What Monte Carlo Methods Cannot Do

Scott Ferson
Applied Biomathematics, Setauket, NY*

* 100 North Country Road
Setauket, New York 11733
613-751-4350, fax -3435
scott@ramas.com

Running head: What Monte Carlo methods cannot do

ABSTRACT

Although extremely flexible and obviously useful for many risk assessment problems, Monte Carlo methods have four significant limitations that risk analysts should keep in mind. (1) Like most methods based on probability theory, Monte Carlo methods are data-intensive. Consequently, they usually cannot produce results unless a considerable body of empirical information has been collected, or unless the analyst is willing to make several assumptions in the place of such empirical information. (2) Although appropriate for handling variability and stochasticity, Monte Carlo methods cannot be used to propagate partial ignorance under any frequentist interpretation of probability. (3) Monte Carlo methods cannot be used to conclude that exceedance risks are no larger than a particular level. (4) Finally, Monte Carlo methods cannot be used to effect deconvolutions to solve backcalculation problems such as often arise in remediation planning. This paper reviews a series of ten exemplar problems in risk analysis for which classical Monte Carlo methods yield an incorrect answer.

KEYWORDS: data requirements, nonstatistical uncertainty, ignorance, exceedance risks, deconvolution, backcalculation

INTRODUCTION

As discussed in other articles in this special issue commemorating their modern use in scientific calculation, Monte Carlo methods provide flexible and extremely powerful techniques for solving many of the central problems in risk analysis. Facilitated by the widespread availability of microcomputers, Monte Carlo methods, like statistical resampling approaches in general (Efron and Diaconis, 1983; Simon and Bruce, 1991), have lately come to be perceived as a fundamental tool by which a many formerly very difficult estimation problems become trivial. Nevertheless, Monte Carlo methods cannot do everything, and they cannot solve all the problems in risk analysis involving calculation. In the interest of maintaining a certain sobriety as we celebrate the utility of Monte Carlo methods, it is important to consider the limitations that constrain the use of the approach. This paper reviews four of the most important ones in risk analysis. The following section lists ten risk analysis problems, most of which appear straightforward or even simple, which cannot be formally solved using the Monte Carlo methods that most risk analysts would use.

RISK ANALYSIS PROBLEMS THAT MONTE CARLO METHODS CANNOT SOLVE

Consider the following exemplar problems:

(1) Suppose we need to estimate the product of two imperfectly known inputs. We know that the first, *A*, can be no smaller than 0.2 and no larger than 0.4. We also know the second, *B*, can be no smaller than 0.3 and no larger than 0.5. But no further information is available about *A* or *B*. How should the product *AB* be characterized?

(2) Suppose we know that *A* is no smaller than 0.2 and no larger than 0.3, and that *B* is lognormally distributed with median 5 ($\mu=\ln(\text{median})\approx 1.6$) and coefficient of variation 0.2 ($\sigma=\sqrt{\ln(CV^2+1)}\approx 0.2$). What can be said about the product *AB*?

(3) Suppose we know that *A* is lognormally distributed with median of 2 and coefficient of variation 0.1 ($\mu\approx 0.69$, $\sigma\approx 0.1$), and that *B* is lognormally distributed with median 5 and coefficient of variation 0.2 ($\mu\approx 1.6$, $\sigma\approx 0.2$). How large is the frequency with which the product *AB* exceeds 14?

(4) Given *A* and *B* as above, and the correlation between *A* and *B* is zero, how large is the frequency with which the product *AB* exceeds 14?

(5) Given *A* and *B* as above, and the correlation between *A* and *B* is 0.3, how large is the frequency with which the product *AB* exceeds 14?

(6) Given *A* and *B* as above, and Spearman's (rho) correlation between *A* and *B* is 0.3, how large is the frequency with which the product *AB* exceeds 14?

(7) Suppose $A\times B=C$ where *A*, *B*, and *C* are random variables with *A* and *B* independent. If *A* has a lognormal distribution with median 2 and coefficient of variation 0.1 ($\mu\approx 0.69$, $\sigma\approx 0.1$), and *C* has a lognormal distribution with median 5 and coefficient of variation 0.2 ($\mu\approx 1.6$, $\sigma\approx 0.2$), what is the distribution of *B*?

(8) Suppose we know well the marginal distributions for random variables *A* and *B* but cannot specify exactly their joint distribution. How can we characterize the product *AB*?

(9) Suppose we cannot specify the marginal distributions for random variables *A* and *B* exactly. How can we characterize the product *AB*?

(10) Suppose we know the full joint distribution for random variables *A* and *B* but are not sure of the precise function that should be used to combine them. How can we make a reasonable characterization of the function of *A* and *B*?

Each of these exemplars is similar in form to common risk analysis problems which are routinely addressed with Monte Carlo methods. Nevertheless, *none of these problems can be solved by a straightforward application of Monte Carlo methods.* The reasons the methods cannot be applied (or yield wrong answers) are discussed below for each problem in turn. Most of the problems do have solutions, however, and these are also discussed in the following sections.

It is important to note first that specific details have been included in the first seven problems to make the discussions concrete. For instance, some problems refer to lognormal distributions with given medians and coefficients of variation, or ask about the product of multiplying random variables. It should be clear that analogous problems could be phrased in terms of distributions with other parameters or other shapes and other mathematical operations on them. There is no detail about a parameter or distribution shape that renders a problem especially difficult. The details permit calculation of the correct answer which will enable us to discern at a glance how the Monte Carlo method fails in each case.

I. MONTE CARLO CANNOT PROPAGATE NONSTATISTICAL UNCERTAINTY

The first of the exemplar problems above asks about the product of two parameters for which only minima and maxima are known. I recently posed this problem on the RISKANAL list server (accessible by subscription through LISTSERV@LISTSERV.PNL.GOV). About two-thirds of the fifty-five respondents gave the solution

$$AB = [0.2, 0.4] \times [0.3, 0.5] = [0.2 \times 0.3, 0.4 \times 0.5] = [0.06, 0.2]$$

which is the answer obtained by interval analysis (Moore, 1966; Neumaier, 1990). Almost one third of respondents, however, suggested that a probabilistic approach should be used. Most of the solutions they offered explicitly described a Monte Carlo method that modeled *A* and *B* with uniform distributions over their respective ranges. Figure 1 depicts the probability density function that results from convolving these two uniform distributions together under an assumption of independence, which is the default assumption about dependence in most software packages including Crystal Ball (Burmaster and Udell, 1990) and @Risk (Salmento, Rubin, and Finkel, 1989; Barton, 1989).

The interval analysis answer and the Monte Carlo distribution agree in the sense that they both say the answer must lie somewhere in the range between 0.06 and 0.2. However, the probability distribution says quite a bit more than this. It asserts that the probability that the product is toward one of the extreme values is much less than the probability that it has a more central value. But where in the statement of the problem can we find the justification for this concentration of probability in the center of the range? In fact, of course, *any* probability distribution over the range 0.06 to 0.2 might be the true distribution of products. There is nothing given in the statement of the problem which we can use to narrow this set. Even a delta distribution at 0.06 or 0.2 cannot be excluded.

