ExplorationChanges

Rasmus Brostrøm

2024-04-16

Exploration time

In the algorithm it is given that the exploration time should be of the form:

$$S_T = T^{2/3}$$

to balance the regret from exploring and the regret from exploiting. We will now look at how changing the order of the exploration time affects the average regret. To do so we will look at 5 different exploration formulas, where each exploration formula is simulated 100 times for the time horizons $T \in \{100, 200, \ldots, 5000\}$ with the four combinations of drift and payoff function given by the the linear drift function:

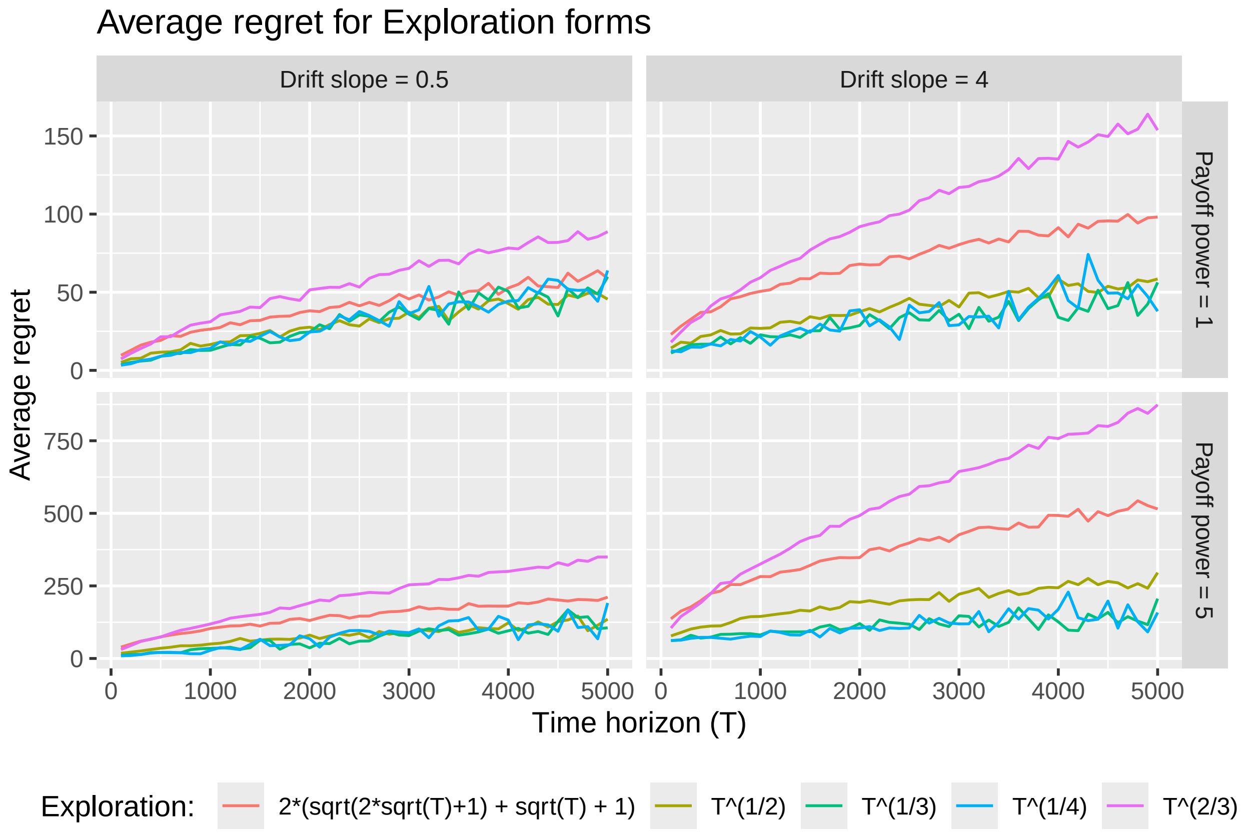
$$b(x) = -Cx, \quad C \in \{0.5, 4\}$$

and the payoff function:

