

截断分布族的方差与熵

张晓伟*, 马冰心*, 马赛鹏

中国矿业大学(北京)理学院, 北京

收稿日期: 2024年3月3日; 录用日期: 2024年3月22日; 发布日期: 2024年4月30日

摘要

熵和方差都是用来量化信息或数据的特性的重要工具, 截断型分布是一类常见分布。本文研究了截断型分布族中熵和方差之间的关系: 首先研究了一般形式的截断性分布中熵和方差, 并分析了它们之间的关系; 之后研究了三种特殊截断型分布: 截断均匀分布、截断正态分布和截断Pareto分布, 分别计算它们的熵和方差, 并深入探讨了它们之间的关联性与相互影响。

关键词

方差, 熵, 截断型分布

Variance and Entropy of Truncated Family of Distributions

Xiaowei Zhang*, Bingxin Ma*, Saipeng Ma

School of Science, China University of Mining & Technology (Beijing), Beijing

Received: Mar. 3rd, 2024; accepted: Mar. 22nd, 2024; published: Apr. 30th, 2024

Abstract

Entropy and variance are essential metrics employed to quantify the properties of information or data, with truncated distribution being a prevalent type of distribution. This paper delves into the interrelation between entropy and variance within the truncated distribution family. Initially, the investigation pertains to entropy and variance in the context of general truncated distribution, analyzing their relationship. Then we study three special truncation distributions: truncated uni-

*相同贡献。

form distribution, truncated normal distribution and truncated Pareto distribution, calculate their entropy and variance respectively, and deeply discuss their correlation and mutual influence.

Keywords

Variance, Entropy, Truncated Distribution

Copyright © 2024 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

截断型分布族, 作为一类在特定区间外取值为零的概率分布, 是统计学中研究的重要对象。这些分布通过截断原始分布, 改变其形状和数字特征, 如均值、方差和偏度等, 从而更准确地描述现实世界中受到限制的数据。从截断正态分布到截断指数分布, 各种截断型分布在实际应用中发挥着不可或缺的作用。

随着统计学的不断发展, 对截断型分布族的研究已经从基本的定义和性质扩展到了参数估计、假设检验等更深层次的问题。研究者们不仅探索了多种参数估计方法, 还关注了截断操作对分布特性的影响。例如, Nadarajah S [1]从长尾分布出发, 研究五种常见长尾分布截断形式, 并求出矩的显式表达式。李红玲[2]对截断正态分布、截断指数分布等常见的截断分布的数字特征进行了深入的研究。师义民[3]研究 Linex 损失函数下, 一类双边截断型分布族参数的经验 Bayes (EB)估计问题, 并给出了该估计的收敛速度; 卢昆亮、赵林城[4]研究了双边截断参数的最小方差无偏估计, 给出参数的 UMVUE 的表达式及存在的充分必要条件; 文献《Estimating the Parameters of Truncated Distributions》[5]研究一些特殊截断分布的参数的估计, 为实际应用提供了理论支持。

截断型分布族在统计学中具有重要的地位。这类分布通过限制变量的取值范围来形成新的分布形态, 使得模型更加符合实际问题的约束条件。截断型分布不仅存在于理论研究中, 更广泛应用于实际生活中多个领域。在金融学中, 截断分布可用于描述风险调整后的收益分布; 在生物学和医学中, 它们可用于分析生命周期受限的数据。这些应用不仅展现了截断型分布族的实用性, 也凸显了对其进行深入研究的必要性。

在描述概率分布特性时, 方差和熵是两个不可或缺的统计量。方差作为衡量数据离散程度的关键指标, 反映了随机变量与其均值之间的偏离程度; 而熵则用于量化数据的不确定性或信息的混乱程度。对于截断分布族而言, 由于截断操作改变了数据的取值范围, 其方差和熵的大小也会受到影响。因此, 研究截断分布族的方差和熵的表达式, 以及它们之间的关系, 对于深入理解截断数据的特性具有重要意义。文章《On the Functional Relationship between Entropy and Variance with Related Applications》[6]深入探讨了熵与方差之间的函数关系, 为我们理解截断分布族的特性提供了新的视角。在具有共同方差的分布类内, 熵可以作为分布之间亲和性的度量。然而, 对于截断分布族而言, 其数字特征熵与方差的具体形式需要根据具体的分布类型来确定。在计算过程中, 我们必须充分考虑截断操作所带来的影响, 以确保结果的准确性。

本文旨在深入研究截断型分布族的方差和熵的特性及其关系, 通过推导截断分布族的方差和熵的表达式, 我们可以深入了解截断操作对它们的影响, 从而为实际应用提供理论支持。此外, 通过比较不同截断条件下的方差和熵的变化, 我们可以评估截断操作的效果, 为实际应用提供理论支持。

2. 截断型分布族

2.1. 定义

在统计学中，截断型分布族是由限制一些其他概率分布的域而产生的条件分布。在实际统计中，如果记录或了解事件仅限于高于或低于指定范围内的值，则会出现截断型分布。

设随机变量 X 的概率密度函数为 $f(x)$ ，分布函数为 $F(x)$ ，其中 $x \in (-\infty, +\infty)$ ，若对 x 的取值范围限制为 $a < x < b$ ，则其对应的截断型分布为

$$f(x|a < X < b) = \frac{g(x)}{F(b) - F(a)} = \frac{f(x)I(\{a < x < b\})}{F(b) - F(a)}$$

其中 $g(x) = \begin{cases} f(x), & a < x < b \\ 0, & \text{其他} \end{cases}$ ，特别地，当我们只进行左截断时 $F(b) = 1$ ，只进行右截断时 $F(a) = 0$ ，我们注意到[7]

$$\int_a^b f(x|a < X < b) dx = \frac{1}{F(b) - F(a)} \int_a^b g(x) dx = 1$$

下面给出截断型分布族概率密度函数的一般表达式：

双边截断型分布族

对于双边截断型分布族，其概率密度形式为

$$f(x; \theta_1, \theta_2) = k(\theta_1, \theta_2) h(x) I_{(\theta_1, \theta_2)}(x), \quad -\infty < \theta_1 < \theta_2 < \infty$$

其中 $h(x)$ 为 $(-\infty, +\infty)$ 上的非负可测函数，它在任一有限区间上可积，其中 $k(\theta_1, \theta_2) = 1 / \int_{\theta_1}^{\theta_2} h(u) du$ [8]

对每一对 (θ_1, θ_2) ， $f(x; \theta_1, \theta_2)$ 是 $[\theta_1, \theta_2]$ 上的一概率密度，分布函数 $F(x; \theta_1, \theta_2) = \int_{\theta_1}^x f(u; \theta_1, \theta_2) du$

