Chapter 7

Monte Carlo experiments

Reading:


7.1 Introduction

A Monte Carlo experiment is a computer-based simulation experiment. To learn about Monte Carlo experiments we will:

1. Review experimental designs, and then
2. Try out some Monte Carlo experiments for ourselves.

7.2 A summary of experimental design

7.2.1 Terminology

Consider an experiment. The individuals of the experiment are called the experimental units. Typically we call human experimental units subjects. The conditions we apply to the individuals in the study is called the treatment. The explanatory variables which (you believe) explain the effect of the treatments are called factors. The different values of the factors are called levels. An experimental design explains how the treatments are assigned to the experimental units in the experiment.
7.2.2 Why design at all?

Experiments are often designed to generalize to a population of interest. People will not believe the results of a study if the experiment is not realistic. Bad experimental designs encourage biased responses.

We want to minimize the effect of lurking variables, common responses and confounders. Then we can accurately know about the outcomes of interest, and can then ask

- “What is the treatment effect”?

In a good design:

1. We should randomly assign the treatments to experimental units (thus minimizing the selection bias).
2. We should control for the effect of unknown effects (“lurkers”) by comparing multiple treatments.
3. To get a notion of the variability of the response under the different treatments we should repeat or replicate the experiment for a number of individuals.

7.2.3 Randomization

It is hard to control for all the factors in an experiment. So, we use randomization to reduce the chances of selection bias. By randomly assigning the individuals to each treatment group, we create treatment groups that are similar. (There will still be some chance variability in the samples). This is called a completely randomized design.

To reduce biases we also have either:

1. (single) blind experiments – subjects do not know the treatment assignments, OR
2. double blind experiments – both the experimenter and the subjects do not know the assignments.

7.2.4 Experimental control

The control group are the individuals in a study which are given “no treatment”. Control also refers to the making sure that the other conditions in an experiment are held as constant as possible. Reason for using control:
1. A method of control provides a basis for comparison (e.g., how well did the treatment group do relative to the control?)

2. Control reduces experimental biases, e.g., selection of subjects and the placebo effect.

A placebo is a dummy treatment which does nothing, but “levels the playing field” in some sense. (people like to be treated!) We often observe a change in the response under a placebo treatment (this is called the placebo effect). Use of placebo is considered ethical if there is no effective standard treatment to give the control group. (Ethics have less of a rôle in Monte Carlo experiments!)

Another tool we can use for control is the block design. A block is a group of experimental units that are similar in some way (with respect to the response). A block design accounts for the fact that some individuals in an experiment may be similar:

1. We split the experimental units into blocks.
2. Within each block, we assign the treatments randomly to the experimental units.

7.2.5 Replication

We replicate an experiment to reduce the variability due to chance. As an example, consider these four situations:

<table>
<thead>
<tr>
<th>Situation</th>
<th>#receive treatment</th>
<th>#receive placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>150</td>
<td>150</td>
</tr>
</tbody>
</table>

Which situation has the most information about the variability of the response in the two treatment groups?

7.2.6 Detecting a difference

Key idea: we wish to detect a difference in the responses (typically the average response) so large that it is unlikely to happen by chance – a statistically significant result.

Remember that collecting data can be hard or expensive. We often have to compromise between the power to detect a difference and the sample size we collect. People often make sample size calculations to decide upon the size of sample they would need to detect a specific difference.
7.3 Example Monte Carlo experiments

7.3.1 Confidence intervals for a binomial proportion

For a positive integer $n$, let $X_1, \ldots, X_n$ be a random sample of Binomial(1, $p$) (Bernoulli) RVs. Suppose we are interested in estimating the population proportion parameter $p$ ($0 < p < 1$).

One point estimate of $p$ is
\[ \hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}, \]
and a $100(1 - \alpha)$% large sample confidence interval for $p$ is given by
\[ \hat{p} \pm z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \]
where $z_{1-\alpha/2}$ is the $1 - \alpha/2$ quantile of the standard normal distribution.

Write a Monte Carlo experiment to estimate the coverage probability for the large sample confidence interval for a binomial proportion.

7.3.2 Demonstrating the central limit theorem

Let $X_1, X_2, \ldots X_n$ be a sample of size $n$ drawn randomly from a population with mean $\mu$ and variance $\sigma^2$. We assume the mean and variance are finite numbers.

The Central Limit Theorem says that
\[ \sqrt{n}(\bar{X} - \mu) \rightarrow_d N(0, \sigma^2), \]
as $n \rightarrow \infty$.

Suppose that $X_1, \ldots, X_n$ is a random sample of size $n$ drawn from a population that can be described using a $t$ distribution on 5 degrees of freedom.

Write a simulation experiment to demonstrate the central limit theorem. For which values of $n$ would you feel confident in using a normal distribution as an approximate sampling distribution for $\bar{X}$?

7.3.3 A Monte Carlo P-value

For the fev dataset, carry out a t-test to see if the mean female fev value for those aged over 10 is equal to 3. Design a simulation (experiment) to estimate the P-value for this t-test. How well does it match with the actual P-value calculated using the t-test?