Simulation incorporates the randomness of the real world. How do we decide upon the nature of this randomness? How do we specify distributions for our random variables? To make these decisions and specifications, we start from the data available. We have:

(a) either lots of data or "little" data;
(b) from either the distribution of interest or a "related" distribution.

Having little data is bad; having little data on the wrong distribution is worse. "Little" data means any combination of

(L1) small sample;
(L2) summary statistics only, e.g., mean, variance, min, max, median, mode;
(L3) indirect qualitative information, based for example on interviews with informed people or on experience in related situations.

Wrong but "related" distributions are sometimes the only sources of information available:

(R1) Wrong amount of aggregation. An inventory simulation might need daily demand data, but our client has only monthly demand data.
(R2) Wrong distribution in time. Almost always our data are historical, but we really want next month's demand distribution—or, worse, next year's.
(R3) Wrong distribution in space. We want to simulate the New York City Fire Department but we have data only from Los Angeles.
(R4) Censored distributions. We want the demand data but have only the sales data. Sales understate demand when there are stockouts. (See Problem 4.9.21.)

(R5) Insufficient distribution resolution. In a message switching system that used the U.S. national phone system, the length of a typical phone call was of the same order of magnitude as the resolution accuracy of the phone system for timing calls. Hence, data on call durations was inaccurate.

With "little" or "wrong, but related" data, be skeptical about any distributions specified. Sensitivity analysis is particularly called for in such bold-inference situations. Sections 4.1 to 4.5 discuss what qualitative information might lead us to conjecture that a normal, lognormal, exponential, Poisson, or Weibull distribution is appropriate. Because such conjectures are tentative, check sensitivity both to the parameters of the distribution conjectured and to its form. Do sensitivity analyses using paired simulation runs with common random numbers.

The goal of a sensitivity analysis is to show that output depends only weakly on which of a set of plausible distributions is used. A theoretical distribution usually has only one or two parameters, which can be varied continuously. This makes sensitivity analysis easy, provided one considers only the (limited) shapes that the theoretical distribution can take. Fitting theoretical distributions to data is popular, but in Section 4.8 we argue that this is not necessarily wise. Section 4.6 discusses a quasi-empirical distribution that blends a continuous, piecewise-linear distribution which closely mimics the data (interpolating the usual empirical distribution) with an exponential tail. For empirical comparisons, see Section 8.1.

Quite often (wrongly, we believe) sensitivity studies consist solely in varying the mean and variance of the input distributions, simply by using transformations of the form \( Y = a + bX \). With rare exceptions (like the mean wait in a stationary \( M/G/1 \) queue), these two parameters do not suffice to determine the expectation of the output distribution, let alone the distribution itself. Especially in the tails of the input distributions, the data give only meager support to the assumed forms and there are significant qualitative differences in the behavior, for example, of the exponential, Weibull, and normal. Usually a meaningful robustness investigation must check sensitivity of the performance measure to the form (e.g., skewness) of the input distributions. Compared to focusing only on the mean and variance, this extended sensitivity analysis may be more difficult to explain to the user and may require more effort to carry out. But, for credibility, it is needed.

Depending on the random number generator, there may be a positive probability of generating the numbers 0 and \( \infty \) from some "continuous" distributions. When transforming uniform random numbers, underflow or overflow errors can occur. We mention the overflow problem here because distributions with nonfinite support are artifices; to avoid statistical anomalies we may wish to reject transformed numbers that are more than a few standard deviations away from their expected values. Fox and Glynn (1986c) study this situation. If the uniform variates are transformed monotonely, this can be