(39) ## 4.2 Linear Bayes estimation

Recall that the Bayes estimator $\hat{\mu}_B$ is defined as the minimizer of the minimization problem:

$$S(\hat{\mu}) := E[(\mu(\theta_i) - \hat{\mu})^2] \longrightarrow \min \qquad (4.2.1)$$

w.r.t. $\hat{\mu} \in \mathcal{M}$, where totality of claim size information

$$\mathcal{M} := \{Z : Z \text{ r.v. w.r.t. } \mathcal{A} := \sigma(X_1, \dots, X_r) \text{ with } E[Z^2] < \infty\} \qquad (4.2.2)$$

Here $\mu(\theta_i) \overset{\text{def}}{=} E[X_{i1} | \theta_i]$.

Rem. 4.2.1: One shows that

$\hat{\mu} \in \mathcal{M} \Longrightarrow$ there ex. a (measurable) function g s.t. $\hat{\mu} = g(X_1, \dots, X_r)$. (4.2.3)

$\longrightarrow$ $\hat{\mu}_L = f(X_1, \dots, X_r)$ for a function f

$\longrightarrow$ problem: It is difficult to calculate the (non-linear) function f in general

However, if $r=1$, $n_1 = 1$ (sample size) and $f \in C^2$, then it follows from Taylor's theorem that

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R_n(x)$$

remainder term

$\longrightarrow$ $f(x) \approx f(x_0) + f'(x_0)(x - x_0)$, if $R_n(x)$ small

$\longrightarrow$ $\hat{\mu}_L \approx f(x_0) + f'(x_0)(X_1 - x_0)$, if $R_n(x)$ small

$$= a_0 + a_1 X_1 \qquad (4.2.4)$$

for $a_0 := f(x_0) - f'(x_0) \cdot x_0$, $a_1 := f'(x_0)$

$\longrightarrow$ Because of (4.2.4) we may also somewhat expect for general r, n_i that a minimizer of the minimization problem (4.2.1) w.r.t

$$\hat{\mu} \in \mathcal{L} := \{Z : Z = a_0 + \sum_{i=1}^{r} \sum_{t=1}^{n_i} a_{it} X_{i,t},\ a_0, a_{it} \in \mathbb{R}\} \subseteq \mathcal{M} \qquad (4.2.5)$$

could be a "good" approximation of $\mu(\theta_i)$

$\longrightarrow$ It turns out that an explicit minimizer w.r.t. (4.2.5) for even less restrictive

(40) models than the heterogeneity model, that is the Bühlmann and Bühlmann-Straub model exists

Such a minimizer is called linear Bayes estimator of $\mu(\Theta_i)$ and denoted by $\hat{\mu}_{LB}$

In order to find $\hat{\mu}_{LB}$ we need some notation and a general result on minimum risk estimation by linear functions:

Notation: X, Y_i, $i=1,\dots,m$ with $Var[X], Var[Y_i] < \infty$, $i=1,\dots,m$

$$Y := (Y_1,\dots,Y_m)' \leftarrow \text{transpose}$$

$$\mathcal{L}' := \{Z : Z = a_0 + a' \cdot Y,\ a_0 \in \mathbb{R},\ a \in \mathbb{R}^m\} \quad (4.2.6)$$

$$a = (a_1,\dots,a_m)' \in \mathbb{R}^m$$

$$E[Y] := (E[Y_1],\dots,E[Y_m])' \quad (4.2.7)$$

$$\Sigma_{X,Y} := (Cov[X,Y_1],\dots,Cov[X,Y_m])' \quad (4.2.8)$$

$$\Sigma_Y := ((Cov[Y_i,Y_j]))_{1\le i,j\le m} \quad (4.2.9)$$

(covariance matrix of Y)

Prop. 4.2.2 (Minimum risk estimation by linear functions)

Let $Var[X], Var[Y_i] < \infty$, $i=1,\dots,m$. Then

(i) Let $\hat{Y} = a_0 + a' \cdot Y$ with (a_0, a) $\leftarrow$ there exists always such a solution solves the equations

$$a_0 = E[X] - a' \cdot E[Y], \quad \Sigma'_{X,Y} = a' \cdot \Sigma_Y \quad (4.2.10)$$

($\leftarrow$ transpose)

Then for all $Z \in \mathcal{L}'$

$$E[(X-Z)^2] \ge E[(X-\hat{Y})^2] = Var[X] - a' \Sigma_Y a \quad (4.2.11)$$

So $\hat{Y}$ is a minimizer of $E[(X-Z)^2]$.

Conversely, (4.2.10) is a necessary cond. for a minimizer $\hat{Y}$.

