arXiv:2302.02486v1 [stat.ME] 5 Feb 2023

# 对数正态差分布：性质、估计与增长

Robert Parham*

星期二 7th 2023年2月

**摘要**

本文介绍对数正态差分布（DLN）。配套论文 Parham （2023年）论证了 DLN 是自然界中的一种基本分布，并展示了如何通过简单应用中心极限定理，在许多迥异的现象中得到 DLN。在本文中，我刻画其概率密度函数、累积分布函数、矩和参数估计量；利用球面分布理论将其推广到 N 维；介绍处理其标志性“双指数”特征的方法；并用它将增长测度推广到可能取负值且服从 DLN 的变量。我还进行了蒙特卡洛实验，以验证文中所述估计量和度量的一些性质。

*JEL*分类：C13、C46、C65关键词：重尾、分布、对数正态分布、增长。

*弗吉尼亚大学（robertp@virginia.edu）。用于处理DLN 分布的代码可向作者索取。

# 1 引言

对数正态差分布（以下简称 DLN）是由一个对数正态随机变量（RV）减去另一个对数正态随机变量而得到的分布。为定义 DLN，考虑一个满足以下条件的随机变量 $W$

$$W = Y_p - Y_n = \exp(X_p) - \exp(X_n) \quad \text{with} \quad \boldsymbol{X} = (X_p, X_n)^T \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \tag{1}$$

其中 $\boldsymbol{X}$ 服从二元正态分布，且

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_p \\ \mu_n \end{bmatrix} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_p^2 & \sigma_p \cdot \sigma_n \cdot \rho_{pn} \\ \sigma_p \cdot \sigma_n \cdot \rho_{pn} & \sigma_n^2 \end{bmatrix} \tag{2}$$

我们称 $W$ 服从五参数对数正态差分布，记作 $W \sim \text{DLN}(\mu_p,\ \sigma_p,\ \mu_n,\ \sigma_n,\ \rho_{pn})$。

配套论文 Parham (2023年) 论证了，双对数正态分布是自然界中的一种基本分布，因为它会自然地出现在众多迥异的自然现象中，类似于正态分布和对数正态分布。该文表明，企业收入、收益和增长都能用双对数正态分布很好地描述；此外还表明，城市人口增长、县级GDP增长以及都市区分行业GDP增长都与双对数正态分布高度吻合。 Parham (2023年) 说明，将中心极限定理和“吉布拉特定律”应用于各种经济现象，会直接导致双对数正态分布的出现。由于对双对数正态分布的研究几乎还是一片空白，[1] 本文旨在填补这一空白。

下一节将全面刻画DLN分布，推导其概率密度函数、累积分布函数、中心矩，以及根据数据估计分布参数的估计量。此外，本节还运用椭圆分布理论，将DLN扩展到多元N维情形。文中还提供了一整套计算机代码，供日后使用。

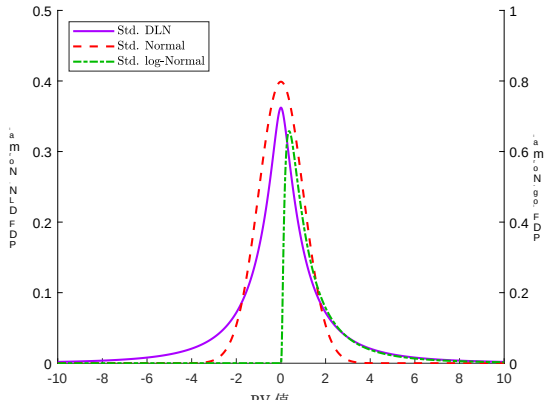
[1]据我所知，目前只有两篇统计学研究探讨了这一问题：Lo (2012) 以及Gulisashvili 和 Tankov (2016)。这两篇论文主要研究对数正态变量之和，但也表明，在某些条件下，其结论同样适用于对数正态变量之差。

接下来，第 3 节将讨论直接处理DLN分布的困难，这源于其特有的“双指数”厚尾。为缓解这一困难，我将探讨DLN与双曲正弦（$sinh$）函数及其反函数（$asinh$）之间的密切联系，并介绍ADLN分布——经过asinh变换的DLN。随后，本节将讨论如何衡量服从DLN分布的随机变量的增长问题。为此，我将增长概念从目前仅适用于严格正随机变量的定义推广到有时为负的DLN随机变量。我将说明，对于某个随机变量，恰当的增长概念（例如百分比、对数差或DLN增长）与该随机变量所服从的统计分布密切相关。

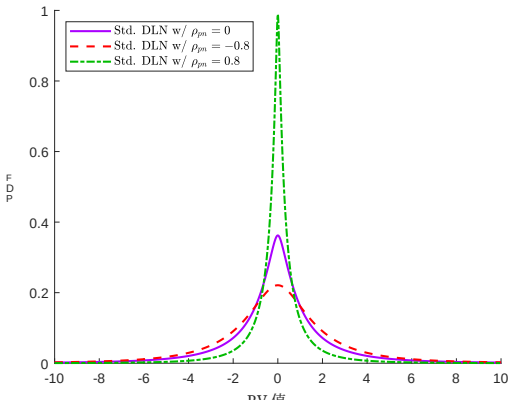
第4 节通过大量蒙特卡洛实验考察所介绍估计量的性质。具体而言：(i) 报告矩估计量和极大似然参数估计量的经验偏差与方差；(ii) 确定针对 DLN 随机变量的柯尔莫哥洛夫-斯米尔诺夫和安德森–达林分布检验的临界值；(iii) 展示第3节提出的增长度量之间的关系。

## 2 定义和性质

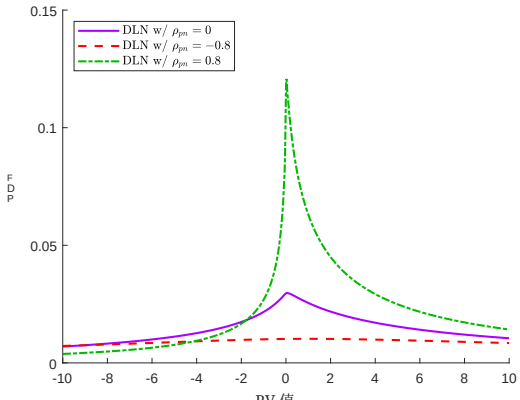
在继续讨论之前，为了明确说明问题，图1 展示了双对数正态分布的若干实例。面板（a）展示并对比了标准正态分布、标准DLN和标准对数正态分布。不相关的标准DLN定义为 DLN(0,1,0,1,0)，即两个互不相关的标准正态随机变量取指数后之差。面板（b）展示相关系数 $\rho_{pn}$ 在标准DLN中的作用，它控制尾部厚度与峰度。面板（c）对不同的参数化方式重复面板（b）的分析；这种参数化方式在实际应用中很常见，并凸显了处理厚尾分布的问题。面板（d）以反双曲正弦空间展示面板（c）的数据，说明反双曲正弦变换如何解决厚尾数据绘图的问题，以及ADLN分布为何在实践中有用。



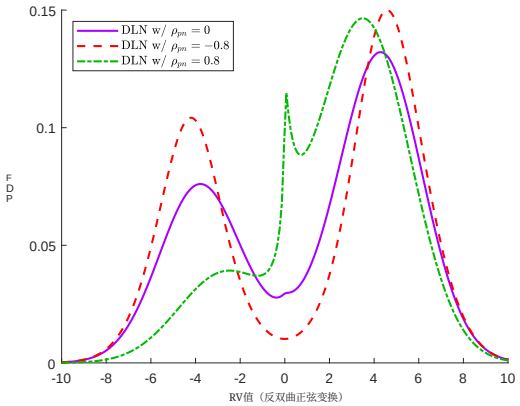

(a) 标准DLN、正态分布、对数正态分布




(b) 带相关性的标准DLN




（c）具有相关性的DLN




（d）具有相关性的ADLN

图1. DLN示例。该图展示了DLN分布的示例。面板（a）绘制了标准正态分布、对数正态分布和DLN的概率密度函数。面板（b）绘制了具有不同相关系数 $\rho_{pn}$的标准DLN的概率密度函数。面板（c）展示了参数为（3，2，2，2）的DLN的概率密度函数，该参数组合在实践中很常见，并具有不同的相关系数$\rho_{pn}$。面板（c）展示了范围 ±10内的概率密度函数；由于该DLN具有长尾，这构成了显著的截断。面板（d）展示了与面板（c）相同的概率密度函数，但x轴经过反双曲正弦变换，因此其范围从sinh(-10) ≈ -11,000到sinh(10) ≈ 11,000。

