Bacterial Growth: Not So Simple

John Chase[1] and Matthew Wright[2]
April 2021

When students encounter exponential functions for the first time, bacterial growth is often used as a standard example because it seems straightforward. However, we will show that hidden assumptions in the time-to-division for individual bacteria lead to unexpected consequences in the exponential population model.

We begin with the following simple-sounding problem:

The Classic Problem. Suppose a bacterium has an average division time of 1 hour. Write a model that gives the population size after t hours if the initial population is 1 bacterium.

The naïve solution to this problem is that the population size is given by $P(t) = 2^t$. However, this model is incorrect because it fails to consider the inherent randomness of the splitting times. In our example, some bacteria will divide in less time than one hour and some will divide in more. This seems like a minor issue that would still give rise to approximately the same model, but we will find that this is not the case at all. Small amounts of randomness in the splitting time for bacteria leads to faster than expected growth in the population size.

Deterministic Time-to-Division

Our first goal is to reexamine the naïve solution to the classic problem and uncover any hidden assumptions. We will see that assuming a deterministic time-to-division, in which each bacteria splits at exactly 1 hour, quickly leads to absurdity.

Notice first, with this assumption, if someone asks for the population after 30 minutes, our answer must still be "1 bacterium," since the first bacterium hasn't yet split. Likewise, at 1 hour and 30 minutes ($t = 1.5$), the population must have size 2, and so on. This behavior is captured in the exponential step function shown in Figure 1.

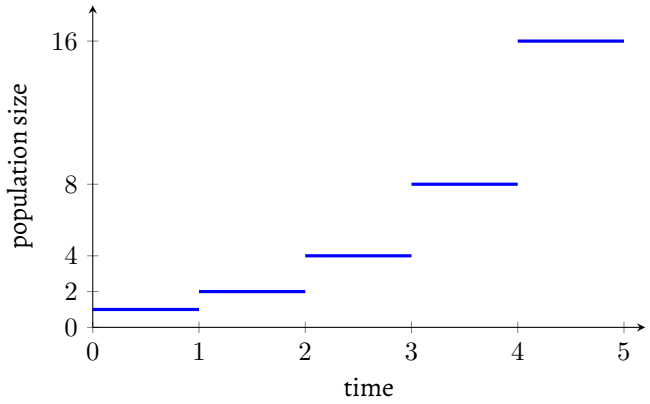

Figure 1: Bacterial population growth with deterministic splitting times—an "exponential step function."

[1]Walter Johnson High School, Bethesda MD, john.chase@mcpsmd.net
[2]St. Olaf College, Northfield, MN, wright5@stolaf.edu

You might say, "This doesn't bother me, it's still roughly exponential and agrees with our classic model at all positive integer values." Yet this model is still deeply troubling for three reasons:

1. We know from empirical data that the behavior of the population smooths out because of randomness and eventually approaches something that is more continuous in nature [3, 6]. It seems absurd, especially as the population gets very large, that there would be hour-long stretches of time in which no single bacterium splits.

2. The deterministic model makes the unstated assumption that the current "age" of the first bacterium is zero. We need to know more than just how often a bacterium splits; we also need to know exactly when it first comes into existence. If we start with a population of size 10, then we need to assume that all ten bacteria have an "age of zero" and split after exactly 1 hour. It seems absurd that the deterministic model requires us to assume a current age for each member of the bacteria colony. Alternatively, perhaps the ages of each member of the original colony follow some distribution; then the aggregate growth is a sum of individual step functions of the type that appear in Figure 1. Unfortunately, this simply creates multiple subcolonies, each with the dilemma in item 1.

3. It turns out that even a small amount of randomness in the splitting times results not in smoothing out the growth function in Figure 1, but in much faster aggregate growth. To the authors, this seems to be a relatively unknown phenomenon. The main goal of this article is to substantiate this point with analysis and simulation.

Some readers, after hearing reason 3 above, might ask if the underlying distribution of splitting time matters, not just the mean splitting time. Indeed, the underlying splitting distribution is critical, and so we present the following definition.

