# TRICK OR TREAT PROBLEM

AKA "LILY MEETS THE CAT"

140 CONDOS, EACH WITH PORCH LIGHT

6pm - 7 pm

NOISY INDICATION

UNCERTAINTY:

- SOMEONE AT HOME?
- IF HOME, TIME TO ANSWER?

ANSWER ≡ PROVIDE CANDY ⇒ REWARD += 1

ACTIONS:

- RING: ONLY ONCE PER HOME [RB]
- WAIT [WH]
- NEXT [NH]

TIME: DISCRETIZED (10 sec)

PABLO RODRIGUEZ BERTORELLO

**From:** Pablo Rodriguez Bertorello <pablo@1now.ai>
**Sent:** Monday, December 4, 2017 8:56 PM
**To:** usa5105@fedex.com
**Subject:** ANOTHER

```
#########################################################
############ TRICK OR TREAT? LILY MEETS CAT
#########################################################

############
# QUESTION 4
############

### Given facts
reward_if_answer = 1
probability_being_home = 0.85 # light is on

### Theta
# defined in Q1, maximum likelihood calculation Q2
theta = 0.29

### Probability of answer
# Function of Theta, and iteration number
function probability_of_answer(t)
   if t==0
      return theta
   end
   return (1-theta)^t * theta
end

### Discounted utility
function discounted_utility(utility_table, t, discount_factor)
   if t==1
      return 0    # before first time increment, utility is 0
   end
   return discount_factor * utility_table[t-1]
end

### POLICY ITERATION
function policy_utility(discount_factor, number_of_increments)
   utility_table = fill(0., number_of_increments)    # dynamic programming; and track it over time
   for t = 1:number_of_increments
      instant_reward = probability_being_home * probability_of_answer(t) * reward_if_answer
      println("t=", t, " instant_reward: ", instant_reward)
      utility_table[t] = instant_reward + discounted_utility(utility_table, t, discount_factor)
   end
   return utility_table
end

############ PART 4a: policy to wait for eternity at house with light (discounted utility)
### Facts
discount_factor = 0.9          # given
number_of_increments = 100     # Run for eternity

## utility over infinite time converges to zero, due to discounting
utility = policy_utility(discount_factor, number_of_increments)

println("UTILITY OVER TIME: ", utility)

## instant reward over time, with decreasing probability of answer
#t=1 instant_reward: 0.20589999999999997
#t=2 instant_reward: 0.14618899999999999
#t=3 instant_reward: 0.10379418999999997
```

