DOI:10.19789/j.1004-9398.2019.01.001

首都师范大学学报(自然科学版)
第40卷 第1期 Journal of Capital Normal University No.1
2019年2月 (Natural Science Edition) Feb.,2019

Lovász 延拓权值下的伪泊松混合分布的风险期望模型*

陈奕延[1] 李 晔[1,2]

(1. 北京理工大学自动化学院,北京 100081;
2. 华北理工大学河北省数据科学与应用重点实验室,河北 唐山 063210)

摘 要

本文在风险损失量为自然数且服从泊松分布的条件下,将泊松分布进行截断和均化的处理生成伪泊松分布,然后根据有限可数混合分布的表达式,利用从集函数转换而来的多线性形式的 Pseudo-Boolean 函数的 Lovász 延拓得到新的权值并构建伪泊松混合分布,最后根据期望的定义和性质得到相应的伪泊松混合分布的风险期望模型.该模型为今后研究混合分布在风险分析中的应用提供了依据.

关键词: 风险损失额; 有限可数混合分布; Pseudo-Boolean 函数; Lovász 延拓
中图分类号: O211.3

0 引 言

传统的金融风险(本文以下简称"风险")测量方法包括历史模拟法、分析方法以及蒙特卡洛模拟法[1].历史模拟法的核心在于根据市场因子的历史样本变化模拟证券组合的未来损益分布,利用分位数给出一定置信水平下的 VAR 估计.历史模拟法是一种非参数方法,它不需要假定市场因子的统计分布,因而可以较好的处理非正态分布,缺点市场因子的历史样本变化与未来样本变化未必一致,因子变动假设可能不满足,且模拟结果受限于历史资料的期数,容易造成低信度风险值; 分析方法是 VAR 计算中最为常用的方法.它利用证券组合的价值函数与市场因子间的近似关系、市场因子的统计分布(方差-协方差矩阵)简化 VAR 计算,其缺点与历史模拟法一样,即过于依赖于历史数据,此外,由于对分布形式的特殊假定和灵敏度的局部特征,分析方法很难有效处理实际金融市场的厚尾性和大幅波动的非线性问题,往往产生各种误差和模型风险; 蒙特卡洛法,又称为统计试验法,该方法的思路是反复模拟决定金融估计价格的随机过程,每次模拟都可以得到组合在持有期末的一个可能值,如果进行大量的模拟则组合价值的模拟分布将收敛于组合的真实分布.这样通过模拟发布会可以导出真实分布,从而求出 VAR,蒙特卡洛法也有一些缺陷,比如对确定性问题需要转化成随机性问题处理,误差是概率误差,计算步骤复杂.此三类方法均基于 VAR 计算,则可统一简称为 VAR 方法,另外还有一些基于 VAR 的改进型模型,比如 TVP-VAR[2],Copula-VAR[3],La-VAR[4],WCVAR[5] 等.有别于传统的 VAR 方法,本文采用构建伪泊松混合分布模型的方法,通过混合分布的风险期望模型测量风险,避免了各型 VAR 方法过于依赖历史数据的缺陷及部分其它缺陷,从而为今后的风险测量问题提供了新的研究方向.

1 模型描述与假设

假设一组风险的损失量 $X_i \in \mathbf{N}$,且 $X_i \sim PN_i(\lambda_i)$,任意 X_i 之间相互独立,$PN_i(\lambda_i)$ 为参数 λ_i 的泊松分布,在此令 $\exists \tilde{\zeta}_{PN}=\{PN_1(\lambda_1),\cdots,PN_n(\lambda_n)\}$,其对于 $\forall g,h \in \mathbf{Z}^+$ 有 $\lambda_g \neq \lambda_h$,$\tilde{\zeta}_{PN}$ 是一个由独立不同参数的泊松分布组成的有限离散随机分布簇,PN 为英文 Poisson 的缩写,代表泊松分布.已知 X_i 对

收稿日期: 2017-10-15
* 河北省数据科学与应用重点实验室开放课题(20170320002)

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$\tilde{\zeta}_{PN}$中的任一元素 $PN_i(\lambda_i)$,$\lambda_i \in \mathbf{Z}^+$皆有不等的概率来自于它们,假设这种概率为 ρ_i,则 X_i 服从混合分布:

$$\overset{ME}{\langle PN \rangle} = \sum_{i=1}^{n} \rho_i PN_i(\lambda_i), \tag{1}$$

其中 $\sum_{i=1}^{n} \rho_i = 1$,$0 \leqslant \rho_i \leqslant 1$,则称 $\overset{ME}{\langle PN \rangle}$ 为有限可列的混合泊松分布,对 $\tilde{\zeta}_{PN}$ 中的任意泊松分布,均 $\exists X_i \sim PN_i(\lambda_i)$,$X_i \in \mathbf{N}$ 且 X_i 为离散型随机变量.其经济意义即风险损失量 X_i 有$(100\rho_i)\%$的概率来自泊松分布 $PN_i(\lambda_i)$,由前述可知这种组合是线性的,可根据性质求解相应的期望.