Definition. The *underlying splitting distribution* is the distribution of the random variable that gives the time from the birth of a single bacterium until it divides. If the distribution has mean μ, then the *underlying splitting rate* is $2^{1/\mu} - 1$ and the *underlying growth factor* is $2^{1/\mu}$.

Said another way, "underlying splitting rate" refers to the the average number of bacteria that are "added" to a single bacterium per unit of time, and "underlying growth factor" refers to the average *multiplicative* growth that a single bacterium experiences per unit of time. For example, in the previous discussion of the classic problem, the assumed underlying splitting distribution is deterministic with underlying splitting rate 1 and growth factor 2. Each hour every bacterium becomes 2 bacteria.

What happens to the overall model if we instead use an underlying distribution that is nondeterministic? We now consider underlying distributions that are random.

Exponentially Distributed Time-to-Division

An exponential model is a good first guess for the underlying splitting distribution, since we often find that exponential distributions govern lifespans, time to failure, or time intervals between random (Poisson-distributed) events. Simple properties of exponential distributions lead us to a basic understanding of the resulting population growth function, which we confirm by simulation.

In this context, the time until any given bacterium splits will be modeled by an exponential random variable with rate $\lambda = 1$ (mean $1/\lambda = 1$). That is, let the random variable X_i be the time until division for the ith bacterium, with $X_i \sim \text{Exp}(1)$. We assume that the lifespans of all bacteria are independent and have identical exponential distributions.

If there are n bacteria at a given time, then the amount of time until the next division will be the minimum of all the X_i. Next, we invoke the *memoryless* property of the exponential distribution. A distribution has this property if the probability of an event happening after time b is the same as the probability of the event happening after time $a + b$ given that time a has elapsed. The fact that the exponential distribution is memoryless implies we do not have to worry about the time each bacterium was "born."

Let the random variable Y_n represent the time until the next division. Thus, Y_n is given by

$$Y_n = \min\{X_1, X_2, \ldots, X_n\}.$$

As is well known [4, Sec. 4.9], the minimum of n independent exponential variables with rate λ is exponential with rate $n\lambda$. We have $\lambda = 1$, so Y_n has an exponential distribution with rate n.

Since $Y_n \sim \text{Exp}(n)$, when the population is of size n it grows at approximately rate n. This suggests that 2^t is not a good model for the population size, because this would imply that the growth rate at population size n would be $n \ln(2)$. Recalling that $\frac{dn}{dt} = n$ is satisfied by e^t, we suppose that population size might be approximated by e^t. Given this intuition, we now propose the population model

$$Q(t) = e^t.$$

We use simulation to compare population growth with the models $P(t) = 2^t$ and $Q(t) = e^t$. Figure 2 shows the growth of five simulated populations, each with underlying splitting distribution $\text{Exp}(1)$. These plots confirm that $P(t) = 2^t$ severely underestimates the population growth, while $Q(t) = e^t$ seems to be a roughly average growth curve for these populations. In the following sections, we verify this result with more rigor.

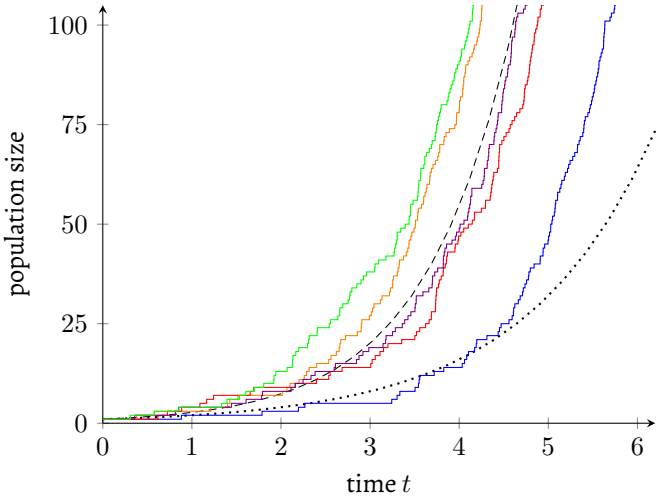

Figure 2: Simulated growth of five populations, each with underlying splitting distribution $\text{Exp}(1)$. The model e^t (plotted as a dashed curve) appears to be an average case, while the naïve model 2^t (dotted curve) underestimates the population growth.

