

1, Answer to the question in Part 1. Show your calculations. (Why does the optimal policy...in Figure 17.3?)

Expected utility for going left $0.655 * 0.8 + 0.611 * 0.1 + 0.660 * 0.1 = 0.6511$

Expected utility for going up $0.660 * 0.8 + 0.655 * 0.1 + 0.388 * 0.1 = 0.6323$

Since the Expected value of going left (go to state (2, 1)) is higher than going up, this is why the optimal policy for $R(s) = -0.04$ say to move left from the state(3, 1).

2, Answers to the following: Compare the means/standard devs. with the expected utility of the initial state. Are the means similar to the expected utility and to each other? How do the results change as we have more runs? The values of the means/standard deviations and the expected value should be included in the answer.

Expected utility: 0.387925

First run: 0.4

10-run average utility: 0.632

10-run standard dev: 0.118389

100-run average utility: 0.3064

100-run standard dev: 0.672719

1000-run average utility: 0.371

1000-run standard dev: 0.601363

Answer: the expected value of initial state is 0.387925, although the average(mean) change when we have more run, we can still notice that as we have more run, the average utility is getting closer to the expected utility. For example, in 10 run, since the randomness, the means are 0.632 and standard deviation is 0.118389, which is quite far from the expected utility, however, when we have 100 run, the means change to 0.3064, and standard deviation become 0.672719. We can see that the means is getting closer to 0.387. When we have 1000 run, the means is 0.371, which is even more closer to the expected utility. Also, the standard deviation change is also become smaller, which means the standard deviation is also start to converge to some value.

3, Full output from P1-output.txt (the thresholds and policies)

Up Bound(approximate 0): -0.0001

(Up bound is no a threshold, there are 8 threshold here)

Optimal Policy:

> > > +1
^ X < -1
^ < < v

Threshold value: -0.0221455

Optimal Policy:

> > > +1
^ X < -1
^ < < <*

Threshold value: -0.0273574

Optimal Policy:

> > > +1
^ X ^* -1
^ < < <

Threshold value: -0.0448332

Optimal Policy:

> > > +1
^ X ^ -1
^ < ^* <

Threshold value: -0.0849889

Optimal Policy:

> > > +1
^ X ^ -1
^ >* ^ <

Threshold value: -0.452624

Optimal Policy:

> > > +1
^ X ^ -1
^ > ^ ^*

Threshold value: -0.731139

Optimal Policy:

> > > +1
^ X ^ -1
>* > ^ ^

Threshold value: -1.56426

Optimal Policy:

> > > +1
^ X ^ -1
> > >* ^

Threshold value: -1.64971

Optimal Policy:

> > > +1
^ X >* -1
> > > ^

4, Full output from P2-output.txt. (means, standard devs, expected val)

