# MLE - MAP

| | |
|---|---|
| Type | Lecture Notes |
| Reviewed | ☐ |
| Available Summary? | In progress |
| Week | 4 |

## MLE (Maximum Likelihood Estimation)

- Choose $\theta$ that maximizes the probability of observed data (likelihood of the data)

### Let's consider a flipping coin case: (H: heads and T: tails)

- Density estimation is a learning problem too:
  - Data: Observed set of flips with $\alpha_H$ heads and $\alpha_T$ tails
  - Hypothesis: Bernoulli distribution
  - Learning: Finding $\theta$, which is an optimization problem
- The likelihood of observing this data is the joint probability:

  $P(D|\theta) = \theta^{\alpha_H}(1-\theta)^{\alpha_T}$

### Maximum likelihood estimate of $\theta$:

$\hat{\theta} = argmax_\theta P(D|\theta)$ —> probability of D given theta

$\hat{\theta} = lnP(D|\theta)$

$\hat{\theta} = ln\theta^{\alpha_H}(1-\theta)^{\alpha_T}$

take the derrivative and set to zero:

$\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H+\alpha_T}$

#### from Hoeffding's Inequality:

- Let $N = \alpha_H + \alpha_T$ and $\hat{\theta}_{MLE} = \frac{\alpha_H}{\alpha_H+\alpha_T}$
  - Let $\theta^*$ be the true parameter. For any $\epsilon > 0$;

    $P(|\hat{\theta} - \theta^*| \geq \epsilon) \leq 2e^{-2Ne^2}$
  - If we are measuring a continuous variable:

    $N(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}}e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$

Probably Approximately Correct (PAC)

#### Learning a gaussian distribution:

- Probability of i.i.d. samples $D = \{x_1, ..., x_n\}$:

  $P(D|\mu,\sigma) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \prod_{i=1}^{N} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$

  $\mu_{MLE}, \sigma_{MLE} = argmax_{\mu,\sigma} P(D|\mu,\sigma)$
- Log-likelihood of data:

  $lnP(D|\mu,\sigma) = ln\left[\frac{1}{\sigma\sqrt{2\pi}}^N \prod_{i=1}^{N} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}\right]$

  $= -N.ln\sigma\sqrt{2\pi} - \sum_{i=1}^{N}\frac{(x_i-\mu)^2}{2\sigma^2}$
- MLE for the mean:

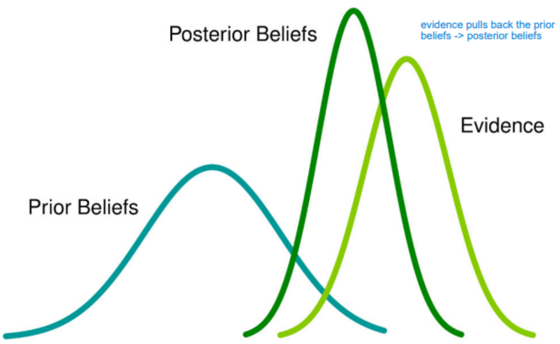
  $\hat{\mu}_{MLE} = \frac{d}{d\mu}lnP(D|\mu,\sigma) = \frac{1}{N}\sum_{i=1}^{N} x_i$
- MLE for the variance:

  $\hat{\sigma}^2_{MLE} = \frac{d}{d\sigma}lnP(D|\mu,\sigma) = \frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$
- MLE for the variance of a gaussian is biased. That is expected result of the estimator is not the true parameter. An unbiased variance estimator:

  $\hat{\sigma}^2_{unbiased} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2$

## Bayesian Rule

$$\underbrace{p(\theta|D)}_{\text{posterior}} = \frac{\overbrace{p(D|\theta)}^{\text{likelihood}} . \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(D)}_{\text{normalization (evidence)}}}$$




### prior

- it can represent some expert knowledge
- it can be form a uniform distribution
- it can be estimated from data (mostly in naive bayes)

### MAP estimation (Maximum A Posteriori)