# MATH 4281 Risk Theory–Ruin and Credibility

Start of Module 3: Credibility Theory

March 16th, 2021

# Introduction

# Recall from the first lecture

Q1: What do you do when $L$ is equal to a sum of smaller RVs?
$\Rightarrow$ Module 1: Aggregate Loss Models

Q2: How do you introduce **time** to this model?
$\Rightarrow$ Module 2: Ruin Theory

Q3: How do I estimate the parameters of the model for $L$...if I don't have a nice heterogeneous sample?
$\Rightarrow$ Module 3: Credibility

# Simple/Classical Example: Auto Insurance

Say I am insuring auto losses. I want the net premium for an individual policy but...

- Lots of ways to segment drivers e.g. age, location, car make/model, education, climate etc...
- Every relevant subdivision creates smaller and smaller sub-samples.
- Very quickly I can start to run into a lack of data on each sub-sample. Not advisable to estimate using simple mean of sub-sample.
- How can I incorporate data from the total sample of all drivers?

# Credibility Theory

- Need to set a premium for different groups of insurance contracts when:
  1. there are reasons to believe that groups have different risks (heterogeneous), but there is only limited experience (data) for each group of contracts,
  2. But there is quite a lot of experience when combined with other contracts which are more or less related.

- Claim amounts $X_{jt}$ are known for group (or individual) $j = 1, 2, ..., J$ and time periods $t = 1, 2, ..., T$.

- How to find the optimal estimators of claims for the group for next period.

# Two extreme approaches

Premium for group $j$ can be based on two extreme positions:

1. Use overall mean $\overline{X}$ of the data [makes sense only if the portfolio is *homogeneous*]. → e.g. All Drivers in ON
2. Use the average $\overline{X}_j$ in group $j$ [makes sense only if the group is sufficiently large and arguably different from other groups]. → e.g. All drivers in a town or neighbourhood

Can we combine these in some way?

# Reconciling these approaches

- Around 1900, American actuaries got the idea to use a weighted average of these extremes as a compromise:

$$\text{Credibility Premium } = z_j \overline{X}_j + (1 - z_j)\,\overline{X},$$

  where $z_j$ is called the credibility factor representing the weight attached to individual data.

- The credibility weight will be a value between 0 and 1, with it being close to 1 if:
  - group $j$ is large enough; and/or
  - claims for the group are very predictable; and/or
  - the variability between the groups is very large.

# The problem

Assume:

- every risk $j$ in the collective is characterized by its individual risk profile $\theta_j \in \Theta$ that does not change over time and that we can't observe.
- $\Theta$ may be either qualitative (e.g. good/bad) or quantitative (e.g. average number of accidents per year).
- we have $T$ observations $X_{j1}, \ldots, X_{jT}$

We want to estimate

$$\mu(\theta_j) = E\left[X_{j,T+1} | \theta_j\right]$$

but $\theta_j$ is unknown to the insurer...

# Two random variables

1st RV

- It is obvious that the losses $X_{j1}, X_{j2}, \ldots$ are random and depend on $\theta_j$.

Assume

- ~~Given~~ $\theta_j$, the losses $X_{j1}, X_{j2}, \ldots$ are independent.
- Since the risk profile can't be observed, we will also model it as random[1].

2nd RV ; $\theta$

- Thus $\mu(\theta)$ becomes a random variable we will use Bayesian techniques to estimate.

[1] Another interpretation of probability i.e. a measure of belief or certainty

# Crash Course on Bayesian Estimation

# First we define some notation

- The prior distribution is denoted by a CDF $\Pi(\theta)$ and pdf $\pi(\theta)$.
- The likelihood function has CDF $F(x|\theta)$ and pdf $f(x|\theta)$.
- The probability of the data has CDF $F(x)$ and pdf $f(x)$.
- The posterior distribution has CDF $\Pi(\theta|x)$ and pdf $\pi(\theta|x)$.
- And they are all related through Bayes' Theorem:

$$\pi(\theta|x) = \frac{f(x \mid \theta)\pi(\theta)}{f(x)}$$

$$= \frac{f(x \mid \theta)\pi(\theta)}{\int_\phi f(x|\phi) d\phi}$$

$$= \frac{f(x \mid \theta)\pi(\theta)}{\int_\phi f(x \mid \phi)\pi(\phi) d\phi}$$

# Estimation

Let $\theta$ be a variable we want to estimate

- we don't know the value of $\theta$
- it is drawn out of a population distributed like $\Pi(\theta)$

We want to create an estimator $\hat{\delta}$ of $\theta$.

