Computing for Numeracy: How Safe is Your COVID-19 Social Bubble?

Charles Connor
University of South Florida, cbconnor@usf.edu

Computing for Numeracy: How Safe is Your COVID-19 Social Bubble?

Abstract

The COVID-19 pandemic has led many people to form social bubbles. These social bubbles are small groups of people who interact with one another but restrict interactions with the outside world. The assumption in forming social bubbles is that risk of infection and severe outcomes, like hospitalization, are reduced. How effective are social bubbles? A Bayesian event tree is developed to calculate the probabilities of specific outcomes, like hospitalization, using example rates of infection in the greater community and example prior functions describing the effectiveness of isolation by members of the social bubble. The probabilities are solved for two contrasting examples: members of an assisted living facility and members of a classroom, including their teacher. A web-based calculator is provided so readers can experiment with the Bayesian event tree and learn more about these probabilities by modeling their own social bubble.

Keywords

COVID-19, social bubble, infection, bayes, Bayesian, event tree, probability tree, classroom

Creative Commons License

Cover Page Footnote

C. Connor is a professor of geology at the University of South Florida, former Chair of the Geology Department and former Associate Dean of Research for the College of Arts and Sciences. He studies volcanoes mostly, especially how to forecast volcanic eruptions and their hazards. Connor teaches students about quantitative methods in geology and about writing computer code to solve geology and geophysics problems. He is a Fellow of the American Association for the Advancement of Science.

<div style="border: 2px solid green;">

Computers for Numeracy

Numeracy, computing and computers, go hand-in-hand. This column explores topics in science, technology, engineering and mathematics (STEM!) that are of social importance by taking an algorithmic approach to problem solving. Each column defines a problem, develops a recipe for solving the problem, and implements a solution, with an example in computer code, often provided as an accompanying website.

</div>

Introduction

One useful strategy that has emerged this year to fight the spread of the novel coronavirus (COVID-19) is the formation of a social bubble. The idea behind a social bubble is to create a group of family members or friends who associate freely among one another while socially distancing from the rest of the world (Block et al. 2020; Leng et al. 2020). Such bubbles are intended to provide a safety net so that people within them can share tasks, socialize, and hopefully avoid depression and related anxieties that might become more prevalent when isolating alone or in a much smaller group. Bubbles rely on trusttrust that all members of the bubble will effectively socially distance from others in the general community by avoiding gatherings outside of the bubble and by exercising precautions while in public, such as wearing masks and avoiding crowds. Social bubbles are also formed around professional relationships, among those working together in an office or among students and teachers in a classroom. This latter type of bubble is necessarily more permeable, since an office worker, student, or teacher will go home, likely interacting frequently with an entirely different group of people.

Just how safe is your COVID-19 social bubble? That is a question best framed in numeracy. The first step is to recognize this is a probability problem. No one knows for sure if they will contract COVID-19 or not because their exposure depends on a host of variables, such as the prevalence of infection in the community, who are themselves uncertain (Dowd et al. 2020). The question also depends on the definition of safety. One might ask what the probability of infection of a member of the bubble is, given rates of infection in the general community outside the bubble, or one might want to know the probability of transmission to themselves or to another individual within the bubble or the probability that someone in the bubble will be hospitalized due to the disease. These are hierarchical questions, with the probability of one outcome depending on the probability of previous outcomes. Safety can be viewed in a hierarchical way (Figure 1).
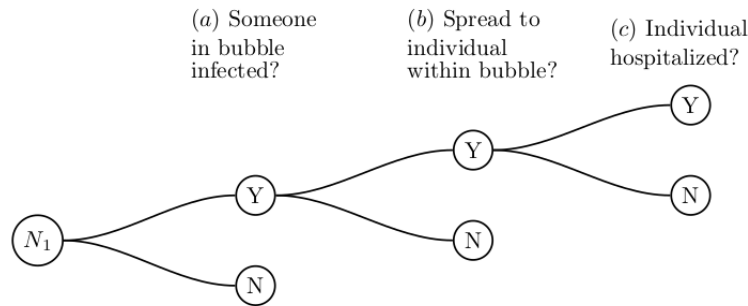
Figure 1. Visualize a binary event tree for the safety of members of a social bubble isolated against COVID-19. Events are sequential: (*a*) an individual within the bubble may become infected through interaction with the outside community, (*b*) the virus spreads to a specific individual within the bubble, and (*c*) that individual becomes seriously ill and is hospitalized. Each node (circle) represents a state represented by yes (Y) and no (N). The terminal nodes (ends of the tree branches) represent all contingencies, so the tree is a probability tree and the probabilities of terminal nodes sum to one.

