

混合截尾数据下艾拉姆咖分布参数的后验分布研究

何贵阳, 吴俊欢

乌鲁木齐职业大学公共教学部, 新疆 乌鲁木齐

收稿日期: 2026年6月11日; 录用日期: 2026年7月19日; 发布日期: 2026年7月24日

摘要

利用混合截尾方法, 得到艾拉姆咖分布参数在无信息Jeffreys先验分布和有信息共轭先验分布下的后验分布结果的具体表达式, 同时利用优化理论, 给出以基本线性组合与加权线性组合下的参数的均方表达式, 并利用MCMC原理, 对优化结果进行数值模拟, 验证优化结果的准确性与真值拟合度; 针对混合截尾方法, 以参数的极大似然估计结果初步确定截尾数的取值区间, 同时在给定两类先验分布条件下, 利用随机模拟得到不同参数条件下的后验分布初始值与优化值的散点分布图, 并根据分布图确定截尾数的具体取值。

关键词

混合截尾, 后验分布, 优化理论, MCMC, 截尾数

Study on the Posterior Distribution of Erlang-K Parameters under Mixed Censored Data

Guiyang He, Junhuan Wu

Public Teaching Department, Urumqi Vocational University, Urumqi Xinjiang

Received: June 11, 2026; accepted: July 19, 2026; published: July 24, 2026

Abstract

Using the hybrid truncation method, the specific expressions of the posterior distribution results of the Erlang-K distribution parameters under non-informative prior distribution and informative conjugate prior distribution are obtained. At the same time, using optimization theory, the mean

squared expressions of parameters under basic linear combination and weighted linear combination are given. Additionally, based on MCMC principles, numerical simulation of the optimization results is conducted to verify the accuracy of the optimization results and the fit to the true values. For the hybrid truncation method, the range of truncation numbers is preliminarily determined using the maximum likelihood estimation results of the parameters. Simultaneously, under the conditions of two types of given prior distributions, stochastic simulation is used to obtain scatter plots of the initial and optimized values of the posterior distribution under different parameter conditions, and the specific values of truncation numbers are determined according to the distribution plots.

Keywords

Hybrid Truncation, Posterior Distribution, Optimization Theory, MCMC, Truncation Number

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在参数估计的发展历程中, 对在给定样本数据下, 讨论满足其数据特点的分布函数中相应参数的贝叶斯估计的研究已形成相应的体系, 即对满足数据特点的分布函数参数的贝叶斯推断问题[1]-[3]。在贝叶斯推断问题中, 作为一类基于贝叶斯推断最先被研究讨论的, 后验分布是具有较多重要的性质与研究意义[4]。在经典的参数估计的问题中, 常以样本数据作为最基本的一类信息来源。考虑样本数据的随机性与复杂性, 在估计问题中以截取部分样本数据来进行整体估计, 主要原则为截断与删失, 常用的截取方案有: 左截断, 右删失, 截断混删失混合方法等, 文献[5][6]介绍了在截尾样本下参数回归问题的相应研究[5][6]; 文献[7]在右删失数据下对相应参数回归问题进行了讨论[7]; 文献[8][9]在基于左截断右删失数据下对分布的相应参数进行了研究[8][9]。根据数据特点并结合研究内容也有两大截尾方案, 即定时, 定数, 也称为一型, 二型截尾方案。常见的截尾模型包括定数截尾与定时结尾。在相应模型的截尾数据下, 有许多学者也给出了其在各类分布下的参数估计问题与实际应用[10]-[15]。对比两类模型的特点, 可以确定两类模型研究的出发点是不同的, 即以给定失效数为常数下截止的实验是定数截尾; 以给定时间下失效数量为随机变量下截止的实验是定时截尾。在实际应用下对寿命实验的研究中, 考虑样本特性的情况, 即样本容量小、数据不完全、实验时间不可预知等, 常限定在定数截尾试验下对特定分布进行可靠性统计推断。

本文研究从截尾方法出发, 对比原有的单边删失数据模型, 即只截断控制一侧数据, 未能充分考虑数据始端或末端的分布特点。在此基础上, 文章新增双边混合截尾方式。借助双边混合截尾能更好的控制来源数据的初混乱性与长尾性等特点, 精准定位核心数据, 增强数据来源的普遍性优势。同时针对数据来源的不同特点, 可选取不同的双边混合截尾模型, 即双边混合截尾模型, 其中包括双定时混合截尾与双定数混合截尾模型。因两类模型最初是由龙兵与张忠占于文献[16]首次提出, 并对 Pareto 分布进行可靠性的统计分析和实际数据的相应模拟[16]; 文献[17]利用新模型得到了两参数 Pareto 分布参数的极大似然估计, 根据 Fisher 信息量推导出参数 θ 的渐近置信区间[17]。本文将在双定数混合截尾模型下对艾拉姆咖分布后验分布结果进行研究。在模型条件下给出参数的极大似然估计, 并通过数值模拟给出参数估计的简单截尾数区间; 同时研究艾拉姆咖分布在相应模型条件下, 以两类基本的先验分布为前提下的后验分布, 并对所得到的后验分布结果进行结构上的近似优化。最后在对所得到的后验分布结果进行相应的

数据模拟, 并对比参数的极大似然估计下参数估计的简单截尾数区间, 确定出截尾数据下的艾拉姆咖分布参数的最优截尾数值, 为后续研究给出准确截尾数参照。

2. 预备知识

1) 分布信息

艾拉姆咖分布的密度函数与分布函数

$$F(x, \lambda) = 1 - (1 + \lambda x)e^{-\lambda x}; \quad (1)$$

$$f(x, \lambda) = \lambda^2 x e^{-\lambda x}; \quad (2)$$

其中, $x > 0, \lambda > 0$ 。由(1)式可以看出艾拉姆咖分布属于单一参数的分布函数, 即形状参数确定为 2 的一族单峰的伽马分布函数。

2) 模型描述[16]

两类双边混合截尾模型的构建较为简单, 即双定数混合截尾试验是对两个定数截尾实验模型的进行相应混合; 双定时混合截尾是对两个定时截尾实验模型的进行相应混合。两类模型结构上有较强的相似性, 以双定数混合截尾模型为例, 其模型描述如下:

将独立同分布的 n 件样品同时投入试验, 事先给定标准失效时刻 t_0 和正整数 m_1, m_2 , 且满足 $m_1 < m_2 < n$ 。试验中取 t_{m_1} 表示第 m_1 件样品的失效时刻; 取 t_{m_2} 表示第 m_2 件样品的失效时刻。规定:

若 $t_{m_1} \geq t_0$, 则试验在 t_{m_1} 时刻停止, 说明 n 件样品有 m_1 件参与试验, 同时没有失效的 $n - m_1$ 件元件退出试验, 且 $0 < t_1 < t_2 \cdots < t_{m_1}$ 为次序失效时刻;

若 $t_{m_1} < t_0$, 则试验继续进行到有 m_2 件元件失效后停止, 同时没有失效的 $n - m_2$ 件元件退出试验, 且 $0 < t_1 < t_2 \cdots < t_{m_2}$ 为次序失效时刻。

从上面的描述可以看出, 该模型实际上就是在两个定数截尾方案之间的一种随机选择。即存在两种不同情形分别表示为 Case I 和 Case II, 并且称其为双定数混合截尾试验, 从而得到如下的观测数据

