第 5 讲：分类费率厘定 1

高光远

中国人民大学 统计学院

主要内容

1. 风险分类
   - 精算角度
   - 经营角度
   - 社会和法律角度
2. 确定性分类费率厘定模型
   - 单变量分析法
   - 迭代法

- 总平均费率厘定 (overall indication) 从整体上保证了保费的充足性. 但它没有考虑不同个体间的风险差异.
- 对于大部分保险产品, 个体的经验损失数据很少, 需要把具有相近潜在损失的个体划分为一个风险集合, 并对其进行费率厘定, 称为分类费率厘定 (classification rating).
- 对个体风险进行分类的特征通常称为分类变量或费率因子 (rating factor).
- 风险分类是通过分类变量将所有个体风险划分为若干个风险集合的过程.

注: 这里的分类变量指**风险分类变量**和统计学上的分类变量 (categorical variable) 不同.

- 风险分类有助于克服个体损失数据不足的问题.
- 风险分类的目的是获得相对同质 (homogeneous) 的风险子集.
- 分类费率厘定有利于消除逆选择 (anti-selection)和道德风险 (morality risk).
- 选择分类变量时, 要从精算、经营、社会和法律等不同角度进行评价.

Statistical significance, Homogeneity, Credibility

- 通过建立统计模型，评估不同风险类别的潜在损失是否具有显著的统计差异 (statistical significance).
- 每个风险类别中个体的潜在损失足够接近 VS 每个风险类别中的风险单位数足够多，即

同质性 (homogeneity) VS 可靠性 (credibility)

市场机制和逆选择

恰当地选择分类变量有利于消除逆选择 (anti-selection)

例:
A 类被保险人的潜在损失为 100,
B 类被保险人的潜在损失为 200.
保险公司 1 没发现 AB 的差异性, 使用费率 150.
保险公司 2 发现了 AB 的差异性, 使用费率 100(A) 和 200(B).
结果, A 类被保险人更倾向于购买保险公司 2 的产品, B 类被保险人更倾向于购买保险公司 1 的产品.
最后, 保险公司 1 **亏损**, 保险公司 2 **盈利**.

Objective, Inexpensive to administer, Verifiable

- Objective: 分类变量的定义要清晰明了.
- Inexpensive to administer: 分类变量的度量成本不能过高.
- Verifiable: 分类变量容易度量, 不易受到人为操纵, 避免道德风险.

分类变量的度量成本

例:
A 类潜在损失 800,
B 类潜在损失 1000,
C 类潜在损失 1200,
度量该分类变量的费用为 200.
如果不使用该分类变量, 每类的保险成本为 1000.
如果使用该分类变量, 则 A, B, C 的保险成本为 1000, 1200, 1400, 都高于 1000.
保险公司和被保险人都不愿使用该分类变量.

Affordability, Causality, Controllability, Privacy

- Affordability: 在强制性保险中, 风险分类需要考虑到所有被保险人对费率的承受能力.
- Causality: 分类变量和潜在损失有直接的因果关系.
- Controllability: 理想的分类变量满足可控性, 即被保险人可以控制自己所属的风险集合.
- Privacy: 分类变量不宜涉及个人隐私.
- 很多国家的法律规定, 风险分类不能有种族歧视和性别歧视, 即使它们和潜在损失有显著的关系.

乘法模型和加法模型

假设有两个费率因子 x_1, x_2, 它们为分类变量, 分别有 I 和 J 个水平. 这两个分类变量可以把被保险人划分为 $I \times J$ 个风险集合. 计属于 $(i, j)_{i=1:I, j=1:J}$ 风险集合的被保险人风险单位数为 n_{ij}, 经验赔款为 C_{ij}, 经验纯保费为 $Y_{ij} = C_{ij}/n_{ij}$, 损失期望为 $\mu_{ij} = \mathbb{E}(Y_{ij})$.

乘法模型 (multiplicative model) 假设

$$\mu_{ij} = \alpha_i \times \beta_j.$$

加法模型 (additive model) 假设

$$\mu_{ij} = \alpha_i + \beta_j.$$

乘法模型和加法模型

- 基于以上的模型, 可以用 $I+J$ 个相对费率$(\alpha_i)_{i=1:I}$, $(\beta_j)_{j=1:J}$ 对 $I\times J$ 个风险集合进行分类费率厘定.
- 以上的模型有一个隐含假设是: 费率因子 x_1, x_2 之间没有交互作用 (interaction effect).

例

原则：**每次**仅计算**一个**分类变量的不同水平所对应的相对费率.

