# Bayesian Statistics and Machine Learning

Dr. Eve Bayesian

May 20, 2025

## 1 Bayes' Theorem

The foundation of Bayesian statistics is Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

In the context of machine learning, this becomes:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} \tag{2}$$

where $\theta$ represents model parameters and $D$ represents data.

## 2 Prior, Likelihood, and Posterior

- **Prior** $P(\theta)$: Our beliefs about parameters before seeing data
- **Likelihood** $P(D|\theta)$: Probability of data given parameters
- **Posterior** $P(\theta|D)$: Updated beliefs after seeing data

## 3 Conjugate Priors

A prior is **conjugate** to a likelihood if the posterior belongs to the same family as the prior.

### 3.1 Beta-Binomial Example

For a binomial likelihood with Beta prior:

$$\text{Likelihood: } P(x|n,p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{3}$$

$$\text{Prior: } P(p) = \text{Beta}(\alpha, \beta) \tag{4}$$

$$\text{Posterior: } P(p|x) = \text{Beta}(\alpha + x, \beta + n - x) \tag{5}$$

## 4 Maximum A Posteriori (MAP)

The MAP estimate maximizes the posterior:

$$\hat{\theta}_{MAP} = \arg\max_{\theta} P(\theta|D) = \arg\max_{\theta} P(D|\theta)P(\theta) \tag{6}$$

## 5 Regularization

Bayesian methods naturally provide regularization. For linear regression with Gaussian priors:

$$P(\mathbf{w}|D) \propto \exp\left(-\frac{1}{2\sigma^2}\|y - X\mathbf{w}\|^2 - \frac{\lambda}{2}\|\mathbf{w}\|^2\right) \tag{7}$$

This is equivalent to L2 regularization (Ridge regression).

## 6 Markov Chain Monte Carlo

For complex posteriors, we use MCMC methods like the Metropolis-Hastings algorithm:

1. Sample $\theta'$ from proposal distribution $q(\theta'|\theta^{(t)})$
2. Accept with probability $\min\left(1, \frac{P(\theta'|D)q(\theta^{(t)}|\theta')}{P(\theta^{(t)}|D)q(\theta'|\theta^{(t)})}\right)$
3. If accepted, set $\theta^{(t+1)} = \theta'$; otherwise, $\theta^{(t+1)} = \theta^{(t)}$

## 7 Applications in ML

Bayesian methods are used in:

- Gaussian Process regression
- Bayesian neural networks
- Latent Dirichlet Allocation (LDA)
- Variational inference
- Uncertainty quantification

## 8 Conclusion

Bayesian statistics provides a principled framework for incorporating uncertainty and prior knowledge into machine learning models.