单边截断型分布族

1) 左截断型分布族

对于左截断型分布族，其概率密度形式为

$$f(x|\theta) = k(\theta) h(x) I_{(\theta, b)}(x)$$

其中 $h(x)$ 在 (θ, b) 上非负可积， $-\infty \leq a < \theta$ ， $b \leq \infty$ ， $k(\theta) = 1 / \int_{\theta}^b h(u) du$

2) 右截断型分布族

对于右截断型分布族，其概率密度形式为

$$f(x|\theta) = k(\theta) h(x) I_{(a, \theta)}(x)$$

其中 $h(x)$ 在 (a, θ) 上非负可积， $\theta < b \leq \infty$ ， $a \geq -\infty$ ， $k(\theta) = 1 / \int_a^{\theta} h(u) du$ [9]。

2.2. 截断分布族的熵与方差

熵和方差为衡量数据不确定性和分散程度的两个相关但不完全相同的指标。熵主要关注数据的概率分布，而方差则关注数据与其均值的偏离程度。分布截断会引入新的不确定性。这种不确定性可能来自于截断边界的选择、原始数据的分布特性等因素。为了更深入地理解截断型分布族的熵与方差的关系，我们通过数学推导，探究截断型分布族熵与方差的关系。

2.2.1. 方差

由概率密度函数可得出

$$E(x) = \int_{\theta_1}^{\theta_2} xf(x; \theta_1, \theta_2) dx = \int_{\theta_1}^{\theta_2} xk(\theta_1, \theta_2)h(x)I_{(\theta_1, \theta_2)}(x) dx = \int_{\theta_1}^{\theta_2} xh(x) dx \cdot k(\theta_1, \theta_2)$$

$$E(x^2) = \int_{-\infty}^{+\infty} x^2 \int(x; \theta_1, \theta_2) dx = k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} x^2 h(x) dx$$

根据方差计算公式

$$\sigma^2 = E(x^2) - [E(x)]^2$$

得出双边截断型分布族的方差为

$$\sigma^2 = k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} x^2 h(x) dx - k^2(\theta_1, \theta_2) \left[\int_{\theta_1}^{\theta_2} xh(x) dx \right]^2$$

2.2.2. 熵

定理 1: 双边截断型分布族中熵的表达式为

$$H(x) = -\ln k(\theta_1, \theta_2) - k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) \ln h(x) dx$$

定理证明如下:

根据熵的计算公式

$$H = -\int p(x) \ln p(x) dx$$

$$H(x) = -\int_{\theta_1}^{\theta_2} f(x, \theta_1, \theta_2) \ln f(x, \theta_1, \theta_2) dx$$

$$= -\int_{\theta_1}^{\theta_2} k(\theta_1, \theta_2) h(x) (\ln k(\theta_1, \theta_2) + \ln h(x)) dx$$

$$= -k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} [h(x) \ln k(\theta_1, \theta_2) + h(x) \ln h(x)] dx$$

$$= -k(\theta_1, \theta_2) \ln k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) dx - k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) \ln h(x) dx$$

$$= -\ln k(\theta_1, \theta_2) - k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) \ln h(x) dx$$

得出双边截断型分布族的熵为

$$H(x) = -\ln k(\theta_1, \theta_2) - k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) \ln h(x) dx$$

□

2.2.3. 方差与熵的关系

方差 $\sigma^2 = k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} x^2 h(x) dx - k^2(\theta_1, \theta_2) \left[\int_{\theta_1}^{\theta_2} xh(x) dx \right]^2$ 与 $h(x), k(\theta_1, \theta_2)$ 有关, 而 $k(\theta_1, \theta_2)$ 可由 $h(x)$ 求出, 因此可以说方差取值取决于 $\theta_1, \theta_2, h(x)$ 。

熵 $H(x) = -\ln k(\theta_1, \theta_2) - k(\theta_1, \theta_2) \int_{\theta_1}^{\theta_2} h(x) \ln h(x) dx$ 与 $h(x), k(\theta_1, \theta_2)$ 有关, 而 $k(\theta_1, \theta_2)$ 可由 $h(x)$ 求出, 故熵依赖于 $\theta_1, \theta_2, h(x)$ 。

若函数 $h(x)$ 为已知函数, 熵与方差可根据上述已推导得到的公式求得。 $h(x)$ 为 $(-\infty, +\infty)$ 上的非负可测函数, 它在任一有限区间上可积。因此对于截断型分布族, 其熵和方差的计算依赖于其分布类型。

3. 特殊截断型分布

通过对于一般型截断型分布中熵和方差性质的研究,我们发现这两者均紧密依赖于具体的分布类型,且有一定的关联性。为了进一步地研究这种关系,选取三种特殊的截断分布:截断均匀分布、截断正态分布以及截断 Pareto 分布,深入探讨熵与方差的关系。

3.1. 截断均匀分布

3.1.1. 定义

1) 均匀分布

若连续型随机变量 x 的概率密度函数为 $p(x) = \begin{cases} 1/\beta, & \alpha - \beta/2 < x < \alpha + \beta/2 \\ 0, & \text{其它} \end{cases}$, 则称 x 服从

$[\alpha - \beta/2, \alpha + \beta/2]$ 上的均匀分布, 记作 $X \sim U[\alpha - \beta/2, \alpha + \beta/2]$ 。

$$\text{分布函数为 } F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < \alpha - \beta/2 \\ \frac{x - (\alpha - \beta/2)}{\beta}, & \alpha - \beta/2 \leq x \leq \alpha + \beta/2 \\ 1, & x \geq \alpha + \beta/2 \end{cases}$$

2) 截断均匀分布

截断型均匀分布是一种概率分布,它是在给定一定范围内的均匀分布的一种变体。通常均匀分布在一个特定的区间内具有相等的概率密度,但截断型均匀分布则限制了这个区间,使得随机变量只能在某个子区间内取值。

3.1.2. 截断均匀分布的熵与方差

1) 方差

由方差计算公式 $Var(x) = E(x^2) - E^2(x)$, 我们先分别计算 $E(x)$ 和 $E(x^2)$

$$E(x) = \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} xp(x) dx = \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} x \frac{1}{\beta} dx = \alpha$$

$$E(x^2) = \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} x^2 p(x) dx = \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} \frac{x^2}{\beta} dx = \alpha^2 + \frac{1}{12} \beta^2$$

则得出截断均匀分布的方差为

$$\sigma^2 = Var(x) = E(x^2) - E^2(x) = \frac{\beta^2}{12}$$

2) 熵

根据熵的计算公式

$$H = - \int p(x) \ln p(x) dx$$

得到截断均匀分布的熵为

$$H(x) = - \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} p(x) \ln p(x) dx = - \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} \frac{1}{\beta} \ln \frac{1}{\beta} dx = \frac{1}{\beta} \int_{\alpha - \frac{\beta}{2}}^{\alpha + \frac{\beta}{2}} \ln \beta dx = \ln \beta$$