对于 ρ_i,本文希望得到 $\rho_i = f_i(X_i)$,$\mathbf{N}$ 为 X_i 的定义域,即找到对应法则 $f_i: \mathbf{N} \mapsto [0,1]$,使得 ρ_i 与随机变量 X_i 形成映射关系,这样可以通过一组 X_i 找到一组 ρ_i,使其在满足约束条件 $\sum_{i=1}^{n} \rho_i = 1$ 下生成混合分布表达式 $\overset{ME}{\langle PN \rangle}$,$\overset{ME}{\langle PN \rangle}$ 的表达式中可不再包含 ρ_i,只包含 X_i 或由 X_i 对应的函数及组成的数学结构,最后再求其期望即可得到相应的伪泊松混合分布关于风险损失量的风险期望模型.

2 建模顺序

首先在一系列约束条件的基础上对混合分布中的单个泊松分布进行截断和均化处理后得到伪泊松分布,然后定义其集函数(Set Function),写出相应的表达式后再转化为 Pseudo-Boolean 函数的多线性形式(Multilinear Form),然后利用莫比乌斯变换(Möbius Transform)得出 Pseudo-Boolean 函数的 Lovász 延拓(Lovász Extension),将其进行变换组合后作为新的权值来构建伪泊松混合分布,使得伪泊松混合分布的权值(结构系数)与其随机变量形成映射关系,最后根据期望的定义和性质,求出伪泊松混合分布的风险期望表达式,即为模型,建模顺序如下:

Step1.模型假设及约束条件;

Step2.对单个泊松分布进行截断处理;

Step3.在截断处理的基础上,再对其进行算术均化处理,得到伪泊松分布;

Step4.定义伪泊松混合分布的集函数;

Step5.根据集函数写出相应的多线性 Pseudo-Boolean 函数;

Step6.根据多线性 Pseudo-Boolean 函数,求出其 Lovász 延拓;

Step7.根据 Lovász 延拓的表达式构造伪泊松混合分布表达式的新权值;

Step8.根据新的权值,写出多元伪泊松混合分布模型的数学表达式;

Step9.根据伪泊松混合分布求得期望表达式,即伪泊松混合分布的风险期望模型.

3 模型设计及理论研究

(1) 泊松分布的截断处理.

$X_i \in \mathbf{N}$,X_i 在服从泊松分布的条件下其取值的个数是无穷多个,由于现实经济生活中风险损失量是有限的,其不可能存在无穷大的情况,这里需对原始泊松分布 $PN_i(\lambda_i)$ 进行截断处理.

截断(Truncate)即设定阈值来限制调查对象的范围,超过阈值的对象不做调查,但将其收入分布总体中用于保证随机分布的完整性.例如对收入调查而言,小于5 000元的用实际收入表示,大于5 000元不作调查,只是将其全部收入分布作为其一部分,这便是截断.

假设服从泊松分布的随机变量 X_i 的阈值为 t_i-1,则 $X_i = 0,1,2,\cdots,t_i-1$,令 $v_{ij}(x_{ij}) = x_{ij}+1$,则集合 $V_i = \{1,2,\cdots,t_i\}$,这里将 X_i 的取值 x_{ij} 自第 k_i 项以后截断,$k_i \in \mathbf{Z}^+$,$1 \leqslant k_i \leqslant t_i$,将第 k_i+1 项以后的各点的概率值 $p_{i(k_i+a_i)}$,$a_i \in \mathbf{Z}^+$ 累积到第 k_i 项上,此时假设 $X_i = x_{i1},x_{i2},\cdots,x_{ik},x_{ik} \in \mathbf{N}$,令 $x_{i1} < x_{i2} < \cdots < x_{ik_i}$,设 $\| X_i \|$ 为随机变量 X_i 中取值 x_{ij} 的个数,$\| \{x_{i1},x_{i2},\cdots,x_{ik_i}\} \| = k_i \leqslant \| X_i \| = \| V_i \| = t_i$.假设 $\sum P_{(-k_i)} = \sum_{a_i=1}^{\infty} p_{i(k_i+a_i)}$,则 x_{ik_i} 的概率值为 $p^*_{ik_i} = p_{ik_i} + \sum P_{(-k_i)}$,$p_{ij} = \frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i}$,$0 \leqslant x_{ij} \leqslant t_i - 1$,此时 X_i 的分布函数为:

$$P_i(X_i = x_{ij}) = \begin{cases} \dfrac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i}, & x_{ij} < x_{ik_i}, \\ \sum\limits_{j=k_i}^{\infty} \dfrac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i}, & x_{ij} = x_{ik_i}. \end{cases} \tag{2}$$

(2) 泊松分布的均化处理.