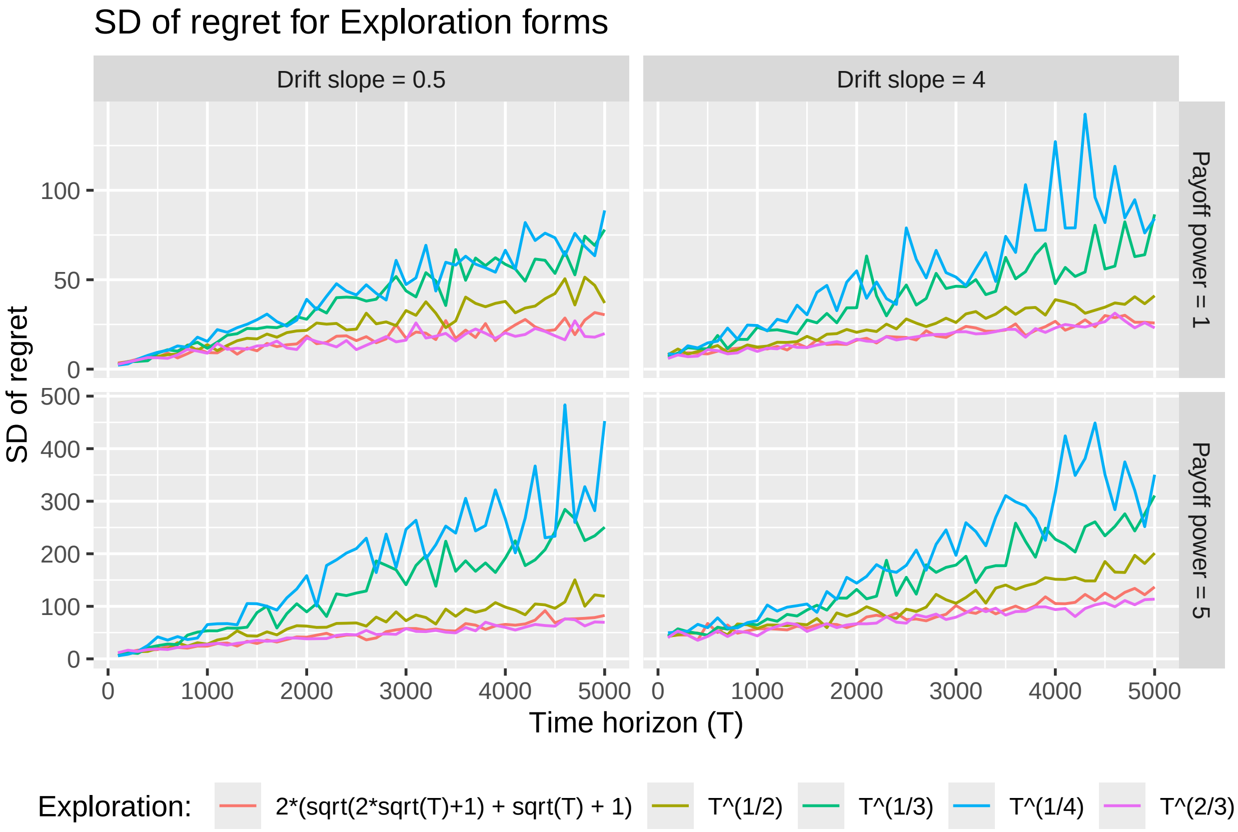
$$g(x) = 0.9 - |1 - x|^p, \quad p \in \{1, 5\}$$

The exploration forms that we are looking at are:

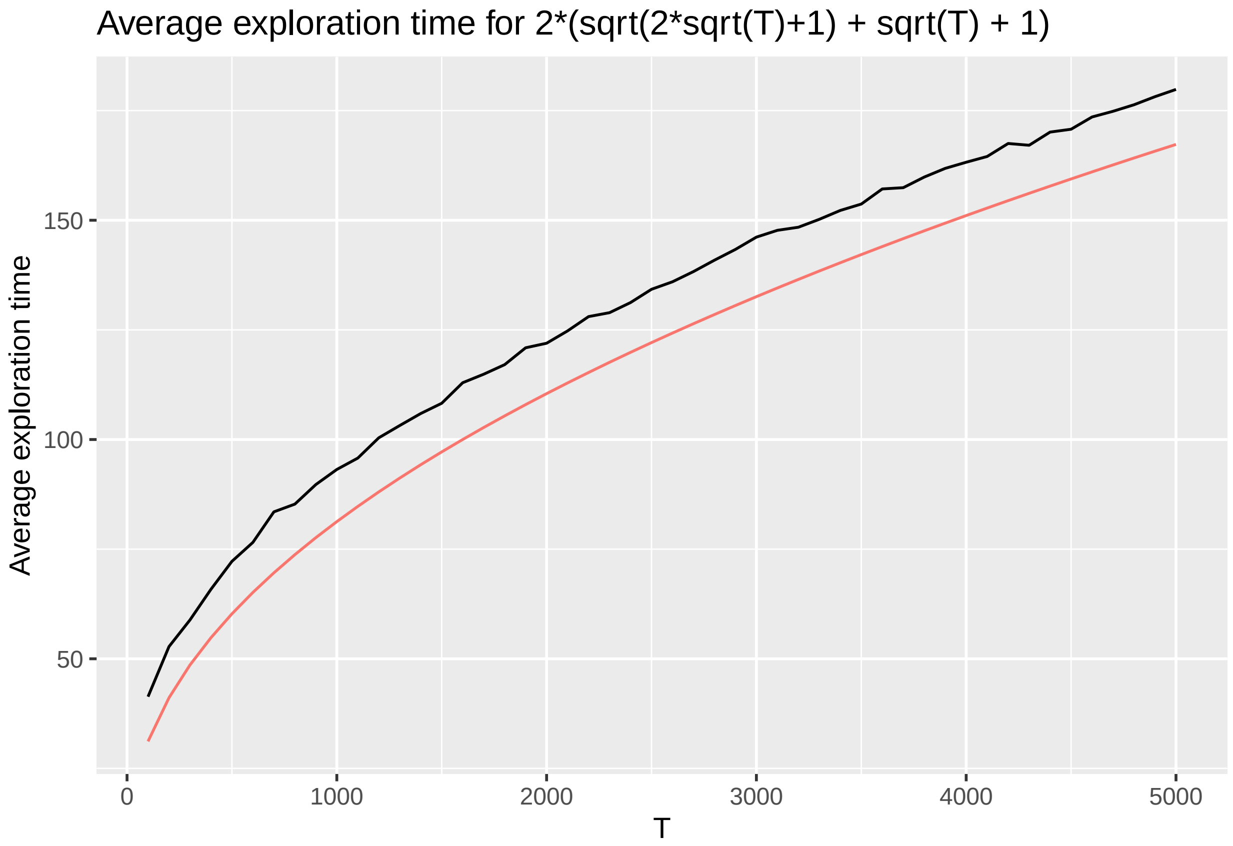
$$S_T = 2 \cdot (\sqrt{2 \cdot \sqrt{T} + 1} + \sqrt{T} + 1) S_T = T^k, \quad k \in \{1/4,\ 1/3,\ 1/2,\ 2/3\}$$