Case I: $(t_1, t_2, \cdots, t_{m_1})$, 若 $t_{m_1} \geq t_0$;

Case II: $(t_1, t_2, \cdots, t_{m_2})$, 若 $t_{m_1} < t_0$ 。

3. 参数的极大似然估计

通过上文给出的双定数混合截尾实验模型构建的方法, 我们可以确定出一组满足艾拉姆咖分布的随机样本 $x_1, x_2, \cdots, x_k$, 其中

$$k = \begin{cases} m_1, \text{Case I} \\ m_2, \text{Case II} \end{cases}, \quad t = \begin{cases} t_{m_1}, \text{Case I} \\ t_{m_2}, \text{Case II} \end{cases};$$

令 $x = (x_1, x_2, \cdots, x_k)$, 则样本 x 的似然函数为:

$$L(x | \lambda) = c \prod_{i=1}^k f(x_i) [1 - F(t)]^{n-k};$$

其中, c 是与参数无关的常数, 则将分布函数(1)式与密度函数(2)式代入, 可得

$$\begin{aligned} L(x | \lambda) &= c \prod_{i=1}^k f(x_i) [1 - F(t)]^{n-k} \\ &= c \prod_{i=1}^k \lambda^2 x_i e^{-\lambda x_i} \left[1 - (1 - (1 + \lambda t) e^{-\lambda t}) \right]^{n-k} \\ &= c (1 + \lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-\left(\sum_{i=1}^k x_i + t(n-k) \right) \lambda}; \end{aligned}$$

令 $P = \left(\sum_{i=1}^k x_i + t(n-k) \right)$, 则有

$$L(x|\lambda) = c(1+\lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-P\lambda}. \quad (3)$$

由上式可知, 其对数似然函数为:

$$\ln L(x|\lambda) = \ln c + (n-k) \ln(1+\lambda t) + 2k \ln \lambda + \sum_{i=1}^k \ln x_i - P\lambda;$$

并对参数 λ 取一阶导数

$$\frac{d \ln L(x|\lambda)}{d\lambda} = \frac{(n-k)t}{1+\lambda t} + \frac{2k}{\lambda} - P;$$

当 $\frac{d \ln L(x|\lambda)}{d\lambda} = 0$, 同时知 $\lambda > 0$ 时, 有:

$$\begin{aligned} \frac{(n-k)t}{1+\lambda t} + \frac{2k}{\lambda} - P &= 0 \\ \Rightarrow \left(\lambda - \frac{nt+kt-P}{2Pt} \right)^2 &= \frac{(nt+kt-P)^2 + 8Ptk}{4P^2t^2}, \\ \hat{\lambda} &= \frac{\sqrt{(nt+kt-P)^2 + 8Ptk} + (nt+kt-P)}{2Pt}. \end{aligned}$$

则确定参数 λ 的极大似然估计为

$$\hat{\lambda} = \frac{\sqrt{(nt+kt-P)^2 + 8Ptk} + (nt+kt-P)}{2Pt}. \quad (4)$$

4. 双定数混合截尾数据下的后验分布

由 Bayes 定理可知, 在处理特定分布的某个参数估计问题时, 因参数的不确定性, 需要利用参数自身的特性, 寻找合理的先验分布去建立参数与样本的关系, 即参数的后验分布。文章将以对比常用的先验分布方法, 即无信息下的 Jeffreys 先验、有信息下的共轭先验, 确定其相应条件下的艾拉姆咖分布中未知参数的后验分布。

4.1. 无信息下的 Jeffreys 先验

根据 Jeffreys 原则确定未知参数先验分布的理论, 由(2)式可知 $f(x, \lambda) = \lambda^2 x e^{-\lambda x}$, 可得

$\ln f(x, \lambda) = 2 \ln \lambda + \ln x - \lambda x$, 则 $\frac{\partial^2 \ln f(x, \lambda)}{\partial \lambda^2} = -\frac{2}{\lambda^2}$; 同时得到 Fisher 信息阵

$$I(\lambda) = -E_{\lambda} \left[\frac{\partial^2 \ln f(x, \lambda)}{\partial \lambda^2} \right] = -\frac{2}{\lambda^2}; \text{ 确定参数 } \lambda \text{ 的先验分布为 } \pi(\lambda) = \frac{1}{\lambda}, \lambda > 0.$$

由确定的先验分布与(3)式, 可以确定参数的后验分布为:

$$\begin{aligned} h(\lambda|x) &= \frac{\pi(\lambda) \cdot L(x|\lambda)}{\int \pi(\lambda) \cdot L(x|\lambda) d\lambda} \\ &= \sum_{j=0}^{n-k} \left(\binom{n-k}{j} t^j \lambda^{2k+j-1} \right) e^{-P\lambda} \left(\sum_{j=0}^{n-k} \frac{\binom{n-k}{j} t^j \Gamma(2k+j)}{P^{2k+j}} \right)^{-1} \end{aligned}$$

对上式结果进行整理, 令 $A_1 = \left(\sum_{j=0}^{n-k} \frac{\binom{n-k}{j} t^j \Gamma(2k+j)}{P^{2k+j}} \right)^{-1}$ 。可得

$$h_1(\lambda | x) = A_1 \cdot \sum_{j=0}^{n-k} \left(\binom{n-k}{j} t^j \lambda^{2k+j-1} \right) e^{-P\lambda}. \quad (5)$$

4.2. 有信息下的共轭先验

在研究一类分布函数中参数未知情况下的估计问题时, 对参数先验分布的选取是一大难题。而共轭先验分布恰是一种解决方案, 共轭先验分布是利用先验分布与后验分布在分布函数上的同一性的特点, 事先假设出先验分布的分布特点。本文取未知参数 λ 的共轭先验分布为指数分布 $\exp(a)$, 则有 $\pi(\lambda) = a^2 e^{-a\lambda}$, 其中参数 a 为超参数, 且 $a > 0$; 由共轭先验分布表达式结合(3)式, 可得未知参数 λ 的后验分布为:

$$h(\lambda | x) = \frac{\pi(\lambda) \cdot L(x | \lambda)}{\int \pi(\lambda) \cdot L(x | \lambda) d\lambda}$$

$$= \sum_{j=0}^{n-k} \left(\binom{n-k}{j} t^j \lambda^{2k+j} \right) e^{-(P+a)\lambda} \left(\sum_{j=0}^{n-k} \frac{\binom{n-k}{j} t^j \Gamma(2k+j+1)}{(P+a)^{2k+j+1}} \right)^{-1}$$

对上式结果进行整理, 令 $A_2 = \left(\sum_{j=0}^{n-k} \frac{\binom{n-k}{j} t^j \Gamma(2k+j+1)}{(P+a)^{2k+j+1}} \right)^{-1}$ 。可得

$$h_2(\lambda | x) = A_2 \cdot \sum_{j=0}^{n-k} \left(\binom{n-k}{j} t^j \lambda^{2k+j} \right) e^{-(P+a)\lambda}. \quad (6)$$

