Bayesian intrinsic point estimation

Teng Zhang    Fan Wu

Department of Statistics
North Carolina State University

December 5, 2012
Bayes estimator

In Bayesian point estimation, loss function $L\{\tilde{\theta}, (\theta, \lambda)\}$ denotes the loss incurred when the parameter is $(\theta, \lambda)$ and the action taken is $\tilde{\theta}$. Given data $\mathbf{x}$, the Bayes estimator $\theta^*(\mathbf{x})$ is the value in $\Theta$ that minimizes the posterior expected loss $L\{\tilde{\theta}|\mathbf{x}\}$, i.e.

$$
\theta^*(\mathbf{x}) = \arg \min_{\tilde{\theta} \in \Theta} L\{\tilde{\theta}|\mathbf{x}\}
$$

$$
= \arg \min_{\tilde{\theta} \in \Theta} \int_{\Lambda} \int_{\Theta} L\{\tilde{\theta}, (\theta, \lambda)\} f(\theta, \lambda|\mathbf{x}) d\theta d\lambda.
$$
Motivation of intrinsic estimation

The loss function has various kinds of form, for example, squared loss,

\[ L\{\tilde{\theta}, \theta\} = (\tilde{\theta} - \theta)^2. \]

Under squared loss, the Bayes estimator is not invariant under one-to-one transformations of the data or the parameter space. However, intrinsic loss functions, which shifts attention from the discrepancy between the estimate \(\tilde{\theta}\) and the true value \(\theta\), to the more relevant discrepancy between the statistical models they label.
Kullback-Leibler divergence

Suppose $p$ and $q$ are two probability distribution functions. The Kullback-Leibler divergence [4] of $p$ and $q$ is defined as

$$\kappa(p|q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx.$$  

The Kullback-Leibler divergence of two probability distributions $p$ and $q$ is non-negative, equal to zero if and only if $p(x) = q(x)$ almost everywhere. But it is not symmetric.
Suppose $p$ is the probability distribution function of $X$, the intrinsic discrepancy [2] [3] between two models $p(x|\theta_1)$ and $p(x|\theta_2)$ for data $x \in \chi$ is defined by

$$\delta\{\theta_1, \theta_2\} = \min\{\kappa(\theta_1|\theta_2), \kappa(\theta_2|\theta_1)\},$$

where

$$\kappa(\theta_i|\theta_j) = \int_{\chi} p(x|\theta_j) \log \frac{p(x|\theta_j)}{p(x|\theta_i)} \, dx.$$

The intrinsic discrepancy is symmetric.
Intrinsic loss function

Suppose $p_x(\cdot | \theta, \lambda)$ is the probability distribution function of $X$, then the intrinsic loss function [1] of parameter $\theta$ is

$$L\{\tilde{\theta}, (\theta, \lambda)\} = \inf_{\tilde{\lambda} \in \Lambda} \delta\{ (\tilde{\theta}, \tilde{\lambda}), (\theta, \lambda) \},$$

where the actual parameter values are $(\theta, \lambda)$. 
Computation of the intrinsic discrepancy loss

Let $\mathcal{F}$ be a parametric family of probability distributions

$$\mathcal{F} = \{ p(x|\theta, \lambda), \theta \in \Theta, \lambda \in \Lambda, x \in \chi(\theta, \lambda) \},$$

with convex support $\chi(\theta, \lambda)$ for all $\theta$ and $\lambda$. Then, [1]

$$L\{\tilde{\theta}, (\theta, \lambda)\} = \inf_{\tilde{\lambda} \in \Lambda} \min \{ \kappa\{\tilde{\theta}, \tilde{\lambda}|\theta, \lambda\}, \kappa\{\theta, \lambda|\tilde{\theta}, \tilde{\lambda}\}\}$$

$$= \min \{ \inf_{\tilde{\lambda} \in \Lambda} \kappa\{\tilde{\theta}, \tilde{\lambda}|\theta, \lambda\}, \inf_{\tilde{\lambda} \in \Lambda} \kappa\{\theta, \lambda|\tilde{\theta}, \tilde{\lambda}\} \}. $$
Bayesian intrinsic estimator

The intrinsic discrepancy loss is given by

\[ d(\tilde{\theta}|x) = \int_{\Theta} \int_{\Omega} L\{\tilde{\theta}, (\theta, \lambda)\} f(\theta, \lambda|x) d\theta d\lambda, \]

and the intrinsic estimator of \( \theta \) is the solution which minimizes \( d(\tilde{\theta}|x) \).
Intrinsic loss functions

Suppose $x_1, \cdots, x_n$ are i.i.d from $N(\mu, \sigma^2)$. By [1], the loss function of mean $\mu$ and variance $\sigma^2$ are

$$L(\tilde{\mu}, (\mu, \sigma^2)) = \frac{n}{2} \log(1 + \frac{(\tilde{\mu} - \mu)^2}{\sigma^2}),$$

$$L(\tilde{\sigma}^2, (\mu, \sigma^2)) = \left\{ \begin{array}{ll}
\frac{n}{2} g(\theta) \\
\frac{n}{2} g(1/\theta)
\end{array} \right.,$$

where $g(t) = (t - 1) - \log(t)$, and $\theta = \tilde{\sigma}^2 / \sigma^2$. 
Prior selected

We use Jeffrey’s prior of $\mu$ and $\sigma^2$, i.e.

$$f(\mu) \propto 1,$$

$$f(\sigma^2) \propto \frac{1}{\sigma^2}.$$

Thus, we can get the posterior sample of $\mu$ and $\sigma^2$ by Gibbs sampling and Metropolis-Hastings. For each fixed $\tilde{\mu}$ and $\tilde{\sigma}^2$, we can get $E[L(\tilde{\mu}, (\mu, \sigma^2))]$ and $E[L(\tilde{\sigma}^2, (\mu, \sigma^2))]$ by MCMC.
Simulation process

Here, we choose four different pair parameters: \( \mu = 0, 1, \sigma^2 = 1, 4 \). For each pair, our random sample will be \( n = 50 \). We calculate the three estimators through Gibbs sampling \( (N = 5000, \text{burn in number is 1000}) \). We repeat this process \( N_{\text{rep}} = 100 \) times, calculating the estimated Bias, variance and MSE.

