XOR Blackmail

Lecture By Natalie Melas-Kyriazi

Thursday March 30, 2023

Problem:

There is a rumor that your house has termites, and you hate termites! Getting rid of them would cost −100.

But! There is also an agent Ω who knows whether you have termites or not. They send you a letter in the mail if & ONLY if

1. you have termites and will not pay the blackmail

OR

2. you do NOT have termites and will pay the blackmail (−10)

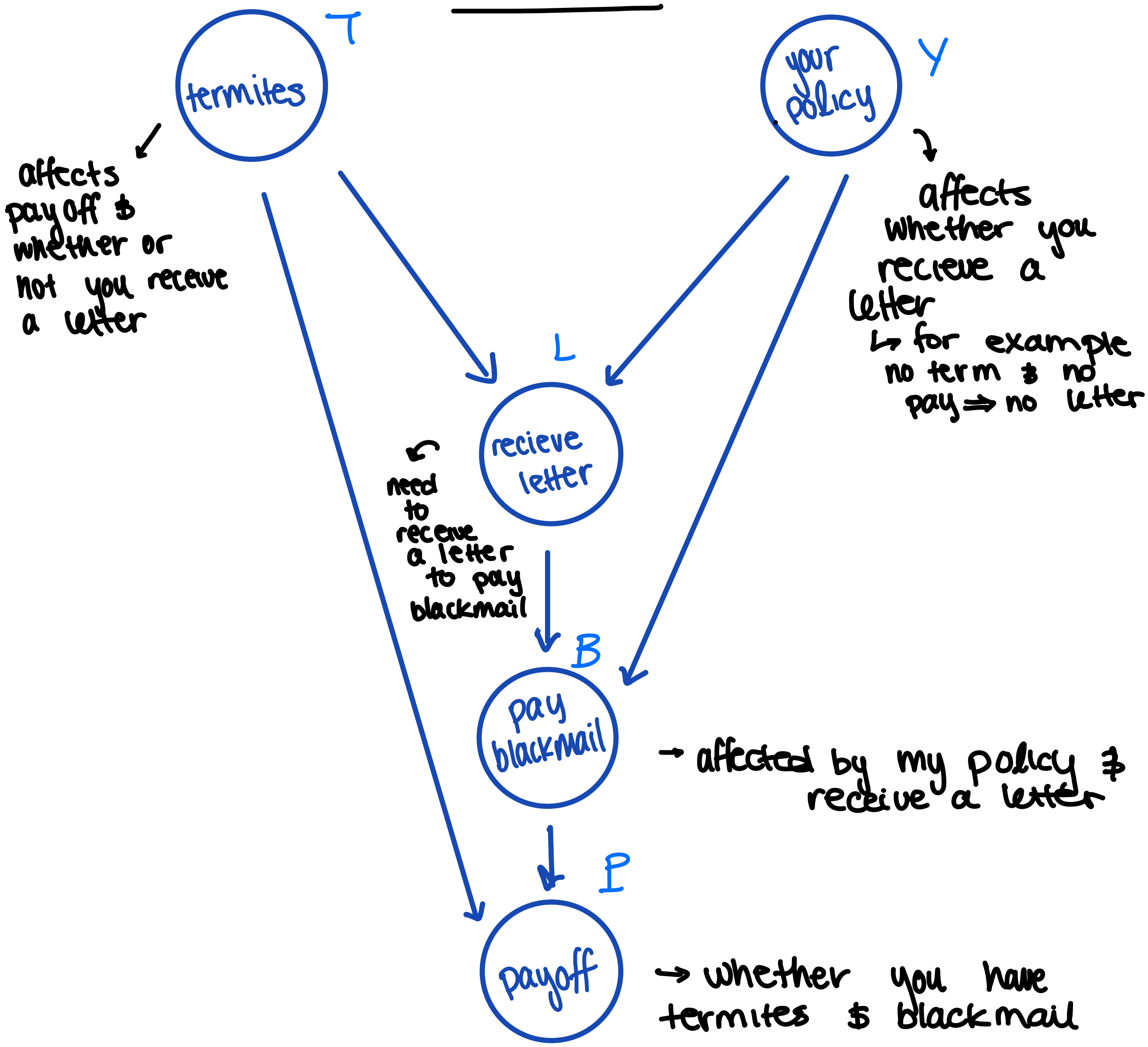
(hence the "XOR"). What do you do?