3) 熵与方差的关系

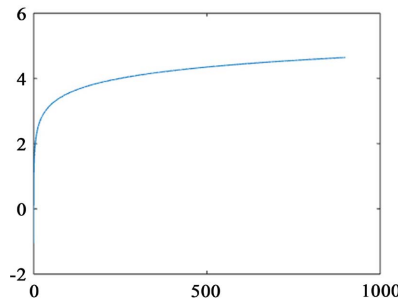
截断均匀分布中熵和方差具有如下的函数关系:

$$H(\sigma^2) = \frac{1}{2} \ln \sigma^2 + \ln 2\sqrt{3}$$

根据方差的表达式 $\sigma^2 = \frac{\beta^2}{12}$ ，可得 $\beta^2 = \frac{\sigma^2}{12}$ ，代入熵的表达式 $H(x) = \ln \beta$

$$H(x) = \ln \beta = \ln \sqrt{\frac{\sigma^2}{12}} = \frac{1}{2} \ln \sigma^2 + \ln 2\sqrt{3}$$

因此截断均匀分布中熵与方差的函数关系为： $H(\sigma^2) = \frac{1}{2} \ln \sigma^2 + \ln 2\sqrt{3}$



由截断均匀分布熵与方差的表达式画图拟合可知，截断均匀分布熵与方差之间存在线性关系，当截断均匀分布的方差增大时，熵会随着方差的增大而增大。

3.2. 截断正态分布

3.2.1. 定义

1) 正态分布

若随机变量 x 服从一个数学期望为 μ ，方差为 σ^2 的概率分布，且其概率密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

，则这个随机变量就称为正态随机变量，正态随机变量服从的分布就称为正态分布，

记作 $X \sim N(\mu, \sigma^2)$ 。设 x 是正态分布随机变量，其分布函数分别为 $\Phi(\mu, \sigma^2; x) = F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$

当 $\mu=0$ ， $\sigma^2=1$ 时，正态分布是标准正态分布。记 $\phi(x)$ 为标准正态分布的密度函数， $\Phi(x)$ 为标准正态分布的累积分布函数； $\phi(\mu, \sigma^2; x)$ 为正态分布的密度函数， $\Phi(\mu, \sigma^2; x)$ 为正态分布的累积分布函数。

2) 截断正态分布

截断正态分布(Truncated Normal Distribution)是在正态分布中界定随机变量进而从正态分布的分布函数中导出概率分布。它是截断分布(Truncated Distribution)的一种，即限制变量取值范围的一种分布。具体来说，如果一个随机变量 X 在 a 和 b 之间的取值服从正态分布 $N(\mu, \sigma^2)$ ，且 a 和 b 两个端点处的取值概率为 0，那么这个随机变量 X 就服从截断正态分布。记为 $X \sim \psi(\mu, \sigma, a, b; x)$ 。在数学上，截断正态分布的概率密度函数可以通过对正态分布的概率密度函数进行截断并归一化得，其概率密度函数为[10]

$$p(x) = \psi(\mu, \sigma, a, b; x) = \begin{cases} 0, & x \leq a \\ \frac{\phi(\mu, \sigma^2; x)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)}, & a < x < b \\ 0, & x \geq b \end{cases}$$

• 双边截断正态分布

设 x 是在左边点 a 和右边点 b 截断的正态分布随机变量，其密度函数为

$$p(x) = \frac{\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)}, \quad a \leq x \leq b$$

• 单边截断正态分布

1) 左截断正态分布：设 x 是在左边点 a 截断的正态分布随机变量，其密度函数为

$$p(x) = \frac{\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{1 - \Phi(\mu, \sigma^2; a)}, \quad a \leq x < \infty$$

2) 右截断正态分布：设 x 是在右边点 b 截断的正态分布随机变量，其密度函数为

$$p(x) = \frac{\frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\Phi(\mu, \sigma^2; b)}, \quad -\infty < x \leq b$$

其中， $\Phi(x)$ 是标准正态分布的累积分布函数。需要注意的是，这里的概率密度函数是归一化后的结果，即保证了在截断区间内的积分为 1。

3.2.2. 截断正态分布的熵与方差

目前对于截断正态分布的研究，张维铭[11]研究截断正态分布的性质、参数估计，及根据截断正态分布制定计量抽样方案的方法。Héctor J. Gómez、Neveka M. Olmos 等人[12]研究截尾正态分布的矩、风险函数和熵等性质，并利用矩和极大似然方法进行了参数估计。G. Baikunth Nath [13]研究二元正态分布的矩，从线性截断的观测值中得到二元正态分布参数的估计。李红玲[2]计算了截断正态函数的数字特征一期望及方差。对于截断正态分布的数字特征：熵与方差，许多研究者李红玲[2]，Burkardt J [10]等已推导出方差的表达式，但较少关注截断正态分布的熵，以及熵与方差关系的相关研究，因此我们通过数学推导截断正态分布的方差与熵，观察截断对两者带来的变化，深入了解两者的关系。

1) 方差

正态分布方差为： $Var(X) = \sigma^2$

双边截断正态分布的方差为

$$Var(X) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right], \quad \text{其中 } \alpha = \frac{a - \mu}{\sigma}; \quad \beta = \frac{b - \mu}{\sigma}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xp(x)dx = \int_{-\infty}^a xp(x)dx + \int_a^b xp(x)dx + \int_b^{\infty} xp(x)dx \\ &= 0 + \int_a^b \frac{x\phi(\mu, \sigma^2; x)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} dx + 0 = \int_a^b \frac{(x - \mu + \mu)\phi(\mu, \sigma^2; x)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} dx \\ &= \frac{\int_a^b \mu\phi(\mu, \sigma^2; x) dx}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} + \frac{\int_a^b (x - \mu)\phi(\mu, \sigma^2; x) dx}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} \end{aligned}$$

$$\begin{aligned}
 &= \mu + \frac{\int_a^b \frac{1}{\sigma} (x-\mu) \phi\left(\frac{x-\mu}{\sigma}\right) d(x-\mu)x}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} = \mu + \frac{\int_a^b \frac{1}{\sigma} (x-\mu) \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} d(x-\mu)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} \\
 &= \mu + \frac{-\sigma^2 \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \Big|_a^b}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} = \mu + \frac{-\sigma \phi\left(\frac{x-\mu}{\sigma}\right) \Big|_a^b}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} \\
 &= \mu - \sigma \frac{\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} = \mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}
 \end{aligned}$$