At this point, we need to highlight an apparent paradox. For any individual bacterium, the average time-to-division is 1 hour. This seems to indicate that the population will double at about time $t = 1$, and this is what the naïve $P(t) = 2^t$ model shows. But according to our $Q(t) = e^t$ model, the population doubles at time $t = \ln(2) \approx 0.69$. A typical population doubles *before* the expected splitting time of the bacteria!

One way to see what's really happening here is to consider the *median* of the exponential distribution instead of the expected value. The median of $\text{Exp}(1)$ is, in fact, $\ln(2)$. In other words, if we consider a large number of populations, each of which starts with a single bacterium at time $t = 0$, approximately half of these populations will grow to size two (or more) by time $t = \ln(2) \approx 0.69$, even though the expected time of the first split is $t = 1$! The paradox is resolved when we understand that, even though the average splitting time is 1 hour, most bacteria (in fact, about 63 percent of them) split earlier than this. With exponential time-to-division, a small portion of the bacteria have a long splitting times, but many split quickly.

We see that the underlying *distribution* for the splitting time matters quite a bit. Our model $Q(t)$ based on exponentially distributed time-to-division gives an aggregate growth factor of $e \approx 2.718$ each hour rather than 2. With the understanding that the underlying splitting distribution results in a different aggregate rate, we introduce the following helpful definition:

Definition. If population size can be approximated by the exponential function $y = ab^t$, with $b > 1$, then we call this the *aggregate growth function*. Furthermore, the *aggregate growth factor* is b and the *aggregate growth rate* is $b - 1$.

The definitions here are meant to be parallel to the definitions given earlier for underlying splitting distribution, underlying growth factor, and underlying growth rate. Once again, b conveys the multiplicative growth and $b-1$ conveys the additive growth per bacterium.

In the present discussion, the underlying splitting distribution is exponential with growth factor 2 and rate 1, and the aggregate growth function is $Q(t) = e^t$ with aggregate growth factor e and aggregate growth rate $e - 1 \approx 1.718$.

Having looked at underlying splitting times that are governed by an exponential distribution, we might wonder what aggregate growth functions arise from other underlying distributions. We now turn to this question.

Gamma Distributed Time-to-Division

When we look at bacterial growth in the laboratory, we actually find that the time-to-division for bacteria is more closely modelled by a gamma distribution [1, 3, 6]. This also makes sense a priori, since the gamma distribution is the sum of exponential distributions, and bacterial maturation can be thought of as a sum of independent exponential distributions, each governing a part of development.

Let the splitting time for each bacterium have a gamma distribution with density function $\frac{x^{\alpha-1}}{\Gamma(\alpha)\beta^\alpha}e^{-x/\beta}$ for $x > 0$, where $\Gamma(\alpha)$ denotes the gamma function. With this parametrization, the gamma distribution has mean $\alpha\beta$ (equivalently, rate $\frac{1}{\alpha\beta}$) and variance $\alpha\beta^2$.

For the simplicity of our model, we assume that when a bacterium divides, it "dies" and is replaced by two new bacteria.

Let the random variable $X_i \sim \mathrm{Gamma}(\alpha, \beta)$ be the time from the birth of the ith bacterium until it dies. We call X_i the time-to-division of the ith bacterium. Thus, $X_1, X_2, \ldots$ are independent identically distributed random variables from the $\mathrm{Gamma}(\alpha, \beta)$ distribution. Note that the X_i do not have the memoryless property.

Let c_i be the birth time of the ith bacterium, and we assume that $c_1 = 0$. Then the first bacterium dies at time X_1, at which time the second and third bacteria are born. Thus $X_1 = c_2 = c_3$.

Let T_n be the time at which the population reaches size n. Suppose that at some instant, the population size is n, and the indexes of the currently-alive bacteria are $i_1, i_2, \ldots, i_n$. Then the time of the next split, when the population grows to size $n + 1$, is

$$T_{n+1} = \min_{k=1,\ldots,n} \{c_{i_k} + X_{i_k}\}.$$

Because the gamma distribution does not have the memoryless property, no closed-form expression for the random variable T_n is available. So we employ simulation to further our analysis.