The idea that it is appropriate to assume probabilistic uniformity when only range information is available dates back to Laplace himself. The idea has come to be known as the 'principle of insufficient reason'. Although it has been justified and generalized by sophisticated theoretical development under the rubric of maximum entropy (Jaynes, 1957; Levine and Tribus, 1979; Grandy and Schick, 1991; Lee and Wright, 1994), the idea is widely viewed with some skepticism, especially by those holding to a frequentist view of probability theory (see Jaynes, 1979).

Ferson and Ginzburg (1996) have argued in some detail that interval analysis provides the only reasonable solution to the first exemplar problem, at least from the point of view of risk analysis. In the context of seeking a representation for a single underlying number, the approach of using a uniform (or maximally entropic) distribution may be reasonable. In the context of risk analysis, however, the approach yields answers that are overconfident. This overconfidence generates results that are at least potentially non-protective. For instance, it may be the case that larger (or smaller) values will occur far more frequently than is implied by the particular distribution obtained. The Monte Carlo approach—and to be fair, any classical probabilistic approach—cannot comprehensively propagate nonstatistical uncertainty, at least under a frequentist interpretation required by risk analysis.

A small minority of the RISKANAL respondents (2 out of 55) suggested that the solution to the problem should be a uniform distribution over the range [0.06, 0.2]. Of course, this solution also overstates what is justifiably deducible about the product *AB*, but, in the sense that a uniform distribution 'represents' an interval, it is perhaps a little more reasonable than the triangular shape shown in the figure. Nevertheless, from the perspective of a risk analyst, it still potentially underestimates the tails of the distribution in a way that cannot be justified by appeal to any empirical facts or assumptions stated in the problem.

Some readers may criticize the first exemplar problem as unrealistic and argue that, in most real-world circumstances, we actually have more information about a parameter than just its minimum and maximum. It is true, of course, that such additional information is often available. Nevertheless, this problem is certainly not impossible in real settings and, because of its apparent simplicity, seems to deserve a cogent and defensible solution even if it is only rarely encountered. One could argue further that there always remains some degree of ignorance about a parameter that is not the result of variability in virtually all real problems. Insofar as this problem offers insight into that issue, it also deserves discussion. The more general case when other details about a distribution are available but do not in themselves precisely specify a single distribution is treated in the discussion below of the ninth exemplar problem.

The second exemplar problem seeks to combine an interval and a probability distribution. Again, using a uniform distribution as a surrogate for an interval will allow Monte Carlo methods to be used to obtain a solution. But doing so will yield a result that is not justifiable given the stated foreknowledge. Figure 2 depicts both the result from a Monte Carlo analysis and bounds on the true solution obtained by a direct probability bounds analysis (Ferson, Ginzburg, and Akçakaya, 1996). Given the information stated in the problem, these bounds are pointwise optimally narrow (Williamson and Downs, 1990).

In other words, they could not be any narrower and yet still enclose all the probability distributions that might actually arise as the product of *A* and *B*. We can see from the figure that the Monte Carlo result suggests the frequency that the product is less than 1.0 is about 15%. In fact, this frequency could be as large as 50%, depending on the distribution or value that *A* actually has within the interval. We again see that the Monte Carlo approach cannot comprehensively propagate nonstatistical uncertainty.

The bounds on the product of an interval and a probability distribution could also be obtained by an alternative strategy using sets of intervals to represent distributions (cf. Berleant, 1993; 1996). If sections are taken systematically from the lognormal distribution for *B* and multiplied by the interval [0.2, 0.3] according to the elementary rules of interval analysis (Moore, 1966), the products will all be intervals. The cumulation of the left bounds of these products would yield the left-hand bound of the region depicted in the figure and the cumulation of the right bounds of the products would approximate the right-hand bound.

II. MONTE CARLO CANNOT BOUND EXCEEDANCE RISKS

An exceedance risk is the frequency with which a random variable may exceed some specified value. Exceedance risks are a focal concern for analysts doing environmental or public health risk assessments such as exposure or cancer incidence modeling, sometimes because there is a critical level beyond which consequences become intolerable. A central task in risk analysis is to estimate how often such consequences may occur. The third through sixth exemplar problems are couched as questions about computing exceedance risks.

The dotted line in Figure 3 depicts the results of a Monte Carlo simulation of the third exemplar problem in which lognormally distributed random deviates sampled independently from the distributions for *A* and *B* are multiplied together. The result is shown as a complementary cumulative distribution function (which is also known as the survival function). From this curve we can read off the estimated frequency of the product *AB* being larger than 14 as 6.5%. But this estimation is contingent on the assumption of independence between *A* and *B*. The problem omitted any mention of the dependence between *A* and *B*. Yet their dependence can have a substantial effect on the resulting distribution. For instance, if the dependence is strongly positive, the exceedance risk could well be twice as large. If the dependence between *A* and *B* is strongly negative, the exceedance risk could be almost surely zero.

Williamson and Downs (1990; see also Frank, Nelsen, and Schweizer, 1987; Ferson and Long, 1995) describe a numerical method for computing the possible range of outcomes for given marginal distributions when their dependence is unknown. Their representation scheme uses lower and upper discrete approximations to the quantile function (the quasi-inverse of the distribution function) as bounds on a distribution. The method is based fundamentally on the classical Fréchet (1935) inequalities

$$\max(0, \Pr(E) + \Pr(F) - 1) \;\le\; \Pr(E \text{ and } F) \;\le\; \min(\Pr(E), \Pr(F)),$$
$$\max(\Pr(E), \Pr(F)) \;\le\; \Pr(E \text{ or } F) \;\le\; \min(1, \Pr(E) + \Pr(F)),$$

which give best possible bounds for the probabilities of conjunctions or disjunctions of

events when only the total probabilities of the events Pr(E) and Pr(F) are given and no further information about the dependence between the events E and F is available. Although the passage from events to random variables requires the invocation of copulas (Schweizer and Sklar, 1983; Nelsen 1991; 1995) which capture the dependency relations embodied in joint distributions, the algorithms developed by Williamson and Downs are rather simple to use in risk analysis (Ferson and Long, 1995). Applying their method yields the solid lines depicted in the figure. The solid lines describe bounds on the true exceedance risk given only information about the marginal distributions of A and B. We see from the figure that the risk that the product AB is larger than 14 could be anywhere between zero and 25%. These bounds are known to be optimal in the sense that they could not be any narrower without more specific information about the dependency between A and B (Williamson and Downs, 1990). How then should we interpret the Monte Carlo estimate of 6.5%?