$$\begin{aligned}
 E(X^2) &= \int_{-\infty}^{\infty} x^2 p(x) dx = \int_a^b x^2 \frac{\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx \\
 &= \int_a^b \left[(x-\mu)^2 + 2x\mu - \mu^2 \right] \frac{\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx \\
 &= \int_a^b (x-\mu)^2 \frac{\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx + 2\mu \int_a^b \frac{x\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx \\
 &\quad - \mu^2 \int_a^b \frac{\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx \\
 &= \int_a^b (x-\mu)^2 \frac{\phi(\mu, \sigma^2, x)}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} dx + 2\mu E(x) - \mu^2 \\
 &= \frac{\int_a^b (x-\mu)^2 \phi(\mu, \sigma^2; x) dx}{\Phi(\mu, \sigma^2, b) - \Phi(\mu, \sigma^2, a)} + \sigma^2 + 2\mu \left[\mu - \sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right] - \mu^2 \\
 &= \frac{-\sigma^2 [\beta\phi(\beta) - \alpha\phi(\alpha)] + \sigma^2 [\Phi(\beta) - \Phi(\alpha)]}{\Phi(\beta) - \Phi(\alpha)} + \sigma^2 + \mu^2 - 2\mu\sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \\
 &= -\sigma^2 \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} + \sigma^2 + \mu^2 - 2\mu\sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)}
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(X) &= \left[-\sigma^2 \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} + \sigma^2 + \mu^2 - 2\mu\sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right] \\
 &\quad - \left[\mu^2 - 2\mu\sigma \frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} + \sigma^2 \left[\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right]^2 \right] \\
 &= -\sigma^2 \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} + \sigma^2 - \sigma^2 \left[\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right]^2 \\
 &= \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]
 \end{aligned}$$

因此可得出双边截断正态分布的方差为

$$\text{Var}(X) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]$$

对于单边截断正态分布，可视为是双边截断正态分布的特殊情况。

a) 左截断正态分布：对于左截断正态分布，可以看作是双边截断正态分布右截断点 b 为正无穷的特殊情况，则 $\Phi(\beta) = \Phi\left(0, 1; \frac{b-\mu}{\sigma}\right) = 1$, $b \rightarrow +\infty$ ，那么左截断正态分布的方差为：

$$\text{Var}(X) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{1 - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{1 - \Phi(\alpha)} \right)^2 \right], a \leq x \leq b = \infty$$

b) 右截断正态分布：对于右截断正态分布，可以看作是双边截断正态分布左截断点 a 为负无穷的特殊情况，则 $\Phi(\alpha) = \Phi\left(0, 1; \frac{a-\mu}{\sigma}\right) = 0$, $a \rightarrow -\infty$ ，那么右截断正态分布的方差为：

$$\text{Var}(X) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta)} \right)^2 \right], -\infty = a \leq x \leq b$$

2) 熵

正态分布中熵的表达式为：

$$H(\sigma^2) = \frac{1}{2} \ln(2\pi e \sigma^2)$$

$$\begin{aligned} H(x) &= -E \left[\ln \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \right] = -E \left[\ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) + \ln \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \right] \\ &= -E \left[\ln \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \right] - E \left[\ln \left(e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) \right] = \frac{1}{2} \ln(2\pi\sigma^2) - E \left[-\frac{1}{2\sigma^2} (x-\mu)^2 \ln e \right] \\ &= \frac{1}{2} \ln(2\pi\sigma^2) + \frac{\ln e}{2\sigma^2} E[(x-\mu)^2] = \frac{1}{2} \ln(2\pi\sigma^2) + \frac{\log e}{2\sigma^2} \sigma^2 \\ &= \frac{1}{2} \ln(2\pi\sigma^2) + \frac{1}{2} \ln e = \frac{1}{2} \ln(2\pi e \sigma^2) \end{aligned}$$

定理 2：双边截断正态分布中熵的表达式为：

$$H(x) = \frac{\ln \sqrt{2\pi}\sigma + \ln M}{M} [\Phi(\beta) - \Phi(\alpha)] + \frac{1}{M} \left[\frac{\text{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^b \right]$$

定理证明如下：

$$\text{令 } M = \Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)$$

$$\begin{aligned}
 H(x) &= -\int_{-\infty}^{\infty} p(x) \ln p(x) dx = -\int_a^b \frac{\phi(\mu, \sigma^2; x)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} \ln \frac{\phi(\mu, \sigma^2; x)}{\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)} dx \\
 &= -\int_a^b \frac{\phi(\mu, \sigma^2; x)}{M} \ln \frac{\phi(\mu, \sigma^2; x)}{M} dx = -\int_a^b \frac{\phi(\mu, \sigma^2; x)}{M} \ln [\phi(\mu, \sigma^2; x) - M] dx \\
 &= -\frac{1}{M} \int_a^b \phi(\mu, \sigma^2; x) \ln [\phi(\mu, \sigma^2; x)] dx + \frac{\ln M}{M} \int_a^b \phi(\mu, \sigma^2; x) dx \\
 &= -\frac{1}{M} \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \ln \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{\ln M}{M} \int_a^b \phi(\mu, \sigma^2; x) dx \\
 &= -\frac{1}{M} \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left(-\ln \sqrt{2\pi\sigma} - \frac{(x-\mu)^2}{2\sigma^2} \right) dx + \frac{\ln M}{M} \int_a^b \phi(\mu, \sigma^2; x) dx \\
 &= \frac{1}{M} \int_a^b \frac{\ln \sqrt{2\pi\sigma}}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx + \frac{1}{M} \int_a^b \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{(x-\mu)^2}{2\sigma^2} dx + \frac{\ln M}{M} \int_a^b \phi(\mu, \sigma^2; x) dx \\
 &= \frac{\ln \sqrt{2\pi\sigma}}{M} \int_a^b \phi(\mu, \sigma^2; x) dx + \frac{1}{\sqrt{2\pi M \sigma}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{(x-\mu)^2}{2\sigma^2} dx + \frac{\ln M}{M} \int_a^b \phi(\mu, \sigma^2; x) dx \\
 &= \frac{\ln \sqrt{2\pi\sigma} + \ln M}{M} \Phi(\mu, \sigma^2; x) \Big|_a^b + \frac{1}{\sqrt{2\pi M \sigma}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{(x-\mu)^2}{2\sigma^2} dx \\
 &= \frac{\ln \sqrt{2\pi\sigma} + \ln M}{M} (\Phi(\mu, \sigma^2; b) - \Phi(\mu, \sigma^2; a)) + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^b \right] \\
 &= \frac{\ln \sqrt{2\pi\sigma} + \ln M}{M} [\Phi(\beta) - \Phi(\alpha)] + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^b \right]
 \end{aligned}$$

