# Analytical inference

# Posterior distribution

Likelihood

Prior

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

Evidence

**What is $P(X)$?**

# Posterior distribution

Likelihood

Prior

This is the posterior. >>

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

Evidence

**What is $P(X)$?**

P(X) is the evidence.

Van Gogh Starry night

P(X) is the evidence. It is van gogh's starry night painting. When we know P(X) distribution, we may be able to generate other paintings that van gogh might have drawn. like.. (turn to next slide)..

# Posterior distribution

$$\overbrace{P(X|\theta)}^{\text{Likelihood}}\ \overbrace{P(\theta)}^{\text{Prior}}$$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\underbrace{P(X)}_{\text{Evidence}}}$$

**What is $P(X)$?**

Van Gogh Starry night

Van Gogh, Starry night over the Rhone

# Maximum a posteriori

We want to maximise the posterior.

$$\theta_{\mathrm{MP}} = arg\max_{\theta} P(\theta|X)$$

By bayes rule..

$$\theta_{\mathrm{MP}} = arg\max_{\theta} \frac{P(X|\theta)P(\theta)}{P(X)}$$

P(X) doesn't depend on theta, so we don't need it:

$$\theta_{\mathrm{MP}} = arg\max_{\theta} P(X|\theta)P(\theta)$$

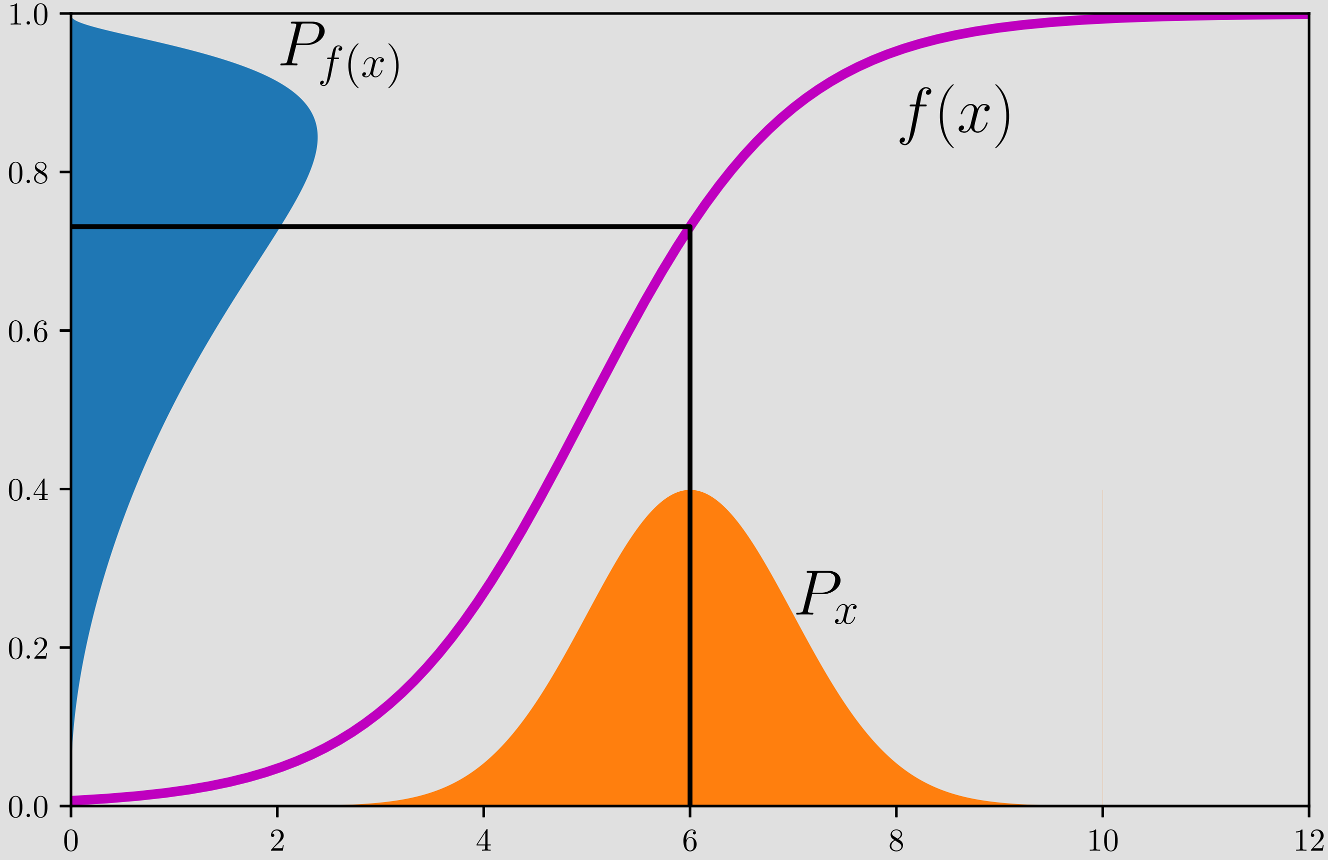
P(X | theta) is sort of like the opposite 'loss' or 'cost' function here? So a good fitting theta will lead to high probability?