综上, 观察后验分布结果可以确定, 本文在对所研究的双定数混合截尾下艾拉姆咖分布中未知参数 λ 的先验分布进行差异选取时, 所得到的后验分布是不同的, 但结果形式是具有近似性的, 即都是由 $n-k+1$ 个特定分布的加权线性组合构成的, 具体举例(5)式的形式:

$$h_1(\lambda | x) = \left(\sum_{j=0}^{n-k} \frac{\binom{n-k}{j} t^j \Gamma(2k+j)}{P^{2k+j}} \right)^{-1} \sum_{j=0}^{n-k} \left(\binom{n-k}{j} t^j \lambda^{2k+j-1} \right) e^{-P\lambda}$$

$$= \frac{\left(\lambda^{2k-1} + \dots + \binom{n-k}{j} t^j \lambda^{2k+j-1} + \dots + t^{n-k} \lambda^{2k+n-k-1} \right) e^{-P\lambda}}{\left(\frac{\Gamma(2k)}{P^{2k}} + \dots + \frac{\binom{n-k}{j} t^j \Gamma(2k+j)}{P^{2k+j}} + \dots + \frac{t^{n-k} \Gamma(2k+n-k)}{P^{2k+n-k}} \right)}$$

$$\begin{aligned}
&= \frac{\left(1 + \dots + \binom{n-k}{j} t^j \lambda^j + \dots + t^{n-k} \lambda^{n-k}\right) \frac{P^{2k+n-k}}{\Gamma(2k)} \lambda^{2k-1} e^{-P\lambda}}{\left(P^{n-k} + \dots + P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} + \dots + t^{n-k} \frac{\Gamma(2k+n-k)}{\Gamma(2k)}\right)} \\
&= \frac{P^{n-k} \frac{P^{2k}}{\Gamma(2k)} \lambda^{2k-1} e^{-P\lambda}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)} + \dots + \frac{P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)} \\
&\quad + \dots + \frac{t^{n-k} \frac{\Gamma(2k+n-k)}{\Gamma(2k)} \frac{P^{2k+n-k}}{\Gamma(2k+n-k)} \lambda^{2k+n-k-1} e^{-P\lambda}}{\left(\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)\right)};
\end{aligned}$$

即有:

$$\begin{aligned}
h_1(\lambda|x) &= \frac{\pi(\lambda) \cdot L(x|\lambda)}{\int \pi(\lambda) \cdot L(x|\lambda) d\lambda} \\
&= \frac{P^{n-k} \frac{P^{2k}}{\Gamma(2k)} \lambda^{2k-1} e^{-P\lambda}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)} + \dots + \frac{P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)} \\
&\quad + \dots + \frac{t^{n-k} \frac{\Gamma(2k+n-k)}{\Gamma(2k)} \frac{P^{2k+n-k}}{\Gamma(2k+n-k)} \lambda^{2k+n-k-1} e^{-P\lambda}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)};
\end{aligned}$$

由上式可以分析得到, 后验分布是一组关于尺度参数为 P , 形状参数为 $2k, 2k+1, \dots, 2k+n-k$ 的 $n-k+1$ 个伽马分布的加权线性组合, 且各项系数均为其相应的权重, 分别记为 $B_0, B_1, \dots, B_{n-k}$, 且有 $B_0 + B_1 + \dots + B_{n-k} = 1$, 则后验分布表达式为

$$\begin{aligned}
h_1(\lambda|x) &= B_{0^1} \frac{P^{2k}}{\Gamma(2k)} \lambda^{2k-1} e^{-P\lambda} + \dots + B_{j^1} \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda} + \dots \\
&\quad + B_{(n-k)^1} \frac{P^{2k+n-k}}{\Gamma(2k+n-k)} \lambda^{2k+n-k-1} e^{-P\lambda};
\end{aligned}$$

$$\text{其中, } B_{j^1} = \frac{P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)}\right)}.$$

同理, (6)式结果形式近似, 即有

$$\begin{aligned}
h_2(\lambda|x) &= B_{0^2} \frac{(P+a)^{2k}}{\Gamma(2k+1)} \lambda^{2k} e^{-(P+a)\lambda} + \dots + B_{j^2} \frac{(P+a)^{2k+j}}{\Gamma(2k+j+1)} \lambda^{2k+j} e^{-(P+a)\lambda} + \dots \\
&\quad + B_{(n-k)^2} \frac{(P+a)^{2k+n-k}}{\Gamma(2k+n-k+1)} \lambda^{2k+n-k} e^{-(P+a)\lambda};
\end{aligned}$$

$$\text{其中, } B_{j^2} = \frac{(P+a)^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+1)}{\Gamma(2k+1)}}{\sum_{j=0}^{n-k} \left((P+a)^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+1)}{\Gamma(2k+1)} \right)}.$$

5. 后验结果的结构优化

优化方式 1 利用后验分布核的性质进行结果近似。

在贝叶斯估计理论中, 为了突出主要部分, 引入了核的定义。即只要知道分布函数的核, 利用分布密度函数在全样本空间下的积分, 则可以确定相应的常数因子属性。因此, 对于要讨论相应随机变量分布函数的基本性质, 最为重要的是确定其分布函数的核[18]。通过第二章介绍的, 在以无信息下的 *Jeffreys* 先验、有信息下的共轭先验, 确定艾拉姆咖分布中未知参数的后验分布的原则下, 以提取后验分布常数因子, 确定核函数的结构形式作为后验分布结果的一种优化方案。

无信息下的 *Jeffreys* 先验。确定参数 λ 的先验分布为 $\pi(\lambda) = \frac{1}{\lambda}, \lambda > 0$, 得到参数的后验分布为:

$$h_1(\lambda|x) = \frac{\pi(\lambda) \cdot L(x|\lambda)}{\int \pi(\lambda) \cdot L(x|\lambda) d\lambda} = \frac{\frac{1}{\lambda} \cdot c(1+\lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-P\lambda}}{\int \frac{1}{\lambda} \cdot c(1+\lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-P\lambda} d\lambda}$$

提取常数因子 A_1 , 确定

$$h_1(\lambda|x) \propto A_1 \sum_{j=0}^{n-k} (\lambda^{2k+j-1} e^{-P\lambda}).$$

其中, A_1 由 $\int h_1(\lambda|x) = 1$ 所确定。

若设 $x_j = \lambda^{2k+j-1} e^{-P\lambda}$, 则 $h_1(\lambda|x) \propto A_1 \sum_{j=0}^{n-k} x_j$, 则后验分布 $h_1(\lambda|x)$ 的核是由 $n-k+1$ 个近似伽马分布线性组成。

根据伽马分布线性可加性的基本性质, 假设 $n-k+1$ 个近似伽马分布独立同分布, 且尺度参数相同。根据优化理论, 确定后验分布中形状参数的取值, 即取 $n-k+1$ 个近似伽马分布形状参数的均值, 有

$$h_1'(\lambda|x) \propto A_1 \lambda^{\sum_{j=0}^{n-k} \left(\frac{2k+j-1}{n-k+1} \right)} e^{-P\lambda} \sim \text{Gam} \left(\frac{\sum_{j=0}^{n-k} (2k+j-1)}{n-k+1} + 1, P \right); \quad (7)$$

有信息下的共轭先验。取未知参数 λ 的共轭先验分布为 $\pi(\lambda) = a^2 e^{-a\lambda}$, 其中参数 a 为超参数, 且 $a > 0$, 得到参数的后验分布为:

$$h_2(\lambda|x) = \frac{\pi(\lambda) \cdot L(x|\lambda)}{\int \pi(\lambda) \cdot L(x|\lambda) d\lambda} = \frac{a^2 e^{-a\lambda} \cdot c(1+\lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-P\lambda}}{\int a^2 e^{-a\lambda} \cdot c(1+\lambda t)^{n-k} \lambda^{2k} \prod_{i=1}^k x_i e^{-P\lambda} d\lambda}$$

