Pure Mathematics 理论数学, 2018, 8(1), 105-112
Published Online January 2018 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2018.81013

The Basic Properties of Accumulated Residual Entropy and Its Application in Stock Analysis

Dan Zhou

School of Mathematics & Physics Science and Engineering, Anhui University of Technology, Ma'anshan Anhui
Email: zhoudan880919@163.com

Received: Jan. 9th, 2018; accepted: Jan. 24th, 2018; published: Jan. 31st, 2018

Abstract

On the basis of information entropy, we further study some properties of the cumulative residual entropy, and the cumulative residual entropy model applied in risk measure. In practical applications, the selection of eight stocks, collect the daily closing price data, using SPSS software to draw the cumulative residual entropy and the standard deviation of the diagram. It is concluded that the cumulative residual entropy and the standard deviation of the linear relationship, and because of the cumulative residual entropy of a broader scope than variance, namely in the risk measurement, the cumulative residual entropy model has more advantages than the variance.

Keywords

Entropy, Cumulative Residual Entropy, Risk Measurement, Variance

累积剩余熵的若干基本性质及其在股票分析中的应用

周　丹

安徽工业大学数理科学与工程学院，安徽　马鞍山
Email: zhoudan880919@163.com

收稿日期：2018年1月9日；录用日期：2018年1月24日；发布日期：2018年1月31日

摘　要

在信息熵的基础上，我们进一步研究累积剩余熵的一些性质，并将累积剩余熵模型应用于风险度量中。

文章引用: 周丹. 累积剩余熵的若干基本性质及其在股票分析中的应用[J]. 理论数学, 2018, 8(1): 105-112.
DOI: 10.12677/pm.2018.81013

再在实际应用中，选取8支股票，收集每日收盘价格数据，利用SPSS软件绘制累积剩余熵与标准差的关系图，得出累积剩余熵与标准差的线性关系，又因为累积剩余熵的适用范围比方差更广，即在度量风险中，累积剩余熵模型比方差更具有优势。

关键词

熵，累积剩余熵，风险度量，方差

1. 引言

金融市场随着时间的演化，越来越多的因素影响着它的变化，随着金融市场的复杂化，越来越多的人试图运用各种方法探索它的变化规律，以求获得利益。投资者获得的收益可能与期望不一致，因为投资具有很大的风险，通常巨大的利益常伴随着巨大的风险，金融市场中存在的各种各样的风险，这种风险越来越引起人们的重视，只有正确理解什么是金融风险，掌握影响金融风险的因素有哪些，才能恰当的度量金融风险，为金融活动提供有力保障。

风险一般有两种理解：一种强调了风险表现为收益不确定性；另一种则强调风险表现为成本或代价的不确定性。金融风险是指金融资产在未来时期内预期收入遭受损失的可能性，即金融风险与损失的不确定性有关。无论是对损失还是收益，风险都表现出一种不确定性，这种不确定性体现出风险的本质。为了减少损失，扩大收益，人们试图采用各种方法去度量风险，即利用已知数据来分析未来不确定事件的不确定性以及可能带来的损失或收益。

现有的度量风险方法有很多种，例如：波动性方法即方差(均方差)方法[1]、*Var* 方法[2]、灵敏度方法[3]、一致性风险度量方法[4]和失真函数方法[5]等等。每种方法都有它的优点和它的局限性。

20 世纪，shannon 提出了一个对离散随机变量不确定性的度量，即 shannon 熵，这与风险的本质相吻合。近年来，越来越引起人们的关注与研究，并将熵应用于金融投资组合、风险度量和数据压缩等领域中。本文中如无特别说明，总假定 log 是以 2 为底的对数。对于离散随机变量 X，其 shannon 熵[6]的定义为：$H(X)=-\sum_{x} p(x)\log p(x)$，其中 $p(x)$ 是随机变量 X 的概率密度函数，$0\le p(x)\le 1$，$\sum_{x} p(x)=1$，并约定 $0\ln 0=0$。当随机变量 X 是连续型时，我们称其不确定性的度量为微分熵，其定义为：$h(X)=-\int_{s} f(x)\log f(x)\mathrm{d}x$，其中 $f(x)$ 是随机变量 X 的概率密度函数，s 是 X 的支撑集(即使 $f(x)>0$ 的所有 x 构成的集合)。