Each step in this hierarchy of safety questions has a binary outcome. Will someone in the bubble become infected or not during some time window, such as the duration of the pandemic? The answer to this question is either yes or no. Similarly, most people within the bubble are likely primarily concerned with their personal safety, or with the safety of another specific individual. Given that someone in the bubble is infected, will another specific individual within the bubble become infected? This is also a binary, yes or no, proposition. Finally, safety outcomes might be framed in terms of the seriousness of the infection, measured in terms of hospitalization. Given that a specific individual is infected, will they become hospitalized?

This hierarchy gives rise to four *contingencies*[1]: (1) no one within the bubble becomes infected—the best contingency; (2) someone within the bubble becomes infected, but they do not spread the virus within the group, possibly because they take action; (3) they may spread infection to other individuals within the group, but these infections do not result in hospitalization; (4) an infected individual in the group may become seriously ill and is hospitalized. This hierarchy is best viewed as a probability tree. This probability tree is an *event tree* because the terminal nodes of the tree are specific contingencies and the probability of reaching a node depends on the probability of specific outcomes, called *transition probabilities*.

Binary questions about contingencies like these are addressed probabilistically using the *binomial distribution* (Shafer 1996). Consider a common binary question. When a fair coin is tossed multiple times, the expected outcome is that the coin will

---

[1]Italicized words are defined in the glossary at the end of this article.

land heads about 50 percent of the time. Even if the number of tails outcomes is not counted, we know that the rate of tails is one minus the rate of heads because no other outcome is possible. If the coin is unfair, say heads appears 60 percent of the time, the process is still modeled with a binomial distribution. The outcomes possible are heads and tails, but the expected outcome is different with an unfair coin. If 100 trials are made with the unfair coin, and 59 percent of the tosses result in heads, we call that 59 percent the sample proportion, which will generally be close to the expected outcome (60 percent in this case, provided enough trials are conducted). The binomial distribution is applied to probability of infection in the same way. For example, if the prevalence of infection in the general community is 20 percent, then the expected outcome for a representative individual in that community is that there is a 20 percent chance they will be infected, or stated another way, the *expected value* of infected individuals in the community is 20 percent.

The binomial distribution, however, is insufficient to model the probability of someone within a social bubble being infected with COVID-19 because bubble participants have taken action to change the expected outcome: they have isolated from the community. This action means that the event tree needed to forecast a series of events, from infection of someone in the bubble to hospitalization of an individual, is *Bayesian*. A Bayesian event tree (BET) can account for the prevalence of infection in the community, as well as the nuances of social bubble behavior, and even the anticipated resilience of individuals within the bubble due to better-than-average health, comorbidity, or similar considerations.

In the following, I describe the numerical approach to create a BET for a COVID-19 social bubble. A javascript code given in the supplementary material runs in a web browser and allows readers to input their own model parameters, calculate their bubble safety, and see how bubble safety changes with changes in specific parameters.

Constructing a BET

Using the visualization of the event tree (Figure 1), we can see the problem is to estimate the probability of each node, especially the four nodes which end the branches of the tree. These nodes are the contingencies. They are meant to account for all the possible outcomes in the scenario represented by the tree. Other contingencies can be imagined (e.g., death), but our tree ends with hospitalization, which is serious enough. In a binary event tree, all contingencies are accounted for, so the probabilities of the terminal nodes must sum to one. Solving the problem of the safety of the bubble reduces to estimating these probabilities.

Before these probabilities are estimated, a nomenclature is needed to describe the problem. Let's designate the nodes of the tree using the symbol N. We keep track of the nodes by assigning each node an index. The first node is the initial state of the COVID-19 bubble, a proclamation that it exists, and is N_1. The first branch of the tree ends in nodes $N_{2,1}$ and $N_{2,2}$, the yes and no outcomes, respectively. Subsequent branches are labeled using the same index scheme (Figure 2).
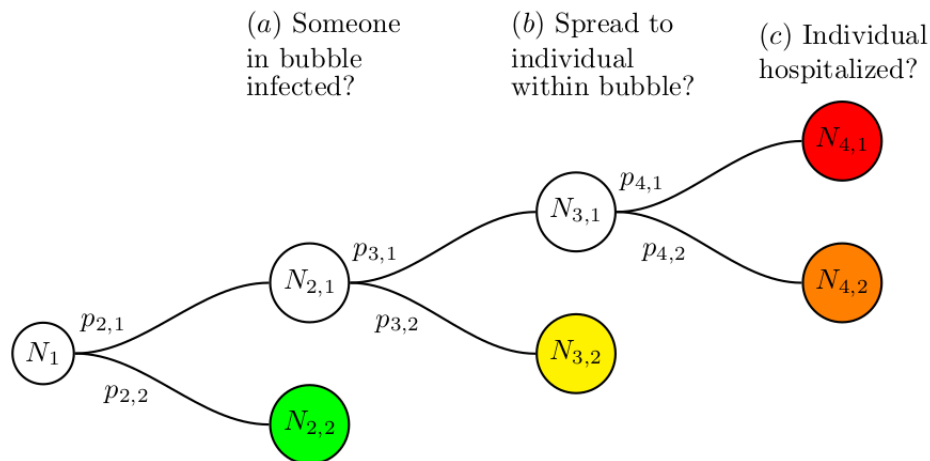

