

8.5: Bayesian Set Estimation

Basic Theory

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 \mathbf{X} taking values in a set S . In general, \mathbf{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

$$\mathbf{X} = (X_1, X_2, \dots, X_n) \quad (8.5.1)$$

where X_i is the vector of measurements for the i th object.

Suppose also that the distribution of \mathbf{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 $k \in \mathbb{N}_+$ and the parameter has the form $\theta = (\theta_1, \theta_2, \dots, \theta_k)$.

The Bayesian Formulation

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 \mathbf{X} given $\theta \in \Theta$ is denoted $f(\mathbf{x} | \theta)$ for $\mathbf{x} \in S$. Moreover, the parameter θ is given a *prior distribution* with probability density function h on Θ . (The prior distribution is often subjective, and is chosen to reflect our knowledge, if any, of the parameter.) The joint probability density function of the data vector and the parameter is

$$(\mathbf{x}, \theta) \mapsto h(\theta)f(\mathbf{x} | \theta); \quad (\mathbf{x}, \theta) \in S \times \Theta \quad (8.5.2)$$

Next, the (unconditional) probability density function of \mathbf{X} is the function f given by

$$f(\mathbf{x}) = \sum_{\theta \in \Theta} h(\theta)f(\mathbf{x} | \theta), \quad \mathbf{x} \in S \quad (8.5.3)$$

if the parameter has a discrete distribution, or by

$$f(\mathbf{x}) = \int_{\Theta} h(\theta)f(\mathbf{x} | \theta) d\theta, \quad \mathbf{x} \in S \quad (8.5.4)$$

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

$$h(\theta | \mathbf{x}) = \frac{h(\theta)f(\mathbf{x} | \theta)}{f(\mathbf{x})}, \quad \theta \in \Theta \quad (8.5.5)$$

In some cases, we can recognize the posterior distribution from the functional form of $\theta \mapsto h(\theta)f(\mathbf{x} | \theta)$ without having to actually compute the normalizing constant $f(\mathbf{x})$, and thus reducing the computational burden significantly. In particular, this is often the case when we have a *conjugate parametric family* of distributions of θ . Recall that this means that when the prior distribution of θ belongs to the family, so does the posterior distribution given $\mathbf{x} \in S$.

Confidence Sets

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

$$\mathbb{P}[\theta \in C(\mathbf{x}) | \mathbf{X} = \mathbf{x}] = 1 - \alpha \quad (8.5.6)$$

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

$$\mathbb{P}[\theta \in C(\mathbf{X})] = 1 - \alpha \quad (8.5.7)$$

In this definition, \mathbf{X} and θ are both random, and so the probability above would be computed using the joint probability density function $(\mathbf{x}, \theta) \mapsto h(\theta)f(\mathbf{x} | \theta)$. 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 θ is deterministic, but unknown. *Before* the data are collected, the confidence set $C(\mathbf{X})$ (which is random by virtue of \mathbf{X}) will contain the parameter with probability $1 - \alpha$. *After* the data are collected, the computed confidence set $C(\mathbf{x})$ either contains θ 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 $C(\mathbf{x})$ with probability $1 - \alpha$.

Real Parameters

Suppose that θ is real valued, so that $\Theta \subseteq \mathbb{R}$. For $r \in (0, 1)$, we can compute the $1 - \alpha$ level Bayesian confidence interval as $[U_{(1-r)\alpha}(\mathbf{x}), U_{1-r\alpha}(\mathbf{x})]$ where $U_p(\mathbf{x})$ is the quantile of order p for the posterior distribution of θ given $\mathbf{X} = \mathbf{x}$. As in past sections, r is the fraction of α in the right tail of the posterior distribution and $1 - r$ is the fraction of α in the left tail of the posterior distribution. As usual, $r = \frac{1}{2}$ gives the symmetric, two-sided confidence interval; letting $r \rightarrow 0$ gives the confidence lower bound; and letting $r \rightarrow 1$ gives the confidence upper bound.

Random Samples

In terms of our data vector \mathbf{X} the most important special case arises when we have a basic variable X with values in a set R , and given θ , $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from X . That is, given θ , \mathbf{X} is a sequence of independent, identically distributed variables, each with the same distribution as X given θ . Thus $S = R^n$ and if X has conditional probability density function $g(x | \theta)$, then

$$f(\mathbf{x} | \theta) = g(x_1 | \theta)g(x_2 | \theta) \cdots g(x_n | \theta), \quad \mathbf{x} = (x_1, x_2, \dots, x_n) \in S \quad (8.5.8)$$

Applications

The Bernoulli Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the Bernoulli distribution with unknown success parameter $p \in (0, 1)$. In the usual language of reliability, $X_i = 1$ means success on trial i and $X_i = 0$ means failure on trial i . The distribution is named for Jacob Bernoulli. Recall that the Bernoulli distribution has probability density function (given p)

$$g(x | p) = p^x (1 - p)^{1-x}, \quad x \in \{0, 1\} \quad (8.5.9)$$

Note that the number of successes in the n trials is $Y = \sum_{i=1}^n X_i$. Given p , random variable Y has the binomial distribution with parameters n and p .

In our previous discussion of Bayesian estimation, we showed that the beta distribution is conjugate for p . Specifically, if the prior distribution of p is beta with left parameter $a > 0$ and right parameter $b > 0$, then the posterior distribution of p given \mathbf{X} is beta with left parameter $a + Y$ and right parameter $b + (n - Y)$; the left parameter is increased by the number of successes and the right parameter by the number of failure. It follows that a $1 - \alpha$ level Bayesian confidence interval for p is $[U_{\alpha/2}(y), U_{1-\alpha/2}(y)]$ where $U_r(y)$ is the quantile of order r for the posterior beta distribution. In the special case $a = b = 1$ the prior distribution is uniform on $(0, 1)$ and reflects a lack of previous knowledge about p .

Suppose that we have a coin with an unknown probability p of heads, and that we give p the uniform prior, reflecting our lack of knowledge about p . We then toss the coin 50 times, observing 30 heads.

1. Find the posterior distribution of p given the data.
2. Construct the 95% Bayesian confidence interval.
3. Construct the classical Wald confidence interval at the 95% level.

Answer

1. Beta with left parameter 31 and right parameter 21.
2. $[0.461, 0.724]$
3. $[0.464, 0.736]$

The Poisson Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the *Poisson distribution* with parameter $\lambda \in (0, \infty)$. Recall that the Poisson distribution is often used to model the number of “random points” in a region of time or space and is studied in more detail in the chapter on the Poisson Process. The distribution is named for the inimitable Simeon Poisson and given λ , has probability density function

$$g(x | \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x \in \mathbb{N} \quad (8.5.10)$$

As usual, we will denote the sum of the sample values by $Y = \sum_{i=1}^n X_i$. Given λ , random variable Y also has a Poisson distribution, but with parameter $n\lambda$.

In our previous discussion of Bayesian estimation, we showed that the gamma distribution is conjugate for λ . Specifically, if the prior distribution of λ is gamma with shape parameter $k > 0$ and rate parameter $r > 0$ (so that the scale parameter is $1/r$), then the posterior distribution of λ given \mathbf{X} is gamma with shape parameter $k + Y$ and rate parameter $r + n$. It follows that a $1 - \alpha$ level Bayesian confidence interval for λ is $[U_{\alpha/2}(y), U_{1-\alpha/2}(y)]$ where $U_p(y)$ is the quantile of order p for the posterior gamma distribution.

Consider the alpha emissions data, which we believe come from a Poisson distribution with unknown parameter λ . Suppose that *a priori*, we believe that λ is about 5, so we give λ a prior gamma distribution with shape parameter 5 and rate parameter 1. (Thus the mean is 5 and the standard deviation $\sqrt{5} = 2.236$.)

1. Find the posterior distribution of λ given the data.
2. Construct the 95% Bayesian confidence interval.
3. Construct the classical t confidence interval at the 95% level.

Answer

1. Gamma with shape parameter 10104 and rate parameter 1208.
2. (8.202, 8.528)
3. (8.324, 8.410)

The Normal Distribution

Suppose that $\mathbf{x} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the normal distribution with unknown mean $\mu \in \mathbb{R}$ and known variance $\sigma^2 \in (0, \infty)$. Of course, the normal distribution plays an especially important role in statistics, in part because of the central limit theorem. The normal distribution is widely used to model physical quantities subject to numerous small, random errors. Recall that the normal probability density function (given the parameters) is

$$g(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\left(\frac{x - \mu}{\sigma} \right)^2 \right], \quad x \in \mathbb{R} \quad (8.5.11)$$

We denote the sum of the sample values by $Y = \sum_{i=1}^n X_i$. Recall that Y also has a normal distribution (given μ and σ), but with mean $n\mu$ and variance $n\sigma^2$.

In our previous discussion of Bayesian estimation, we showed that the normal distribution is conjugate for μ (with σ known). Specifically, if the prior distribution of μ is normal with mean $a \in \mathbb{R}$ and standard deviation $b \in (0, \infty)$, then the posterior distribution of μ given \mathbf{X} is also normal, with

$$\mathbb{E}(\mu | \mathbf{X}) = \frac{Yb^2 + a\sigma^2}{\sigma^2 + nb^2}, \quad \text{var}(\mu | \mathbf{X}) = \frac{\sigma^2 b^2}{\sigma^2 + nb^2} \quad (8.5.12)$$

It follows that a $1 - \alpha$ level Bayesian confidence interval for μ is $[U_{\alpha/2}(y), U_{1-\alpha/2}(y)]$ where $U_p(y)$ is the quantile of order p for the posterior normal distribution. An interesting special case is when $b = \sigma$, so that the standard deviation of the prior distribution of μ is the same as the standard deviation of the sampling distribution. In this case, the posterior mean is $(Y + a)/(n + 1)$ and the posterior variance is $\sigma^2/(n + 1)$.

The length of a certain machined part is supposed to be 10 centimeters but due to imperfections in the manufacturing process, the actual length is a normally distributed with mean μ and variance σ^2 . The variance is due to inherent factors in the process, which remain fairly stable over time. From historical data, it is known that $\sigma = 0.3$. On the other hand, μ may be set by adjusting various parameters in the process and hence may change to an unknown value fairly frequently. Thus, suppose that we give μ with a prior normal distribution with mean 10 and standard deviation 0.03. A sample of 100 parts has mean 10.2.

1. Find the posterior distribution of μ given the data.
2. Construct the 95% Bayesian confidence interval.
3. Construct the classical z confidence interval at the 95% level.

Answer

1. Normal with mean 10.198 and standard deviation 0.0299.
2. (10.14, 10.26)
3. (10.14, 10.26)

The Beta Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the beta distribution with unknown left shape parameter $a \in (0, \infty)$ and right shape parameter $b = 1$. The beta distribution is widely used to model random proportions and probabilities and other variables that take values in bounded intervals. Recall that the probability density function (given a) is

$$g(x | a) = ax^{a-1}, \quad x \in (0, 1) \quad (8.5.13)$$

We denote the product of the sample values by $W = X_1 X_2 \cdots X_n$.

In our previous discussion of Bayesian estimation, we showed that the gamma distribution is conjugate for a . Specifically, if the prior distribution of a is gamma with shape parameter $k > 0$ and rate parameter $r > 0$, then the posterior distribution of a given \mathbf{X} is also gamma, with shape parameter $k + n$ and rate parameter $r + \ln(W)$. It follows that a $1 - \alpha$ level Bayesian confidence interval for a is $[U_{\alpha/2}(w), U_{1-\alpha/2}(w)]$ where $U_p(w)$ is the quantile of order p for the posterior gamma distribution. In the special case that $k = 1$, the prior distribution of a is exponential with rate parameter r .

Suppose that the resistance of an electrical component (in Ohms) has the beta distribution with unknown left parameter a and right parameter $b = 1$. We believe that a may be about 10, so we give a the prior gamma distribution with shape parameter 10 and rate parameter 1. We sample 20 components and observe the data

0.98, 0.93, 0.99, 0.89, 0.79, 0.99, 0.92, 0.97, 0.88, 0.97, 0.86, 0.84, 0.96, 0.97, 0.92, 0.90, 0.98, 0.96, 0.96, 1.00. (8.5.14)

1. Find the posterior distribution of a .
2. Construct the 95% Bayesian confidence interval for a .

Answer

1. Gamma with shape parameter 30 and rate parameter 2.424.
2. (8.349, 17.180)

The Pareto Distribution

Suppose that $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random sample of size n from the Pareto distribution with shape parameter $a \in (0, \infty)$ and scale parameter $b = 1$. The Pareto distribution is used to model certain financial variables and other variables with heavy-tailed distributions, and is named for Vilfredo Pareto. Recall that the probability density function (given a) is

$$g(x | a) = \frac{a}{x^{a+1}}, \quad x \in [1, \infty) \quad (8.5.15)$$

We denote the product of the sample values by $W = X_1 X_2 \cdots X_n$.

In our previous discussion of Bayesian estimation, we showed that the gamma distribution is conjugate for a . Specifically, if the prior distribution of a is gamma with shape parameter $k > 0$ and rate parameter $r > 0$, then the posterior distribution of a given \mathbf{X} is also gamma, with shape parameter $k + n$ and rate parameter $r + \ln(W)$. It follows that a $1 - \alpha$ level Bayesian confidence interval for a is $[U_{\alpha/2}(w), U_{1-\alpha/2}(w)]$ where $U_p(w)$ is the quantile of order p for the posterior gamma distribution. In the special case that $k = 1$, the prior distribution of a is exponential with rate parameter r .

Suppose that a financial variable has the Pareto distribution with unknown shape parameter a and scale parameter $b = 1$. We believe that a may be about 4, so we give a the prior gamma distribution with shape parameter 4 and rate parameter 1. A random sample of size 20 from the variable gives the data

1.09, 1.13, 2.00, 1.43, 1.26, 1.00, 1.36, 1.03, 1.46, 1.18, 2.16, 1.16, 1.22, 1.06, 1.28, 1.23, 1.11, 1.03, 1.04, 1.05. (8.5.16)

1. Find the posterior distribution of a .
2. Construct the 95% Bayesian confidence interval for a .

Answer

1. Gamma with shape parameter 24 and rate parameter 5.223.
2. (2.944, 6.608)

This page titled [8.5: Bayesian Set Estimation](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist](#) ([Random Services](#)) via [source content](#) that was edited to the style and standards of the LibreTexts platform.