<1>Chapter 3: Markov models

In this Chapter, we will examine one of the most common and useful ways to understand complex public health and healthcare interventions: the *Markov model.* A Markov model is a representation of health or disease that requires us to analyze multiple processes in sequence, such as multiple stages of disease, to identify how effective or cost-effective a public health or healthcare program might be. Markov models are highly flexible and allow for an infinite variety of diseases or interventions to be simulated and understood, which is why they are among the most popular tools for public health and healthcare research.

<2> Principles of Markov modeling

You no-doubt opened this book hoping to gain skills to conquer major life-threatening diseases. Fortunately, we will now focus on a major pathology of traumatic, global significance, causing substantial suffering through severe psychological trauma: male pattern baldness (a.k.a., "MPB"). My personal experiences witnessing the suffering from this disease among colleagues has caused me nightmares. In my own community of university professors, who frequently suffer from the disfiguring disorder, there are several stages of illness that illustrate the typical progression of MPB. First, the victims begin in a happy, hairy state. Next, they progress to a stage of "frontal bossing", in which a receding hairline might be ignored as it ascends up to the northern pole of the head. Third, the bossing progresses to cover the full top surface of the head, leaving a U-shaped ring of hair around the edge of the head, or—in the most horrifying case—the "comb over" hairstyle. Finally, as the terminal state of disease, there is the complete bowling ball. The rate of progression from one stage to the subsequent stage is about 10% per year.

Naturally, many of the victims of this disorder do not take the problem lightly. They struggle, often in vain, to reverse the progression of MPB. Many use a hair regrowth treatment, which has a small probability of reversing the progression of disease—typically, a rate of regression from one stage to the prior stage of about 5% per year.

With this information, we might try to draw a decision tree to estimate the probability that any given person at risk for MPB might end up completely bald. Figure 3.1 provides a first attempt at this decision tree. As shown in the tree, a person starting out in the *hairy* state has a small probability per year of moving to *frontal bossing*. Once a person is in the *frontal bossing* stage, they can also try the hair regrowth treatment and have a small probability of becoming *hairy* again, or may just stay in the *frontal bossing* stage for that year, or they can progress to *combover*. This makes the decision tree complicated; if the person becomes *hairy* again, the tree could become infinitely long ….

[FIGURE 3.1 HERE]

The dilemma of trying to draw an infinitely-long decision tree illustrates an inherent limitation to our approach: that diseases with multiple progression and regression possibilities can be difficult to diagram through the decision tree approach. Many diseases, such as cancers, heart diseases, inflammatory disorders, and substance abuse, are diseases that have multiple stages, with possibilities for improving disease sometimes and worsening disease at other times. These diseases are not easily diagrammed through a decision tree.

Instead of trying to draw an infinitely-long decision tree, we can use an alternative approach of drawing a *Markov model* to answer key questions about a disease, such as how likely a person is to become completely bald in the case of MPB.

A Markov model can be conceptualized as a series of lily pads on which a frog is jumping. The pads are different states of health or disease. The frog jumping across the pads is like a person going between these states of health or disease. The key features of a Markov model are a *state diagram*, which clearly designates the possible health states that can be occupied as a sequence of connected circles; and *state transitions*, or arrows that identify the probability of moving from one state to another.

Figure 3.2 illustrates a state diagram and state transitions for male pattern baldness.

[FIGURE 3.2 HERE]

Hence, we can calculate *self-loops*, or probabilities of staying in the same state each month, by calculating the probability that a person does not move to a different state but simply remains in their current state from one month to the next. The self-loops are equal to 1 minus the sum of all transition probabilities leaving a given state.

We can check that we haven't made a mistake in our model by adding all of the transition probabilities that leave a given state, including self-loop transition probabilities; the sum of all probabilities must add to 1, since people cannot disappear from our diagram.

Now that we have established the state diagram for our Markov model, we can use the model to answer several critical questions about male pattern baldness. To start with, we can ask: in a male population of about 100 million people, how many people would we expect to have the horrific combover?

At first, this question may seem impossible to answer, since we only have limited information about the probabilities that any individual might transition between different levels of cocaine use, not information about the overall "epidemic" of male pattern baldness. But one of the key strengths of a Markov model is its ability to translate data from the level of the individual to the level of the population. To see this advantage of Markov modeling, we can translate our state diagram into a series of equations.

The final equation, Equation 5.4, tells us that people cannot magically disappear from our state diagram. A male who is alive in the simulated population must be a

and therefore the sum of all three probabilities must be equal to 1.

Equations 5.1 through 5.4 are commonly termed *difference equations* because they express the difference in the probability of being in a state at time t+1 and the probability of being in a state at time t.

<2> Using Markov models to address disease uncertainties

How do we use these equations to help us understand the male pattern baldness epidemic?

The easiest strategy to utilize these equations is to first solve for the probability of being in each state under a *steady-state situation*, that is over the long-term after much time has passed and the probability of being in each state doesn't vary much over time. After we perform an algebraic solution to the problem, we will later illustrate that Markov models typically have a long-term *stationary distribution*, or a long-term steady state probability for an individual to be in any one of the given states. For now, we can take it on faith that such a long-term probability exists, and later we will demonstrate its existence more rigorously.

Over the long-term, we can reason that the probability of being in any given state is so stable that it does not change from one time step to another. Hence, we no longer need to think of transitions from time t to time t+1, and can rewrite Equations 5.1 and 5.4 as their steady-state versions, Equations 5.5 through 5.8:

Equations 5.5 through 5.8 assume that over the long run, the epidemic of male pattern baldness will stabilize such that the probability of being a non-user will not vary dramatically from month to month, nor will the probability of being a rare user or the probability of being a routine user.

Because equations 5.5 through 5.8 provide us with three unknown terms

Those students who recall linear algebra can also choose to express Equations 5.5 through 5.8 as a series of matrices as shown in Equation 5.9, which can be solved with matrix multiplication to obtain the same solution.

The four-by-four matrix of transition probabilities is commonly referred to as a *transition matrix*.

Using our solution to the steady state equations, we can solve our problem regarding how many men we expect to be totally bald in our population. Over the long-term, we anticipate that the prevalence of

What would be the typical incidence rate of becoming a

The calculation of incidence reveals an important property of Markov models: the so-called Markov assumption, or *memoryless* property. A Markov model does not capture a person's prior history when estimating the person's probability of transitioning from one state to another. A person who has never

Sometimes, modelers will add more states to a Markov model to differentiate people with a prior history of a condition from people without such a history, although this adds more states and more equations to a model, making it more complex. In a latter chapter, we will introduce microsimulation modeling methods, which can resolve this problem. For now, let us use the simplest model for illustrative purposes.

<2> Markov models for intervention planning

The simple Markov model we have created can assist us to plan an intervention

We can depict the change in our model following this intervention by rewriting Equations 5.5 through 5.8 as Equations 5.13 through 5.16:

If we were to solve these equations, we would see a subtle change in the solution for the stationary distribution transition probabilities.

In addition to calculating the lower *prevalence*