Figure 2. Nomenclature for solving the BET.

There is a transition probability calculated for each branch. That is, given that a node is reached in the tree, what is the probability that the one of the subsequent nodes on the tree will be reached. These transition probabilities always link two nodes, so they can be indexed using the node index. For example, $p_{2,1}$ is the transition probability between node N_1 and node $N_{2,1}$ (Figure 2).

The terminal nodes are the nodes of most interest because these are the contingencies. We add to the nomenclature by color-coding the terminal nodes in a stoplight arrangement, from best outcome (●) to worst outcome (●). The four probabilities that are ultimately calculated with the BET are as follows.

- P[●] is the probability that the COVID-19 bubble remains intact and no one within the bubble is infected.

- P[○] is the probability that the infection is not spread to an individual within the bubble, even though someone else in the bubble is infected by contact with the outside world (e.g., someone breaks the isolation bubble).

- P[●] is the probability that an individual is infected with COVID-19 within the bubble because of spread of the infection inside the bubble, but that individual is not sufficiently ill to become hospitalized.

- P[●] is the probability that the individual is hospitalized because of spread of COVID-19 within their bubble.

Given this nomenclature for the BET, it is practical to consider how to calculate the probabilities on each branch of the tree and the contingency probabilities.

Calculating a BET for People in an Assisted Living Facility

This BET is binomial (Figure 2), so we can start by specifying the sample proportions that are part of our calculations. The sample proportion is calculated for each branch of the tree, and these calculations are related. Consider a theoretical group of older adults, well isolated from the greater community, say in an assisted living facility. Twenty individuals are isolated inside this COVID-19 social bubble. The rate of infection in the larger community is taken to be 10 percent. What is the likelihood that the bubble will be broken, an individual infected and hospitalized?

Let n_{21} be the total number of individuals within the COVID-19 social bubble. For the assisted living facility this number of individuals is 20. Then y_{21} is the expected number of individuals to be infected, given that no precautions are taken. That is, the bubble is not isolated effectively from the greater community. The sample proportion is:

$$\bar{\mu}_{21} = \frac{y_{21}}{n_{21}} \tag{1}$$

We think we know the infection rate in the broader community is 10 percent, so in this example $\bar{\mu}_{21} = 0.1$, which means that $y_{21} = 2$ for people in the assisted living facility. Of course, we hope they isolate better than the community generally!

The subscripts mean that the variables n_{21} and y_{21}, and the sample proportion $\bar{\mu}_{21}$, are calculated for the first branch of the tree. The sample proportion has a value between zero and one, so $y_{21} \leq n_{21}$, and both are greater than zero.

If the probability tree were not Bayesian then the transition probability for this branch would be $p_{2,1} = \bar{\mu}_{21}$. But, importantly, we have additional information that the individuals in the bubble of the assisted living facility are isolating. For a BET, this additional information is called a prior function, a statistical model to describe the additional information about the social bubble.

For a BET, the most commonly used and easiest prior function to use is a Beta distribution, with a capital "B" by convention, which depends on two parameters (Gelman et al. 2013). For this branch of the tree, the two parameters are designated α_{21} and β_{21}. Both $\alpha_{21} > 0$ and $\beta_{21} > 0$.

These two parameters and the Beta distribution are used to weight the sample proportion in the BET. The prior distribution mean is:

$$\mu_{p(2,1)} = \frac{\alpha_{2,1}}{\alpha_{2,1} + \beta_{2,1}} \tag{2}$$

and the transition probability for this branch of the tree is

$$p_{2,1} = \frac{\alpha_{2,1} + y_{2,1}}{\alpha_{2,1} + \beta_{2,1} + n_{2,1}} \tag{3}$$

The transition probability $p_{2,1}$ is also called the expected value of the *posterior probability* because it is the probability that results from weighting the sample proportion by the prior distribution. The weighting always results in a transition probability that is between the values of the prior mean $\mu_{p(2,1)}$ and the sample proportion $\bar{\mu}_{21}$.