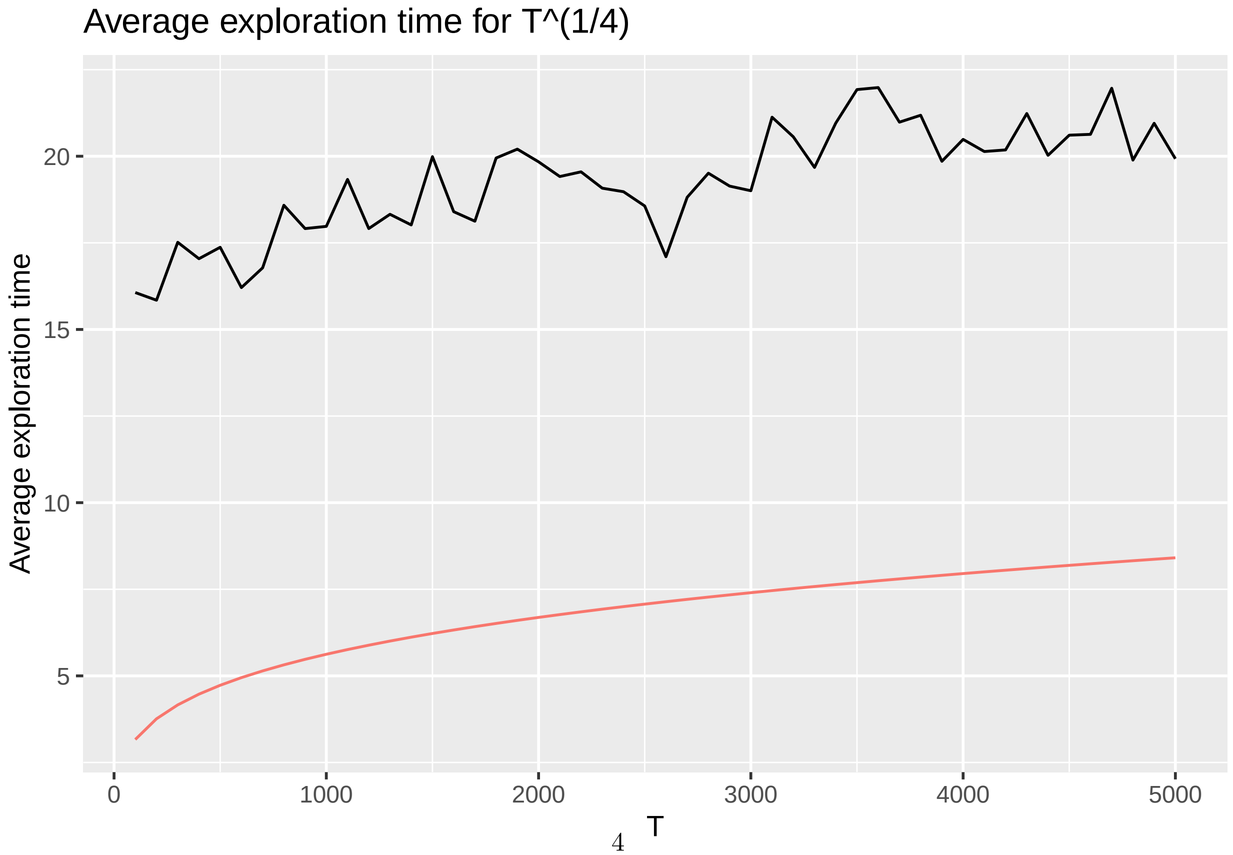


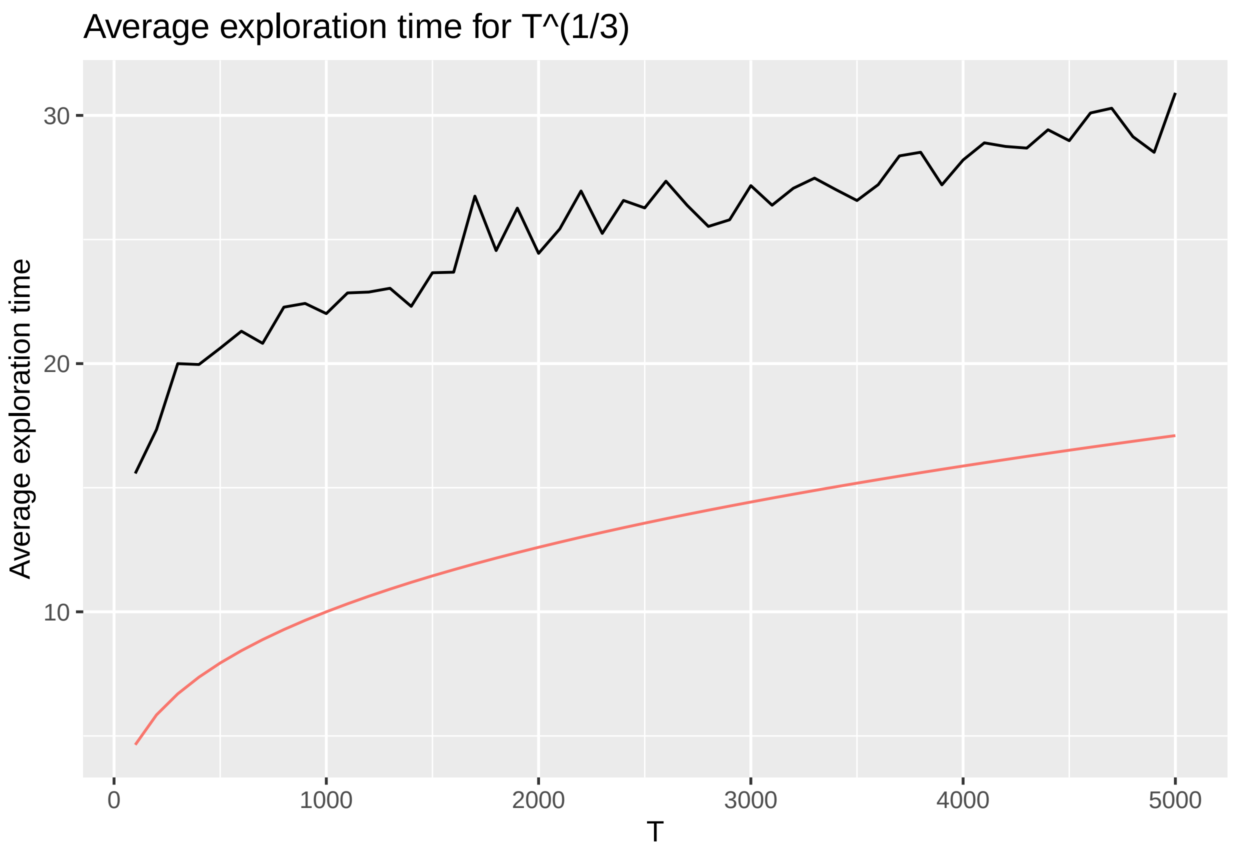
Lets also look at the standard deviation of the regret to see how changing the exploration times affects the stability of the regret

