Introduction to Bayesian Statistics

Wenda Zhou

June 14, 2017

What is Bayesian statistics

Alternative statistical paradigm.
Considers parameters as random quantities (in the frequentist world, parameters are fixed).

Example

Suppose you toss a coin twice, and observe two heads. The frequentist estimate of p is given by $\hat{p} = 1$.
Want to use your experience that coins are usually not too biased.

Priors

Priors

Belief about the state of the parameter is expressed through a distribution called the prior, often written $\pi(\theta)$.

Inference

Inference is conducting using Bayes' rule to update the distribution of the parameter after seeing the data.

Bayes rule

The Bayes rule for conditional probability gives us the following:

$$\pi(\theta \mid X) = \frac{f_X(X \mid \theta)\pi(\theta)}{f_X(x)} \tag{1}$$

$\pi(\theta \mid X)$ is called the posterior distribution.

Bayes rule

The Bayes rule for conditional probability gives us the following:

$$\pi(\theta \mid X) = \frac{f_X(X \mid \theta)\pi(\theta)}{f_X(x)} \tag{1}$$

$\pi(\theta \mid X)$ is called the posterior distribution.
Equivalently, we may write it in terms of the likelihood to obtain

$$\pi(\theta \mid x) \propto L(X \mid \theta)\pi(\theta) \tag{2}$$

as $f_X(x)$ is simply a constant.

Bayes rule

Example: binomial distribution

Consider X binomial with parameters n and p, Suppose that p has a prior given by a $\mathrm{Beta}(\alpha, \beta)$ distribution with p.d.f.:

$$\pi(p) = \frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha, \beta)} \tag{3}$$

Then the posterior distribution of p is a $\mathrm{Beta}(\alpha + X, \beta + (n - X))$ distribution.

Bayes rule

Example: normal distribution

Consider X normal with parameters μ and $\sigma^2 = 1$. Suppose that μ has a prior given by $\mu \sim \mathcal{N}(0, 1)$.
The posterior distribution of μ given by:

$$\mu \mid X \sim \mathcal{N}(X/2, \frac{1}{2}) \tag{4}$$

Conjugate priors

In general, the posterior need not be of the same family as the prior distribution. If it does happen, the prior is called conjugate
If the prior is not conjugate, need to resort to numerical methods.

Bayesian estimators

Obtain posterior distribution. However, often want a number.

Posterior mean

Most often use posterior mean as an estimator, the mean of the parameter under the posterior distribution.

Bayesian estimator

Example: binomial

If $p \sim \mathrm{Beta}(\alpha, \beta)$ distribution, and $X \sim \text{Binom}(n, p)$ then the posterior mean is given by:

$$\hat{p}^B = \frac{\alpha + X}{\alpha + \beta + n} \tag{5}$$

Bayesian estimator

Example: normal

If $\mu \sim \mathcal{N}(0, 1)$ and $X \sim \mathcal{N}(\mu, 1)$.
The posterior mean is given by:

$$\hat{\mu}^B = X/2 \tag{6}$$

Bayesian estimator

Bayesian estimators are biased and tend to be regularizing. Can often achieve good performance with appropriate prior.

Markov chain Monte Carlo

Generic method to simulate from the posterior for any model.
Numerous existing tools: e.g. Stan.