IIT Jodhpur

Biological Vision and Applications
Module 03-07: Parameter Estimation

Hiranmay Ghosh

How to estimate a parameter ?

Maximum Likelihood estimation

- Bayesian framework of reasoning assumes some conditional probabilities (priors)
  - e.g., $P(\text{Red} \mid \text{Banana}) = 0.1$
- Where do you get the number from?
- Maximum likelihood estimation (purely data-driven):
  - You observe 20 bananas; 2 are red
  - $P(\text{red} \mid \text{banana}) = \frac{2}{20} = 0.1$
- Not reliable, if the sample size is small
  - Does not tell you how reliable the estimate is

Bayesian Theory provides a more reliable methods

Combines prior belief with observations

- Let the parameter $\theta = P(\text{red} \mid \text{banana})$
  - $\theta \in [0, 1]$
- **Prior hypotheses**: Without any further information, we may assume
  - All values of $\theta \in [0, 1]$ are equi-probable
  - i.e. the pdf $p(\theta)$ has a uniform distribution. $p(\theta) = 1$
- Now, we depend on data (observations) to update the belief
- By Baye's law

$$p(\theta \mid d) = \frac{P(d|\theta).p(\theta)}{P(d)}$$

$$\text{where } P(d) = \int_0^1 P(d \mid \theta).d\theta$$

$P(x)$ is probability (discrete), $p(x)$ is probability density function (continuous)

Bernoulli's Theorem

- Consider the problem
  - Suppose, you toss a coin t times
  - Probability of head is θ on every toss (known)
  - What is the probability of the outcome d: h heads and $t - h$ tails?
- Bernoulli's theorem:

$$P(d \mid \theta) = \theta^h.(1-\theta)^{t-h}$$

- Easy to derive. ... Try it out!

Back to our problem

Tossing a coin

- Using Bernoulli's theorem

$$
\begin{aligned}
P(d) &= \int_0^1 P(d \mid \theta).d(\theta) = \int_0^1 \theta^h.(1-\theta)^{t-h}.d\theta \\
&= \frac{h!.(t-h)!}{(t+1)!}
\end{aligned} \tag{1}
$$

- We have

$$
\begin{aligned}
&p(\theta) = 1 && \text{[Uniform probability assumption]} && (2)\\
&P(d \mid \theta) = \theta^h.(1-\theta)^{t-h} && \text{[Bernoulli's theorem]} && (3)\\
&p(\theta \mid d) = \frac{P(d|\theta).p(\theta)}{P(d)} && \text{[Bayes Theorem]} && (4)
\end{aligned}
$$

- Substituting (1), (2), (3) in (4)

$$
\boxed{p(\theta \mid d) = \frac{P(d|\theta).p(\theta)}{P(d)} = \frac{(t+1)!}{h!.(t-h)!}.\theta^h.(1-\theta)^{t-h}}
$$

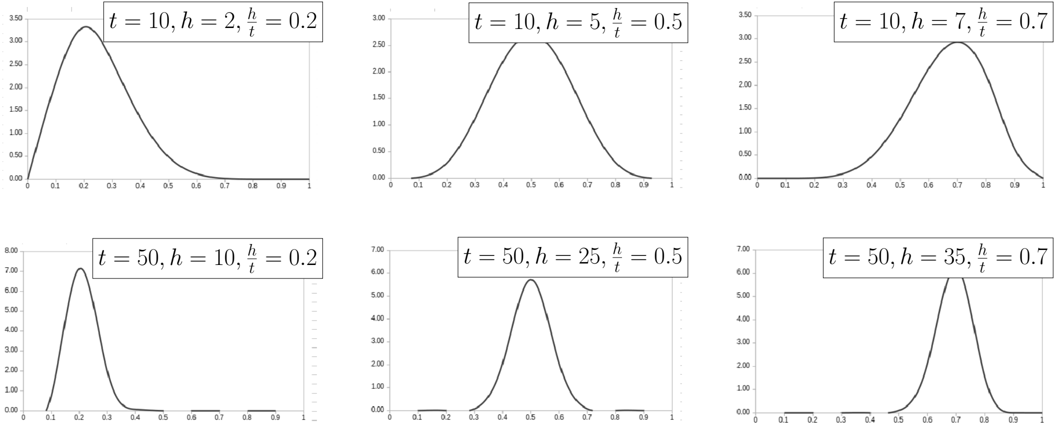
Discussions

- We get a pdf for θ, rather than a single value
  - More informative
- Expected value for θ is