Table 1: 汽车保单的经验赔付率

	车型 1		车型 2		Total	
	Exposure	Loss Ratio	Exposure	Loss Ratio	Exposure	Loss Ratio
地区 A	2000	40%	8000	80%	10000	72%
地区 B	8000	40%	2000	80%	10000	48%
Total	10000	40%	10000	80%	12500	60%

表1给出的赔付率是基于总平均费率, 即每个风险集合都采用一样的总平均费率:

$$\text{赔付率} = \frac{\text{赔款}}{\text{风险单位数} \times \text{总平均费率}}$$

例 (续)

地区 A 的赔付率为 72%, 地区 B 的赔付率为 48%. 所以总平均费率对于地区 A 显得过少, 对于地区 B 显得过多.

以地区 A 车型 1 为基础类别, 地区 B 的相对费率应该为 $0.48/0.72 = 0.6667$, 车型 2 的相对费率应该为 $0.8/0.4 = 2$. 这样可得到如下的相对费率:

Table 2: 应用单变量分析法计算的相对费率

	车型 1	车型 2
地区 A	1	2
地区 B	0.6667	1.3334

例 (续): 结果讨论

- 相对费率厘定的理想目标是: 使得每个风险集合中的赔付率相等.
- 使用表2的相对费率计算新的赔付率. A1 的赔付率为 0.4, A2 的赔付率为 0.8/2=0.4, B1 的赔付率为 0.4/0.6667=0.6000, B2 的赔付率为 0.8/1.3334=0.6000.
- 使用表2的相对费率, 并没达到我们的理想目标. 原因是: 地区 A 比地区 B 有更多的"高风险"车型 2, 这使得地区 A 的赔付率远高于地区 B 的赔付率. 所以 0.48/0.72 不能完全归结于 A 和 B 的差异, 它还包含了在 A 和 B 地区车型分布的差异.
- 实际上, 地区 AB 无差异, 车型 2 与车型 1 的相对费率为 2.

例 (续)：单变量分析法适用的情形

Table 3: 单变量分析法适用的情形

	车型 1		车型 2		Total	
	Exposure	Loss Ratio	Exposure	Loss Ratio	Exposure	Loss Ratio
地区 A	2000	40%	8000	80%	10000	72%
地区 B	500	40%	2000	80%	2500	72%
Total	2500	40%	10000	80%	12500	72%

- 在地区 A 和 B, 车型 2 的风险单位数都为车型 1 的两倍.
- 赔付率 72% 和 72% 的差异完全归结于地区分类变量, 即地区的相对费率为 1.
- 同理, 车型 2 的相对费率为 2.

- 迭代法 (iteration method)又称作最小偏差法 (minimum bias method).
- 迭代法可以克服单变量分析法的缺点.
- 迭代法的关键在于建立迭代公式.
- 根据迭代公式的含义和性质, 可分为
    1. 直接法
    2. 边际总和法
    3. 最小卡方法
    4. 最小二乘法
    5. 极大似然法
    6. 加权边际总和法

Setting

假设有两个费率因子 x_1, x_2, 它们为分类变量, 分别有 I 和 J 个水平. 这两个分类变量可以把被保险人划分为 $I \times J$ 个风险集合. 计属于 $(i,j)_{i=1:I,j=1:J}$ 风险集合的被保险人风险单位数为 n_{ij}, 经验赔款为 C_{ij}, 经验纯保费为 $Y_{ij} = C_{ij}/n_{ij}$, 损失期望为 $\mu_{ij} = \mathbb{E}(Y_{ij})$.

乘法模型 (multiplicative model) 假设

$$\mu_{ij} = \alpha_i \times \beta_j.$$

加法模型 (additive model) 假设

$$\mu_{ij} = \alpha_i + \beta_j.$$

直接法

原理: $\alpha_i = \frac{\mathbb{E}(Y_{ij})}{\beta_j} = \mathbb{E}\left[\frac{Y_{ij}}{\beta_j}\right]$, $\beta_j = \frac{\mathbb{E}(Y_{ij})}{\alpha_i} = \mathbb{E}\left[\frac{Y_{ij}}{\alpha_i}\right]$. 这两个式子可用 $Y_{ij}/\beta_j, Y_{ij}/\alpha_i$ 的加权平均进行估计, 权重为 n_{ij} (风险单位数越多, 经验损失越能反映潜在损失).

