5. Bayesian Set Estimation

Definitions

As usual, our starting point is a random experiment with an underlying sample space and a probability measure $\mathbb{P}$. In the basic statistical model, we have an observable random variable $\boldsymbol{X}$ taking values in a set S. In general, $\boldsymbol{X}$ can have quite a complicated structure. For example, if the experiment is to sample n objects from a population and record various measurements of interest, then

$$\boldsymbol{X} = (X_1, X_2, ..., X_n)$$

where X_i is the vector of measurements for the i^{th} object. The most important special case occurs when $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ are independent and identically distributed. In this case, we have a random sample of size n from the common distribution.

Suppose also that the distribution of $\boldsymbol{X}$ depends on a parameter θ taking values in a parameter space Θ. The parameter may also be vector valued, in which case $\Theta \subseteq \mathbb{R}^k$ for some k and the parameter has the form $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_k)$

Recall that in Bayesian analysis, the unknown parameter θ is treated as a random variable. Specifically, suppose that the conditional probability density function of the data vector $\boldsymbol{X}$ given θ is denoted $g(\boldsymbol{x}|\theta)$. Moreover, the parameter θ is given a **prior distribution** with probability density function h. (The prior distribution is chosen to reflect our knowledge, if any of the parameter). The joint probability density function of the data vector and the parameter is

$$f(\boldsymbol{x}, \theta) = h(\theta)\, g(\boldsymbol{x}|\theta), \quad \boldsymbol{x} \in S, \theta \in \Theta$$

Next, the (unconditional) probability density function of $\boldsymbol{X}$ is the function g given by

$$g(\boldsymbol{x}) = \sum_{\theta \in \Theta} f(\boldsymbol{x}, \theta) = \sum_{\theta \in \Theta} h(\theta)\, g(\boldsymbol{x}|\theta), \quad \boldsymbol{x} \in S$$

if the parameter has a discrete distribution, or by

$$g(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x}, \theta) d\theta = \int_\Theta h(\theta)\, g(\boldsymbol{x}|\theta) d\theta, \quad \boldsymbol{x} \in S$$

if the parameter has a continuous distribution. Finally, by Bayes' theorem, the **posterior probability density function** of θ given $\boldsymbol{x}$ is

$$h(\theta|\boldsymbol{x}) = \frac{h(\theta)\, g(\boldsymbol{x}|\theta)}{g(\boldsymbol{x})}, \quad \theta \in \Theta, \boldsymbol{x} \in S$$

Now let $C(\boldsymbol{X})$ be a confidence set (that is, a subset of the parameter space that depends on the data variable $\boldsymbol{X}$. but no unknown parameters). One possible definition of a $1 - \alpha$ level **Bayesian confidence set** requires that

$$\mathbb{P}(\theta \in C(\boldsymbol{x}) | \boldsymbol{X} = \boldsymbol{x}) = 1 - \alpha$$

In this definition, only θ is random and thus the probability above is computed using the posterior probability density function $h(\theta|\boldsymbol{x})$. Another possible definition requires that

$$\mathbb{P}(\theta \in C(\boldsymbol{X})) = 1 - \alpha$$

In this definition, $\boldsymbol{X}$ and θ are both random, and so the probability above would be computed using the joint probability density function $f(\boldsymbol{x}, \theta)$. Whatever the philosophical arguments may be, the first definition is certainly the easier one from a computational viewpoint, and hence is the one most commonly used.

Let us compare the classical and Bayesian approaches. In the classical approach, the parameter is deterministic, but unknown. *Before* the data are collected, the confidence set (which is random) will contain the parameter with probability $1 - \alpha$. *After* the data are collected, the computed confidence set either contains the parameter or does not, and we will usually never know which. By contrast in a Bayesian confidence set, the random parameter θ falls in the computed, deterministic confidence set with probability $1 - \alpha$.

Applications

The Bernoulli Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample from the Bernoulli distribution with success parameter p. Thus, $X_i = 1$ if trial i resulted in success, and $X_i = 0$ if trial i resulted in failure. Moreover, suppose that p has a prior beta distribution with left parameter $a > 0$ and right parameter $b > 0$. Denote the number of successes by

$$Y = \sum_{i=1}^{n} X_i$$

Recall that for a given value of p, Y has the binomial distribution with parameters n and p.

1. Show that given $Y = y$,

   a. The posterior distribution of p is beta with left parameter $a + y$ and right parameter $b + (n - y)$
   b. A $1 - \alpha$ level Bayesian confidence interval for p is $(L(y), U(y))$ where $L(y)$ is the quantile of order $\frac{\alpha}{2}$ and $U(y)$ is the quantile of order $1 - \frac{\alpha}{2}$ for the beta distribution in (a).

2. Specifically, suppose that we have a coin with an unknown probability p of heads and that we give p the uniform prior. We then toss the coin 10 times, observing 7 heads. Compute the 90% Bayesian confidence interval for p.

The Poisson Distribution

Suppose that $\boldsymbol{X} = (X_1, X_2, ..., X_n)$ is a random sample of size n from the Poisson distribution with parameter μ. Moreover, suppose that μ has a prior gamma distribution with shape parameter $k > 0$ and scale parameter $b > 0$. Denote the sum of the sample values by

$$Y = \sum_{i=1}^{n} X_i$$

.

3. Show that given $Y = y$,

a. The posterior distribution of μ is gamma with shape parameter $k + y$ and scale parameter $\frac{b}{n\,b+1}$

b. A $1 - \alpha$ level Bayesian confidence interval for μ is $(L(x),\ U(x))$ where $L(x)$ is the quantile of order $\frac{\alpha}{2}$ and $U(x)$ is the quantile of order $1 - \frac{\alpha}{2}$ for the gamma distribution in (a).

4. Suppose that the number of defects in a manufactured item has the Poisson distribution with parameter μ and we give μ the exponential distribution with parameter 1. We sample 5 items and observe a total of 8 defects. Compute the 90% Bayesian confidence interval.