Bayesian inference

sábado, 3 de outubro de 2020 21:10

2.2. Bayes Theorem

- Developed in the presence of observations x whose values are initially uncertain and described through $f(x|\theta)$
- The parameter θ serves as an index of the possible distributions for the observations
- The canonical situation is when we draw c.i.i.d. samples $\underset{\sim}{x} = (x_1, \dots, x_n)^T$ from $f(\underset{\sim}{x}|\theta)$

2.2.1. Prior, Posterior and predictive distribution

- Typically, the parameters $\underset{\sim}{\theta}$ are uncertain and the researcher may have an intuition about its value
- Its possible to incorporate those intuitions using a parametrization on θ
- Frequentist inference does not admit this information to be incorporated in the model because its not subject to empirical verification
- The Bayesian approach incorporates this information through the prior
- The basic ingredients of Bayesian inference are:

(a) The prior $p(\theta)$
(b) The likelihood $f(x|\theta)$, which can have constants (hyper parameters)

- The last ingredient of Bayesian inference is the distribution of the parameter of interest θ after seeing the data $\underset{\sim}{x}$:

$$p(\theta|\underset{\sim}{x}) = \frac{f(\underset{\sim}{x}|\theta)\, p(\theta)}{p(\underset{\sim}{x})} \quad , \quad p(\underset{\sim}{x}) = \int f(\underset{\sim}{x}|\theta)\, p(\theta)\, d\theta$$

$$\pi(\theta) \propto f(\underset{\sim}{x}|\theta)\, p(\theta)$$ posterior

Example 2.1

- Measurements about a physical constant μ with measurement errors e_i described by the $N(0, \sigma^2)$, where σ^2 is unknown.

$$x_i = \mu + e_i \quad , \quad i = 1, \dots, n$$

$$\Rightarrow f(\underset{\sim}{x}|\underset{\sim}{\theta}) = \prod_{i=1}^{n} f(x_i|\underset{\sim}{\theta}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}\right\} \quad , \quad \underset{\sim}{\theta} = (\mu, \sigma^2)$$

- Suppose for now that σ^2 is known $\Rightarrow \theta = \mu$. Furthermore, assume the following prior:

$$p(\mu) = f_N(\mu; \mu_0, \tau_0^2) = \frac{1}{\sqrt{2\pi\tau_0^2}} \exp\left\{\frac{1}{2\tau_0^2}(\mu-\mu_0)\right\}$$

$$l(\theta) = f(\underset{\sim}{x}|\theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\} \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\}$$

$$\Rightarrow \pi(\mu) \propto \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} \exp\left\{\frac{1}{2\tau_0}(\mu-\mu_0)^2\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\frac{(\bar{x}-\mu)^2}{\sigma^2} + \frac{(\mu-\mu_0)^2}{\tau_0^2}\right]\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\frac{(\mu-\mu_1)^2}{\tau_1^2}\right\} \quad , \quad \tau_1^{-2} = n\sigma^{-2} + \tau_0^{-2}$$

$$\mu_1 = \tau_1^2(n\sigma^{-2}\bar{x} + \tau^{-2}\mu_0)$$

$$= f_N(\mu; \mu_1, \tau_1^2)$$

- There is controversy among Bayesians about the use of non-informative distributions as priors
- One of the reasons for this is that some specifications of vague priors leads to improper distributions (do not integrate to one)
- One of the most commonly used vague prior is the Jeffreys prior, given by:

$$p(\theta) \propto |I(\theta)|^{\frac{1}{2}} \quad , \quad I(\theta) = E\left[-\frac{\partial^2 \log f(x|\theta)}{\partial\theta\,\partial\theta'} \,\middle|\, \theta\right]$$

which is the expected value of the Fisher information matrix about θ.

- Another important element of Bayesian inference is the marginal distribution of x with density

$$f(\underset{\sim}{x}) = \int f(\underset{\sim}{x}|\underset{\sim}{\theta})\, p(\underset{\sim}{\theta})\, d\underset{\sim}{\theta} = E[f(\underset{\sim}{x}|\underset{\sim}{\theta})]$$

- A similar derivation can be applied to the prediction of y given observations $\underset{\sim}{x}$

$$f(y|\underset{\sim}{x}) = \int f(y, \underset{\sim}{\theta}|\underset{\sim}{x})\, d\underset{\sim}{\theta} = \int f(y|\underset{\sim}{\theta})\, \underbrace{f(\underset{\sim}{x}|\underset{\sim}{\theta})\, p(\underset{\sim}{\theta})}_{\pi(\underset{\sim}{\theta})}\, d\underset{\sim}{\theta}$$

2.2.2. Summarising the information

- Once the posterior is available, one may summarize its information through a few elements: → most probable value for θ

(a) Location measures: Mean, mode, and median
(b) Dispersion measures: Variance, standard deviation, precision, and inter-quantile range

→ expected value for θ

- In multivariate spaces, marginal densities are useful for concentrating inference on a component of the parameter space
- One useful measure for these components are credible intervals

$$C \text{ is a } 100(1-\alpha)\% \text{ credible interval for } \theta \text{ s.t. } \int_C \pi(\theta)\, d\theta = 1-\alpha$$

- C is the minimal length interval that include points of higher posterior points, and are called highest posterior density (HPD) intervals
- For parameters in higher dimensions, this idea can be extended to credible regions