(33) (3.4.21) $\longrightarrow$ <u>interpretation</u>:
$\varphi(u)$ distr. of a compound geometrically distr. r.v. with M for $q = (1+\rho)^{-1}$ and $X_i' = X_{I,i}$

$\longrightarrow$ <u>Th. 3.4.13</u> (Cramér's ruin bound for subexponential $F_{X_1,I}$)
Assume the Cramér-Lundberg model for $S(t)$.
Require that $E[X_1] < \infty$ and that the NPC holds.
Further, suppose that F_{X_1} has a prob. density and that $F_{X_1,I}$ is subexponential. Then

$$\lim_{u \to \infty} \frac{\psi(u)}{\rho^{-1} \overline{F}_{X_1,I}(u)} = 1 \qquad (3.4.23)$$

<u>Rem. 3.4.14</u>:
(i) (3.4.23) $\longrightarrow$ $\psi(u)$ is asymptotically of the same order as $\overline{F}_{X_1,I}(u)$, which is subexponential
$\longrightarrow$ dangerous compared to the light-tailed case (Th. 3.3.1):

$$\psi(u) \sim C \cdot e^{-r u} \quad \text{for } u \to \infty$$

like the exponential distr.

<u>Spontaneous ruin</u> through single large claims is <u>characteristic</u> for <u>heavy-tailed</u> (i.e. subexponential) distr.

(ii) If X_1 is <u>regularly varying</u> with index $\alpha > 1$ (i.e. $\overline{F}_{X_1}(x) = x^{-\alpha} L(x)$ ← slowly varying)
then $F_{X_1,I}$ is <u>subexponential</u> (see exercises)

$\longrightarrow$ example: $\overline{F}_{X_1}(x) = \left(\frac{\kappa}{\kappa + x}\right)^{\alpha}$, $\alpha > 1$ $\left(\overset{(3.4.14)}{\Longrightarrow} E[X_1] < \infty\right)$
Pareto distr. $\longrightarrow$ $\overline{\overline{F}}_{X_1,I}$ regularly varying with index $\alpha - 1$
$\overset{\text{Prop. 3.4.6}}{\Longrightarrow}$ $F_{X_1,I}$ subexponential

(54) Proof of Th. 3.4.13 : Idea: Application of Prop. 3.4.11

$$\frac{\psi(u)}{\overline{F}_{X_{1,I}}(u)} \overset{\text{Prop. 3.4.11}}{=} \frac{\rho}{1+\rho} \sum_{n\geq 1} (1+\rho)^{-n} \frac{P(X_{I,1}+\ldots+X_{I,n}>u)}{\overline{F}_{X_{1,I}}(u)}$$

On the other hand, we know from (3.4.16) that

$$\lim_{u\to\infty} \frac{P(X_{I,1}+\ldots+X_{I,n}>u)}{\overline{F}_{X_{1,I}}(u)} = n, \qquad (*)$$

since $F_{X_{1,I}}$ is subexponential.

Further, (iii) in L.3.4.9 yields for all $\varepsilon>0$

$$\frac{P(X_{I,1}+\ldots+X_{I,n}>u)}{\overline{F}_{X_{1,I}}(u)} \leq \overset{\text{constant}}{K_\varepsilon} (1+\varepsilon)^n$$

Choose $\varepsilon<\rho$. Then

$$\sum_{n\geq 1} \underbrace{(1+\rho)^{-n} K_\varepsilon (1+\varepsilon)^n}_{=K_\varepsilon\left(\frac{1+\varepsilon}{1+\rho}\right)^n} < \infty$$

$$\overset{\text{dom. convergence}}{\Longrightarrow} \lim_{u\to\infty} \frac{\psi(u)}{\overline{F}_{X_{1,I}}(u)} = \frac{\rho}{1+\rho} \sum_{n\geq 1}(1+\rho)^{-n} \overbrace{\lim_{u\to\infty} \frac{P(X_{I,1}+\ldots+X_{I,n}>u)}{\overline{F}_{X_{1,I}}(u)}}^{\overset{(*)}{=}n}$$