A lot can be learned about the BET by considering how equation 3 behaves with different values of its parameters. First, let's say the isolation bubble size $n_{2,1}$ is very large compared to the values of α_{21} and β_{21}. In this case, the transition probability will be quite close to the sample proportion. Physically, this means that isolation in a bubble is not really helping because the isolation group is too large. Now consider that $\beta_{2,1}$ is large compared to the other three parameters. If the social bubble is small (e.g., $n_{2,1} = 5$) then large values of $\beta_{2,1}$ and $\beta_{2,1} > \alpha_{2,1}$ will mean that the prior mean will be small and the value of the transition probability will be close to the prior mean. This means that the group is well-isolated, with a lower chance of infection than the community at large.

Rules of the road for estimating α_{21} and β_{21} are provided in a following section. In the example of the assisted living community, the hope is the community is very well isolated. If this is so, β_{21} should have a large value. Because the community is well isolated in their facility, the value of the prior distribution mean is small (e.g., by setting $\beta_{2,1} = 100$), and $p_{2,1} \approx 2.5$ percent (Figure 3).

Because the tree is binary:

$$p_{2,2} = 1 - p_{2,1} \tag{4}$$

and the probability that the COVID-19 bubble remains intact and no one within the bubble is infected is: $\mathrm{P}[\bullet] = p_{2,2}$. For the assisted living facility bubble, that works out to $p_{2,2} \approx 97.5$ percent. Because of isolation, people in the assisted living facility bubble are safer than the broader community.

That is the good news. Unfortunately, once someone in the assisted living facility bubble is infected, there might be a greater risk of spread than generally in the broader community. That penalty is reflected in the next node.
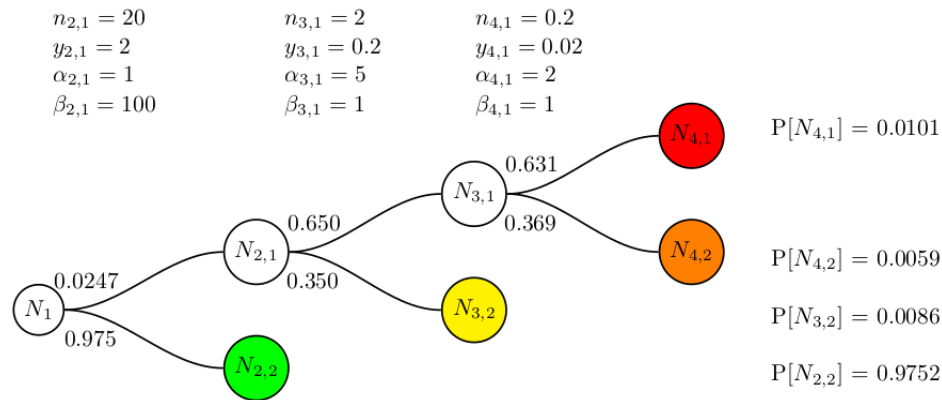
$$n_{2,1} = 20 \qquad n_{3,1} = 2 \qquad n_{4,1} = 0.2$$
$$y_{2,1} = 2 \qquad y_{3,1} = 0.2 \qquad y_{4,1} = 0.02$$
$$\alpha_{2,1} = 1 \qquad \alpha_{3,1} = 5 \qquad \alpha_{4,1} = 2$$
$$\beta_{2,1} = 100 \qquad \beta_{3,1} = 1 \qquad \beta_{4,1} = 1$$

$\mathrm{P}[N_{4,1}] = 0.0101$

$\mathrm{P}[N_{4,2}] = 0.0059$

$\mathrm{P}[N_{3,2}] = 0.0086$

$\mathrm{P}[N_{2,2}] = 0.9752$

Figure 3. A BET for a group of 20 older adults, well isolated from the greater community, but quite susceptible to transmission and more likely to be hospitalized if infected compared to members of the general population.

The next transition probability is calculated in exactly the same way by specifying values for $n_{3,1}$, $y_{3,1}$, $\alpha_{3,1}$, and $\beta_{3,1}$ with these caveats:

$$n_{3,1} = y_{2,1} \tag{5}$$

and

$$y_{3,1} < n_{3,1} \tag{6}$$

because the BET is hierarchical. For the assisted living facility bubble, $n_{3,1} = 2$ and $y_{3,1} = 0.2$, keeping the expected rate of infection at 10 percent. If the members of the assisted living facility are not well-isolated from each other, then $\alpha_{3,1} > \beta_{3,1}$. Using $\alpha_{3,1} = 5$ and $\beta_{3,1} = 1$, gives $p_{3,1} = 0.65$. Infection spreads much more quickly in this group than in the broader community. Once $p_{3,1}$ is calculated:

$$p_{3,2} = 1 - p_{3,1} \tag{7}$$

and $\mathrm{P}[\mathrm{o}] = p_{2,1} \times p_{3,2}$. $\mathrm{P}[\mathrm{o}]$ is a relatively good outcome since no additional people in the social bubble are infected, but for people in the assisted living community it is a low probability contingency (Figure 3).