In some cases, one can explore the range of possible exceedance risks by varying the correlation coefficient over all possible values between –1.0 and +1.0. Several workers (e.g., Mosleh and Bier, 1992; Burgman, Ferson, and Akçakaya, 1993; Bukowski, Korn and Wartenberg, 1995) have suggested this strategy which might be called 'dispersive Monte Carlo sampling' because of its relation to variance maximization (Bratley, Fox, and Shrage, 1983; Whitt, 1976). This strategy does not work in general however. Even varying the correlation coefficient over all possible values *cannot* yield the entire range of the possible values for the exceedance risk (Ferson, 1994; Ferson and Burgman, 1995). This is because correlation is a very limited kind of *linear* dependency and such a strategy explores only a small fraction of the space of possible dependencies between the two variables.

If, as in the case for the third exemplar problem, we do not know the dependence between the variables, what justifies an assumption of independence, other than computational convenience? The whole point of our interest in exceedance risks seems to necessitate attention to this kind of problem. Without knowing the full joint distribution of the two variables which expresses their dependence, the classical methods of probability theory including Monte Carlo methods cannot compute an estimate of the product distribution, even in principle. Unfortunately, this fact has not slowed many analysts who routinely assume independence among all variables in the risk expression. The statement of such assumptions is sometimes explicitly made, although it is often not.

The fourth exemplar problem assumes that the correlation between the two factors is zero. But having zero correlation is not the same as being independent (Flury, 1986; Nelsen, 1995). It is well known, for instance, that two normal random variables may have a sum that is not even close to normal, even though their correlation is zero. Likewise, the product of two lognormal random variables need not yield a lognormal distribution even if they have zero correlation. There are infinitely many ways to achieve a zero correlation, but there is only one way to be independent.

Similar difficulties persist in the fifth and sixth exemplar problems. Having a particular (non-zero) correlation does not imply a particular joint distribution (Nelsen, 1991; 1992; 1995). The same is true for rank correlation. Even though Monte Carlo software packages can *simulate* such correlations (Scheuer and Stoller, 1962; Iman and Conover, 1982; Nelsen, 1986; 1987), they cannot *describe* such correlations in the sense of

capturing the variety of distributions that may result when variables are combined arithmetically. This means that, even if we have measured the correlation and dutifully included it in a Monte Carlo simulation, we cannot be sure that the estimate of the exceedance probabilities are not underestimates, perhaps substantial underestimates. To put it most simply, the risks of adverse consequences may be decidedly higher than would be predicted by Monte Carlo simulation, even in the unusually good circumstance of knowing well both marginal distributions and correlation structure, and even with infinitely many replications. These facts seem not to be widely appreciated in the risk analysis community. How much higher the true frequencies may be is still an open research question whose resolution merits serious attention.

III. MONTE CARLO CANNOT COMPUTE BACKCALCULATIONS

The seventh exemplar problem involves backcalculation (Burmaster, Lloyd, and Thompson, 1995; Burmaster and Thompson, 1995). If, as is stated in the problem, $A \times B = C$ and we want to estimate B from A and C, an analyst might try to solve for B by rearranging the equation to get $B = C / A$ and estimating the answer by Monte Carlo. The result of this approach is depicted in Figure 4 as a dotted line. This answer is wrong, however, as can be easily checked by putting it back into the original equation and computing C. The correct answer, i.e., the distribution that will yield the observed distribution for C when convolved with that for A, is depicted as a solid line in the figure.

The reason for the discrepancy is that we tried to compute B by assuming independence between A and C. But of course A and C cannot be independent of one another. Indeed, one is a function of the other, so independence is manifestly an inappropriate assumption. Solving such backcalculation problems requires a special operation called deconvolution (Jansson, 1984; Ferson, 1995; Burmaster, Lloyd, and Thompson, 1995). Deconvolution is used to untangle distributions that had been convolved together under independence. If paired data (a_i, c_i) were available, it would be a simple matter to compute the distribution of $(b_i = c_i / a_i)$. But when paired data are not available, it is generally more difficult to compute the distribution for B and requires a deconvolution. Except in special cases, using Monte Carlo instead of deconvolution will yield a wrong answer.

In a few special circumstances, it may be possible to approximate a deconvolution using Monte Carlo methods. For instance, in the case where we know that $C = A + B$ and A and B are independent, we can estimate B as $C - A$ if we use a correlation between A and C of $r = s_A / s_C$ where s_A and s_C are the observed standard deviations of the respective distributions. This approach is precise for large number of replications provided that all the distributions are normal and provided that the sampling algorithm simulates Pearson correlation (Scheuer and Stoller, 1962). It should be mentioned that Crystal Ball (Burmaster and Udell, 1990) and @Risk (Salmento, Rubin, and Finkel, 1989; Barton, 1989) both simulate Spearman's rank correlation between deviates, and several other commercial software tools have no provision for simulating correlations at all. It should also be emphasized that distributions that are neither normal nor lognormal often arise in risk analysis problems.

In many circumstances, a simple trial-and-error approach (Ferson, 1995) may be adequate for approximating deconvolution. This approach is based on the idea that the mean obtained by Monte Carlo assuming independence will be approximately true but the spread of the distribution needs correction. An iterative process may be able to find a solution that works reasonably well.

Deconvolution is required in many risk analysis problems. One important area is remediation planning to compute the cleanup levels that will be sufficient to yield a planned level of protection (Burmaster, Lloyd, and Thompson, 1995; Burmaster and Thompson, 1995; Ferson, 1995). Of course, deconvolution is not always required in regulatory planning, but it has many applications in this area and in other kinds of problems throughout risk analysis. For instance, suppose we define the risk of some carcinogen to be the product of its potency for an individual and the dose that individual receives. Dose and potency might be independent of one another if the former depends on behavior and the latter on physiological health. Certainly, risk is dependent on both dose and potency as they are what determine it. Suppose we can estimate the distribution of dose from measurements of urinary metabolites and can estimate the distribution of risk from published mortality data. How might we go about estimating the distribution of potency in a population? This problem is exactly analogous to the seventh exemplar problem, where dose is *A*, potency is *B*, and risk is *C*. Using ordinary Monte Carlo with its implicit assumption of independence would yield a incorrect result.

Note that if we had *paired* data on risk and dose for exposed individuals, we could have correctly estimated the distribution of the potency factor, simply by forming a histogram of the quotients of the pairs. It is the ignorance about which individual risk comes from which dose that trips us up and invalidates the use of a Monte Carlo simulation using division and independence to estimate the potency factor. Unfortunately, such individual-based data are often expensive or even impossible to collect. In this case, the nature of mortality data prevents us from measuring risk for a particular individual, so we cannot generate pairs of dose and risk for individuals. The circumstance of having only information on marginal distribution is probably the usual case in many risk analyses.

IV. MONTE CARLO CANNOT YIELD AN ANSWER UNLESS MUCH IS KNOWN

By this point, it should be clear that Monte Carlo methods require a great deal of information to yield answers. There are three corollaries of this general observation embodied in the last three exemplar questions. They might be expressed as

1. Monte Carlo methods cannot yield an answer when the statistical dependencies among the variables are unknown or uncertain.
2. Monte Carlo methods cannot yield an answer when input distributions are unknown or uncertain.
3. Monte Carlo methods cannot yield an answer when the model structure is unknown or uncertain.

