Bayesian Estimation: loss, risk, Bayes risk, and Bayes estimator

Notes by Guoqi Qian
(edited by Heejung Shim)

Loss and risk

The Mean Squared Error (MSE) of an estimator T of $\tau(\theta)$, namely $E[T - \tau(\theta)]^2$, measures the *expected squared error loss*. One can consider other types of loss function.

- **Loss function** $L(t;\theta)$ is a real-valued function such that $L(t;\theta) \geq 0$ for all t and $L(t;\theta) = 0$ when $t = \tau(\theta)$.
- **Risk function** $R_T(\theta)$ is the expectation of the loss function w.r.t. the data, i.e.

$$R_T(\theta) = E[L(T;\theta)] = \int L(t(\mathbf{x});\theta) f(\mathbf{x}|\theta)\, d\mathbf{x}$$

where the estimator $T = t(\mathbf{X})$ is a function of the data $\mathbf{X} = (X_1, \cdots, X_n)^T$, and $f(\mathbf{x}|\theta)$ is the joint pdf (or pmf) of $\mathbf{X}$.

Minimization of loss and risk?

- Given R_T, the best estimator would be the one which minimises the risk function for the true value of θ, except that we don't know θ.
  We would settle for an estimator T^* which minimizes $R_T(\theta)$ for all possible values of θ.
  Unfortunately such estimators rarely exist.
- An estimator T_1 **dominates** another estimator T_2 iff $R_{T_1}(\theta) \leq R_{T_2}(\theta)$ for all $\theta \in \Theta$, and $R_{T_1}(\theta) < R_{T_2}(\theta)$ for at least some $\theta \in \Theta$. T is an **admissible estimator** iff no other estimators dominate it.
  It is usually possible to identify a class of admissible estimators

Bayesian approach

Regard the parameter θ as a random variable having a pdf $p(\theta)$, where $p(\theta)$ is called the **prior distribution** or **prior density**.

- Then **Bayes risk** is defined to be

$$E_\theta[R_T(\theta)] = \int_\Theta R_T(\theta) p(\theta) d\theta,$$

  which is just an expected risk w.r.t. the prior distribution.

- **Bayes estimator** is defined to be the estimator T^* which minimizes the Bayes risk:

$$E_\theta[R_{T^*}(\theta)] \leq E_\theta[R_T(\theta)] \quad \text{for every estimator } T \text{ of } \tau(\theta).$$

  Namely, $T^* = \arg\min_T E_\theta[R_T(\theta)]$.

How to find the Bayes estimator?

$$
\begin{aligned}
E_\theta[R_T(\theta)] &= \int_\Theta R_T(\theta)p(\theta)\,d\theta = \int_\Theta \left[\int_{\mathbf{x}} L(t(\mathbf{x});\theta) f(\mathbf{x}|\theta)\,d\mathbf{x}\right] p(\theta)\,d\theta \\
&= \int_\Theta \int_{\mathbf{x}} L(t(\mathbf{x});\theta) f(\mathbf{x}|\theta) p(\theta)\,d\mathbf{x}\,d\theta \\
&= \int_{\mathbf{x}} \left[\int_\Theta L(t(\mathbf{x});\theta) \frac{f(\mathbf{x}|\theta)p(\theta)}{\int_\Theta f(\mathbf{x}|\theta)p(\theta)\,d\theta}\,d\theta\right] \left[\int_\Theta f(\mathbf{x}|\theta)p(\theta)\,d\theta\right] d\mathbf{x} \\
&= \int_{\mathbf{x}} \left[\int_\Theta L(t(\mathbf{x});\theta) p(\theta|\mathbf{x})\,d\theta\right] f(\mathbf{x})\,d\mathbf{x} \\
&= \int_{\mathbf{x}} E_{\theta|\mathbf{x}}[L(T;\theta)] f(\mathbf{x})\,d\mathbf{x} = E_{\mathbf{x}}\left(E_{\theta|\mathbf{x}}[L(T;\theta)]\right)
\end{aligned}
$$

Thus if an estimator minimizes $E_{\theta|\mathbf{x}}[L(T;\theta)]$ for each $\mathbf{x}$, it must also minimize the Bayes risk $E_\theta[R_T(\theta)]$.

Bayes estimator under squared loss

The Bayes estimator is that which minimizes $E_{\theta|\mathbf{x}}[L(T;\theta)]$.

Theorem (Bayes estimator under squared loss)

Suppose we choose to use the squared loss function $L(t;\theta) = [t - \tau(\theta)]^2$, *then*

$$T^* = E_{\theta|\mathbf{x}}[\tau(\theta)] = \int_\Theta \tau(\theta) p(\theta|\mathbf{x}) d\theta$$

is the Bayes estimator of $\tau(\theta)$.

Proof:

Remarks

1. $f(\mathbf{x}) = \int_\Theta f(\mathbf{x}|\theta)p(\theta)\,d\theta$ is the marginal pdf of $\mathbf{X}$.
2. The conditional pdf $p(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)p(\theta)}{\int_\Theta f(\mathbf{x}|\theta)p(\theta)\,d\theta}$ is called the **posterior pdf** of θ.
3. Using squared loss, the Bayes estimator T^* is the posterior mean of $\tau(\theta)$.

Bayes estimator under absolute loss

Theorem (Bayes estimator under absolute loss)

Suppose we choose to use the absolute loss function $L(t;\theta) = |t - \tau(\theta)|$, then the Bayes estimator of $\tau(\theta)$ is the median of the posterior distribution.

Proof:

Remarks

1. The idea involved in Bayes estimation is very appealing. Without any information or with only prior information about θ, we would estimate $\tau(\theta)$ by its prior mean. Once the data are observed, new information about θ is available, we then would estimate $\tau(\theta)$ by its posterior mean.
2. The posterior mean may be analytically intractable if the posterior pdf is mathematically complicated. This difficulty may be overcome by using numerical techniques to approximate the posterior mean.
3. The Bayesian approach also gives a natural way of producing confidence intervals (called credible intervals) .