但 shannon 熵也存在一些不足，例如：假设 X 和 Y 是两个离散的随机变量，X 的取值空间是 $\{1,2,3,4,5,6,7,8\}$，每一个的取值概率都为 $\frac{1}{8}$，Y 的取值空间是 $\{1,2,3,4,5,6,7,10000\}$，每一个的取值概率也都为 $\frac{1}{8}$，则 $H(X)=-\sum_{i=1}^{8}\frac{1}{8}\log\frac{1}{8}=\frac{27}{8}=H(Y)$，即 X 和 Y 的 shannon 熵是一样的，但如果 X 和 Y 代表的是博弈游戏中两个不同的回报方案，显然随机变量 X 和 Y 包含的信息量是不同的，但 shannon 熵对此并加以区分。为了克服 shannon 熵的一些不足，Murali Rao 等提出了一种新的熵概念，即累积剩余熵(Cumulative Residual Entropy，简记为 CRE)。

2. 累积剩余熵的基本性质

定义：[7]设 $X\in\mathbf{R}^N$ 是一个随机向量，则 X 的 CRE 定义为：

$$\varepsilon(X)=-\int_{\mathbf{R}_+^N}P(|X|>\lambda)\log P(|X|>\lambda)\mathrm{d}\lambda$$

其中 $X=(X_1,X_2,\cdots,X_N)$，$\lambda=(\lambda_1,\lambda_2,\cdots,\lambda_N)$，并且 $|X|>\lambda$ 意味着

$$|X_i|>\lambda_i,i=1,\cdots,N,\mathbf{R}_+^N=\{x_i\in\mathbf{R}^N;x_i\ge 0\}.$$

性质 1：若存在 $p>N$，使得对任意的 i，$E(|X_i|^p)<+\infty$ 恒成立，则 $\varepsilon(X)<+\infty$。

性质 2：[7]对于任意的非负且独立的随机变量 X 和 Y，有 $\max(\varepsilon(X),\varepsilon(Y))\le\varepsilon(X+Y)$。

性质 3：(条件使熵减少) 对任意的 $X,Y\in\mathbf{R}^N$，有 $E_X[\varepsilon(Y|X)]\le\varepsilon(X+Y)$，当且仅当 X,Y 相互独立时，等号成立。

此性质说明当已知条件越多时，未来事件的不确定性就下降了，即风险降低了。

此外，累积剩余熵中，由于 $p(x>\lambda)=1-F(\lambda)$，其中 $F(\lambda)$ 是其经验分布函数，在实际应用中，经验分布函数更容易得到，若对已知的分布，只需按定义即可计算出其剩余熵，若随机变量的分布形式未知，则可依据下面的定理，用经验 CRE$(\varepsilon(F_n))$ 逼近 CRE$(\varepsilon(F))$。

设 $X_1,X_2,\cdots,X_n$ 是正的且是独立同分布的随机变量，F 是其分布函数，令 F_n 是 n 个样本 $X_1,X_2,\cdots,X_n$ 的经验分布，且每个样本 $X_1,X_2,\cdots,X_n$ 的概率都是 $\frac{1}{n}$，令 $G(n)=1-F(n)$，则经验分布 F_n 的 CRE 为 $\varepsilon(F_n)=-\int_0^\infty G_n(x)\log G_n(x)\mathrm{d}x$。

