1B. Bayesian Estimation

Notes by Guoqi Qian
(eds by Owen Jones)

If a man will begin with certainties, he shall end in doubts; but if he will be content to begin with doubts, he shall end in certainties.

Francis Bacon, The Advancement Of Learning
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.

$$R_T(\theta) = E[L(T; \theta)] = \int L(t(x); \theta)f(x|\theta)dx$$

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

1B. Bayesian Estimation
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 $\theta$, except that we don’t know $\theta$. We would settle for an estimator $T^*$ which minimizes $R_T(\theta)$ for all possible values of $\theta$. 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 \( \theta \) 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, \( T^* = \arg\min_T E_\theta[R_T(\theta)] \).
Mini-max approach

An alternative approach to choosing a “best” estimator w.r.t. $R_T$ is to take the *minimax* estimator.

- An estimator $T_1$ is a **minimax estimator** of $\tau(\theta)$ if

$$\max_{\theta} R_{T_1}(\theta) \leq \max_{\theta} R_T(\theta) \quad \text{for every estimator } T \text{ of } \tau(\theta).$$

Namely, $T_1 = \arg\min_T \{ \max_{\theta} R_T(\theta) \}$.

- The minimax approach is conservative in general. Not much is known about its performance and we won’t be considering it any further. It is more commonly seen in computer science than statistics.
How to find the Bayes estimator?

\[
E_\theta[R_T(\theta)] = \int_\Theta R_T(\theta) p(\theta) d\theta = \int_\Theta \left[ \int_x L(t(x); \theta) f(x|\theta) dx \right] p(\theta) d\theta \\
= \int_\Theta \int_x L(t(x); \theta) f(x|\theta) p(\theta) dx d\theta \\
= \int_x \left[ \int_\Theta L(t(x); \theta) \frac{f(x|\theta)p(\theta)}{\int_\Theta f(x|\theta)p(\theta) d\theta} d\theta \right] \left[ \int_\Theta f(x|\theta)p(\theta) d\theta \right] dx \\
= \int_x \left[ \int_\Theta L(t(x); \theta) p(\theta|x) d\theta \right] f(x) dx \\
= \int_x E_\theta[L(T; \theta)|x] f(x) dx = E_x \left( E_\theta[L(T; \theta)|x] \right)
\]

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

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

**Theorem (Bayes estimator under squared loss)**

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

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

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

**Proof:**
Remarks

1. \( f(x) = \int_{\Theta} f(x|\theta)p(\theta) d\theta \) is the marginal pdf of \( X \).
2. The conditional pdf \( p(\theta|x) = \frac{f(x|\theta)p(\theta)}{\int_{\Theta} f(x|\theta)p(\theta) d\theta} \) is called the posterior pdf of \( \theta \).
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:**
Example 1

Unless stated otherwise we will use the squared loss function.

Consider a random sample \( X = (X_1, \ldots, X_n) \) from a Bernoulli distribution with pdf \( f(x|\theta) = \theta^x(1-\theta)^{1-x}; \ x = 0, 1 \). Let the prior pdf of \( \theta \) be Uniform\((0, 1)\), i.e. \( p(\theta) = I(0 < \theta < 1) \).

1. Find the Bayes estimator of \( \theta \).
2. Find the risk of the Bayes estimator of \( \theta \).
3. Find the Bayes risk of the Bayes estimator of \( \theta \).
4. Find the Bayes estimator of \( \theta^2 \).
5. Formulate the risk of the Bayes estimator of \( \theta^2 \).
Example 2

Consider a random sample \( \mathbf{X} = (X_1, \cdots, X_n) \) from a Poisson distribution with pdf

\[
f(x|\theta) = \frac{\theta^x}{x!} e^{-\theta}; \quad x = 0, 1, \cdots.
\]

Let the prior pdf of \( \theta \) be \( \text{Gamma}(\beta, \kappa) \), i.e.

\[
p(\theta) = \frac{1}{\beta^\kappa \Gamma(\kappa)} \theta^{\kappa-1} e^{-\theta/\beta}; \quad \theta > 0; \beta > 0, \kappa > 0.
\]

Find the Bayes estimator of \( \theta \) and the associated risk.
Remarks

1. The idea involved in Bayes estimation is very appealing. Without any information or with only prior information about $\theta$, we would estimate $\tau(\theta)$ by its prior mean. Once the data are observed, new information about $\theta$ 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).
Example 3

Consider a random sample $Y = (Y_1, \cdots, Y_n)$ with $Y_i \overset{d}{=} \text{Poisson}(e^{\beta x_i})$ and $x_i$ being given, $i = 1, \cdots, n$. Let the prior pdf of $\beta$ be $N(0, 1)$.

Then the posterior pdf of $\beta$ is

$$p(\beta | y, x) = \frac{f(y | \beta, x)p(\beta)}{\int_B f(y | \beta, x)p(\beta) d\beta} = \frac{e^{-\sum_{i=1}^n e^{\beta x_i}} e^{\beta \sum_{i=1}^n x_i y_i} e^{-\beta^2/2}}{\int_{-\infty}^{\infty} e^{-\sum_{i=1}^n e^{\beta x_i}} e^{\beta \sum_{i=1}^n x_i y_i} e^{-\beta^2/2} d\beta}.$$

The Bayes estimator of $\beta$ is $T = E(\beta | y, x) = \int_{-\infty}^{\infty} \beta p(\beta | y, x) d\beta$, which does not have a closed form and will have to be calculated numerically.