TAM 598 Lecture 12:

Bayesian Inference - Analytical Examples

Announcements:

- HW 3 covers lectures 8-12; due on Mar 12

Bayesian Inference - a way to fit a model to data.

Say our model predicts the result of a random experiment.
we describe the model using pdf

$p(x \mid \theta) \Longrightarrow$ probability of observing outcome x, given model parameters θ

our data $X_{1:N} = (x_1, x_2, \ldots, x_N)$

our problem: how do we use the data to learn about θ?

Maximum Likelihood Principle - choose the parameters θ that maximize the likelihood of measuring data $X_{1:N}$

likelihood of a single data point x_n is $\boxed{p(x_n|\theta)}$

likelihood of the entire dataset $x_{1:N}$ is the joint pdf of all observations. Since all observations are independent, conditioned on θ

$$p(x_{1:N}|\theta) = p(x_1|\theta)\, p(x_2|\theta) \dots p(x_N|\theta)$$
$$= \boxed{\prod_{n=1}^{N} p(x_n|\theta)}$$

Choose θ' inside all possible θ

Max. Likelihood: find the θ that maximizes $p(x_{1:N}|\theta)$

$$\boxed{\theta_{ML} = \underset{\theta}{\operatorname{argmax}}\; p(x_{1:N}|\theta)}$$

This gives us a point estimate of parameters θ. But there should be some uncertainty in our selected θ, given limited data $x_{1:N}$. How do we quantify this? Bayes rule.

we start with our prior $p(\theta)$ in the absence of any data (principle of max entropy)

$$\theta \sim p(\theta)$$

Bayes:

$$p(A|B) = \frac{p(AB)}{p(B)} = \frac{p(B|A)\,p(A)}{p(B)}$$

A = model parameters are θ

B = the data are $x_{1:N}$

$$\underbrace{p(\text{model parameters} \mid \text{data})}_{\text{posterior}\ p(\theta \mid x_{1:N})} = \frac{\overbrace{p(\text{data} \mid \text{model params})}^{\text{likelihood}\ p(x_{1:N} \mid \theta)}\ \overbrace{p(\text{model params})}^{\text{prior}\ p(\theta)}}{\underset{p(x_{1:N})}{p(\text{data})}}$$

so: $$p(\theta \mid x_{1:N}) = \frac{p(x_{1:N} \mid \theta)\, p(\theta)}{p(x_{1:N})}$$

and $$\boxed{p(\theta \mid x_{1:N}) \propto p(x_{1:N} \mid \theta)\, p(\theta)}$$

Example: inferring the probability of a coin toss from data.

- toss a coin N times, $x_{1:N} = (x_1, \dots, x_N)$ are the results
  tails $= 1$, heads $= 0$. θ = prob. of heads
- say we know nothing initially: $\theta \sim \mathcal{U}([0,1])$
- likelihood of data is $p(x_{1:N} \mid \theta) = \prod_{n=1}^{N} p(x_n \mid \theta)$
- each measurement $x_n \mid \theta \sim \text{Bernoulli}(\theta)$ $\quad p(x_n \mid \theta) = \begin{cases} \theta & x_n = 1 \\ 1-\theta & x_n = 0 \end{cases}$

① each measurement $X_n \mid \theta \sim \text{Bernoulli}(\theta)$ so $p(x_n \mid \theta) = \begin{cases} \theta & x_n = 1 \\ 1-\theta & x_n = 0 \end{cases}$

$$p(x_n \mid \theta) = \theta^{x_n} (1-\theta)^{1-x_n}$$

if $x_n = 1$, $p \to \theta$
if $x_n = 0$, $p \to 1-\theta$

② so then likelihood

$$p(x_{1:N} \mid \theta) = \prod_{n=1}^{N} p(x_n \mid \theta) = \prod_{n=1}^{N} \theta^{x_n} (1-\theta)^{1-x_n}$$

$$= \boxed{\theta^{\sum_{n=1}^{N} x_n} (1-\theta)^{N - \sum_{n=1}^{N} x_n}}$$

called the Beta dist.