Recall: $f(q) := \sum_{k\geq 0} q^k, \ |q|<1 \Longrightarrow$ $\qquad (**)$

$f'(q) = \sum_{k\geq 1} k q^{k-1}$

$f(q) = \frac{1}{1-q} \Longrightarrow f'(q) = \frac{1}{(1-q)^2}$ on $(-1,1)$

$$\underset{(**)}{\overset{q=(1+\rho)^{-1}}{\Longrightarrow}} \lim_{u\to\infty} \frac{\psi(u)}{\overline{F}_{X_{1,I}}(u)} = \frac{\rho}{(1+\rho)}(1+\rho)^{-1} \cdot \frac{1}{\left(1-\frac{1}{(1+\rho)}\right)^2} = \frac{\rho}{(1+\rho)^2}\frac{(1+\rho)^2}{\rho^2} = \frac{1}{\rho}$$

$\Longrightarrow$ proof.

4. Experience rating

In view of the risk assessment of claims in an insurance portfolio the objective of collective risk theory is the analysis of the order of magnitude of the total claim amount $S(t)$. Here it is assumed that the insurance portfolio is homogeneous, that is the portfolio consists of policies for similar risks (e.g. theft insurance for households or water damage insurance for one-family houses)

→ basic (collective risk) model (w.r.t. $S(t)$):

$$S(t) = \sum_{i=1}^{N(t)} X_i, \qquad (4.1)$$

where

(i) X_i (i-th claim size) i.i.d. (which reflects the homogeneous probabilistic structure in the portfolio)

(ii) $N(t) = \#\{i \geq 1 : T_i \leq t\}$, T_i i-th arrival time of claim X_i

(iii) $X_i, i \geq 1$ and $T_i, i \geq 1$ independent

→ Deficiency: premiums w.r.t. (4.1) are the same for all policyholders and calculated e.g. by means of the equivalence principle (i.e. $P_{net}(t) := E[S(t)]$) or the expected value principle (i.e. $P_{EV}(t) := (1+\overset{\text{safety loading}}{\delta}) E[S(t)]$)

→ unfair, since a driver with poor driving record pays the same as one with a good one

→ possible solution is to introduce an individual model for every policyholder, which takes into account the individual claim history, in view of premium calculation

→ basic idea in credibility theory, which goes back to H. Bühlmann

→ method:

4.1 Bayes estimation

In order to explain this method from Bayesian statistics we need to recall some additional notions and

(3) results w.r.t. conditional expectations/probabilities:

(i) $E[Y|\xi]$, $P(A|\xi)$: Y, ξ r.v's (or rand. vectors) ← e.g. stock price at time t

$\mathcal{A} := \sigma(\xi) :=$ smallest σ-algebra containing all events $\{\xi \in C\}$, C measurable set

$\longrightarrow$ $E[Y|\xi] := E[Y|\mathcal{A}]$, $P(A|\xi) := E[\mathbb{1}_A|\xi]$ (4.2)

← Def. 2.10

(ii) $E[Y|\xi = x]$, $P(A|\xi = x)$:

$g(x) = E[Y|\xi = x]$ is the unique (measurable) function s.t.

$$E[Y|\xi] = g(\xi) \qquad (4.3)$$

$E[Y|\xi = x]$ expected value of Y given $\xi = x$

$P(A|\xi = x) := E[\mathbb{1}_A|\xi = x]$ probability of A given $\xi = x$

(iii) $P(A|\xi)$ or $P(Y \in A|\xi)$ are in general not prob. meas. for all ω, however: There exists a regular conditional probability of Y given ξ $Q(\omega, A)$

1. for each $\omega \in \Omega$ $P^*(A) := Q(\omega, A)$ defines a prob. meas. (on $\mathbb{R}^n$)
2. for each A $X(\omega) := Q(\omega, A)$ is a r.v. (w.r.t. $\mathcal{A}$ in (4.2))
3. for each A: $P(Y \in A|\xi) = Q(\omega, A)$ with prob. 1

$\longrightarrow$ property: $$E[f(\underset{\in \mathbb{R}^n}{Y})|\xi] = \int_{\mathbb{R}^n} f(y)\, \underbrace{Q(\omega, dy)}_{\text{prob. meas. for each } \omega} \qquad (4.4)$$

(iv) $f_Y(y|\xi = x)$: If $Y = (Y_1, \ldots, Y_n)$, and

$$P(Y_1 \le y_1, \ldots, Y_n \le y_n|\xi = x) = \int_{-\infty}^{y_n} \cdots \int_{-\infty}^{y_1} g_x(z_1, \ldots, z_n)\, dz_1 \ldots dz_n \qquad (4.5)$$