因此可得出双边截断正态分布的熵为

$$H(x) = \frac{\ln \sqrt{2\pi\sigma} + \ln M}{M} [\Phi(\beta) - \Phi(\alpha)] + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^b \right]$$

□

对于单边截断正态分布，可视为是双边截断正态分布的特殊情况。

1) 左截断正态分布：对于左截断正态分布，可以看作为双边截断正态分布右截断点 b 为正无穷的特殊情况，则 $\Phi(\beta) = \Phi\left(0, 1; \frac{b-\mu}{\sigma}\right) = 1$, $b \rightarrow +\infty$, 那么左截断正态分布的方差为：

$$H(X) = \frac{\ln \sqrt{2\pi\sigma} + \ln M}{M} [1 - \Phi(\alpha)] + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^{+\infty} \right], \quad a \leq x \leq b = +\infty$$

2) 右截断正态分布: 对于右截断正态分布, 可以看作是双边截断正态分布左截断点 a 为负无穷的特殊情况, 则 $\Phi(\alpha) = \Phi\left(0, 1; \frac{a-\mu}{\sigma}\right) = 0, a \rightarrow -\infty$, 那么右截断正态分布的方差为:

$$H(X) = \frac{\ln\sqrt{2\pi}\sigma + \ln M}{M} \Phi(\beta) + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_{-\infty}^b \right], -\infty = a \leq x \leq b$$

3) 熵与方差的关系

- 正态分布

熵的大小与方差有着直接的联系, 可以用数学表达式直接表示:

$$H(\sigma^2) = \frac{1}{2} \ln(2\pi e\sigma^2)$$

对于正态分布来说, 正态分布的熵随着方差的增大而增大, 方差越大, 意味着数据的离散程度越高, 分布越广泛, 因此需要更多的信息来描述这个分布, 即熵越大。

- 截断正态分布

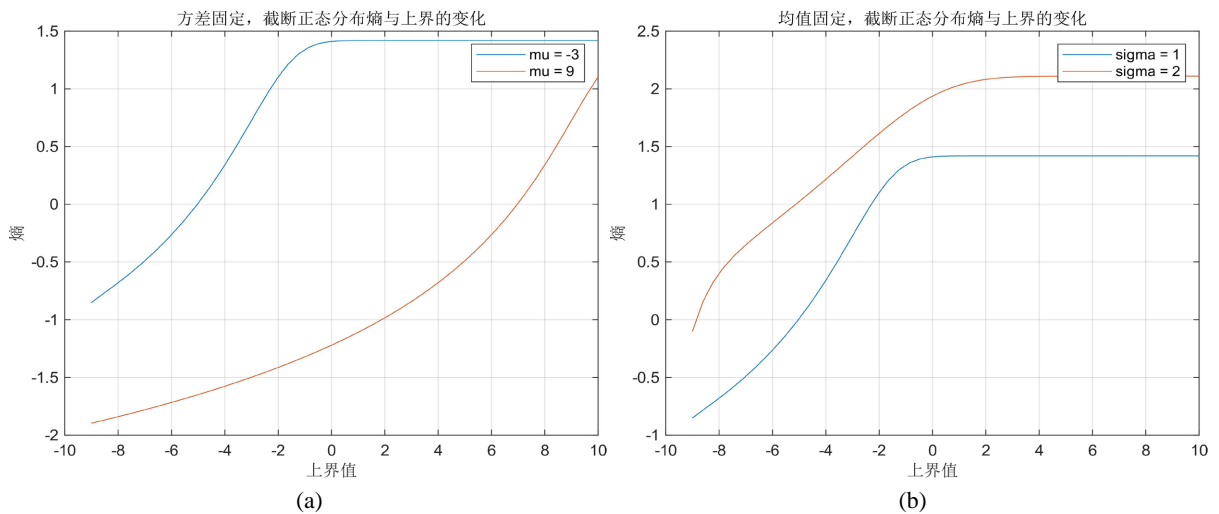
双边截断正态分布的方差与熵为:

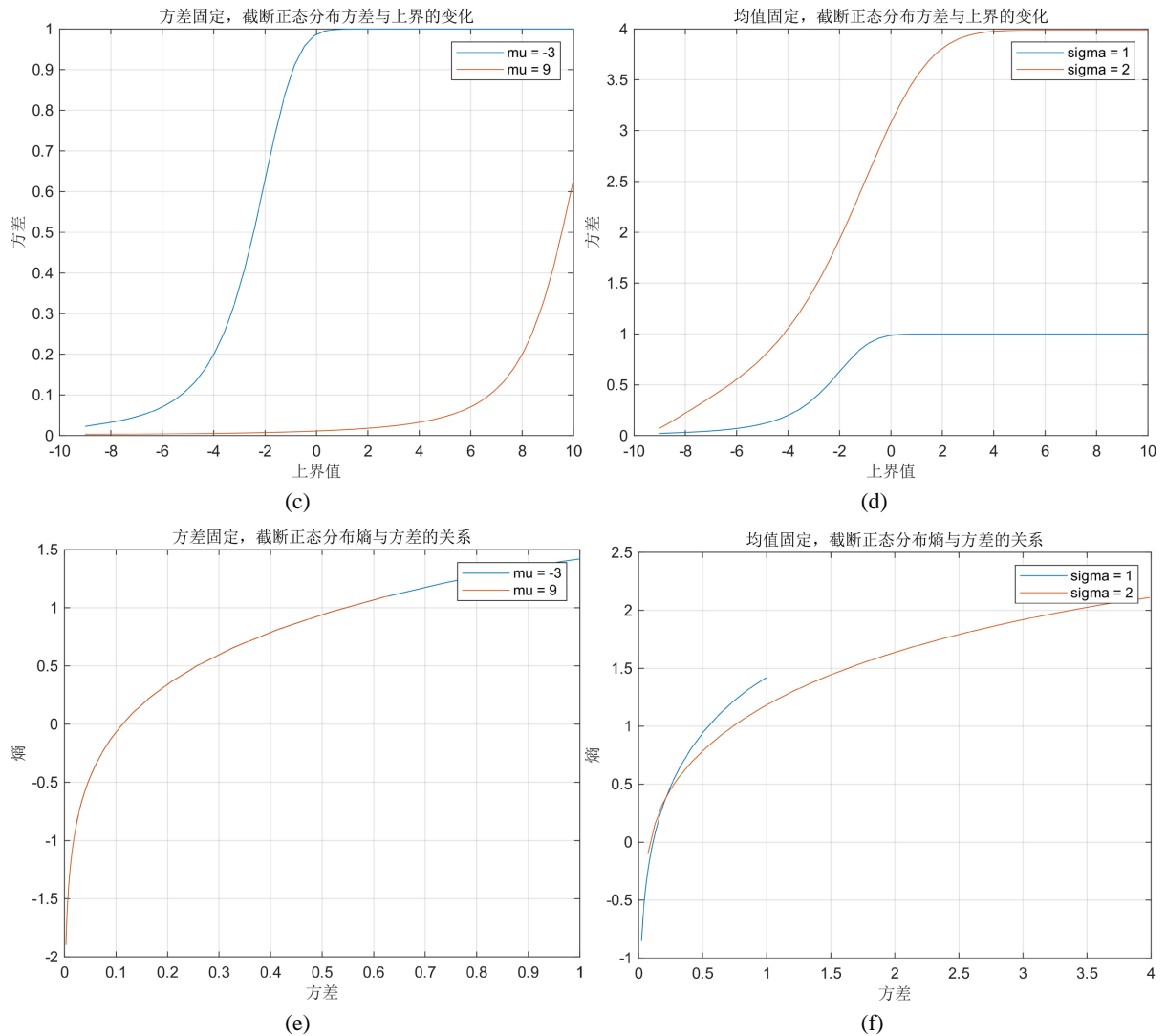
$$\operatorname{Var}(X) = \sigma^2 \left[1 - \frac{\beta\phi(\beta) - \alpha\phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} - \left(\frac{\phi(\beta) - \phi(\alpha)}{\Phi(\beta) - \Phi(\alpha)} \right)^2 \right]$$