PABLO RODRIGUEZ BERTORELLO

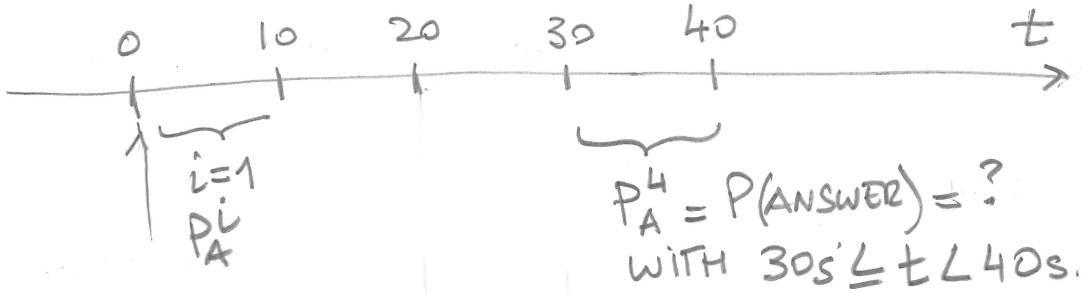
```
#t=4 instant_reward: 0.07369387489999998
#t=5 instant_reward: 0.052322651178999986
#t=6 instant_reward: 0.03714908233708998
#t=7 instant_reward: 0.02637584845933389
#t=8 instant_reward: 0.01872685240612706
#t=9 instant_reward: 0.013296065208350211
#t=10 instant_reward: 0.00944020629792865

# UTILITY OVER TIME (converges to zero): [0.2059, 0.331499, 0.402143, 0.435623, 0.444383, 0.437094, 0.41976, 0.396511, 0.370156, 0.342581,
0.315025,
#0.288282, 0.262832, 0.238948, 0.216756, 0.19629, 0.17752, 0.160377, 0.144772, 0.130602, 0.11776, 0.106139, 0.0956353, 0.0861498,
#0.0775903, 0.0698706, 0.0629115, 0.0566402, 0.0509903, 0.0459012, 0.0413182, 0.0371914, 0.0334759, 0.0301308, 0.0271196, 0.0244089,
#0.0219689, 0.0197727, 0.0177959, 0.0160166, 0.0144152, 0.0129738, 0.0116765, 0.010509, 0.00945814, 0.00851236, 0.00766116, 0.00689506,
#0.00620557, 0.00558503, 0.00502653, 0.00452388, 0.0040715, 0.00366435, 0.00329792, 0.00296813, 0.00267132, 0.00240418, 0.00216377,
#0.00194739, 0.00175265, 0.00157739, 0.00141965, 0.00127768, 0.00114992, 0.00103492, 0.000931431, 0.000838288, 0.000754459, 0.000679013,
#0.000611112, 0.000550001, 0.000495001, 0.000445501, 0.000400951, 0.000360856, 0.00032477, 0.000292293, 0.000263064, 0.000236757,
#0.000213082, 0.000191773, 0.000172596, 0.000155337, 0.000139803, 0.000125823, 0.00011324, 0.000101916, 9.17247e-5, 8.25522e-5,
#7.4297e-5, 6.68673e-5, 6.01805e-5, 5.41625e-5, 4.87462e-5, 4.38716e-5, 3.94845e-5, 3.5536e-5, 3.19824e-5, 2.87842e-5]

############ PART 4b: policy to wait for eternity at house with light (no discounting)
### Facts
discount_factor = 1.0          # given
number_of_increments = 100     # Run for eternity

## utility over infinite time converges to zero, due to discounting
utility = policy_utility(discount_factor, number_of_increments)

println("UTILITY OVER TIME: ", utility)
# UTILITY OVER TIME (converges to >0): [0.2059, 0.352089, 0.455883, 0.529577, 0.5819, 0.619049, 0.645425, 0.664151, 0.677448,
#0.686888, 0.69359, 0.698349, 0.701728, 0.704127, 0.70583, 0.707039, 0.707898, 0.708508, 0.70894, 0.709248, 0.709466, 0.709621,
#0.709731, 0.709809, 0.709864, 0.709904, 0.709932, 0.709951, 0.709966, 0.709976, 0.709983, 0.709988, 0.709991, 0.709994, 0.709996,
#0.709997, 0.709998, 0.709998, 0.709999, 0.709999, 0.709999, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71,
#0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71,
#0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71, 0.71,
#0.71, 0.71, 0.71]


############ PART 4c: policy to wait for 1 minute at house with light (no discounting)
### Facts
discount_factor = 1.0          # given
number_of_increments = 6       # Run for eternity

## utility over infinite time converges to zero, due to discounting
utility = policy_utility(discount_factor, number_of_increments)

println("UTILITY OVER TIME: ", utility)
# UTILITY OVER TIME (converges to >0 about same as infinite wait): [0.2059, 0.352089, 0.455883, 0.529577, 0.5819, 0.619049]


############
# QUESTION 5
############

############ Part 5a: policy to wait up to eternity at set of houses (no discounting)
### Facts
discount_factor = 1.0          # given
theta = 0.5                    # given

number_of_increments = 18      # Run for eternity; if get candy go to next house

## utility over infinite time converges to zero, due to discounting
utility = policy_utility(discount_factor, number_of_increments)

println("UTILITY OVER TIME: ", utility)
# UTILITY OVER TIME (converges to >0 about same as infinite wait): [0.2059, 0.352089, 0.455883, 0.529577, 0.5819, 0.619049]
```

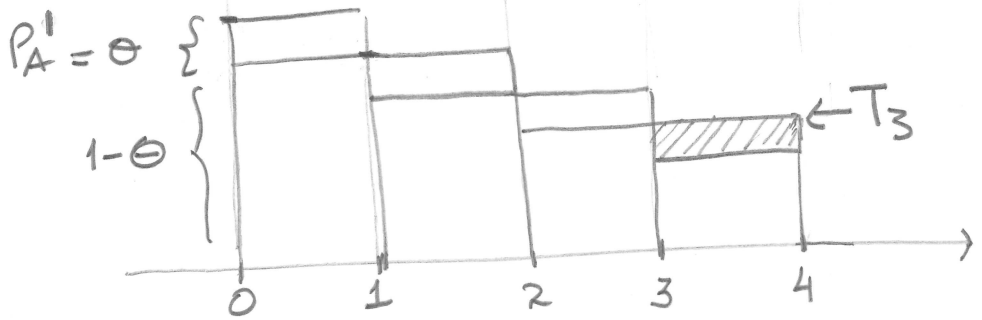
PABLO RODRIGUEZ BERTORELLO 3/19

(1)

$\theta = P(\text{OCUPANT ANSWER})$ WITHIN TIME INTERVAL (10s)

$\forall$ INTERVAL (IF NOT OPENED ALREADY)




$T_0 = 1$ ASSUME OCCUPANT KNOWN TO BE HOME

$T_1 = T_0 \cdot (1-\theta)$

$T_n = T_{n-1}(1-\theta) = (1-\theta)^n$

$P_A^n = \frac{T_{n-1}}{T_0} \cdot \theta = \boxed{(1-\theta)^{n-1} \cdot \theta = P_A^n}$ FOR $n > 1$