迭代公式:

$$\begin{aligned}
\alpha_i &= \frac{\sum_{j=1}^{J} n_{ij} \frac{Y_{ij}}{\beta_j}}{\sum_{j=1}^{J} n_{ij}}, && \text{for } i = 1, \ldots, I \\
\beta_j &= \frac{\sum_{i=1}^{I} n_{ij} \frac{Y_{ij}}{\alpha_i}}{\sum_{i=1}^{I} n_{ij}}, && \text{for } j = 1, \ldots, J
\end{aligned} \tag{1}$$

边际总和法 (marginal total method)

原理: 每个费率因子的不同水平计算的纯保费之和等于对应的经验赔款.

$$\begin{aligned} \sum_{j=1}^{J} C_{ij} &= \sum_{j=1}^{J} n_{ij}\alpha_i\beta_j, \qquad \text{for } i = 1,\dots,I \\ \sum_{i=1}^{I} C_{ij} &= \sum_{i=1}^{I} n_{ij}\alpha_i\beta_j, \qquad \text{for } j = 1,\dots,J \end{aligned} \tag{2}$$

上面 $I+J$ 个方程等价于

$$\begin{aligned} \alpha_i &= \frac{\sum_{j=1}^{J} C_{ij}}{\sum_{j=1}^{J} n_{ij}\beta_j}, \qquad \text{for } i = 1,\dots,I \\ \beta_j &= \frac{\sum_{i=1}^{I} C_{ij}}{\sum_{i=1}^{I} n_{ij}\alpha_i}, \qquad \text{for } j = 1,\dots,J \end{aligned} \tag{3}$$

边际总和法 (marginal total method)

- 给定初始值 $\beta_j^0 = 1, j = 1, \ldots, J$. 通过迭代(3), 可以得到方程组(2)的解.
- 可以假设 $\mu_{ij} = \mu\alpha_i\beta_j$, 这里 μ 表示总平均费率, 即

$$\mu = \frac{\sum_{ij} C_{ij}}{\sum_{ij} n_{ij}}$$

- 加法模型的迭代公式如下:

$$\begin{aligned}
\alpha_i &= \frac{\sum_{j=1}^{J} C_{ij} - n_{ij}\beta_j}{\sum_{j=1}^{J} n_{ij}}, \qquad \text{for } i = 1, \ldots, I \\
\beta_j &= \frac{\sum_{i=1}^{I} C_{ij} - n_{ij}\alpha_i}{\sum_{i=1}^{I} n_{ij}}, \qquad \text{for } j = 1, \ldots, J
\end{aligned} \tag{4}$$

最小卡方法

原理: 估计相对费率使得如下的目标函数 (或损失函数) 最小.

$$\arg\min_{\alpha,\beta} \sum_{ij} \frac{n_{ij}(Y_{ij} - \alpha_i\beta_j)^2}{\alpha_i\beta_j}$$

- 上述目标函数类似于 χ^2 检验中的统计量.
- 上述目标函数关于 α, β 求导, 可得如下迭代公式

$$\begin{aligned} \alpha_i &= \left[\frac{\sum_{j=1}^{J} \frac{n_{ij}Y_{ij}^2}{\beta_j}}{\sum_{j=1}^{J} n_{ij}\beta_j}\right]^{1/2}, \qquad \text{for } i = 1, \ldots, I \\ \beta_j &= \left[\frac{\sum_{i=1}^{I} \frac{n_{ij}Y_{ij}^2}{\alpha_i}}{\sum_{i=1}^{I} n_{ij}\alpha_i}\right]^{1/2}, \qquad \text{for } j = 1, \ldots, J \end{aligned} \tag{5}$$

最小卡方法

求证: 保费总额不会小于经验赔款总额, 即

$$\sum_{j=1}^{J} n_{ij}\alpha_i\beta_j \geq \sum_{j=1}^{J} C_{ij} \text{ and } \sum_{i=1}^{I} n_{ij}\alpha_i\beta_j \geq \sum_{i=1}^{I} C_{ij}$$

证明:

$$\alpha_i^2 = \frac{\sum_{j=1}^{J} \frac{n_{ij}Y_{ij}^2}{\beta_j}}{\sum_{j=1}^{J} n_{ij}\beta_j} = \sum_{j=1}^{J}\left[\frac{n_{ij}\beta_j}{\sum_{h=1}^{J} n_{ih}\beta_h}\left(\frac{Y_{ij}}{\beta_j}\right)^2\right]$$

假设随机变量 $U_j = \frac{Y_{ij}}{\beta_j}$, 其 pmf 为 $\Pr(U_j) = \frac{n_{ij}\beta_j}{\sum_{h=1}^{J} n_{ih}\beta_h}$. 所以

$$\alpha_i^2 = \mathbb{E}(U^2) \geq \mathbb{E}(U)^2.$$

最小卡方法

So

$$\alpha_i \geq \mathbb{E}(U) = \sum_{j=1}^{J} \left[\frac{n_{ij}\beta_j}{\sum_{h=1}^{J} n_{ih}\beta_h} \frac{Y_{ij}}{\beta_j} \right] = \frac{\sum_{j=1}^{J} n_{ij} Y_{ij}}{\sum_{j=1}^{J} n_{ij}\beta_j}$$