性质 4：[7]设 $p>1$，X 是 L^p 空间中任意的随机变量，则它的经验 CRE 收敛于 CRE，即

$$\varepsilon(F_n)\to\varepsilon(F)\ \text{a.s.}$$

性质 5：设 X 是随机变量，若 X 的方差 $Var(X)$ 存在，则 X 的累积剩余熵 $\varepsilon(X)$ 存在。

证明：由于 X 的方差 $Var(X)$ 存在，即 $Var(X)=E(X^2)-E^2(X)<+\infty$，

所以

$$E(X^2)<E^2(X)<+\infty,$$

由性质 1 可得，$\varepsilon(X)<+\infty$，即 X 的累积剩余熵 $\varepsilon(X)$ 存在。

由性质 5，我们可以看出，只要随机变量的方差存在，则其累积剩余熵必存在，反之，当累积剩余熵存在时，方差是否存在呢？为此我们以 Pareto 分布为例比较一下。

Pareto 分布是一种比较常见的厚尾分布，厚尾分布是金融市场中比较常见的分布。首先，我们了解一下 Pareto 分布的密度函数，

$$p(X)=(\alpha/X)^k,X\in[\alpha,\infty),k>0,\alpha>0$$

由此可得，其方差为：

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

其累积剩余熵为：

$$\text{CRE}(X)=\frac{k\alpha}{(k-1)^2},k>1$$

由以上方差与累积剩余熵的算式，我们可以看出，当$1<k\le 2$时，其累积剩余熵存在，但其方差不存在，即累积剩余熵的适用范围比方差更广。

下面我们研究一下，当随机变量的方差与累积剩余熵都存在时，方差与累积剩余熵是否有关系。首先，我们比较一下几种常见分布的方差与累积剩余熵。

1) 均匀分布的 CRE。

设$X\sim U[0,a]$，则

$$\sigma^2=Var(X)=\frac{a^2}{12},\ \varepsilon(X)=\frac{a}{4},$$

即

$$\varepsilon(X)=\frac{\sqrt{3}}{2}\sigma$$

2) 指数分布的 CRE。

设$X\sim \exp\left(\frac{1}{\lambda}\right)$，则

$$\sigma^2=Var(X)=\frac{1}{\lambda^2},\ \varepsilon(X)=\frac{1}{\lambda},$$

则

$$\varepsilon(X)=\sigma$$

3) Pareto 分布的 CRE。

由上面的分析，当$k>2$时，

$$\varepsilon(X)=\frac{\sqrt{k(k-2)}}{k-1}\sigma$$

由以上几个例子，可以看出随机变量的累积剩余熵与它的标准差呈线性关系。

3. 实证分析

下面我们举例看一下它的应用。

例现有两个转盘 A,B 的抽奖游戏，中奖的钱X,Y(单位：元)以及概率如表 1，表 2 所示：

Table 1. The amount and probability of A winning the turntable A
表 1. 转盘 A 中奖的钱数及概率

X	6	7	8
p	1/3	1/3	1/3

Table 2. The amount of money and probability of winning the turntable B
表 2. 转盘 B 中奖的钱数及概率

Y	0	8
p	1/8	7/8

由表 1，表 2，我们可以计算出它们的均值、方差、熵以及累积剩余熵，比较一下，哪一个更准确。

首先计算出均值：$E(X)=7, E(Y)=7$，即从均值的角度看，两个转盘游戏是一样的，风险相同。

其次计算出方差：$D(X)=\frac{2}{3}, D(Y)=7$，即从方差的角度看，转盘 2 的波动比较大，风险更高。

下面计算熵：

$$H(X)=-\sum_{i=1}^{3} p_i \log p_i=-\left(\frac{1}{3}\log\frac{1}{3}+\frac{1}{3}\log\frac{1}{3}+\frac{1}{3}\log\frac{1}{3}\right)=\log 3\approx 1.5849625$$

$$H(Y)=-\sum_{i=1}^{2} p_i \log p_i=-\left(\frac{1}{8}\log\frac{1}{8}+\frac{7}{8}\log\frac{7}{8}\right)=3-\frac{7}{8}\log 7\approx 0.54356$$

因为 $H(X)>H(Y)$，即转盘 1 的风险更高，但此与实际不符。

最后我们看一下它们的累积剩余熵：

首先，设 $F(X), F(Y)$ 为它们的分布，$\bar{F}(X)=1-F(X)$，$\bar{F}(Y)=1-F(Y)$，

$$\bar{F}(X)=\begin{cases}0, & X\ge 8\\ \frac{1}{3}, & 7\le X<8\\ \frac{2}{3}, & 6\le X<7\\ 1, & X<6\end{cases},\quad \bar{F}(Y)=\begin{cases}0, & Y\ge 8\\ \frac{7}{8}, & 0\le Y<8\\ 1, & Y<0\end{cases}$$