提取常数因子 A_2 , 确定

$$h_2(\lambda|x) \propto A_2 \sum_{j=0}^{n-k} (\lambda^{2k+j} e^{-(P+a)\lambda})$$

利用(7)式理论, 确定

$$h'_2(\lambda|x) \propto A_2 \lambda^{\sum_{j=0}^{n-k} \binom{2k+j}{n-k+1}} e^{-(P+a)\lambda} \sim \text{Gam} \left(\frac{\sum_{j=0}^{n-k} (2k+j)}{n-k+1} + 1, P+a \right). \quad (8)$$

优化方式 2 利用线性规划原理与伽马分布的性质进行结果优化。

下面以(5)式为例, 利用最优化方法, 对得到的加权线性组合的后验分布进行优化讨论。把构成后验分布中的 $n-k+1$ 个伽马分布当作加权线性组合中的自变量, 去预测后验分布的最优解。

考虑后验分布 $h_1(\lambda|x)$, 因对其所得结果进行分解后, 得到以权系数为 B_{j^1} 下的一组线性加权组合,

$$\begin{aligned} h_1(\lambda|x) = & B_{0^1} \frac{P^{2k}}{\Gamma(2k)} \lambda^{2k-1} e^{-P\lambda} + \cdots + B_{j^1} \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda} + \cdots \\ & + B_{(n-k)^1} \frac{P^{2k+n-k}}{\Gamma(2k+n-k)} \lambda^{2k+n-k-1} e^{-P\lambda}; \end{aligned}$$

若限定后验分布中的 $\frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda} = X_j$, 其中 X_j 是相互独立的且 $j=0, \dots, n-k$; 根据分布函数的性质, 确定 X_j 是服从伽马分布的, 即有

$$X_j = \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda} = \text{Gam}(2k+j, P).$$

在以上限定下, 后验分布 $h_1(\lambda|x)$ 在结构上看作是由 $n-k+1$ 个尺度参数为 P , 形状参数为 $2k, 2k+1, \dots, k+n$ 的伽马分布的线性加权组合, 即

$$\begin{aligned} h_1(\lambda|x) &= B_{0^1} X_0 + \cdots + B_{j^1} X_j + \cdots + B_{(n-k)^1} X_{n-k} \\ &\Rightarrow h_1(\lambda|x) = \sum_{j=0}^{n-k} B_{j^1} X_j. \end{aligned}$$

其中, B_{j^1} 为对应组合下的权重值。

确定后验分布 $h_1(\lambda|x)$ 的一组伽马分布的线性加权组合下, 利用伽马分布线性组合的性质进行优化。

性质 1: 假设有 n 个独立的伽玛分布 $X_1, X_2, \dots, X_n$, 它们分别服从参数为 $(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_n, \beta_n)$ 的伽玛分布。那么它们的线性组合可以表示为:

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n;$$

其中, $a_1, a_2, \dots, a_n$ 为权重系数。

线性组合后的 Y 仍然是一个伽玛分布, 它的参数为 (α, β) , 其中 α 和 β 的值可以通过计算得到。

$$\alpha = a_1 \alpha_1 + a_2 \alpha_2 + \cdots + a_n \alpha_n;$$

$$\beta = a_1 \beta_1 + a_2 \beta_2 + \cdots + a_n \beta_n.$$

即后验分布优化的结果为

$$h''_1(\lambda|x) = \sum_{j=0}^{n-k} \left(B_{j^1} \frac{P^{2k+j}}{\Gamma(2k+j)} \lambda^{2k+j-1} e^{-P\lambda} \right) = \text{Gam} \left(\sum_{j=0}^{n-k} (B_{j^1} (2k+j)), P \right) \quad (9)$$

综上, 在给定参数取值的情况下, 可考虑将预测值 $h''_1(\lambda|x)$, 作为后验分布的优化的最终结果, 即

$$\begin{aligned}
h_1''(\lambda | x) &= \text{Gam} \left(\sum_{j=0}^{n-k} (B_{j^1}(2k+j)), P \right) \\
&= \text{Gam} \left(\frac{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \cdot (2k+j) \right)}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \right)}, P \right) \\
&= \text{Gam} \left(\frac{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+1)}{\Gamma(2k)} \right)}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \right)}, P \right)
\end{aligned}$$

根据以上结论, 对已得到的在不同损失函数下的后验分布结果优化可知:

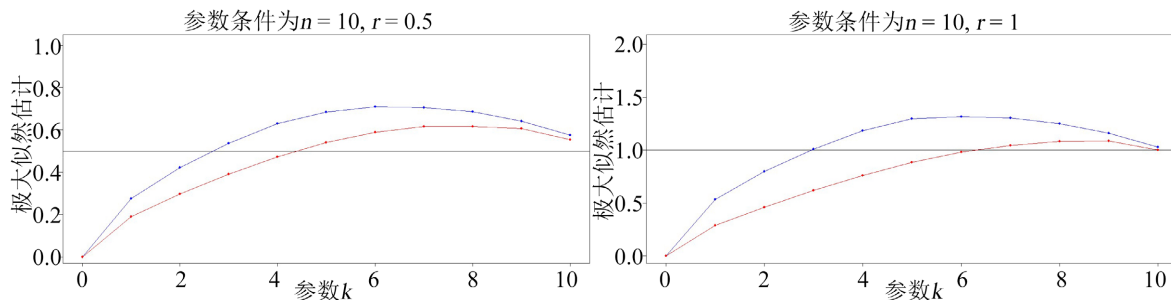
$$\begin{aligned}
h_1''(\lambda | x) &= \text{Gam} \left(\frac{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+1)}{\Gamma(2k)} \right)}{\sum_{j=0}^{n-k} \left(P^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j)}{\Gamma(2k)} \right)}, P \right); \\
h_2''(\lambda | x) &= \text{Gam} \left(\frac{\sum_{j=0}^{n-k} \left((P+a)^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+2)}{\Gamma(2k+1)} \right)}{\sum_{j=0}^{n-k} \left((P+a)^{n-k-j} \binom{n-k}{j} t^j \frac{\Gamma(2k+j+1)}{\Gamma(2k+1)} \right)}, P+a \right).
\end{aligned}$$

