Bayesian Statistics. Decision Theory

Evgeny Burnaev

Skoltech, Moscow, Russia

Outline

1. The Bayesian Philosophy

Frequentist approach

- Probability refers to limiting relative frequencies
- Probabilities are objective properties of the real world
- Parameters are fixed, unknown constants $\Rightarrow$ no useful probability statements can be made about parameters
- Statistical procedures should have well-defined long run frequency properties
- E.g. a 95 percent confidence interval should contain value of a parameter with limiting frequency at least 95%

Bayesian inference

- Probability describes degree of belief, not limiting frequency
- $\Rightarrow$ we can make probability statements like: "the probability that Albert Einstein drank a cup of tea on August 1, 1948" is .35. This does not refer to any limiting frequency. It reflects person's strength of belief
- We can make probability statements about parameters, even though they are fixed constants
- We make inferences about a parameter θ by producing a probability distribution for θ
- Inferences, such as point estimates and interval estimates, may then be extracted from this distribution

Bayesian inference inherently embraces a subjective notion of probability (due to prior distributions)!

Bayesian Method

- We choose a probability density $f(\theta)$ — the prior distribution. It expresses our beliefs about θ prior to getting any data
- We select a statistical model $f(x|\theta)$ that reflects our beliefs about x given θ
- After observing data $x_1, \ldots, x_n$ we update our beliefs and calculate the posterior

$$f(\theta|x_1, \ldots, x_n)$$

Bayesian Method

- Assume that θ is discrete and we have a single discrete observation x
- Let us denote r.v. by bold letters, then

$$\mathbb{P}(\boldsymbol{\theta} = \theta | \mathbf{x} = x) = \frac{\mathbb{P}(\mathbf{x} = x, \boldsymbol{\theta} = \theta)}{\mathbb{P}(\mathbf{x} = x)}$$
$$= \frac{\mathbb{P}(\mathbf{x} = x | \boldsymbol{\theta} = \theta)\mathbb{P}(\boldsymbol{\theta} = \theta)}{\sum_\theta \mathbb{P}(\mathbf{x} = x | \boldsymbol{\theta} = \theta)\mathbb{P}(\boldsymbol{\theta} = \theta)}$$

- In case of density functions Bayes formula has the form

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int f(x|\theta)f(\theta)d\theta}$$

Bayesian Method

- In i.i.d. case we get that the likelihood

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = L_n(\theta)$$

- We denote by $\mathbf{x}^n = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and by $x^n = (x_1, \ldots, x_n)$, then

$$f(\theta|x^n) = \frac{f(x^n|\theta)f(\theta)}{\int f(x^n|\theta)f(\theta)d\theta} = \frac{L_n(\theta)f(\theta)}{c_n} \sim L_n(\theta)f(\theta),$$

where $c_n = \int L_n(\theta)f(\theta)d\theta$ is the normalizing constant

Bayesian Method

- We can get a point estimate of θ, e.g. by using the posterior mean
$$\overline{\theta}_n = \int \theta f(\theta|x^n) d\theta = \frac{\int \theta L_n(\theta) f(\theta) d\theta}{\int L_n(\theta) f(\theta) d\theta}$$
- We can also obtain a Bayesian interval estimate, i.e. find a and b, such that $\int_{-\infty}^{a} f(\theta|x^n) d\theta = \int_{b}^{\infty} f(\theta|x^n) d\theta = \alpha/2$. Denote $C = (a, b)$, then
$$\mathbb{P}(\theta \in C|x^n) = \int_a^b f(\theta|x^n) d\theta = 1 - \alpha,$$
i.e. C is a $1 - \alpha$ posterior interval