$\boxed{P_A^4 = (1-\theta)^3 \cdot \theta}$

(2) TIME TO ANSWER (SECONDS) AT 8 HOMES:

13, 17, 5, 41, 11, 89, 22, 34

| INTERVAL | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| COUNT$_i$ | 1 | 3 | 1 | 1 | 1 | 0 | 0 | 0 | 1 |
| OUT OF | 8 | 7 | 4 | 3 | 2 | 1 | 1 | 1 | 1 |
| $\theta_i$ | $\frac{1}{8}$ | $\frac{3}{7}$ | $\frac{1}{4}$ | $\frac{1}{3}$ | $\frac{1}{2}$ | 0 | 0 | 0 | 1 |

$\theta_i \equiv$ P(ANSWER) WITHIN INTERVAL

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\, P(D|\theta) = \frac{1}{n}\sum_{i=1}^{n}\theta_i$$

$$\hat{\theta} = \frac{1}{9}\cdot\left(\frac{1}{8}+\frac{3}{7}+\frac{1}{4}+\frac{1}{3}+\frac{1}{2}+1\right)$$

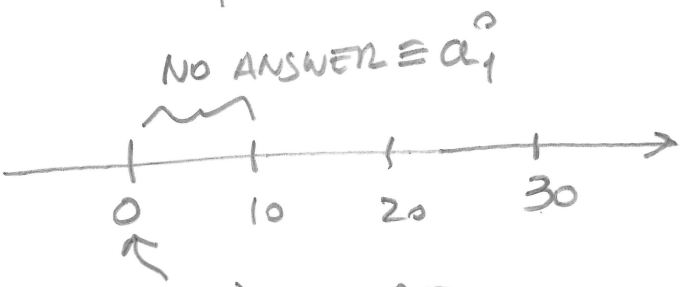
$$\hat{\theta} = \frac{443}{1512} = 0.29$$

PABLO RODRIGUEZ BERTORELLO

PRIOR PROBABILITY

(3')

| LIGHT | P(H\|L) |
|---|---|
| ON | 0.85 |
| OFF | 0.05 |



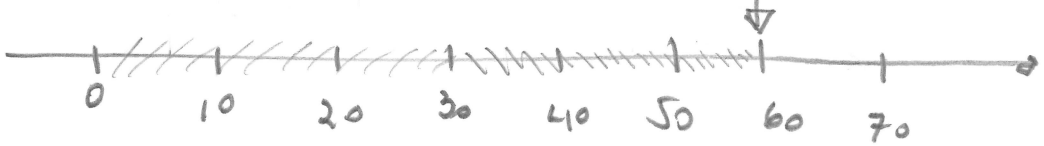

LIGHT [ON] $\Rightarrow P(H) = 0.85$

$$P(H \mid a_1^\circ a_2^\circ a_3^\circ) = P(H) = 0.85$$

CONDITIONAL INDEPENDENCE

(3'')

$$P(\text{ANSWER} \mid a_4^\circ, a_5^\circ, a_6^\circ) = P_A^7 \cdot \underbrace{P(H \mid L)}_{0.85 \text{ LIGHT=ON}}$$