由公式(2)可知,因为原始泊松分布的取值有无穷多项,故完成截断处理后的分布在点 x_{ik_i} 处有一个相较前 k_i-1 个点而言非常巨大的概率值 $p^*_{ik_i}$,这会造成概率在某一点处的过度堆积以致于 $p^*_{ik_i} \to 1$,为方便作图起见,假设 $k_i \equiv k$,截断处理后的分布为

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$(PN_i)^{\Lambda}$,其概率分布示意图如图 1 所示,省略中间项的概率,只突出关键点的概率.

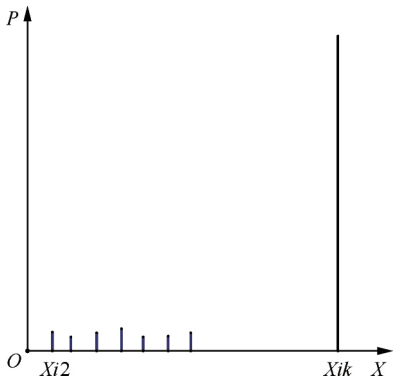


图 1　$(PN_i)^{\Lambda}$ 分布的概率分布示意图

如图可知截断处理后,$p_{ik}^{*} \gg P_i(X_i = x_{ij} \mid 1 \leqslant j \leqslant k-1)$,为保证随机波动的均匀性,须再次对其进行均化处理.

均化(Homogenization)原本是化学工业上的一个概念,狭义的均化特指通过采用一定的工艺措施,使物料化学成分的波动振幅降低,使物料的化学成分均匀一致的过程,这种方法多用于化工作业的操作中.这里的均化是广义的均化,特指为了防止离散随机波动过于极端而将极端的概率值进行处理后均匀(算术平均,几何平均等等)分散到其余各点的行为.在离散条件下,若有 l 个极端概率,则舍弃这 l 个概率值对应的点,将这 l 个点的概率累积后再进行均化处理.特别注意,若先进性截断处理后再进行均化处理,则被截断处理的点的概率未必是极端值,其累积也未必很大,这需要通过计算进行判断.因为泊松分布有无穷多项,从某一项 k_i 处进行截断处理,显然剩余的 $m-k_i$ 项依然是无穷多项,且 $\sum P_{(-k_i)}$ 是单调非减的,则显然可知 $\sum P_{(-k_i)} \gg p_{ik_i}$.

设 $\| \{x_{i1}, x_{i2}, \cdots, x_{ik}\} \|$ 为截断处理后随机变量 X_i 中取值 x_{ij} 的个数,由于采取了截断措施,故 $\| \{x_{i1}, x_{i2}, \cdots, x_{ik_i}\} \| = k_i \leqslant \| X_i \| = t_i$,将第 k_i 项以后(不包括第 k_i 项)的概率累积值 $\sum P_{(-k_i)}$ 按算术平均数平均分割成 k 份,每一份为 $p_i^{\hbar} = (\sum P_{(-k_i)})/k_i$,$p_i^{\hbar}$ 称为均化增量(Homogenized Increment)将 $p_i^{\hbar}$ 分别加到前 k_i 项概率值 $P_i(X_i = x_{ij})$ 上可得:

$$P_i(X_i = x_{ij}) = \frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i} + \frac{(\sum P_{(-k_i)})}{k_i}. \quad (3)$$

设均化处理后的离散随机分布为 $\Omega_i^{\tau}(\lambda_i)$,可知 $X_i \sim \Omega_i^{\tau}(\lambda_i)$,为便于行文,本文称 $\Omega_i^{\tau}(\lambda_i)$ 为伪泊松分布(Pseudo-Poisson Distribution).

截断处理和均化处理的目的是为了在原有泊松分布的基础上得到新的分布,即得到新的伪泊松分布.为使图中文字公式清晰便于查阅,这里更改一下表示符,设 $p_{ij} = p(i, j)$,具体操作从而生成伪泊松分布的流程如图 2 所示.

(3)有限可列的伪泊松混合分布.

原始泊松分布经过截断和均化处理后的分布 $\Omega_i^{\tau}(\lambda_i)$ 称为伪泊松分布,假设 $\exists\, \psi_{\tau}^{\sim} = \{\Omega_1^{\tau}(\lambda_1), \Omega_2^{\tau}(\lambda_2), \cdots, \Omega_n^{\tau}(\lambda_n)\}$ 为分布 $\Omega_i^{\tau}(\lambda_i)$ 的簇,对 $\forall g, h \in \mathbf{Z}^{+}$ 有 $\lambda_g \neq \lambda_h$,

其中 $\lambda_i \in \mathbf{Z}^{+}$,对簇中任一 $\Omega_i^{\tau}(\lambda_i)$,$\exists\, \Omega^{M}(\lambda_1, \lambda_2, \cdots, \lambda_n)$,使得:

$$\Omega^{M}(\lambda_1, \lambda_2, \cdots, \lambda_n) = \sum_{i=1}^{n} \rho_i^{M} \Omega_i^{\tau}(\lambda_i), \quad (4)$$

其中 $\sum_{i=1}^{n} \rho_i^{M} = 1, 0 \leqslant \rho_i^{M} \leqslant 1$.