### 2.1 概率密度函数和累积分布函数

二元正态分布（BVN）随机变量 $\boldsymbol{X}$ 的概率密度函数众所周知为

$$f_{BVN}(\boldsymbol{x}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2\pi} \cdot \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2\pi} \cdot \exp\left(-\frac{1}{2}||\boldsymbol{x}-\boldsymbol{\mu}||_{\boldsymbol{\Sigma}}\right) \tag{3}$$

其中，$|\boldsymbol{\Sigma}|$ 是 $\boldsymbol{\Sigma}$ 的行列式，$||\boldsymbol{x}||_{\boldsymbol{\Sigma}}$ 是在由 $\boldsymbol{\Sigma}$诱导的马氏距离下 $\boldsymbol{x}$ 的欧几里得范数。

二元对数正态分布（BVLN）随机变量 $\boldsymbol{Y} = (Y_p, Y_n)^T$ 的概率密度函数可通过多元变量变换定理求得。若 $\boldsymbol{Y} = g(\boldsymbol{X})$，则

$$f_Y(\boldsymbol{y}) = f_X(g^{-1}(\boldsymbol{y})) \cdot ||J_{g^{-1}}(\boldsymbol{y})|| \tag{4}$$

其中，$J_{g^{-1}}$ 是 $g^{-1}(\cdot)$的雅可比矩阵，$||J_{g^{-1}}||$ 是其行列式的绝对值。将该定理应用于 $\boldsymbol{Y} = g(\boldsymbol{X}) = (\exp(X_p),\exp(X_n))^T$，可得 $g^{-1}(\boldsymbol{y}) = (log(y_p)\ log(y_n))^T$ 以及 $||J_{g^{-1}}(\boldsymbol{y})|| = (y_p \cdot y_n)^{-1}$。因此，BVLN随机变量的概率密度函数为

$$f_{BVLN}(\boldsymbol{y}) = \frac{|\boldsymbol{\Sigma}|^{-\frac{1}{2}}}{2\pi y_p y_n}\exp\left(-\frac{1}{2}||\log(\boldsymbol{y})-\boldsymbol{\mu}||_{\boldsymbol{\Sigma}}\right) \tag{5}$$

现在，我们可以利用两个随机变量之差的累积分布函数的定义，来定义 DLN 分布的累积分布函数（CDF）。

$$\begin{aligned} F_{DLN}(w) &= P[W \le w] = P[y_p - y_n \le w] = P[y_p \le y_n + w] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y_n+w} f_{BVLN}(y_p, y_n)dy_p dy_n \end{aligned} \tag{6}$$

对 $w$ 求导即可得到概率密度函数

$$f_{DLN}(w) = \int_{-\infty}^{\infty} f_{BVLN}(y+w, y)dy = \int_{-\infty}^{\infty} f_{BVLN}(y, y-w)dy \tag{7}$$

但由于 $f_{BVLN}(\boldsymbol{y})$ 仅在 $\boldsymbol{y} > 0$ 时非零，因此我们限制积分范围

$$f_{DLN}(w) = \int_{\max(0,w)}^{\infty} f_{BVLN}(y, y-w)dy \tag{8}$$

由此得到 DLN 分布的概率密度函数。

然而，众所周知，方程8中的积分没有闭式解。随附的代码套件会对其进行数值计算，并根据累积分布函数的定义对其进行数值计算

$$F_{DLN}(w) = \int_{-\infty}^{w} f_{DLN}(y)dy \tag{9}$$

对于不相关对数正态分布之差这一较简单的情况，即 $\rho_{pn} = 0$，我们也可以通过特征函数（CF）方法推导 DLN 分布的概率密度函数。在这种情况下，DLN 的特征函数可写为 $\varphi_{DLN}(t) = \varphi_{LN}(t) \cdot \varphi_{LN}(-t)$，其中 $\varphi_{LN}(t)$ 是对数正态分布的特征函数。接下来，我们可以应用傅里叶变换来得到概率密度函数，

$$f_{DLN}(w) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i \cdot t \cdot w} \cdot \varphi_{DLN}(t)dt \tag{10}$$

遗憾的是，对数正态分布不存在解析特征函数，使用方程10还需要对 $\varphi_{LN}(t)$ 进行数值近似。 Gubner (2006) 提出了一种快速且准确的对数正态分布特征函数近似方法；使用该方法计算 $f_{DLN}(w)$ 时，我采用了这种方法。

## 2.2 矩

### 2.2.1 矩母函数

双对数正态分布的矩母函数（MGF）可写为

$$M_W(t) = \mathbb{E}\left[e^{tW}\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{tw} f_{BVLN}(y+w, y)dydw \tag{11}$$

但由于这些积分不存在闭式解，这种表达式的实用性有限。相反，直接刻画矩更为有用，因为我们可以得到它们的闭式解。

### 2.2.2 均值和方差

根据 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$ 在 2中的定义，定义 BVLN 随机变量的均值和协方差，$\hat{\boldsymbol{\mu}}$ 和 $\boldsymbol{\Sigma}$（逐元素）如下

$$\begin{aligned}\hat{\boldsymbol{\mu}}_{(i)} &= \exp\left(\boldsymbol{\mu}_{(i)} + \frac{1}{2}\boldsymbol{\Sigma}_{(i,i)}\right) \\ \hat{\boldsymbol{\Sigma}}_{(i,j)} &= \exp\left(\boldsymbol{\mu}_{(i)} + \boldsymbol{\mu}_{(j)} + \frac{1}{2}\left(\boldsymbol{\Sigma}_{(i,i)} + \boldsymbol{\Sigma}_{(j,j)}\right)\right) \cdot \left(\exp\left(\boldsymbol{\Sigma}_{(i,j)}\right) - 1\right)\end{aligned} \tag{12}$$

注意，如果 $\boldsymbol{\Sigma}$ 是对角的（即，$X_p$ 和 $X_n$ 不相关），那么 $\boldsymbol{\Sigma}$ 也将是对角的。不过，我们关注的是 DLN 分布的一般形式。关于随机变量之和的期望和方差的恒等式给出

$$\mathbb{E}[W] = \mathbb{E}[Y_p] - \mathbb{E}[Y_n] = \hat{\boldsymbol{\mu}}_{(1)} - \hat{\boldsymbol{\mu}}_{(2)} = \exp(\mu_p + \frac{\sigma_p^2}{2}) - \exp(\mu_n + \frac{\sigma_n^2}{2}) \tag{13}$$

and

$$\begin{aligned}\mathrm{Var}[W] &= \mathbb{C}[Y_p, Y_p] + \mathbb{C}[Y_n, Y_n] - 2 \cdot \mathbb{C}[Y_p, Y_n] = \hat{\boldsymbol{\Sigma}}_{(1,1)} + \hat{\boldsymbol{\Sigma}}_{(2,2)} - 2 \cdot \hat{\boldsymbol{\Sigma}}_{(1,2)} \\ &= \exp\left(2\mu_p + \sigma_p^2\right) \cdot \left(exp\left(\sigma_p^2\right) - 1\right) + \exp\left(2\mu_n + \sigma_n^2\right) \cdot \left(exp\left(\sigma_n^2\right) - 1\right) \\ &- 2\exp\left(\mu_p + \mu_n + \frac{1}{2}(\sigma_p^2 + \sigma_n^2)\right) \cdot \left(\exp\left(\sigma_p \sigma_n \rho_{pn}\right) - 1\right)\end{aligned} \tag{14}$$

其中 $\mathbb{C}$ 两个一般随机变量的协方差算子 $U_1$，$U_2$

$$\mathbb{C}[U_1, U_2] = \mathbb{E}[(U_1 - \mu_1)(U_2 - \mu_2)] \tag{15}$$

### 2.2.3 偏度和峰度

双对数正态分布的偏度和峰度同样可以通过共偏度和共峰度来确定（概述可参见，例如 Miller (2013)）。三个一般随机变量的共偏度 $U_1$， $U_2$， $U_3$定义为

$$\mathbb{S}\left[U_1, U_2, U_3\right] = \frac{\mathbb{E}\left[(U_1 - \mu_1)(U_2 - \mu_2)(U_3 - \mu_3)\right]}{\sigma_1 \sigma_2 \sigma_3} \tag{16}$$