As the group is comprised of older individuals, it is also much more likely that an individual in the group will be hospitalized once they are infected (e.g., $\alpha_{4,1} > \beta_{4,1}$) compared to the general population (McMichael et al. 2020). Exactly the same calculation is done to estimate this third transition probability:

$$n_{4,1} = y_{3,1} \tag{8}$$

and

$$y_{4,1} < n_{4,1} \tag{9}$$

and the remaining probabilities calculated:

$$p_{4,2} = 1 - p_{4,1} \qquad (10)$$

$$\mathrm{P}[\bullet] = p_{2,1} \times p_{3,1} \times p_{4,2} \qquad (11)$$

$$\mathrm{P}[\bullet] = p_{2,1} \times p_{3,1} \times p_{4,1} \qquad (12)$$

Comparing all of the contingencies (Figure 3), this group is relatively safe only because of the strength of their isolation. Once there is infection within the bubble, there is a very high chance of hospitalization. That is, P[•] is greater than P[•] or P[•]. By comparison, in a non-Bayesian approach, the probability of hospitalization would be the product of the sample proportions. In this example, the non-Bayesian probability of hospitalization would be $0.1 \times 0.1 \times 0.1 = 0.001$, or about $1/10^{th}$ the value of P[•] calculated using the BET. This is a large difference in probabilities and illustrates how the knowledge of the group improves the assessment of their safety using a BET.

Calculating a BET for a Classroom

Consider a very different group of people, a classroom of 50 young and healthy students (Connor 2020). The rate of infection in the community at large is the same as in the previous example, assumed to be 10 percent. Students being students, this isolation bubble is much more permeable than the group of isolated older adults. In fact, we have little knowledge of their isolation because we do not keep track of the students outside of class. To capture this state, ignorance of the effectiveness of isolation, $\alpha_{2,1} = \beta_{2,1} = 1$. Because the group is large, the transition probability is close to the value of the sample proportion. But a penalty is paid for ignorance, so the transition probability is slightly greater than the sample proportion. The rest of this BET is quite similar to the previous example for older adults, except for the prior function for the last transition probability. Because the students are likely quite resilient, $\beta_{4,1} \gg \alpha_{4,1}$. The resulting BET for the student social bubble with example values is shown in Figure 4. The resulting probability of hospitalization for the students in the classroom is good, P[•] $= 0.0003$, or the chances of hospitalization are about 3 in 10,000 despite the relative lack of isolation of this group. The same cannot be said for the teacher, who might be more vulnerable to hospitalization, given infection, than her students. If the teacher has the same rate of hospitalization as the community generally given infection, 10 percent in this example, $\alpha_{4,1} = \beta_{4,1} = 0$. Her probability of hospitalization is P[•] $= 0.0036$, roughly 10 times higher than the probability of her students being hospitalized. She pays a cost for interacting with a permeable bubble. The non-Bayesian probability
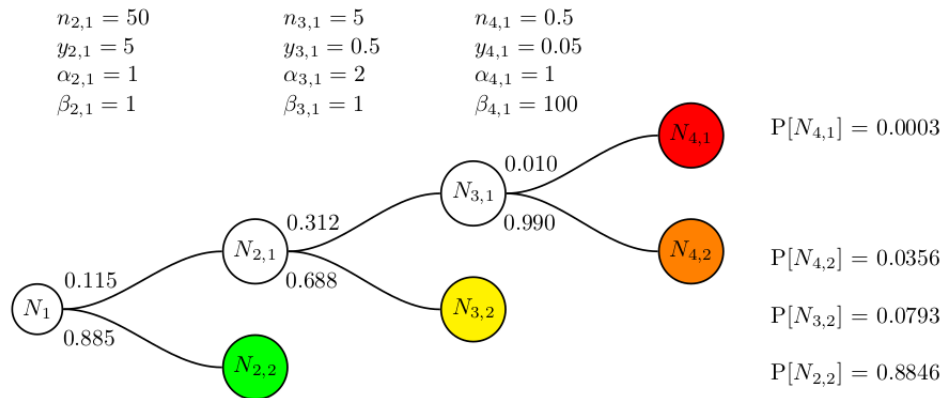
$n_{2,1} = 50$
$y_{2,1} = 5$
$\alpha_{2,1} = 1$
$\beta_{2,1} = 1$