6. 数据对比

6.1. 极大似然估计的数值模拟

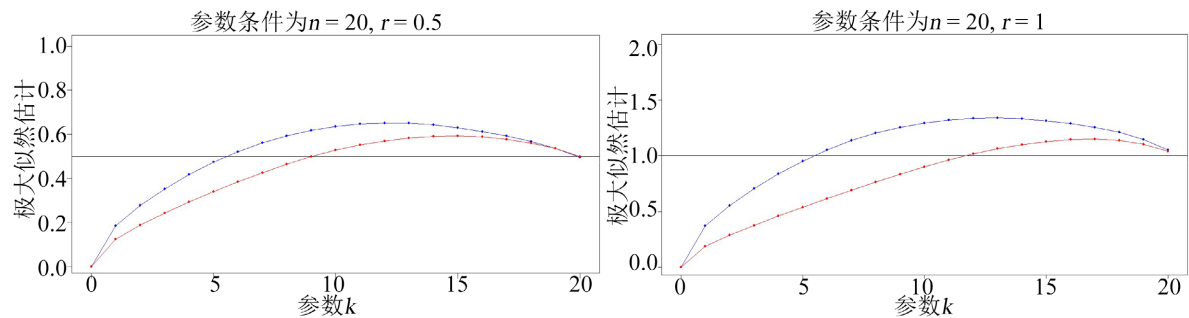
利用蒙特卡洛随机数生成原理, 以样本总数 n 取 10, 20, 50, 尺度参数 λ 真值取 0.5, 1 为限定, 生成一组艾拉姆咖分布的一组随机数。讨论在双定数混合截尾模型下, 给定模型参数 t 取值, 对模型参数 k 在样本总数 n 下的取值情况进行分析, 并利用尺度参数 λ 的极大似然估计数据, 预估最优估计下的模型参数 k 的取值, 并对模拟过程进行 100,000 次重复计算取均值, 最后对所得数据在相应参数取值下, 绘制可视区间分布下的数据分布图。

以样本总数 $n=10$, 尺度参数 λ 真值取 0.5, 1, 生成的艾拉姆咖分布的随机数; 在限定模型参数 t 取 1, 2, 3 下, 双定数混合截尾模型下艾拉姆咖分布中尺度参数 λ 的极大似然估计数据分布图, 如下:



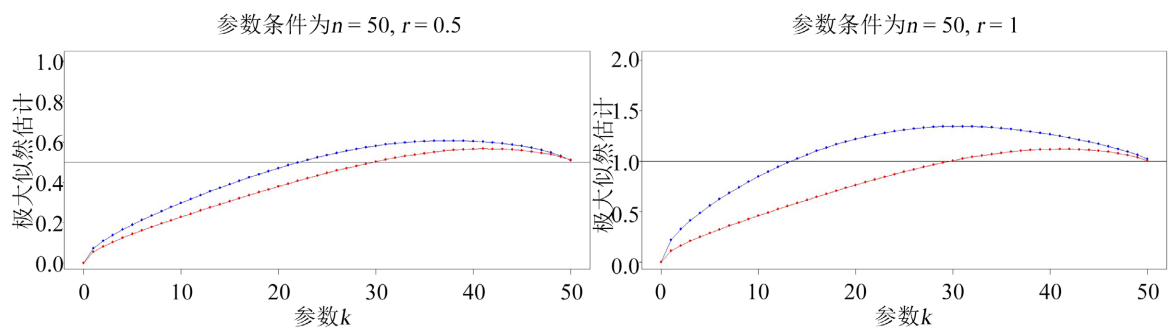
注: 在分布参数真值为 1 时, 蓝色数据为模型参数 $t=1$, 红色数据为 $t=2$; 在分布参数真值为 0.5 时, 蓝色数据为模型参数 $t=2$, 红色数据为 $t=3$ 。

以样本总数 $n=20$ ，尺度参数 λ 真值取 0.5, 1，生成的艾拉姆咖分布的随机数；在限定模型参数 t 取 1, 2, 3 下，双定数混合截尾模型下艾拉姆咖分布中尺度参数 λ 的极大似然估计数据分布图，如下：



注：在分布参数真值为 1 时，蓝色数据为模型参数 $t=1$ ，红色数据为 $t=2$ ；在分布参数真值为 0.5 时，蓝色数据为模型参数 $t=2$ ，红色数据为 $t=3$ 。

以样本总数 $n=50$ ，尺度参数 λ 真值取 0.5, 1，生成的艾拉姆咖分布的随机数；在限定模型参数 t 取 1, 2, 3, 4 下，双定数混合截尾模型下艾拉姆咖分布中尺度参数 λ 的极大似然估计数据分布图，如下：



注：在分布参数真值为 1 时，蓝色数据为模型参数 $t=1$ ，红色数据为 $t=2$ ；在分布参数真值为 0.5 时，蓝色数据为模型参数 $t=3$ ，红色数据为 $t=4$ 。

总结：

1) 根据数据分布情况表明：随着模型中截尾参数 k 的增大，艾拉姆咖分布尺度参数 λ 的极大似然估计结果逐步增大，且靠近真值；在全样本数据下，即模型中截尾参数 $k=n$ 下，模拟结果等于真值，符合参数估计结果；模型建立准确，有较高的数据拟合度。

2) 对比三组不同总量的数据样本的分布图，可以确定：除全样本 $k=n$ 条件下，在模型参数 k 的取值范围内，一定存在着某个截尾数，使得艾拉姆咖分布在双定数混合截尾模型下，尺度参数 λ 的极大似然估计值近似于真值。本文就是根据此结论，确定模型参数 k 的取值的情况，并在以此为已知条件下，进行了后续研究。

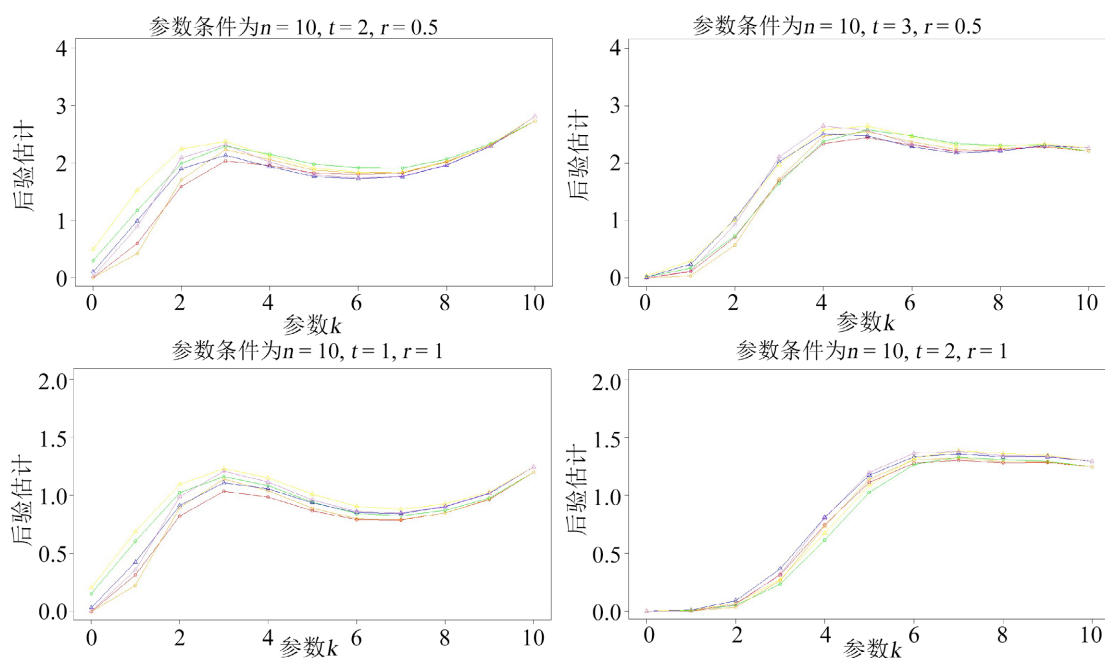
3) 观察三组数据分布图，可得下表：

样本总数	$n=10$				$n=20$				$n=50$			
分布参数真值 r	0.5		1		0.5		1		0.5		1	
模型参数 t	2	3	1	2	2	3	1	2	3	4	1	2
模型参数 k 的估计值范围	2~3	4~5	2~3	6~7	5~6	8~10	5~6	11~13	21~23	29~31	12~14	29~31

