arXiv:2304.02723v1 [stat.AP] 5 Apr 2023

无限定义域上的离散风险度量：理论基础、条件五数概括与数据分析

Daoping Yu[1]
威斯康星大学密尔沃基分校

Vytaras Brazauskas[2]
威斯康星大学密尔沃基分校

Ričardas Zitikis[3]
西安大略大学

摘要。为适应众多实际场景，本文将平滑分位数估计量的统计推断从有限域扩展到无限域。借助一种新设计的、针对具有无限域的离散损失分布的截断方法，我们完成了这一任务。一项模拟研究以泊松分布、负二项分布及其零膨胀版本等多个分布为例，展示了该方法的应用；这些分布通常用于保险行业中的理赔频率建模。此外，我们提出了一种非常灵活的基于自助法的方法，以便在实践中使用。利用汽车事故数据及其变体，我们计算了称为尾部风险条件五数概括（C5NS）的量，并为构成C5NS的五个分位数分别构造置信区间，随后计算尾部概率。结果表明，与线性插值方法相比，平滑分位数方法不仅能够更准确地对投资组合的尾部风险进行分类，而且在尾部概率估计中产生的变异系数也更低。

关键词：自助法；索赔次数；平滑分位数；风险价值；截断分布。

[1] Corresponding Author：Daoping Yu，博士、ASA，威斯康星大学密尔沃基分校数学科学系，美国威斯康星州密尔沃基。电子邮箱：dyu@uwm.edu[2] Vytaras Brazauskas，博士、ASA，威斯康星大学密尔沃基分校数学科学系，美国威斯康星州密尔沃基。电子邮箱：vytaras@uwm.edu[3] Ri˘cardas Zitikis，博士，西安大略大学数学与统计科学学院，加拿大安大略省伦敦市。电子邮箱：zitikis@stats.uwo.ca

1 引言与动机

风险价值（VaR）一直是保险和银行业中一种重要的风险度量（例如 BCBS， 2019）。在保险损失的情形下，损失为非负随机变量，水平为 $p \in (0, 1)$，其接近于 1，且通常由监管机构设定（例如 BCBS， 2019），是以不低于 p:

$$\mathrm{VaR}_p(Y) = F_Y^{-1}(p) = \inf\big\{y : F_Y(y) \geq p\big\},$$

其中， Y 是损失随机变量， F_Y 是其累积分布函数（c.d.f.）。对于取实值的损益（P&L）变量，类似的公式及其解释同样成立，只是此时损失位于实数轴的“负”侧。

显然，VaR 并未告诉我们尾部深处会发生什么。此外，如果相干性很重要，那么 VaR就不具备这一性质。另一方面，Artzner 等人，1999所称的期望损失（ES）也被称为条件尾部期望以及其他几个名称，是一种相干风险度量（Wang 和 Zitikis（2021），其首个同类公理化刻画由 p 提出。ES 为研究人员提供了一个急需的线索，帮助其了解在预先指定的概率水平 p 之外、VaR 之上的尾部可能正在发生什么。

虽然上述ES的性质很有吸引力，但有人可能会认为，由于损失分布通常具有偏斜性，（条件）期望并不是“总结”尾部的最佳方式，因此人们自然会想到使用（条件）中位数，正如统计文献所建议的那样。这样，作为水平 p处ES的替代，我们自然得到水平$(1+p)/2$处的VaR。这个VaR仍然只是一个参数，并不能提供对尾部更全面和令人满意的描述，这一事实已被许多作者注意到，例如以下引文:

> VaR 不会考虑置信水平之外的分布特征。这意味着，$\mathrm{VaR}_\alpha(X)$ 可能会随着 α的小幅增加而大幅上升。（Sarykalin 等人，2008，第 283 页）

我们建议使用一种向量值风险度量，称为条件五数概括（C5NS），其定义为

$$\mathrm{C5NS}_p(Y) = \Big(F_Y^{-1}(u_1), F_Y^{-1}(u_2), F_Y^{-1}(u_3), F_Y^{-1}(u_4), F_Y^{-1}(u_5)\Big),$$

其中

$$u_1 = 0.90p + 0.10, \quad u_2 = 0.75p + 0.25, \quad u_3 = 0.50p + 0.50,$$
$$u_4 = 0.25p + 0.75, \quad u_5 = 0.10p + 0.90.$$

这五个 u_i 分别产生分布中高于 $\text{VaR}_p(Y)$ 的 Y 的条件 10th、25th、50th、75th 和 90th 百分位数，从而相当充分地描述了损失变量 Y 的分布尾部。

为说明“条件”这一修饰语为何恰当，不妨以分位数$F_Y^{-1}(u_3)$为例。容易验证，当累积分布函数 F_Y 连续时，该分位数与我们所称的条件尾部中位数（CTM）一致。对于任意损失变量 Y 和任意概率水平 $p \in \in (0, 1)$，条件尾部中位数定义如下：

$$\text{CTM}_p(Y) = \text{median}\big(Y \mid Y > \text{VaR}_p(Y)\big).$$

然而，本文聚焦于离散损失随机变量，因此有必要区分 $F_Y^{-1}(u_3)$ 与 $\text{CTM}_p(Y)$，尽管从直观上讲，我们仍可方便地认为两者含义相同。

事实上，实践中研究人员常常会遇到离散损失随机变量（例如 Denuit 等，2007；Boucher 等，2009）。这些变量的累积分布函数呈阶梯状，由平坦区段和跳跃组成。现在不难看出，水平 p 即使只有无穷小的下降，也可能导致监管资本大幅减少；同样，水平 p 即使只有无穷小的上升，也可能导致监管资本大幅增加。这种对 u 的敏感性并不自然，可能对保险公司、监管机构或双方都造成严重不利影响。通过对 VaR 进行平滑处理可以解决这一问题，统计文献中早已有多种此类方法（Stigler，1977；Harrell 和 Davis，1982；Machado 和 Santos Silva，2005；Wang 和 Hutson，2011）。如今，数位专注于保险领域的研究人员已采用、改进并探索了这些方法（Boucher 等，2009；Alemany 等，2013；Bolanc´e 和 Guill´en，2021；Brazauskas 和 Ratnam，2023+）。

除了提出替代性风险度量 C5NS，以适应高度偏斜的分布之外，本文还提供了一种平滑技术，为在任意预先设定的概率水平下对 VaR 开展统计推断开辟了一条技术上便利的途径。更进一步地，该技术使研究者能够同时估计任意有限数量的 VaR，无论所选定或规定的概率水平为何，从而使研究者能够获得置信区间以及针对所提出的向量值风险度量 C5NS 的其他统计推断结果。

本文遵循 Wang 和 Hutson (2011) 提出的 Brazauskas 和 Ratnam (2023+)方法论，以充分解决当离散风险存在于无限域上时所产生的理论值挑战。此类分布（例如泊松、负二项）及其零膨胀版本通常用于对索赔频率建模。我们首先指出，本文提出的平滑技术不同于传统的基于核的方法（Alemany 等人，2013；Bolanc´e 和 Guill´en，2021），后者必须 假设带宽和核函数的形式。采用我们的方法，可以避免此类假设，因为分位数估计量的计算公式可直接由独立同分布随机变量顺序统计量的现有定理得出。此外，平滑分位数可以轻松转换为尾部概率，因此可用于尾部概率的估计。正如我们将在第 5 节看到的，与基于线性近似的估计（Klugman 等人，2012，第13.1 节）相比，此类“平滑”估计可将尾部估计的变异性降低多达 40-60%。

本文其余部分的组织如下。在第 2 节中，我们介绍截断方法的三部分设计，并为其奠定理论基础。在第 3 节中，我们使用常规泊松分布和负二项分布及其零膨胀版本开展模拟研究，以阐释所建立的理论。在第 4 节中，我们设计了一种对平滑分位数进行自助法处理的算法，从而再次验证我们的理论，同时也为近似更复杂的问题提供一种灵活的工具。在第 5 节中，我们展示了新方法在捕捉保险组合尾部风险方面的实际优势。最后，在第 6 节中，我们总结全文并给出结论性评论。

2 Smoothed Discrete Risks

2.1 Finite Domains

考虑一个离散随机变量 Y，其累积分布函数（c.d.f.）为 F_Y，概率质量函数（p.m.f.）为 $p_j = \mathrm{P}\{Y = y_{j:d}\}$，其中 $y_{j:d}$ 是 Y 可以取到的最小不同值。记作 $F_j := F_Y(y_{j:d}) = \sum_{i=1}^{j} p_i$，其中 $F_0 \equiv 0$。当 $\sum_{j=1}^{d} p_j = 1$ 且 $1 < d < \infty$（可能的不同值总数）时，该离散随机变量 Y 的平滑总体分位数函数定义为

$$Q_Y(u) \;=\; \sum_{j=1}^{d} \Big(B_{\alpha_u,\beta_u}(F_j) - B_{\alpha_u,\beta_u}(F_{j-1}) \Big) y_{j:d} \;=:\; \sum_{j=1}^{d} w_{j(u)}\, y_{j:d}, \tag{2.1}$$

其中 $B_{\alpha_u,\ \beta_u}$ 表示具有参数 $\alpha_u = (d+1)u$ 和$\beta_u = (d+1)(1-u)$ 的贝塔随机变量的累积分布函数。注意权重满足 $w_{j(u)} \geq 0$ 和 $\sum_{j=1}^{d} w_{j(u)} = 1$。为了获得直观

