Note: The references mentioned in the exercises refer to the textbook (Sutton and Barto) in the 2nd edition.

7 n-step Bootstrapping

42. Exercise 7.1 In Chapter 6 we noted that the Monte Carlo error can be written as the sum of TD errors (6.6) if the value estimates don't change from step to step. Show that the n-step error used in (7.2) can also be written as a sum TD errors (again if the value estimates don't change) generalizing the earlier result. (page 143)

43. Implementation Task: n-step Algorithm

    Exercise 7.2 With an n-step method, the value estimates do change from step to step, so an algorithm that used the sum of TD errors (see previous exercise) in place of the error in (7.2) would actually be a slightly different algorithm. Would it be a better algorithm or a worse one? Devise and program a small experiment to answer this question empirically.

8 Planning and Learning

44. Implementation Task: n-step Algorithm vs Planning

    Try to reproduce Figure 8.2 on page 165. Then apply a multi-step algorithm to this problem. What does the performance comparison look like?

45. Exercise 8.1 The nonplanning method looks particularly poor in Figure 8.3 because it is a one-step method; a method using multi-step bootstrapping would do better. Do you think one of the multi-step bootstrapping methods from Chapter 7 could do as well as the Dyna method? Explain why or why not.