

# Basics of stochastic modeling

Biological processes display inherent variability. Take height, for example: height for human males can range from 5 to 7 feet (1.5 to 2.1 meters), although the average is close to 5 feet, 9 inches (1.77 meters) in the United States. If you were to develop a computer model to predict the amount of cloth needed to make some number of pairs of pants, and you needed to include human male height as a variable, you could do so in two ways. The first method would incorporate height as a single value, i.e. 5 feet 9 inches, and assume that all men are that height. This is called *deterministic modeling*, because the value of the variable height has already been determined. If you wanted to estimate the amount of fabric needed to make 1000 pairs of pants with a deterministic model, you would run the model once to calculate the amount of fabric needed for one pair, then multiply the results by 1000. Running the model additional times would not provide any new information: the input variable is fixed, and the result will be the same each time.

A second approach would be to incorporate height as a distribution of values representative of the natural range (see Figure 0-1, below). Now, if the model is run several times, and if a different value is drawn at random from this distribution for each run, the results will be slightly different each time. This is a *stochastic model* (stochastic is defined as “being or having a random variable”) because height is no longer a fixed value, but is drawn randomly from a distribution. Values with higher probability of occurring (i.e., heights close to 5’9”) are more likely to be drawn than values at the extreme ends (“tails”) of the distribution. In Figure 0-1, we can see that values below 1.61 meters (5’3”) are drawn 5% of the time, and values above 1.93 meters (6’4”) are drawn 5% of the time. Thus, 90% of the time, the value of height will be between 5’3” and 6’4” (1.61 and 1.93 meters).