四个一般随机变量 $U_1$、 $U_2$、 $U_3$、 $U_4$ 的共峰度定义为

$$\mathbb{K}\left[U_1, U_2, U_3, U_4\right] = \frac{\mathbb{E}\left[(U_1 - \mu_1)(U_2 - \mu_2)(U_3 - \mu_3)(U_4 - \mu_4)\right]}{\sigma_1 \sigma_2 \sigma_3 \sigma_4} \tag{17}$$

其性质为 $\mathbb{S}\left[U, U, U\right]$ =偏度 $\left[U\right]$且 $\mathbb{K}\left[U, U, U, U\right]$ =峰度 $\left[U\right]$。更重要的是，很容易证明

$$\text{Skew}\left[U - V\right] = \frac{\sigma_U^3 \mathbb{S}\left[U, U, U\right] - 3\sigma_U^2 \sigma_V \mathbb{S}\left[U, U, V\right] + 3\sigma_U \sigma_V^2 \mathbb{S}\left[U, V, V\right] - \sigma_V^3 \mathbb{S}\left[V, V, V\right]}{\sigma_{U-V}^3} \tag{18}$$

类似地

$$\begin{aligned} \text{Kurt}\left[U - V\right] = \frac{1}{\sigma_{U-V}^4} [&\sigma_U^4 \mathbb{K}\left[U, U, U, U\right] - 4\sigma_U^3 \sigma_V \mathbb{K}\left[U, U, U, V\right] \\ &+ 6\sigma_U^2 \sigma_V^2 \mathbb{K}\left[U, U, V, V\right] - 4\sigma_U \sigma_V^3 \mathbb{K}\left[U, V, V, V\right] + \sigma_V^4 \mathbb{K}\left[V, V, V, V\right]] \end{aligned} \tag{19}$$

其中 $\sigma_{U-V}$ =方差 $\left[U - V\right]^{\frac{1}{2}}$根据公式 14计算。对于双对数正态分布的情形，计算算子 $\mathbb{S}$ 和 $\mathbb{K}$需要计算一般形式的表达式 $\mathbb{E}\left[Y_p^i Y_n^j\right]$，这可通过双变量正态分布的矩母函数完成

$$\mathbb{E}\left[Y_p^i Y_n^j\right] = \mathbb{E}\left[e^{iX_p} e^{jX_n}\right] = \text{MGF}_{BVN}\left(\begin{bmatrix} i \\ j \end{bmatrix}\right) = \mathbb{E}\left[Y_p^i\right] \mathbb{E}\left[Y_n^j\right] e^{ij\boldsymbol{\Sigma}_{(1,2)}} \tag{20}$$

其中 $\mathbb{E}\left[Y_p^i\right] = \exp\left(i\mu_p + \frac{1}{2} i^2 \sigma_p^2\right)$。至此，推导的技术细节介绍完毕 .

所述方法也可扩展至更高阶中心矩。配套的代码套件包含实现上述方程的函数，并利用这些方程

计算给定参数（$\mu_p$，$\sigma_p$，$\mu_n$，$\sigma_n$，$\rho_{pn}$）时双对数正态分布的前五阶矩。后文第 4 节将介绍蒙特卡洛实验的结果，这些实验用于检验矩的经验方差和偏差如何随样本量变化。

## 2.3 估计

给定数据 $\boldsymbol{D} \sim$ 双对数正态分布$(\boldsymbol{\Theta})$，其中 $\boldsymbol{\Theta} = (\mu_p, \sigma_p, \mu_n, \sigma_n, \rho_{pn})$，我们希望找到参数向量$\boldsymbol{\Theta}$ 的一个估计值 $\hat{\boldsymbol{\Theta}}$。实验表明，给定适当的初始猜测，$\hat{\boldsymbol{\Theta}}$ 的最大似然估计在实践中表现良好。最难估计的主要参数是 $\rho_{pn}$。该参数类似于稳定分布中的形状参数，作用相似，估计起来也同样困难，参见例如 Fama 和 Roll (1971)。因此，估计该参数需要特别谨慎。

所提供的估计代码使用多起点算法，针对双对数正态分布的概率密度函数，最小化数据的负对数似然。前四个参数在所有起点上的初始值固定如下：

$$\begin{bmatrix} \mu_p \\ \sigma_p \\ \mu_n \\ \sigma_n \end{bmatrix} = \begin{bmatrix} \text{Median}\left[\log\left(\boldsymbol{D}\right)\right] \quad \text{for} \quad \boldsymbol{D} > 0 \\ \text{IQR}\left[\log\left(\boldsymbol{D}\right)\right] / 1.35 \quad \text{for} \quad \boldsymbol{D} > 0 \\ \text{Median}\left[\log\left(-\boldsymbol{D}\right)\right] \quad \text{for} \quad \boldsymbol{D} < 0 \\ \text{IQR}\left[\log\left(-\boldsymbol{D}\right)\right] / 1.35 \quad \text{for} \quad \boldsymbol{D} < 0 \end{bmatrix} \tag{21}$$

$\hat{}\hat{}\hat{}$ 而 $\rho_{pn}$ 的初始猜测值为 $(-0.8,\ -0.3,\ 0,\ 0.3,\ 0.8)$。随后，估计量 $\hat{\boldsymbol{\Theta}}$ 即为多起点算法中使负对数似然最小的数值。由于 $\boldsymbol{\Theta}$ 的维数固定、似然函数光滑且其支撑集为 $\mathbb{R}\ \forall\boldsymbol{\Theta}$，该估计量继承了一般 M-$\hat{}\hat{}\hat{}$ 估计量理论中的渐近正态性、相合性和有效性等性质。或许有必要为双对数正态分布的参数设计更好的估计方法，但这留待未来研究。

## 2.4 椭圆多元双对数正态分布

双对数正态分布的实际应用要求能够处理多元双对数正态随机变量。因此，我借助椭圆分布理论，将双对数正态分布推广到多元情形；该领域的标准参考文献是 Fang、Kotz 和 Ng (1990)。

椭圆分布方法要求对称基准分布。因此，我们将重点讨论对称双对数正态分布情形，其中 $\mu_p = \mu_n \equiv \mu$且 $\sigma_p = \sigma_n \equiv \sigma$，由此得到三参数单变量对称分布 SymDLN$(\mu, \sigma, \rho)$ = 双对数正态分布$(\mu, \sigma, \mu, \sigma, \rho)$。首先，我使用 SymDLN 和 Cambanis、Huang 和 Simons 的球面分解来定义标准化 N 维椭圆双对数正态随机变量，随后将其扩展为位置尺度分布族。Cambanis、Huang 和 Simons (1981)

令 $\mathbf{U}$ 为一个在 $\mathbb{R}^N$中的单位超球面上均匀分布的 N 维随机变量，并将其写作列向量。令 $R \geq 0$ 为一个与 $\mathbf{U}$ 独立的单变量随机变量，其概率密度函数 $f_R(r)$ 稍后推导；并令 $\mathbf{Z} = R \cdot \mathbf{U}$ 为标准化 N 维椭圆双对数正态随机变量。$\mathbf{U}$的一种常见取法是 $\mathbf{U}/||\mathbf{U}||_2$，其中 $\mathbf{U} \sim MVN(\mathbf{0}_N, \mathbf{1}_N)$。$\mathbf{U}$表示 $\mathbb{R}^N$中的一个方向，并且有 $\sqrt{\mathbf{U}^T \cdot \mathbf{U}} = ||\mathbf{U}||_2 \equiv 1$，这意味着 $\sqrt{\mathbf{Z}^T \cdot \mathbf{Z}} = ||\mathbf{Z}||_2 = R$。我们还知道，半径为 $R$ 的 N 维球面的表面积为

$$S_N(R) = \frac{2 \cdot \pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)} \cdot R^{N-1} \tag{22}$$

因此可以将$\mathbf{Z}$ 的概率密度函数写为

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{f_R(||\mathbf{z}||_2)}{S_N(||\mathbf{z}||_2)} = \frac{\Gamma\left(\frac{N}{2}\right) \cdot f_R(||\mathbf{z}||_2)}{2 \cdot \pi^{\frac{N}{2}} \cdot ||\mathbf{z}||_2^{N-1}} \tag{23}$$