则

$$\begin{aligned}\varepsilon(X)&=-\int_{-\infty}^{+\infty}\bar{F}(X)\log\bar{F}(X)\mathrm{d}X\\&=-\left(\int_{-\infty}^{6}1\times\log 1\mathrm{d}X+\int_{6}^{7}\frac{2}{3}\log\frac{2}{3}\mathrm{d}X+\int_{7}^{8}\frac{1}{3}\log\frac{1}{3}\mathrm{d}X+\int_{8}^{+\infty}0\log 0\mathrm{d}X\right)\\&=\log 3-\frac{2}{3}\\&\approx 0.9182958\end{aligned}$$

$$\begin{aligned}\varepsilon(Y)&=-\int_{-\infty}^{+\infty}\bar{F}(Y)\log\bar{F}(Y)\mathrm{d}Y\\&=-\left(\int_{-\infty}^{0}1\times\log 1\mathrm{d}Y+\int_{0}^{8}\frac{7}{8}\log\frac{7}{8}\mathrm{d}Y+\int_{8}^{+\infty}0\log 0\mathrm{d}Y\right)\\&=21-7\log 7\\&\approx 1.3485\end{aligned}$$

显然，$\varepsilon(X)<\varepsilon(Y)$，即转盘 2 的风险大于转盘 1 的风险。

对于独立同分布的随机变量 $X_1, X_2, \cdots, X_n$，$F(X)$ 是其分布函数，随机变量 $X_1, X_2, \cdots, X_n$ 的经验分布函数为 $F_n(X)$，$F_n(X)=\frac{1}{n}\sum_{i=1}^{n} I_{\{X\ge X_i\}}$，其中 $I_{\{X\ge X_i\}}$ 是示性函数，$G_n(X)=1-F_n(X)$，首先，可以先将 n 个样本按从小到大的顺序重新排列，使得 $X_1\le X_2\le\cdots\le X_n$。则

$$\begin{aligned}\varepsilon(F_n)&=-\int_{0}^{\infty}G_n(X)\log G_n(X)\mathrm{d}X\\&=-\frac{1}{n}\sum_{i=1}^{n-1}\frac{n-i}{n}\log\frac{n-i}{n}\left(X_{i+1}-X_i\right)\end{aligned}$$

为了分析股票市场中的不确定性，下面选了 8 支股票(海螺水泥，长城汽车，工商银行中国国航，青

岛啤酒股份，马鞍山钢铁股份，大唐发电以及中国石油股份)的每日收盘价格(从 2014 年 1 月 3 日到 2017 年 8 月 30 日)为样本(数据来源：东方财富网)。

根据相应的公式，我们可以计算出 8 支股票对应的平均值、标准差、偏度(偏态的系数)、峰度(峰态的系数)以及累积剩余熵(CRE)，如表 3 所示。

由表 3 可以看出，8 支股票都有一定的偏度和峰度，不是标准的正态分布，而且由上表，我们可以看出标准差与累积剩余熵有一定的联系，标准差大的股票，其对应的累积剩余熵也较大；标准差小的股票，其对应的累积剩余熵也较小。

为了进一步观察标准差与累积剩余熵的关系，我们在 8 支股票中任选一支股票，例如中国石油股份，将其 900 个数据(从 2014 年 1 月 3 日到 2017 年 8 月 30 日)按时间顺序分成 15 份，如表 4 所示。

