RL – שאלה 2

*שאלה 2*

1. *נגדיר MDP לבעיה:*
2. *נבחין שעל מנת למקסם את הסיכוי לסיים ב-0, אנו רוצים להיות במצבים הקרובים ל-0 עד כמה שאפשר. לכן, המדיניות האופטימלית תהיה כזו ה"שואפת" להתקרב ל-0 עד כמה שאפשר, ולהישאר סביב 0 כאשר נמצאים במצבים הקרובים ל-0.*

*נוכל להגדיר באופן הבא:*

*עבור מדיניות זו, כל עוד אנו רחוקים ממצב 0, נתקדם לכיוון 0 במסלול הקצר ביותר, ואם אנו נמצאים במצב 0, נתקדם למצב 2k-1, ובצעד הבא נחזור למצב 0.*

1. *נבצע איטרציה אחת של :*

*מכיוון שלכל מתקיים לכל , ועבור מתקיים , נקבל:*

1. *נבצע איטרציה נוספת:*

*עבור , מתקיים או לכל (הם לא סמוכים למצב 0, לכן ההסתברות לעבור למצב עבורו היא 0).*

*נתבונן במצבים שנותרו:*

*סה"כ:*

1. *נחשב את לכל :*

*חישבנו את לכל s, נמשיך בחישוב (אבחנה, משיקולי סימטריה, , לכן אחשב רק את ):*

*ניתן לראות דפוס, ולחשב במקרה הכללי:*

*כאשר עבור זוגי.*

*אזי:*

*ולכן בגבול לאינסוף:*

*זהו טור הנדסי ו- לכן*

*באופן דומה, נקבל:*

*עבור הנוסחה הכללית היא:*