对定义的理解（2.1），我们参考Brazauskas 和 Ratnam（2023+）及其中的参考文献。

当获得独立同分布实现时 Y被获得，例如 y_1，…，y_n，则 $y_{1:d} < \cdots < y_{d:d}$表示具有相应频率的不同数据点 r_1，…，r_d。样本概率质量函数为 $\widehat{p}_i = r_i/n$，经验累积分布函数在 $y_{j:d}$处由 $\widehat{F}_j = \widehat{F}_Y(y_{j:d}) = \sum_{i=1}^{j} \widehat{p}_i = n^{-1}\sum_{i=1}^{j} r_i$给出。因此，平滑 u^{th} 离散数据分位数的样本估计量定义为通过将 F_j 替换为 $\widehat{F}_j$ 在定义（2.1）的 $Q_Y(u)$。这样我们得到估计量

$$\widehat{Q}_Y(u) = \sum_{j=1}^{d} \Big(B_{\alpha_u,\beta_u}(\widehat{F}_j) - B_{\alpha_u,\beta_u}(\widehat{F}_{j-1}) \Big) y_{j:d} =: \sum_{j=1}^{d} \widehat{w}_{j(u)} y_{j:d}, \tag{2.2}$$

其中，$F_0 = 0$和B_{α_u}，$_{\beta_u}$是贝塔累积分布函数，其 $\alpha_u = (d+1)u$和 $\beta_u = (d+1)(1-u)$。正如Wang 和 Hutson（2011，第 4.1定理）所证明的，$\widehat{Q}_Y(u)$一致估计 $Q_Y(u)$，并且渐近正态。此外，$\widehat{Q}_Y(u_1)$……，$\widehat{Q}_Y(u_l)$是一致的，并且联合渐近正态（Brazauskas 和 Ratnam，2023+，第 3.1 定理）。

2.2 Infinite Domains

无限情形 $d = \infty$包括精算研究中使用的许多相关分布。例如，泊松分布、负二项分布及其零膨胀版本通常用于对索赔频率建模。为了复现第 2.1 节中引入的估计量的设计和性质，Brazauskas 和 Ratnam (2023+，第 3.4 节）提出构造无限可数离散分布的截断版本。目标是找到一个有限区间，使大部分概率质量位于其中，然后模拟有限情形 $d < \infty$。

具体来说，将 Y的均值和方差分别记为 μ_Y和 σ_Y^2，假设 $\sigma_Y^2 < \infty$。定义区间

$$\big[L_k;\, U_k\big] := \big[\mu_Y - k\sigma_Y;\, \mu_Y + k\sigma_Y\big]. \tag{2.3}$$

根据切比雪夫不等式，Y 落入区间 (2.3) 的概率至少为$1 - 1/k^2$。此外，当用相应的样本版本 $\overline{Y}$ 和 S^2（基于大小为 n 的样本）估计 μ_Y 和 σ_Y^2 时，覆盖概率下界仍相当接近 $1 - 1/k^2$，且等于 $1 - \frac{1}{n+1}\left[\frac{n+1}{n}\left(\frac{n-1}{k^2} + 1\right)\right]$，其中 $[\cdot]$ 表示最大整数部分（见 Kab´an，2012）。例如，如果 $k = 5$，那么当 $n = 10$ 时，这个经验下界等于 0.909；当 $n = 50$ 时，等于 0.941；当 $n = 100$ 时，等于 0.950；而切比雪夫界等于 0.960。Brazauskas 和 Ratnam（2023+）在其模拟研究中考虑了 k 的多种取值，并推荐将 $k = 3$、4 或 5 作为最合理的实际取值。

考虑到这一点，不同的点的总数，以及最小和最大的不同的点

在定义（2.1）和（2.2）中使用的定义如下：

$$\text{Population:} \quad d_k = y_{d_k:d_k} - y_{1:d_k} + 1, \quad y_{1:d_k} = \max\left\{0, \left[L_k\right]\right\}, \quad y_{d_k:d_k} = \left[U_k\right] + 1, \tag{2.4}$$

$$\text{Sample:} \quad \widehat{d}_k = \widehat{y}_{\widehat{d}_k:\widehat{d}_k} - \widehat{y}_{1:\widehat{d}_k} + 1, \quad \widehat{y}_{1:\widehat{d}_k} = \max\left\{0, \left[\widehat{L}_k\right]\right\}, \quad \widehat{y}_{\widehat{d}_k:\widehat{d}_k} = \left[\widehat{U}_k\right] + 1, \tag{2.5}$$

其中 $[\cdot]$ 表示最大整数部分，L_k 且 U_k 根据定义给出（2.3），并且 $\widehat{L}_k$ 和 $\widehat{U}_k$分别是 L_k 和 U_k的样本版本，即将 μ_Y 替换为 $\overline{Y}$，并将 σ_Y^2 替换为 S^2。相应的截断分布为

$$F^*_{j(k)} = \mathbf{P}\left\{Y \le y_{j:d_k} \mid y_{0:d_k} < Y \le y_{d_k:d_k}\right\} = \frac{F_{j(k)} - F_{0(k)}}{F_{d_k(k)} - F_{0(k)}} \tag{2.6}$$

and

$$\widehat{F}^*_{j(k)} = \widehat{\mathbf{P}}\left\{Y \le \widehat{y}_{j:\widehat{d}_k} \mid \widehat{y}_{0:\widehat{d}_k} < Y \le \widehat{y}_{\widehat{d}_k:\widehat{d}_k}\right\} = \frac{\widehat{F}_{j(k)} - \widehat{F}_{0(k)}}{\widehat{F}_{\widehat{d}_k(k)} - \widehat{F}_{0(k)}}. \tag{2.7}$$

这种由量（2.4）–（2.7）指定的截断分布设计在实践中易于实现，但在考虑分位数估计量的理论值性质时难以处理。事实上，不清楚如何证明此类估计量渐近性质的猜想（见 Brazauskas 和 Ratnam， 2023+，猜想3.1），因为点（2.4）和（2.5）使用向下取整或“floor”运算，而截断分布（2.6）和（2.7）涉及在这些点不连续的累积分布函数。

因此，我们建议修订有限区间及相应截断分布的设计如下：

1. 使用与之前相同的 L_k 和 U_k的定义，以及 $\widehat{L}_k$ 和 $\widehat{U}_k$ 的定义。为确保它们始终产生非整数值，选择 k 一个无理数。例如，为了得到大于3的 k，我们可以考虑 $k = \pi \approx 3.1416$、$k = \pi^2 \approx 9.8696$或 $k = \pi^3 \approx 31.0063$，它们分别产生以下切比雪夫界的值：0.899、0.990和0.999。

2. 将点（2.4）和（2.5）替换为

$$y_{1:d_k} = \text{最小值}\left\{y_j \in \mathbb{Z} \mid y_j \in [L_k;\ U_k]\right\}, \qquad y_{d_k:d_k} = \text{最大值}\left\{y_j \in \mathbb{Z} \mid y_j \in [L_k;\ U_k]\right\}, \tag{2.8}$$

$$\widehat{y}_{1:\widehat{d}_k} = \min\left\{\widehat{y}_j \in \mathbb{Z} \mid \widehat{y}_j \in [\widehat{L}_k; \widehat{U}_k]\right\}, \qquad \widehat{y}_{\widehat{d}_k:\widehat{d}_k} = \max\left\{\widehat{y}_j \in \mathbb{Z} \mid \widehat{y}_j \in [\widehat{L}_k; \widehat{U}_k]\right\}. \tag{2.9}$$

此处 $d_k = y_{d_k:\,d_k} - y_{1:d_k} + 1$（总体）和 $d_k = \widehat{y}_{\widehat{d}_k:\,\widehat{d}_k} - \widehat{y}_{1:\widehat{d}_k} + 1$（样本）。请注意，$y_j$ 和 $\widehat{y}_j$ 表示不同的整数，它们可能为负数，也可能超出基础概率分布的支持集。幸运的是，这些情况不会造成问题。对于

例如，如果 $L_k < 0$，那么在 $y_{1:d_k}$、……、$y_{r:d_k}$ 上，对于某个 $1 \le r < d_k$，没有概率质量。这将导致在 $y_{1:d_k}$、……、$y_{r:d_k}$ 的定义（2.1）中，分配给点 $Q_Y(u)$ 的权重为零，并会将数量 d_k 减少到 $d_k - r$。类似的解释适用于 $L_k < 0$的情况。

3.替换截断分布 (2.6) 和 (2.7) 为

$$F^*_{j(k)} = \mathbf{P}\{Y \le y_{j:d_k} \mid L_k < Y \le U_k\} = \frac{F_{j(k)} - F_Y(L_k)}{F_Y(U_k) - F_Y(L_k)} \tag{2.10}$$

and

$$\widehat{F}^*_{j(k)} = \widehat{\mathbf{P}}\left\{Y \le \widehat{y}_{j:\widehat{d}_k} \mid \widehat{L}_k < Y \le \widehat{U}_k\right\} = \frac{\widehat{F}_{j(k)} - \widehat{F}_Y(\widehat{L}_k)}{\widehat{F}_Y(\widehat{U}_k) - \widehat{F}_Y(\widehat{L}_k)}. \tag{2.11}$$

其中，$F_{j(k)} = \mathbf{P}\{Y \le y_{j:d_k}\}$，$F_Y(L_k) = \mathbf{P}\{Y \le L_k\}$，且 $F_Y(U_k) = \mathbf{P}\{Y \le U_k\}$。对于这些函数的经验版本，我们使用 $\widehat{y}_{j:\widehat{d}_k}$、$L_k$ 和 U_k。注意，由于这些点是非整数，F_Y 在 L_k 处连续且，F_Y 在 L_k 处连续且可微。