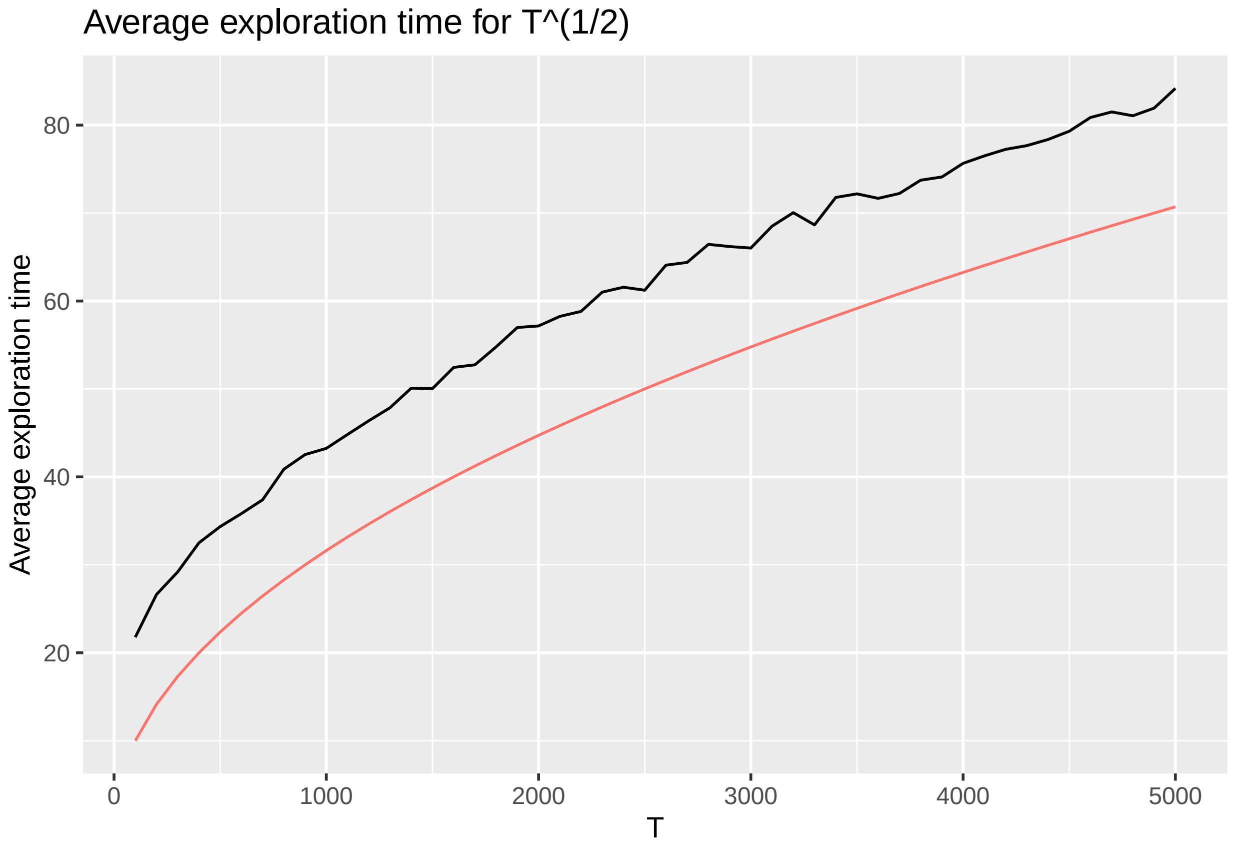

SD of regret for Exploration forms
Drift slope = 0.5
Drift slope = 4
Payoff power = 1
Payoff power = 5
SD of regret
0
50
100
0
100
200
300
400
500
0
1000
2000
3000
4000
5000
