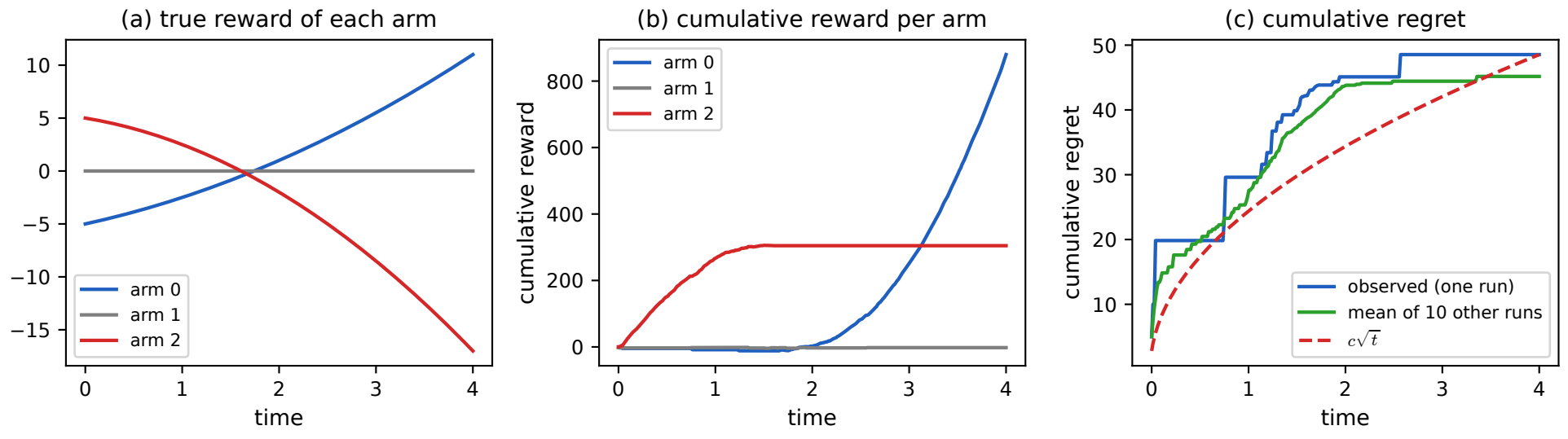

(a) true reward of each arm
10
5
0
−5
−10
−15
arm 0
arm 1
arm 2
0
1
2
3
4
time
(b) cumulative reward per arm
cumulative reward
800
600
400
200
0
arm 0
arm 1
arm 2
0
1
2
3
4
time
(c) cumulative regret
cumulative regret
50
40
30
20
10
observed (one run)
mean of 10 other runs
$c\sqrt{t}$
0
1
2
3
4
time