Surprisingly, these limitations seem not to have restrained risk analysts from using Monte Carlo methods even in the absence of formally sufficient knowledge. Of course, the analyses do not escape the limitations. If the requisite information is not provided by

empirical study, it must be present in the form of assumptions, whether explicit or only implicit. It is an important question then as to how incomplete knowledge will affect the reliability of the assessments analysts make using Monte Carlo methods.

The eighth exemplar problem in which the joint distribution of the variables is unknown or uncertain arises very commonly in real risk and safety assessments (Smith and Watson, 1980; Ferson, 1994; Bukowski, Korn and Wartenberg, 1995; Ferson and Long, 1995). Indeed, the case for which sufficient data has been collected to estimate a joint probability distribution well is very unusual. Fortunately, methods are available to compute bounds on distributions of arithmetic combinations using only marginal distributions, without making any assumption about the dependencies among the variables (Williamson and Downs, 1990; Ferson and Long, 1995). In many cases these bounds will be optimally narrow, and variables which are known to be independent may be involved in the risk expression as well. As mentioned above, however, the case in which partial information is known about the dependence between variables has not been solved. There is currently no way to compute optimal bounds when some but not all specifying information is known about the dependence or when the correlation is known to be a certain value or to lie within some interval.

The ninth exemplar problem generalizes the first problem in that input distributions are unknown or uncertain. In the first problem, very little information about the input distributions was available. The more general situation, however, is that some information, and maybe a lot, is available that characterizes the input distributions but it is not enough to specify the distributions precisely. Because standard probabilistic assessment including Monte Carlo requires the selection of particular, well-specified probability distributions as inputs even when available empirical information does not justify such specificity, various strategies have been proposed to get around this problem.

One approach to the problem is to use the maximum entropy criterion (Jaynes, 1957) to define the distribution using whatever information is available about the variable. The strategy is to specify the distribution having the largest Shannon-Weaver entropy that is consistent with the available knowledge which forms constraints on the possible shapes. Thus, this approach makes no shape assumptions yet allows one to select the input distribution in an optimal way using only limited information about the variates. For instance, the criterion selects a uniform distribution when only upper and lower bounds on values are known, an exponential distribution when only a lower bound and the mean are known, and a normal distribution when only the mean and standard deviation are known. If other population statistics are known about the variates, the maximum entropy criterion can be used to derive a solution. Lee and Wright (1994) suggest how the criterion can be used in risk analysis.

As we mentioned above, however, the use of the maximum entropy criterion is often criticized. The approach confounds an analyst's uncertainty with stochasticity of the focal population. While a subjectivist interpretation of probability finds no fault with this, in risk analysis it is often considered objectionable, especially when the analysis is conducted by regulatory authorities (inter alia Vesely et al., 1981; also see Feller, 1968; Fisher, 1973). There are other methods that can be used to solve problems like the ninth exemplar, including second-order Monte Carlo simulation (Fisher, 1957; Good, 1965; Iman and

Helton, 1985; 1991; Helton, 1994), robust Bayesian analysis (Huber, 1980; Wasserman, 1990; Walley, 1991) and probability bounds analysis (Ferson et al., 1996). All of these methods accept and propagate the uncertainty about the input distributions through the assessment calculations.

The tenth exemplar problem in which the model itself is unknown or uncertain represents a profound but ubiquitous difficulty in risk analysis. Examples in human health are cancer risk assessments in which the appropriate model for low-dose extrapolations of dose-response information is questionable. Examples in ecological risk assessment include extinction risk analyses for endangered species in which the best model of density dependence relating changes in vital rates to population size is unknown. In both cases, there are several functions with different kinds of nonlinearities that are traditionally used, but without careful and specific study of the underlying biology, an analyst cannot be sure about which model should be used. As might be anticipated, however, the choice can make a big difference. Cullen (1995) explored the consequences on the final result of using different models in a risk assessment. Mosleh and Bier (1992) considered the somewhat more abstract problem of selecting the level of aggregation at which to build a model for use in a risk analysis.

There is very little agreement about how uncertainty about the form of the model should be handled. Some argue that the uncertainty should be averaged into the process in a way very similar to how parametric uncertainty is treated. For instance, Holland and Sielken (1994) have argued that it is reasonable to average the results from different models, weighted by the respective evidence each one has supporting its claim as the truth. Sometimes this weighting degenerates to tallying supporters for the models, or counting numbers of papers published on either side of a debate, as though popularity were a measure of truth.

Others disagree with this approach, suggesting it is nonsensical to average the results of mutually exclusive theories (e.g., Morgan and Henrion, 1990; Committee on Risk Assessment of Hazardous Air Pollutants, 1994; Finkel, 1995). Both probability bounds analysis (Ferson et al., 1996) and second-order Monte Carlo simulation (Iman and Helton, 1985; 1991) have the potential to handle model uncertainty comprehensively if there are a finite number of models to choose from, although the latter grows in complexity combinatorially as the number of choices about model form increases.

DISCUSSION

Like any analytical tool, Monte Carlo methods can yield incorrect or unjustifiable results whenever their assumptions are false or are not justified empirically. The diverse problems exemplified above are the result fundamentally of two basic kinds of mistakes: (1) using precise distributions without empirical justification, and (2) inappropriate modeling of dependencies among variables. Although it is obvious that improper use of precise distributions and independence assumptions is the fault of the analyst rather than the analysis, it is also fair to say that Monte Carlo analysis, and indeed probability theory in general, would not be of much practical use without such assumptions. These issues have long been recognized. As Jaynes (1957, page 622) noted, ‘‘This problem of specification of probabilities in cases where little or no information is available, is as old as the theory of probability.’’ Whittaker (1990, page 23) explained, ‘‘Independence is

cemented into the very foundation of probability theory: it is a theme that recurs in the far frontiers of research, and one that permeates all applications of probability to scientific investigation.''

Moreover, the discrepancies between the Monte Carlo results and the more comprehensive solutions to the exemplar problems should not be dismissed as trivial or unimportant. In some cases, the Monte Carlo result may be approximately true; in other cases, it may be decidedly far from the truth. But either way, it is important for risk analysts at least to know the formally correct solution to a problem even if it is not used every time. Moreover, there emerges an issue of good professional practice. As a general rule, analysts should not make unjustified assumptions merely for the sake of computational convenience. Risk assessments should not be back-of-envelope calculations. Whatever is known empirically should be stated explicitly and, to the fullest extent possible, no further assumptions should be made merely for the sake of making calculation easier. Whenever further assumptions are necessary to tame an otherwise intractable problem, they should be discussed explicitly in the assessment documentation and it should be emphasized that this is their justification.

