results mean that it is usually practical to perform only a small number of experiments. These must be designed to shed as much light as possible on the issues of interest.

The first thing to recognize is that results from Monte Carlo experiments are necessarily random. At a minimum, this means that results must be reported in a way which allows readers to appreciate the extent of experimental randomness. Moreover, it is essential to perform enough replications so that the results are sufficiently accurate for the purpose at hand. The number of replications that is needed can sometimes be substantially reduced by using variance reduction techniques, which will be discussed in the next two sections. Such techniques are by no means always readily available, however. In this section, we consider various other aspects of the design of Monte Carlo experiments.

We first consider the problem of determining how many replications to perform. As an example, suppose that the investigator is interested in calculating the size of a certain test statistic (i.e., the probability of rejecting the null hypothesis when it is true) at, say, the nominal .05 level. Let us denote this unknown quantity by $p$. Each replication will generate a test statistic that either exceeds or does not exceed the nominal critical value. These can be thought of as independent Bernoulli trials. Suppose $N$ replications are performed and $R$ rejections are obtained. Then the obvious estimator of $p$, which is also the ML estimator, is $R/N$. The variance of this estimator is $N^{-1}p(1-p)$, which can be estimated by $R(N-R)/N^3$.

Now suppose that one wants the length of a 95% confidence interval on the estimate of $p$ to be approximately .01. Using the normal approximation to the binomial, which is surely valid here since $N$ will be a large number, we see that the confidence interval must cover $2 \times 1.96 = 3.92$ standard errors. Hence we require that

$$3.92 \left( \frac{p(1-p)}{N} \right)^{1/2} = .01. \tag{21.02}$$

Assuming that $p$ is .05, the nominal level of the test being investigated, we can solve (21.02) for $N$. The result is $N \cong 7299$. To be on the safe side (since $p$ may well exceed .05, implying that $R/N$ may have a larger variance) the investigator would probably choose $N = 8000$. This is a rather large number of replications and may be expensive to compute. If one were willing to let the 95% confidence interval on $p$ have a length of .02, one could make do with a sample one-quarter as large, or roughly 2000 replications.

If the objective of an experiment is to compare two or more estimators or two or more test statistics, fewer replications may be needed to obtain a given level of accuracy than would be needed to estimate the properties of either of them with the same level of accuracy. Suppose, for example, that we are interested in comparing the biases of two estimators, say $\hat{\theta}$ and $\tilde{\theta}$, of a

parameter the true value of which is $\theta_0$. On each replication, say the $j^{\text{th}}$, we obtain realizations $\hat{\theta}_j$ and $\tilde{\theta}_j$ of each of the two estimators. The biases of the two estimators are

$$B(\hat{\theta}) \equiv E(\hat{\theta} - \theta_0) \quad \text{and} \quad B(\tilde{\theta}) \equiv E(\tilde{\theta} - \theta_0),$$

which may be estimated by

$$\hat{B}(\hat{\theta}) = \frac{1}{N}\sum_{j=1}^{N}(\hat{\theta}_j - \theta_0) \quad \text{and} \quad \tilde{B}(\tilde{\theta}) = \frac{1}{N}\sum_{j=1}^{N}(\tilde{\theta}_j - \theta_0).$$

The difference between $B(\hat{\theta})$ and $B(\tilde{\theta})$ is

$$E(\hat{\theta} - \theta_0) - E(\tilde{\theta} - \theta_0) = E(\hat{\theta} - \tilde{\theta}), \tag{21.03}$$

which may be estimated by

$$\frac{1}{N}\sum_{j=1}^{N}(\hat{\theta}_j - \tilde{\theta}_j). \tag{21.04}$$

It is possible and indeed likely that the variance of (21.04) will be substantially smaller than the variance of either $\hat{B}(\hat{\theta})$ or $\tilde{B}(\tilde{\theta})$, because both $\hat{\theta}_j$ and $\tilde{\theta}_j$ depend on the same pseudo-random vector $\boldsymbol{u}^j$. The variance of (21.04) is

$$\frac{1}{N}V(\hat{\theta}) + \frac{1}{N}V(\tilde{\theta}) - \frac{2}{N}\text{Cov}(\hat{\theta}, \tilde{\theta}),$$

which will be smaller than the variance of either $\hat{B}(\hat{\theta})$ or $\tilde{B}(\tilde{\theta})$ whenever $\text{Cov}(\hat{\theta}, \tilde{\theta})$ is positive and large enough. This will very often be the case, since it is likely that $\hat{\theta}_j$ and $\tilde{\theta}_j$ will be strongly positively correlated. Thus it may require far fewer replications to estimate (21.03) than to estimate $B(\hat{\theta})$ and $B(\tilde{\theta})$ with the same level of accuracy. Of course, this assumes that $\hat{\theta}_j$ and $\tilde{\theta}_j$ are obtained using the same set of pseudo-random variates, but that is how the Monte Carlo experiment would normally be designed. We will encounter an idea similar to this one when we discuss the method of antithetic variates in the next section.

The second important thing to keep in mind when designing Monte Carlo experiments is that the results will often be highly sensitive to certain aspects of the experimental design and largely or totally insensitive to other aspects. Obviously, one will want to vary the former across the experiments while fixing the latter in a more or less arbitrary fashion. For example, many test statistics related to regression models are invariant to the variance of the error terms. Consider the ordinary $t$ statistic for $\alpha = 0$ in the regression

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \alpha\boldsymbol{z} + \boldsymbol{u}. \tag{21.05}$$