Bayesian Estimates

1 Background and Motivation

Suppose that

$$y_1, y_2, ..., y_n \quad iid \quad f(y|\theta)$$

and we wish to estimate θ.
Let T be an estimate for θ. How good is T ? The mean square estimation error,

$$MSE(\theta, T) = E_\theta \left\{ (T - \theta)^2 \right\} = \int (t - \theta)^2 f(y|\theta)\, dy,$$

may be used to answer this question.

Notice that $MSE(\theta, T)$ is a non-negative function of θ, and T, . For the trivial estimate

$$T \equiv t_0$$

we have

$$MSE(t_0, t_0) = 0.$$

Therefore, it is not possible to find (except for trivial cases) an optimal estimate T^* with the property

$$MSE(\theta, T^*) \leq MSE(\theta, T), \quad \text{for all } T \text{ and } \theta.$$

Possible approaches to circumvent this problem are:

1) **Restrict the class of possible estimates** by imposing some "reasonable" restriction such as

- **Unbiasedness**

  $$E_\theta(T) = \theta, \quad \text{for all } \theta$$

  In this case

  $$MSE(\theta, T) = E_\theta\left\{(T - \theta)^2\right\} = Var_\theta(T)$$

  and we search for the minimum variance unbiased estimator (MUVE).

- **Invariance**

2) **Operate out the value of** θ. For example, average θ out. That is, assume some "weighting" distribution

over the parameter space, Θ, and consider

$$mse\left(T\right) = \int_{\Theta} MSE\left(\theta, T\right) \pi\left(\theta\right) d\theta$$

In this case, the optimal estimate minimizes $mse\left(T\right)$, that is:

$$T^{*} = \arg\min\left[mse\left(T\right)\right]$$

We can write:

$$mse\left(T\right) = \int MSE\left(\theta, T\right) \pi\left(\theta\right) d\theta$$

$$= \int \left[\int \left(t-\theta\right)^{2} f\left(\mathbf{y}|\theta\right) dy\right] \pi\left(\theta\right) d\theta$$

$$= \int \int \left(t-\theta\right)^{2} f\left(\mathbf{y}|\theta\right) \pi\left(\theta\right) d\mathbf{y} d\theta$$

$$= \int \int \left(t-\theta\right)^{2} h\left(\theta|\mathbf{y}\right) m\left(\mathbf{y}\right) d\theta d\mathbf{y}$$

$$\int (t-\theta)^2 h(\theta|\mathbf{y})\,d\theta = \int (\theta - t)^2 h()\,d\theta$$

$$\geq \int (\theta - E[\theta|\mathbf{y}])^2 h(\theta|\mathbf{y})\,d\theta$$

with equality if and only if $h(\theta|\mathbf{y})$ puts all its mass at $E[\theta|\mathbf{y}]$.

Therefore, we solve the optimality problem by choosing

$$T^*(\mathbf{y}) = E[\theta|\mathbf{y}]$$

More precisely,

$$mse(E[\theta|\mathbf{y}]) \leq mse(T), \quad \text{for all } T$$

Step by step derivation of T^*:

1. Chose an appropriate weighting density $\pi(\theta)$, called "*prior density*"

2. Write down the joint density for the data and the parameter

$$h\left(\mathbf{y},\theta\right) = f\left(\mathbf{y}|\theta\right)\pi\left(\theta\right) \tag{1}$$

3. Compute the marginal density of the data, by integrating out the parameter in (1)

$$m\left(\mathbf{y}\right) = \int_{\Theta} h\left(\mathbf{y},\theta\right) d\theta$$

4. Compute the posterior density of the parameter given the data

$$h\left(\theta|\mathbf{y}\right) = \frac{h\left(\mathbf{y},\theta\right)}{m\left(\mathbf{y}\right)} = \frac{f\left(\mathbf{y}|\theta\right)\pi\left(\theta\right)}{\int f\left(\mathbf{y}|\theta\right)\pi\left(\theta\right) d\theta}$$

   which is called "*posterior density*"

5. Compute the mean of the posterior density

$$T^{*} = E\left[\theta|\mathbf{y}\right] = \frac{\int \theta f\left(\mathbf{y}|\theta\right)\pi\left(\theta\right) d\theta}{\int f\left(\mathbf{y}|\theta\right)\pi\left(\theta\right) d\theta}$$

Example 1: Suppose that $y_1, y_2, ..., y_n$ are iid Bernoulli(p) and we wish to estimate p.

Step 0. The likelihood

$$f(\mathbf{y}|p) = p^{\sum y_i}(1-p)^{n-\sum y_i}$$

$$= p^s(1-p)^{n-s}, \quad s = \sum y_i$$

Step 1. The prior density:

$$\pi(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}$$

for some **specific values** of $\alpha > 0$, and $\beta > 0$.

It can be verified that

$$E_\pi(p) = \frac{\alpha}{\alpha+\beta}, \quad Var_\pi(p) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Step 2. The joint density:

$$h(\mathbf{y},p) = f(\mathbf{y}|p)\,\pi_{\alpha,\beta}(p)\,, \qquad \text{(hyperparameters } \alpha, \beta\text{)}$$

$$p^{\sum y_i}(1-p)^{n-\sum y_i}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{\alpha-1}(1-p)^{\beta-1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{s}(1-p)^{n-s}\,p^{\alpha-1}(1-p)^{\beta-1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}p^{s+\alpha-1}(1-p)^{n-s+\beta-1}$$

where

$$s = \sum y_i.$$

Step 3. The marginal density:

$$m(\mathbf{y}) = \int_{-\infty}^{\infty} h(\mathbf{y},p)\,dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{s+\alpha-1}(1-p)^{n-s+\beta-1}\,dp$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(s+\alpha)\Gamma(n-s+\beta)}{\Gamma(n+\alpha+\beta)}$$

