk20] Moshe Zukerman. "Introduction to Queueing Theory and Stochastic Teletraffic Models". Feb. 21, 2020. URL: http://arxiv.org/abs/1307.2968.