6.2. 后验分布的数值模拟

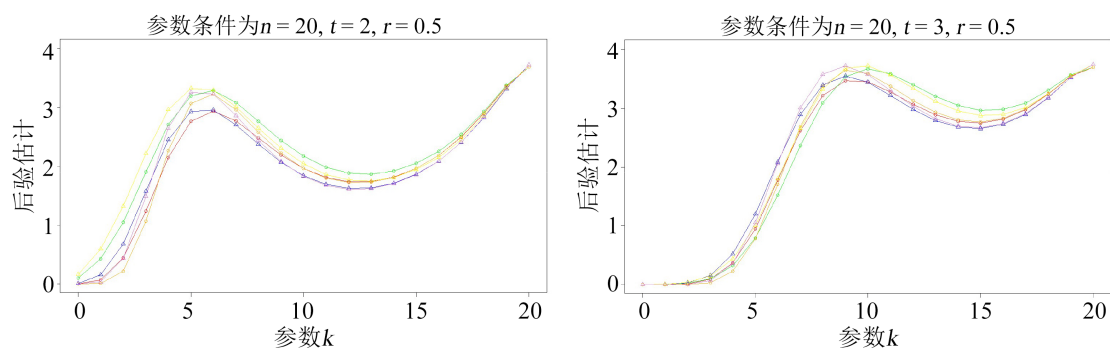
讨论在双定数混合截尾模型下, 给定模型参数 t 取值, 对模型参数 k 在样本总数 n 下的取值情况进行分析, 并对比上文提出的尺度参数 λ 的后验估计与优化结果的分布数据, 评估优化结果的准确性与真实性, 并对模拟过程进行 2000 次重复计算取均值, 最后对所得数据在相应参数取值下, 绘制可视区间分布下的数据分布图。

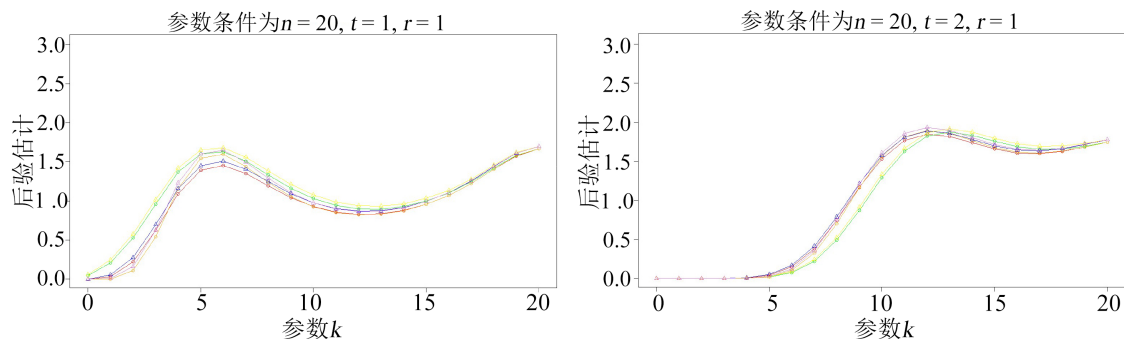
以样本总数 $n=10$, 尺度参数 λ 真值取 0.5, 1, 生成的艾拉姆咖分布的随机数; 在限定模型参数 t 取 1, 2, 3 下, 双定数混合截尾模型下艾拉姆咖分布的尺度参数 λ , 在三类不同先验分布下的后验估计与优化数据分布图, 如下:



注: 红圈数据是无信息 Jeffreys 先验分布下的后验分布数据; 蓝角数据是共轭先验分布下的后验分布数据; 绿圈数据是无信息 Jeffreys 先验分布下的后验分布优化方案 1 的数据; 黄角数据是共轭先验分布下的后验分布优化方案 1 的数据; 橙圈数据是无信息 Jeffreys 先验分布下的后验分布优化方案 1 的数据; 紫角数据是共轭先验分布下的后验分布优化方案 1 的数据。

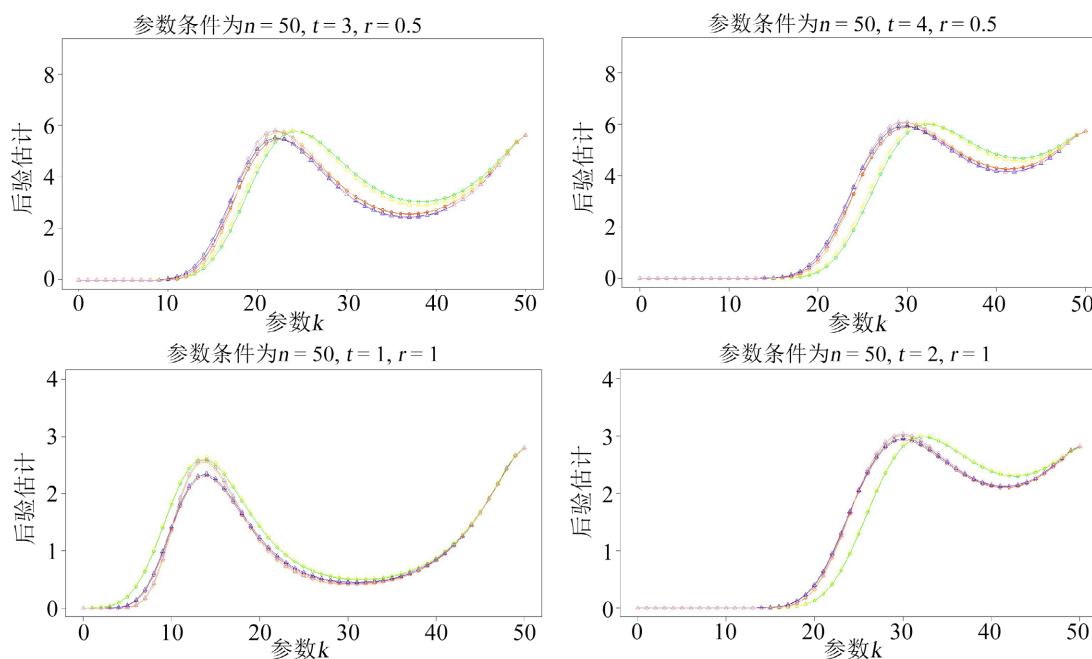
以样本总数 $n=20$, 尺度参数 λ 真值取 0.5, 1, 生成的艾拉姆咖分布的随机数; 在限定模型参数 t 取 1, 2, 3 下, 双定数混合截尾模型下艾拉姆咖分布的尺度参数 λ , 在三类不同先验分布下的后验估计与优化数据分布图, 如下:





注: 红圈数据是无信息 *Jeffreys* 先验分布下的后验分布数据; 蓝角数据是共轭先验分布下的后验分布数据; 绿圈数据是无信息 *Jeffreys* 先验分布下的后验分布优化方案 1 的数据; 黄角数据是共轭先验分布下的后验分布优化方案 1 的数据; 橙圈数据是无信息 *Jeffreys* 先验分布下的后验分布优化方案 1 的数据; 紫角数据是共轭先验分布下的后验分布优化方案 1 的数据。