Step 4. The posterior density:

$$h(p|\mathbf{y}) = \frac{h(\mathbf{y},p)}{m(\mathbf{y})}$$

$$= \frac{\Gamma(n+\alpha+\beta)}{\Gamma(s+\alpha)\Gamma(n-s+\beta)} p^{s+\alpha-1}(1-p)^{n-s+\beta-1}$$

which is a Beta$(s+\alpha, n+\beta-s)$

Step 5. The posterior mean:

We will use that if $X \sim Beta(a, b)$ then

$$E(X) = a/(a+b).$$

$$
\begin{aligned}
T_n^* &= E\left[p|\mathbf{y}\right] \\
&= \frac{s+\alpha}{(s+\alpha)+(n+\beta-s)} \\
&= \frac{s+\alpha}{\alpha+\beta+n} \\
&= \left(\frac{n}{\alpha+\beta+n}\right)\frac{s}{n} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta} \\
&= w\hat{p} + (1-w)\,E\left(p\right)
\end{aligned}
$$

Notice that

$$T_n^* \to \hat{p} \quad \text{as} \quad n \to \infty.$$

On the other hand, if

$$\alpha = \tau\beta, \qquad \beta \to \infty$$

then

$$T_n^* = \left(\frac{n}{\alpha+\beta+n}\right)\frac{s}{n} + \left(\frac{\alpha+\beta}{\alpha+\beta+n}\right)\frac{\alpha}{\alpha+\beta}$$

$$\to \frac{\alpha}{\alpha+\beta} = \frac{\tau\beta}{\tau\beta+\beta} = \frac{\tau}{\tau+1}$$

Therefore, we can make

$$T_n^* \approx \frac{\tau}{\tau+1}$$

for any chosen value of τ.

Remark 1 *The "non-informative" prior density corresponds to the choice*

$$\alpha = \beta = 1$$

as in this case

$$\pi(p) = 1, \quad \textit{for} \quad 0 < p < 1$$

In this case the posterior density is given by

$$h\left(p|\mathbf{y}\right)=\frac{\Gamma\left(n+2\right)}{\Gamma\left(s+1\right)\Gamma\left(n-s+1\right)}p^{s}\left(1-p\right)^{n-s}$$

which is a Beta$(s+1,n-s+1)$, *with mean*

$$\frac{s+1}{n+2}$$

That is, the sample proportion for modified sample which has been increased by two observations, one being a "success" and the other a "failure". For example, if the original sample is

$$n=10,\quad s=1,$$

the MLE of p *is*

$$\hat{p}=\frac{1}{10}=0.10$$

and the Bayesian estimate is has posterior mean

$$\frac{2}{11}=0.181\,82.$$

Example 2: Suppose that $y_1, y_2, ..., y_n$ are *iid* $Exp(\theta)$ and we are interested in estimating the parameter θ.

Step 0. The likelihood:

$$f(\mathbf{y}|\theta) = \theta^n \exp\left\{-\theta \sum_{i=1}^{n} y_i\right\}$$

$$= \theta^n e^{-\theta s}, \qquad s = \sum_{i=1}^{n} y_i$$

Step 1. The prior:

$$\pi(\theta) = \lambda \exp\{-\lambda\theta\} = Gamma(1, \lambda)$$

for some **specific values** of $\lambda > 0$.

It can be verified that

$$E_\pi(\theta) = \frac{1}{\lambda}, \quad Var_\pi(\theta) = \frac{1}{\lambda^2}$$

Step 2. The joint density of **y and** θ:

$$h(\mathbf{y},\theta) = f(\mathbf{y}|\theta)\,\pi(\theta)$$

$$= \theta^n \exp\left\{-\theta\sum_{i=1}^{n} y_i\right\}\lambda\exp\{-\lambda\theta\}$$

$$= \theta^n \lambda \exp\left\{-\theta\left[\lambda + \sum_{i=1}^{n} y_i\right]\right\}$$

$$= \theta^n \lambda \exp\{-[\lambda + s]\,\theta\}$$

where

$$s = \sum_{i=1}^{n} y_i.$$

Step 3. The marginal density:

$$m(\mathbf{y}) = \int_{-\infty}^{\infty} h(\mathbf{y}{,}\theta)\, d\theta$$

$$= \frac{\lambda}{[\lambda + s]^{n+1}} \Gamma(n+1) \frac{[\lambda + s]^{n+1}}{\Gamma(n+1)} \int_0^{\infty} \theta^n \exp\{-[\lambda + s]\theta\}\, d\theta$$

$$= \frac{\lambda}{[\lambda + s]^{n+1}} \Gamma(n+1)$$

Step 4. The posterior density:

$$h(\theta|\mathbf{y}) = \frac{h(\mathbf{y}{,}p)}{m(\mathbf{y})}$$

$$= \frac{\theta^n \lambda \exp\{-[\lambda + s]\theta\}}{\frac{\lambda}{[\lambda+s]^{n+1}} \Gamma(n+1)}$$

$$= \frac{[\lambda + s]^{n+1}}{\Gamma(n+1)} \theta^{(n+1)-1} \exp\{-[\lambda + s]\theta\}$$

which is a $Gamma\left(n+1, \lambda+s\right)$

Step 5. The posterior mean:

$$T_n^* = E\left[\theta|\mathbf{y}\right]$$

$$= \frac{n+1}{\lambda+s}$$

$$= \frac{1}{\frac{\lambda}{n+1} + \frac{n}{n+1}\bar{y}}$$

The effect of this prior distribution is to create a modified sample by adding a single observation $y_{n+1} = \lambda$. Moreover, notice that

$$T_n^* \to \frac{1}{\bar{y}} \quad \text{as} \quad n \to \infty.$$