我们要求 $f_R(r)$ 和 $f_{\mathbf{Z}}(\mathbf{z})$ 是有效的概率密度函数，由此得到以下条件

$$\begin{aligned} & f_R(r) \geq 0 \ \forall \ r \in \mathbb{R} \\ & f_{\mathbf{Z}}(\mathbf{z}) \geq 0 \ \forall \ \mathbf{z} \in \mathbb{R}^N \\ & \int_{-\infty}^{\infty} f_R(r) \ dr = 1 \\ & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{\mathbf{Z}}(\mathbf{z}) \ d\mathbf{z}_{(N)} \cdots d\mathbf{z}_{(1)} = 1 \end{aligned} \tag{24}$$

此外，我们还可以添加条件： $f_R(r)$ 经适当归一化后的分布为对称双对数正态分布，

$$f_R(r) = \widetilde{M}_N(r) \cdot f_{DLN}(r) \tag{25}$$

其中 $\widetilde{M_N(r)}$ 的选取应使方程 24 中的条件成立。求解这组条件可得

$$f_R(r) = \frac{r^{N-1}}{\int_0^{\infty} \widetilde{r}^{N-1} \cdot f_{DLN}(\widetilde{r}) \ d\widetilde{r}} \cdot f_{DLN}(r) \tag{26}$$

and

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{\Gamma\left(\frac{N}{2}\right)}{2 \cdot \pi^{\frac{N}{2}} \cdot \int_0^{\infty} \widetilde{r}^{N-1} \cdot f_{DLN}(\widetilde{r}) \ d\widetilde{r}} \cdot f_{DLN}(||\mathbf{z}||_2) = M_N \cdot f_{DLN}(||\mathbf{z}||_2) \tag{27}$$

其中， $M_N$ 是归一化常数，仅取决于所使用的基准对称对数正态分布$(\mu, \sigma, \rho)$的维数 N 和参数。我们还可以利用$\mathbf{Z}$ 的累积分布函数的定义来写出

$$\begin{aligned} F_{\mathbf{Z}}(\mathbf{z}) &= \int_{-\infty}^{\mathbf{z}_{(1)}} \cdots \int_{-\infty}^{\mathbf{z}_{(N)}} f_{\mathbf{Z}}(\widehat{\mathbf{z}}) \ d\widehat{\mathbf{z}}_{(N)} \cdots d\widehat{\mathbf{z}}_{(1)} \\ &= \int_{-\infty}^{\mathbf{z}_{(1)}} \cdots \int_{-\infty}^{\mathbf{z}_{(N)}} M_N \cdot f_{\mathbf{DLN}}(||\mathbf{z}||_2) \ d\widehat{\mathbf{z}}_{(N)} \cdots d\widehat{\mathbf{z}}_{(1)} \end{aligned} \tag{28}$$

至此，标准化 N 维椭圆对数正态随机变量的刻画便告完成。

将标准化的 N 维双对数正态分布扩展为位置尺度分布族现在非常简单。设 $\widetilde{\boldsymbol{\mu}} = (\mu_1, \mu_2, \ldots\ldots \mu_N)^T$ 为位置列向量，并设 $\widetilde{\boldsymbol{\Sigma}}$

为秩为 $N$ 的半正定缩放矩阵。定义

$$\mathbf{W} = \widetilde{\boldsymbol{\mu}} + \widetilde{\boldsymbol{\Sigma}}^{\frac{1}{2}} \cdot \mathbf{Z} \tag{29}$$

$\widetilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}$ 其中 $\widetilde{\boldsymbol{\Sigma}}^{\frac{1}{2}}$ 表示 $\widetilde{\boldsymbol{\Sigma}}$的特征值分解。$\mathbf{W}$ 的概率密度函数为

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{w}) &= |\widetilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \cdot f_{\mathbf{Z}}\left(\widetilde{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \cdot (\mathbf{w} - \widetilde{\boldsymbol{\mu}})\right) \\ &= |\widetilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \cdot M_N \cdot f_{DLN}\left(\sqrt{(\mathbf{w} - \widetilde{\boldsymbol{\mu}})^T \cdot \widetilde{\boldsymbol{\Sigma}}^{-1} \cdot (\mathbf{w} - \widetilde{\boldsymbol{\mu}})}\right) \\ &= |\widetilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \cdot M_N \cdot f_{DLN}\left(||\mathbf{w} - \widetilde{\boldsymbol{\mu}}||_{\widetilde{\boldsymbol{\Sigma}}}\right) \end{aligned} \tag{30}$$

$\mathbf{W}$ 的累积分布函数同样可写为

$$F_{\mathbf{W}}(\mathbf{w}) = |\widetilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} \cdot M_N \cdot \int_{-\infty}^{\mathbf{w}_{(1)}} \cdots \int_{-\infty}^{\mathbf{w}_{(N)}} f_{\mathbf{DLN}}\left(||\mathbf{w} - \widetilde{\boldsymbol{\mu}}||_{\widetilde{\boldsymbol{\Sigma}}}\right) \, d\widehat{\mathbf{w}}_{(N)} \cdots d\widehat{\mathbf{w}}_{(1)} \tag{31}$$

这刻画了一般的椭圆型多元双对数正态随机变量。

最后请注意，缩放矩阵 $\boldsymbol{\Sigma}$ 并非 $\mathbf{W}$ 的协方差矩阵，这是由于 $\mathbf{W}$具有重尾特性，与其他重尾椭圆分布（如多元稳定分布、t 分布或拉普拉斯分布）类似。还需注意，当 N 取值较大时（例如 $N \geq 5$），方程 26 中的归一化积分在数值上不稳定，因此在推导高维双对数正态随机变量的概率密度函数时应格外谨慎。

# 3 重尾分析方法

如上文所述，使用DLN分布的一个主要难点源于其“双指数”特性，即它在正向和负向都表现出指数尾部。对于单侧指数尾部，通常采用对数变换来缓解这一问题，但由于对数在负数上无定义，这种方法在此失效。本节将介绍如何将应用于单侧指数尾部的方法扩展到双指数分布。

## 3.1 反双曲正弦空间与 ADLN

使用对数变换的一个常见替代方法是使用反双曲正弦（asinh）对数据进行变换。有关asinh在经济应用中的使用综述，请参阅Bellemare 和 Wichman (2020)。双曲正弦及其反函数定义如下

$$\begin{aligned} \sinh(x) &= \frac{e^x - e^{-x}}{2} \\ \operatorname{asinh}(x) &= \log\left(x + \sqrt{1 + x^2}\right) \end{aligned} \tag{32}$$

asinh 变换具有以下有用性质：

1. 关于 x 可微且严格递增。

2. asinh($x$) $\approx$ sign($x$)(log$|x|$ + log(2))，且随着$|x|$ 增大，近似误差迅速趋于零。[2]

3. 是奇函数，因此 asinh($-x$) $=$ $-$asinh($x$)。

4. 以零为基准，因此 asinh(0) $= 0$

也就是说，反双曲正弦是一种风格上类似于负对数变换的双射：

$$\operatorname{neglog}(x) = \operatorname{sign}(x)\log(1 + |x|) \tag{33}$$

但在 0 附近的失真比负对数变换更小 , 代价是固定偏差 log(2) $\approx 0$ 7 。。值得注意的是，任何指数差函数都可以分解为一个指数函数与一个双曲正弦的乘积，即，

$$y = \exp(x_1) - \exp(x_2) = 2 \cdot \exp\left(\frac{x_1 + x_2}{2}\right) \cdot \sinh\left(\frac{x_1 - x_2}{2}\right) \tag{34}$$

这凸显了 sinh 函数与双对数正态分布和拉普拉斯分布之间亲密而直观的关系。三者都以指数函数之差的形式表示，从而导致[2]在 $|x|$=4 处的近似误差约为 1%，在 $|x|$=10 处约为 0.1%。

其特有的“双指数”性质。Sinh 的逆函数，即 asinh，因此是应用于服从双对数正态分布和拉普拉斯分布的随机变量的自然变换。

由于反双曲正弦函数可微且严格递增，因此适用变量变换法。若 $Z = \text{asinh}(W)$，其中 $W \sim DLN$，则 $Z \sim ADLN$，$W = \sinh(Z)$，且 $\frac{dZ}{dW} = (1 + \sinh(Z)^2)^{-1/2}$。现在可以写出 ADLN 分布的概率密度函数

$$f_{ADLN}(z) = \frac{f_{DLN}(\sinh(z))}{\text{asinh}'(\sinh(z))} = f_{DLN}(\sinh(z))\sqrt{1 + \sinh(z)^2} \tag{35}$$