Table 3. The values of each of the eight stocks
表 3. 8 支股票对应的各参数值

股票	平均值	标准差	偏度	峰度	CRE
海螺水泥	25.15471111	4.143943572	−0.34573552	−0.378607978	0.00567103
长城汽车	21.85918889	15.81356261	0.560897073	−1.085948676	0.023137935
工商银行	4.998355556	0.625932511	0.859293615	1.208585424	0.000954
中国国航	6.075777778	1.320075905	1.025236892	0.371892073	0.002048
青岛啤酒股份	42.03461111	11.99408778	0.317773978	−1.505081472	0.016755
马鞍山钢铁股份	2.108133333	0.59367379	1.272127076	1.002785755	0.000928
大唐发电	2.958333333	0.83296801	0.543866136	−0.996340603	0.001232
中国石油股份	7.009133333	1.925877095	0.46147846	−1.230233326	0.002801

Table 4. Standard deviation and CRE of petrochina
表 4. 中国石油股份的标准差与 CRE

组别	平均值	标准差	CRE
1	7.942	0.25514	0.000856
2	9.1665	0.362332	0.001023
3	10.534333	0.491367	0.001696
4	9.0878333	0.673325	0.002053
5	8.6088333	0.214477	0.000643
6	9.3696667	0.638645	0.001987
7	7.4008333	0.96352	0.002661
8	5.83	0.365434	0.001049
9	4.8816667	0.360391	0.000845
10	5.3373333	0.231135	0.000909
11	5.2951667	0.093183	0.00032
12	5.272	0.185963	0.000653
13	5.963667	0.228265	0.000559
14	5.521333	0.232689	0.000594
15	4.925833	0.091989	0.000273

由表 4 我们可以画出标准差与累积剩余熵的图形，如图 1~3 所示。

由图 3，我们可以看出标准差与累积剩余熵呈线性关系。由以上分析，我们可以看出，当方差波动较大时，累积剩余熵[5]也变大，反之，则累积剩余熵变小。当经验分布是重尾的时，累积剩余熵比方差更有利。特别是当经验分布服从 Levy 分布时，其均值与方差不存在，无法利用方差度量时，累积剩余熵仍然可以计算出来。累积剩余熵的计算比较简单，容易实现，由于经验分布趋向于总体分布，而经验分布是比较容易得到的，而且不论是连续型随机变量还是离散型随机变量，都可以利用累积剩余熵计算出来。虽然金融市场的复杂性很大，针对不同的风险源有不同的度量方法，但累积剩余熵在金融投资中有一定的实际意义。