$$\boxed{P = 0.85 \cdot (1-\theta)^6 \theta}$$

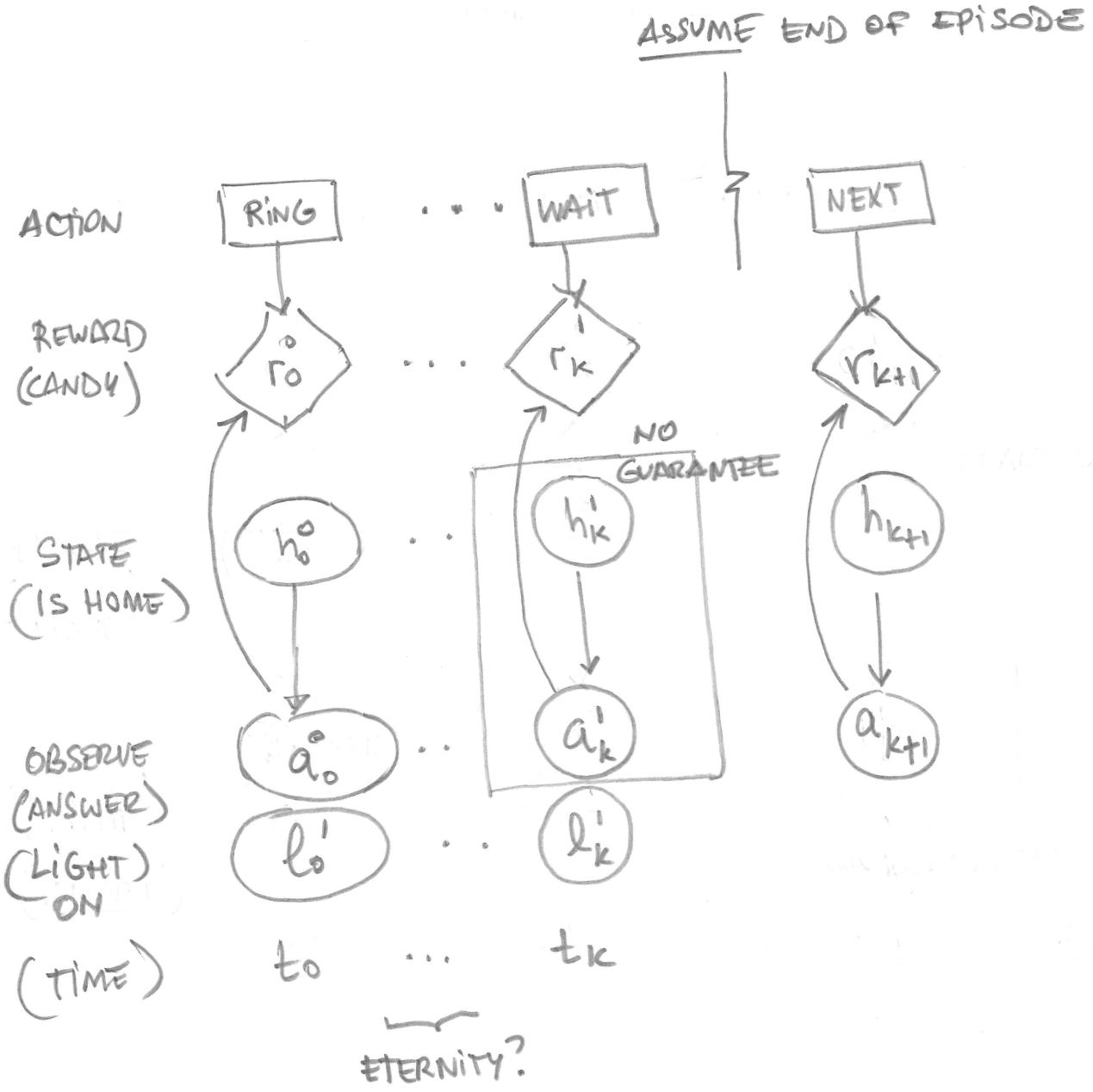
$P_A^7$ PER Q1 ANSWER

④

DOOR W/ LIGHT ON

WILLING TO WAIT ETERNITY (NO GUARANTEE)

—EXACT ANSWER IN Θ—



(41) EU = ? $\gamma = 0.9$ LIGHT = ON

$$U_t^{\pi}(s) = R\left[s, \pi(s)\right] + \gamma \sum_{s'} T\left[s' \mid s, \pi(s)\right] U_{t-1}^{\pi}(s'), \qquad 1 \equiv (s = s')$$

$R\left[s, \pi(s)\right]$: A = RING/WAIT

$T\left[s' \mid s, \pi(s)\right]$: WAIT FOR ETERNITY

$$R_t\left[s, \pi(s)\right] = r(t) \cdot P_A^t \cdot P(H \mid L), \qquad P(H \mid L) = 0.85 \text{ (LIGHT)}$$

$r(t) = 1$ GIVEN

$P_A^t$: PROBABILITY OF OCCUPANT ANSWER IN INTERVAL $t$ (Q1)

(4a) WAIT ETERNITY OK

$$\boxed{U_t^{\pi}(s) = 0.85 P_A^t + \gamma\, U_{t-1}^{\pi}(s)}$$

$\gamma$: GIVEN — $U_{t-1}^{\pi}(s)$: $U_0^{\pi}(s) = 0$

$$U_1^{\pi}(s) = 0.85 \cdot \underbrace{\theta}_{P_A^1}$$

$$U_{2(s)}^{\pi} = 0.85 \underbrace{(1-\theta)}_{P_A^2} \cdot \theta + 0.9\, \underbrace{\theta}_{U_1^{\pi}(s)}$$

$$\boxed{U_{\infty}^{\pi}(s) = 0} \quad \begin{cases} \lim_{t \to \infty} P_A^t = 0 \\ \lim_{t \to \infty} \gamma^t = 0 \end{cases}$$

[CODE ATTACHED WITH VALUES]

PABLO RODRIGUEZ BERTORELLO

(4") NO DISCOUNTING

(b) ETERNITY OK NO DISCOUNT

$$U_t^{\pi}(s) = 0.85\, P_A^t + U_{t-1}^{\pi}(s)$$ FROM PART (a)

$\uparrow$

$\gamma = 1$

$$\boxed{U_t^{\pi}(s) = 0.85\,(1-\theta)^{t-1}\,\theta + U_{t-1}^{\pi}(s)} \quad t>1$$

FROM Q1

$$\boxed{U_{\infty}^{\pi}(s) = 0.60}$$ $>0$ CONVERGENCE W/O DISCOUNTING

$\uparrow$ CODE / VALUES ATTACHED

(c) NO DISCOUNTING
MAX WAIT ONE MINUTE

$$\boxed{U_6^{\pi}(s) = 0.85.(1-\theta)^5\,\theta + U_5^{\pi}(s)}$$

$\uparrow$ Q4 b

$$\boxed{U_6^{\pi}(s) = 0.52}$$ NEARLY AS MUCH AS 4b WITHOUT WAITING ETERNITY

$\uparrow$ CODE / VALUES ATTACHED.