- What criteria should it respect?
- For example: unbiased
- and..?

# The concept of loss function

Estimation error:

- when we estimate something, we (almost surely) make an error
- of course, we want to minimize that error
- are there errors we dislike more than others?
- (for example it might be better to overestimate a loss than underestimate it)

$\Rightarrow$ The loss function $L(\theta, \hat{\delta})$

# Loss functions

The loss function $L(\theta, \hat{\delta})$:

estimator

- is a function of $\theta$ and $\hat{\delta}$
- reflects the weight we want to give to estimation errors
- is the function we want to minimize
- minimization (of its expectation) yields the associated estimator

# The absolute error (or deviation) loss function

The function:

$$L(\theta, \hat{\theta}) = |\theta - \hat{\delta}| \tag{1}$$

The idea:

- the importance of the error is proportional to the distance between $\theta$ and $\hat{\delta}$
- positive errors and negative errors have the same weight (symmetrical)

# The quadratic (MSE) loss function

The function:

$$L(\theta, \hat{\theta}) = (\theta - \hat{\delta})^2 \tag{2}$$

The idea:

- the farther $\hat{\delta}$ is from $\theta$, the (exponentially) worse it is
- positive errors and negative errors of identical magnitude have the same weight (symmetrical)

# Almost constant loss functions

The function:

$$L(\theta, \hat{\theta}) = \begin{cases} c & \hat{\delta} \neq \theta \\ 0 & \hat{\delta} = \theta \end{cases} \tag{3}$$

The idea:

- the result of the estimation is binary: right or wrong
- if it is wrong, the cost is $c$

When $c = 1$, the above loss function is also called an **all-or-nothing** loss function.

# Other loss functions

- any loss function could be used
- the only restriction is the creativity of the user
- (of course, some restrictions just make sense – such as $L(\theta, \hat{\delta})$ positive for all $\theta$)

For example,

$$L(\theta, \hat{\delta}) = \begin{cases} \alpha(\hat{\delta} - \theta) & \hat{\delta} > \theta \\ \beta(\theta - \hat{\delta}) & \hat{\delta} < \theta \end{cases} \tag{4}$$

- the cost is $\alpha$ times the error if $\theta$ is overestimated and $\beta$ times the error if $\theta$ is underestimated
- this is a generalization of the absolute error loss function, which is the case $\alpha = \beta = 1$

# The Risk Function

- $L(\theta, \hat{\delta}(\mathbf{x}))$, the loss function, if $\theta$ is the "true" parameter and $\hat{\delta}(\mathbf{x})$ is the value taken by the estimator if $\mathbf{X}$ is observed

- The risk function of the estimator $\hat{g}$

$$R_{\hat{\delta}}(\theta) := E_{\mathbf{X}|\theta}[L(\theta, \hat{\delta}(\mathbf{X}))] = \int_{\mathbb{R}^n} L(\theta, \hat{\delta}(\mathbf{x}))\, \mathrm{d}F_{X|\theta}(\mathbf{x}|\theta) \quad (5)$$

# Frequentist Risk function

- In the case of frequentist inference we (can!) use the empirical distribution for our likelihood.
- In the case of the MSE we have:

Empirical pdf: $d\hat{F} = \frac{1}{n}\sum_{i=1}^{n} \delta(x - x_i)$ where $\{x_i\}_{i=1}^{n}$ is our sample

$L(\theta, \hat{\delta}) = (\theta - \hat{\delta})^2$ In the EDF-x is the parameter itself i.e. theta=x!

$$\int L(\theta; \delta)\, d\hat{F} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \delta)^2 = \overline{x^2} - 2\bar{x}\delta + \delta^2$$

$$= (\bar{x} - \delta)^2$$

minimized when $\delta = \bar{x}$

# Bayes Risk

However in the Bayesian case we augment our risk function with our prior:

$$
\begin{aligned}
R(\hat{\delta}) &= \int_{\Theta} R_{\hat{\delta}}(\theta)\, \mathrm{d}\Pi(\theta) = \int_{\Theta} E_{X|\theta}[L(\theta, \hat{\delta})]\, \mathrm{d}\Pi(\theta) \\
&= \int_{\Theta} \int_{\mathbb{R}^n} L(\theta, \hat{\delta}(\mathbf{x}))\, \mathrm{d}F(\mathbf{x}|\theta)\, \mathrm{d}\Pi(\theta) \\
&= \int_{\mathbb{R}^n} \left[ \int_{\Theta} L(\theta, \hat{\delta}(\mathbf{x}))\, \mathrm{d}\Pi(\theta|x) \right] \mathrm{d}F(\mathbf{x}). \qquad (6)
\end{aligned}
$$

Where in the last step we apply Bayes theorem to obtain the quantity in brackets.

