PHYS 177
winter 2019

lecture 20: Bayesian inference

learning goals
- recall Bayes' theorem
- conceptualize prior distributions

in the last example, we saw pathological results w/ maximum likelihood w/ little data

w/ one observation, $\hat{\epsilon} = \pm\infty$! this seems unreasonable

what is our expectation for ϵ before we measure?
can we average over different ϵ, weighted by their likelihood?

Bayesian inference is this measured approach

results follow from Bayes' theorem, a simple observation about joint probabilities

$$\underbrace{P(A,B)}_{\text{prob of A and B}} = \underbrace{P(B)}_{\text{prob of B only}} \underbrace{P(A|B)}_{\text{prob of A given B ("conditional probability")}} = P(A)P(B|A)$$

example: A = it is raining, B = I am carrying an umbrella

rearrangement is Bayes' theorem: $P(A|B) = \dfrac{P(B|A)P(A)}{P(B)}$

in the context of inference,

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

probability distribution for parameters θ, depending on the data X!

$P(X|\theta)$ is described by our stochastic model

$P(\theta)$ is the prior distribution for parameters, could reflect real knowledge (informative prior) or mainly serve as regularization (uninformative or "weakly informative" prior)

in Bayesian inference, we choose prior distributions and then average over $P(\theta|X)$, the posterior distribution

* notebook example