PABLO RODRIGUEZ BERTORELLO

(5) IT's 6:57 pm (3 MINUTES REMAIN)

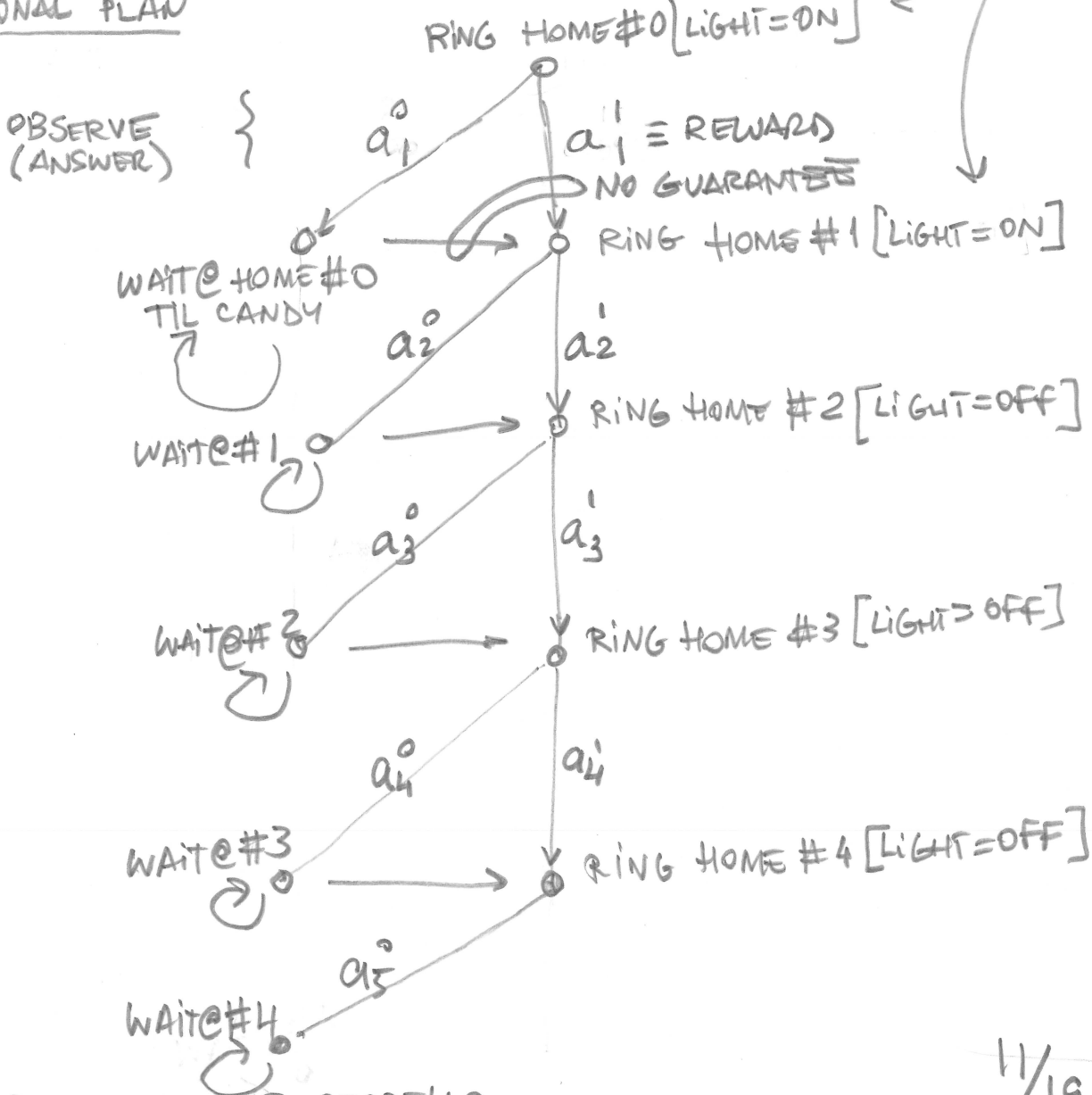
5 HOMES REMAIN
- 2 LIGHTS ON
- 3 LIGHT OFF

⟹ $\pi \equiv$ WAIT TIL GET CANDY. $\exists U$?