遵循上述三部分设计，尤其是利用公式（2.8）和（2.10），我们定义平滑的 u^{th} 分位数

$$Q^{(k)}_*(u) = \sum_{j=1}^{d_k} \Big(B_{\alpha_u,\beta_u}(F^*_{j(k)}) - B_{\alpha_u,\beta_u}(F^*_{j(k)-1})\Big) y_{j:d_k} =: \sum_{j=1}^{d_k} w^{(k)}_{j(u)} y_{j:d_k} \tag{2.12}$$

对于截断离散总体，其中 B_{α_u,β_u} 表示贝塔累积分布函数，其参数为$\alpha_u = (d_k + 1)u$ 和 $\beta_u = (d_k + 1)(1 - u)$. 同样，使用公式 (2.9) 和 (2.11)，我们得到平滑的 u^{th} 分位数

$$\widehat{Q}^{(k)}_*(u) = \sum_{j=1}^{\widehat{d}_k} \Big(B_{\widehat{\alpha}_u,\widehat{\beta}_u}(\widehat{F}^*_{j(k)}) - B_{\widehat{\alpha}_u,\widehat{\beta}_u}(\widehat{F}^*_{j(k)-1})\Big) \widehat{y}_{j:\widehat{d}_k} =: \sum_{j=1}^{\widehat{d}_k} \widehat{w}^{(k)}_{j(u)} \widehat{y}_{j:\widehat{d}_k} \tag{2.13}$$

对于截断离散样本，其中$B_{\widehat{\alpha}_u,\widehat{\beta}_u}$是参数为 $\widehat{\alpha}_u = (d_k + 1)u$ 和 $\beta_u = (d_k + 1)(1 - u)$ 的贝塔累积分布函数。

重要的是，函数 $\widehat{Q}^{(k)}_*$和 $Q^{(k)}_*$与它们的未截断版本直接相关，$\widehat{Q}^{(k)}_Y$ 和 $Q^{(k)}_Y$，分别。例如，$Q^{(k)}_*$可以解释为 F^*的平滑版本的反函数，该平滑版本通过方程（2.10）与 F_Y 的类似平滑版本相关。对截断分布（2.10）进行反演，对于平滑的 F^* 和 F_Y 导致方程

$$Q^{(k)}_*(u) = Q^{(k)}_Y\Big(F_Y(L_k) + u\big(F_Y(U_k) - F_Y(L_k)\big)\Big) \tag{2.14}$$

将 $Q_*^{(k)}$与 $Q_Y^{(k)}$ 相关联，后者是平滑 F_Y的反函数。从公式（2.14）可以清楚地看出，如果F_Y（L_k）≈ 0 和 F_Y（U_k）≈ 1，则 $Q_*^{(k)}$（u）$\approx Q_Y^{(k)}$（u）。

在图2.1中，我们描绘了平滑分位数函数 $Q_Y^{(k)}$ (u) 以及泊松($\lambda = 9$) 和负二项($r = 9, \beta = 1$) 分布及其零膨胀版本以$c = 0.2$ 和 0.8 的离散分位数函数。参数 λ、r 和 β 已选择，使得两个分布具有相同的均值，但负二项的方差会大两倍。（关于这些分布的具体参数化，我们参考 Klugman 等人 (2012)。）平滑曲线使用数据截断区间 $\mu_Y \pm k\sigma_Y$ 构建，其中 $k = \pi$、π^2、π^3。从图中可以看出，k 的所有三种选择对于标准泊松和负二项分布都效果良好。然而，对于它们的零膨胀版本 ZIP 和零膨胀负二项分布，$k = \pi$ 的情况错过了分布的两个尾部，并且 $k = \pi^2$ 可以通过 $k = \pi^3$ 在从 0 到 1 、 2 、 3 或 4 的跳跃处显著改进。由于典型的索赔计数数据包含约 80% 的零（即 $c = 0.8$），我们建议使用量级为 $\pi^3 \approx 31$ 的 k。

2.3 Theoretical Foundations

其中 $Q_*^{(k)}(u)$ 和 $\widehat{Q}_*^{(k)}(u)$ 分别由方程（2.12）和（2.13）定义，向量

$$\left(Q_*^{(k)}(u_1), \ldots, Q_*^{(k)}(u_l)\right) \tag{2.15}$$

of smoothed 总体分位数 ($0 < u_1 < \cdots < u_l < 1$) 可以通过 vect 来估计 or

$$\left(\widehat{Q}_*^{(k)}(u_1), \ldots, \widehat{Q}_*^{(k)}(u_l)\right) \tag{2.16}$$

平滑后的 样本分位数。根据我们的下一个定理，后一个向量是向量（2.15）的一致且（联合）渐近正态估计量。

定理 2.1。 给定一个大小为 n 的离散分布 F_Y，其支持集为无限集$y_{1:d} < y_{2:d} < y_{3:d} < \cdots$，且 $d = \infty$，令其截断版本 $y_{1:d_k} < y_{2:d_k} < \cdots < y_{d_k:\ d_k}$ 其中 $d_k < \infty$使用第 2.2节的三部分设计构造。然后，当 $n \to \infty$时，

(i) $\left(\widehat{Q}_*^{(k)}(u_1), \ldots, \widehat{Q}_*^{(k)}(u_l)\right) \xrightarrow{\mathbf{P}} \left(Q_*^{(k)}(u_1), \ldots, Q_*^{(k)}(u_l)\right)$,

(ii) $\left(\widehat{Q}_*^{(k)}(u_1), \ldots, \widehat{Q}_*^{(k)}(u_l)\right) \sim \mathcal{AN}\left(\left(Q_*^{(k)}(u_1), \ldots, Q_*^{(k)}(u_l)\right), \frac{1}{n}\mathbf{HDH}'\right)$,

其中 $\mathbf{D} := [d_{ij}]_{(d_k-1)\times(d_k-1)}$，其中 $d_{ij} = d_{ji} = F_i^*(k)(1 - F_j^*(k))$，$i(k) \le j(k)$，且 $\mathbf{H} := [h_{ij}]_{l\times(d_k-1)}$，其中 $h_{ij} = (y_{j:d_k} - y_{j+1:d_k})\, b_{\alpha_{u_i},\ \beta_{u_i}}(F_j^*(k))$。这里，$b_{\alpha_{u_i},\beta_{u_i}}$是贝塔概率密度函数，其参数为 $\alpha_{u_i} = (d_k + 1)u_i$ 和 $\beta_{u_i} = (d_k + 1)(1 - u_i)$，而累积分布函数 $F_{j(k)}^*$ 由公式（2.10）定义。

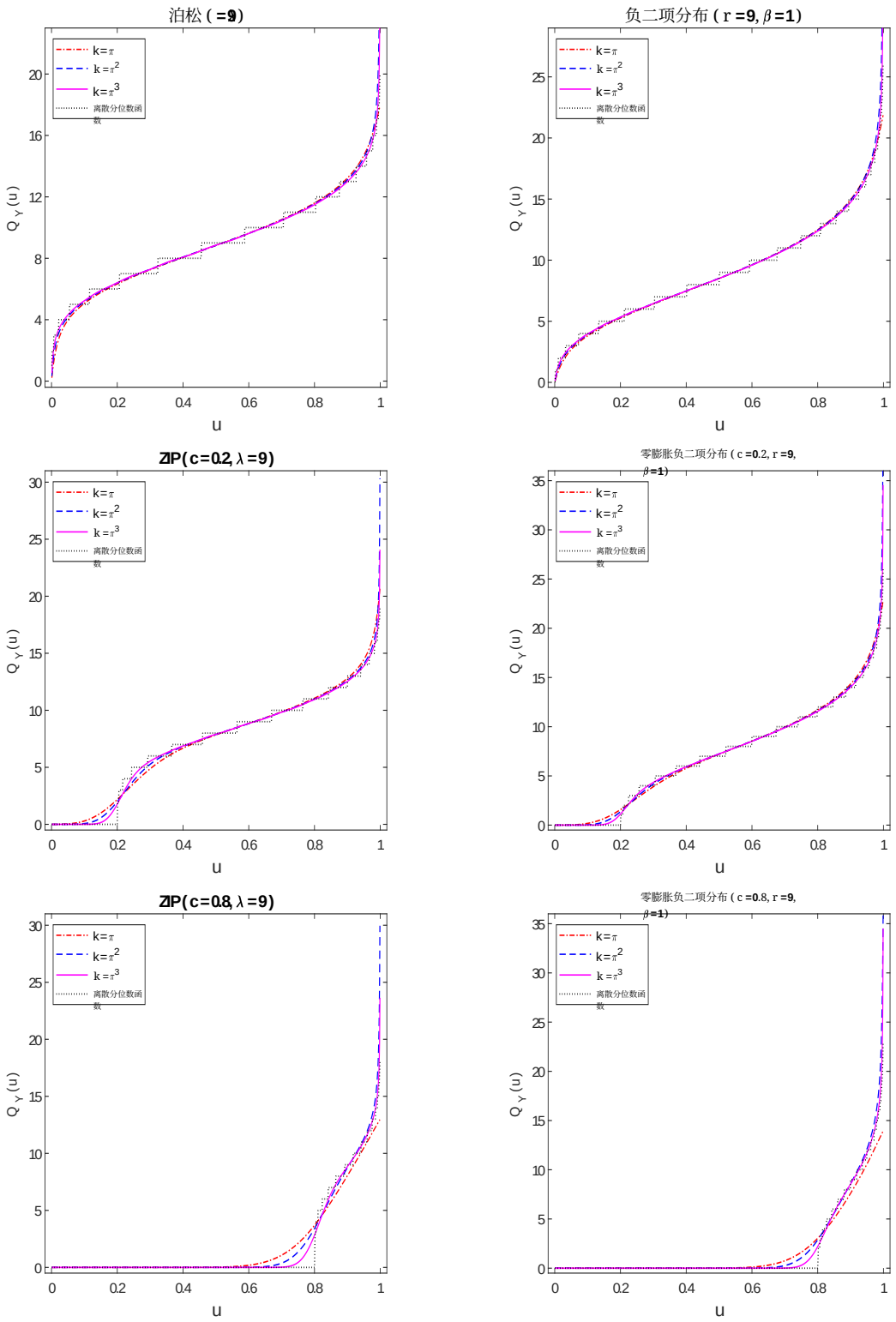

图2.1：泊松（左上）和负二项（右上）的离散（点状）和平滑分位数函数，以及相应零膨胀分布在数据截断区间 $\mu_Y \pm k\sigma_Y$ 内的分位数函数。