All of the arguments in this paper rest on the presumption that risk analysis should be based on a frequentist conception of probability. This may be a controversial presumption, so some discussion about it is in order. Among probabilists, the frequentist interpretation of probability has been largely abandoned. Most hold that probability theory is not the mathematical science of frequencies, despite the widespread popular view that it is, and notwithstanding the occasional protestation that it ought to be from empirical scientists (e.g., Rowlinson, 1970). Jaynes (1979) likened the assertion that probability measures frequencies to someone claiming, 500 years after Columbus, that the world is flat. Most theorists consider probability theory a calculus of subjective uncertainty, which is to say the uncertainty of some analyst rather than the objective variation that exists independent of humans (see Morgan and Henrion, 1990). Neapolitan (1992) cogently explains the divergence of frequentist and subjectivist interpretations over the history of probability theory and explains why the subjectivist view has come to dominate the field.

The now-out-of-favor frequentist interpretation holds that probability is a measure of the frequency of an event occurring. Under this strict interpretation, probability theory can only be applied when the events can, at least in principle, be repeated, and it depends on the idea of a statistical ensemble. For unique situations, where there is no ensemble and repetition is not possible, this interpretation cannot be made. Morgan and Henrion (1990, page 49) assert, ''[the frequentist interpretation] renders the theory of probability virtually inapplicable to real-world decision making, outside games of chance involving dice or cards.''

The subjectivist interpretation holds that probability is a measure of a person's subjective belief that an event will occur. While observed frequencies will often be useful in establishing values for probabilities, subjectivists argue that it is silly to restrict probability theory to only those problems that admit frequentist interpretations when it is clearly useful in a far wider range of problems. Subjectivists believe that all probabilities are conditional on available information which may be different for different people. For

instance, while most spectators at a horse race might regard the probability of a particular unfavored horse's winning as low, the probability of this event is much different for a person who has just fixed the race.

The logical conclusion of this idea is that there is no such thing as a *true* probability distribution that is an attribute of the physical world. This is an extension of de Finetti's (1974) idea that probability does not exist as an objective quantity and only has meaning as a characterization of subjective feelings of uncertainty. Frequencies are only useful in informing these feelings and for estimating probabilities. One of the results of this, as Savage (1954; 1971) and others have argued, is that there is no distinction to be made between frequency and uncertainty due to ignorance (cf. Hoffman and Hammonds, 1994). Probability is probability, and probability includes both forms of uncertainty. Consequently, attempts by risk analysts to distinguish the two are regarded as misguided. Kyburg (1989) showed that all second-order formulations that would treat variability and lack of knowledge separately are unnecessary under a subjectivist interpretation and should not be conducted.

Despite the attraction of being able to use probability theory in much wider contexts, it seems that there are some problems that require a purely frequentist interpretation. For instance, in applications of risk analysis, especially for problems in public health or environmental protection, it does not seem prudent to convolve an analyst's *subjective feelings* into an analysis, no matter how well trained and nobly motivated the analyst may be. Although we are obviously keen to include whatever dependable information exists about a problem in the analysis, the purpose of risk analysis as it is applied to, say, public health issues is not to compute expectations based on the subjective knowledge of an individual analyst. The knowledge and uncertainty of one analyst might, and should, be different from those of another. What relevance for a risk analysis could there be in an analyst's *beliefs*, as distinguished from any evidence he or she has collected? Why should anyone's beliefs be muddled with objective scientific information? The manipulation of such beliefs could only result in a formal calculation of suspicions, rather than true frequencies of adverse events.

If probability theory is not the mathematical science of frequencies, then perhaps risk analysis should be given this duty. One need not assert that all forms of uncertainty can be expressed in terms of precisely known frequencies. Nevertheless, Morgan and Henrion's pessimism on the issue seems overwrought. Just because quantitative aspects of probabilistic events are unknown does not mean that we must use subjective estimates for them. We can objectively *bound* our ignorance using the ordinary methods of science. It is clear that an approach based fundamentally on frequencies but generalized to admit bounded ignorance about them can serve as the mathematical underpinning of a risk analysis with wide applicability. It may be of practical interest to use beliefs in secondary 'what-if' analyses and this use is perfectly reasonable. It is important, however, to carefully distinguish cases when subjective estimates are employed from those that purport to summarize the state of empirical knowledge.

Acknowledgments

This paper was improved by thoughtful comments by Daniel Berleant of the University of Arkansas, David Burmaster of Alceon Corporation, Matthew Spencer of Applied Biomathematics, Clark Carrington of the Food and Drug Administration, and three anonymous reviewers.

REFERENCES

Barton, W.T. 1989. Response from Palisade Corporation. *Risk Analysis* **9,** 259-260.

Berleant, D. 1993. Automatically verified reasoning with both intervals and probability density functions. *Interval Computations*, no. 2, 48-70.

Berleant, D. 1996. Automatically verified arithmetic on probability distributions and intervals. In: *Applications of Interval Computations*, pages 227-244. (Kearfott, R.B. and Kreinovich, V., Eds.) Dordrecht, Kluwer Academic Publishers.

Bukowski, J., Korn, L., and Wartenberg, D. 1995. Correlated inputs in quantitative risk assessment: the effects of distributional shape. *Risk Analysis* 15:215-219.

Burgman, M., S. Ferson and H.R. Akçakaya. 1993. *Risk Assessment in Conservation Biology*. Chapman & Hall, London.

Burmaster, D.E. and Thompson, K.M. 1995. Backcalculating cleanup targets in probabilistic risk assessments when the acceptability of cancer risk is defined under different risk management policies. *Human and Ecological Risk Assessment* **1,** 101-120.

Burmaster, D.E. and Udell, E.C. 1990. A review of Crystal Ball. *Risk Analysis* **10,** 343-345.

Burmaster, D.E., Lloyd, K.J., and Thompson, K.M. 1995. The need for new methods to backcalculate soil cleanup targets in interval and probabilistic cancer risk assessments. *Human and Ecological Risk Assessment* **1,** 89-100.

Bratley, P., Fox, B.L., and Shrage, L.E. 1983. *A Guide to Simulation*. Springer Verlag, New York.

Committee on Risk Assessment of Hazardous Air Pollutants 1994. *Science and Judgment in Risk Assessment*. Washington, National Research Council.

Cullen, A.C. 1995. The sensitivity of probabilistic risk assessment results to alternative model structures: a case study of municipal waste incineration. *Journal of the Air & Waste Management Association* **45,** 538-546.

de Finetti, B., 1974. *Theory of probability; a critical introductory treatment*. Translated by A. Machi and A. Smith. Wiley, New York.

Efron, B. and Diaconis, P. 1983. Computer intensive methods in statistics. *Scientific American* (May), 116-130.

Feller, W. 1968. *An Introduction to Probability Theory and Its Applications*, Volume 1, third edition. New York, John Wiley & Sons.

Ferson, S. 1994. Naive Monte Carlo methods yield dangerous underestimates of tail probabilities. In: *Proceedings of the High Consequence Operations Safety Symposium*, SAND94-2364, pages 507-514. (Cooper, J.A., Ed.) Albuquerque, Sandia National Laboratories.