(ii) Let $\hat{Y}$ be a minimizer. Then (4.2.10) is equivalent to the equations

$$E[X] = E[\hat{Y}], \quad Cov[X,Y_i] = Cov[\hat{Y},Y_i],\ i=1,\dots,m \quad (4.2.12)$$

(iii) If Σ_Y^{-1} exists, then there is a unique minimizer $\hat{Y} \in \mathcal{L}'$ given by

$$\hat{Y} = E[X] + \Sigma'_{X,Y} \cdot \Sigma_Y^{-1} (Y - E[Y]) \quad (4.2.13)$$

with risk

(41) $E[(X-\hat{Y})^2] = Var[X] - \Sigma_{X,Y}' \Sigma_Y^{-1} \Sigma_{X,Y}$ (4.2.14)

$= Var[X] - Var[\hat{Y}]$

Rem. 4.2.3 : (4.2.14) $\Rightarrow$ $\hat{Y}, X-\hat{Y}$ uncorrelated (exercise)

Proof:

(i) necessary cond.: We have to show that a minimizer $\hat{Y}$ has the form $\hat{Y} = a_0 + a' \cdot Y$ with (a_0, a) solving (4.2.10):

Since
$$\inf_{a, a_0} E[(X-(a_0+a'Y))^2] = \inf_a \inf_{a_0} E[(X-(a_0+a'Y)^2],$$
we can try to verify (4.2.10) by first minimizing w.r.t. a_0 and then w.r.t. a.

Let a be fixed.

Since $E[(X-Z)^2] \geq Var[X-Z]$

for all Z, we have that
$$E[(X-Z)^2] \geq E[(X-\hat{Y})^2] \geq Var[X-\hat{Y}]$$
for all $Z \in \mathcal{L}'$ $\Rightarrow$ $\hat{Y}$ and $\hat{Y} + E[X] - E[\hat{Y}]$ minimizers

$\Rightarrow$ $E[(X-\hat{Y})^2] = Var[X-\hat{Y}] = E[(X-\hat{Y})^2] - (E[X-\hat{Y}])^2$

$\Rightarrow$ $(E[X-\hat{Y}])^2 \Rightarrow E[X] = E[\hat{Y}] = a_0 + a' E[Y]$

$\Rightarrow$ $a_0 = E[X] - a' \cdot E[Y]$ in (4.2.10).

So $E[(X-\hat{Y})^2] = Var[X-\hat{Y}]$
$$= E[((X-E[X]) - \sum_{t=1}^m a_t (Y_t - E[Y_t]))^2]$$
$$= Var[X] + Var[\sum_{t=1}^m a_t Y_t] - 2 Cov[X, \sum_{t=1}^m a_t Y_t]$$
$$= Var[X] + \sum_{t=1}^m \sum_{s=1}^m a_t a_s Cov[Y_t, Y_s] - 2 \sum_{t=1}^m a_t Cov[X, Y_t]$$
$=: f(a)$ (4.2.14)

$\xRightarrow[a=(a_1,\dots,a_m)]{\text{minimum in}}$ $0 = \frac{\partial}{\partial a_k} f(a) = \sum_{t=1}^m a_t Cov[Y_k, Y_t] - Cov[X, Y_k]$

for all $k = 1, \dots, m$

$\Leftrightarrow$ $\Sigma_{X,Y}' = a' \Sigma_Y$ in (4.2.10) $\Rightarrow$ necessary cond.

sufficient cond.: Let $\hat{Y}$ be as in (4.2.10) with (a_0, a) and $Z = b_0 + b' \cdot Y \in \mathcal{L}'$ $\Rightarrow$

$E[(X-Z)^2] \geq Var[X-Z] = Var[(X-a' \cdot Y) + (a-b)' Y]$ (*)

$X - a'Y$ and $(a-b)'Y$ uncorrelated, since e.g. for $m=1$

(42) $\mathrm{Cov}[X-aY,(a-b)Y] = (a-b)(\underbrace{\mathrm{Cov}[X,Y]-a\,\mathrm{Cov}[Y,Y]}_{\overset{4.2.10}{=}0})$

$\Longrightarrow (*) = \mathrm{Var}[X-a'\cdot Y] + \mathrm{Var}[(a-b)'\cdot Y]$

$\geq \mathrm{Var}[X-a'Y] = E[(X-\hat{Y})^2] \Longrightarrow$ sufficient cond.

As for the (simpler) proofs of (ii) and (iii) see T. Mikosch.

Rem. 4.2.4: (i) The equivalent equations (4.2.10) and (4.2.12) are called normal equations.

(ii) The minimizer $\hat{Y}$ in Prop. 4.2.2 is called linear Bayes estimator of X.

Def. 4.2.5 (Bühlmann model)

(i) The i-th policy is described by the r.v.'s $(\theta_i, (X_{i,t})_{t\geq 1})$. Here θ_i is the heterogeneity parameter and $X_{i,t}, t\geq 1$ is the sequence of claim sizes or claim numbers.

