L11 Bayes estimator

1. Bayes estimator

   (1) Five distributions
   In Bayesian statistics parameter $\theta \in R^p$ is treated as a random vector with a prior distribution that reflects the prior knowledge on θ.
   There will be the distribution of sample Y, the joint distribution of θ and Y, the conditional distribution of Y given θ and the conditional distribution of θ given Y.
   The traditional distribution family of Y with parameter θ is now treated as a conditional distribution of Y given θ.

   (2) Bayes estimator
   The conditional distribution of θ given Y is the posterior distribution of θ. The mean of this distribution, $E(\theta|Y)$, is a function of Y, a statistic. This statistic is often used to estimate θ, and is called the Bayes estimator of θ.

   (3) Finding the posterior pdf
   Suppose $f_\theta(\theta) \propto f_0(\theta)$ which means the prior pdf $f_\theta(\theta)$ is proportional to a given function $f_0(\theta)$, i.e., $f_\theta(\theta) = c_0 f_0(\theta)$ where $c_0 > 0$
   Suppose $L(\theta) = f_{Y|\theta}(y) \propto f_1(\theta)$ which means the likelihood function, the conditional pdf of Y given θ treated as a function of θ, is proportional to a given function $f_1(\theta)$, i.e., $f_{Y|\theta}(y) = c_1(y) f_1(\theta, y)$ where $c_1(y) > 0$.
   Then $f_{\theta|Y}(\theta) \propto f_0(\theta) f_1(\theta)$, i.e., the posterior pdf of θ is proportional to $f_0(\theta) f_1(\theta)$.
   Proof. The joint pdf of θ and Y is $f(y, \theta) = f_\theta(\theta) f_{Y|\theta}(y) = c_0 f_0(\theta) c_1(y) f_1(\theta, y)$.
   Hence $f_{\theta|Y}(\theta) = \frac{f(\theta, y)}{f_Y(y)} = \frac{c_0 c_1(y)}{f_Y(y)} f_0(\theta) f_1(\theta, y) \propto f_0(\theta) f_1(\theta, y)$.

Ex1: For $X \sim N_p(\mu, \Sigma)$ let $f_0(x) = \exp\left(\frac{x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu}{-2}\right)$. The pdf of X

$$\begin{aligned} f(x) &= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left[-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right] \\ &= \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(\frac{\mu'\Sigma^{-1}\mu}{-2}\right) \exp\left(\frac{x'\Sigma^{-1}x - 2x'\Sigma^{-1}\mu}{-2}\right) = c f_0(x) \end{aligned}$$

So $X \sim N(\mu, \Sigma) \Longrightarrow f(x) \propto f_0(x)$. On the other hand,

$$\begin{aligned} f(x) \propto f_0(x) &\Longrightarrow f(x) = c f_0(x) \Longrightarrow 1 = \textstyle\int_R f(x)dx = c \int_R f_0(x)dx \\ &\Longrightarrow f(x) = \frac{f_0(x)}{\int_R f_0(x)dx} \Longrightarrow X \sim N(\mu, \Sigma). \end{aligned}$$

2. Bayes estimator in linear model

   (1) Prior and posterior distributions for β
   For $Y|\beta \sim N(X\beta, \sigma^2 I_n)$ assign prior $\beta \sim N(\beta_0, \Sigma_0)$. Then the posterior

$$\beta|Y \sim N\left(\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'Y}{\sigma^2}\right), \left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\right)$$

   Because the posterior and prior are in the same distribution family, the prior is called a conjugate prior.

Pf: $\beta \sim N(\beta_0,\, \Sigma_0) \Longrightarrow f_\beta(\beta) \propto f_0(\beta) = \exp\left[\frac{\beta'\Sigma_0^{-1}\beta - 2\beta'\Sigma_0^{-1}\beta_0}{-2}\right]$.

$Y|\beta \sim N(X\beta,\, \sigma^2 I_n) \Longrightarrow L(\beta) \propto f_1(\beta) = \exp\left[\frac{(X\beta)'(X\beta) - 2(X\beta)'Y}{-2\sigma^2}\right]$.

$$\begin{aligned} f_{\beta|Y}(\beta) &\propto f_0(\beta)f_1(\beta) = \exp\left[\frac{\beta'\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)\beta - 2\beta'\left(\Sigma_0^{-1}\beta_0 + \frac{X'Y}{\sigma^2}\right)}{-2}\right] \\ &= \exp\left[\frac{\beta'\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)\beta - 2\beta'\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'Y}{\sigma^2}\right)}{-2}\right] \end{aligned}$$

Thus $\beta|Y \sim N\left(\left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'Y}{\sigma^2}\right),\, \left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\right)$