$\Theta = 0.5$ $\gamma = 1$ SMART ORDER $\equiv$ LIGHT ON FIRST

— EXACT #, NOT APPROXIMATION —

CONDITIONAL PLAN




PABLO RODRIGUEZ BERTORELLO

11/19

(5') $\gamma = 1 \quad \theta = 0.5$ WAIT TIL GET CANDY

$$\vec{b(s)} = \left[P(h^0), P(h^1)\right]$$

$$b(H \mid \text{LIGHT=ON}) = [0.15, 0.85]$$

$$b(H \mid \text{LIGHT=Off}) = [0.95, 0.05]$$

$P(h^0)$ GIVEN $P(h^1)$

$$R_t^{\pi}[s, \text{WAIT}] = 1 \cdot P_A^t$$

CANDY — ANSWER IN INTERVAL $t$

$$\alpha_{\text{WAIT}}^{(t)} = \left[0, (1-\theta)^{t-1}\cdot\theta\right]$$

$$\alpha_{\substack{\text{RING}\\\text{NEXT}}}^{(t=1)} = [0, \theta]$$

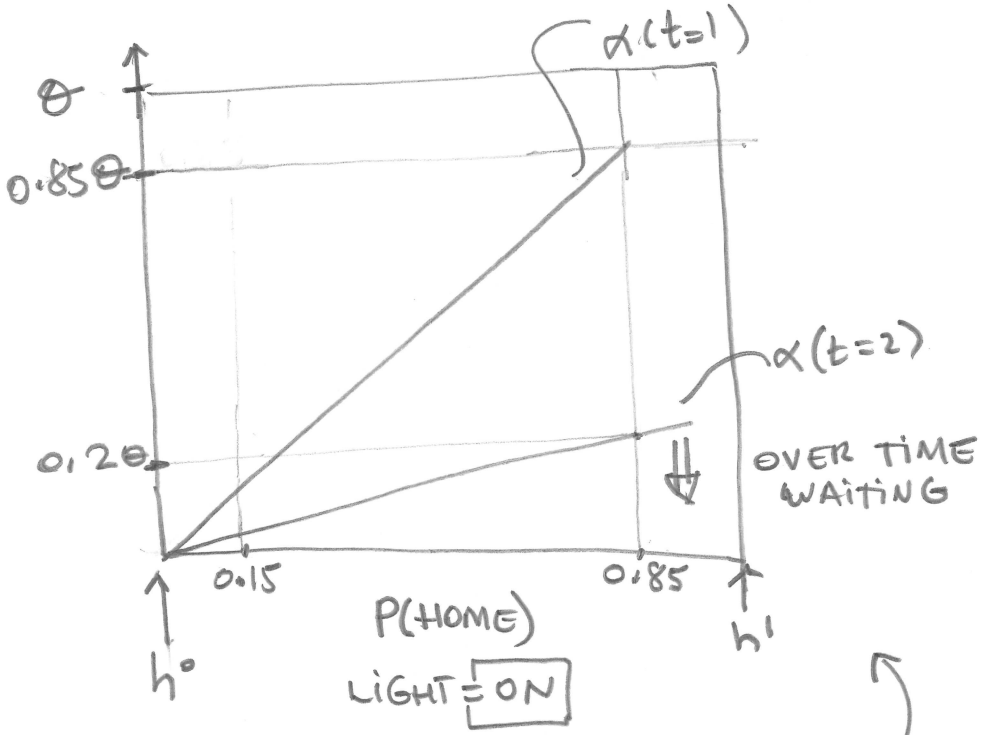
$h^0 \quad h^1$

REWARDS FOR ACTIONS

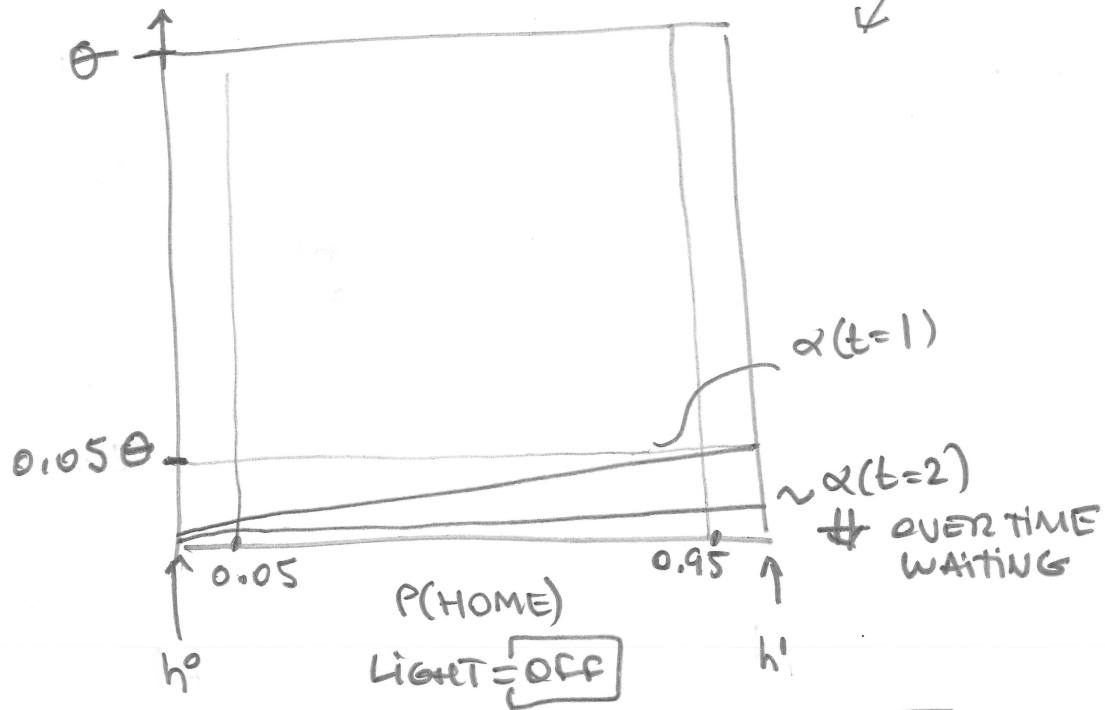
PABLO RODRIGUEZ BERTORELLO

12/19

5ª




U(t=1) OVERWHELMS U(t>1)




U(LIGHT=ON) OVERWHELMS U(LIGHT=OFF)

(5th) POLICY: WAIT TIL GET CANDY [MAX 3 MINUTES]

SMART: GO TO HOUSES WITH LIGHT ON FIRST

$\theta = 0.5 \quad \gamma = 1$

$P(H \mid \text{LIGHT ON})$

$$U^{\pi}_{t(s)} = 0.85(1-\theta)^{t-1}\theta + U_{t-1}(s)$$

for $t > 1$

$U_0 = 0$

$U_1 = 0.85\theta$

SEE CODE AND VALUES

$U_{18}(s) \geq 0,42$

UTILITY MAY BE GREATER IF CAN GO TO OTHER HOMES (NO GUARANTEE)

3 MINUTES, 10 SEC INTERVALS

PABLO RODRIGUEZ BERTORELLO

(6) POLICY: WAIT NO MORE THAN 30 SEC AT A HOME

↓