**Optimization problem!**

# MAP: problems

for example. if we have a gaussian distribution P_x, and we apply the sigmoid on it to form P_{f(x)}, we have a problem that the position of the maximum has changed. This is what is meant by 'not invariant to reparametrization'

## Not invariant to reparametrization



# MAP: problems

Can't use as prior

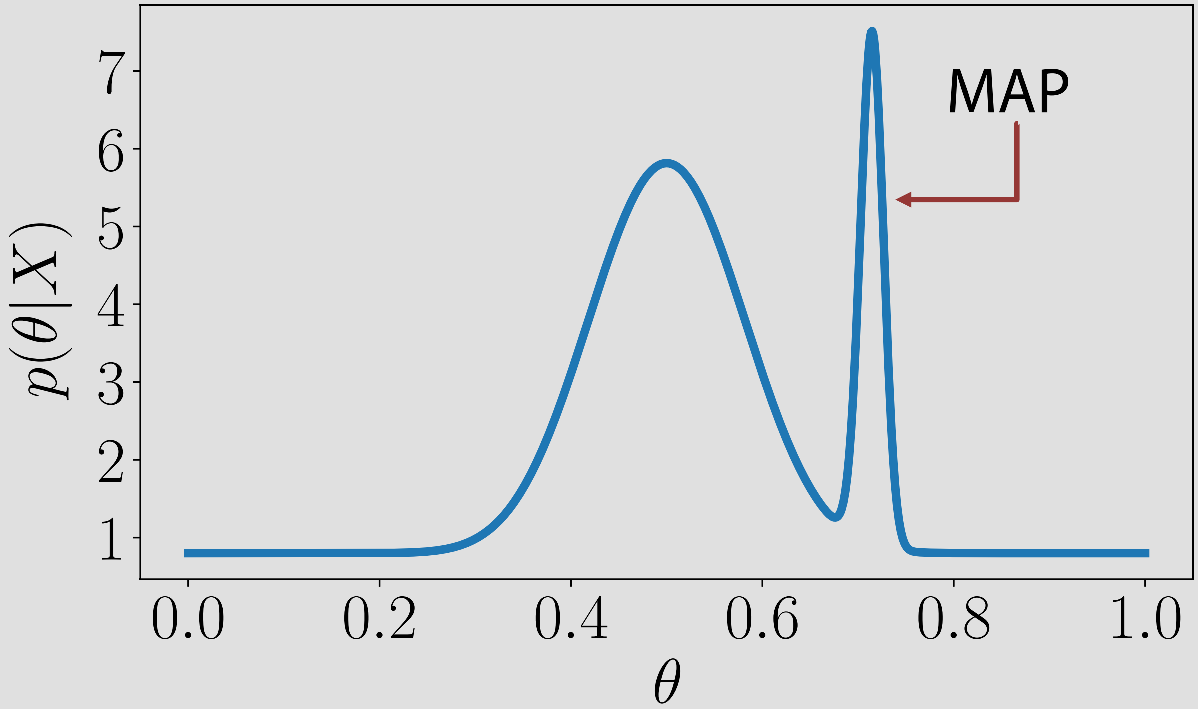
We will not get any new information.

$$P_k(\theta) = \frac{P(x_k|\theta)P_{k-1}(\theta)}{P(x_k)}$$

$$P_k(\theta) = \frac{P(x_k|\theta)\delta(\theta - \theta_{\mathrm{MP}})}{P(x_k)} = \delta(\theta - \theta_{\mathrm{MP}})$$

# MAP: problems

MAP is a solution to $L(\theta) = \mathbb{I}[\theta \neq \theta^*] \to \min_\theta$



# MAP: problems

| Objectives | Solution |
| --- | --- |
| $L(\theta) = \mathbb{I}[\theta \neq \theta^*] \to \min\limits_{\theta}$ | Mode |
| $L(\theta) = \mathbb{E}(\theta - \theta^*)^2 \to \min\limits_{\theta}$ | Mean |
| $L(\theta) = \mathbb{E}\|\theta - \theta^*\| \to \min\limits_{\theta}$ | Median |

# MAP: problems

Can't compute credible regions

$$\theta_{\mathrm{MP}} = 12.53$$

# MAP: problems

Can't compute credible regions

$$\theta_{\text{MP}} = 12.53$$

$$\theta_{\text{MP}} = 12.53 \pm 1000$$

$$\theta_{\text{MP}} = 12.53 \pm 0.001$$

# Summary

**Pros:**

- Easy to compute

**Cons:**

- Not invariant to reparametrization
- Can’t use as a prior
- Finds untypical point
- Can’t compute credible intervals