ˆ ˆ证明。 将 L_k 和 U_k 代入方程（2.9），用 L_k 和 U_k分别替换它们；由于它们是已知常数，我们得到方程（2.8）。类似地，修改后的方程（2.11）变为 $\widetilde{F}^*_j$（k） = $\widehat{F}_{j(k)}-\widehat{F}_Y(L_k)\ \widehat{F}_Y(U_k)$ $-\widehat{F}_Y(L_k)$ ’从而得到如下修改

$$\widetilde{Q}^{(k)}_*(u) \;=\; \sum_{j=1}^{d_k}\Big(B_{\alpha_u,\beta_u}(\widetilde{F}^*_{j(k)}) - B_{\alpha_u,\beta_u}(\widetilde{F}^*_{j(k)-1})\Big)y_{j:d_k}$$

估计量（2.16）。这构成了第 2.1 节的有限域情形。因此，向量估计量

$$\Big(\widetilde{Q}^{(k)}_*(u_1),\ldots,\widetilde{Q}^{(k)}_*(u_l)\Big)$$

满足陈述 (i) 和 (ii)。这可以通过遵循Brazauskas 和 Ratnam (2023+,定理 3.1)的证明来确立。首先注意到 $(\overline{Y},S^2)$ P$\longrightarrow$ (μ_Y,σ^2_Y) 意味着 $(\widehat{L}_k,\widehat{U}_k)$ P$\longrightarrow$ (L_k,U_k)，并且由于按照设计， $L_k\in($ $y_{1:d_k}-1;\ y_{1:d_k})$ 和 $U_k\in(y_{d_k:d_k};\ y_{d_k:d_k}+1)$ 都是累积分布函数 $\widetilde{F}^*_{j(k)}$ 和 $F^*_{j(k)}$ 连续的点，因此 $\widehat{F}^*_j$(k) P $\longrightarrow$ $F^*_{j(k)}$。现在注意到 $(\widehat{Q}^{(k)}_*(u_1)\ldots,\widehat{Q}^{(k)}_*(u_l))$ 是 $F^*_{1:d_k},\ldots,F^*_{d_k:d_k}$ 的ˆ ˆ连续变换。因此，连续映射定理的应用（例如，Serfling，1980，第 1.7 节）意味着

$$\Big(\widehat{Q}^{(k)}_*(u_1),\ldots,\widehat{Q}^{(k)}_*(u_l)\Big)\xrightarrow{\mathbf{P}}\Big(Q^{(k)}_*(u_1),\ldots,Q^{(k)}_*(u_l)\Big),$$

这就证明了定理的（i）部分。

为证明（ii），我们注意到已建立的结果

$$\Big(\widetilde{Q}^{(k)}_*(u_1),\ldots,\widetilde{Q}^{(k)}_*(u_l)\Big)\;\sim\;\mathcal{AN}\left(\big(Q^{(k)}_*(u_1),\ldots,Q^{(k)}_*(u_l)\big),\frac{1}{n}\mathbf{HDH}'\right).$$

由于 $\widehat{Q}^{(k)}_*(u)$ 是将 (L_k,U_k) 替换为其相合估计量 $(\widehat{L}_k,\widehat{U}_k)$ 后得到的 $\widetilde{Q}^{(k)}_*(u)$，广义 Slutsky 定理（Demidenko，2004，第 13.1.2 节）保证，估计量 $(\widetilde{Q}^{(k)}_*(u_1)\ldots,\widetilde{Q}^{(k)}_*(u_l))$ 和 $(\widehat{Q}^{(k)}_*(u_1)\ldots,\widehat{Q}^{(k)}_*$ $(u_l))$ 具有相同的渐近正态分布。至此，(ii) 部分的证明完成，并且定理 2.1 得以确立。

□

3 模拟数据示例

本节通过开展蒙特卡洛模拟研究，旨在说明定理 2.1 中确立的理论性质。我们首先介绍研究设计（第 3.1 节），然后提供泊松分布和负二项分布的汇总表及相关讨论（第 3.2 节），并介绍它们的零膨胀版本 ZIP 和零膨胀负二项分布（第 3.3 节）。

3.1 研究设计

研究设计基于以下选择。

模拟设计

- 离散分布。泊松($\lambda = 9$) 和 负二项($r = 9$, $\beta = 1$)。
- 零膨胀离散分布。零膨胀泊松分布、ZIP($\lambda = 1$, $c = 0.8$)，以及零膨胀负二项分布 ZINB($r = 1$, $\beta = 1$, $c = 0.8$)。
- 关键公式和参数（用于生成数据和计算理论目标）。– 均值： $\lambda = 9$ (泊松)、$r\beta = 9$ (负二项)、$\lambda(1-c)\frac{}{1-e^{-\lambda}} = 0.32$ (ZIP)，以及 $r\beta(1-c)\,1-(1+\beta)^{-r}$ =0.40 (零膨胀负二项分布)。– 方差： $\lambda = 9$ (泊松)、 $r\beta(1+\beta) = 18$ (负二项)、 $\frac{1-c}{1-e^{-\lambda}}\left(\lambda + \frac{\lambda^2}{(}\right.$ $\left.\frac{c-e^{-\lambda}}{1-e^{-\lambda}}\right) = 0.53$) (ZIP)，以及 $1-\overset{1-c}{}(1+\beta)^{-r}\left(r\beta(1+\beta) + (r\beta)^2(c-(1+\beta)^{-r}\right) = 1.04$)1–($1+\beta$) (零膨胀负二项分布)。– 常规零比例： $p_0 = e^{-\lambda} = 0.0001$ (泊松)、 $p_0 = (1+\beta)^{-r}$ =0.002 (负二项)、 $\overline{p_0(}\,{}_{1-c} = e^{-\lambda})\,(1-c)\frac{}{1-e^{-\lambda}} = 0.12$(ZIP)，以及 $p_0(1-c)$ $1-p_0 = (1+\beta)^{-r}(1-c)1-(1+\beta)^{-r}\overset{1-p_0}{} = 0.2$(零膨胀负二项分布)。– 过量零比例：0(泊松)、0 (负二项)、$\frac{c-p_0}{1-p_0} = \frac{c-e^{-\lambda}}{1-e^{-\lambda}} = 0.68$ (ZIP)，以及 $\frac{c-p_0}{1-p_0} = {}^{c-}(\frac{}{1+\beta\ -r})\,(1+\beta)_{-r} = 0.6$ (零膨胀负二项分布)。
- 截断区间。$\overline{Y} \pm k\,S$ 用于模拟数据和 $\mu_Y \pm k\sigma_Y$ 用于理论目标，带有 $k = \pi, \pi^2, \pi^3$。
- *Sample sizes.* $n = 10^2, 10^3, 10^4$.

对于任何给定的分布，我们生成 10,000 个指定长度的随机样本 n。对于每个样本，我们使用公式（$(Q_*0。25)$ Q_*（0。50） Q_*（0。75）$)$ 估计分布的四分位数，公式为（2.13）。随后，我们计算这10,000个四分位数估计值的均值和协方差-方差（$\times n$）矩阵。请注意，生成零膨胀数据需要一个两步过程。在第一步中，过量零部分从零值或非零值中生成，其中零的概率为$\frac{c-p_0}{1-p_0}$。在第二步中，使用相应的常规泊松或负二项分布生成剩余的零部分（以及所有其他正整数值），其概率现在等于 $p_0\ (1-c)\ 1-p_0$。上述两个零的比例之和为 c，即样本中零的总比例。

3.2 Poisson and NB Distributions

泊松和负二项分布的参数经过选择，以匹配图 2.1中绘制的模型。两个分布具有相同的均值，但负二项分布的方差是泊松分布方差的两倍。我们从表 3.1 中注意到估计均值和协方差-的快速收敛。

表3.1：在点 $\overline{Y} \pm k\,S$ 截断、参数为 $k = \pi$、 π^2、 π^3 的泊松($\lambda = 9$)和负二项($r = 9$, $\beta = 1$)分布的平滑四分位数估计量 $\widehat{Q}_*^{(k)}$(0.25)、 $\widehat{Q}_*^{(k)}$(0.50)、 $\widehat{Q}_*^{(k)}$(0.75) 的估计均值和协方差-方差($\times n$)矩阵。