PART A

Assume:

probability of having termites is 10% and your policy is that you flip a fair coin (so a 50/50 chance) on whether you pay

BAYES NET:



Probability tables:

T

Yes termites	No termites
0.1	0.9

Y

Pay	Do not pay
0.5	0.5

L

T	Y	Receieve letter	Do not receive letter
Yes	Pay	0	1
Yes	No pay	1	0
No	pay	1	0
No	No pay	0	1

"XOR" <

B

L	Y	Pay blackmail	Do not pay blackmail
Yes	Pay	1	0
Yes	No pay	0	1
No	pay	0	1
No	No pay	0	1

ⓟ B	Y	-110	-100	-10	0
Yes	Pay	1	0	0	0
Yes	No pay	0	1	0	0
No	pay	0	0	1	0
No	No pay	0	0	0	1

PART B

Now we want to calculate

$$\mathbb{E}[P \mid y = \text{pay}] \quad \text{and} \quad \mathbb{E}[P \mid y = \text{don't pay}]$$

↳ expected payoffs given that you pay or don't pay (note: you don't know if you've received a letter)

By previous class work we know

$$\mathbb{E}[P \mid y] = \underbrace{\sum_p p \Pr(p \mid y)}_{\text{weighted average of payoffs}}$$

so this is sum over
$p \in \{0, -10, -100, -110\}$

we know:

$$\Pr(p|y) = \sum_{t,l,b} \Pr(t, l, b, p \mid y)$$

$$= \sum_{t,l,b} \Pr(t|y)\Pr(l|t,y)\Pr(b|t,l,y)\Pr(p|b,t,l,y)$$

$$= \sum_{t,l,b} \Pr(t)\Pr(l|t,y)\Pr(b|l,y)\Pr(p|t,b)$$

can look these up in our table!

We have 8 possible combinations of t, l, b because they're binary. If you do it all out you will see the only non-zero terms in the sum conditioning on y = pay are

① t = yes, l = no, b = no
↳ payoff: −100, prob: 0.1

② t = no, l = yes, b = yes
↳ payoff: −10, prob: 0.9

so $\mathbb{E}[P \mid y=\text{pay}] = \sum_p p \Pr(p|y)$

$$= 0.1(-100) + 0.9(-10)$$

$$= \boxed{-19}$$

Now for $\mathbb{E}[p \mid y = \text{don't pay}]$ we have different values of $t, b, \& l$ which produce non-zero probabilities those are

① $t = \text{yes}$, $l = \text{yes}$, $b = \text{no}$

↳ prob: 0.1, payoff: -100

② $t = \text{no}$, $l = \text{no}$, $b = \text{no}$

↳ prob: 0.9, payoff: 0

so $\mathbb{E}[p \mid y = \text{don't pay}] = \sum_p p \Pr(p \mid y)$

$$= 0.9(0) + 0.1(-100)$$

$$= \boxed{-10}$$

<u>CONCLUSION</u>:

$$\mathbb{E}[p \mid y = \text{don't pay}] > \mathbb{E}[p \mid y = \text{pay}]$$

PART C

Now! We condition <u>also</u> on the fact that you recieved a letter so we want to find:

i) $\mathbb{E}[p \mid y = \text{pay},\ \ell = \text{yes}]$

ii) $\mathbb{E}[p \mid y = \text{don't pay},\ \ell = \text{yes}]$

like earlier, we use

$$\mathbb{E}[p \mid y = \text{don't pay},\ \ell = \text{yes}] = \sum_p p \Pr(p \mid y, \ell)$$

but this time,

$$\Pr(p \mid y, \ell) = \sum_{t,b} \Pr(t \mid \ell, y) \Pr(b \mid t, \ell, y) \Pr(p \mid t, \ell, b, y)$$

$$= \sum_{t,b} \Pr(t \mid \ell, y) \Pr(b \mid \ell, y) \Pr(p \mid t, b)$$

these we can look up in our probability tables

we don't know this one!

$\Rightarrow$ Baye's Rule:

$$\Pr(t \mid \ell, y) = \frac{\Pr(\ell \mid t, y) \Pr(t \mid y)}{\Pr(\ell \mid y)}$$

$$= \frac{\Pr(\ell \mid t, y) \Pr(t)}{\Pr(\ell \mid y)}$$

Okay now let's look at (i) which is $\mathbb{E}[p \mid y = pay, l = yes]$. What do we think this should be from our understanding of the problem?

↳ -10 because if we recieve a letter & pay then we must have paid the blackmail & must not have termites!

following this logic, we can do out the math and see that $t = no$ & $b = yes$ is the only combination that gives you a non-zero term

↳ $Pr(t \mid l, y) = 1$ (by Bayes)

$Pr(b \mid l, y) = 1$

$Pr(p \mid t, b) = -10$

$$\mathbb{E}[p \mid l = yes, y = pay] = 1 \cdot -10 = \boxed{-10}$$

Now we consider the same question when our policy is not to pay.

↳ If we think about it, if we get a letter and don't pay, then we must have termites!

We can check this with math & we see the only non-zero term in the $\sum_p p \Pr(t \mid l, y)$ is when

t = yes, b = no

↳ $\Pr(t \mid l, y) = 1$

$\Pr(b \mid l, y) = 1$

$\Pr(p \mid t, b) = -100$

$$E[p \mid l = \text{yes}, y = \text{not pay}] = 1 \cdot -100$$
$$= \boxed{-100}$$

so we see when we condition on receiving the letter, it is in your interest to pay the blackmail, aka

$$E[p \mid l = \text{yes}, y = \text{pay}] > E[p \mid l = \text{yes}, y = \text{not pay}]$$