GETS TO RING ALL 5 HOMES IN REMAINING 3'

$$U^{\pi}(b) = \sum_{s} \vec{\alpha}_p^{\,T} \cdot \vec{b}$$

$$\approx [0,\theta]^T [0.15, 0.85] + \text{HOUSE \#0 (LIGHT=ON)}$$

$$[0,\theta]^T [0.15, 0.85] + \text{HOUSE \#1 (LIGHT=ON)}$$

$$[0,\theta]^T [0.95, 0.05] + \text{HOUSE \#3 (LIGHT=OFF)}$$

$$[0,\theta]^T [0.95, 0.05] + \text{HOUSE \#4 (LIGHT=OFF)}$$

$$[0,\theta]^T [0.95, 0.05] + \text{HOUSE \#5 (LIGHT=OFF)}$$

$$x \cdot [0, (1-\theta)^t \cdot \theta] [0.15, 0.85] + \text{WAITS AT LIGHT=ON}$$

$$y \cdot [0, (1-\theta)^t \theta] [0.95, 0.05] \quad \text{WAITS AT LIGHT=OFF}$$

APPROXIMATE = 0 AS $\theta \gg (1-\theta)^t \theta$

$$U^{\pi}(b) \approx 0.85\,\theta \times 3 + 0.05 \cdot \theta \times 2$$

$$\boxed{U^{\pi}(b) \approx 1.85\,\theta = 0.93}$$

## (7) POMDP MODEL

12 STATE VARIABLES:

[h] - IS SOMEONE AT A PARTICULAR HOME: 5 BINARY

[r] - RUNG DOOR AT A PARTICULAR HOME: 5 BINARY

[l] - CURRENT LOCATION

[t] - TIMESTEPS REMAINING

ASSUMPTION: NO KNOWLEDGE TRANSFER ABOUT PREVIOUS CONDO VISITS

5 ACTIONS:

- WAIT: =
- NEXT DOOR: ←↑→↓

REWARD:

- VISIT SAME HOME AS BEFORE: $-\infty$

FAST INFORMED BOUND:

⇒ HOW MANY $\alpha$ VECTORS: 5 (ONE PER ACTION)

- DIMENSIONALITY OF $\alpha$ VECTORS: # OF STATES

$STATE = [h, r, l, t]^T \Rightarrow |S| = |h| \times |r| \times |l| \times |t|$

@SCENARIO OF Q5.