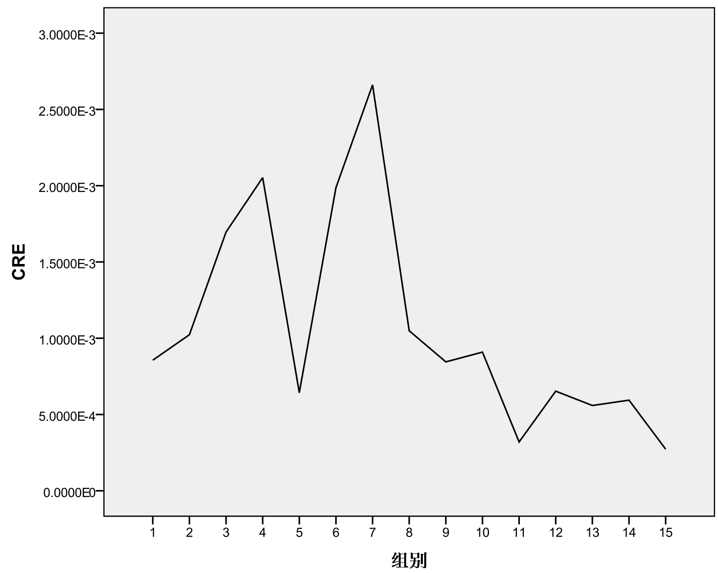

Figure 1. Sequence diagram of CRE
图 1. CRE 序列图

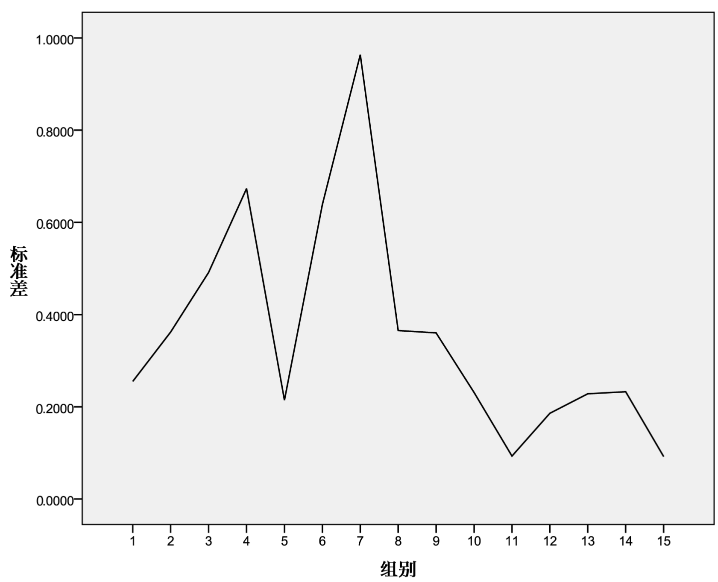

Figure 2. The standard deviation sequence diagram
图 2. 标准差的序列图

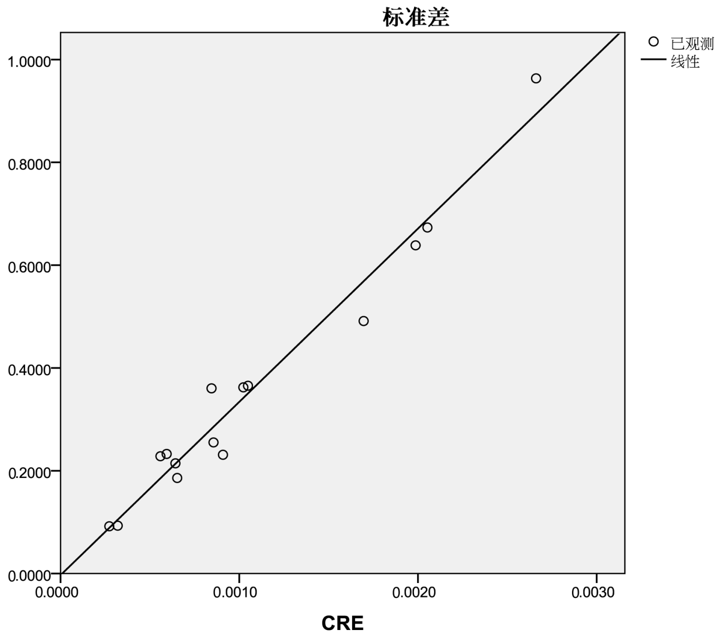

Figure 3. The linear correlation between standard deviation and CRE
图 3. 标准差与 CRE 之间的线性相关性

致　谢

感谢我的导师对我的悉心教导，感谢我的同学和朋友在学习、生活上的关心和帮助，最后，感谢我的家人，一直以来，无论是学习和生活，都离不开他们的支持、照顾和理解。

基金项目

安徽工业大学研究生创新基金资助(2016137)。

参考文献 (References)

[1] Mar kowitz, H. (1952) Portfolio selection. *Journal of Finance*, **7**, 71-91.

[2] Jorion, P. (1996) Measure the Risk in Value at Risk. *Financial Analysts Journal*, **11**, 47-55.

[3] Sharpe, W.F. (1964) Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, **19**, 425-442.

[4] Wirch, J.I. and Hardy, M.R. (1999) A Synthesis of Risk Measures for Capital Adequacy. *Insurance: Mathematics and Economics*, **25**, 337-347.

[5] 杨丽娟, 李兴斯. 度量尾部风险的剩余熵模型[J]. 运筹与管理, 2010, 19(6): 88-103.

[6] Thomas, M.C. and Thomas, J.A. 信息论基础[M]. 第 2 版. 北京: 机械工业出版社, 2017.

[7] Rao, M., Chen, Y., Vemuri, B.C., *et al*. (2004) Cumulative Residual Entropy, A New Measure of Information & Its Application to Image Alignment. *IEEE Transactions on Information Theory*, **50**, 1220-1228.

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2160-7583，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：pm@hanspub.org