This is all that's relevant for minimization

# Bayes estimator

The Bayes estimator is defined such that Bayes risk $R(\hat{\delta})$ is minimal:

$$\widetilde{\delta} \;:\; \hat{\delta} \mid R(\hat{\delta}) \text{ is minimal} \tag{7}$$

From (6) it follows that given $\mathbf{X} = \mathbf{x}$, $\widetilde{g(\mathbf{x})}$ takes the value which minimizes the error

$$\int_{\Theta} L(\theta, \hat{\delta}(\mathbf{x}))\, \mathrm{d}\Pi(\theta|x).$$

In other words,

- $\widetilde{\delta(\mathbf{x})}$ is the estimator which minimizes Bayes risk
- $\widetilde{\delta(\mathbf{x})}$ is the estimator which minimizes the expected loss with respect to the posterior distribution of $\theta$
- $\widetilde{\delta(\mathbf{x})}$ is "the best" estimator with respect to the loss function

- **Example 1** The Bayes estimator under the quadratic loss is the mean of the posterior distribution, i.e.

$$\widetilde{\delta(\mathbf{x})} = \int \theta d\Pi(\theta|x) = E[\theta|x]$$

**Example 2** The Bayes estimator under the absolute error loss is

the median of the posterior distribution, i.e.

$$\widetilde{\delta(\mathbf{x})} = \theta \mid \Pi(\theta|x) = 0.5$$

**Example 3** The Bayes estimator under the all-or-nothing loss is the mode of the posterior distribution, i.e.

$$\widetilde{\delta(\mathbf{x})} = \theta \mid \Pi(\theta|x) \text{ is maximal}$$

# Exercise

Prove example 1. In other words, show that the Bayes estimator under the quadratic loss is the mean of the posterior distribution.

We want to find $\arg\min_{\delta(x)} \left\{ \int_\Theta L(\theta, \hat{\delta}(x)) \pi(\theta|x) d\theta \right\}$

if $L(\theta, \hat{\delta}) = (\theta - \delta)^2$ then

$$\int L \pi d\theta = \int (\theta - \delta)^2 \pi(\theta|x) d\theta = \int [\theta^2 - 2\theta \delta(x) + (\delta(x))^2] \pi(\theta|x) d\theta$$

$$= \underbrace{\int \theta^2 \pi(\theta|x) d\theta}_{\text{Not relevant for min}} - 2\delta(x) \int \theta \pi(\theta|x) d\theta + \delta(x)^2$$

$$\underset{\delta(X)}{\arg\min} \left\{ \int L(\theta, \delta)\, d\pi(\theta|x) \right\}$$

$$= \underset{\delta(x)}{\arg\min} \left\{ -2\delta(x) E[\theta|x] + (\delta(x))^2 \right\}$$

$$\Longrightarrow \delta^*(x) = E[\theta | X] = \int \theta \, d\pi(\theta|x)$$

# Bayes estimate for the mean

Assume that:

- $\theta \sim \mathcal{N}(\mu, \tau)$
- $x|\theta \sim \mathcal{N}(\theta, \sigma)$
- What is the Bayes estimate for the mean given some sample of $x_i$ for $i = 1 \ldots n$?

You can show (Exercise) that:

$$\theta | x \sim \mathcal{N}\left( \frac{\frac{n}{\sigma^2}\bar{x} + \frac{\mu}{\tau^2}}{n/\sigma^2 + 1/\tau^2}, \frac{1}{\frac{n}{\sigma^2} + \frac{1}{\tau^2}} \right)$$

what this means

$$\delta^*(X) = \underbrace{\left( \frac{1/\sigma^2}{1/\sigma^2 + 1/\tau^2} \right)}_{\text{"}z\text{"}} \bar{X} + \underbrace{\left( \frac{\frac{1}{\tau^2}}{1/\sigma^2 + 1/\tau^2} \right)}_{\text{"}1-z\text{"}} \mu$$

Note

$n \to \infty$ or $\sigma \to 0$ $\qquad \delta^* \to \bar{X}$

$\frac{1}{\tau} \to \infty$ $\qquad \delta^* \to \mu$

$(\tau \to 0)$