Ferson, S. 1995. Using approximate deconvolution to estimate cleanup targets in probabilistic risk analyses. In: *Hydrocarbon Contaminated Soils*, Volume 5, pages 245-254. (Kostecki, P.T., Calbrese, E.J., and Bonazountas, M., Eds.) Amherst, Amherst Scientific Publishers.

Ferson, S. and M. Burgman. 1995. Correlations, dependency bounds and extinction risks. *Biological Conservation* **73,** 101-105.

Ferson, S. and Ginzburg, L.R. 1996. Different methods are needed to propagate ignorance and variability. *Reliability Engineering and Systems Safety* [in press].

Ferson, S. and Long, T.F. 1995. Conservative uncertainty propagation in environmental risk assessments. In: *Environmental Toxicology and Risk Assessment - Third Volume, ASTM STP 1218*, pages 97-110. (Hughes, J.S., Biddinger, G.R. and Mones, E., Eds.) Philadelphia, American Society for Testing and Materials.

Ferson, S., Ginzburg, L.R. and Akçakaya, H.R. 1996. Whereof one cannot speak: when input distributions are unknown. *Risk Analysis* [to appear].

Finkel, A.N. 1995. A second opinion on an environmental misdiagnosis: the risky prescriptions of *Breaking the Viscious Circle. New York University Environmental Law Journal* **3,** 295-381.

Fisher, R.A. 1957. The underworld of probability. *Sankhya* **18,**201-210.

Fisher, R.A. 1973. *Statistical Methods and Scientific Inference*. New York, Hafner Press.

Flury, B.K. 1986. On sums of random variables and independence. *The American Statistician* **40,** 214-215.

Frank, M.J., Nelsen, R.B., and Schweizer, B. 1987. Best-possible bounds for the distribution of a sum—a problem of Kolmogorov. *Probability Theory and Related Fields* **74**, 199-211.

Fréchet, M., 1935. Généralisations du théorème des probabilités totales. *Fundamenta Mathematica* **25**, 379-387.

Good, I.J. 1965. *The Estimation of Probabilities*. Cambridge, MIT Press.

Grandy, W.T., Jr. and Schick, L.H. 1991. *Maximum Entropy and Bayesian Methods*. Dordrecht, Kluwer Academic Publishers.

Helton, J.C. 1994. Treatment of uncertainty in performance assessments for complex systems. *Risk Analysis* **14,** 483-511.

Hoffman, F.O. and Hammonds, J.S. 1994. Propagation of uncertainty in risk assessments: The need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Analysis* **14,** 707-712.

Holland, C.D. and Sielken, R.L., Jr. 1994. *Quantitative Cancer Modeling and Risk Assessment*. Englewood Cliffs, Prentice Hall.

Huber, P.J. 1980. *Robust Statistics*. New York, John Wiley.

Iman, R.L. and Conover, W.J. 1982. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics* **B11,** 311-334.

Iman, R.L. and Helton, J.C. 1985. *A Comparison of Uncertainty and Sensitivity Analysis Techniques for Computer Models*, NUREG/CR-3904, SAND84-1461. Albuquerque, Sandia National Laboratories.

Iman, R.L. and Helton, J.C. 1991. The repeatability of uncertainty and sensitivity analyses for complex probabilistic risk assessments. *Risk Analysis* 11: 591-606.

Jansson, P.A. 1984. *Deconvolution with Applications in Spectroscopy*. Orlando, Academic Press.

Jaynes, E.T. 1957. Information theory and statistical mechanics. *Physical Review* 106:620-630.

Jaynes, E.T. 1979. Where do we stand on maximum entropy? In: *The Maximum Entropy Formalism*. (Levine, R.D. and Tribus, M., Eds.) Cambridge, MIT Press.

Kyburg, H.E., Jr. 1989. Higher order probabilities. In: *Uncertainty in Artificial Intelligence*, volume 3, pages 15-22. (Kanal, L.N., Levitt, T.S., and Lemmer, J.F., Eds.) Amsterdam, Elsevier Science Publishers.

Lee, R.C. and Wright, W.E. 1994. Development of human exposure-factor distributions using maximum-entropy inference. *Journal of Exposure Analysis and Environmental Epidemiology* **4**, 329-341.

Levine, R.D. and Tribus, M. 1979. *The Maximum Entropy Formalism*. Cambridge, MIT Press.

Moore, R.E. 1966. *Interval Analysis*. Englewood Cliffs, Prentice-Hall.

Morgan, M.G. and Henrion, M. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, Cambridge University Press.

Mosleh, A. and Bier, V. 1992. On decomposition and aggregation error in estimation: some basic principles and examples. *Risk Analysis* **12,** 203-214.

Neapolitan, R.E. 1992. A survey of uncertain and approximate inference. In: *Fuzzy Logic for the Management of Uncertainty*, pages 55-82. (Zadeh, L. and Kacprzyk, J., Eds.) New York, John Wiley and Sons.

Nelsen, R.B. 1986. Properties of a one-parameter family of bivariate distributions with specified marginals. *Communications in Statistics (Theory and Methods)* **A15,** 3277-3285.

Nelsen, R.B. 1987. Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics (Simulation and Computation)* **B16,** 199-208.

Nelsen, R.B. 1991. Copulas and association. In: *Advances in Probability Distributions with Given Marginals: Beyond the Copulas*, pages 51-74. (Dall'Aglio, G., Kotz, S., and Salinetti, S., Eds.) Dordrecht, Kluwer Academic Publishers.

Nelsen, R.B. 1992. On measures of association as measures of positive dependence. *Statistics and Probability Letters* **14,** 269-274.

Nelsen, R.B. 1995. Copulas, characterization, correlation, and counterexamples. *Mathematics Magazine* **68,** 193-198.

Neumaier, A. 1990. *Interval Methods for Systems of Equations*. Cambridge, Cambridge University Press.

Rowlinson, J.S. 1970. Probability, information and entropy. *Nature* **225,** 1196.

Salmento, J.S., Rubin, E.S., and Finkel, A.M. 1989. A review of @Risk. *Risk Analysis* **9,** 255-257.

Savage, L.J. 1954. *The Foundations of Statistics*. New York, Wiley.

Savage, L.J. 1971. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association* **66,** 783-801.

Scheuer, E.M. and Stoller, D.S. 1962. On the generation of normal random vectors. *Technometrics* **4,** 278-281.

Schweizer, B. and A. Sklar, 1983. *Probabilistic Metric Spaces*. North-Holland, New York.

Simon, J.L. and Bruce, P. 1991. Resampling: a tool for everday statistical work. *Chance* **4,** 22-32.

Smith, A.M. and Watson, I.A. 1980. Common cause failure—a dilemma in perspective. *Reliability Engineering* **1,** 127-142.

Vesely, W.E., Goldberg, F.F., Roberts, N.H., Haasl, D.F. 1981. *Fault Tree Handbook*, NUREG-0492. Washington, U.S. Nuclear Regulatory Commission.

Walley, P. 1991. *Statistical Reasoning with Imprecise Probabilities*. London, Chapman & Hall.