Assume, for example, that the underlying splitting distribution is $\mathrm{Gamma}\left(2, \frac{1}{2}\right)$. For this distribution, the mean splitting time for one bacterium is 1 hour, so we can easily compare with an exponential random variable with rate 1 division per hour. Our simulation chooses a $\mathrm{Gamma}\left(2, \frac{1}{2}\right)$ random splitting time for each new bacterium and when that time elapses, two new random splitting times are initiated. We continue the simulation in this manner, recording the time at which the population reaches each size. We average our results over $10{,}000$ simulations to approximate the expected population size at any time. Since the Gamma distributions are not memoryless, the initial age of the bacterium matters; thus, we choose the age of the initial bacterium for each population from the ages of bacteria alive at the end of a previous simulation.

We likewise simulate the population size with the time-to-division given by other gamma distributions of the form $\mathrm{Gamma}\left(\alpha, \frac{1}{\alpha}\right)$. Each of these gamma distributions has mean 1, so that we can maintain the comparison to the 1 division per hour rate from the exponential distribution.

Figure 3 displays the average population growth curves resulting from these simulations for six values of α. We fit an exponential function $\hat{y}(t) = r^t$ to each of these curves, as shown in Table 1. We see that when the underlying splitting distribution is a gamma distribution, the population size is well approximated by an exponential function.

Moreover, Table 1 shows that the gamma distributions give rise to a continuum of exponential functions. For $\mathrm{Gamma}(1, 1)$ we see the aggregate growth factor is very close to e, which is exactly as expected,

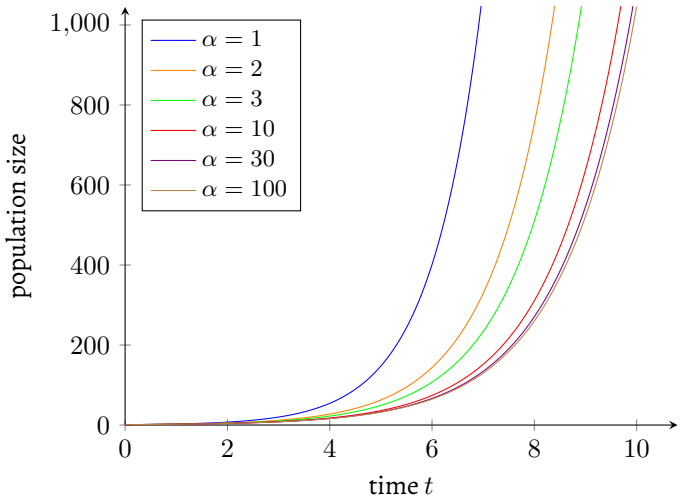

Figure 3: Average population size over time when underlying splitting distribution is $\text{Gamma}(\alpha, \frac{1}{\alpha})$. Each curve is the average of 10,000 simulations.

Underlying Splitting Distribution	Aggregate Growth Function
Deterministic	$P(t) = 2^t$
$\text{Exp}(1)$	$Q(t) = e^t$
$\text{Gamma}(1, 1)$	$\hat{y}(t) = 2.718^t$
$\text{Gamma}(2, \frac{1}{2})$	$\hat{y}(t) = 2.290^t$
$\text{Gamma}(3, \frac{1}{3})$	$\hat{y}(t) = 2.181^t$
$\text{Gamma}(10, \frac{1}{10})$	$\hat{y}(t) = 2.050^t$
$\text{Gamma}(30, \frac{1}{30})$	$\hat{y}(t) = 2.016^t$
$\text{Gamma}(100, \frac{1}{100})$	$\hat{y}(t) = 2.004^t$

Table 1: Letting the underlying splitting distribution vary gives rise to different aggregate growth functions. The first two results in the table are from earlier in this article and were established analytically. The other results were obtained through simulation and regression with $0 \leq t \leq 10$.

since the $\text{Gamma}(1, 1)$ distribution is the same as the $\text{Exp}(1)$ distribution! As α increases, the aggregate growth factor decreases towards 2. This also makes sense, because the variance of the $\text{Gamma}(\alpha, \frac{1}{\alpha})$ is given by $\alpha\beta^2 = \frac{\alpha}{\alpha^2} = \frac{1}{\alpha}$. As α increases, this variance decreases towards zero, and thus the population behaves more and more like our deterministic model.