(4)伪泊松分布的集函数.

集函数是计算机技术领域内常用的一类函数,国内外对其也有一定研究[6-11].设离散型随机变量 X_i 的全体取值为 V_i,根据前述定义,则 $V_i = \{1, 2, 3, \cdots, t_i\}$,若 A_i 为 V_i 的任意子集,则对 $\forall A_i$,$\exists A_i \subseteq V_i$ 使得 f_i^{δ} 为集函数,$f: 2^{V_i} \mapsto \mathbf{R}$,设 $i \in \mathbf{Z}^{+}$ 且 $1 \leqslant i \leqslant n$,假设空集的集函数 $f_i^{\delta}(\varphi) = 0$,继续定义其余子集的集函数,令:

$$f_i^{\delta}(\{x_{i1}, \cdots, x_{ik_i}\}) = \sum_{j=1}^{k_i} \frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i} + \frac{k_i(\sum P_{(-t_i)})}{t_i}, \quad (5)$$

这里 $\frac{(\sum P_{(-t_i)})}{t_i}$ 是对原始泊松分布进行截断处理后 t_i+1 项及其以后的概率累积的算术平均值,存在排序 $x_{i1} < x_{i2} < \cdots < x_{ik}$,$\| \{x_{i1}, x_{i2}, \cdots, x_{ik_i}\} \| = k_i \leqslant \| V_i \| = t_i$ 特别地,当 $k_i = t_i$ 时集函数有:

$$f_i^{\delta}(\{x_{i1}, x_{i2}, \cdots, x_{it_i}\}) = 1. \quad (6)$$

根据集函数的定义及性质,由此可知,随机变量 X_i 在定义域 2^{V_i} 上的集函数为:

$$f_i^{\delta}(A_i) = \begin{cases} 0, & A_i = \varphi \\ \sum_{j=1}^{k_i} \frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!} e^{-\lambda_i} + \frac{k_i(\sum P_{(-t_i)})}{t_i}, & A_i \neq \varphi \end{cases} \quad (7)$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

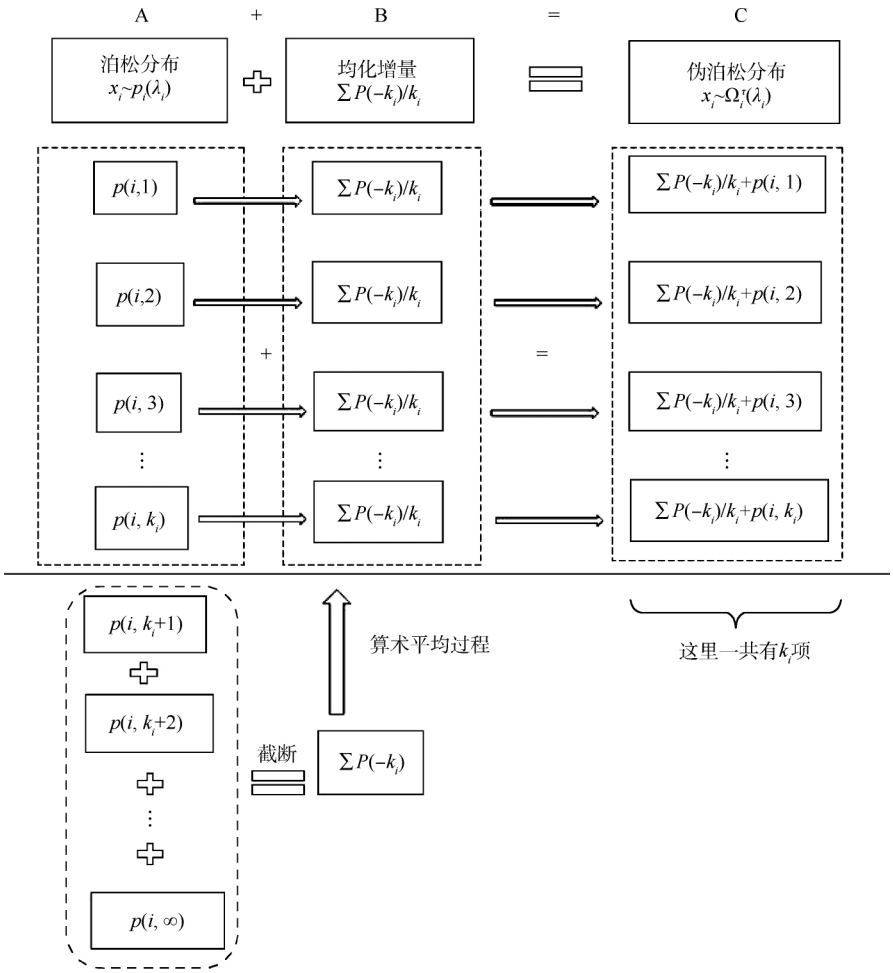


图 2 生成伪泊松分布的流程图

(5) **伪泊松分布的多线性** Pseudo-Boolean **函数.**

Pseudo-Boolean **函数也是一种计算机技术领域常用的函数**[12-14]**,其定义域为** 0-1 **变量且对应到实数域上的函数** $f:\mathbf{B}^n \mapsto \mathbf{R}$**,这里** $n=t_i$**,则相应的多线性**(Multilinear Form) Pseudo-Boolean **函数的表达式为:**

$$f_i^B(x_{i1},\cdots,x_{it_i}) = \sum_{A_i \subseteq V_i} \hat{f}_i^S(A_i) \prod_{j \in A_i} x_{ij}. \tag{8}$$

