2) "I have 3 kids. Each of them rolled a pair of dice. I have at least one boy who rolled (1,1)."

What is the prob. of 3 boys?

BBB = all 3 are boys

$\tilde{B}_s$ = at least one is a boy who rolled (1, 1)

$$P(BBB \mid \tilde{B}_s) = \frac{P(\tilde{B}_s \mid BBB)\, P(BBB)}{P(\tilde{B}_s)}$$

$$P(BBB) = \frac{1}{8} \qquad P(\tilde{B}_s \mid BBB) = 1-(1-\epsilon)^3$$

$$\epsilon = \text{prob. of } (1,1) = \frac{1}{36}$$

$(1-\epsilon)^3$ = prob. no one got (1,1)

trick to calculate denom. in Bayes rule:

$$1 = \sum_A P(A \mid B) = \sum_A \frac{P(B \mid A)\, P(A)}{P(B)}$$

$$= \frac{1}{P(B)} \sum_A P(B \mid A)\, P(A)$$

$$\Rightarrow \quad P(B) = \sum_A P(B \mid A)\, P(A)$$

here: $$P(\tilde{B}_s) = \sum_{\substack{XYZ \\ \text{all 3 kid combo}}} P(\tilde{B}_s \mid XYZ)\, \overbrace{P(XYZ)}^{\frac{1}{8}}$$

$$\mathcal{P}(\tilde{B}_s \mid GGG) = 0$$

$$\mathcal{P}(\tilde{B}_s \mid BGG) = \epsilon$$

$\circ\circ\circ$

$\downarrow$

final result :

$$\mathcal{P}(\tilde{B}_s) = \frac{1}{8}\epsilon\left(12 - 6\epsilon + \epsilon^2\right)$$

plug in Bayes: $\mathcal{P}(BBB \mid \tilde{B}_s) = \dfrac{1-(1-\epsilon)^3}{\epsilon\left(12 - 6\epsilon + \epsilon^2\right)}$

$$\epsilon = \frac{1}{36} \Rightarrow \quad \approx 0.25 > \frac{1}{7}$$

limit of small ϵ $\quad \approx \dfrac{1}{4} - \dfrac{\epsilon}{8}$

---

more abstractly:

numerator: "likelihood"

$$\mathcal{P}(BBB \mid \tilde{B}_s) = \underbrace{\left[\frac{\mathcal{P}(\tilde{B}_s \mid BBB)}{\mathcal{P}(\tilde{B}_s)}\right]}_{\text{rule for updating knowledge}} \mathcal{P}(BBB)$$

"posterior"
knowledge after accounting for data $\tilde{B}_s$

"prior"
knowledge before knowing $\tilde{B}_s$
(hypothesis)

"likelihood":
prob. of data given the prior knowledge (hypothesis)

model fitting: dataset $\mathcal{D}$

model w/ some params $\vec{w} = (w_1, w_2, \ldots)$

$$\mathcal{P}(\vec{w} \mid \mathcal{D}) = \frac{\mathcal{P}(D \mid \vec{w})}{\mathcal{P}(D)} \, p(\vec{w})$$

posterior

prior
(reflects physical ranges of params)

Two major applications:

1) find "best" model parameters:

maximize $\mathcal{P}(\vec{w} \mid \mathcal{D})$ w/ respect to $\vec{w}$

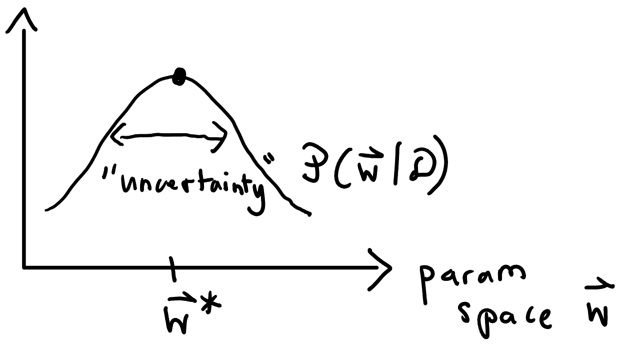
$\Rightarrow$ $\vec{w}^*$ is the "best" parameter set

$\Rightarrow$ MAP: maximum a posteriori fitting

(typical model fitting: i.e. training neural networks)

2) find $\mathcal{P}(\vec{w} \mid \mathcal{D})$ directly:

gives info about uncertainty of $\vec{w}^*$ estimates



$\Rightarrow$ generally hard: Bayesian neural networking

$\Rightarrow$ neat trick from stat. mech that enables this (we will return to it!)

$y = ax^2 + bx + c$