$n_{3,1} = 5$
$y_{3,1} = 0.5$
$\alpha_{3,1} = 2$
$\beta_{3,1} = 1$

$n_{4,1} = 0.5$
$y_{4,1} = 0.05$
$\alpha_{4,1} = 1$
$\beta_{4,1} = 100$

$P[N_{4,1}] = 0.0003$

$P[N_{4,2}] = 0.0356$

$P[N_{3,2}] = 0.0793$

$P[N_{2,2}] = 0.8846$

0.010

0.312

0.115

0.990

0.688

0.885

Figure 4. A BET for a classroom of 50 students.

of her hospitalization based on the products of sample proportions is still 0.001. The teacher's probability of infection because she is associated with the students is roughly four times the probability of her infection compared to the community generally. To reduce her probability of hospitalization, she would have to isolate more effectively from students within the bubble, changing the values of $\alpha_{3,1}$ and $\beta_{3,1}$, say by more effective social distancing within the bubble than practiced by students.

Calculating Your Own BET

The circumstances for COVID-19 social bubbles are highly variable depending on how many people are involved, how well they isolate from the outside world and from each other, and how likely individuals within the bubble are to be hospitalized compared with the general population. What about the safety of your COVID-19 social bubble? I have developed a web-based calculator to solve the BET for COVID-19 bubbles as described here. Access the web-based calculator at:

computers for numeracy: the covid social bubble calculator

Enter your bubble size, the sample proportion appropriate for your community, prior parameters for your situation, and calculate the contingency probabilities. While using this calculator, you are getting to know equation 3 and how prior knowledge influences probable outcomes.

There are some guidelines for making sure your calculation makes sense for your particular social bubble. First, larger values of $n_{2,1}$ mean there are more people in the bubble. As the bubble size gets bigger, the prior mean decreases relatively (equation 3), and the isolated group behaves more like the general community. If

the isolation group is small then the prior function (values of α and β) is more important.

At each node, the sample proportion depends on y and n. The sample proportion y/n should reflect values in the general community in your area, which might change with time or might be poorly known. It is important to consider a range of values and investigate their effect on probabilities. You might try $y_{2,1}/n_{2,1} = y_{3,1}/n_{3,1}$. Similarly, for each node $n \geq y$, since the proportion of people infected or hospitalized must be less than or equal to the total number of people in the group. Since the BET is hierarchical, once the first column is completed, values of n are updated automatically in successive columns of the web-based calculator.

The parameters α and β in a Beta distribution are used to calculate the prior mean (your knowledge of the bubble that makes it different from the population generally). Examples of possible values:

- $\alpha = \beta = 0$. The transition probability (expected value of the posterior probability) will be the same as the sample proportion. If this is the case for all three columns then the BET is actually not Bayesian.

- $\alpha = \beta = 1$. The prior reflects "maximum ignorance," meaning we have no knowledge of how well the group actually isolates (for node 1 on the BET) or how a individual in the group will react to infection (node 3).

- $\alpha > \beta$. The prior means that the probability of a bad outcome (infection, hospitalization) is greater than estimated for the community in general. For an older group with comorbidities, $\alpha > \beta$ for the third transition probability.

- $\alpha < \beta$. The prior means that the probability of a good outcome (no infection, no hospitalization) is greater than in the general community. A well-isolated group will have $\alpha < \beta$ (node 1). A healthy group will have $\alpha < \beta$ (node 3).

The important thing is to experiment with alternative values. One technical feature of the Beta distribution is that its variance decreases with increasing values of α and β. That means that using larger values of α and β indicates greater confidence in the state of isolation of the bubble. The web-calculator does not include the variance in the posterior probabilities, essentially the confidence in the probability estimate. In general confidence is low because of the complexity of the social situation (e.g., the classroom case uses lower values of α and β at node 1 compared to the assisted-living facility case). One way to explore these effects is to develop a set of parameters for your social bubble, then change individual parameters one at a time to see if the change is significant for the contingency probabilities.

Concluding Remarks

Our lives are uncertain. We may be unable to reduce this uncertainty, but we can gain understanding of uncertainty through numeracy (Vacher 2016; Paulos 2018).

The importance of a numerate approach in our lives is placed in stark relief by the COVID-19 pandemic. Numerate people react to the pandemic by assessing their risk of infection and by attempting to reduce their risk by assessing alternative strategies, like forming a social bubble and evaluating its effectiveness.