$\hat{f}_i^S(A_i)$ **是** $f_i^S(A_i)$ **的估计值,根据莫比乌斯变换**(Möbius Transform) **可知:**

$$\hat{f}_i^S(A_i) = \sum_{U_i \subseteq A_I} (-1)^{\|A_i\|-\|U_i\|} f_i^S(1_{U_i}). \tag{9}$$

$\|A_i\|$ **与** $\|U_i\|$ **分别代表集合** A_i **与** U_i **中元素的个数,**1_{U_i}**为指示向量,任意集函数都可以用其指示向量标记出来,设** $A_i \subseteq V_i 1_{A_i} \in \mathbf{B}^{t_i} = \{0,1\}^{t_i}$**,令未知函数为:**

$$1_{A_i}(x_{ij}) = \begin{cases} 1, x_{ij} \in A_i, \\ 0, x_{ij} \notin A_i. \end{cases} \tag{10}$$

例如对 $\{x_{ij}\} \subseteq V_i$ **其指示向量的表达式为:**

$$1_{\{x_{ij}\}} = \begin{pmatrix} 1_{A_i}(x_{i1}) \\ 1_{A_i}(x_{i2}) \\ \vdots \\ x_{ij} \\ \vdots \\ 1_{A_i}(x_{it_i}) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_{ij} \\ \vdots \\ 0 \end{pmatrix}. \tag{11}$$

由此可求出 1_{A_i}**的形式,根据定义,任意一个集函数** $f:2^{V_i} \mapsto \mathbf{R}$ **都可等价于一个** Pseudo-Boolean **函数** $f:\mathbf{B}^{t_i} \mapsto \mathbf{R}$**,即** $f(A_i) = f(1_{A_i})$**,这里为统一标记符号,表示为:**

$$f_i^S(A_i) = f_i^S(1_{A_i}). \tag{12}$$

将公式(12) **代入公式**(9)**,可得:**

$$\hat{f}_i^S(A_i) = \sum_{U_i \subseteq A_i} (-1)^{\|A_i\|-\|U_i\|} f_i^S(U_i). \tag{13}$$

然后将公式(13) **代入公式**(8) **中,可得伪泊松分布的多线性** Pseudo-Boolean **函数,函数表达式为公式**(14):

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$f_i^B(x_{i1},x_{i2},\cdots,x_{it_i}) = \sum_{A_i\subseteq V_i}\left[\sum_{U_i\subseteq A_i}(-1)^{\|A_i\|-\|U_i\|}f_i^S(U_i)\right]\prod_{j\in A_i}x_{ij}. \quad (14)$$

(6) **伪泊松分布的多线性** Pseudo-Boolean **函数的** Lovász **延拓.**

Lovász **延拓主要用于求解离散条件下很难处理的极值问题**[15-18]**.由定义可知:** Lovász **延拓的表达式可写为:**

$$f_i^L(x_{i1},\cdots,x_{it_i}) = \sum_{A_i\subseteq V_i}\hat{f}_i^S(A_i)\min_{j\in A_i}(x_{ij}). \quad (15)$$

此时 $f:[0,1]^{t_i}\mapsto\mathbf{R}$**,定义域从离散变成了连续区间** $[0,1]$**,对于** $f_i^B(x_{i1},x_{i2},\cdots,x_{it_i})$ **而言,其表达式中所含** $\prod_{j\in A_i}x_{ij}$ **与** $\min_{j\in A_i}(x_{ij})$ **等价,因为在** $f_i^B(x_{i1},x_{i2},\cdots,x_{it_i})$ **中,** x_{ij} **的值非** 0 **即** 1**,假设** $\exists x_{ig}=1,x_{ih}=0$**,则** $x_{ig}x_{ih}=0=x_{ih}=\min\{x_{ig},x_{ih}\}$**.这里可先求得** $\hat{f}_i^S(\varphi)=(-1)^{0-0}f_i^S(\varphi)=0$**,继续求单个元素的子集的集函数值,** $0\leqslant x_{ij}\leqslant t_i-1$**,则:**

$$\hat{f}_i^S(\{x_{ij}\}) = \sum_{U_i\subseteq\{x_{ij}\}}(-1)^{\|\{x_{ij}\}\|-\|U_i\|}f_i^S(U_i) = f_i^S(\{x_{ij}\}), \quad (16)$$

$$f_i^S(\{x_{ij}\}) = \frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!}e^{-\lambda_i}+\frac{\left(\sum P_{(-t_i)}\right)}{t_i}. \quad (17)$$