(2) Bayes estimator for β

Let $\widehat{\beta} = (X'X)^{-1}X'Y$ be the BLUE for β in classical statistics. Then the Bayesian estimator for β

$$\begin{aligned} \widehat{\beta}_B &= E(\beta|Y) = \left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'Y}{\sigma^2}\right) \\ &= \left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'X}{\sigma^2}(X'X)^{-1}X'Y\right) \\ &= \left(\Sigma_0^{-1} + \frac{X'X}{\sigma^2}\right)^{-1}\left(\Sigma_0^{-1}\beta_0 + \frac{X'X}{\sigma^2}\widehat{\beta}\right). \end{aligned}$$

Let $W_1 = \Sigma_0^{-1}$, $W_2 = \frac{X'X}{\sigma^2}$ and $W = W_1 + W_2$. Then

$$\widehat{\beta}_B = W^{-1}(W_1\beta_0 + W_2\widehat{\beta}).$$

So the Bayesian estimator for β is the weighted average of prior mean β_0 and classical BLUE $\widehat{\beta}$ with weight matrices W_1 and W_2.

3. Loss functions and risk functions

(1) Risk function in Bayesian statistics

In Bayesian statistics, the risk is calculated with respect to the posterior distribution of θ, i.e.,

$$r(\widehat{\theta},\, \theta) = E\left[(\widehat{\theta} - \theta)(\widehat{\theta} - \theta)'|Y\right] \in R^{p\times p}$$

where $\widehat{\theta}(Y)$ is treated as non-random. With Bayesian estimator $\widehat{\theta}_B = E(\theta|Y)$,

$$r(\widehat{\theta}_B,\, \theta) = E[(\theta - \widehat{\theta}_B)(\theta - \widehat{\theta}_B)'] = \mathrm{Cov}(\theta|Y).$$

(2) Bayes estimator dominates all estiamtors

Let $\widehat{\theta}$ be an estimator for θ. Then

$$\begin{aligned} r(\widehat{\theta},\, \theta) &= E\left[(\theta - \widehat{\theta})(\theta - \widehat{\theta})'|Y\right] \\ &= E\left[(\theta - \widehat{\beta}_B + \widehat{\beta}_B - \widehat{\theta})(\theta - \widehat{\beta}_B + \widehat{\beta}_B - \widehat{\theta})'|Y\right] \\ &= E\left[(\theta - \widehat{\beta}_B)(\theta - \widehat{\beta}_B|Y\right] + E\left[(\widehat{\beta}_B - \widehat{\theta})(\widehat{\beta}_B - \widehat{\theta})'|Y\right] \\ &= r(\widehat{\theta}_B,\, \theta) + (\widehat{\beta}_B - \widehat{\theta})(\widehat{\beta}_B - \widehat{\theta})' \\ &\geq r(\widehat{\theta}_B,\, \theta). \end{aligned}$$

Thus for Bayes estimator $\widehat{\theta}_B$, $E[L(\widehat{\theta}_B,\, \theta)|Y] = \mathrm{Cov}(\theta|Y) \leq E[L(\widehat{\theta},\, \theta)|Y]$.

L12 Confidence regions

1. $\frac{\text{SSE}}{\sigma^2} \sim \chi^2(n-p)$

   (1) Expression of SSE
   In $Y = X\beta + \epsilon$, $\epsilon \sim N(0,\, \sigma^2\Sigma)$, $E(Y) = X\beta$ is estimable with BLUE
   $$\widehat{Y} = X\widehat{\beta} = X\left(\Sigma^{-1/2}X\right)^{+}\Sigma^{-1/2}Y.$$
   So $\epsilon = Y - X\beta$ is predicted by the residual vector
   $$e = Y - \widehat{Y} = Y - X\left(\Sigma^{-1/2}X\right)^{+}\Sigma^{-1/2}Y = \Sigma^{1/2}\left[I - \left(\Sigma^{-1/2}X\right)\left(\Sigma^{-1/2}X\right)^{+}\right]\Sigma^{-1/2}Y.$$
   $$\begin{aligned}\text{SSE} &\overset{def}{=\!=} \|Y-\widehat{Y}\|^2_{\Sigma^{-1}} = \|\Sigma^{-1/2}(Y-\widehat{Y})\|^2 \\ &= \left\|\left[I - \left(\Sigma^{-1/2}X\right)\left(\Sigma^{-1/2}X\right)^{+}\right]\left(\Sigma^{-1/2}Y\right)\right\|^2 = Z'AZ.\end{aligned}$$
   where $Z = \Sigma^{-1/2}Y \sim N\left(\Sigma^{-1/2}X\beta,\, \sigma^2 I_n\right)$, $A = I - DD^{+}$ with $D = \Sigma^{-1/2}X$.

