I have 3 kids. Each roll a pair of dice, + I have at lecst one boy who got (1,1). What is the prob. that all three are boys?

BBB = all 3 kids are boys

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

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

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

$$\mathcal{P}(BBB) = \frac{1}{8}$$

$$\mathcal{P}(\hat{\tilde{B}}_s \mid BBB) = 1 - \underbrace{(1-\epsilon)^3}_{\text{prob. of none get. (1,1)}}$$

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

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

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

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

$$\mathcal{P}(\tilde{B}_s) = \sum_{\substack{XYZ \\ \in \text{ all 3 kid combos}}} \mathcal{P}(\tilde{B}_s \mid XYZ)\, \underbrace{\mathcal{P}(XYZ)}_{1/8}$$

$$\mathcal{P}(\tilde{B}_s \mid GGG) = 0 \qquad \mathcal{P}(GGG) = \frac{1}{8}$$

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

$$\circ \circ \circ$$

$$\mathcal{P}(\tilde{B}_s \mid BBG) = 1 - (1-\epsilon)^2$$

$$\vdots$$

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

$$\Rightarrow \quad \mathcal{P}(BBB \mid \tilde{B}_s) = \frac{1-(1-\epsilon)^3}{\epsilon(12 - 6\epsilon + \epsilon^2)}$$

$$\epsilon = \frac{1}{36} \Rightarrow \quad \approx 0.247 > \frac{1}{7}$$

limit of small ϵ — Taylor expand $\approx \frac{1}{4} - \frac{\epsilon}{8}$

---

more abstract interpretation

$$\underbrace{\mathcal{P}(BBB \mid \underset{\text{data}}{\tilde{B}_s})}_{\text{"posterior"}} = \underbrace{\left[\frac{\mathcal{P}(\tilde{B}_s \mid BBB)}{\mathcal{P}(\tilde{B}_s)}\right]}_{\text{rule for updating knowl.}} \underbrace{\mathcal{P}(BBB)}_{\text{"prior" knowl. before knowing data}}$$

"posterior" knowledge after accounting for data

$\mathcal{P}(\tilde{B}_s \mid BBB)$ = "likelihood" : prob. of data given knowledge

$\mathcal{P}(\tilde{B}_s)$ = normaliz. constant

---

model fitting : dataset $\mathcal{D}$

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

$$\underbrace{\mathcal{P}(\vec{w} \mid \mathcal{D})}_{\text{posterior}} = \frac{\mathcal{P}(\mathcal{D} \mid \vec{w})}{\mathcal{P}(\mathcal{D})} \underbrace{\mathcal{P}(\vec{w})}_{\text{prior}}$$

prior usually reflects physical ranges of possible params

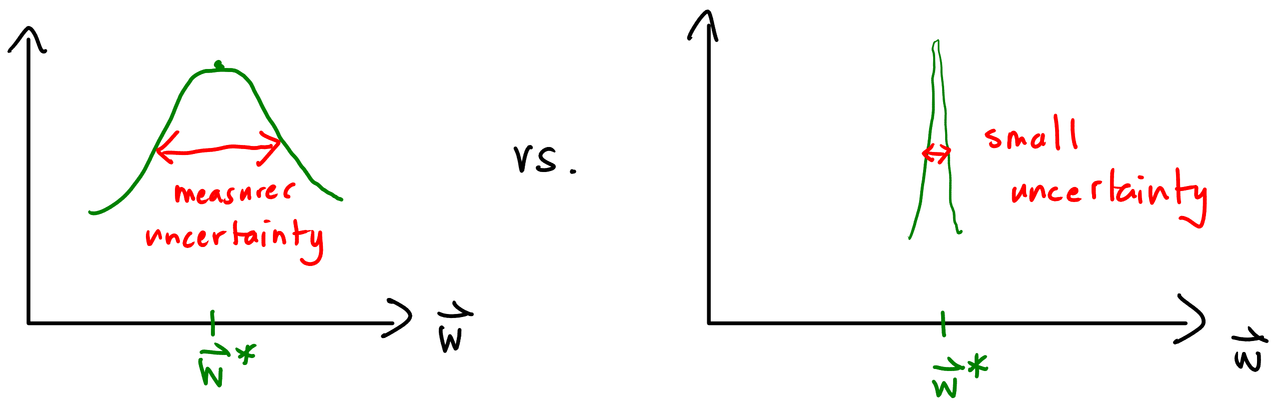
alternative: $\mathcal{P}(\vec{w}) = \text{const.}$
$\Rightarrow$ no clue what params are ("uniform" prior)

Two major applications:

1) find "best" model parameters
$\Rightarrow$ maximize $\mathcal{P}(\vec{w} \mid \mathcal{D})$ w/ respect to $\vec{w}$
$\Rightarrow$ $\vec{w}^*$ is the "best" param. 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:



$\Rightarrow$ gives you info. about uncertainty of $\vec{w}^*$ estimate

$\Rightarrow$ generally hard: Bayesian neural networking

$\Rightarrow$ later on: use a neat from stat. mech to do this