然后求多元素的子集的集函数值,先求 $\{x_{ig},x_{ih}\}=A_i$ **的情况,省略推导过程,可知:** $\hat{f}_i^S(\{x_{ig},x_{ih}\})=0$**,即** $\hat{f}_i^S(\{x_{ig},x_{ih}\})=0$**,依此类推,设** $w_i\geqslant 2$ **且** $w_i\in\mathbf{Z}^+$ **可知:**

$$\hat{f}_i^S(\{x_{i1},x_{i2},\cdots,x_{iw_i}\})=0. \quad (18)$$

故由上述计算结果可知:

$$\hat{f}_i^S(A_i)=\begin{cases}0, & \|A_i\|=0,\\ \dfrac{(\lambda_i)^{x_{ij}}}{(x_{ij})!}e^{-\lambda_i}+\dfrac{\left(\sum P_{(-t_i)}\right)}{t_i}, & \|A_i\|=1,\\ 0, & \|A_i\|\geqslant 2.\end{cases} \quad (19)$$

根据公式(19)的结果,将其代入公式(15),则 Lovász **延拓的表达式可写为:**

$$f_i^L(x_{i1},\cdots,x_{it_i}) = \sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{x_{ij}}}{(x_{ij})!}e^{-\lambda_i}+\frac{\left(\sum P_{(-t_i)}\right)}{t_i}\right]x_{ij}. \quad (20)$$

此时 $x_{ig}\in[0,1]$ $f:[0,1]^{t_i}\mapsto\mathbf{R}$**,为与之前符号区别表示,设** $z_{ig}\in[0,1]$**,则:**

$$f_i^L(z_{i1},\cdots,z_{it_i}) = \sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{z_{ij}}}{(z_{ij})!}e^{-\lambda_i}+\frac{\left(\sum P_{(-t_i)}\right)}{t_i}\right]z_{ij}. \quad (21)$$

根据 Gamma **函数的余元公式来求广义阶乘** $(z_{ij})!$**,可知:**

$$\Gamma(1-z_{ij})\Gamma(z_{ij})=\frac{\pi}{\sin(\pi z_{ij})}. \quad (22)$$

此时 $z_{ij}\in(0,1)$**,因为** $(z_{ij})!=\Gamma(z_{ij}+1)$**,则:**

$$(z_{ij})!=\frac{\pi(1+z_{ij})}{\sin(\pi z_{ij})\Gamma(1-z_{ij})}. \quad (23)$$

根据麦克劳林(Maclaurin**)公式,设** $b_{ij}=\pi z_{ij}\ m\in\mathbf{Z}^+$**,有:**

$$\sin b_{ij}=b_{ij}-\frac{(b_{ij})^3}{3!}+\frac{(b_{ij})^5}{5!}+\cdots+(-1)^{2m-1}\frac{(b_{ij})^{2m-1}}{(2m-1)!}+o\left[(b_{ij})^{2m-1}\right]. \quad (24)$$

将公式(24)代入(23),可得:

$$(z_{ij})!=\frac{\pi(1+z_{ij})}{\left\{b_{ij}-\dfrac{(b_{ij})^3}{3!}+\dfrac{(b_{ij})^5}{5!}+\cdots+(-1)^{m-1}\dfrac{(b_{ij})^{2m-1}}{(2m-1)!}+o\left[(b_{ij})^{2m-1}\right]\right\}\Gamma(1-z_{ij})}, \quad (25)$$

这里 $b_{ij}=\pi z_{ij},m\in\mathbf{Z}^+,z_{ij}\in(0,1)$**,**Gamma **函数的值可以查看** Gamma **函数表求得.对于** $z_{ij}=0$ **以及** $z_{ij}=1$ **时可直接根据公式代入求解.**

(7) Lovász **延拓下的伪泊松混合分布的权值及混合分布表达式**

根据公式(20**)可知,对于** n **个伪泊松混合分布而言,均存在其相应的** Lovász **延拓,对于任意随机变量** X_i **的取值** x_{ij}**,令** $z_{ij}=\dfrac{x_{ij}}{10^{\theta_{ij}}},\theta_{ij}\in\mathbf{N}$**,设** $0\leqslant\rho_i^M\leqslant 1$**,根据公式(**20**),这里令:**

$$\rho_i^M=\frac{\left\{\sum_{j=1}^{t_i}\left[\dfrac{(\lambda_i)^{(x_{ij}/10^{\theta_{ij}})}}{(x_{ij}/10^{\theta_{ij}})!}e^{-\lambda_i}+\dfrac{\left(\sum P_{(-t_i)}\right)}{t_i}\right](x_{ij}/10^{\theta_{ij}})\right\}}{\left\{\sum_{i=1}^{n}\sum_{j=1}^{t_i}\left[\dfrac{(\lambda_i)^{(x_{ij}/10^{\theta_{ij}})}}{(x_{ij}/10^{\theta_{ij}})!}e^{-\lambda_i}+\dfrac{\left(\sum P_{(-t_i)}\right)}{t_i}\right](x_{ij}/10^{\theta_{ij}})\right\}}. \quad (26)$$