	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = \infty$
Poisson ($\lambda = 9$), $k = \pi$				
$\widehat{means}$	(6.82, 8.84, 11.02)	(6.82, 8.84, 11.02)	(6.82, 8.84, 11.02)	(6.815, 8.835, 11.021)
$\widehat{\mathbf{HDH'}}$	11.24 8.51 5.91 8.51 11.63 10.10 5.91 10.10 16.00	11.64 8.65 5.78 8.65 11.79 10.20 5.78 10.20 16.29	11.42 8.39 5.54 8.39 11.45 9.67 5.54 9.67 15.45	11.367 8.360 5.539 8.360 11.497 9.753 5.539 9.753 15.478
泊松 ($\lambda = 9$), $k = \pi^2$				
$\widehat{means}$	(6.86, 8.84, 10.98)	(6.86, 8.84, 10.98)	(6.86, 8.84, 10.98)	(6.856, 8.838, 10.982)
$\widehat{\mathbf{HDH'}}$	12.05 8.38 5.71 8.38 12.28 9.73 5.71 9.73 16.52	12.42 8.58 5.69 8.58 12.45 9.95 5.69 9.95 17.00	12.20 8.34 5.57 8.34 12.23 9.62 5.57 9.62 16.51	12.153 8.309 5.526 8.309 12.289 9.714 5.526 9.714 16.579
泊松（$\lambda = 9$）， $k = \pi^3$				
$\widehat{means}$	(6.89, 8.84, 10.94)	(6.89, 8.85, 10.95)	(6.89, 8.85, 10.95)	(6.893, 8.853, 10.951)
$\widehat{\mathbf{HDH'}}$	12.90 8.20 5.61 8.20 13.17 9.55 5.61 9.55 17.79	11.32 7.55 5.08 7.55 12.00 9.01 5.08 9.01 16.91	10.61 7.09 4.83 7.09 11.45 8.49 4.83 8.49 15.77	10.533 7.033 4.695 7.033 11.401 8.415 4.695 8.415 15.631
负二项（$r = 9, \beta = 1$）， $k = \pi$				
$means$	(5.88, 8.51, 11.64)	(5.86, 8.50, 11.63)	(5.86, 8.50, 11.63)	(5.859, 8.504, 11.628)
$\widehat{\mathbf{HDH'}}$	18.66 15.35 11.46 15.35 23.08 21.45 11.46 21.45 36.85	18.48 15.12 11.22 15.12 23.17 21.48 11.22 21.48 37.13	18.10 14.53 10.21 14.53 22.35 20.27 10.21 20.27 35.13	18.038 14.458 10.384 14.458 22.085 20.054 10.384 20.054 34.815
负二项 ($r = 9$, $\beta = 1$), $k = \pi^2$				
$\widehat{means}$	(5.92, 8.52, 11.61)	(5.90, 8.52, 11.61)	(5.90, 8.52, 11.60)	(5.904, 8.515, 11.604)
$\widehat{\mathbf{HDH'}}$	20.08 15.10 11.04 15.10 24.18 20.33 11.04 20.33 37.06	19.84 14.90 10.84 14.90 24.36 20.41 10.84 20.41 37.45	19.51 14.46 10.25 14.46 23.91 20.10 10.25 20.10 37.35	19.552 14.467 10.507 14.467 23.833 20.212 10.507 20.212 37.975
负二项 ($r = 9$, $\beta = 1$), $k = \pi^3$				
$\widehat{means}$	(5.93, 8.51, 11.56)	(5.93, 8.51, 11.56)	(5.93, 8.50, 11.55)	(5.928, 8.504, 11.554)
$\widehat{\mathbf{HDH'}}$	21.65 14.95 10.99 14.95 25.66 20.11 10.99 20.11 39.37	19.16 14.41 10.20 14.41 27.72 20.75 10.20 20.75 40.10	17.74 13.80 9.44 13.80 28.35 20.84 9.44 20.84 40.31	17.673 13.777 9.675 13.777 28.408 20.920 9.675 20.920 40.813

NOTE: $n < \infty$ 的条目是估计四分位数的平均值和样本协方差。结果基于10,000个模拟样本。这些条目的标准误差为 ≤ 0.01。

方差（$\times n$）矩阵。即使是 $n = 100$ 对应的矩阵元素，也接近各自的理论目标，相关数值列于 $n = \infty$列中。此外，正如从图 2.1可以预料的那样，对于这些分布，选择 k 并不重要。我们建议选择 $k = \pi^2$ ，这对于大多数典型的离散分布来说已经足够（但不包括零膨胀分布）

情形）。最后请注意，对于泊松分布和负二项分布，估计量的估计均值看起来相似，而协方差-方差矩阵的元素则相差约两倍。这是预期之中的，原因在于泊松分布和负二项分布的参数选择不同。

3.3 ZIP and ZINB Distributions

ZIP 和零膨胀负二项分布的参数经过选择，以匹配图 2.1 （第三列）所描绘的模型。和 $c = 0.8$ 代表现实的保险数据场景。我们从表 3.2 中注意到，估计均值和协方差-的收敛

表 3.2：在截断点 $\overline{Y} \pm k\,S$ 处截断的 ZIP($\lambda = 1$, $c = 0.8$) 和 ZINB($r = 1$, $\beta = 1$, $c = 0.8$) 分布的平滑四分位数估计量 $\widehat{Q}_*^{(k)}$(0.25)、 $\widehat{Q}_*^{(k)}$(0.50)、 $\widehat{Q}_*^{(k)}$(0.75) 的估计均值和协方差-方差（$\times n$）矩阵，其中 $k = \pi$、 π^2、 π^3。

	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = \infty$
ZIP ($\lambda = 1$, $c = 0.8$), $k = \pi$				
$\widehat{means}$	(0.01, 0.10, 0.62)	(0.01, 0.10, 0.62)	(0.01, 0.10, 0.62)	(0.006, 0.095, 0.616)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.02 & 0.06 \\ 0.02 & 0.23 & 0.80 \\ 0.06 & 0.80 & 3.80 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.02 & 0.06 \\ 0.02 & 0.22 & 0.69 \\ 0.06 & 0.69 & 2.50 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.02 & 0.06 \\ 0.02 & 0.21 & 0.66 \\ 0.06 & 0.66 & 2.23 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.015 & 0.044 \\ 0.015 & 0.150 & 0.461 \\ 0.044 & 0.461 & 1.522 \end{bmatrix}$
ZIP ($\lambda = 1$, $c = 0.8$), $k = \pi^2$				
$\widehat{means}$	(0.00, 0.03, 0.53)	(0.00, 0.03, 0.52)	(0.00, 0.03, 0.51)	(0.000, 0.026, 0.514)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.06 & 0.31 \\ 0.00 & 0.31 & 2.63 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.31 \\ 0.00 & 0.31 & 2.88 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.01 \\ 0.00 & 0.05 & 0.41 \\ 0.01 & 0.41 & 3.60 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.004 \\ 0.000 & 0.041 & 0.318 \\ 0.004 & 0.318 & 2.709 \end{bmatrix}$
ZIP（$\lambda = 1$, $c = 0.8$）, $k = \pi^3$				
$\widehat{means}$	(0.00, 0.00, 0.35)	(0.00, 0.00, 0.32)	(0.00, 0.00, 0.31)	(0.000, 0.001, 0.315)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.06 \\ 0.00 & 0.06 & 3.65 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.02 \\ 0.00 & 0.02 & 4.07 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.02 \\ 0.00 & 0.02 & 3.94 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.021 \\ 0.000 & 0.021 & 3.400 \end{bmatrix}$
零膨胀负二项分布（$r = 1$, $\beta = 1$, $c = 0.8$）, $k = \pi$				
$\widehat{means}$	(0.00, 0.08, 0.64)	(0.00, 0.07, 0.65)	(0.00, 0.07, 0.64)	(0.003, 0.069, 0.642)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.01 & 0.04 \\ 0.01 & 0.19 & 0.79 \\ 0.04 & 0.79 & 4.87 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.01 & 0.03 \\ 0.01 & 0.18 & 0.72 \\ 0.03 & 0.72 & 4.40 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.01 & 0.04 \\ 0.01 & 0.17 & 0.75 \\ 0.04 & 0.75 & 3.76 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.007 & 0.029 \\ 0.007 & 0.119 & 0.519 \\ 0.029 & 0.519 & 2.534 \end{bmatrix}$
零膨胀负二项分布 ($r = 1$, $\beta = 1$, $c = 0.8$), $k = \pi^2$				
$\widehat{means}$	(0.00, 0.02, 0.53)	(0.00, 0.01, 0.50)	(0.00, 0.01, 0.49)	(0.000, 0.012, 0.489)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.28 \\ 0.00 & 0.28 & 3.96 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.24 \\ 0.00 & 0.24 & 4.25 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.29 \\ 0.00 & 0.29 & 4.74 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.001 \\ 0.000 & 0.014 & 0.223 \\ 0.001 & 0.223 & 3.781 \end{bmatrix}$
ZINB ($r = 1$, $\beta = 1$, $c = 0.8$), $k = \pi^3$				
$\widehat{means}$	(0.00, 0.00, 0.33)	(0.00, 0.00, 0.28)	(0.00, 0.00, 0.27)	(0.000, 0.000, 0.270)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.03 \\ 0.00 & 0.03 & 4.83 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.01 \\ 0.00 & 0.01 & 5.22 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 5.11 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.003 \\ 0.000 & 0.003 & 4.155 \end{bmatrix}$

