Conditional, Marginal, Joint probability, Bayes: "roaring tiger" case study

Presentation in video: http://personales.upv.es/asala/YT/V/tiger3EN.html

The code executed in Matlab **R2023b**

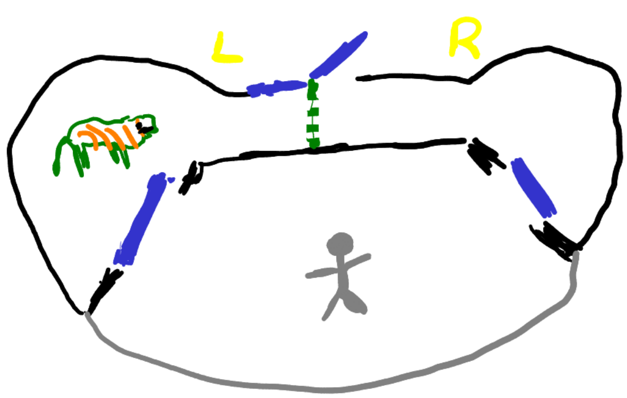
Objectives: Illustrate some issues on conditional, joint, marginal probabilities and Bayes formulae in the case study "Tiger behind two doors" where noisy measurements about tiger location need to be accumulated to improve chances of survival (correctly assessing where the tiger is before "acting" opening a door).

Deciding the action to take will NOT be considered in this material, just dealing with updating the "belief" on tiger's location as observations are gathered.

Table of Contents

Problem statement

A zoo keeper cleans the tiger attraction and, when finishing, he must decide which door to open to get out without being eaten by the tiger.



Tiger lies either left or right when central gate is closed, and roaring is heard either at the left or right doors... We'll assume "binary" setup:

Measurements (also denoted as "observations") are binary: `hear_left` (1 HL), `hear_right` (2 HR).

Tiger "states" are binary: `tiger_left` (1 TL), `tiger_right`(2 TR).

```
hear_L=1;  hear_R=2; %numbers indicate position on a list, variables are cathegorical and cannot be ordered/quantified.
tiger_L=1; tiger_R=2; %numbers indicate position on a list, variables are cathegorical and cannot be ordered/quantified.
```

Case 0: Central gate closes and I hear just ONE roar.

We'll define PM(m,s) is the probability of measuring "m" when state is "s". So, it is a CONDITIONAL probability table:

PM= [$p(HL|TL)$ $p(HL|TR)$;

$p(HR|TL)$ $p(HR|TR)$]

Rows represent "observations", columns represent tiger location "states".

COLUMNS must add 1. PM does not need to be symmetric.

```
%PM=[.99 .9;.01 .1]; %non symmetric... different measurement noise at each state... most times I hear "left" even if it is "right".
PM=[.8 .1;.2 .9];
%PM=[.65 .35;.35 .65];
%PM=[.55 .001;.45 .999];
PM
```

```
PM = 2×2
    0.8000    0.1000
    0.2000    0.9000
```

Note that this is a table of ***conditional*** probabilities, not a "joint" probability distribution.

It will be our "model of the tiger world" and we will do "statistical inference" from it (estimate where the tiger is, decide which door to open, decide how many roars to hear until we are "safe enough" to open the door -- that depends on fear that you have, which is modeled with additional utility/cost functions--)...

Joint versus conditional probability table

We started with a CONDITIONAL probability table.

PM= [$p(HL|TL)$ $p(HL|TR)$;

$p(HR|TL)$ $p(HR|TR)$] ;

The "JOINT" probability table is the probability table of the 4 possible "individual" events... Considered as "a random variable that can take 4 values":

[(HL∧TL),   (HL∧TR),   (HR∧TL),   (HR∧TR)]

Considered as a "two-dimensional random variable", we can put them in a 2x2 table, with probabilities that we will denote:

PJOINT= [*p*(HL,TL)   *p*(HL, TR) ;

p(HR,TL)   *p*(HR,TR)];

We changed notation with "and" connective (∧) to 2D pairs (a,b)... but it's basically the same.

But at this moment we CANNOT write down the above joint probability table: ONE PIECE OF DATA IS MISSING.

Conditional probability is defined as $p(m|t) := \frac{p(m \wedge t)}{p(t)}$, i.e.: $p(m \wedge t) = p(m|t) \cdot p(t)$.

We must then know the value of $p(TL)$, $p(TR)$ ... obviously they add one so we just need to know $b := p(TR)$ and we'll have $p(TL) = 1 - b$.

Once b is available, the joint probability table reads:

PJOINT= [$p(HL|TL) \cdot (1 - b)$ $p(HL|TR) \cdot b$;

$p(HR|TL) \cdot (1 - b)$ $p(HR|TR) \cdot b$] ;

```
b=0.60; % b=0.5 would the the most "uncertain" non-informative prior.
PJOINT=[PM(:,1)*(1-b) PM(:,2)*b]
```

```
PJOINT = 2×2
    0.3200    0.0600
    0.0800    0.5400
```

```
sum(sum(PJOINT)) %adds 1, it's a "probability measure" on this four-thing
set.
```

```
ans = 1.0000
```

```
PM %conditional, to have it seen
```

```
PM = 2×2
    0.8000    0.1000
    0.2000    0.9000
```

Interpretation of $pjoint(m, e)$**:** It is the probability of a certain measurement (observation) and of a certain place where the tiger is... considering both as random variables... "rolling the dice at the same time"...