Reorganizing the above equation leads to

$$\sum_{j=1}^{J} n_{ij}\alpha_i\beta_j \geq \sum_{j=1}^{J} C_{ij}$$

最小二乘法

原理: 估计相对费率使得如下的目标函数 (或损失函数) 最小.

$$\arg\min_{\alpha,\beta} \sum_{ij} n_{ij}(Y_{ij} - \alpha_i\beta_j)^2.$$

迭代方程为

$$\begin{aligned} \alpha_i &= \frac{\sum_{j=1}^{J} C_{ij}\beta_j}{\sum_{j=1}^{J} n_{ij}\beta_j^2}, \qquad \text{for } i = 1, \ldots, I \\ \beta_j &= \frac{\sum_{i=1}^{I} C_{ij}\alpha_i}{\sum_{i=1}^{I} n_{ij}\alpha_i^2}, \qquad \text{for } j = 1, \ldots, J \end{aligned} \tag{6}$$

对于加法模型, 其等价于线性模型 (linear model).

极大似然法 (指数分布)

原理: 估计相对费率使得似然函数最大.

假设 C_{ij} 服从参数为 $n_{ij}\alpha_i\beta_j$ 的指数分布, 则全样本 $\mathcal{D}=(C_{ij})_{i=1:I,j=i:J}$ 的似然函数 (likelihood)为

$$L(\alpha,\beta|\mathcal{D})=\prod_{i,j}\frac{1}{n_{ij}\alpha_i\beta_j}\exp\left[-\frac{n_{ij}Y_{ij}}{n_{ij}\alpha_i\beta_j}\right].$$

其对数似然函数 (log likelihood)为

$$l(\alpha,\beta|\mathcal{D})=\log L(\alpha,\beta|\mathcal{D})=\sum_{ij}\left[-\ln n_{ij}-\ln\alpha_i-\ln\beta_j-\frac{Y_{ij}}{\alpha_i\beta_j}\right]$$

极大似然法 (指数分布)

上式关于 α, β 分别求偏导, 并令其等于零, 可得如下的迭代公式

$$\begin{aligned} \alpha_i &= \frac{1}{J}\sum_{j=1}^{J}\frac{Y_{ij}}{\beta_j}, \qquad \text{for } i = 1, \ldots, I \\ \beta_j &= \frac{1}{I}\sum_{i=1}^{I}\frac{Y_{ij}}{\alpha_i}, \qquad \text{for } j = 1, \ldots, J \end{aligned} \tag{7}$$

极大似然法 (伽马分布)

假设 C_{ij} 服从期望为 $n_{ij}\alpha_i\beta_j$ 的伽马分布. 可以证明如下的迭代公式:

$$\begin{aligned} \alpha_i &= \frac{\sum_{j=1}^{J} \frac{C_{ij}}{\beta_j}}{\sum_{j=1}^{J} n_{ij}}, \qquad \text{for } i = 1, \ldots, I \\ \beta_j &= \frac{\sum_{i=1}^{I} \frac{C_{ij}}{\alpha_i}}{\sum_{i=1}^{I} n_{ij}}, \qquad \text{for } j = 1, \ldots, J \end{aligned} \tag{8}$$

伽马分布的极大似然法等价于直接法(1).

加权边际总和法

式(2)等价于

$$
\begin{aligned}
\sum_{j=1}^{J} n_{ij}(\alpha_i\beta_j - Y_{ij}) = 0, &\qquad \text{for } i = 1, \ldots, I \\
\sum_{i=1}^{I} n_{ij}(\alpha_i\beta_j - Y_{ij}) = 0, &\qquad \text{for } j = 1, \ldots, J
\end{aligned} \tag{9}
$$

引入权重 ω_{ij}, 得如下加权边际总和法:

$$
\begin{aligned}
\sum_{j=1}^{J} \omega_{ij} n_{ij}(\alpha_i\beta_j - Y_{ij}) = 0, &\qquad \text{for } i = 1, \ldots, I \\
\sum_{i=1}^{I} \omega_{ij} n_{ij}(\alpha_i\beta_j - Y_{ij}) = 0, &\qquad \text{for } j = 1, \ldots, J
\end{aligned} \tag{10}
$$

加权边际总和法

前面的五种迭代方法都可以表述为特殊的加权边际总和法

1. 直接法: $\omega_{ij} = 1/(\alpha_i\beta_j)$
2. 边际总和法: $\omega_{ij} = 1$
3. 最小卡方法: $\omega_{ij} = 1 + Y_{ij}/(\alpha_i\beta_j)$
4. 最小二乘法: $\omega_{ij} = \alpha_i\beta_j$
5. 极大似然法 (伽马分布): $\omega_{ij} = 1/(\alpha_i\beta_j)$

1. 阅读教材 4.1- 4.3.
2. 自测课后习题.
3. 选做: 证明式(5).