$$\rho_i^M=\frac{f_i^L\left[(x_{i1}/10^{\theta_{i1}}),(x_{i2}/10^{\theta_{i2}}),\cdots,(x_{it_i}/10^{\theta_{it_i}})\right]}{\sum_{i=1}^{n}\left[(x_{i1}/10^{\theta_{i1}}),(x_{i2}/10^{\theta_{i2}}),\cdots,(x_{it_i}/10^{\theta_{it_i}})\right]}. \quad (27)$$

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

由于公式(27)是混合分布的权值,故 P_i^M 需要满足非负性与归一性.对于非负性而言,由于

$$\sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{(x_{ij}/10^{\theta ij})}}{(x_{ij}/10^{\theta ij})!}e^{-\lambda_i}+\frac{(\sum P_{-t_i})}{t_i}\right](x_{ij}/10^{\theta ij})\geqslant 0,$$

显然有 $0\leqslant\rho_i^M\leqslant 1$;对于归一性而言,显然有

$$\sum_{i=1}^{n}P_i^M=\sum_{i=1}^{n}\frac{f_i^L[x_{i1}/10^{\theta_{i1}}),x_{i2}/10^{\theta_{i2}}),\cdots,x_{it_i}/10^{\theta_{it_i}})]}{\sum_{i=1}^{n}f_i^L[x_{i1}/10^{\theta_{i1}}),x_{i2}/10^{\theta_{i2}}),\cdots,x_{it_i}/10^{\theta_{it_i}})]}=1,$$

由此可证明 $\sum_{i=1}^{n}\rho_i^M=1$ 由此可知,$0\leqslant\rho_i^M\leqslant 1$ 且 $\sum_{i=1}^{n}\rho_i^M=1$,设伪泊松混合分布为 $\Omega^M(\lambda_1,\lambda_2,\cdots,\lambda_n)$,则:

$$\Omega^M(\lambda_1,\lambda_2,\cdots,\lambda_n)=\sum_{i=1}^{n}\frac{\left\{\sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{(x_{ij}/10^{\theta ij})}}{(x_{ij}/10^{\theta ij})!}e^{-\lambda_i}+\frac{(\sum P_{(-t_i)})}{t_i}\right](x_{ij}/10^{\theta ij})\right\}}{\left\{\sum_{i=1}^{n}\sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{(x_{ij}/10^{\theta ij})}}{(x_{ij}/10^{\theta ij})!}e^{-\lambda_i}+\frac{(\sum P_{(-t_i)})}{t_i}\right](x_{ij}/10^{\theta ij})\right\}}\Omega_i^T(\lambda_i).\tag{28}$$

可知公式已达到研究目的,即得到函数关系 $f(X_i)=\rho_i^M$,使得随机变量与系数形成映射关系 $f:X_i\in\{0,1,2,\cdots,t_i-1\}\mapsto[0,1]$,即得到了随机变量 X_i 与 ρ_i^M 之间的映射.

(8)伪泊松混合分布的风险期望模型

假设 $\Omega^M(\lambda_1,\lambda_2,\cdots,\lambda_n)$ 的风险期望为 $E^M(X_1,X_2,\cdots,X_n|\lambda_1,\lambda_2,\cdots,\lambda_n)$,根据期望的定义,以及期望的运算性质由公式(28)可知伪泊松混合分布的风险期望模型的表达式为:

$$E^M(X_1,\cdots,X_n|\lambda_1,\cdots,\lambda_n)=\sum_{i=1}^{n}\frac{\left\{\sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{(x_{ij}/10^{\theta ij})}}{(x_{ij}/10^{\theta ij})!}e^{-\lambda_i}+\frac{(\sum P_{(-t_i)})}{t_i}\right](x_{ij}/10^{\theta ij})\right\}}{\left\{\sum_{i=1}^{n}\sum_{j=1}^{t_i}\left[\frac{(\lambda_i)^{(x_{ij}/10^{\theta ij})}}{(x_{ij}/10^{\theta ij})!}e^{-\lambda_i}+\frac{(\sum P_{(-t_i)})}{t_i}\right](x_{ij}/10^{\theta ij})\right\}}E_i^T(X_i|\lambda_i).\tag{29}$$

4 总 结

伪泊松混合分布的风险期望模型不像 VAR 方法一样依赖于历史数据,基于伪泊松混合分布的风险期望模型具备以下优点:(1)完善了离散型随机变量混合分布的情况;(2)根据经济意义,通过截断和均化处理,使分布相较原始泊松分布更加合理;(3)混合分布的权值是一个与随机变量构成映射的数值,减少了参数依赖;(4)考虑了混合分布中各分布随机变量取值不同的情况;(5)考虑了不同随机变量的不同阈值,便于灵活设定.但该模型也存在一些缺点,比如无法处理股票、证券、期货交易数据等连续型随机金融数据,只能解决类似车辆保险、企业个体经营风险、个体信用风险等离散型的随机时点或随机时期类金融数据,影响了模型的适用范围.另外,模型的层次结构较多,计算量较大,复杂度较高,这样无形之中给模型的使用带来了诸多不便.