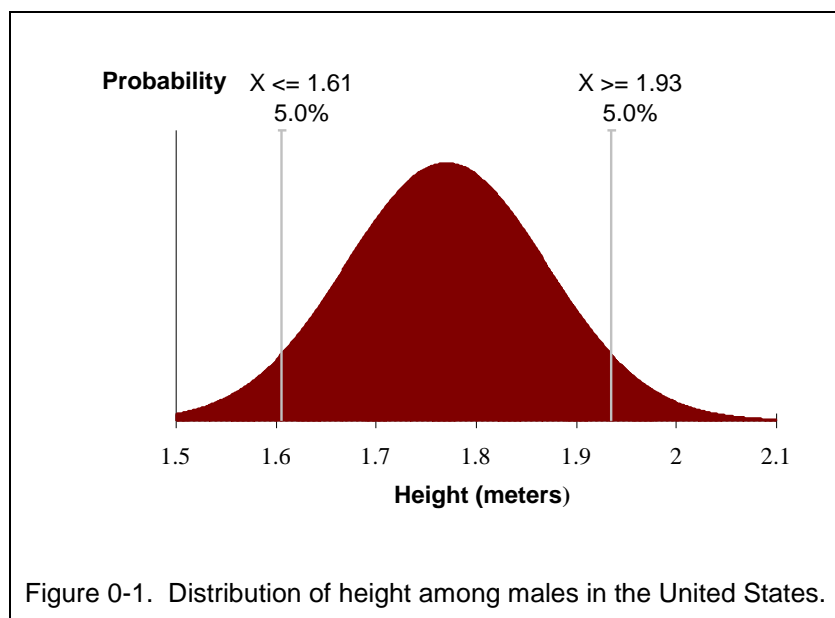
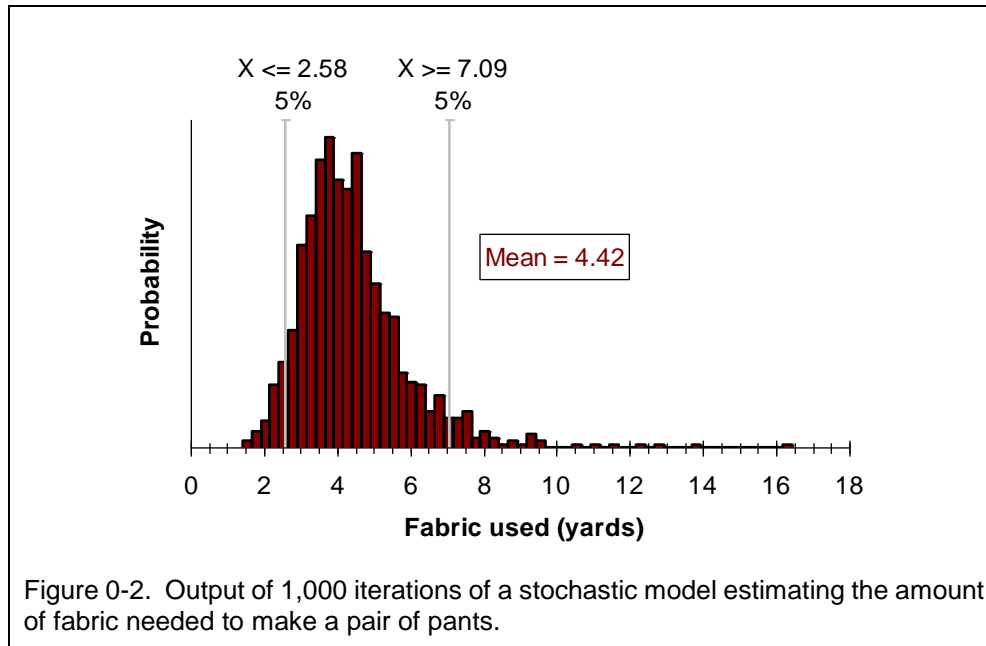
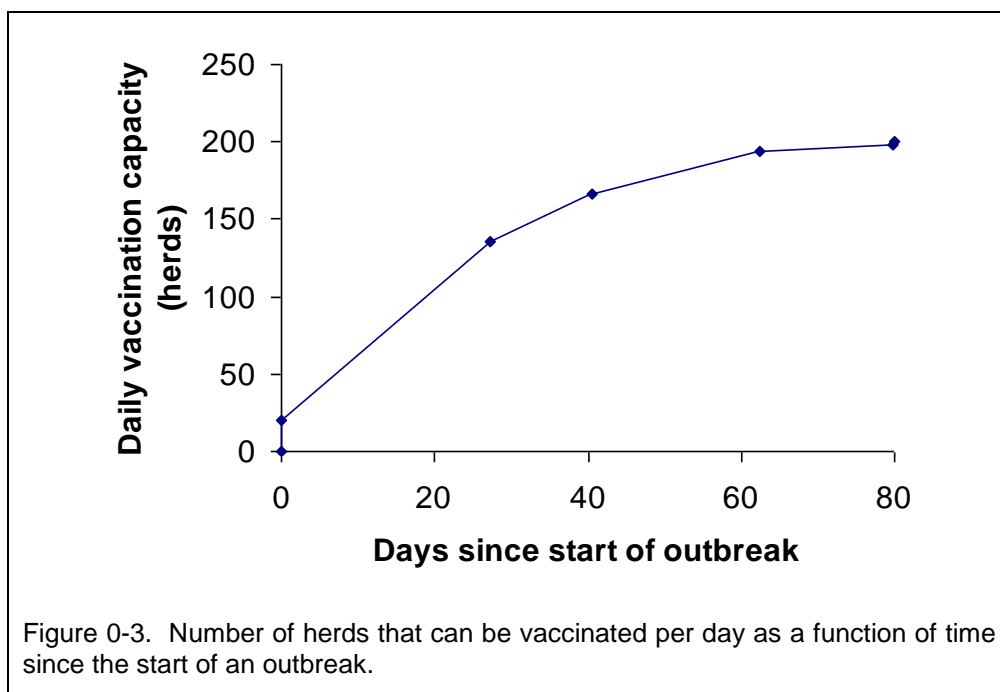


Figure 0-1. Distribution of height among males in the United States.

In a stochastic model, both the inputs and the outputs are distributions. Running the stochastic model from our example once is not particularly informative. However, by running the model 1,000 times (running 1,000 “iterations”), we get 1,000 different estimates of the fabric needed for one pair of pants. These estimates form a distribution (Figure 0-2). To estimate the fabric needed for 1,000 pairs of pants, we can sum the 1,000 estimates. Results of stochastic models incorporate variability, whereas results from deterministic models do not. We can improve our model by incorporating additional stochastic variables, such as variability in the amount of fabric used per inch of height (which is affected by waist size and pants style).



*ADSM* is a stochastic model. Variability is incorporated in the model using *probabilities, distributions, and relationships*. A *distribution* is used when the variability in a parameter is assumed to be random. The distribution describes the range of values a variable can take, and how likely those values are to be selected. Model parameters described as distributions include the length of infectious period and the distance that animals are likely to be transported. A *relationship* is used when the variability in one parameter is due to another factor. A relationship describes one variable as a function of another (Figure 0-3). Model parameters described as relationships include the probability of detecting an infectious herd (which is a function of time since the herd was infected), and number of herds that can be depopulated per day (a function of time since the outbreak was detected).



*Thanks to the NAADSM Development Team for providing the source documentation.*