NOTE: $n < \infty$ 的条目是估计四分位数的均值和样本协方差。结果基于 10,000 个模拟样本。这些条目的标准误为 ≤ 0.01。

方差（$\times n$）矩阵的平滑四分位数估计量的收敛速度不及泊松和负二项分布。它还取决于截断区间的宽度。虽然选择$k = \pi^2$ 可获得合理的结果， $k = \pi^3$ 则有所改进。这一观察与基于图2.1所提出的建议一致。此外，注意到所有四分位数估计量的估计均值都趋近于零。这是预期之中的，因为三个四分位数水平均低于 $c = 0.8$。自然地，接近零的均值估计会使协方差-方差估计量的值相近（几乎为 0）。

4 Bootstrap Approximation

在第 3 节中使用模拟，我们阐释了定理 2.1 的论述。在本节中，我们将进一步借助计算机的力量，构建一种自助法算法，以帮助我们近似第 2节的结果。请注意，如果设计得当，自助法过程甚至可以用于近似比离散分位数平滑更具挑战性的风险度量任务。在第 4.1 节中，我们概述自助法估计的算法。在第 4.2节中，我们验证该算法的性能，并针对泊松、负二项、ZIP 和零膨胀负二项分布，通过定理 2.1 进行交叉检验。

4.1 The Algorithm

The bootstrap algorithm requires specifications of the following inputs:

- *Data* is a sample generated by some distribution
- n is the sample size
- m is the number of bootstrapped resamples
- k is the number of standard deviations in interval (2.3)
- $u_1, \ldots, u_l$是向量（2.15）和（2.16）的分位数水平。

在算法的描述中，我们使用 $\boldsymbol{y}$ 表示重采样数据集，其样本均值和样本标准差分别记为 $\overline{\boldsymbol{y}}$ 和 $s_{\boldsymbol{y}}$。此外，给定任意 $y \in \boldsymbol{y}$，我们使用记号 $\mathsf{freq}(y)$ 表示数据集 $\boldsymbol{y}$ 中 y 的数量。

Algorithm for Bootstrap Approximation

输入数据； n； m； k； $u_1, \ldots, u_l$

for $i = 1, \ldots, m$ **do**

ResampledData (=: $\boldsymbol{y}$) = bootstrap(Data, with Replacement)

$\widehat{L}_k = \max(-0.5, \overline{\boldsymbol{y}} - k s_{\boldsymbol{y}}), \quad \widehat{U}_k = \overline{\boldsymbol{y}} + k s_{\boldsymbol{y}}$

$\widehat{y}_{1:\widehat{d}_k} = \text{ceiling}(\widehat{L}_k), \quad \widehat{y}_{\widehat{d}_k:\widehat{d}_k} = \text{floor}(\widehat{U}_k)$ / see (2.9) /

$\widehat{d}_k = \widehat{y}_{\widehat{d}_k:\widehat{d}_k} - \widehat{y}_{1:\widehat{d}_k} + 1, \quad \widehat{y}_{j:\widehat{d}_k} = \widehat{y}_{1:\widehat{d}_k} + j - 1$

对于 $j = 1, \ldots, \widehat{d}_k$ **do**

$\widehat{F}_{j(k)} = \Big(\text{freq}(\widehat{y}_{1:\widehat{d}_k}) + \cdots + \text{freq}(\widehat{y}_{j:\widehat{d}_k})\Big) \Big/ n$

结束循环

$\widehat{F}_Y(\widehat{L}_k) = \Big(\text{freq}\big(y \in \boldsymbol{y} \text{ such that } y < \widehat{y}_{1:\widehat{d}_k}\big)\Big) \Big/ n$

$\widehat{F}_Y(\widehat{U}_k) = \widehat{F}_Y(\widehat{L}_k) + \Big(\text{freq}(\widehat{y}_{1:\widehat{d}_k}) + \cdots + \text{freq}(\widehat{y}_{\widehat{d}_k:\widehat{d}_k})\Big) \Big/ n$

$\widehat{F}^*_{j(k)} = \Big(\widehat{F}_{j(k)} - \widehat{F}_Y(\widehat{L}_k)\Big) \Big/ \Big(\widehat{F}_Y(\widehat{U}_k) - \widehat{F}_Y(\widehat{L}_k)\Big)$ / see (2.11) /

for $u = u_1, \ldots, u_l$ **do**

$\widehat{\alpha}_u = (\widehat{d}_k + 1)u, \quad \widehat{\beta}_u = (\widehat{d}_k + 1)(1 - u)$

$\widehat{Q}^{(k)}_*(u) = \sum_{j=1}^{\widehat{d}_k} \Big(B_{\widehat{\alpha}_u, \widehat{\beta}_u}(\widehat{F}^*_{j(k)}) - B_{\widehat{\alpha}_u, \widehat{\beta}_u}(\widehat{F}^*_{j(k)-1})\Big)\widehat{y}_{j:\widehat{d}_k}$ / see (2.13) /

结束循环

将 $\widehat{Q}^{(k)}_*(u)$ 存入 $m \times l$ 矩阵

end for

列均值($\widehat{\mathrm{Q}}^{(\mathrm{k})}_*(\mathrm{u})$) = colMeans($m \times l$ matrix)

$\widehat{\Sigma}(\widehat{\mathrm{Q}}^{(\mathrm{k})}_*(\mathrm{u}))$ = 协方差($m \times l$ 矩阵)

4.2 Validation of the Algorithm

根据第 4.1 节的自助法算法，并给定一个大小为 n 的某个分布的样本，样本通过有放回抽样（自助法）进行经验重抽样，并计算平滑四分位数估计值。该循环重复10,000 次。随后计算这 10,000 个估计值的均值和协方差-方差矩阵，并进行汇总。结果报告于表4.1–4.2中。

As we see from the 表格中，自助法算法近似理论值 esta b-

表 4.1：在截断点 $\overline{y} \pm k\,s$ 处（ $k = \pi$、 π^2、 π^3）的泊松($\lambda = 9$)和负二项($r = 9$, $\beta = 1$)分布的平滑四分位数估计量 $\widehat{Q}_*^{(k)}$(0.25)、 $\widehat{Q}_*^{(k)}$(0.50)、 $\widehat{Q}_*^{(k)}$(0.75) 的自助法均值和协方差-方差($\times n$)矩阵。

	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = \infty$
Poisson ($\lambda = 9$), $k = \pi$				
$\widehat{means}$	(7.29, 9.06, 10.98)	(6.87, 8.80, 11.05)	(6.81, 8.82, 11.08)	(6.815, 8.835, 11.021)
$\widehat{\mathbf{HDH'}}$	9.51 6.90 3.79 6.90 10.49 7.26 3.79 7.26 8.46	11.13 7.73 5.58 7.73 10.89 9.80 5.58 9.80 16.04	11.95 8.58 5.84 8.58 11.92 10.45 5.84 10.45 16.46	11.367 8.360 5.539 8.360 11.497 9.753 5.539 9.753 15.478
泊松 ($\lambda = 9$), $k = \pi^2$				
$\widehat{means}$	(7.37, 9.03, 10.95)	(6.91, 8.78, 11.01)	(6.86, 8.81, 11.04)	(6.856, 8.838, 10.982)
$\widehat{\mathbf{HDH'}}$	10.63 7.32 3.44 7.32 12.75 7.08 3.44 7.08 7.83	13.30 8.28 5.71 8.28 12.09 9.52 5.71 9.52 15.87	12.83 8.59 5.88 8.59 12.80 10.45 5.88 10.45 17.71	12.153 8.309 5.526 8.309 12.289 9.714 5.526 9.714 16.579
泊松 ($\lambda = 9$), $k = \pi^3$				
$\widehat{means}$	(7.44, 9.01, 10.95)	(6.93, 8.77, 10.97)	(6.89, 8.82, 11.00)	(6.893, 8.853, 10.951)
$\widehat{\mathbf{HDH'}}$	12.76 8.25 3.15 8.25 15.14 6.36 3.15 6.36 6.71	14.63 9.08 5.25 9.08 14.10 8.54 5.25 8.54 13.40	11.29 7.72 5.11 7.72 13.10 9.55 5.11 9.55 17.36	10.533 7.033 4.695 7.033 11.401 8.415 4.695 8.415 15.631
负二项 ($r = 9$, $\beta = 1$), $k = \pi$				
$\widehat{means}$	(5.68, 8.08, 10.42)	(5.84, 8.40, 11.78)	(5.82, 8.44, 11.58)	(5.859, 8.504, 11.628)
$\widehat{\mathbf{HDH'}}$	17.82 11.60 8.80 11.60 13.45 12.47 8.80 12.47 25.04	17.14 13.94 12.20 13.94 21.49 23.88 12.20 23.88 48.24	17.21 13.97 10.42 13.97 21.70 20.33 10.42 20.33 35.85	18.038 14.458 10.384 14.458 22.085 20.054 10.384 20.054 34.815
负二项 ($r = 9$, $\beta = 1$), $k = \pi^2$				
$\widehat{means}$	(5.74, 8.11, 10.25)	(5.89, 8.38, 11.72)	(5.87, 8.44, 11.55)	(5.904, 8.515, 11.604)
$\widehat{\mathbf{HDH'}}$	18.13 10.75 8.17 10.75 13.65 11.62 8.17 11.62 24.02	16.26 12.38 11.39 12.38 20.91 22.39 11.39 22.39 51.69	17.84 13.48 10.22 13.48 22.77 19.74 10.22 19.74 37.65	19.552 14.467 10.507 14.467 23.833 20.212 10.507 20.212 37.975
负二项 ($r = 9$, $\beta = 1$), $k = \pi^3$				
$\widehat{means}$	(5.79, 8.13, 10.13)	(5.93, 8.35, 11.67)	(5.90, 8.42, 11.51)	(5.928, 8.504, 11.554)
$\widehat{\mathbf{HDH'}}$	16.54 8.81 6.47 8.81 13.06 9.81 6.47 9.81 20.80	11.35 9.61 9.30 9.61 22.65 22.73 9.30 22.73 58.05	15.05 12.05 8.99 12.05 26.48 20.26 8.99 20.26 40.64	17.673 13.777 9.675 13.777 28.408 20.920 9.675 20.920 40.813

NOTE: The entries for $n < \infty$ are the averages and sample covariances of estimated quartiles.
Results are based on 10,000 bootstrap resamples.