   (2) $\frac{\text{SSE}}{\sigma^2} \sim \chi^2(n-r)$ where $r = \text{rank}(X)$
   Recall: $X \sim N(\mu,\, \Sigma)$, $A\Sigma A = A = A' \Longrightarrow X'AX \sim \chi^2(\mu' A\mu,\, \text{tr}(A\Sigma))$.
   With $\frac{\text{SSE}}{\sigma^2} = Z'\frac{A}{\sigma^2}Z$ where $Z \sim N(\Sigma^{-1/2}X\beta,\, \sigma^2 I)$ and $A = I - DD^{+}$, $D = \Sigma^{-1/2}X$,
   $\frac{A}{\sigma^2}\sigma^2 I\frac{A}{\sigma^2} = \frac{I-DD^{+}}{\sigma^2}\sigma^2 I\frac{I-DD^{+}}{\sigma^2} = \frac{I-DD^{+}}{\sigma^2} = \frac{A}{\sigma^2}$,
   $\left(\Sigma^{-1/2}X\beta\right)'\frac{A}{\sigma^2}\left(\Sigma^{-1/2}X\beta\right) = \left(\Sigma^{-1/2}X\right)'\frac{I-DD^{+}}{\sigma^2}(D\beta) = 0$ and
   $\text{tr}\left(\frac{A}{\sigma^2}\sigma^2 I\right) = \text{tr}(I - DD^{+}) = n - \text{rank}(D) = n - r$ where $r = \text{rank}(X)$.
   Hence $\frac{\text{SSE}}{\sigma^2} \sim \chi^2(n-r)$.

   (3) Consequences
   $\text{MSE} \overset{def}{=\!=} \frac{\text{SSE}}{n-r}$ is an UE for σ^2 since $E\left(\frac{\text{SSE}}{\sigma^2}\right) = n - r \Longrightarrow E(\text{MSE}) = \sigma^2$.
   $\left(\frac{\text{SSE}}{\chi^2_{\alpha/2}(n-r)},\, \frac{\text{SSE}}{\chi^2_{1-\alpha/2}(n-r)}\right)$ is a $1-\alpha$ CI for σ^2 since
   $$\begin{aligned}&P\left(\frac{\text{SSE}}{\chi^2_{\alpha/2}(n-r)} \le \sigma^2 \le \frac{\text{SSE}}{\chi^2_{1-\alpha/2}(n-r)}\right) = P\left(\frac{\chi^2_{1-\alpha/2}(n-r)}{\text{SSE}} \le \frac{1}{\sigma^2} \le \frac{\chi^2_{\alpha/2}(n-r)}{\text{SSE}}\right) \\ =\ &P\left(\chi^2_{1-\alpha/2}(n-r) \le \frac{\text{SSE}}{\sigma^2} \le \chi^2_{\alpha/2}(n-r)\right) = 1-\alpha.\end{aligned}$$

   Ex1: For $Y = X\beta + \epsilon$, $\epsilon \sim N(0,\, \sigma^2 I)$, $\text{SSE} = Y'(I - XX^{+})Y$, $\text{MSE} = \frac{\text{SSE}}{n-p}$ and $\frac{\text{SSE}}{\sigma^2} \sim \chi^2(n-p)$.

2. $(H\widehat{\beta} - H\beta)'\frac{[H(X'\Sigma^{-1}X)^{-1}H']^{-1}}{\sigma^2}(H\widehat{\beta} - H\beta) \sim \chi^2(q)$.

   (1) $H\widehat{\beta} - H\beta \sim N(0,\, \sigma^2 H(X'\Sigma^{-1}X)^{-1}H)$.
   In $Y = X\beta + \epsilon$, $\epsilon \sim N(0,\, \sigma^2\Sigma)$, X has full column rank so $\theta = H\beta$ is estimable for all H. With full row rank H, $\theta = H\beta \in R^q$ has BLUE
   $H\widehat{\beta} = H\left(\Sigma^{-1/2}X\right)^{+}\Sigma^{-1/2}Y = H(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \sim N\left(H\beta,\, \sigma^2 H(X'\Sigma^{-1}X)^{-1}H\right)$.
   So $H\widehat{\beta} - H\beta \sim N(0,\, \sigma^2 H(X'\Sigma^{-1}X)^{-1}H)$.