The problem of assessing the safety of a COVID-19 social bubble has a great deal in common with risk assessment in many STEM disciplines (Woo 2011). BETs are used to forecast volcanic eruptions (Connor et al. 2001; Newhall and Hoblitt 2002; Baxter et al. 2008; Marzocchi and Bebbington 2012), earthquakes (Marzocchi et al. 2012), and tsunami (Grezio et al. 2010). Engineered systems, characterized by potential sequential failures and cascading events, are assessed with Bayesian methods (Apostolakis 1990; Connor 2011; Zuccaro et al. 2018). The medical community has a long history of using Bayesian methods in areas like diagnosis, and is applying these methods to improve our understanding of COVID-19 at a furious rate (Iwendi et al. 2020; Roda et al. 2020). These analyses can become quite complex, but these examples also point to the need to understand Bayesian methods and BETs at local or personal levels (Wang 2015).

Not everyone is in a position to develop their own BET or to deal with all the complexities of a sophisticated Bayesian analysis, but many people can use a BET to forecast outcomes that are important to them, like the safety of their COVID-19 social bubble. The trick is to present the core calculations (e.g., equation 3–12) in a sequential fashion, following the hierarchy of the tree itself. These sequential calculations might be implemented in a spreadsheet or using a web-based calculator like the one presented here. The main point is to do the calculation and react to the new quantitative perspective it provides.

Acknowledgment

Glossary

Bayesian statistics: named for Thomas Bayes, a paradigm for statistical inference in which a prior distribution is specified for all unknown parameters in the model. The prior distribution is based on knowledge independent of the frequency of past events. Bayes' theorem is then applied to calculate probability using both the frequency of past events and the prior distribution. In this column, the frequency of

past events is the sample proportion. Bayesian statistics have a checkered reputation because of the subjective nature of the prior distribution, in this column the Beta distribution, which depends on two estimated parameters: α and β. Advocates for Bayesian statistics argue that the difficulty in estimating the prior is offset by the insights gained from incorporation of that information. In models like the one presented in this column, it is best to explore the model using alternative values of α and β. That is, it is best to see how the probabilities of events change with different assumptions about the prior distribution.

Beta distribution: used to weight probability as a prior distribution in Bayesian event trees. The Beta distribution has two parameters, α and β. When α is large, the weighted probability tends to increase. When β is large, the weighted probability tends to decrease. Large values of both α and β give a low variance to the Beta distribution, meaning there is great confidence in the mean of the prior distribution. Small values of both α and β give high variance to the Beta distribution and there is low confidence in the mean value of the prior distribution.

binomial distribution: arises when there are only two possibilities associated with a trial or event. The two possibilities might be that an event occurs or does not occur. The two possibilities do not necessarily have equal probability.

contingencies: represent all the possible outcomes of a series of events. That is, if one sums the probabilities of all possible contingencies, they sum to one.

event tree: a diagram that illustrates how one event might lead to another event in a hierarchical fashion, attempting to capture all possible outcomes. A probabilistic event tree assigns probabilities to all events. A Bayesian event tree calculates probability using Bayesian statistics.

expected value: for any probability distribution, the theoretical mean value of the distribution. An average is the mean value of a sample, whereas the expected value is calculated for a probability distribution.

posterior probability: calculated by mathematically combining a probability based on the frequency of events, in this column the sample proportion, as well as prior information, in this column information or assumptions about the isolation of the social bubble and health of members within that bubble.

prior distribution: accounts for additional information associated with calculation of a probability. Probability can be calculated using the frequency of past measured

events. For instance, by finding how many people are infected in a community. Caution is needed in applying this frequency to a smaller group, members of which might have special circumstances. In this case the probability is weighted using a prior distribution. A drawback of this approach is that the prior distribution must also be estimated!

transition probability: on an event tree, the probability that one event will follow another. An event tree has a series of transition probabilities. These are multiplied together to find the probability of a contingency.

References

Apostolakis, G. 1990. "The Concept of Probability in Safety Assessments of Technological Systems." *Science* 250(4986): 1359–1364. https://science.sciencemag.org/content/250/4986/1359

Baxter, P., W. Aspinall, A. Neri, G. Zuccaro, R. Spence, R. Cioni, and G. Woo. 2008. "Emergency Planning and Mitigation at Vesuvius: A New Evidence-Based Approach." *Journal of Volcanology and Geothermal Research* 178(3): 454–473. https://doi.org/10.1016/j.jvolgeores.2008.08.015

Block, P., M. Hoffman, I. J. Raabe, J. B. Dowd, C. Rahal, R. Kashyap, and M. C. Mills. 2020. "Social Network-Based Distancing Strategies to Flatten the COVID-19 Curve in a Post-Lockdown World." *Nature Human Behaviour* 4(June 2020): 588–596. https://www.nature.com/articles/s41562-020-0898-6

