Probability Models for Epidemics and Materials

By Peter Fortini

Q How does knowledge of materials help us to understand an epidemic?

A Throughout the current COVID-19 pandemic, we have seen graphs in newspapers showing numbers of new cases in different countries, regions, or states rise, level off and fall, then perhaps rise again. In seeking to understand what is going on, we draw on the training and knowledge we have. There are close analogies when we consider the structure of some of the materials of interest to ASTM International and in ASTM standards.

At ASTM, we mainly see statistics in the context of defining repeatability and reproducibility of test methods, control charting for processes, acceptance sampling, calibration, and perhaps design of experiments. However, the study of statistics also includes a large dose of probability theory and stochastic processes.

A classic stochastic process example is the Galton-Watson branching process. This process starts with one individual, which generates none, one, or more (a random number) of "progeny" for the next generation. The number of progeny is given by a probability distribution, p_0, p_1, p_2 etc. Each of these, in turn, generates a random number of individuals for a third generation, and so on. In the Galton-Watson branching process, the number follows the same probability distribution.

This probability model, created by Francis Galton, was originally applied to the survival or spread of family names, in which the relevant progeny of an individual is the number of sons who will carry the family name. In genetics, an additional case is the survival and spread of a mutant gene, in which the mean number of progeny is a direct measure of the "fitness" of the mutant gene.

Further examples given in classic texts on probability are nuclear chain reactions and lengths of waiting lines (where the "progeny" of an individual being served are those who join the queue while that person is being served).[1] An epidemic is yet another example of the process, where the progeny of an infected individual are additional people that person infects.

The basic features of this process are:

— There is a probability p_0 that the initial individual gives rise to no progeny for the next generation, so the process dies out immediately.

— There is a probability that the chain dies out even if the initial individual has progeny. This will always happen if the mean μ (in epidemiology, the reproductive number R_0) of the distribution of progeny of an individual is less than one.

— If the mean of the progeny distribution is greater than one, there is a chance that the number will grow exponentially. There is still a chance of dying out. The probability of eventual die-out can be calculated using the theory of the Galton-Watson process. The calculation is a solution to the following equation, where *P(x)* is the probability-generating function of the progeny distribution.

$$x = P(x) = p_0 + p_1 x + p_2 x^2 + p_3 x^3 + \dots$$

With mean greater than one, once the number of individuals in a generation is high enough, the growth is for practical purposes exponential with:

number of individuals in n-th generation $= k \exp(\mu\, n)$

where k depends on the variation that occurs in the initial stages.

Of course, unlimited exponential growth cannot continue for long with any real process. There will be a resource limit. For an epidemic, it is the population of people potentially exposed. For a chain reaction, it is the amount of nuclear material to start with as well as its dispersion in the subsequent explosion.

Where, in the context of materials, do we see similar processes? Important additional examples are part of polymer science and are described in the classic text by Flory.[2] Many materials, such as resins, coatings, plastics, and rubber are made of molecules that start as long linear chains and are joined by crosslinks.

The branching process applies to materials in the following way.

Imagine starting from a randomly chosen (monomer) unit in such a material. We consider the linear chain that it is on to be the initial individual. Its "progeny" are additional chains that are connected to it by crosslinks or branches.

Figures 1 and 2 show the comparison. For simplicity, it shows a case where the process dies out. The entire set of chains that are reached by following the crosslinks (combining generations) is the portion of the material connected to the initially chosen one. In an oligomeric, uncured resin, the mean number of additional chains connected to an initial one is less than one. When a material is cured, the mean number of crosslinks or branches on a chain is increased to greater than one. Exponential growth then means that the initial unit is part of an effectively infinite network. The gel point is the point of curing at which the infinite network first appears. Not far into curing of crosslinked polymeric materials, the majority of the material comprises a single large molecule. The probability of ultimate die-out from the initially random monomer unit gives us the fraction of the material not connected to this network. This is the sol fraction of the polymer, which might be extractible.

Applying this analogy to the spread of an epidemic or pandemic, it illustrates that in the uncontrolled case, essentially all of a susceptible population will eventually be infected from a first case, and only a relatively small fraction will not be connected to the contagion network.

Many of the graphs that follow the progress of the current COVID pandemic show declines after initial rises in case numbers. Based on the analogy to the networks in polymeric materials, we should only expect this to occur when the vast majority of the population has been infected. However, the total numbers are still only a fraction of that.

What is behind the declines we see? Declines may partly be due to geographic distance between regions reporting their numbers and partly to general social fragmentation, with the members of our society belonging to subgroups that have relatively little mutual contact and so are not (yet) affected. Undoubtedly, however, the declines are mainly due to the measures that have been taken: canceling large gatherings, mask wearing when we do have contact with the general public, working from home, and other social distancing, to reduce the mean of the progeny distribution directly.

Similarities between apparently very different phenomena become clear when they are described mathematically. What you already know about one or more of these gives you a starting point for understanding new ones. ■

REFERENCES

1. Feller, William, *An Introduction to Probability Theory and Its Applications*, Vol. 1, Wiley, New York, 3rd edition, 1968 (first edition 1950).
2. Flory, Paul J. *Principles of Polymer Chemistry*, Cornell University Press, Ithaca, New York, 1953.

Peter E. Fortini is vice chair of the committee on quality and statistics (E11), which he has also served as committee chair. He is an ASTM International fellow and Award of Merit honoree.

Dean V. Neubauer, the Data Points column coordinator, is engineering fellow and chief statistician, Corning Inc. A member at large on the executive subcommittee of the committee on quality and statistics (E11), he is an ASTM International fellow and a past chair of the E11 committee.

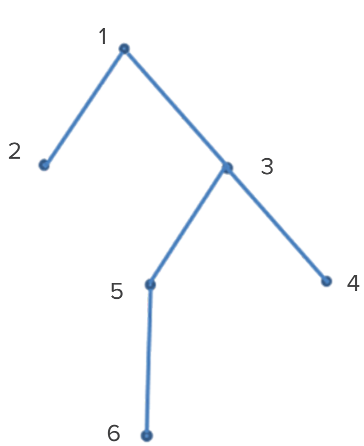

Figure 1 — Realization of a Branching Process

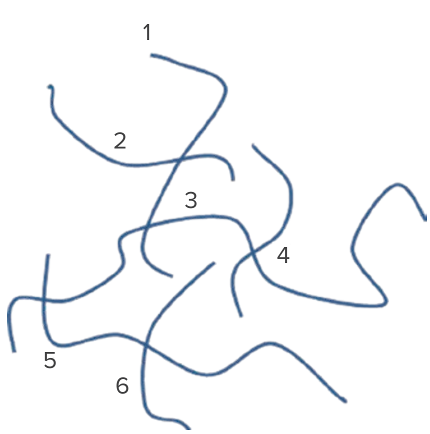

Figure 2 — A Polymer Network