Contents

5
Loss functions

5.1 Decision analysis

Bayesian analysis is combined with decision analysis via explicit recognition
of a loss function. Relative to classical statistics this is a strength as a loss
function always exists but is sometimes not acknowledged. For simplicity
and brevity we'll explore symmetric versions of a few loss functions.[1]

Let $\widehat{\theta}$ denote an estimator for θ, $c\left(\widehat{\theta}, \theta\right)$ denote a loss function, and
$p\left(\theta \mid y\right)$ denote the posterior distribution for θ given evidence y. A sensible
strategy for consistent decision making involves selecting the estimator, $\widehat{\theta}$,
to minimize the average or expected loss.

$$\min_{\widehat{\theta}} E\left[loss\right] = \min_{\widehat{\theta}} E\left[c\left(\widehat{\theta}, \theta\right)\right]$$

Briefly, for *symmetric* loss functions, we find the expected loss minimizing
estimator is the posterior mean for a quadratic loss function, is the median
of the posterior distribution for a linear loss function, and is the posterior
mode for an all or nothing loss function.

---

[1] The more general case, including asymmetric loss, is addressed in chapter 4 of *Accounting and Causal Effects: Econometric Challenges*.

5.2   Quadratic loss

Let $c\left(\widehat{\theta}, \theta\right) = \alpha\left(\widehat{\theta} - \theta\right)^2$ for $\alpha > 0$, then (with support from a to b)

$$\min_{\widehat{\theta}} E\left[c\left(\widehat{\theta}, \theta\right)\right] = \min_{\widehat{\theta}} \int_a^b \alpha\left(\widehat{\theta} - \theta\right)^2 p\left(\theta \mid y\right) d\theta$$

The first order conditions are

$$\frac{d}{d\widehat{\theta}} \int_a^b \alpha\left(\widehat{\theta} - \theta\right)^2 p\left(\theta \mid y\right) d\theta = 0$$

Expansion of the integrand gives

$$\frac{d}{d\widehat{\theta}} \int_a^b \alpha\left(\widehat{\theta}^2 - 2\widehat{\theta}\theta + \theta^2\right) p\left(\theta \mid y\right) d\theta = 0$$

or

$$\frac{d}{d\widehat{\theta}} \alpha\left(\widehat{\theta}^2 - 2\widehat{\theta} \int_a^b \theta p\left(\theta \mid y\right) d\theta + \int_a^b \theta^2 p\left(\theta \mid y\right) d\theta\right) = 0$$

Differentiation gives

$$\alpha\left(2\widehat{\theta} - 2\int_a^b \theta p\left(\theta \mid y\right) d\theta\right) = 0$$

and the solution is

$$\widehat{\theta} = \int_a^b \theta p\left(\theta \mid y\right) d\theta$$

where $\int_a^b \theta p\left(\theta \mid y\right) d\theta$ is the posterior mean.

5.3   Linear loss

Let $c\left(\widehat{\theta}, \theta\right) = \alpha\left|\widehat{\theta} - \theta\right|$ for $\alpha > 0$, then

$$\min_{\widehat{\theta}} E\left[c\left(\widehat{\theta}, \theta\right)\right] = \min_{\widehat{\theta}} \int_a^b \alpha\left|\widehat{\theta} - \theta\right| p\left(\theta \mid y\right) d\theta$$

The first order conditions are

$$\frac{d}{d\widehat{\theta}} \int_a^b \alpha\left|\widehat{\theta} - \theta\right| p\left(\theta \mid y\right) d\theta = 0$$

Rearrangement of the integrand gives

$$\frac{d}{d\widehat{\theta}} \left[\int_a^{\widehat{\theta}} \alpha\left(\widehat{\theta} - \theta\right) p\left(\theta \mid y\right) d\theta - \int_{\widehat{\theta}}^b \alpha\left(\widehat{\theta} - \theta\right) p\left(\theta \mid y\right) d\theta\right] = 0$$

or

$$\frac{d}{d\widehat{\theta}} \left[\alpha \left(\begin{array}{c} \widehat{\theta} F\left(\widehat{\theta} \mid y\right) - \int_a^{\widehat{\theta}} \theta p\left(\theta \mid y\right) d\theta - \widehat{\theta}\left(1 - F\left(\widehat{\theta} \mid y\right)\right) \\ + \int_{\widehat{\theta}}^b \theta p\left(\theta \mid y\right) d\theta \end{array} \right) \right] = 0$$

where $F\left(\widehat{\theta} \mid y\right)$ is cumulative posterior probability evaluated at $\widehat{\theta}$. Differentiation gives

$$\alpha \left(\begin{array}{c} F\left(\widehat{\theta} \mid y\right) + \widehat{\theta} p\left(\widehat{\theta} \mid y\right) - \widehat{\theta} p\left(\widehat{\theta} \mid y\right) \\ -\left(1 - F\left(\widehat{\theta} \mid y\right)\right) - \widehat{\theta}\left(1 - p\left(\widehat{\theta} \mid y\right)\right) + \widehat{\theta}\left(1 - p\left(\widehat{\theta} \mid y\right)\right) \end{array} \right) = 0$$

or

$$\alpha \left(F\left(\widehat{\theta} \mid y\right) - \left(1 - F\left(\widehat{\theta} \mid y\right)\right) \right) = 0$$
$$\alpha \left(2F\left(\widehat{\theta} \mid y\right) - 1 \right) = 0$$

and the solution is

$$F\left(\widehat{\theta} \mid y\right) = \frac{1}{2}$$

or the median of the posterior distribution.

5.4   All or nothing loss

Let

$$c\left(\widehat{\theta}, \theta\right) = \left\{ \begin{array}{cc} \alpha > 0 & \text{if } \widehat{\theta} \neq \theta \\ 0 & \text{if } \widehat{\theta} = \theta \end{array} \right.$$

Then, we want to assign $\widehat{\theta}$ the maximum value of $p\left(\theta \mid y\right)$ or the posterior mode.

5.5   Experimentation

The science of experimentation and evaluation of evidence is a deep and subtle art. Essential ingredients include careful framing of the problem (*theory development*), matching of *data* to be collected with the frame, and iterative *model specification* that complements the data and problem frame so that causal effects can be inferred. Consistent evaluation of evidence draws from the posterior distribution which is proportional to the product of the likelihood and prior. When they combine to generate a recognizable posterior distribution, our work is made simpler. Next, we briefly discuss conjugate families which produce recognizable posterior distributions.