But can that probability be interpreted as "percentage of times the event occurs if it is repeated many times"?

- **YES and NO:**

-- If repeating means that "the intermediate passage door is opened and a long time elapses (the tiger switchs sides if he wants), and the roar is heard only once" then **YES**;

-- If repeating means "close the intermediate passage door so that the tiger is always on one side and wait until you hear 10 roars" then **NO**. In that situation (or if I open the gate but allow "very little time" to elapse before closing again), the probability that the tiger is on a certain side depends on where it was "a minute before"... and "past roars" give me information about where it could be in the past that can be extrapolated to "now". Then $pjoint(m,t)$ $p(m|t)$ and $p(t)$ require a different interpretation/usage.

Belief

Belief b is my "**prior**" probability of `tiger_right`. Of course, prior probability of `tiger_left` will be $1-b$.

```
syms b real
```

Probability of observations (MARGINAL)

```
PJOINT
```

```
PJOINT = 2×2
    0.3200    0.0600
    0.0800    0.5400
```

```
marg_TLTR=sum(PJOINT) %sum of COLUMS, TIGERLOCATION marginal
```

```
marg_TLTR = 1×2
    0.4000    0.6000
```

```
marg_HLHR=sum(PJOINT,2) %sum of ROWS, HEAR marginal
```

```
marg_HLHR = 2×1
    0.3800
    0.6200
```

If we symbollically compute the marginals, for later use:

```
prob_hearL(b)  = PM(hear_L,tiger_R)*b + PM(hear_L,tiger_L)*(1-b);
vpa(prob_hearL)
```

ans(b) = $0.8-0.7\,b$

```
prob_hearR(b)  = PM(hear_R,tiger_R)*b + PM(hear_R,tiger_L)*(1-b);
vpa(prob_hearR)
```

ans(b) = $0.7\,b+0.2$

In matrix notation, the marginal is "conditional probability matrix" times "vector of prior probabilities":

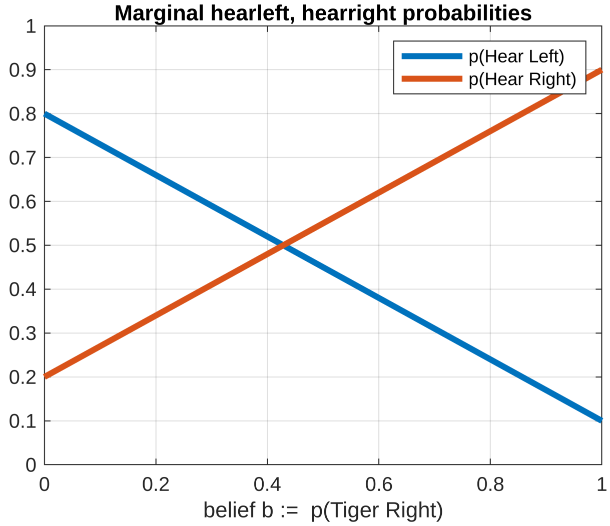
```
vpa(  PM*[(1-b); b]   , 4)
```

ans =

$$\begin{pmatrix}0.8-0.7\,b\\0.7\,b+0.2\end{pmatrix}$$

```
fplot([prob_hearL;prob_hearR],[0 1],LineWidth=3), grid on, ylim([0 1])
```

```
legend("p(Hear Left)","p(Hear Right)") %OJO interpretacion, b no es
"evento"...
xlabel("belief b :=  p(Tiger Right)"), title("Marginal hearleft, hearright
probabilities")
```



Recovering conditionals from joint and marginal probabilities:

```
PJOINT %repeat it for convenience
```

```
PJOINT = 2×2
    0.3200    0.0600
    0.0800    0.5400
```

```
CondHearGivenTiger=PJOINT./marg_TLTR
```

```
CondHearGivenTiger = 2×2
    0.8000    0.1000
    0.2000    0.9000
```

```
sum(CondHearGivenTiger) %columns add 1
```

```
ans = 1×2
    1.0000    1.0000
```

The result is no surprise, because those were the initial data.

The other conditional can also be computed... which is its meaning?:

```
CondTigerGivenHear=PJOINT./marg_HLHR
```

```
CondTigerGivenHear = 2×2
    0.8421    0.1579
    0.1290    0.8710
```

```
sum(CondTigerGivenHear,2) %rows add 1
```

```
ans = 2×1
    1.0000
    1.0000
```

The conditional probability "p(<tiger side> | observation)" makes mathematical sense (it fulfills the definitions), but (at first glance) not "conceptual" sense in the tiger problem. Observation is the "effect" and the tiger is the "cause" (condition)... the tiger stays in one side and then the roar is heard depending on the side the tiger is, and not the other way around: stating that "*you hear the noise and the tiger decides where it goes based on noise*" has no "physical sense"...

