This is an explanation of:

multivariate probability distributions

marginal probability distributions

conditional probability

Bayes' theorem

events

11
01
00
11
11
00
10
01
11
⋮

$p(x_1, x_2)$

x_1 \ x_2	0	1	
0	0.3	0.1	0.4
1	0.2	0.4	0.6
	0.5	0.5	

$p(x_1)$ marginal

$p(x_2)$ marginal of $p(x_1, x_2)$

Now, what happens if we know $x_1 = 0$?
How update with this new information?

filtered events:

0 1
0 0
0 0
0 0
0 1
⋮

$p(x_1, x_2 \mid x_1 = 0)$

x_2

x_1 \ x_2	0	1	$p(x_1 \mid x_1 = 0)$
0	0.75	0.25	1
1	0	0	0
	0.75	0.25	

from original table!

$$p(x_2 \mid x_1 = 0) = \frac{p(x_1 = 0, x_2)}{p(x_1 = 0)}$$

Same if $x_1 = 1$

$$p(x_2 \mid x_1 = 1) = \frac{p(x_1 = 1, x_2)}{p(x_1 = 1)}$$

0	0
0.33	0.67

0.33 0.67

So generally

$$p(x_2 \mid x_1) = \frac{p(x_1, x_2)}{p(x_1)}$$

0.75	0.25
0.33	0.67

and similarly

$$p(x_1 \mid x_2) = \frac{p(x_1, x_2)}{p(x_2)}$$

marginal multiplied back to $p(x_2|x_1)$ to get the original $p(x_1,x_2)$!

$\Rightarrow$

$$p(x_1)\,p(x_2|x_1) = \underbrace{p(x_1, x_2)}_{\text{this part often left out!}} = p(x_2)\,p(x_1, x_2)$$

Bayes theorem!

Think in whole tasks and
not just in single numbers!

$$p(x_1, x_2, x_3, x_4)$$

if independent

$$p(x_1, x_2, x_3, x_4) = \overbrace{p(x_1)\, p(x_2)\, p(x_3)\, p(x_4)}^{\text{we can then just multiply the marginals!}}$$

if pairwise dependent in chain (Markov)

$$p(x_1, x_2, x_3, x_4) = p(x_1)\, p(x_2 | x_1)\, p(x_3 | x_2)\, p(x_4 | x_3)$$

(follows from repeated application of Bayes theorem and the Markov property)