$$H(x) = \frac{\ln\sqrt{2\pi}\sigma + \ln M}{M} [\Phi(\beta) - \Phi(\alpha)] + \frac{1}{M} \left[\frac{\operatorname{erf}\left(\frac{\sqrt{2}x}{2}\right)}{4} - \frac{xe^{-\frac{x^2}{2}}}{2} \Big|_a^b \right]$$

由其表达式可知, 截断正态分布熵与方差之间的关系并非简单的线性关系。

为了更直观地探索截断均匀分布中熵与方差的关系, 利用 MATLAB 软件进行画图拟合。在这一过程中, 固定截断下界为-10, 然后变化上界以改变截断范围。观察截断正态分布中熵与方差随截断范围改变而产生的变化。





经过实验模拟发现,随着截断范围的变化(即 a 和 b 的变化),截断正态分布的熵和方差都会发生变化。通常,当截断范围变窄时,分布的离散程度减小,方差降低,但熵也可能降低。当截断范围变宽时,分布的离散程度增加,方差增大,同时熵也可能增大。从熵与方差的关系图可以看出,方差增大,熵同时也增大,成正相关关系。但这并不是简单的线性关系,都受到截断正态分布具体参数(如均值和标准差)的影响。

3.3. 截断 Pareto 分布

3.3.1. 定义

1) Pareto 分布

帕累托分布是以意大利经济学家维弗雷多·帕雷托命名的,从大量真实世界的现象中发现的幂次定律分布。帕累托分布可以归纳为一个非常简洁的表述:通过市场交易,20%的人将占有80%的社会财富。由于 Pareto 分布具有递减的失效率函数,经常用来描述诸如个人收入(收入越高,获得更高收入的能力就会增加)、某种药理过程后病人的存活时间(存活时间越长,能够继续存活更长时间的可能性就越高)等模型。其它模型例如城市人口容量、自然现象的发生、股票价格波动、保险风险、商业失效等,也都可以

用 Pareto 分布来描述。

设 x 为一个非负连续随机变量, 如果 x 具有如下的概率密度函数 $f(x)$ 和分布函数 $F(x)$, 对任意 $\alpha > 0, \beta > 0$,

$$f(x) = \begin{cases} \alpha\beta^\alpha x^{-\alpha-1}, & x \geq \beta \\ 0, & x < \beta \end{cases}, \quad F(x) = \begin{cases} 1 - \left(\frac{\beta}{x}\right)^\alpha, & x \geq \beta \\ 0, & x < \beta \end{cases}$$

则称 x 服从参数为 α 与 β 的 Pareto 分布, 记做 $X \sim \text{Pareto}(\alpha, \beta)$ [14]

2) 截断 Pareto 分布

原 Pareto 分布的概率密度函数 $p(x|\alpha, \beta) = \alpha\beta^\alpha x^{-\alpha-1}, \alpha > 0, x \geq \beta > 0$ 属于左截型分布。

现假定 x 取值上界为 γ , 则右截断 Pareto 分布与双截断 Pareto 分布的概率密度函数均为[7]

$$p(x|\alpha, \beta, \gamma) = \frac{\alpha x^{-\alpha-1}}{\beta^{-\alpha} - \gamma^{-\alpha}}, 0 < \beta \leq x \leq \gamma < \infty$$

3.3.2. 截断 Pareto 分布的熵与方差

关于 Pareto 分布的研究, 相关学者已经做了很多工作, 李海芬, 茆诗松[14]给出了 Pareto 分布检验的图检验法、相关系数 R2 检验法, 并且讨论了 R2 检验对常用非 Pareto 分布的功效; 单国栋[15]研究了对于截断数据的 Pareto 分布的 Bayes 分析, Aban, I. B., Meerschaert, M. M.和 Panorska, A. K. [16]研究了截断 Pareto 分布的估计量, 研究了它们的性质并说明了一种检验拟合的方法; 对于截断 Pareto 分布的数字特征, 李红玲在文献[2]已计算截断 Pareto 分布的数学期望及方差, 然而截断 Pareto 分布的熵以及熵和方差关系的研究, 文献中较少涉及, 本次研究主要针对截断 Pareto 分布进行熵和方差的计算并且对二者的关系进行讨论。

1) 方差

- 左截断 Pareto 分布

由概率密度函数得出

$$E(x) = \int_{\beta}^{\infty} xp(x|\alpha, \beta)dx = \int_{\beta}^{\infty} \alpha\beta^\alpha x^{-\alpha}dx = \frac{\alpha\beta}{\alpha-1}$$

$$E(x^2) = \int_{\beta}^{\infty} x^2 p(x|\alpha, \beta)dx = \int_{\beta}^{\infty} \alpha\beta^\alpha x^{-\alpha+1}dx = \frac{\alpha\beta^2}{\alpha-2}$$

根据方差计算公式

$$\sigma^2 = E(x^2) - [E(x)]^2$$

得出左截断 Pareto 分布的方差为

$$\sigma^2 = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$$

- 右截断 Pareto 分布和双边截断 Pareto 分布中方差的表达式为

$$\sigma^2 = \frac{\frac{\alpha}{2-\alpha}(\gamma^{2-\alpha} - \beta^{2-\alpha})}{\beta^{-\alpha} - \gamma^{-\alpha}} - \frac{\left(\frac{\alpha}{1-\alpha}\right)^2(\gamma^{1-\alpha} - \beta^{1-\alpha})^2}{(\beta^{-\alpha} - \gamma^{-\alpha})^2}, \alpha > 2$$