Connor, C. 2011. "A Quantitative Literacy View of Natural Disasters and Nuclear Facilities." *Numeracy* 4(2): Article 2. http://dx.doi.org/10.5038/1936-4660.4.2.2

Connor, C. 2020. "Factors in the Probability of COVID-19 Transmission in University Classrooms." *Numeracy* 13(2): Article 7. https://doi.org/10.5038/1936-4660.13.2.1368

Connor, C. B., B. E. Hill, B. Winfrey, N. M. Franklin, and P. C. La Femina. 2001. "Estimation of Volcanic Hazards from Tephra Fallout." *Natural Hazards Review* 2(1): 33–42. https://doi.org/10.1061/(ASCE)1527-6988(2001)2:1(33)

Dowd, J. B., L. Andriano, D. M. Brazel, V. Rotondi, P. Block, X. Ding, Y. Liu, and M. C. Mills. 2020. "Demographic Science Aids in Understanding the Spread and Fatality Rates of COVID-19." *Proceedings of the National Academy of Sciences* 117(18): 9696–9698. https://doi.org/10.1073/pnas.2004911117

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian Data Analysis.* London: CRC press.

Grezio, A., W. Marzocchi, L. Sandri, and P. Gasparini. 2010. "A Bayesian Procedure for Probabilistic Tsunami Hazard Assessment." *Natural*

Hazards 53(1): 159–174.
https://link.springer.com/article/10.1007/s11069-009-9418-8

Iwendi, C., A. K. Bashir, A. Peshkar, R. Sujatha, J. M. Chatterjee, S. Pasupuleti, R. Mishra, S. Pillai, and O. Jo. 2020. "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm." *Frontiers in Public Health* 8(3 July 2020): 357. https://doi.org/10.3389/fpubh.2020.00357

Leng, T., C. White, J. Hilton, A. J. Kucharski, L. J. Pellis, H. Stage, N. G. Davies, M. J. Keeling, and S. Flasche. 2020. "The Effectiveness of Social Bubbles as Part of a COVID-19 Lockdown Exit Strategy, a Modelling Study." *medRxiv* https://doi.org/10.1101/2020.06.05.20123448

Marzocchi, W., and M. S. Bebbington 2012. "Probabilistic Eruption Forecasting at Short and Long Time Scales." *Bulletin of Volcanology* 74(8): 1777–1805. https://link.springer.com/article/10.1007/s00445-012-0633-x

Marzocchi, W., J. D. Zechar, and T. H. Jordan. 2012. "Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting." *Bulletin of the Seismological Society of America* 102(6): 2574–2584. https://doi.org/10.1785/0120110327

McMichael, T. M., D. W. Currie, S. Clark, S. Pogosjans, M. Kay, N. G. Schwartz, J. Lewis, A. Baer, V. Kawakami, M. D. Lukoff, and others. 2020. "Epidemiology of COVID-19 in a Long-Term Care Facility in King County, Washington." *New England Journal of Medicine* 382(21): 2005–2011. https://www.nejm.org/doi/full/10.1056/NEJMoa2005412

Newhall, C., and R. Hoblitt 2002. "Constructing Event Trees for Volcanic Crises." *Bulletin of Volcanology* 64(1): 3–20. https://link.springer.com/article/10.1007/s004450100173

Paulos, J. A. 2018. "A Few Reflections on A Numerate Life." *Numeracy* 11(1): Article 8. https://doi.org/10.5038/1936-4660.11.1.8

Roda, W. C., M. B. Varughese, D. Han, and M. Y. Li. 2020. "Why Is It Difficult to Accurately Predict the COVID-19 Epidemic?" *Infectious Disease Modelling* 5: 271–281. https://doi.org/10.1016/j.idm.2020.03.001

Shafer, G. 1996. *The Art of Causal Conjecture.* Cambridge: MIT press.

Vacher, H. 2016. "Grassroots Numeracy." *Numeracy* 9(2): Article 2. http://dx.doi.org/10.5038/1936-4660.9.2.2

Wang, F. 2015. "Numeracy Infusion Course for Higher Education (NICHE), 2: Development of Students' Bayesian Reasoning Skill." *Numeracy* 8(2): Article 6. http://dx.doi.org/10.5038/1936-4660.8.2.7

Woo, G. 2011. *Calculating Catastrophe.* London: World Scientific.

Zuccaro, G., D. De Gregorio, and M. F. Leone. 2018. "Theoretical Model for Cascading Effects Analyses." *International Journal of Disaster Risk Reduction* 30(part B): 199–215. https://doi.org/10.1016/j.ijdrr.2018.04.019