Wasserman, L.A. 1990. Prior envelopes based on belief functions. *Annals of Statistics* **18,** 454-464.

Whitt, W. 1976. Bivariate distributions with given marginals. *The Annals of Statistics* **4,** 1280-1289.

Whittaker, J. 1990. *Graphical Models in Applied Multivariate Statistics*. Wiley, Chichester, UK.

Williamson, R.C. and Downs, T. 1990. Probabilistic arithmetic I: Numerical methods for calculating convolutions and dependency bounds. *International Journal of Approximate Reasoning* **4,** 89-158.

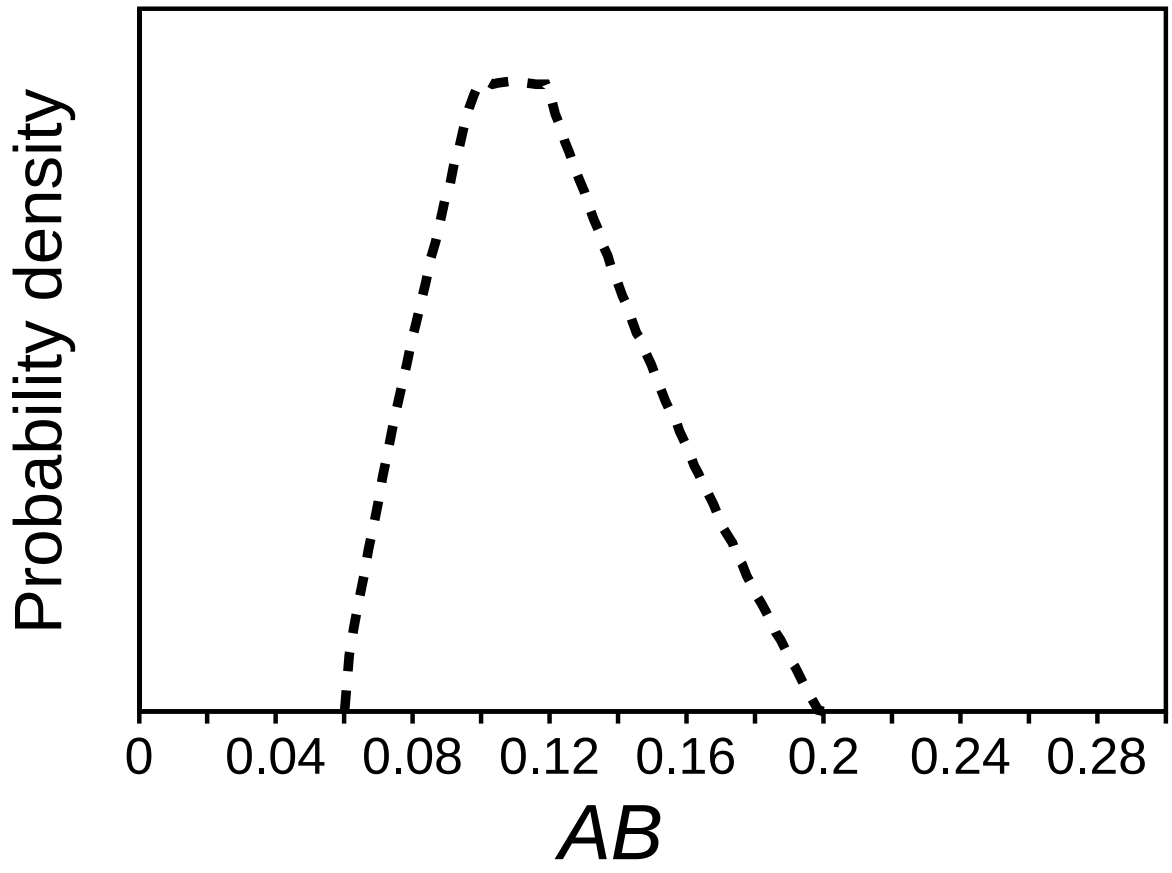

FIGURE 1. Monte Carlo estimate for distribution of the product in the first exemplar problem.

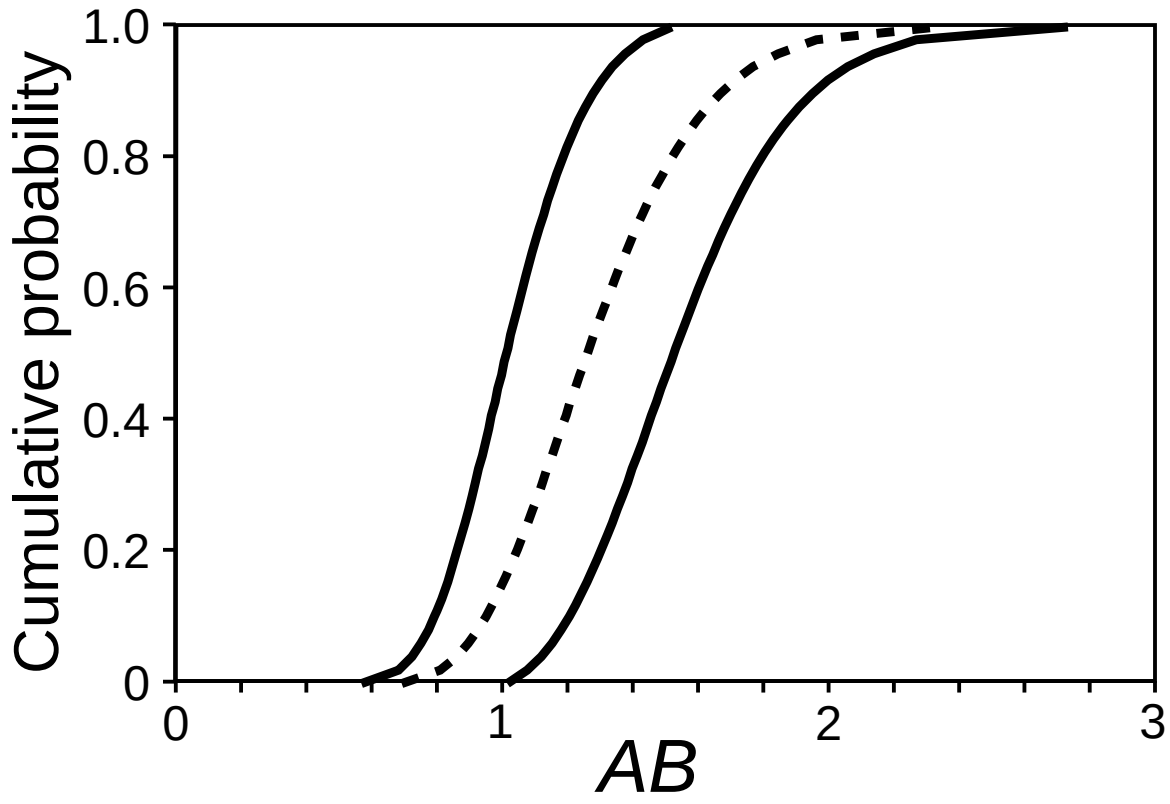

FIGURE 2**.** Bounds (solid lines) on the cumulative distribution function of the product in the second exemplar problem and the Monte Carlo estimate (dotted line) assuming a uniform distribution for *A*.