Expected utility: 0.387925

First run: 0.4

10-run average utility: 0.632

10-run standard dev: 0.118389

100-run average utility: 0.3064

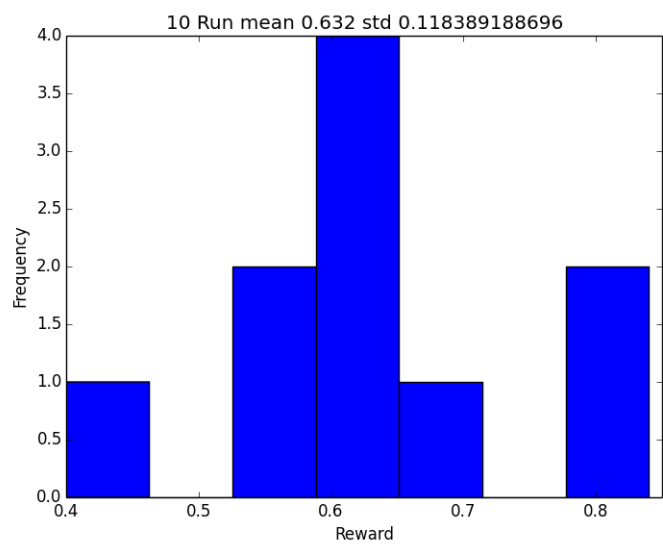
100-run standard dev: 0.672719

1000-run average utility: 0.371

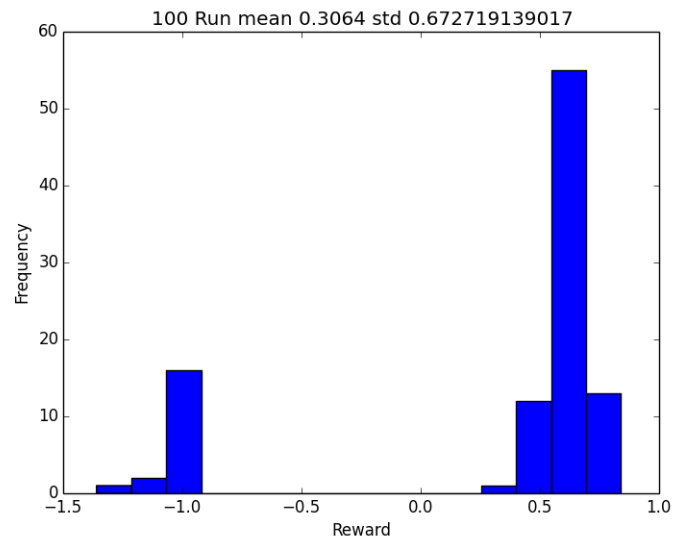
1000-run standard dev: 0.601363

5, 3 histograms generated from Part 2.

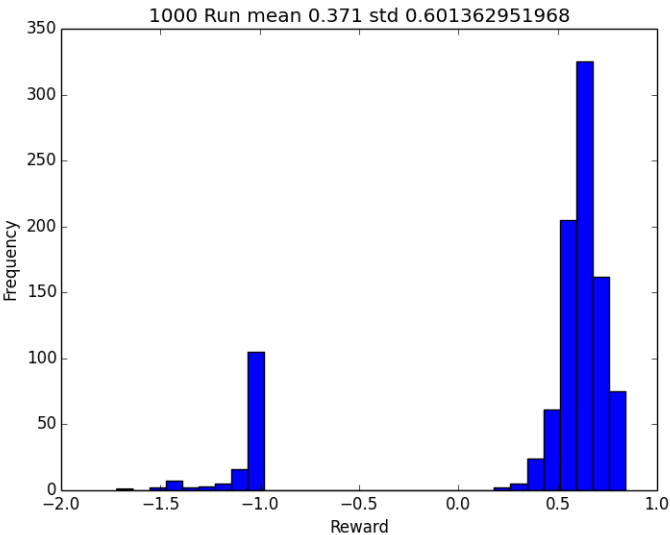
P2-histogram-10.png



P2-histogram-100.png



P2-histogram-1000.png



6, Full output from P3-output.txt. (thresholds and policies)

Lower bound: (approximate 0) 0.001

(there are only 5 threshold here, the lower and up bound are not threshold)

Optimal Policy:

^ > +10

^ ^ ^

^ ^ ^

Utilities:

3.00260312	-0.99219932	10.00000000
-0.99779780	-1.00099276	-0.99219932
-1.00099844	-1.00100099	-1.00099396

Threshold value: 0.73306751

Optimal Policy:

^ > +10

^ ^ ^

^ ^ ^

Utilities:

9.99984816	5.48852551	10.00000000
5.48842373	3.02345223	5.48852551
2.48418460	1.13733544	2.48425475

Threshold value: 0.85973232

Optimal Policy:

^ < +10

^ ^ ^

^ ^ ^

Utilities:

18.50436039	13.80024525	10.00000000
13.80023700	10.31433752	7.40089422
9.98609148	7.39713097	5.17072931

Threshold value: 0.85984427

Optimal Policy:

^ < +10

^ ^ <

^ ^ ^

Utilities:

18.51865257	13.81434146	10.00000000
13.81433424	10.32742835	7.40913575
9.99939091	7.40907797	5.17893087

Threshold value: 0.87706932

Optimal Policy:

^ < +10

^ < <

^ ^ <

Utilities:

21.02958576	16.30007819	10.00000000
16.30007014	12.71194311	9.42996398
12.36697538	9.63756069	7.22280950

Threshold value: 0.97101281

Optimal Policy:

^ < +10

^ < <

^ < <

Utilities:

87.14797699	82.26837856	10.00000000
82.26836915	78.08932276	67.41748078
77.63988124	74.08807335	69.88457848

Upper bound: (approximate 1) 0.99900000

Optimal Policy:

$\hat{v} < +10$

$\hat{v} < v$

$\hat{v} < <$

Utilities:

2507.58085371	2502.65183301	10.00000000
2502.65182303	2498.28125725	2486.36604412
2497.79616726	2493.96399565	2489.23896360

