DL4-3 ベルマン方程式

3

確率:

条件付き確率:

報酬:

期待値:

報酬の期待値

4

報酬:

収益:

状態価値関数:

5-6

方策・環境の状態遷移ともに確率的である場合の収益を考える。

より、

従って、とを求めればよい。

8

決定論的に行動が決定されるため、

9

に値を代入し、

整理すると、

11

12

14-15

ベルマン方程式

最適方策をとすると、ベルマン方程式は

となる。(ただし、を省略してと表現している)

ここで、方策は最適なので、

16

例えば

のとき、

17

Q関数のベルマン方程式

最適方策をとすると、同様に

となる。(ただし、を省略してと表現している)

ここで、方策は最適なので、

18

決定論的に行動が決定されるため、

各値を代入すると、

19

例えば

のとき、

20

最適方策は

L1にいるとき、

Left :

Right:

L2にいるとき、

Left :

Right: