Classical and Bayesian inference

AMS 132

January 27, 2018

Nature of an Estimation Problem

- How do we choose a single number as an estimate of θ, the mean of the distribution of normally distributed r.v.s.

Definition (Estimator and Estimate)

Let $X_1, \ldots, X_n$ be r.v.s whose joint distribution is indexed by a parameter θ taking values in a subset Ω of the real line. An *estimator* of the parameter θ is a real-valued function $\delta(X_1, \ldots, X_n)$. If $X_1 = x_1, \ldots, X_n = x_n$ are observed, then $\delta(x_1, \ldots, x_n)$ is called the *estimate* of θ.

- It is not required the value of $\delta(X_1, \ldots, X_n)$ to lie in Ω. If not, the experimenter will need to decide in the specific problem whether that seems appropriate or not.
- An estimator is a random variable and has a probability distribution.
- An estimate is a specific value of the estimator, determined by the specific observed values.
- We can use vector notation and write $\delta(\boldsymbol{X})$ and $\delta(\boldsymbol{x})$.

Loss Function

- A good estimator is one for which it is highly probable that the error $\delta(\boldsymbol{X}) - \theta$ will be close to 0.
- We shall assume that for each $\theta \in \Omega$ and each possible estimate a, there is a number $L(\theta, a)$ that measures the loss or cost to the statistician when the true value of the parameter is θ and the estimate is a.

Definition (Loss Function)

A *loss function* is a real-valued function of two variables, $L(\theta, a)$, where $\theta \in \Omega$ and a is a real number. The interpretation is that the statistician loses $L(\theta, a)$ if the parameter equals θ and the estimate equals a.

- Before observing any values in the random sample, the expected loss of choosing a as an estimate is given by

$$E[L(\theta, a)] = \int_{\Omega} L(\theta, a)\xi(\theta)d\theta.$$

Definition of a Bayes Estimator

- Suppose that we observe $\boldsymbol{x}$ of the random vector $\boldsymbol{X}$. The expected loss of choosing a as an estimate will be

$$E[L(\theta, a) \mid \boldsymbol{x}] = \int_{\Omega} L(\theta, a)\xi(\theta \mid \boldsymbol{x})d\theta.$$

Definition (Bayes Estimator and Estimate)

Let $L(\theta, a)$ be a loss function. For each possible value $\boldsymbol{x}$ of $\boldsymbol{X}$, let $\delta^{\star}(\boldsymbol{x})$ be a value of a such that $E[L(\theta, a) \mid \boldsymbol{x}]$ is minimized. Then $\delta^{\star}$ is called a *Bayes estimator* of θ. Once $\boldsymbol{X} = \boldsymbol{x}$ is observed, $\delta^{\star}$ is called a *Bayes estimate* of θ.

- Other way to describe Bayes estimator $\delta^{\star}$ is to note that, for each possible value $\boldsymbol{x}$ of $\boldsymbol{X}$,

$$E[L(\theta, \delta^{\star}(\boldsymbol{x})) \mid \boldsymbol{x}] = \min_{Alll\ a} E[L(\theta, a) \mid \boldsymbol{x}].$$

- Bayes estimator depends on $\xi(\theta)$ and $L(\theta, a)$.
- Bayes estimator might not exist!

Different Loss Functions

Definition (Square Error Loss)

The loss function $L(\theta, a) = (\theta - a)^2$ is called *square error loss*.

- When the square error loss function is used, the Bayes estimate $\delta^\star(\boldsymbol{x})$ for each observed value $\boldsymbol{x}$ will be the value of a for which $E[(\theta - a)^2 \mid \boldsymbol{x}]$ is minimum.

Corollary

Let θ be a real-valued parameter. Suppose that the squared error loss function $L(\theta, a) = (\theta - a)^2$ is used and that the posterior mean of θ, $E(\theta \mid \boldsymbol{X})$, is finite. Then, a Bayes estimator of θ is $\delta^\star(X) = E(\theta \mid \boldsymbol{X})$.

Example

a) If $X_i \mid \theta \overset{i.i.d.}{\sim} Bernoulli(\theta)$, and $\theta \sim Beta(a, b)$, show that under square error loss $\delta^\star(X) = (a + \sum_{i=1}^n X_i)/(a + b + n)$.

Different Loss Functions

Definition (absolute Error Loss)

The loss function $L(\theta, a) = \mid \theta - a \mid$ is called *absolute error loss*.

- When the absolute error loss function is used, the Bayes estimate $\delta^\star(\boldsymbol{x})$ for each observed value $\boldsymbol{x}$ will be the value of a for which $E[\mid \theta - a \mid \mid \boldsymbol{x}]$ is minimum.

Corollary

Let θ be a real-valued parameter. When the absolute error loss function $L(\theta, a) = \mid \theta - a \mid$ is used, a Bayes estimator of θ is $\delta^\star(X)$ equal to the median of the posterior distribution of θ.

Example

a) If $X_i \mid \theta \overset{i.i.d.}{\sim} Bernoulli(\theta)$, and $\theta \sim Beta(a, b)$, then $\delta^\star(X)$ has not a simple expression, numerical approximation is needed.

- Many other loss functions can be considered!

The Bayes Estimate for Large samples

Example

Consider a large shipment of iPhone has arrive, and the proportion of defective iPhone, θ, is unknown. A *Beta*(α, β) prior is assumed and $n = 100$ randomly selected items are inspected from which 10 were defective. Suppose that θ must be estimated, and that the square error loss is used.
Find the estimate of θ when $\alpha = \beta = 1$ and when $\alpha = 1$ and $\beta = 2$.

The Bayes Estimate for Large samples

Definition

A sequence of estimators that converges in probability to the unknown value of the parameter being estimated, as $n \to \infty$, is called a *consistent sequence of estimators*.

- Recall that, if $X_1, \ldots, X_n$ is a random sample such that $E(X_i) = \theta$, then the from the law of large numbers (Section 6.2), it follows that $\overline{X}_n$ converges in probability to θ as $n \to \infty$.
- Interpretation: when large numbers of observations are taken, there is high probability that the Bayes estimator will be very close to the unknown value of θ.

Example (Cont.)

Show that Bayes estimator forms a a consistent sequence of estimators in the above problem.

More General Parameters and Estimators

Definition

Let $X_1, \ldots, X_n$ be r.v.s whose joint distribution is indexed by a parameter θ taking values in a subset Ω of k-dimensional space. Let h be a function from Ω into d-dimensional space. Define $\Psi = h(\theta)$. An estimator of Ψ is a function $\delta(X_1, \ldots, X_n)$ that takes values in d- dimensional space. If $X_1 = x_1, \ldots, X_n = x_n$ are observed, then $\delta(x_1, \ldots, x_n)$ is called the estimate of Ψ.

Example (Lifetime of a component)

Consider $X_i \mid \theta \overset{i.i.d.}{\sim} exp(\theta)$ and $\theta \mid a, b \sim Gama(a, b)$. Find Bayes estimate of the mean lifetime of a component under square error loss.