But the formula IS INDEED CORRECT, and it is the basis of Bayesian inference... Although the "causal chain" is in the other direction (first tiger decides side, gate closes, and then roars), the joint probability is what it is, and I can "compute probabilities of where is the tiger after hearing a roar to the left"... we call that "estimation" as "a posteriori probability" or "belief". That is the KEY idea of the "Bayes formula"...

Posterior belief update: Bayes rule

If I measure `hear_left`,

$$p(TR|hearleft) = \frac{p(hearleft \wedge TR)}{p(hearleft)} = \frac{p(hearleft \wedge TR)}{p(hearleft \wedge TR) + p(hearleft \wedge TL)}$$

in terms of conditional tables:

$$p(TR|hearleft) = \frac{p(hearleft|TR) \cdot p(TR)}{p(hearleft|TR) \cdot p(TR) + p(hearleft|TL) \cdot p(TL)} = \frac{p(hearleft|TR) \cdot b}{p(hearleft|TR) \cdot b + p(hearleft|TL) \cdot (1-b)}$$

```
bnext_ifHL(b) = PM(hear_L,tiger_R)*b/
(PM(hear_L,tiger_R)*b+PM(hear_L,tiger_L)*(1-b));
vpa(bnext_ifHL)
```

ans(b) =

$$-\frac{0.1\,b}{0.7\,b - 0.8}$$

```
bnext_ifHL(b) = PM(hear_L,tiger_R)*b/prob_hearL(b);
vpa(bnext_ifHL)
```

ans(b) =

$$-\frac{0.1\,b}{0.7\,b - 0.8}$$

If I measure `hear_right`,

$$p(TR|hearright) = \frac{p(hearright|TR) \cdot p(TR)}{p(hearright|TR) \cdot p(TR) + p(hearright|TL) \cdot p(TL)} = \frac{p(hearright|TR) \cdot b}{p(hearright|TR) \cdot b + p(hearright|TL) \cdot (1-b)}$$

```
bnext_ifHR(b) = PM(hear_R,tiger_R)*b / prob_hearR(b);
vpa(bnext_ifHR)
```

ans(b) =

$$\frac{0.9\,b}{0.7\,b+0.2}$$

Examples:

1.) Uninformative prior $b = 0.5$.

```
eval(bnext_ifHL(0.5))
```

ans = 0.1111

```
eval(bnext_ifHR(0.5))
```

ans = 0.8182

2.) Almost sure that tiger was at the right side:

```
eval(bnext_ifHL(0.98))
```

ans = 0.8596

```
eval(bnext_ifHR(0.98))
```

ans = 0.9955

3.) Almost sure that tiger was at the left side:

```
eval(bnext_ifHL(0.002))
```

ans = 2.5044e-04

```
eval(bnext_ifHR(0.002))
```

ans = 0.0089

4.) "Fully certain" That tiger is at left side:

```
eval(bnext_ifHL(0))
```

ans = 0

```
eval(bnext_ifHR(0))
```

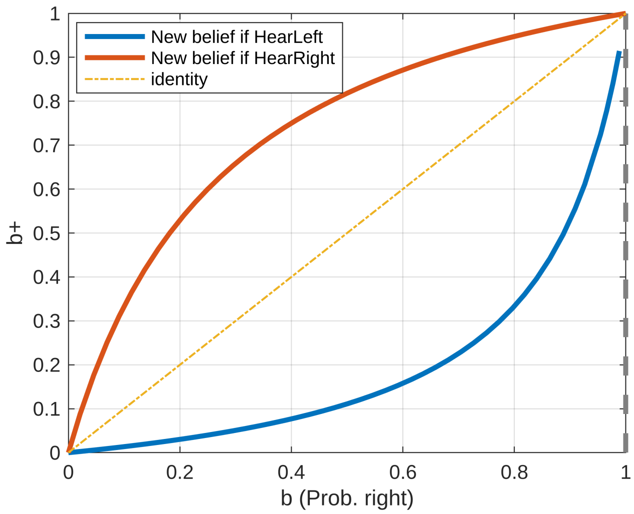
ans = 0

Graphical representation:

From a prior belief and an observation I get a posterior belief.

```
fplot([bnext_ifHL;bnext_ifHR],[0 1],LineWidth=2.5), grid on
hold on, fplot(b,[0 1],'-.',LineWidth=1), hold off
xlabel('b (Prob. right)'), ylabel('b+')
```

```
legend("New belief if HearLeft","New belief if
HearRight",'identity',Location="best")
```



There is no "fixed" point where belief new will be equal to old except total belief on right (b=1) or left (b=0). If I am "100% sure" of something, experimenting will not provide any new information.