Time horizon (T)
Exploration:
2*(sqrt(2*sqrt(T)+1) + sqrt(T) + 1)
T^(1/2)
T^(1/3)
T^(1/4)
T^(2/3)

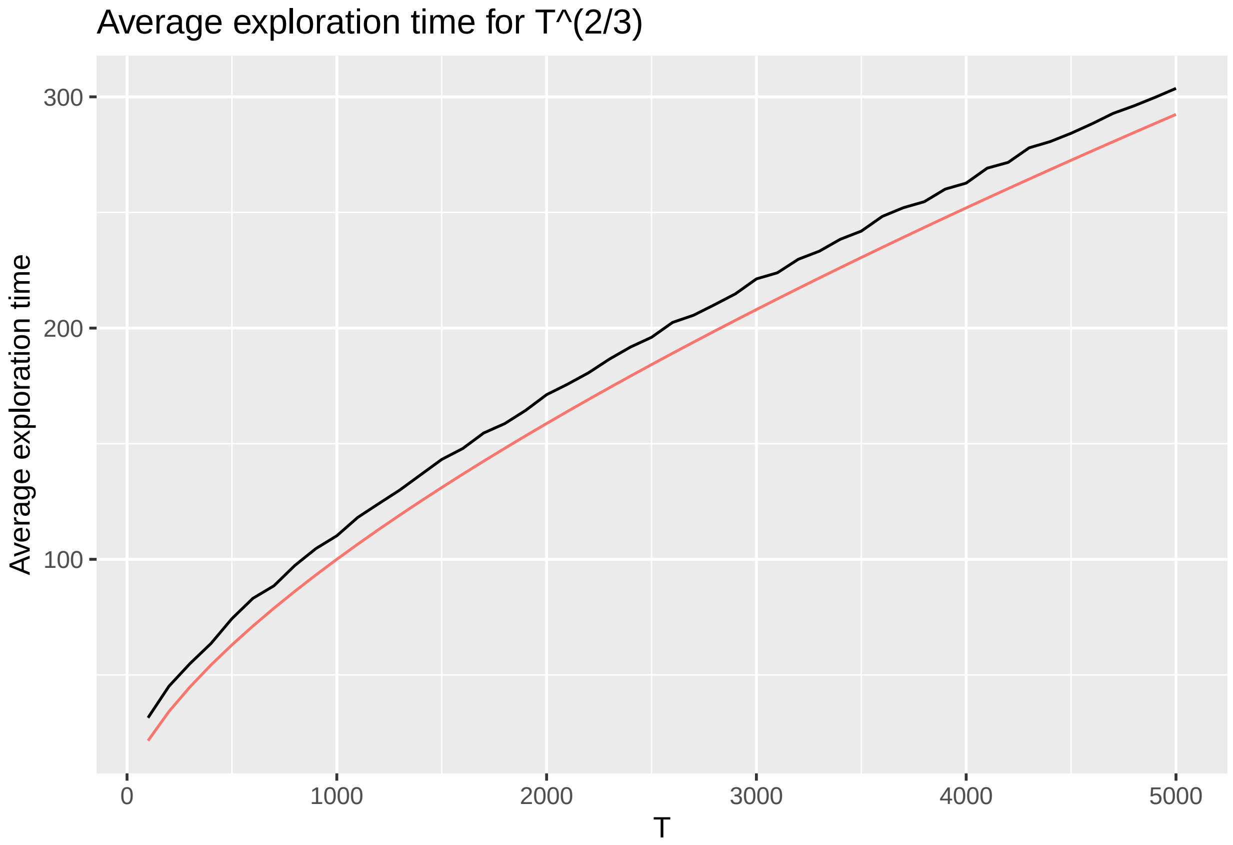
Lets confirm that the exploration times actually fit the corresponding exploration forms:



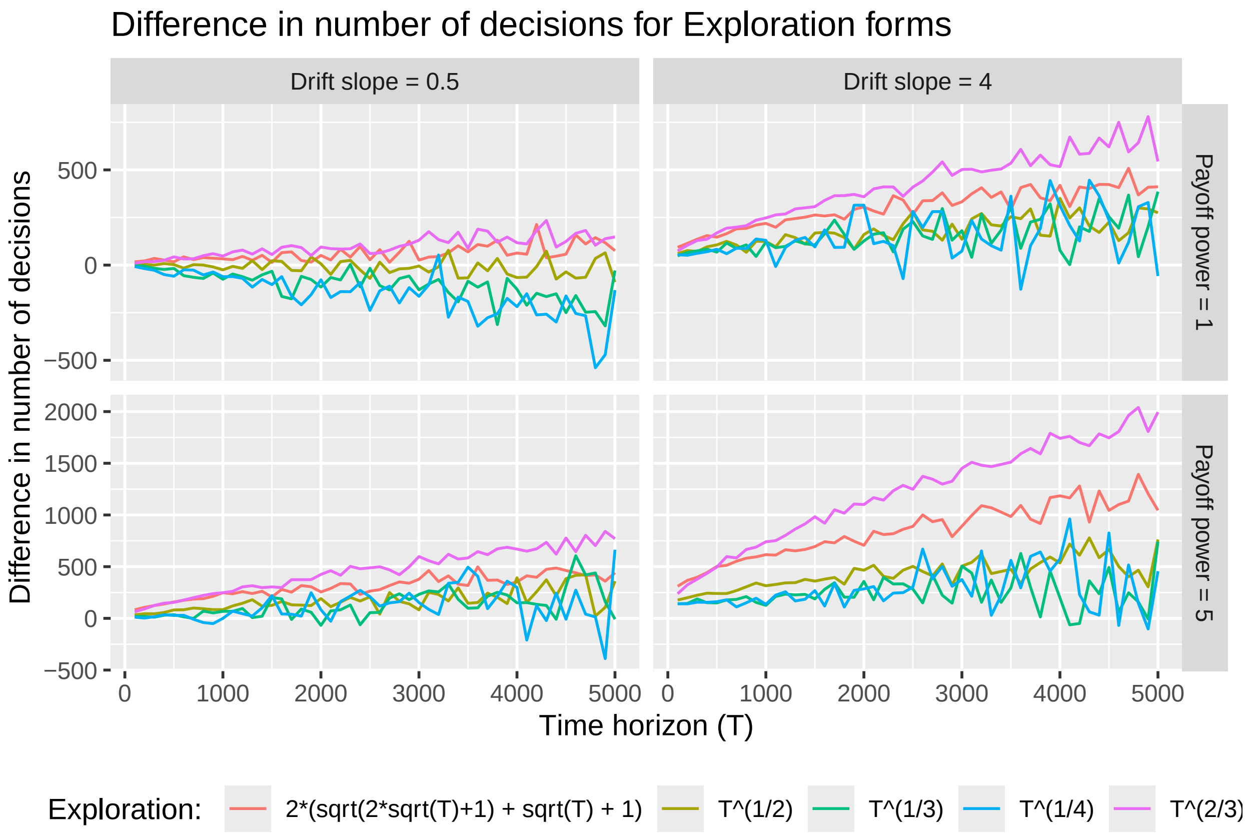



Average exploration time for T^(1/3)
Average exploration time
30
20
10
0
1000
2000
3000
4000
5000
T


Average exploration time for T^(1/2)
Average exploration time
20
40
60
80
0
1000
2000
3000
4000
5000
T


Average exploration time for T^(2/3)
Average exploration time
300
200
100
0
1000
2000
3000
4000
5000
T

Difference in number of decisions:


Difference in number of decisions for Exploration forms
Drift slope = 0.5
Drift slope = 4
Payoff power = 1
Payoff power = 5
Difference in number of decisions
Time horizon (T)
Exploration:
2*(sqrt(2*sqrt(T)+1) + sqrt(T) + 1)
T^(1/2)
T^(1/3)
T^(1/4)
T^(2/3)