在定理中给出2.1 表现得相当好。对于泊松分布和负二项分布， $n \geq 100$ ，且 $k = \pi^2$在大多数情况下已足够。对于 ZIP 和零膨胀负二项分布， $n = 100$ 可能太小，即使 $k = \pi^3$，但当 $n = 1000$ 以及更大的样本量时，该算法表现良好。请注意，为了节省表格空间，协方差-方差矩阵的元素乘以 n。因此，这些数字可能会给人造成自助法近似与理论值近似之间差异很大的误导性印象，但可以检验出实际并非如此。例如，在表4.1 中，对于 $n = 100$、泊松分布、 $k = \pi^2$以及矩阵元素（3，3），实际有 $7.83/100 = 0.0783$（自助法）和 $16.579/100 = 0.1658$ （理论值）。在表 4.2 中，对于 $n = 1000$、零膨胀负二项分布、 $k = \pi^3$以及

表4.2：在点 $\overline{y} \pm k\,s$ 截断且具有 $k = \pi$、 π^2、 π^3 的ZIP($\lambda = 1$, $c = 0.8$)和零膨胀负二项分布($r = 1$, β =1, $c = 0.8$) 的平滑四分位数估计量 $\widehat{Q}_*^{(k)}$(0.25)、 $\widehat{Q}_*^{(k)}$(0.50)、 $\widehat{Q}_*^{(k)}$(0.75) 的自助法均值和协方差-方差 ($\times n$) 矩阵。

	$n = 10^2$	$n = 10^3$	$n = 10^4$	$n = \infty$
ZIP ($\lambda = 1$, $c = 0.8$), $k = \pi$				
$\widehat{means}$	(0.00, 0.08, 0.54)	(0.01, 0.09, 0.58)	(0.01, 0.10, 0.62)	(0.006, 0.095, 0.616)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.01 & 0.03 \\ 0.01 & 0.13 & 0.57 \\ 0.03 & 0.57 & 3.96 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.01 & 0.04 \\ 0.01 & 0.14 & 0.43 \\ 0.04 & 0.43 & 1.46 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.02 & 0.05 \\ 0.02 & 0.16 & 0.48 \\ 0.05 & 0.48 & 1.56 \end{bmatrix}$	$\begin{bmatrix} 0.001 & 0.015 & 0.044 \\ 0.015 & 0.150 & 0.461 \\ 0.044 & 0.461 & 1.522 \end{bmatrix}$
ZIP ($\lambda = 1$, $c = 0.8$), $k = \pi^2$				
$\widehat{means}$	(0.00, 0.03, 0.52)	(0.00, 0.03, 0.49)	(0.00, 0.03, 0.52)	(0.000, 0.026, 0.514)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.22 \\ 0.00 & 0.22 & 2.12 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.03 & 0.19 \\ 0.00 & 0.19 & 1.64 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.33 \\ 0.00 & 0.33 & 2.73 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.004 \\ 0.000 & 0.041 & 0.318 \\ 0.004 & 0.318 & 2.709 \end{bmatrix}$
ZIP ($\lambda = 1$, $c = 0.8$), $k = \pi^3$				
$\widehat{means}$	(0.00, 0.00, 0.36)	(0.00, 0.00, 0.29)	(0.00, 0.00, 0.33)	(0.000, 0.001, 0.315)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.03 \\ 0.00 & 0.03 & 3.11 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.000 \\ 0.00 & 0.00 & 0.02 \\ 0.00 & 0.02 & 2.57 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.000 \\ 0.00 & 0.00 & 0.02 \\ 0.00 & 0.02 & 3.12 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.021 \\ 0.000 & 0.021 & 3.400 \end{bmatrix}$
零膨胀负二项分布 ($r = 1$, $\beta = 1$, $c = 0.8$), $k = \pi$				
$\widehat{means}$	(0.00, 0.06, 0.46)	(0.00, 0.06, 0.58)	(0.00, 0.07, 0.64)	(0.003, 0.069, 0.642)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.01 & 0.02 \\ 0.01 & 0.08 & 0.32 \\ 0.02 & 0.32 & 1.72 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.01 & 0.02 \\ 0.01 & 0.10 & 0.42 \\ 0.02 & 0.42 & 2.49 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.01 & 0.03 \\ 0.01 & 0.12 & 0.53 \\ 0.03 & 0.53 & 2.62 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.007 & 0.029 \\ 0.007 & 0.119 & 0.519 \\ 0.029 & 0.519 & 2.534 \end{bmatrix}$
零膨胀负二项分布 ($r = 1$, $\beta = 1$, $c = 0.8$), $k = \pi^2$				
$\widehat{means}$	(0.00, 0.01, 0.36)	(0.00, 0.01, 0.47)	(0.00, 0.01, 0.48)	(0.000, 0.012, 0.489)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.01 \\ 0.00 & 0.01 & 1.42 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.01 & 0.15 \\ 0.00 & 0.15 & 2.52 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.02 & 0.20 \\ 0.00 & 0.20 & 3.17 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.001 \\ 0.000 & 0.014 & 0.223 \\ 0.001 & 0.223 & 3.781 \end{bmatrix}$
零膨胀负二项分布 ($r = 1$, $\beta = 1$, $c = 0.8$), $k = \pi^3$				
$\widehat{means}$	(0.00, 0.00, 0.17)	(0.00, 0.00, 0.25)	(0.00, 0.00, 0.26)	(0.000, 0.000, 0.270)
$\widehat{\mathbf{HDH'}}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.40 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 3.05 \end{bmatrix}$	$\begin{bmatrix} 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 3.45 \end{bmatrix}$	$\begin{bmatrix} 0.000 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.003 \\ 0.000 & 0.003 & 4.155 \end{bmatrix}$

NOTE: $n < \infty$ 的条目是估计四分位数的平均值和样本协方差。结果基于 10,000 次自助重抽样。

矩阵元素 (3,3) 中，我们实际上有 $3.05/1000 = 0.0031$ (自助法) 和 $4.155/1000 = 0.0042$(理论值)。

5 Real Data Examples

在本节中，新开发的方法应用于汽车事故数据集（Klug-man et al.，2012）及其三个修改。具体而言，在第5.1节中，介绍并描述了数据集。在第5.2节中，计算了四个数据的条件五数概括

集合，并补充了 95%（逐点）置信区间。在第 5.3节中，使用传统的离散概率近似方法以及新的平滑方法，对几个尾部概率的点估计进行了评估。

5.1 数据集

汽车事故数据（Klugman 等人，2012，表 6.2）代表了 9,461 份保险保单的风险概况。遵循 Brazauskas 和 Ratnam（2023+，第 5.3 节）的数值示例，我们还考虑了该数据集的三种尾部修改。具体来说，我们选取 140 份保单（约占投资组合的 1.5%）报告 0 起事故，并将其替换为 140 份报告至少 2 起事故的保单；这导致三种不同的情景。在表 5.1 中，

表 5.1：原始　　和修改后的数据集，由每份保单的事故数量组成　　.

数据集	事故数量									Total Number of Policies
	0	1	2	3	4	5	6	7	≥ 8	
O (original)	7,840	1,317	239	42	14	4	4	1	0	9,461
M1（修改版 #1）	7,700	1,317	*379*	42	14	4	4	1	0	9,461
M2（修改版 #2）	7,700	1,317	*279*	*62*	*34*	*24*	*24*	*21*	0	9,461
M3（修改版 #3）	7,700	1,317	239	42	14	4	4	*141*	0	9,461

提供了原始数据集和三个修改后的数据集。修改后的政策数量以斜体显示。

乍一看，M1、M2 和 M3 似乎比原始数据集 O 的投资组合风险更高。同样明显的是，风险从最低的投资组合（M1）逐步上升到风险最高的投资组合（M3）。我们随后计算的目标，是检验这些初步观察是否得到新方法的支持。

5.2 条件五数概括

为说明平滑分位数的联合行为如何帮助评估投资组合的尾部风险，我们使用表 5.1 中的数据集，并进行 C5NS 计算。结果列于表 5.2。

从表中可以看出，C5NS 方法支持关于 O、M1、M2 和 M3 的直观结论（见第 5.1 节）。事实上，原始数据集 O 中有 140 份保单的事故数为 0，而这些保单报告的事故数更高（在 M1 中均有 2 起事故，在 M3 中均有 7 起事故），因此 C5NS 使用的五个分位数会逐步同步上升。此外，相关置信区间相对较窄，通常互不重叠（O 和 M1 的区间除外）。这表明，可以在统计上区分并将这些组合分类如下：O 的风险最低，M1 的风险略高于 O，