这使得我们能够分析 $Z \sim ADLN$，即经过变换的 DLN 随机变量，其直方图更加“紧凑”，也更易于呈现。

图 1 中的面板（c）和（d）展示了实践中常见的典型 DLN 分布，横轴分别采用线性刻度（面板（c））和反双曲正弦刻度（面板（d））。面板（c）展示了该分布的一个截断片段。得益于 asinh 变换，面板（d）能够呈现整个分布。面板（c）中无法看出 DLN 正值部分和负值部分近似具有对数正态性，而面板（d）中的 asinh 变换则清楚地展现了这一点。

## 3.2 双对数正态分布变量的增长

如何衡量双对数正态分布随机变量的增长？某企业第 1 年收入为 $\$100M$，第 2 年收入为 $\$120M$，其收入显然有所增长。至于说该企业的收入增长了 $\frac{120M}{100M} - 1 = 0.2 = 20\%$，还是增长了 $\log(120M) - \log(100M) = 0.182$ 对数点，或许值得商榷，但这个问题本身是明确的。然而，如果该企业第 1 年的收入为 $-\$100M$（即亏损），而第 2 年的收入为 $\$120M$，那么它的增长是多少？本节旨在对此问题给出严谨的回答。

首先，我们需要给增长下一个定义。巴罗和萨拉-伊-马丁(2003)以及Studen´yand Mezn´ik (2013)将时间连续且严格为正的随机变量$Z(t) > 0$ 的瞬时增长定义为

$$\frac{dZ(t)/dt}{Z(t)} = \frac{Z'(t)}{Z(t)} \approx \frac{Z_{t+1} - Z_t}{Z_t} \tag{36}$$

方程的第二部分将离散变量的一阶差分用作导数 $Z'(t)$ 的近似，由此得到离散变量百分比增长的著名表达式。将此定义推广到 $Z(t) \in \mathbb{R}$ ，可得：

$$d\% \equiv \frac{dZ(t)/dt}{|Z(t)|} = \frac{Z'(t)}{|Z(t)|} \approx \frac{Z_{t+1} - Z_t}{|Z_t|} \quad \text{for} \quad Z(t) \neq 0 \tag{37}$$

这保证了 $Z_{t+1} > Z_t$ 会带来正增长，而无论 $Z_t$ 的符号如何。近似项（$Z_{t+1} - Z_t$） $/|Z_t|$ 是广义百分比增长 （以下记作 d%）；当 $|Z_t| \to 0$ 时，它会呈爆炸式增长，这与“传统”的百分比增长类似。

接下来，考察对数正态分布随机变量的增长很有启发性，因为企业动态（以及其他领域）中遇到的大多数规模度量都近似服从对数正态分布。为此，考虑以下设定：

$$\begin{aligned} X_{t+1} &= (1 - \rho_X) \cdot \mu_X + \rho_X \cdot X_t + \epsilon_t^X \\ \epsilon_t^X &\sim \mathcal{N}(0, \sigma_X^2) \\ Y_t &= \exp(X_t) \end{aligned} \tag{38}$$

其中， $X_t$ 是一个简单的 $AR(1)$随机过程，因此服从正态分布； $Y_t > 0$服从对数正态分布。 $Y_t$的增长是多少?

根据定义，我们有：

$$\frac{Y'(t)}{|Y(t)|} = \frac{Y(t) \cdot X'(t)}{Y(t)} = X'(t) \approx X_{t+1} - X_t = \log(Y_{t+1}) - \log(Y_t) \equiv \text{dlog}(Y_{t+1}) \tag{39}$$

由此得到广为人知的增长表示形式：相邻数值之间的对数差，记作 dlog()。方程 37 与 39 的区别在于，是否在应用一阶差分近似之前先进行微分。请注意，若此处使用百分比

增长（如方程 37 所示），则会得到：

$$\frac{Y_{t+1}}{Y_t} - 1 = \exp(X_{t+1} - X_t) - 1 \tag{40}$$

或者说，百分比增长是对数增长的凸变换这一一般性观察。还值得进一步指出，$\lim_{\rho_X \to 1}(X_{t+1} - X_t) = \epsilon_t^X$。对数增长给出了底层 AR(1) 过程中的创新项，而百分比增长则给出了变换后的数值 $\exp(\epsilon_t^X) - 1$。也就是说，对于对数正态分布的随机变量，百分比增长相较于对数增长会引入凸性偏差。

反过来，即使正态分布随机变量在实际中严格为正，使用对数增长来衡量其增长，也会引入类似但方向相反的“凹性偏差”。为说明这一点，考虑以对数微分项衡量 $X(t) > 0$ 的增长：

$$\mathrm{dlog}(X_{t+1}) = \log(X_{t+1}) - \log(X_t) = \log\left(\frac{X_{t+1}}{X_t} - 1 + 1\right) = \log\left(\frac{X'(t)}{|X(t)|} + 1\right) \tag{41}$$

换言之，使用 dlog() 衡量 $X$ 的增长，得到的是百分比增长的对数；根据方程 36 和 37中的定义，这才是恰当的衡量方式。因此，所采用的增长概念与所考虑的分布密切相关。

接下来，考虑一个类似的情形，但针对双对数正态随机变量：

$$\begin{aligned}
X_{t+1}^p &= (1 - \rho_p) \cdot \mu_p + \rho_p \cdot X_t^p + \epsilon_t^p \\
X_{t+1}^n &= (1 - \rho_n) \cdot \mu_n + \rho_n \cdot X_t^n + \epsilon_t^n \\
(\epsilon_t^p, \epsilon_t^n)^T &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \\
Y_t^p &= \exp(X_t^p) \quad ; \quad Y_t^n = \exp(X_t^n) \\
W_t &= Y_t^p - Y_t^n
\end{aligned} \tag{42}$$

其中 $\mathbf{\Sigma}$ 与方程2中相同。应用广义增长定义37，可得：

$$\begin{aligned}\frac{W'(t)}{|W(t)|} &= \frac{Y^p(t)\cdot dX^p(t)/dt - Y^n(t)\cdot dX^n(t)/dt}{|W(t)|} \approx \frac{Y_t^p\cdot\left(X_{t+1}^p - X_t^p\right) - Y_t^n\cdot\left(X_{t+1}^n - X_t^n\right)}{|W(t)|} \\ &= \frac{Y_t^p\cdot \mathrm{dlog}\left(Y_{t+1}^p\right) - Y_t^n\cdot \mathrm{dlog}\left(Y_{t+1}^n\right)}{|Y_t^p - Y_t^n|}\end{aligned} \tag{43}$$

这意味着，双对数正态随机变量的增长可以定义为其两个组成部分对数正态随机变量的水平和增长率的函数。第4 节通过蒙特卡洛实验，探究上述针对正态分布、对数正态分布和双对数正态分布随机变量所提出的增长度量之间的关系。

# 4 蒙特卡洛实验

本节报告旨在确定上述矩、估计量和度量性质的蒙特卡洛实验结果。

## 4.1 估计量的性质

我首先探究第2.2 节和2.3节中的矩和参数估计量。我将实验集中于实际应用中与企业理论相关的参数空间区域：

$$\boldsymbol{Q}: \quad (\mu_p, \sigma_p, \mu_n, \sigma_n, \rho_{pn}) \in ([-3,3], [0.5,2.5], [-3,3], [0.5,2.5], [-1,1]) \tag{44}$$

蒙特卡洛分析的数据收集/创建过程如下。对于每个 $i \in \{1...N\}$：

1. 以均匀概率抽取参数向量 $\mathbf{\Theta}_i \in \boldsymbol{Q}$ 。

2. 使用 $\mathbf{\Theta}_i$ ，通过第 2.2 节的方法计算理论中心矩。

3. 绘制 $K$ 个观测值 $W_{i,\ k}$ ~双对数正态分布($\Theta_i$)。

4. 计算 $W_{i,k}$ 的前五个经验中心矩。

5. 使用逐渐缩小的子集，重新计算 $K$观测值.[3] 的前五个经验矩。

$\widehat{\ }$6. 使用第 $W_{i\text{节}\ k}$所述方法，估计 $\Theta_i$（记为2.3）的参数。

7. 根据 $\Theta_i$ 和 $W_{i,k}$，计算柯尔莫哥洛夫-斯米尔诺夫（K-S）、卡方（C-2）和安德森–达林$\widehat{\ }$（A-D）检验统计量。