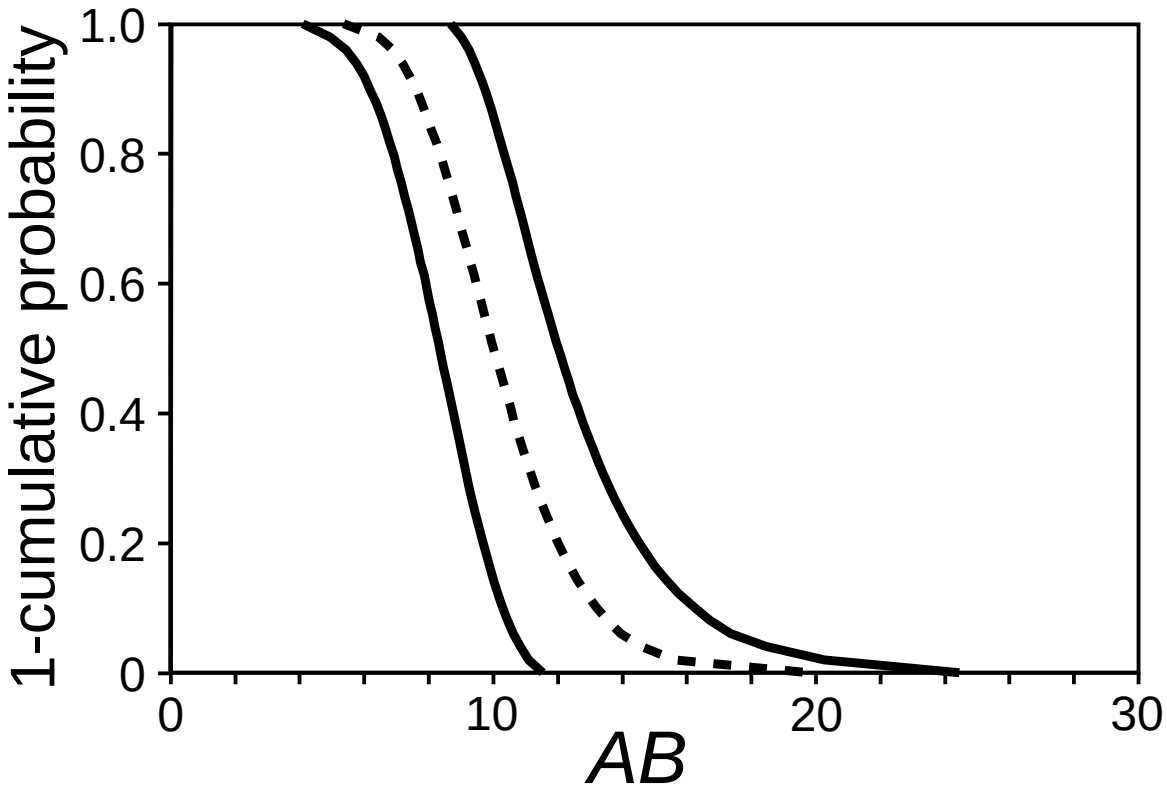

FIGURE 3**.** Bounds (solid lines) on the complementary cumulative distribution function, and the Monte Carlo estimate (dotted line) assuming independence, for the product of two lognormal random variables *A* and *B* specified by parameters in the third exemplar problem.

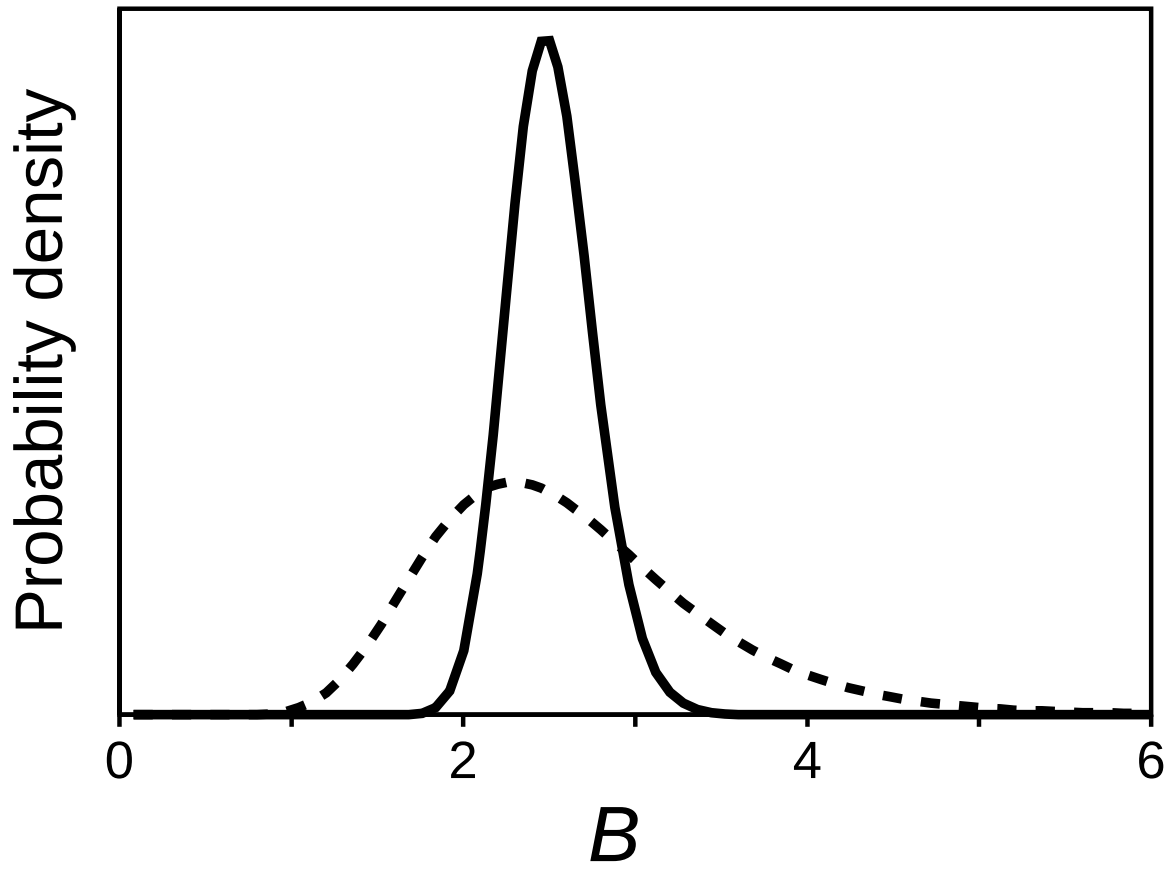

FIGURE 4. Discrepancy between the true distribution (solid line) and the estimate computed using Monte Carlo (dashed line) for the seventh exemplar problem.