(x-a.e.) for all $y_1, \ldots, y_n \in \mathbb{R}$ for a function g_x then $f_Y(y|\xi = x) := g_x(y)$ is called conditional density of Y given $\xi = x$

$\longrightarrow$ objective: Calculation of the premium of the i-th policy based on the claim history, that is the claim data realizations

$$X_{i,1} = x_{i,1}, \ldots, X_{i,n_i} = x_{i,n_i}$$

of the i-th policy (n_i sample size, $X_{i,j}$ j-th claim)

(37) For this purpose we need an individual model for the i-th policy w.r.t. claim sizes or claim numbers

$\longrightarrow$ heterogeneity model

idea: a random parameter Θ_i (heterogeneity parameter) is assigned to the i-th policy, whose outcomes are e.g. given by the different (possible) driving skills of the i-th policyholder

$\longrightarrow$ Def. 4.1.2 (Heterogeneity model)

(i) The i-th policy is desribed by $(\Theta_i, (X_{i,t})_{t\geq 1})$. Here Θ_i is the heterogeneity parameter and $X_{i,t}$ the claim size or claim number in the i-th policy in period t (month, year, ...)

(ii) $(\Theta_i, (X_{i,t})_{t\geq 1})$ i.i.d. ($\Longleftrightarrow$ for all $t_1 < \dots < t_m$ the r.v.'s $Y_i := (\Theta_i, X_{i,t_1}, \dots, X_{i,t_m})$, $i \geq 1$ are i.i.d)

(iii) Given Θ_i, the sequence $X_{i,t}$, $t\geq 1$ is i.i.d. with distr. $F(\cdot|\Theta_i)$, that is

1. for all $t_1 < \dots < t_m$:
$$P(X_{i,t_1} \leq x_1, \dots, X_{i,t_m} \leq x_m | \Theta_i) = \prod_{j=1}^{m} P(X_{i,t_j} \leq x_j | \Theta_i)$$
with prob. 1 for all $x_1, \dots, x_m \in \mathbb{R}$
2. $P(X_{i,t} \leq x | \Theta_i) = F(x|\Theta_i)$ with prob. 1 for all x

Rem.: Dependence of $X_{i,t}$, $t\geq 1$ in the i-th policy is possible, but $X_{i,t}$, $t \geq 1$ is identically distributed

$\longrightarrow$ What could be a reasonable premium of the i-th policy?

$$\longrightarrow \quad \mu(\Theta_i) := E[X_{i,1}|\Theta_i] \quad \left(\overset{(4.4)}{=} \int_{\mathbb{R}} y\, Q(\omega, dy)\right) \qquad (4.6)$$

reg. cond. prob. of $X_{i,1}$ given Θ_i

$\longrightarrow$ expected claim size in the i-th policy given the information on e.g. driving skills

$\longrightarrow$ problem: Θ_i is not known a priori

$\longrightarrow$ $\mu(\Theta_i)$ has to be estimated

$\longrightarrow$ reasonable estimate of $\mu(\Theta_i)$: Bayes estimator $\hat{\mu}_B$

$\longrightarrow$ best approximation to $\mu(\Theta_i)$ in the mean square sense from available data

$$X_i = (X_{i,1}, \dots, X_{i,n_i}), \quad i = 1, \dots, r \qquad (4.7)$$

or more mathematically: If $E[(\mu(\Theta_i))^2] < \infty$, $\hat{\mu}_B$ minimizes

$$\rho(\hat{\mu}) \overset{\text{def}}{=} E[(\mu(\Theta_i) - \hat{\mu})^2] \quad \text{for r.v.'s } \hat{\mu} \text{ w.r.t. } \hat{\mu} := g(X_1, \dots, X_r)$$

with $E[\hat{\mu}^2] < \infty$,

distance between $\mu(\Theta_i)$ and $\hat{\mu}$ — entirely of claim size information (4.8) where $\mathcal{B}(X_1,\dots,X_T)$ = smallest $\mathcal{B}$-algebra generated by information given by $\{X_1 \in C_1, \dots, X_T \in C_T\}$, C_j sets

→ Th. 4.1.3 (Representation of $\hat{\mu}_B$)