我将数据生成过程重复 $N = 70{,}000$ 次。在每次循环中，抽取 $K = 100{,}000$ 个观测值 $W_{i,\ k}$ ~DLN($\Theta_i$)。

表 1 的面板 (a) 展示了第 2.2 节矩估计量的蒙特卡洛结果。该面板比较了蒙特卡洛实验第 2 步推导出的理论矩与第 4 步得出的经验矩，重点关注分布的前五阶矩。由于双对数正态分布的厚尾会导致其矩迅速发散，因此分析在反双曲正弦空间中进行（这与对数正态分布的矩类似，后者也通常在对数空间中考察）。经验矩与理论矩高度相关，且奇数阶矩（均值或 $1^{st}$ 阶矩、偏度或 $3^{rd}$ 阶矩，以及 $5^{th}$阶矩）未表现出显著偏差。偶数阶矩（方差或 $2^{nd}$，以及峰度或$4^{th}$）则呈现偏差，其中峰度的偏差相当严重。对峰度估计量进行小样本偏差校正似乎是必要的，但这超出了本文的讨论范围。理论矩与经验矩之差的四分位距随矩的阶数增加而增大，这符合预期。

表（b）1继续展示第2.3节参数估计量的蒙特卡洛结果。它将步骤 1 中抽取的实际参数与步骤 6 中计算出的估计参数进行比较。结果表明，该估计程序的表现

[3]具体而言，我根据前 $K/2^s$ 个观测值重新计算矩，适用于 $s \in \{1...11\}$。

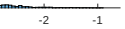
表1
估计量蒙特卡罗实验

本表展示了估计量蒙特卡洛实验的结果，每次重复均抽取 $N = 70$，000 次重复和 $K = 100$，000 个观测值。面板（a）检验第 2.2 节的矩估计量 $\widehat{M}_i \quad i \in \{1...5\}$ 与实际矩 $M_i$，所有分析均在反双曲正弦空间中进行。表中报告了总体准确度 corr(asinh($M_i$), asinh($\widehat{M}_i$))；偏差 median(asinh($\widehat{M}_i$) $-$ asinh($M_i$))；以及准确度 IQR(asinh($\widehat{M}_i$) $-$asinh($M_i$))。面板（b）报告了类似的统计量，用于比较第 2.3 $\boldsymbol{\Theta}$ 节的双对数正态分布参数估计量与实际参数 $\boldsymbol{\Theta}$。面板（c）报告了 Kolmogorov-Smirnov、卡方和安德森-达林检验统计量针对双对数正态随机变量的逆累积分布函数近似式 $ICDF(p) = a \cdot \exp(b \cdot p) + c \cdot \exp(d \cdot p)$ 中参数 a、b、c、d 的取值，以及近似式$R^2$。

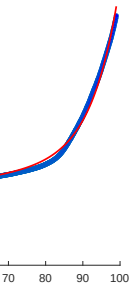
| 面板 *(a)*：矩估计量 | $\widehat{M}_1$ | $\widehat{M}_2$ | $\widehat{M}_3$ | $\widehat{M}_4$ | $\widehat{M}_5$ |
|---|---|---|---|---|---|
| 相关性 | 0.9997 | 0.9929 | 0.9282 | 0.8238 | 0.8478 |
| Bias | -0.0001 | 0.1092 | -0.0002 | 6.3410 | 0.0220 |
| 准确度 | 0.0217 | 0.4785 | 3.4480 | 8.5609 | 32.0236 |
| **面板 *(b)*：参数估计量** | $\widehat{\mu}_p$ | $\widehat{\sigma}_p$ | $\widehat{\mu}_n$ | $\widehat{\sigma}_n$ | $\widehat{\rho}_{pn}$ |
| 相关性 | 0.9408 | 0.9619 | 0.9412 | 0.9623 | 0.9190 |
| Bias | -0.0034 | 0.0019 | -0.0043 | 0.0019 | -0.0048 |
| 准确度 | 0.0588 | 0.0251 | 0.0614 | 0.0259 | 0.0762 |
| **面板（*c*）：逆累积分布函数近似** | a | b | c | d | $R^2$ |
| 柯尔莫哥洛夫-斯米尔诺夫 | 6.75e-7 | 0.1553 | -6.7520 | -0.0011 | 0.9976 |
| 卡方 | 1.88e-8 | 0.1955 | 1.2080 | 0.0044 | 0.9920 |
| 安德森–达林 | 1.18e-5 | 0.1350 | -5.7070 | -0.0060 | 0.9900 |

表现相当不错。实际参数与估计参数之间具有很高的相关性，其中包括难以估计的相关参数。参数估计值也没有表现出系统性偏差，估计误差四分位距相当低。这些结果表明，该估计过程虽然繁琐，但能够正确捕捉双对数正态分布参数。

为进一步探究矩估计量的精确度和小样本偏差，图2展示了估计量质量随样本量的变化关系。图中的面板 (a) 展示了理论矩与经验矩之间的相关性随样本量的变化关系。峰度的精确度甚至低于 $5^{th}$ 矩，并且受样本量的强烈影响。图2 的面板 (b) 随后展示了偏差随样本量的变化关系。$1^{st}$ 和 $3^{rd}$ 矩估计量没有表现出小样本偏差。 $2^{nd}$ 和 $5^{th}$ 表现出较小且快速下降的偏差。峰度同样显示出较高的偏差，并且仅随样本量增加而缓慢下降。




statistic




statistic

periments. Panel (a)
Panel (b) graphs the
ic in the simulations.
es.

## 4.2 检验统计量临界值

蒙特卡洛实验的第二个目标，是为检验某些给定数据是否来自DLN分布这一假设的检验统计量确定临界值。这对于安德森–达林检验统计量尤为重要，因为众所周知，其临界值在很大程度上取决于所检验的分布。例如参见 Stephens (1979年),D'Agostino 和 Stephens (1986) 第4章，以及 J¨antschi 和 Bolboac˘a (2018)。

为此，我按照上文步骤8所述，计算样本中每个 $N$ 抽样值的柯尔莫戈洛夫–斯米尔诺夫统计量、C-2统计量和A-D检验统计量。为便于说明，图 2 的面板（c）展示了蒙特卡洛实验中（取对数后的）柯尔莫戈洛夫–斯米尔诺夫统计量的分布。接着，我计算所得分布中各检验统计量（取对数后）的逆累积分布函数（ICDF）。图 2 的面板（d）、（e）和（f）分别展示了（取对数后的）柯尔莫戈洛夫–斯米尔诺夫统计量、C-2统计量和A-D检验统计量的逆累积分布函数。例如，面板（f）表明，若A-D统计量高于exp($ICDF(95)$) = $\exp(1.135) = 3.110$，则应在5%的置信水平下拒绝给定数据来自DLN分布这一假设。

为了从计算临界值转向推导 p 值与检验统计量数值之间的连续映射，上文讨论的文献通常会提出一种能够很好地近似逆累积分布函数的临时函数形式。使用非线性最小二乘法估计该近似函数形式后，便可用它求出与每个检验统计量数值对应的 p 值，反之亦然。经过试验，最能复现所得逆累积分布函数的函数形式如下：

$$ICDF(p) = a \cdot \exp(b \cdot p) + c \cdot \exp(d \cdot p) \tag{45}$$

这是一个由四个参数构成的指数函数之和（若 $c < 0$，则为差）。

图 2 的面板（d）、（e）和（f） 展示了函数形式的拟合值，表明该函数形式与经验逆累积分布函数之间拟合极佳。表1 的面板（c）列出了每个

（对数）检验统计量逆累积分布函数的近似参数，并进一步报告了拟合的 $R^2$ ，三个统计量的该值均高于 0.99。因此，可以放心地使用这些函数来推导分布假设检验的p值。

## 4.3 增长度量

第二组蒙特卡洛实验检验第 3.2 节所述增长度量之间的关系，涉及服从正态分布、对数正态分布和双对数正态分布的随机变量。为此，我定义了三个随机过程，其平稳分布分别服从正态分布、对数正态分布和双对数正态分布。对于每种随机变量类型，在每次蒙特卡洛迭代中，我都会抽取该分布的随机参数，对其进行向前模拟，使用上述不同度量测量每期增长，并考察随机创新项 $\epsilon_t$ 与各种增长度量之间的关系。