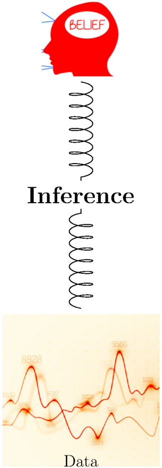
$$\boxed{\hat{\theta} = \int_0^1 \theta . p(\theta \mid d) . d\theta = \frac{h+1}{t+2}}$$

- Bayesian method vs. max likelihood estimate
- Let's assume, we have observed 2 bananas, none is red
  - $h = 0, t = 2$
- By max. likelihood: $\theta = \frac{h}{t} = 0$
- By Bayesian method: $\hat{\theta} = \frac{h+1}{t+2} = \frac{1}{4}$
  - Prior belief in Bayesian method moderates the extreme estimates

Dependence of pdf on data



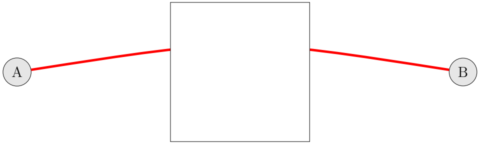
Priors vs. data (observations)

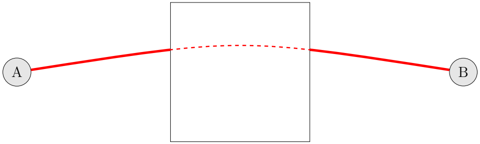


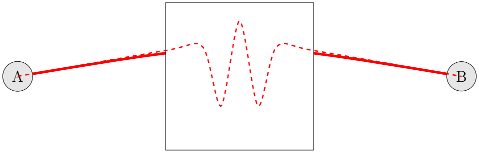
- We have assumed uniform pdf $p(\theta) = 1$ in this example.
  - It is possible of assume other priors
  - What determines the priors?
- Prior belief dominates so long there are less observations
- Data tends to dominate with increased number of observations
- Weak prior $\Rightarrow$ it takes less data to update the parameter
  - Susceptible to noisy data
- Strong prior $\Rightarrow$ it takes more data to update the parameter
  - Susceptible to erroneous prior

How do we get the priors ?

On complexity of models and priors



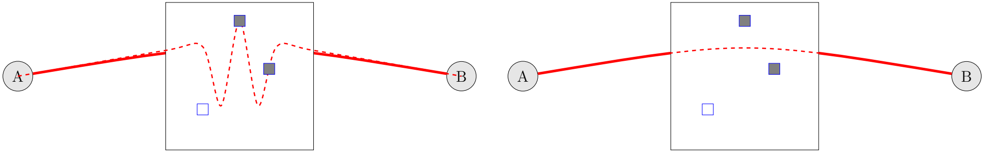




- Human mind tends to choose the simplest model

What do the data say?

Goodness of fit

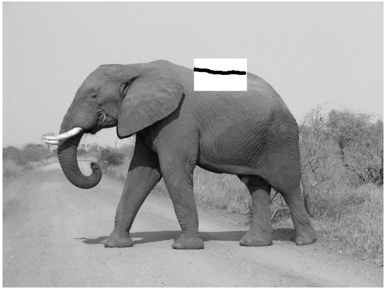


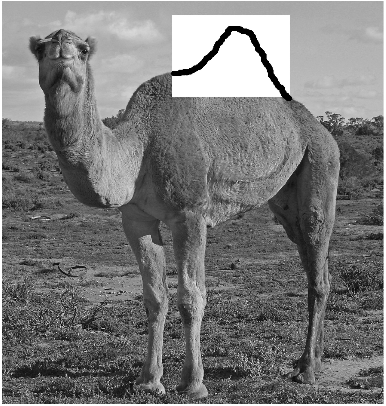
- How well does the fit a model?

Complexity and Prior

- Let $c(M)$ denote the complexity for a model M
  - Prior probability for M can be expressed as: $P(M) = 2^{-c(M)}$ (axiom)
  - Probabilies for the model hypotheses (priors and conditionals) and inference
    - $P(h_i) = 2^{-c(h_i)}$
    - $P(d \mid h_i) = 2^{-c(d|h_i)}$
    - $P(h_i \mid d) = 2^{-c(h_i|d)}$
- Baye's law $P(h_i \mid d) = \kappa.P(h_i).P(d \mid h_i)$
- Substituting, and taking logarithm
  - $c(h_i \mid d) = k + c(h_i) + c(d \mid h_i)$
- Human mind chooses the inference with least complexity
  - Complexity of inference is the sum of complexities of hypotheses
  - Belief maximization $\equiv$ complexity minimization

Complexity (prior probability) is guided by knowledge





Quiz

Quiz 03-07

End of Module 03-07