Bayesian statistics

The Bayesian influence is using a likelihood function $L_n(\theta)$ that is weighted by prior knowledge.

The Bayesian approach is using sample data to update prior beliefs, forming posterior beliefs. To do this, we model the parameter as a random variable, **even though it is not.**

The **prior distribution** is the distribution of the parameter “random variable.” The **posterior distribution** is the distribution of the parameter “random variable” given sample data.

Conjugate prior

A prior distribution is the **conjugate** to the data model if the posterior model is in the same distribution family as the prior model. Having a more general prior and more specific likelihood model makes the prior more likely to be a conjugate prior. Some examples of conjugate models to data models are:

- Gamma prior with exponential data model
- Beta prior with Bernoulli data model
- Gaussian prior with Gaussian data model

Setup of Bayesian statistics problem

$\pi(\cdot)$: prior distribution. It could be uniform, exponential, Gaussian, etc.

$X_1, \ldots, X_n$: sample of n random variables

$L_n(\cdot | \theta)$: joint pdf of $X_1, \ldots, X_n$ conditionally on θ, where $\theta \sim \pi$. It is equal to the likelihood from the frequentist approach.

Applying Bayes' formula, we have:

$$\pi(\theta|X_1, \ldots, X_n) \propto L_n(X_1, \ldots, X_n|\theta) \pi(\theta)$$

$$\pi(\theta|X_1, \ldots, X_n) = \frac{L_n(X_1, \ldots, X_n|\theta) \pi(\theta)}{\int_\Theta L_n(X_1, \ldots, X_n|\theta) \pi(\theta)d\theta}$$

From this updated PDF, we can extract the new parameters (hyperparameters) of the distribution of the parameter.

Bernoulli experiment with Beta prior

Let $X_i \sim {\rm Ber}(\theta)$.

Select a Beta prior for the parameter θ. That is, $\pi(\theta) \sim {\rm Beta}(a, b)$

First, calculate the joint pdf, or the likelihood function.

$$L_n(X_1, \ldots, X_n | \theta) = p_n(X_1, \ldots, X_n | \theta) = \theta^{\sum_{i=1}^n X_i} (1-\theta)^{n-\sum_{i=1}^n X_i}$$

Then, update the distribution.

$$\pi(\theta|X_1, \ldots, X_n) \propto L_n(X_1, \ldots, X_n | \theta) \pi(\theta) $$ $$= \theta^{a-1}(1-\theta)^{b-1}\theta^{\sum_{i=1}^n X_i} (1-\theta)^{n-\sum_{i=1}^n X_i}$$ $$= \theta^{a+\sum_{i=1}^n X_i-1}(1-\theta)^{b+n-\sum_{i=1}^n X_i-1}$$

So the new parameters (for the Beta distribution describing the parameter as a random variable) are:

$$a' = a+\sum_{i=1}^n X_i$$ $$b' = b+n-\sum_{i=1}^n X_i$$

Noninformative prior

If we have no prior information about the parameter, we can choose a prior with constant pdf on Θ.

- If Θ is bounded, the distribution is uniform on Θ.
- If Θ is unbounded, the prior is an **improper prior.** Formally, $\pi(\theta) \equiv 1$.
    - In general, a prior is improper iff $\int \pi(\theta) d\theta = \infty$.
    - Bayes' formula still works.

Bayesian confidence region

A Bayesian confidence region with level α is a random subset $\mathcal{R}$ of Θ such that:

$$\mathbb{P}[\theta \in \mathcal{R} | X_1, \ldots, X_n] = 1 - \alpha$$

The randomness comes from the prior distribution.

Bayesian estimation

One Bayes estimator is the posterior mean:

$$\hat{\theta}^{(\pi)}=\int_\Theta \theta \pi(\theta | X_1, \ldots, X_n) d\theta$$

Another estimator is the point that maximizes the posterior distribution, called the MAP (maximum a posteriori):

$$\hat{\theta}^{\rm MAP} = {\rm argmax}_{\theta \in \Theta} \pi(\theta | X_1, \ldots, X_2) = L_n (X_1, \ldots, X_n | \theta) \pi(\theta)$$