服从正态分布、对数正态分布和双对数正态分布的 $X$、 $Y$ 和 $W$ 的随机过程分别如上文方程 38 和 42 所述。参数区域如下：

$$\begin{aligned}&\boldsymbol{Q}_N\text{: }(\rho_N,\mu_N,sd_N)\in([0.60,0.99],[-100,100],[10,100])\boldsymbol{Q}_{LN}\text{: }(\rho_{LN},\mu_{LN},sd_{LN})\in\\&([0.60,0.99],[-3,3],[0.5,2.5])\boldsymbol{Q}_{DLN}\text{: }(\rho^p_{,\overset{n}{DLN}},\mu^p_{,\overset{n}{DLN}},sd^p_{,\overset{n}{LN}},\rho^{pn}_{DLN})\in([0.60,0.99],\\&[-3,3],[0.5,2.5],[-1,1]),\text{ 其中 }\sigma_\square=\sqrt{sd^2_\square}\cdot(1-\rho^2_\square)\text{ 对于 }\square\in\{N,LN,DLN\}。第\\&二次蒙特卡洛分析的数据收集/创建按以下步骤进行：对于每种随机变量类型\\&\square\in\{N,LN,DLN\}\text{：对于每个 }i\in\{1...N\}\text{：}\end{aligned}\tag{46}$$

1. 以均匀概率抽取一个参数向量 $\boldsymbol{\Theta}_i\in\boldsymbol{Q}_\square$ 。

2. 将 RV $Z_{\square\text{初始化为}0}$ （N）、$\mu_\square$的指数（LN），以及 DLN 中 $\exp(\mu^{p,}_{\square\overset{n}{0},})$ 处的 $Z^{p,}_{\square\overset{n}{0},}$。3. 抽取一个长度为 $K+100$ 的冲击向量（DLN 则抽取两个相关的冲击向量）。

4. 根据其运动规律向前模拟 $K+100$ 期。

5. 删除最初的 100 个观测值作为预热期。

6. 计算第 3.2 节中的增长度量。

我重复数据生成过程 $N=10$ 次，每次运行 $K=1$ 期，从而为每种分布类型生成总计 $10M$ 个增长观测值供分析。

表 2 的面板 (a)、(b)、(c) 分别展示了正态分布、对数正态分布和双对数正态分布随机变量的不同增长度量之间的相关性。各面板还报告了仅针对严格正值的相关性（即，当 $Z_t>0$ 且 $Z_{t+1}>0$ 时），以及进一步排除极小初始值后的相关性（即，$Z_t>1$）。正态分布随机变量的恰当增长概念是 $\epsilon_t/|Z_t|$，面板 (a) 显示它与广义百分比增长度量高度相关。该面板还表明，将对数差分用作正态随机变量的增长度量并不准确。图 3 的面板 (a) 和 (b) 进一步突显了这一点，分别展示了恰当增长度量与广义百分比增长（d%）度量和对数差分度量之间的关系。面板 (a) 显示，d% 能很好地反映正态随机变量的增长；面板 (b) 则突显了使用对数差分度量而非 d% 度量所产生的“凹性偏差”。面板 (a) 中围绕 45度线的离散，是由一阶自回归模型的均值回归项造成的，而增长概念并未考虑该项。

表格的面板 (b)2 转而考察对数正态随机变量。此处，适当的增长概念就是 $\epsilon_t$，面板显示 dlog 能很好地度量增长，而 d% 则受到凸性偏差的影响，是一种较差的增长度量。图 3 的面板 (c) 和 (d) 分别绘出增长与 dlog、增长与 d% 之间的关系，从而清楚地展示了凸性偏差。

最后，表2 的面板（c）展示了双对数正态随机变量的增长与各增长度量之间的相关性。对于双对数正态分布，适当的增长概念是 $(Z_t^p\cdot\epsilon_t^p-Z_t^n\cdot\epsilon_t^n)\ /|Z_t^p-Z_t^n|$，而该面板表明，公式43 中推导出的双对数正态分布增长公式能够很好地刻画这一概念。该面板还表明，dlog 对双对数正态分布的增长测量用途有限

表2 增长蒙特卡洛实验

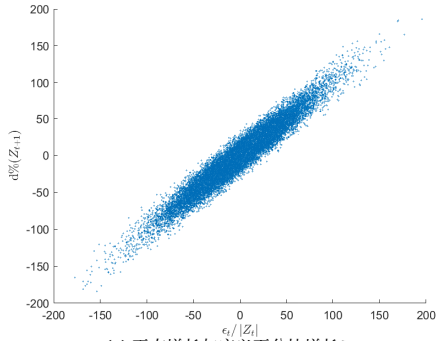
本表展示了增长蒙特卡洛实验的结果，每次实验向前模拟 $N = 10{,}000$ 次重复和 $K = 1{,}000$ 个观测值。面板（a）、（b）和（c）分别展示正态分布、对数正态分布和双对数正态分布的结果。在每个面板中，我报告了以下增长度量之间的相关性： $\epsilon_t$ 时刻 $t$ 的增长所依据的随机创新； $\epsilon_t/|Z_t|$ 相对随机创新；d%（$Z_{t+1}$）=（$Z_{t+1} - Z_t/|Z_t|$ 广义百分比增长；dlog（$Z_{t+1}$）=log（$Z_{t+1}$）-log（$Z_t$）对数点增长；dDLN（$Z_{t+1}$）基于公式43的双对数正态分布增长公式。

| 面板 *(a)*：正态分布 | $\epsilon_t$ | $\epsilon_t/\lvert Z_t\rvert$ | 广义百分比增长($Z_{t+1}$) | 对数差分($Z_{t+1}$) |
|---|---|---|---|---|
| $\epsilon_t$ | 1.000 | 0.010 | 0.009 | $0.659^a$ |
| $\epsilon_t/\lvert Z_t\rvert$ | $0.380^b$ | 1.000 | 0.973 | $0.031^a$ |
| 广义百分比增长($Z_{t+1}$) | $0.357^b$ | $0.960^b$ | 1.000 | $0.033^a$ |
| 对数差分($Z_{t+1}$) | $0.712^b$ | $0.590^b$ | $0.617^b$ | 1.000 |

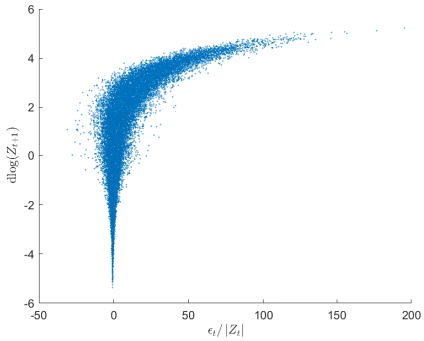
| 面板 *(b)*：对数正态分布 | $\epsilon_t$ | $\epsilon_t/\lvert Z_t\rvert$ | 广义百分比增长($Z_{t+1}$) | 对数差分($Z_{t+1}$) |
|---|---|---|---|---|
| $\epsilon_t$ | 1.000 | $0.023^a$ | $0.269^a$ | $0.931^a$ |
| $\epsilon_t/\lvert Z_t\rvert$ | $0.644^b$ | 1.000 | $0.097^a$ | $0.023^a$ |
| 广义百分比增长($Z_{t+1}$) | $0.381^b$ | $0.363^b$ | 1.000 | $0.295^a$ |
| 对数差分($Z_{t+1}$) | $0.929^b$ | $0.620^b$ | $0.381^b$ | 1.000 |

| 面板（*c*）：双对数正态分布 | $\widehat{\epsilon}_t^c$ | $\widehat{\epsilon}_t/\lvert Z_t\rvert^c$ | 广义百分比增长($Z_{t+1}$) | 对数差分($Z_{t+1}$) | 双对数正态分布($Z_{t+1}$) |
|---|---|---|---|---|---|
| $\widehat{\epsilon}_t^c$ | 1.000 | $0.000^a$ | $0.000^a$ | $0.038^a$ | $0.000^a$ |
| $\widehat{\epsilon}_t/\lvert Z_t\rvert^c$ | $0.043^b$ | 1.000 | 0.652 | $0.022^a$ | 0.944 |
| 广义百分比增长($Z_{t+1}$) | $0.009^b$ | $0.464^b$ | 1.000 | $0.016^a$ | 0.645 |
| 对数差分($Z_{t+1}$) | $0.057^b$ | $0.739^b$ | $0.397^b$ | 1.000 | $0.023^a$ |
| d双对数正态分布($Z_{t+1}$) | $0.040^b$ | $0.931^b$ | $0.455^b$ | $0.797^b$ | 1.000 |