未来可以基于模型的上述缺点对其进行改良,使之不断完善.该风险期望模型可在充分考虑各种约束及参数的条件下精确地求解离散型风险问题中的风险损失量的期望,为今后量化研究随机风险问题提供了参考依据.

参 考 文 献

[1] 袁岗. 我国商业银行交易账户市场风险与测度方法研究[D].南昌:江西财经大学,2012.

[2] 张绿原,陈燕武. 融资融券是否加剧了中国证券市场波动性:基于 TVP-VAR 模型的实证分析[J]. 金融理论与实践,2017,(4):72-76.

[3] 谢忆. 基于 Copula-VaR 模型的 G 证券公司 FICC 业务风险度量优化研究[D].上海:华东师范大学,2017.

[4] 赵茂林. 基于流动性风险 La-VaR 度量模型的股权质押率确定[J]. 财会月刊,2017,(14):119-123.

[5] 朱振星. 基于简化 WCVaR 模型的金融极端风险度量研究[D].合肥:合肥工业大学,2017.

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

[6] WANG D W. Fast hybrid level set model for non-homogenous image segmentation solving by algebraic multigrid [C] // Science And Engineering Research Center. Proceedings of 2016 International Conference on Electrical Engineering and Automation(ICEEA2016) .Xiamen: DEStech Publications,2016.
[7] JIANG D,HAN D F,HU X L. The shape optimization of the arterial graft design by level set methods [J] . Applied Mathematics: A Journal of Chinese Universities,2016,(2) : 205-218.
[8] 师肖静,张兴芳. 概率逻辑、不确定逻辑和模糊逻辑之比较 [J] . 计算机工程与应用,2017,(12) : 50-52,69.
[9] 钱玲,张慧,张化朋. 非可加测度空间上 Egoroff 定理的伪条件 [J] . 模糊系统与数学,2016,(1) : 77-83.
[10] 杨勇,潘伟民,徐春. 水平集函数规则化的 C-V 主动轮廓模型 [J] . 计算机工程与应用,2008,(34) : 166-168.
[11] 欧阳耀,李军. Choquet 积分定义的单调集函数的几个遗传性质 [J] . Journal of Southeast University,2003,(4) : 423-426.
[12] 李劲,岳昆,张德海,等. 社会网络中影响力传播的鲁棒抑制方法 [J] . 计算机研究与发展,2016,(3) : 601-610.
[13] ZHAO Y,CHENG D Z. Calculus of Pseudo-Boolean functions [C] //Technical Committee on Control Theory of Chinese Association of Automation. Proceedings of the 31th Chinese Control Conference (International) . Hefei: IEEE Xplore, 2012.
[14] JOAO M S. On applying lower bound estimates in Pseudo-Boolean optimization [C] //Guangzhou University Proceedings of Guangzhou Symposium on Satisfiability and Its Applications. Guangzhou: Sun Yat-Sen University,2004.
[15] SATORU F. Submodular functions and optimization [M] .Ameterdam: Elsevier Science & Technology, 2005: 253-274
[16] MA X L, WEN Q S. Multi-input multi-output (MIMO) detection systems [EB/OL] . [2017-07-11] .https: //smartech.gatecl.edu/handle/1853/59593.
[17] 邓素娟. 基于次模函数极小化的最优化问题 [J] . 内江师范学院学报,2016,31(12) : 29-32.
[18] 夏传甲,徐辉. GPS 降相关算法的比较研究 [J] . 全球定位系统,2012,37(5) : 66-70.

Risk Expectation Model in Pseudo-Poisson Mixture Distribution with Lovász Extension Weight

CHEN Yiyan[1] LI Ye[1,2]

(1.School of Automation, Beijing Institute of Technology, Beijing 100081;
2.Hebei Key Laboratory of Data Science and Application, North China University of Science and Technology, Tangshan Hebei 063210)

Abstract

Under the condition for risk loss values satisfied possion distribution and belong to natural numbers, poisson distributions will be truncated and homogenized to generate Pseudo-Poisson distributions. According to the representations of finite and countable mixture distributions, the multilinear Pseudo-Boolean functions, which converting from set functions, taking advantages of its Lovász extension to generate new weights and structure Pseudo-Poisson mixture distribution, finally, according to the definitions and characters of expectation to gain the relevant risk expectation model of Pseudo-Poisson mixture distribution. This model provide research directions for future studies of applications in risk analysis.

Keywords: risk losses; finite and countable mixture distributions; Pseudo-Boolean functions; Lovász extension

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net