③ and our posterior $\propto$ (likelihood)(prior)

$$p(\theta \mid x_{1:N}) \propto p(x_{1:N} \mid \theta)\, p(\theta)$$

$$\propto \theta^{\sum_{n=1}^{N} x_n} (1-\theta)^{N - \sum_{n=1}^{N} x_n}$$

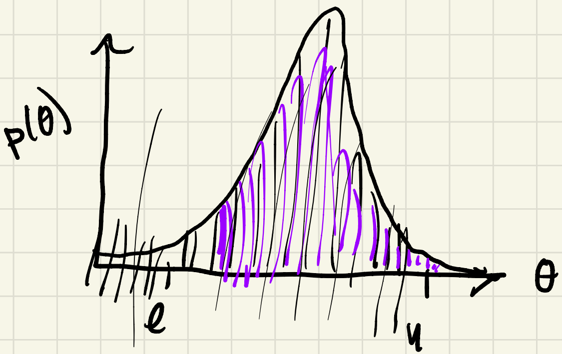
$\text{Beta}(\theta \mid \alpha, \beta)$

$= \frac{1}{B(\alpha,\beta)} x^{\alpha-1} (1-x)^{\beta-1}$

$\alpha = 1 + \sum_{n=1}^{N} x_n$

$\beta = 1 + N - \sum_{n=1}^{N} x_n$

Credible Intervals: an interval inside which the parameter θ lies with high probability. Eg, a 95% credible interval (l, u) for θ is



$$p(l \le \theta \le u \mid x_{1:N}) = 0.95$$

Note: not unique. We often work with a **central credible interval**

$$p(\theta \le l \mid x_{1:N}) = 0.025$$

$$p(\theta \le u \mid x_{1:N}) = 0.975$$

Decision Making: what if you are asked to report a single value for θ? You need to make a decision.

A wrong decision incurs a cost or a loss. The best decision minimizes the cost.

let $\ell(\theta', \theta)$ be the cost incurred if we guess θ' and the actual value is θ.

cost is subjective. some options:

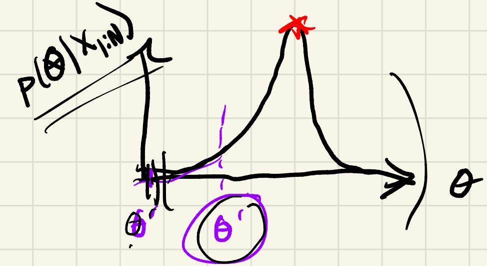
peak max $p(\theta|x_{1:N})$ (1) "0-1 loss" $\quad \ell(\theta', \theta) = \begin{cases} 1 & \text{if } \theta' \neq \theta \\ 0 & \text{if } \theta' = \theta \end{cases}$

expectation or mean (2) square loss $\quad \ell_2(\theta', \theta) = (\theta' - \theta)^2$

median (3) absolute loss $\quad \ell_1(\theta', \theta) = |\theta' - \theta|$

approach: minimize the expected loss

$$\theta^* = \min_{\theta'} \mathbb{E}\left[\ell(\theta', \theta) \,\middle|\, x_{1:N} \right]$$



your decision

$$= \min_{\theta'} \int \ell(\theta', \theta)\, p(\theta \mid x_{1:N})\, d\theta$$

(1) 0-1 loss: $\theta^* = \arg\max_{\theta} p(\theta \mid x_{1:N})$

(2) square loss $\mathbb{E}[(\theta' - \theta)^2] = \mathbb{E}[\theta^2] - 2\theta' \mathbb{E}[\theta] + (\theta')^2$

minimize w.r.t. θ' $\quad \frac{d}{d\theta'}(\quad) = -2\mathbb{E}[\theta] + 2\theta' = 0$

$$\theta' = \mathbb{E}[\theta]$$

$$\Longrightarrow \mathbb{E}[\theta \mid x_{1:N}] = \int \theta\, p(\theta \mid x_{1:N})\, d\theta$$