表5.2：原始和修改后的汽车数据集的$\text{VaR}_{0.90}$之外的C5NS（括号内为95%置信区间）。

Data Set	汇总分位数（高于$\text{VaR}_{0.90}$水平）				
	$q_{0.91}$	$q_{0.925}$	$q_{0.95}$	$q_{0.975}$	$q_{0.99}$
O	1.35 (1.28; 1.41)	1.60 (1.51; 1.68)	2.28 (2.14; 2.43)	3.70 (3.48; 3.92)	5.33 (5.15; 5.50)
M1	1.47 (1.40; 1.53)	1.71 (1.63; 1.80)	2.38 (2.24; 2.52)	3.76 (3.54; 3.97)	5.35 (5.17; 5.52)
M2	1.86 (1.76; 1.96)	2.25 (2.13; 2.37)	3.19 (3.05; 3.34)	4.69 (4.56; 4.82)	5.96 (5.89; 6.04)
M3	2.30 (2.16; 2.43)	2.79 (2.64; 2.93)	3.85 (3.69; 4.00)	5.26 (5.15; 5.37)	6.27 (6.22; 6.33)

NOTE: 结果基于定理 2.1，截断点为 $\overline{Y} \pm \pi^3 S$。

M2 的风险显著高于 M1，而 M3 的风险最高。当然，这类统计推断并未用到分位数的联合渐近正态性。不过，如果要将这些分位数合并（例如，计算其加权平均值），则需要这一结果。

5.3 Tail Probabilities

为了展示平滑方法相较于通常使用的线性插值（Klugman 等人，2012，第 13.1 节）的优势，我们估计若干尾部概率，并评估这些估计值的标准误和变异系数（CV）。具体而言，对于平滑变量 Y^*，我们首先通过反演分位数函数，然后直接计算 $\mathbf{P}\{Y^* > a^*\}$（当 a^* 为非整数时），或者在 a^* 为整数时应用 0.5 连续性校正（Brazauskas 和 Ratnam，2023+，第 5.4 节）。这类估计值的变异性可以通过反演定理 2.1 的结果来评估，但我们将采用自助法算法（第 4.1节），该算法能够产生实际等价的结果（第 4.2节），但更易于实现。对于离散变量 Y，超过整数阈值的尾部概率可以直接估计。对于非整数阈值，则使用两个相邻整数处概率的线性插值。例如，$\mathbf{P}\{Y > 1.29\} = 0.71\mathbf{P}\{Y > 1\} + 0.29\,\mathbf{P}\{Y > 2\}$。这些估计值的变异性度量也通过采用自助法方法进行评估。这些计算结果汇总于表 5.3。

表中，概率 $\mathrm{P}\{Y > 0\}$ 和 $\mathrm{P}\{Y^* > 0.5\}$ 衡量至少发生一次索赔的可能性。数值 $a = a^* = 0.21$ 和 $a = a^* = 1.29$ 分别表示原始组合 O 中事故数量超过均值，以及超过均值加两个标准差的事件。（请注意，对于 O，均值为 0.21，标准差为 0.54。）关于平滑分位数方法，可以得出两点：首先，该方法更加保守，因为与标准离散变量方法相比，它给出的尾部概率估计更高；

表 5.3：原始数据集和修改后数据集的尾部概率。基于离散（Y）和光滑（Y^*）近似的均值、标准差和变异系数的自助法估计。

Data Set	Estimated Quantity	$\mathrm{P}\{Y > a\}$ 与 $\mathrm{P}\{Y^* > a^*\}$					
		$a = 0$	$a^* = 0.5$	$a = 0.21$	$a^* = 0.21$	$a = 1.29$	$a^* = 1.29$
O	Mean	0.172	0.208	0.142	0.301	0.025	0.095
	Std. Dev.	0.004	0.004	0.003	0.006	0.001	0.003
	CV	0.023	0.021	0.023	0.021	0.057	0.031
M1	Mean	0.186	0.226	0.157	0.321	0.035	0.105
	Std. Dev.	0.004	0.005	0.003	0.007	0.002	0.003
	CV	0.022	0.021	0.022	0.021	0.046	0.028
M2	Mean	0.186	0.226	0.157	0.318	0.038	0.122
	Std. Dev.	0.004	0.004	0.003	0.004	0.002	0.003
	CV	0.022	0.016	0.022	0.014	0.046	0.021
M3	Mean	0.186	0.231	0.157	0.319	0.040	0.137
	Std. Dev.	0.004	0.004	0.003	0.004	0.002	0.003
	CV	0.022	0.015	0.022	0.014	0.047	0.021

NOTE: R结果基于 1,000 次自助法重抽样，截断点为 $\overline{Y} \pm \pi^3 S$.

其次，该方法更加精确，因为“平滑”估计的变异系数始终小于“离散”估计的变异系数。

6 Summary and Concluding Remarks

本文研究了在多个分位数水平上同时估计平滑 VaR，以处理用于保险理赔频率建模的离散随机变量。我们将理论从有限域推广到无限域，并证明了截断离散风险的平滑分位数估计量具有一致性和联合渐近正态性。这些理论性质是通过为截断后的底层总体构造非整数的上下界而建立的。

此外，我们还通过在截断方法的理论设计中实现该程序，开展了蒙特卡罗模拟研究，以说明所建立的理论。模拟研究考察了具有无限定义域的常用离散分布，例如泊松分布、负二项分布及其（现实的）零膨胀版本。我们通过估计均值和协方差-方差矩阵向其相应理论值的收敛，成功说明了平滑分位数估计量的一致性和渐近正态性，尽管零膨胀分布的收敛速度比常规分布更慢。

此外，还设计了一个自助法近似，以说明理论值结果。利用该近似，对于来自每个所考察分布的一个原始样本数据集，仅通过有放回重采样，我们便能够看到自助法估-计的均值向量与均值向量的理论值近似之间的一致性，以及自助法估计的协方差-方差矩阵与协方差-方差矩阵的理论值近似之间的一致性。

最后，我们将无限域上的截断方法应用于汽车事故数据，并考察了数据尾部的三种渐进式修改。通过计算向量值风险度量C5NS，以及原始数据集及其三种修改形式的置信区间，我们表明，平滑分位数估计量能够通过适当评估尾部风险，准确地对投资组合的风险程度进行分类。为进一步展示平滑方法的优势，我们比较了使用平滑方法和线性插值方法获得的尾部概率。我们发现，与线性插值方法相比，平滑方法在尾部概率估计中产生的变异系数更低。

Acknowledgments

本研究得到加拿大自然科学与工程研究委员会（NSERC）及加拿大国家研究组织信息技术与复杂系统数学组织（MITACS）的支持，资助项目名称为“风险管理的新秩序：系统性风险时代的理论与应用”。

References

Alemany, R.、Bolanc´e, C. 和Guill´en, M.（2013）。一种计算风险价值的非参数方法。《保险：数学与经济学》，52(2)，255–262。

Artzner, P.、Delbaen, F.、Eber, J.-M. 和 Heath, D.（1999）。相干风险度量。《数学金融》，9(3)，203–228。

BCBS（2019）。《市场风险最低资本要求》。2019年2月。巴塞尔银行监管委员会。国际清算银行，巴塞尔。

https://www.bis.org/bcbs/publ/d457.htm

Bolanc´e, C. 和 Guill´en, M.（2021）。使用非参数估计极端分位数及其在长寿风险中的应用。《Risks》，9（77），23页， https://doi.org/10.3390/risks9040077Boucher, J.-P.、Denuit, M. 和 Guillen, M.（2009）。事故次数还是索赔次数？一种针对面板数据的零膨胀泊松模型方法。《风险与保险杂志》，76（4），821–846。Brazauskas, V. 和 Ratnam, P.（2023+）。用于衡量离散风险的平滑分位数。《北美精算学报》，即将出版。Demidenko, E.（2004）。《混合模型：理论与应用》。Wiley，纽约。Denuit, M.、Mar´echal, X.、Pitrebois, S. 和 Walhin, J.-F.（2007）。《索赔次数精算建模：风险分类、可信度和奖惩系统》。Wiley，奇切斯特。Harrell, F.E. 和 Davis, C.E.（1982）。一种新的无分布假设分位数估计量。《生物计量学》，69（3），635–640。Kab´an, A.（2012）。高维数据中无意义距离的非参数检测。《统计与计算》，22（2），375–385。Klugman, S.A.、Panjer, H.H. 和 Willmot, G.E.（2012）。《损失模型：从数据到决策》，第4版。Wiley，纽约。Machado, J.A.F. 和 Santos Silva, J.M.C.（2005）。计数分位数。《美国统计学会杂志》，100（472），1226–1237。Sarykalin, S.、Serraino, G. 和 Uryasev, S.（2014）。风险管理与优化中的风险价值与条件风险价值。《INFORMS 运筹学教程》，270–294。https://doi.org/10.1287/educ.1080.0052 Serfling, R.J.（1980）。《数理统计近似定理》。Wiley，纽约。Stigler, S.M.（1977）。分数阶次序统计量及其应用。《美国统计学会杂志》，72（359），544–550。Wang, D. 和 Hutson, A.D.（2011）。一种旨在为离散数据定义平滑分位数函数的分数阶次序统计量。《统计规划与推断杂志》，141（9），3142–3150。Wang, R. 和 Zitikis, R.（2021）。期望损失的公理化基础。《管理科学》，67，1413–1429。