The minimizer of the risk $\varrho(\hat{\mu})$ in (4.8) exists and is given by

$$\hat{\mu}_B = E[\mu(\theta_i) | X_i]$$

with corresponding risk

$$\varrho(\hat{\mu}_B) = E[Var[\mu(\theta_i) | X_i]]$$

L. 4.1.4: Let X be a r.v. w.r.t the $\mathcal{B}$-alg. $\mathcal{G}$ with $E[X^2] < \infty$. Let $\mathcal{A} \subseteq \mathcal{G}$ $\mathcal{B}$-algebra. Then $Y := E[X | \mathcal{A}]$ is the unique minimizer of $E[(X-Y)^2]$ for r.v.'s Y w.r.t. $\mathcal{A}$ s.t. $E[Y^2] < \infty$.

Proof of Th. 4.1.3: It follows from Th. 4.1.3 for $\mathcal{A} = \mathcal{B}(X_1,\dots,X_T)$ that $\hat{\mu}_B = E[\mu(\theta_i) | X_1,\dots,X_T] \overset{X_j,\, j \neq i \text{ indep. of } \mu(\theta_i)}{=} E[\mu(\theta_i) | X_i]$ (see exercises)

On the other hand, $\varrho(\hat{\mu}_B) = E[(\mu(\theta_i) - E[\mu(\theta_i)|X_i])^2] \underset{\text{property (ii) of cond. expect.}}{=} E[E[(\mu(\theta_i) - E[\mu(\theta_i)|X_i])^2 | X_i]] = E[Var[\mu(\theta_i)|X_i]]$

$\Longrightarrow$ proof.

The following result is useful for the computation of $\hat{\mu}_B$:

L. 4.1.5 (Calculation of $f_{\theta_i}(y | X_i = x)$)

Assume that θ_i in Def. 4.1.2 has a density f_θ ($= f_{\theta_i}$) and that $f_{\theta_i}(y | X_i = x)$, $y \in \mathbb{R}$ exists for x in the support of X_i (i.e. in R_{X_i} = smallest closed set s.t. $P(X_i \in R_{X_i}) = 1$)

(i) If $X_{i,1}$ is a discrete r.v., then

$$f_{\theta_i}(y | X_i = x) = \frac{f_\theta(y)\, P(X_{i,1} = x_1 | \theta_i = y) \cdots P(X_{i,1} = x_{n_i} | \theta_i = y)}{P(X_i = x)} \qquad (4.9)$$

for all $y \in \mathbb{R}$, $x = (x_1, \dots, x_{n_i})$ s.t. $P(X_i = x) > 0$.

(ii) If (X_i, θ_i) has the joint density $f_{X,\theta}$ ($= f_{X_i,\theta_i}$), then

$$f_{\theta_i}(y | X_i = x) = \frac{f_\theta(y)\, f_{X_{i,1}}(x_1 | \theta_i = y) \cdots f_{X_{i,1}}(x_{n_i} | \theta_i = y)}{f_{X_i}(x_1, \dots, x_{n_i})} \qquad (4.10)$$

for all $y \in \mathbb{R}$, $x = (x_1, \dots, x_{n_i})$ s.t. $f_{X_i}(x_1, \dots, x_{n_i}) > 0$.

Proof: See T. Mikosch.

Ex. 4.1.6: Let $X_{i,t}$, $t \geq 1$ be claim numbers in Def. 4.1.2 with $P(X_{i,1} = k | \theta_i) = e^{-\theta_i} \frac{(\theta_i)^k}{k!}$, $k \geq 0$ (Poisson distr. with (random) intensity $\theta_i > 0$), where $\theta_i \sim \Gamma(\gamma, \beta)$, $\gamma, \beta > 0$, i.e.

$$f_{\theta_i}(x) = \frac{\beta^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-\beta x}, \quad x > 0$$

→ $$\hat{\mu}_B = \frac{\gamma + X_i}{\beta + n} \quad \text{with } X_i := \sum_{t=1}^{n_i} X_{i,t} \quad \text{(see exercises)}$$

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

$$\longrightarrow \hat{\mu}_L \approx f(x_0) + f'(x_0)(X_1 - x_0), \text{ if } R_n(x) \text{ small}$$
$$= a_0 + a_1 X_1 \qquad (4.2.4)$$

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