Therefore, when the time-to-division is governed by an exponential random variable or a gamma random variable, we obtain faster overall population growth than in our deterministic model. In practice, values of α and β that give a near-zero variance, such as those shown in the bottom of Table 3, are unreasonable. Bacterial splitting times are well-fitted by Gamma distributions with parameters such as $\alpha = 9$ and $\beta = 0.6$ [3]. For parameters such as these, the fact that the aggregate growth functions are so different from the deterministic model is significant. This substantiates our point that that even small amounts of randomness in splitting times produce faster aggregate population growth.

Arbitrarily Distributed Time-to-Division

Our simulations reveal that when the underlying splitting distribution is a gamma distribution, the overall population size is eventually well-approximated by an exponential function, with an average aggregate

growth factor that is determined by the underlying splitting distribution. What if we choose a different distribution, such as a Weibull distribution, as the underlying splitting distribution? It turns out that regardless of our particular choice of underlying splitting distribution, the population size will be approximately an exponential function for large times t.

In the literature on stochastic processes, the bacteria-splitting scenario is known as a *Bellman-Harris Branching Process* [2, 5]. For such a process, let $Z(t)$ be the aggregate population growth function, and let b be the aggregate growth factor. It can be shown that if the underlying splitting distribution is continuous and the underlying growth factor is larger than 1, then the ratio

$$\frac{Z(t)}{b^t} \tag{1}$$

converges with probability 1 to a random variable W with expected value 1 [2, 5].

Informally, this says that for any underlying splitting distribution (subject to mild constraints), the aggregate growth function of any individual population will converge to wb^t for some positive number w. Furthermore, if we examine many populations and average their values of w, we will get approximately 1. In other words, the average of the aggregate growth functions of many populations, at a large time value t, will be approximately b^t. In effect, we have already seen this in Figure 3, where we averaged the aggregate growth functions of 10,000 simulations and obtained curves that were very close to exponential functions.

As we have seen in the previous section, the aggregate growth factor b is not obvious from the underlying splitting distribution. However, the following result from the theory of branching processes is critical here [5]. It turns out that the aggregate growth factor b and the underlying splitting distribution with density function $g(x)$ have the relationship

$$2\int_0^\infty \frac{g(x)}{b^x}dx = 1. \tag{2}$$

The leading 2 above is due to our assumption that each bacterium splits into exactly two new bacteria.

To demonstrate this result, we use Equation (2) and numerical integration to compute the aggregate growth factor when the underlying splitting distribution is $\text{Gamma}\left(3, \frac{1}{3}\right)$, and we find that $b \approx 2.18096$. We then simulate 10,000 populations with this underlying splitting distribution, letting each grow until $t = 8$, and record the ratio from Equation (1), $Z(8)/b^8$, which is an instance of the random variable W. Figure 4 shows a histogram of these 10,000 ratios, which reveals the shape of the probability density function of W. The mean of these 10,000 ratios is 0.996, which is close to 1, just as the theory promises!

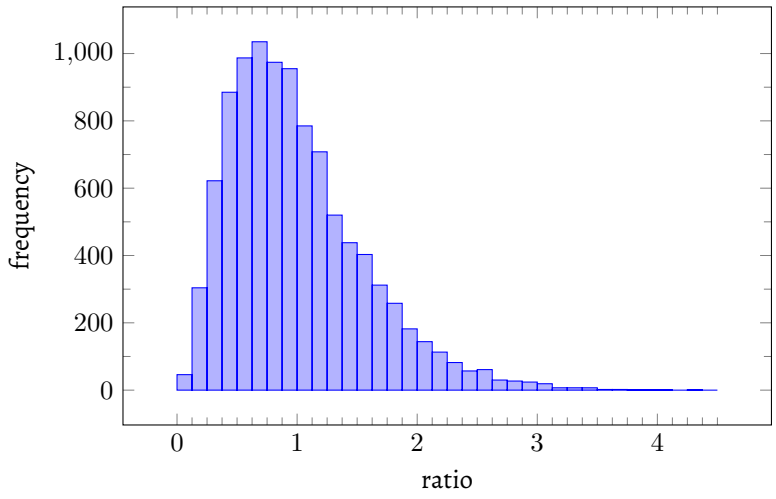

Figure 4: Histogram of 10,000 ratios $Z(t)/b^t$ computed from simulations of populations with underlying splitting distribution $\text{Gamma}\left(3, \frac{1}{3}\right)$ at $t = 8$. The mean of these ratios is $0.996 \approx 1$.