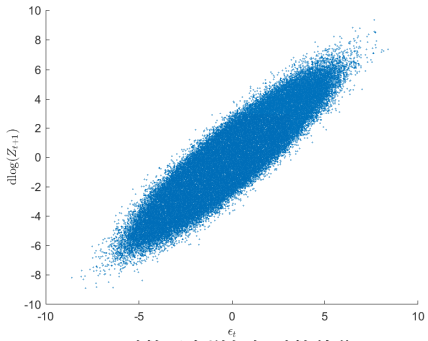
[a] 对于严格正值（$Z_t > 0$ 和 $Z_{t+1} > 0$）[b] 对于严格为正且非极小的初始值（$Z_t > 1$ 和 $Z_{t+1} > 0$）[c] 对于双对数正态分布，我定义 $\widehat{\epsilon}_t = (Z_t^p \cdot \epsilon_t^p - Z_t^n \cdot \epsilon_t^n)$ 和 $Z_t = Z_t^p - Z_t^n$



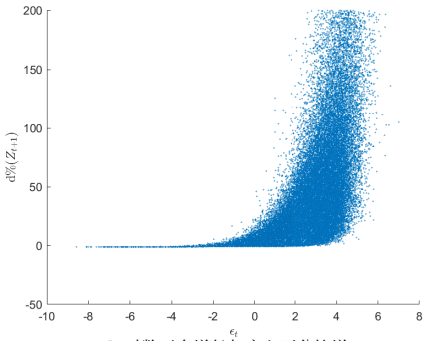

(a) 正态增长与广义百分比增长[a]



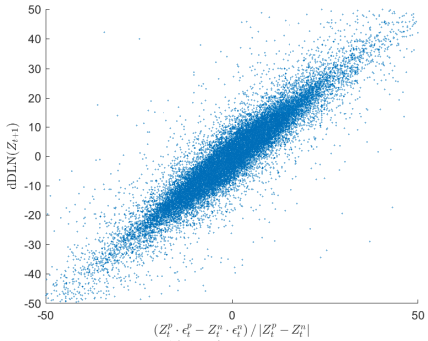

(b) 正态增长与对数差分[b]



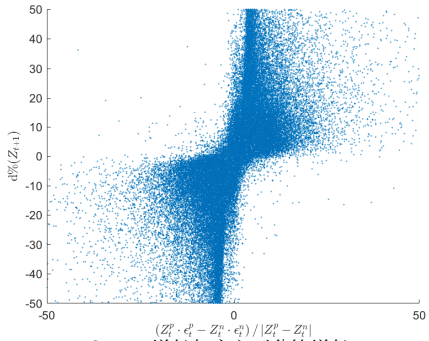

(c) 对数正态增长与对数差分




(d) 对数正态增长与广义百分比增长




(e) DLN增长与 dDLN[a]




(f) DLN增长与广义百分比增长[a]

[a] 对于非微小初始值（$|Z_t| > 1$）[b] 对于严格为正且非微小的初始值（$Z_t > 1$ 和 $Z_{t+1} > 0$）

图 3. 增长蒙特卡洛实验。本图展示了增长蒙特卡洛实验的结果。面板 (a) 和 (b) 描绘了正态随机变量的增长与以下指标之间的关系：(a) 广义百分比增长 d%($Z_{t+1}$)=$(Z_{t+1} - Z_t) / |Z_t|$; (b) 对数点增长 dlog($Z_{t+1}$)=log($Z_{t+1}$)-log($Z_t$)。面板 (c) 和 (d) 描绘了对数正态随机变量的增长与以下指标之间的关系：(c) dlog($Z_{t+1}$)；(d) d%($Z_{t+1}$)。面板 (e) 和 (f) 描绘了双对数正态随机变量的增长与以下指标之间的关系：(e) 方程 43 中的双对数正态增长度量 dDLN($Z_{t+1}$)；(f) d%($Z_{t+1}$)。

增长仅适用于正值；即使限定为正值，其表现也不佳；即使限定为正且非微小的值，与双对数正态分布增长的相关性也仅达到 0.75。图 3 的面板 (e) 和 (f) 表明，dDLN 确实是一种恰当的度量，而 d% 是双对数正态分布增长的无偏但噪声较大的度量。

# 5 总结

本文介绍了源自乘法中心极限定理的对数正态差分布（DLN），并为分析服从 DLN 分布的现象奠定了方法论和定量基础。本文首先刻画该分布，定义其概率密度函数和累积分布函数，介绍其矩和参数的估计量，并将其推广至椭圆多元随机变量。

接下来，本文讨论了分析 DLN 分布时有用的数学方法。首先，本文展示了 DLN 分布与双曲正弦之间密切而直观的联系，并说明了为何在处理 DLN 等“双指数”随机变量时，反双曲正弦（asinh）是一种有用的变换。

接下来，本文探讨了 DLN 随机变量的增长概念。本文将仅适用于正随机变量的增长经典定义扩展至随机变量 $\in \mathbb{R}$。随后，本文指出，所用的增长度量取决于被度量数据的分布。本文论证了正态分布、对数正态分布和 DLN 随机变量的增长应采用不同的增长度量，并为 DLN 随机变量构建了适当的增长度量。

本文报告了大量蒙特卡洛实验的结果，旨在确定所提出的估计量和度量的性质。结果表明，矩估计量的准确度较高，但也凸显了其小样本偏差，尤其是在峰度方面。因此，有必要对峰度估计量采用小样本偏差校正方法。结果还表明，所提出的参数估计量具有较高的准确度，且基本无偏。为准确检验某些数据是否服从双对数正态分布，本文建立了临界值

以及三种分布检验的 p 值估计量：柯尔莫哥洛夫-斯米尔诺夫检验、卡方检验和安德森–达林检验。

第二个蒙特卡洛实验验证，文中讨论的广义增长度量确实能为正态分布、对数正态分布和双对数正态分布找出恰当的增长概念。实验尤其凸显了对随机变量采用错误的增长度量时所产生的“凸性/凹性偏差”。值得注意的是，证据表明，使用百分比增长来衡量对数正态随机变量的增长会导致显著的凸性偏差。

# 参考文献


Barro, R. J., Sala-I-Martin, X. I., 2003. 经济增长。麻省理工学院出版社，马萨诸塞州剑桥，第二版。

Bellemare, M. F., Wichman, C.J., 2020。《弹性与反双曲正弦变换》。《牛津经济学与统计学公报》82，50–61。

Cambanis, S., Huang, S., Simons, G., 1981。《关于椭圆等高分布理论》。《多元分析杂志》11，368–385。

D’Agostino, R.B., Stephens, M. A.（编），1986。《拟合优度检验技术》。统计学、教科书与专著第68卷，M. Dekker，纽约。

Fama, E. F., Roll, R., 1971。《对称稳定分布的参数估计》。《美国统计学会杂志》66，331–338。

Fang, K.-T., Kotz, S., Ng, K. W., 1990。《对称多元分布及相关分布》。CRC出版社。

Gubner, J., 2006. 对数正态特征函数的新公式。《IEEE车辆技术汇刊》55，1668–1671。

Gulisashvili, A., Tankov, P., 2016. 对数正态随机变量之和与差的尾部行为。《伯努利》22，444–493。

J¨antschi, L., Bolboac˘a, S. D., 2018. 与Anderson–Darling统计量相关的概率计算。《数学》6，88。

Lo, C. F., 2012. 两个对数正态随机变量之和与差。《应用数学杂志》2012，1–13。

Miller, M. B., 2013。《金融风险管理数学与统计学》。威利。

Parham, R., 2023年。0. 对数正态差分布是自然界中的基本分布。油印稿。

Stephens, M. A., 1979年。基于经验分布函数的逻辑斯蒂分布拟合优度检验。Biometrika 66,591–595。

Studen´y, V., Mezn´ık, I., 2013。增长率公理模型。Applied Mathe- matics04, 1326–1332。