$|h| = 5 \times 2$ GIVEN

$|r| = 5 \times 2$ GIVEN

$|l| = 5$ (HOMES LEFT)

$|t| = 18$ (3 MINUTES LEFT, 10 SEC INCREMENTS)

$|S| = 10 \times 10 \times 5 \times 18$

$|S| = 9000$

16/19

PABLO RODRIGUEZ BERTORELLO

(8)

ENUMERATED 18-STEP CONDITIONAL PLANS

$\alpha_p$ ALPHA VECTOR (FROM COMPUTING VALUE) FOR EACH POSSIBLE INITIAL PLAN STATE

RELATIONSHIP?

$$\max_p \vec{\alpha}_p^T b \leq \max_{\alpha \in \Gamma_{FIB}} \vec{\alpha}^T b$$

$U^*(\vec{b})$

EXACT OPTIMAL VALUE FUNCTION

(P-SPACE COMPLETE COMPLEXITY)

$U^{\Gamma}(b)$

APPROXIMATE FAST INFORMED BOUND VALUE FUNCTION

$O(|A|^2 |S|^2 |O|)$

(9)

WITH FAST INFORMATION BOUND

$$O(|A|^2|S|^2|O|)$$ AT EACH ITERATION

TEMPTING AS QMPD WOULD BE

$$O(|A|^2|S|^2)$$ AT EACH ITERATION

ANNA WOULD NOT HAVE FULL OBSERVABILITY AT EVERY NEXT STEP

⇒ SARSOP WOULD BE MY RECOMMENDATION

BECAUSE SHE WOULD ONLY HAVE SOLVE FOR THE FEW BELIEF STATES REACHABLE BY THE OPTIMAL POLICY IN THE 3 MINUTES REMAINING.

**From:** Pablo Rodriguez Bertorello <pablo@1now.ai>
**Sent:** Monday, December 4, 2017 8:56 PM
**To:** usa5105@fedex.com
**Subject:** FIRST

```
##################################################
########## TRICK OR TREAT? CAT EATS LILY
##################################################
########## THE HUNGRY CAT NEAR DOOR WAITS
########## WITH TEETH GUARDING HOLY GATES
########## ADORNED IN MASTER'S LOVE SHE SMILES
########## BEEN WAITING TO EAT LILY FOR MILES
##################################################
```

(10) RHYME

```
### REFERENCES
# POMDPs.jl API http://juliapomdp.github.io/POMDPs.jl/latest/api/#POMDPs.state_index
# Tutorial Tiger POMDP http://nbviewer.jupyter.org/github/sisl/POMDPs.jl/blob/master/examples/Tiger.ipynb
# Explicit POMDP http://juliapomdp.github.io/POMDPs.jl/latest/explicit/
# Tiger problem PDF https://www.cs.rutgers.edu/%7Emlittman/papers/aij98-pomdp.pdf
# Tiger problem PPT https://www.techfak.uni-bielefeld.de/~skopp/Lehre/STdKI_SS10/POMDP_tutorial.pdf
# POMDP Simulator https://github.com/JuliaPOMDP/POMDPToolbox.jl

### POMDPS.jl
# The following may be performed prior and separately by own JuliaCommand.jl
importall POMDPs
POMDPs.add("QMDP")              # Design Under Uncertainty 6.4.1
POMDPs.add("SARSOP")
using QMDP
using SARSOP
using POMDPToolbox
using POMDPModels

# TODO: remove
# Pkg.clone("https://github.com/sisl/POMDPs.jl.git")
# Pkg.add("POMDPs")
# using POMDPs
# Pkg.add("SARSOP")
# POMDPs.add("QMDP")              # Design Under Uncertainty 6.4.1
# POMDPs.add("POMDPToolbox")      # implements discrete belief updating
# Pkg.add("Distributions")
# importall POMDPs
# using SARSOP
# using POMDPToolbox


### PROBLEM
# In the tiger POMDP, the agent is tasked with escaping from a room. There are two doors leading out of the room.
# Behind one of the doors is a tiger, and behind the other is sweet, sweet freedom.
# If the agent opens the door and finds the tiger, it gets eaten (and receives a reward of -100).
# If the agent opens the other door, it escapes and receives a reward of 10.
# The agent can also listen. Listening gives a noisy measurement of which door the tiger is hiding behind.
# Listening gives the agent the correct location of the tiger 85% of the time.
# The agent receives a reward of -1 for listening.


### PARTIALLY OBSERVABLE MARKOV DECISION PROBLEM
# definition:
# S: State
# A: Action
# Omega: Observation space
# O: Observation function
# T: Transition
# R: Reward
```

PABLO RODRIGUEZ BERTORELLO



19/19