Furthermore, Equation (2) provides a theoretical basis for the results from Table 1. It can be checked using numerical integration that the aggregate growth factors in the table are approximately the values that satisfy Equation (2) with the density functions for each of the gamma distributions considered.

Conclusion

We repeat how surprising and important these results are, considering how often bacterial growth is used as a standard example of exponential growth. What could be simpler than bacteria that split once an hour, we say? We have shown, however, that populations grow faster than expected when even a small amount of randomness is introduced in the time-to-division. It's not just a matter of "smoothing out" the deterministic solution to the classic problem. Exponential growth does result from realistic scenarios, but the connection between the underlying splitting distribution and the aggregate growth function is much more intricate than it seems at first glance.

It is also worth taking a moment to recognize the real world context in which these models live. It goes without saying that our models assume bacterial growth is not limited by environmental factors. In true bacterial growth, we may see an initial lag, a period of exponential growth, a "stationary phase" in which environmental factors limit population growth, and possibly a death phase in which the population dies out. It also should be mentioned that we have used simple parameters for our underlying splitting distributions that are useful for revealing the underlying mathematics. A more realistic underlying splitting distribution might be $\mathrm{Gamma}(9, 0.6)$ [3], which results in an average time-to-division of 5.4 hours and an aggregate growth factor $b \approx 1.14274$.

For those interested in an introduction to stochastic processes, we recommend the text by Ross [7]. For a deep dive into branching processes, including a discussion of the general result discussed in the previous section, see Kimmel and Axelrod [5].

We invite the reader to apply the results in this paper to the following problem, which we call the *Unreliable Banker Problem*: Your banker agrees to pay you interest every 20 days. However, he is a bit scattered and sometimes pays the interest a few days late, but other times he pays a few days early. Every time he hands you an interest check, he restarts the 20-day clock. This might seem like worrisome behavior, but the average time between interest payments turns out to be exactly 20 days. Is the banker's untimely behavior helping your balance, or not? Should you find a more timely banker? Do you need more information to decide?

Code

The simulations presented in this paper were written in Python. The code is available at `https://github.com/mlwright84/bacgrowth`.

Acknowledgements and Dedication

The authors thank Will Rose for conversations that inspired this article. The authors also thank Dr. Gene Chase for his suggestions. Dr. Chase is the father of one of the authors and passed away from COVID-19 during the review process; this article is dedicated to his memory.

References

[1] Alonso, A., Molina, I, Theodoropoulos, C. (2014). Modeling Bacterial Population Growth from Stochastic Single-Cell Dynamics. *Applied and Environmental Microbiology*. 80(17): 5241–5253. doi.org/10.1128/AEM.01423-14

[2] Athreya, K.B., Ney, P.E. (2004). *Branching Processes.* New York: Dover.

[3] Baranyi, J., George, S., Kutalik, Z. (2009). Parameter estimation for the distribution of single cell lag times. *Journal of Theoretical Biology*. 259: 24–30. doi.org/10.1016/j.jtbi.2009.03.023

[4] Carlton, M.A., Devore, J.L. (2017). *Probability with Applications in Engineering, Science, and Technology, 2nd ed.* New York: Springer.

[5] Kimmel, M., Axelrod, D.E. (2015). *Branching Processes in Biology.* New York: Springer.

[6] Pin, C., Baranyi, J. (2006) Kinetics of Single Cells: Observation and Modeling of a Stochastic Process. *Applied and Environmental Microbiology* 72(3): 2163–2169. doi:10.1128/AEM.72.3.2163–2169.2006

[7] Ross, S.M. (1995). *Stochastic Processes, 2nd ed.* New York: Wiley.