   (2) $(H\widehat{\beta} - H\beta)'\frac{[H(X'\Sigma^{-1}X)^{-1}H']^{-1}}{\sigma^2}(H\widehat{\beta} - H\beta) \sim \chi^2(q)$.
   Write $(H\widehat{\beta} - H\beta)'\frac{[H(X'\Sigma^{-1}X)^{-1}H']^{-1}}{\sigma^2}(H\widehat{\beta} - H\beta)$ as $Z'AZ$ where $Z = H\widehat{\beta} - H\beta$ and $A = \frac{[H(X'\Sigma^{-1}X)^{-1}H']^{-1}}{\sigma^2}$. Then $Z \sim N(0,\, A^{-1})$. Note that
   $AA^{-1}A = A$, $0'A0 = 0$ and $\text{tr}(AA^{-1}) = \text{tr}(I_q) = q$.
   So $Z'AZ \sim\sim \chi^2(q)$.

3. Confidence regions

(1) A pivotal quantity

$F = \frac{(H\beta - H\widehat{\beta})'[H(X'\Sigma^{-1}X)^{-1}H']^{-1}(H\beta - H\widehat{\beta})}{q\,\text{MSE}} \sim F(q,\, n-p)$

Proof. In $Y = X\beta + \epsilon$, $\epsilon \sim N(0,\, \sigma^2\Sigma)$, X has full column rank.

With $\widehat{\beta} = \left(\Sigma^{-1/2}X\right)^{+}\Sigma^{-1/2}Y = D\Sigma^{-1/2}Y$ where $D = \Sigma^{-1/2}X$ and

$\text{SSE} = (\Sigma^{-1/2}Y)'\left[I - \left(\Sigma^{-1/2}X\right)\left(\Sigma^{-1/2}X\right)^{+}\right](\Sigma^{-1/2}Y) = Y'\Sigma^{-1/2}(I - DD^{+})\Sigma^{-1/2}Y$,

$\widehat{\beta}$ and SSE are independent since

$$\left(D\Sigma^{-1/2}\right)\left(\sigma^2\Sigma\right)\left[\Sigma^{-1/2}(I - DD^{+})\Sigma^{-1/2}\right] = 0.$$

$$\begin{aligned}\text{Hence } F &= \frac{(H\beta - H\widehat{\beta})'[H(X'\Sigma^{-1}X)^{-1}H']^{-1}(H\beta - H\widehat{\beta})}{q\,\text{MSE}} \\ &= \frac{(H\beta - H\widehat{\beta})'[H(X'\Sigma^{-1}X)^{-1}H']^{-1}(H\beta - H\widehat{\beta})/(\sigma^2 q)}{\text{SSE}/[\sigma^2(n-p)]} \sim F(q,\, n-p).\end{aligned}$$

(2) $1-\alpha$ confidence region for $\theta = H\beta$.

Because $P\left(\frac{(H\beta - H\widehat{\beta})'[H(X'\Sigma^{-1}X)^{-1}H']^{-1}(H\beta - H\widehat{\beta})}{q\,\text{MSE}} \le F_\alpha(q,\, n-p)\right) = 1-\alpha$,

the collection of all $\theta \in R^q$ satisfying

$$\frac{(\theta - H\widehat{\beta})'[H(X'\Sigma^{-1}X)^{-1}H']^{-1}(\theta - H\widehat{\beta})}{q\,\text{MSE}} \le F_\alpha(q,\, n-p)$$

is a $1-\alpha$ confidence region for $\theta = H\beta$.

This region can be written as $(\theta - H\widehat{\beta})'A^{-1}(\theta - H\widehat{\beta}) \le 1$, and hence is an ellipsoid in R^q with center $H\widehat{\beta}$.

Ex2: With $H = I_p$, the $1-\alpha$ confidence region for β is the collection of β satisfying

$$\frac{(\beta - \widehat{\beta})'(X'\Sigma^{-1}X)(\beta - \widehat{\beta})}{p,\text{MSE}} \le F_\alpha(p,\, n-p)$$

Ex3: With $H = l'$ such that $\theta = H\beta = l'\beta \in R$, the $1-\alpha$ confidence region for θ becomes a confidence interval.

$$\begin{aligned} & \frac{(\theta - l'\widehat{\beta})'\left[l'(X'\Sigma^{-1}X)^{-1}l\right]^{-1}(\theta - l'\widehat{\beta})}{\text{MSE}} \le F_\alpha(1,\, n-p) \\ \iff\ & (\theta - l'\widehat{\beta})^2 \le F_\alpha(1,\, n-p)\,\text{MSE}\ \ l'(X'\Sigma^{-1}X)^{-1}l \\ \iff\ & \theta \in l'\widehat{\beta} \pm \sqrt{F_\alpha(1,\, n-p)}\ \sqrt{\text{MSE}\ l'(X'\Sigma^{-1}X)^{-1}l} \end{aligned}$$