PHYS 177
winter 2019

lecture 19: maximum likelihood

learning goals
- use likelihood to pose a statistical inference problem

last time, loss functions were general (and arbitrary...)

for stochastic models, there is a standard approach – what loss makes sense for stoch. models?

in principle, many sets of parameters could have generated data but which ones are most likely to have generated our data?

$$\max_{\theta} P(X|\theta) \quad \text{"maximum likelihood"}$$

often for numerical reasons we take a log transform,

$$\max_{\theta} \log P(X|\theta)$$

example: Ising spin

B field ↑ ↑ spin σ

$$E = -\epsilon\sigma, \quad P(\sigma=+1) = \frac{e^{\epsilon/T}}{Z}, \quad P(\sigma=-1) = \frac{e^{-\epsilon/T}}{Z}$$

say that we get data $\{+,+,+,-,+,-,-,+,\ldots\}$

$$\text{then } P(\text{data}) = \underbrace{P(\sigma=+1)\times P(\sigma=+1)\times\ldots}_{\text{product of observations}}$$

given many observations, we could try to figure out the strength of interaction b/w B field and spin, divided by T

assume T=1, then we can write log-likelihood

$$\log P(\text{data}) = \log P(\sigma=+1) + \log P(\sigma=+1) + \ldots$$

and note that

$$\log P(\sigma=\pm1) = \pm\epsilon - \log Z$$

so if there are N data points,

$$\log P(\text{data}) = \sum_{i=1}^{N}\epsilon\sigma_i - N\log Z$$

* notebook example

$$\frac{\partial \log P(\text{data})}{\partial\epsilon} = n_+ - n_- + N\frac{e^{-\epsilon}+e^{\epsilon}}{Z}$$

$$= n_+ - n_- - N\tanh\epsilon = 0$$

$$\hat{\epsilon} = \tanh^{-1}\left(\frac{n_+-n_-}{N}\right)$$

↑ ML estimate of ϵ