由概率密度函数得出

$$E(x) = \int_{\beta}^{\gamma} xp(x|\alpha, \beta, \gamma)dx = \int_{\beta}^{\gamma} \frac{\alpha x^{-\alpha}}{\beta^{-\alpha} - \gamma^{-\alpha}} dx = \frac{\alpha}{1-\alpha} \frac{(\gamma^{1-\alpha} - \beta^{1-\alpha})}{\beta^{-\alpha} - \gamma^{-\alpha}}$$

$$E(x^2) = \int_{\beta}^{\gamma} x^2 p(x|\alpha, \beta, \gamma)dx = \int_{\beta}^{\gamma} \frac{\alpha x^{-\alpha+1}}{\beta^{-\alpha} - \gamma^{-\alpha}} dx = \frac{\alpha}{2-\alpha} \frac{(\gamma^{2-\alpha} - \beta^{2-\alpha})}{\beta^{-\alpha} - \gamma^{-\alpha}}$$

根据方差计算公式得出右截断 Pareto 分布和双截断 Pareto 分布的方差为

$$\sigma^2 = \frac{\alpha}{2-\alpha} \frac{(\gamma^{2-\alpha} - \beta^{2-\alpha})}{\beta^{-\alpha} - \gamma^{-\alpha}} - \left(\frac{\alpha}{1-\alpha} \right)^2 \frac{(\gamma^{1-\alpha} - \beta^{1-\alpha})^2}{(\beta^{-\alpha} - \gamma^{-\alpha})^2}, \alpha > 2$$

2) 熵

- 左截断 Pareto 分布

根据熵的计算公式

$$H = -\int p(x) \ln p(x) dx$$

得到左截 Pareto 分布的熵为

$$H = -\int_{\beta}^{\infty} \alpha \beta^{\alpha} x^{-\alpha-1} \ln(\alpha \beta^{\alpha} x^{-\alpha-1}) dx = -\alpha \beta^{\alpha} \int_{\beta}^{\infty} x^{-\alpha-1} \left[\ln(\alpha \beta^{\alpha}) + \ln x^{-\alpha-1} \right] dx$$

$$= -\left[\ln \alpha \beta^{\alpha} - (\alpha + 1) \left(\ln \beta + \frac{1}{\alpha} \right) \right] = \ln \beta - \ln \alpha + \frac{\alpha + 1}{\alpha}$$

定理 3: 右截断 Pareto 分布和双边截断 Pareto 分布中熵的表达式为

$$H = \frac{(\alpha + 1) [\beta^{-\alpha} \ln \beta - \gamma^{-\alpha} \ln \gamma]}{\beta^{-\alpha} - \gamma^{-\alpha}} - \ln \frac{\alpha}{\beta^{-\alpha} - \gamma^{-\alpha}} + \frac{\alpha + 1}{\alpha}$$

定理证明如下:

根据熵的计算公式

$$H = -\int p(x) \ln p(x) dx$$

将右截断 Pareto 分布和双边截断 Pareto 分布的概率密度函数

$$p(x|\alpha, \beta, \gamma) = \frac{\alpha x^{-\alpha-1}}{\beta^{-\alpha} - \gamma^{-\alpha}}, 0 < \beta \leq x \leq \gamma < \infty$$

代入得到

$$H = -\int_{\beta}^{\gamma} \frac{\alpha x^{-\alpha-1}}{\beta^{-\alpha} - \gamma^{-\alpha}} \ln \left(\frac{\alpha x^{-\alpha-1}}{\beta^{-\alpha} - \gamma^{-\alpha}} \right) dx$$

$$= -\frac{\alpha}{\beta^{-\alpha} - \gamma^{-\alpha}} \int_{\beta}^{\gamma} x^{-\alpha-1} \left[\ln \left(\frac{\alpha}{\beta^{-\alpha} - \gamma^{-\alpha}} \right) + \ln x^{-\alpha-1} \right] dx$$

$$= \frac{(\alpha + 1) [\beta^{-\alpha} \ln \beta - \gamma^{-\alpha} \ln \gamma]}{\beta^{-\alpha} - \gamma^{-\alpha}} - \ln \frac{\alpha}{\beta^{-\alpha} - \gamma^{-\alpha}} + \frac{\alpha + 1}{\alpha}$$

□

3) 熵与方差的关系

定理 4: 左截断 Pareto 分布中熵和方差具有如下的函数关系:

$$H(\sigma^2) = \frac{1}{2} \ln \sigma^2 + \frac{1}{2} \ln \left[\frac{(\alpha-1)^2(\alpha-2)}{\alpha^3} \right] + \frac{\alpha+1}{\alpha}$$

定理证明如下:

假设 α 已知, 根据方差的表达式, $\sigma^2 = \frac{\alpha\beta^2}{(\alpha-1)^2(\alpha-2)}, \alpha > 2$

得出

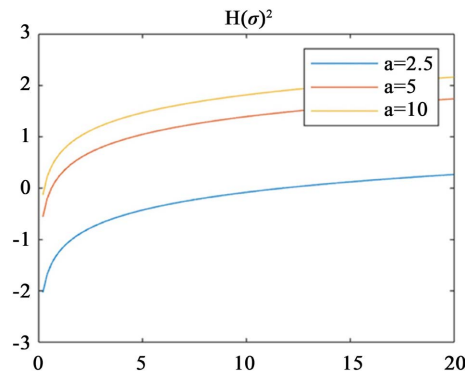
$$\beta^2 = \frac{(\alpha-1)^2(\alpha-2)\sigma^2}{\alpha}$$

代入熵的表达式 $H = \ln \beta - \ln \alpha + \frac{\alpha+1}{\alpha}$, 我们得到

$$\begin{aligned} H &= \ln \beta - \ln \alpha + \frac{\alpha+1}{\alpha} = \frac{1}{2} \ln \beta^2 - \ln \alpha + \frac{\alpha+1}{\alpha} \\ &= \frac{1}{2} \ln \frac{(\alpha-1)^2(\alpha-2)\sigma^2}{\alpha} - \ln \alpha + \frac{\alpha+1}{\alpha} \\ &= \frac{1}{2} \ln \sigma^2 + \frac{1}{2} \ln \left[\frac{(\alpha-1)^2(\alpha-2)}{\alpha^3} \right] + \frac{\alpha+1}{\alpha} \end{aligned}$$

□

由二者的关系式可以看出, 熵和方差的关系还受参数 α 大小的影响, 我们令 α 分别为 2.5, 5 和 10, 得到熵和方差的关系如下所示:



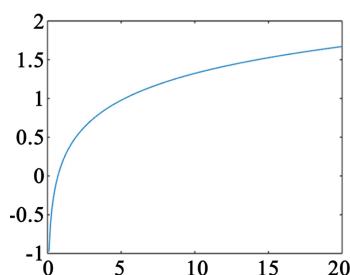
可以看出, 当左截断 Pareto 分布的方差增大时, 熵会随着方差的增大而增大。

在右截断及双截断 Pareto 分布中, 方差 $\sigma^2 = \frac{\alpha}{2-\alpha} \frac{(\gamma^{2-\alpha} - \beta^{2-\alpha})}{\beta^{-\alpha} - \gamma^{-\alpha}} - \left(\frac{\alpha}{1-\alpha} \right)^2 \frac{(\gamma^{1-\alpha} - \beta^{1-\alpha})^2}{(\beta^{-\alpha} - \gamma^{-\alpha})^2}, \alpha > 2$, 熵

$$H = \frac{(\alpha+1)[\beta^{-\alpha} \ln \beta - \gamma^{-\alpha} \ln \gamma]}{\beta^{-\alpha} - \gamma^{-\alpha}} - \ln \frac{\alpha}{\beta^{-\alpha} - \gamma^{-\alpha}} + \frac{\alpha+1}{\alpha}$$

其熵与方差的关系我们无法直观看出, 可以通过对右截断及双截断 Pareto 分布的参数设置具体数值并列举若干数据点, 对它们的熵和方差进行计算从而得出熵与方差的关系。

当我们令形状参数 $\alpha = 9$ ，截断参数 $\gamma = 3\beta$ 时，通过计算熵和方差的表达式我们得到熵和方差的关系图如下所示：



可以看出，与左截断 Pareto 分布中熵和方差的关系类似，右截断及双截断 Pareto 分布中熵和方差的关系同样呈正相关，当方差增大时，熵也会随之增大。

4. 结论

本文深入探讨了截断型分布中熵和方差的性质，发现两者在截断型分布中均展现出对分布类型的依赖性，并显示出一定的相关性。为了进一步理解这种关系，我们特别选取了三种截断分布进行深入研究：截断均匀分布、截断正态分布以及截断 Pareto 分布。通过计算这些特殊截断分布中的熵和方差，截断均匀分布的熵和方差之间呈现出较为简单的数学关系；而截断正态分布和截断 Pareto 分布的熵和方差则相对复杂。这些结果为我们在实际应用中选择合适的截断分布提供了理论依据。综合以上分析，我们可以得出以下结论：

截断型分布中熵和方差之间存在一定的关联性，这种关联性随着截断参数的变化而发生变化。

不同类型的截断分布具有不同的熵和方差特性，因此在实际应用中需要根据具体情况选择合适的截断分布。

通过研究截断型分布中熵和方差的关系，我们可以更深入地理解数据的内在规律和特性，为数据分析和信息处理提供有力的支持。展望未来，我们可以进一步拓展截断型分布中熵和方差关系的研究范围，探索更多类型的截断分布以及它们在实际应用中的表现，还可以将熵和方差与其他信息度量指标相结合，从而构建更全面的信息度量体系。

致 谢

在此，向指导老师李再兴教授表达最诚挚的感谢。感谢李老师在“大创”(大学生创新性训练项目)过程中的耐心指导和帮助。在整个研究过程中，正是李老师的精心指导、无私帮助和宝贵建议，让我们得以顺利完成研究并取得成果。再次感谢李老师的辛勤付出和无私奉献，我们深感荣幸和感激。

基金项目

本论文得到中国矿业大学(北京)大学生创新性训练项目的资助。

参考文献

- [1] Nadarajah, S. (2009) Some Truncated Distributions. *Acta Applicandae Mathematicae*, **106**, 105-123. <https://doi.org/10.1007/s10440-008-9285-4>
- [2] 李红玲. 截断密度函数分布的数字特征[J]. 河北北方学院学报(自然科学版), 2009, 25(3): 6-10.
- [3] 师义民. 双边截断型分布族参数的经验 Bayes 估计[J]. 高校应用数学学报 A 辑(中文版), 2000(4): 475-483.

- <https://doi.org/10.13299/j.cnki.amjcu.000940>
- [4] 卢昆亮, 赵林城. 双边截断参数的最小方差无偏估计[J]. 应用数学学报, 1982(2): 177-185.
- [5] Mittal, M.M. (1984) Estimating the Parameters of Truncated Distributions. PhD Thesis, Old Dominion University, Norfolk.
- [6] Mukherjee, D. and Ratnaparkhi, M.V. (1986) On the Functional Relationship between Entropy and Variance with Related Applications. *Communications in Statistics-Theory and Methods*, **15**, 291-311. <https://doi.org/10.1080/03610928608829122>
- [7] HandWiki (2023) Truncated Distribution. https://handwiki.org/wiki/index.php?title=Truncated_distribution&oldid=3014283
- [8] 陈桂景, 过纪扬, Michael A. Stephens. 关于截断型分布族中估计的经验过程的渐近分布的进一步研究[J]. 系统与数学, 1988(1): 32-41.
- [9] 黄藏红, 王华敏. Linex 损失及 NA 样本下单边截断型分布族参数函数的 EB 估计[J]. 洛阳师范学院学报, 2008(5): 5-8. <https://doi.org/10.16594/j.cnki.41-1302/g4.2008.05.002>
- [10] Burkardt, J. (2014) The Truncated Normal Distribution.
- [11] 张维铭. 截断正态分布的计量抽样方案[J]. 浙江理工大学学报, 2006(2): 182-188.
- [12] Gómez, H.J., Olmos, N.M., Varela, H., *et al.* (2018) Inference for a Truncated Positive Normal Distribution. *Applied Mathematics*, **33**, 163-176. <https://doi.org/10.1007/s11766-018-3354-x>
- [13] Nath, G.B. (1972) Moments of a Linearly Truncated Bivariate Normal Distribution. *Australian Journal of Statistics*, **14**, 97-102. <https://doi.org/10.1111/j.1467-842X.1972.tb00345.x>
- [14] 李海芬, 茆诗松. Pareto 分布的检验[J]. 徐州师范大学学报(自然科学版), 2004(3): 12-16.
- [15] 单国栋. 截断 Pareto 分布的 Bayes 分析[J]. 长春大学学报, 2008(10): 5-7+10.
- [16] Aban, I.B., Meerschaert, M.M. and Panorska, A.K. (2006) Parameter Estimation for the Truncated Pareto Distribution. *Journal of the American Statistical Association*, **101**, 270-277. <http://www.jstor.org/stable/30047455> <https://doi.org/10.1198/016214505000000411>