The result of simulation study is given as following,
\( \mu = 0, \sigma^2 = 1 \)

<table>
<thead>
<tr>
<th></th>
<th>Intrinsic</th>
<th>Bayesian (Squared loss)</th>
<th>UMVUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias(( \mu ))</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0016</td>
</tr>
<tr>
<td>Var(( \mu ))</td>
<td>0.0243</td>
<td>0.0244</td>
<td>0.0239</td>
</tr>
<tr>
<td>MSE(( \mu ))</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>Bias(( \sigma^2 ))</td>
<td>0.0091</td>
<td>0.0317</td>
<td>-0.0108</td>
</tr>
<tr>
<td>Var(( \sigma^2 ))</td>
<td>0.0474</td>
<td>0.0494</td>
<td>0.0451</td>
</tr>
<tr>
<td>MSE(( \sigma^2 ))</td>
<td>0.0083</td>
<td>0.1008</td>
<td>0.0116</td>
</tr>
</tbody>
</table>
\[ \mu = 1, \sigma^2 = 1 \]

<table>
<thead>
<tr>
<th></th>
<th>Intrinsic</th>
<th>Bayesian (Squared loss)</th>
<th>UMVUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias(\mu)</td>
<td>0.0021</td>
<td>0.0018</td>
<td>0.0016</td>
</tr>
<tr>
<td>Var(\mu)</td>
<td>0.0243</td>
<td>0.0244</td>
<td>0.0239</td>
</tr>
<tr>
<td>MSE(\mu)</td>
<td>0.0004</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>Bias(\sigma^2)</td>
<td>0.0091</td>
<td>0.0317</td>
<td>-0.0108</td>
</tr>
<tr>
<td>Var(\sigma^2)</td>
<td>0.0474</td>
<td>0.0494</td>
<td>0.0451</td>
</tr>
<tr>
<td>MSE(\sigma^2)</td>
<td>0.0083</td>
<td>0.1008</td>
<td>0.0116</td>
</tr>
</tbody>
</table>
\[ \mu = 0, \quad \sigma^2 = 4 \]

<table>
<thead>
<tr>
<th></th>
<th>Intrinsic</th>
<th>Bayesian (Squared loss)</th>
<th>UMVUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias((\mu))</td>
<td>0.0045</td>
<td>0.0045</td>
<td>0.0033</td>
</tr>
<tr>
<td>Var((\mu))</td>
<td>0.0967</td>
<td>0.0963</td>
<td>0.0951</td>
</tr>
<tr>
<td>MSE((\mu))</td>
<td>0.0020</td>
<td>0.0021</td>
<td>0.0011</td>
</tr>
<tr>
<td>Bias((\sigma^2))</td>
<td>0.0385</td>
<td>0.1290</td>
<td>-0.0431</td>
</tr>
<tr>
<td>Var((\sigma^2))</td>
<td>0.7557</td>
<td>0.7899</td>
<td>0.7217</td>
</tr>
<tr>
<td>MSE((\sigma^2))</td>
<td>0.1482</td>
<td>1.6633</td>
<td>0.1857</td>
</tr>
</tbody>
</table>
\[ \mu = 1, \sigma^2 = 4 \]

<table>
<thead>
<tr>
<th></th>
<th>Intrinsic</th>
<th>Bayesian (Squared loss)</th>
<th>UMVUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bias ((\mu))</td>
<td>0.0045</td>
<td>0.0045</td>
<td>0.0033</td>
</tr>
<tr>
<td>Var ((\mu))</td>
<td>0.0967</td>
<td>0.0966</td>
<td>0.0955</td>
</tr>
<tr>
<td>MSE ((\mu))</td>
<td>0.0020</td>
<td>0.0021</td>
<td>0.0011</td>
</tr>
<tr>
<td>Bias ((\sigma^2))</td>
<td>0.0385</td>
<td>0.1290</td>
<td>-0.0431</td>
</tr>
<tr>
<td>Var ((\sigma^2))</td>
<td>0.7557</td>
<td>0.7899</td>
<td>0.7217</td>
</tr>
<tr>
<td>MSE ((\sigma^2))</td>
<td>0.1482</td>
<td>1.6633</td>
<td>0.1857</td>
</tr>
</tbody>
</table>
Analysis of results

- The performance of Bias, variance, MSE does not depend on $\mu$;
- The performance of three estimators of $\mu$ are similar;
- Bayesian estimator with squared loss is not good for estimating $\sigma^2$;
- Intrinsic estimator of $\sigma^2$ has smaller bias and MSE, but larger variance than UMVUE.
Pros and Cons

- **Advantages:** Invariant under transformation; good performance on Bias, variance and MSE.
- **Disadvantages:** Computation issues; low efficiency.
J.M. Bernardo. 
Objective bayesian point and region estimation in location-scale models. 

JM Bernardo and M. Juárez. 
Intrinsic estimation. 

J.M. Bernardo and R. Rueda. 
Bayesian hypothesis testing: A reference approach.