(ii) The sequence $(\theta_i, (X_{i,t})_{t\geq 1}), i=1,2,\ldots$ is independent (see Def. 4.1.2 (ii)).

(iii) $\theta_i, i\geq 1$ i.i.d

(iv) Given θ_i, the r.v's $X_{i,t}, t\geq 1$ are independent (see Def. 4.1.2 (iii))

(v) There exist functions μ, v s.t. for all $t\geq 1$:

$$\mu(\theta_i) = E[X_{i,t}|\theta_i] \text{ and } v(\theta_i) = \mathrm{Var}[X_{i,t}|\theta_i]$$

$(\overset{(iii)}{\Longrightarrow} \mu(\theta_i), i\geq 1 \text{ and } v(\theta_i), i\geq 1 \text{ i.i.d.})$ $(\overset{\text{def}}{=} E[(X_{i,t}-E[X_{i,t}|\theta_i])^2|\theta_i])$

Rem. 4.2.6: Differences between the Bühlmann model and the heterogeneity model:

(i) Because of Def. 4.2.5 (ii) the r.v.'s $Z_i := (X_{i,t_1},\ldots,X_{i,t_m})$, $i\geq 1$ are not necessarily identically distributed for $t_1 < \ldots < t_m$

(ii) Because of Def. 4.2.5 (iv) the r.v.'s $Z_t := X_{i,t}, t\geq 1$ are not necessarily identically distributed

(iii) If $E[X_{i,t}^2] < \infty$ in Def. 4.1.2 then $\mu(\theta_i)$ and $v(\theta_i)$ are the same.

Lemma 4.2.7 (properties of $X_{i,t}$ in the Bühlmann model)

Assume the Bühlmann model for $Var[X_{i,t}]<\infty$ for all i,t. Then for all $i \geq 1$ and $t \neq s$:

(i) $E[X_{i,t}] = E[E[X_{i,t}|\Theta_i]] = E[\mu(\Theta_i)] =: \mu$ ← indep. of i, since Θ_i i.i.d.

(ii) $E[(X_{i,t})^2] = E[E[X_{i,t}^2|\Theta_i]] = E[Var[X_{i,t}|\Theta_i]] + E[(E[X_{i,t}|\Theta_i])^2] = \underbrace{E[V(\Theta_i)]}_{=:\,\varphi \text{ ← indep. of } i} + E[(\mu(\Theta_i))^2]$

$= \varphi + \lambda + \mu^2$,

where $\lambda := Var[\mu(\Theta_i)]$

(iii) $Var[X_{i,t}] = \varphi + \lambda$

(iv) $Cov[X_{i,t}, X_{i,s}] = E[E[X_{i,t}-E[X_{i,t}]|\Theta_i]\cdot E[X_{i,s}-E[X_{i,s}]|\Theta_i]]$
$= Var[\mu(\Theta_i)] = \lambda$

(v) $Cov[\mu(\Theta_i), X_{i,t}] = E[(\mu(\Theta_i)-E[X_{i,t}])E[X_{i,t}-E[X_{i,t}]|\Theta_i]]$
$= Var[\mu(\Theta_i)] = \lambda$

Rem. 4.2.8: (iii), (iv) $\Longrightarrow Cov[X_{i,t}, X_{i,s}] = \begin{cases} \lambda+\varphi, & \text{if } t=s \\ \lambda, & \text{if } t\neq s \end{cases}$

$\Longrightarrow \Sigma_{X_i}^{-1}$ exists $\Longleftrightarrow \varphi > 0$ (exercise)

Th. 4.2.9 ($\hat{\mu}_{LB}$ in the Bühlmann model)

Assume the Bühlmann model for $Var[X_{i,t}]<\infty$, $i,t\geq 1$ and $\varphi \overset{\text{def}}{=} E[V(\Theta_i)] > 0$. Then the linear Bayes estimator $\hat{\mu}_{LB}$ of $\mu(\Theta_i) \overset{\text{def}}{=} E[X_{i,t}|\Theta_i]$ w.r.t. the class $\mathcal{L}$ in (4.2.5) is uniquely given by

$$\hat{\mu}_{LB} = (1-w)\mu + w\overline{X}_i, \qquad (4.2.15)$$

where $$w := \frac{n_i \lambda}{\varphi + n_i\lambda} \quad \leftarrow \lambda \overset{\text{def}}{=} Var[\mu(\Theta_i)] \qquad (4.2.16)$$

and $$\overline{X}_i := \frac{1}{n_i}\sum_{t=1}^{n_i} X_{i,t}.$$

The risk of $\hat{\mu}_{LB}$ is given by

$$\varrho(\hat{\mu}_{LB}) = (1-w)\lambda \qquad (4.2.17)$$