以样本总数 $n=50$, 尺度参数 λ 真值取 0.5, 1, 生成的艾拉姆咖分布的随机数; 在限定模型参数 t 取 1, 2, 3, 4 下, 双定数混合截尾模型下艾拉姆咖分布的尺度参数 λ , 在三类不同先验分布下的后验估计与优化数据分布图, 如下:



注: 红圈数据是无信息 *Jeffreys* 先验分布下的后验分布数据; 蓝角数据是共轭先验分布下的后验分布数据; 绿圈数据是无信息 *Jeffreys* 先验分布下的后验分布优化方案 1 的数据; 黄角数据是共轭先验分布下的后验分布优化方案 1 的数据; 橙圈数据是无信息 *Jeffreys* 先验分布下的后验分布优化方案 1 的数据; 紫角数据是共轭先验分布下的后验分布优化方案 1 的数据。

总结:

1) 对比不同样本总量下的数据分布图可以看出, 在某一特定模型参数 k 值下, 不同先验分布函数得到的分布参数 r 的后验分布结果相同, 即在双定数混合截尾模型下艾拉姆咖分布的尺度参数 λ , 在不同的先验分布下的后验估计具有一定相合性。

2) 在不同样本总量的条件下, 艾拉姆咖分布在双定数混合截尾模型下的尺度参数 λ , 不同先验分布

下的后验估计与优化方案数据都具有一定拟合性, 数据分布在结构上有一定相似性。但优化方案随着样本总量的增加, 拟合程度并不一致, 优化方案 1 样本总量的增加偏差度逐步增加, 以致截点数错位; 优化方案 2 较 1 拟合度紧密。即优化方案 1 下的后验估计适用于小样本空间下; 优化方案 2 下的后验估计适用较广, 适用于大小样本空间。

3) 同条件下, 数据分布图内的分布参数结果, 优化方案 1 数据在小样本数据下较原数据偏高; 优化方案 2 数据在不同样本数据下都基本符合原数据数据分布, 优化方案较为准确, 优化过程合理有效。

4) 对比上一节模拟中尺度参数 λ 的极大似然估计数据分布图, 参照得到的最优模型参数 k 值, 可以判断出, 除全样本 $k=n$ 条件下, 在最优模型参数 k_1 值上分布参数的后验分布达到最高值, 基本符合估计结果, 模型模拟符合分布特点, 同时取后验分布最低值为模型参数 k_2 。

无信息 *Jeffreys* 条件下:

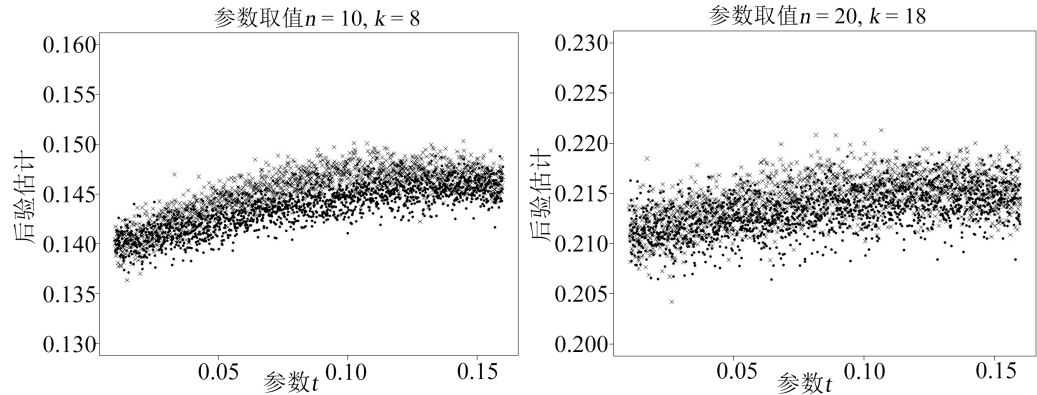
样本总数	$n=10$				$n=20$				$n=50$			
分布参数真值 r	0.5		1		0.5		1		0.5		1	
模型参数 t	2	3	1	2	2	3	1	2	3	4	1	2
模型参数 k_1	3	5	3	7	6	9	6	12	22	29	14	30
模型参数 k_2	6	7	6	8	12	15	12	16	37	41	32	41

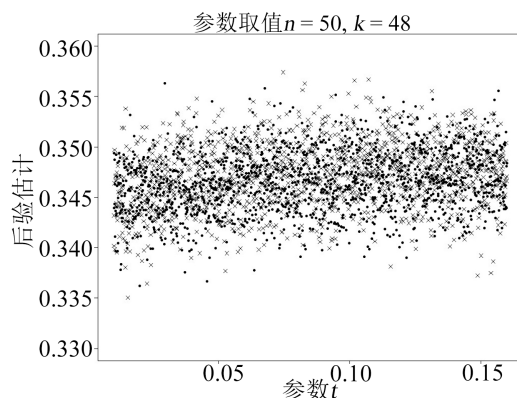
共轭条件下:

样本总数	$n=10$				$n=20$				$n=50$			
分布参数真值 r	0.5		1		0.5		1		0.5		1	
模型参数 t	2	3	1	2	2	3	1	2	3	4	1	2
模型参数 k_1	3	4	3	7	5	9	6	12	22	29	14	30
模型参数 k_2	6	7	6	8	12	15	12	16	37	41	32	41

7. 数据验证

(一) 已知参数取值情况, 在 *Jeffreys* 原则下, 后验分布真实值与优化值的散点对比图:





注：图中 $\cdot$ 表示原真实值的后验分布数据； $\times$ 表示现优化值的后验分布数据。

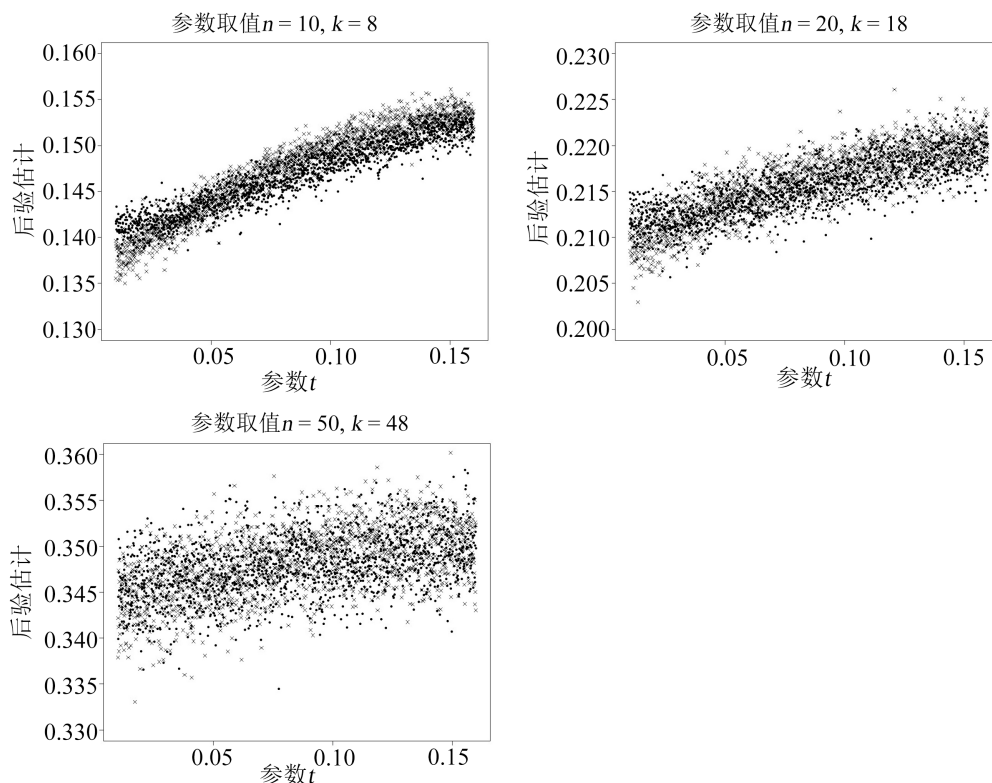
结论：

1) 由上述散点图中数据的分布情况可以确定，在 *Jeffreys* 原则下，后验分布的优化值与真实值间存在相合性并可控范围内的误差值。

2) 由数据图得，在参数 t 取 $(0.01, 0.16)$ 的范围下， n 取 10, 20 的小样本情形下，*Jeffreys* 原则下得到的后验分布数据整体偏小；同条件下， n 取 50 的较大样本情形下，后验分布数据整体较大；对比数据分布， n 取 10 时数据分布较为集中且有较为明显的增长性， n 取 20, 50 时数据分布较分散且稳定。

3) 整体数据对比下，可以确定在在 *Jeffreys* 原则下的后验分布，其真实值与优化值较为近似且数据真实有效。

(二) 已知参数取值情况，在共轭先验分布条件下，后验分布真实值与优化值的散点对比图：



注：图中 $\cdot$ 表示原真实值的后验分布数据； $\times$ 表示现优化值的后验分布数据。

结论:

1) 由上述散点图中数据的分布情况可以确定, 在共轭先验分布条件下, 后验分布的优化值与真实值间存在相合性并可控范围内的误差值。

2) 由数据图得, 在参数 t 取 $(0.01, 0.16)$ 的范围下, n 取 10, 20 的小样本情形下, 共轭先验分布所得到的后验分布数据整体偏小; 同条件下, n 取 50 的较大样本情形下, 后验分布数据整体较大。对比数据分布, n 取 10, 20 时数据分布较为集中且有较为明显的增长性, n 取 50 时数据分布较分散且稳定。

3) 对比整体数据, 可以确定在共轭先验分布条件下的后验分布, 其真实值与优化值较为近似且数据真实有效。

8. 结论

本文利用混合截尾方法, 对艾拉姆咖分布参数的后验分布进行研究。一方面对比两类基本的先验分布, 即无信息 *Jeffreys* 先验分布、有信息共轭先验分布, 给出后验分布结果的具体表达式; 同时利用优化理论, 对后验分布结果进行优化讨论, 给出以基本线性组合与加权线性组合下的参数的均方表达式, 并利用 MCMC 原理, 对结果进行数值模拟, 验证优化结果的准确性与拟合度, 得出基本线性组合的优化方案适用于小样本下, 对于加权线性组合的优化方案适用范围较广, 常见样本数据下均可使用, 且优化数据与原始数据偏差较小; 另一方面, 针对混合截尾方法, 以参数的极大似然估计结果确定截尾数的估计取值范围, 在给定的两类先验分布条件下, 利用数值模拟给出后验分布原始值与优化值的散点分布图, 并根据分布图, 给出截尾数的具体取值。在后续的研究中, 可以利用给定的截尾数对艾拉姆咖分布在不同损失函数下的参数的估计值进行讨论, 为后续相应模拟的研究给出了重要数据截点。

参考文献

- [1] 郑从平, 王涛, 谢有余. 偏正态数据下半参数混合效应模型的贝叶斯估计[J]. 重庆工商大学学报(自然科学版), 2023, 40(4): 93-98.
- [2] 徐宝, 蓝海, 赵仲达. 一种对称损失函数下反向帕累托分布形状参数的估计[J]. 南开大学学报(自然科学版), 2023, 56(1): 76-81.
- [3] 鄢伟安, 李欣忆, 张士杰, 等. 基于指数扩散退化过程的竞争失效模型的客观贝叶斯分析[J]. 华中科技大学学报(自然科学版), 2023, 51(11): 38-46.
- [4] 程伟丽, 李晓霞. 后验分布在贝叶斯统计课堂教学的延伸[J]. 高等数学研究, 2018, 21(1): 74-78.
- [5] 肖金安, 周玉茹. 双定数混合截尾下广义瑞利分布的统计分析[J]. 科学技术创新, 2025(9): 37-40.
- [6] 彭秀云, 赖俊峰, 卢静莉. 广义逐次截尾样本下改进的威布尔模型的研究[Z]. 呼和浩特: 内蒙古工业大学, 2025.
- [7] 李忠桂, 何书元. 右删失数据下分位数回归的光滑经验似然检验[J]. 应用概率统计, 2019, 35(2): 153-164.
- [8] 赵江南, 周菊玲. 基于左截断右删失数据的单参数 Pareto 分布参数估计[J]. 数学的实践与认识, 2017, 47(15): 254-257.
- [9] 周旭, 田茂再. 左截断右删失数据下伽马分布的参数推断[J]. 山东理工大学学报(自然科学版), 2023, 37(6): 1-6.
- [10] 张春雨, 刘禄勤. 定数截尾情形下 Poisson-Lomax 分布的 Bayes 估计[J]. 统计与决策, 2018, 34(10): 70-73.
- [11] 韩荣. 截尾数据下逆 Lomax 分布竞争失效模型的统计分析[D]: [硕士学位论文]. 贵阳: 贵州民族大学, 2025.
- [12] Wu, W.H., Wang, B.X., Chen, J.Y., Miao, J. and Guan, Q. (2022) Interval Estimation of the Two-Parameter Exponential Constant Stress Accelerated Life Test Model under Type-II Censoring. *Quality Technology & Quantitative Management*, 20, 751-762. <https://doi.org/10.1080/16843703.2022.2147688>
- [13] 龙芳, 蔡静, 朱艳. 逐步II型混合截尾下 Lomax 分布多部件应力强度模型的可靠性分析[J]. 广西师范大学学报(自然科学版), 2024, 42(2): 120-130.
- [14] Agnihotri, S., Singh, S.K. and Singh, U. (2023) Inferences for Multiple Interval Type-I Censoring Scheme. *Journal of Applied Statistics*, 50, 86-105.
- [15] Algarni, A., Almarashi, A.M., Kundu, D., Abd-Elmougod, G.A. and Abdel-Khalek, S. (2022) Comparison of Different

Confidence Intervals under Type-I Censoring Scheme. *Journal of Mathematics*, **2022**, Article ID: 1272045.
<https://doi.org/10.1155/2022/1272045>

- [16] 龙兵, 张忠占. 双定数混合截尾下两参数 Pareto 分布的统计分析[J]. 数学物理学报, 2022, 42(1): 269-281.
- [17] 龙兵, 张忠占. 双定时混合截尾下两参数 Pareto 分布的统计推断[J]. 应用概率统计, 2022, 38(5): 633-646.
- [18] 张尧庭, 陈汉峰. 贝叶斯统计推断[M]. 北京: 科学出版社, 1991.