6.3 $\underbrace{\text{Sum, Product}}_{\text{Rule}}$, Bayes T

Once we have assigned prob dists to the correspong uncertainties of the data, it turns out there are only two fundemental rules:

- Sum & Product rules

Recall:

Joint dist = $P(x, y)$

Margin Dist = $P(x)$ $P(y)$

Conditiol Dist = $P(y \mid x)$

~~Sum Rule AKA Marginalization~~

Joint dist = Complete picture of how X & Y relate. Prob of two specific values X & Y simeltaneous

Marginal dist = Prob of single RV (X or Y) taking on value regardles of other

Conditional = Prob of specific Y, given specific X

Sum Rule aka marginilization

is a math rule to get from joint dist (full picture) to marginal

in essence, you "sum out" the variable you are not interested in

given joint distribs:

	Y=high	Y=low
X = Sun	0.4	0.1
X = Rain	0.2	0.3

want find $P(X = \text{Sun}) =$

$P(X=\text{Sun}, Y=\text{high}) + P(X=\text{Sun}, Y=\cancel{\text{high}}\ \text{low}) = 0.4 + 0.1$

$= 0.5$ = marginal dist of X

~~reason~~ to find Prob of X you consider all events where X can happen & sum up the probabilities

(8)

$$P(x) = \begin{cases} \sum_{y \in Y} P(x,y) & \text{if } y \text{ is discrete} \\ \int_y p(x,y)\,dy & \text{if } y \text{ is continuous} \end{cases}$$

Note many of the computation challenges of probabilistic modelling are due to the application of the sum rule

when many vars w/ many states the sum becomes high dimensional

Product Rules

relates the joint dist to the condition

$$P(x,y) = P(y|x)\,P(x)$$

Prod Rule = every joint dist of 2 RVs can be written as a product of two other dists

Two other dists:

1. margin Dist of first var $P(x)$
2. conditional dist of sec var $P(y|x)$

Since order (1st/2nd) is arbitrary it also means:

$$P(x,y) = P(x|y)\,P(y)$$

In ML & stats, we are often interested in making inferences of unobserved RVs given that we have observed an RV

Assume prior knowledge of $P(x)$ given $\underbrace{\text{RV } x}_{\text{unobserved}}$ & relationship $P(y|x)$ between second RV y (observed) if we observe Y, we can use Bayes theorem to draw some conclusions about X

$$BT = \underbrace{P(x|y)}_{\text{Posterior}} = \frac{\overbrace{P(y|x)}^{\text{likelihood}}\ \overbrace{P(x)}^{\text{Prior}}}{\underbrace{P(y)}_{\text{Evidence}}}$$

Prior = $P(x)$ = Subject knowlege about unobserved

likelihood = $P(y|x)$ = describes how y & x are related - w/ DISC = Prob of y if we know unobserved RV x
"Prob of y given x"

Posterior = $P(x|y)$ = this is the Quant of Interest as it expresses what we are interested in - what we know about x after obs y the quantity:

$$P(y) := \int P(y|x)\, P(x)\, dx = E_x[P(y|x)]$$

Margin probability

sum of all y Probs