Pure Mathematics 理论数学, **2023**, **13**(3), 533-540
Published Online March 2023 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2023.133057

Hans汉斯

双截断随机变量的Cox-Czanner散度

赵苏媛

西北师范大学数学与统计学院，甘肃 兰州

收稿日期：2023年2月16日；录用日期：2023年3月16日；发布日期：2023年3月27日

摘 要

Cox和Czanner (2016)提出了生存散度的概念并研究了一些分布的生存散度。作为KL散度的推广，生存散度在统计学、生态学等领域获得了广泛应用。本文提出了**Cox-Czanner**散度在双截断随机变量下的分布差异，通过广义失效率的方法研究了双截断随机变量生存散度的有界性和单调性，并讨论了单调变换对生存散度的影响，最后通过分布的变换将双截断随机变量的生存散度应用到比例优势模型进行实例检验。

关键词

Cox-Czanner散度，广义失效率，特征，比例优势模型

Cox-Czanner Divergence of a Doubly Truncated Random Variable

Suyuan Zhao

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: Feb. 16^{th}, 2023; accepted: Mar. 16^{th}, 2023; published: Mar. 27^{th}, 2023

文章引用: 赵苏媛. 双截断随机变量的Cox-Czanner散度[J]. 理论数学, 2023, 13(3): 533-540.
DOI: 10.12677/pm.2023.133057

Abstract

Cox and Czanner (2016) put forward the concept of survival divergence and studied the survival divergence of some distributions. As a generalization of KL divergence, survival divergence has been widely used in statistics, ecology and other fields. This paper proposes the distribution difference of Cox-Czanner divergence under double-truncated random variables, studies the boundedness and monotonicity of the survival divergence of double-truncated random variables by means of generalized failure rate method, and discusses the influence of monotonic transformation on the survival divergence. Finally, the survival divergence of double truncated random variables is applied to the proportional dominance model by the transformation of distribution.

Keywords

Cox-Czanner Divergence, Generalized Failure Rate, Characteristic, Proportional Odd Model

1. 引言

散度度量的理论知识在近几十年来得到了很好的研究, 并广泛地应用在工程和医学等各个领域, 最常见的度量有Kullback- leibler 散度(KL divergence) [1]、Renyi 散度 [2]、Kagans散度(卡方距离) [3]、Cox-Czanner散度等. 关于散度度量理论和应用的更多细节, 可以参考Basseville, M. [4], Vonta, F.和Karagrigoriou, A. [5]等文献. 其中Cox-Czanner [6]散度因度量两组生存函数之间的差异而得到了广泛应用, 例如可以解释为是一组患者在时间t 死亡而另一组患者在时间t 后存活的绝对概率差的积分. 假设连续随机变量X和Y分别有分布函数$F(x)$ 和$G(x)$，生存函数$\bar{F}(x)$ 和$\bar{G}(x)$, 概率密度函数$f(x)$ 和$g(x)$ 以及失效率函数$h_1(x) = f(x)/\bar{F}(x)$ 和$h_2(x) = g(x)/\bar{G}(x)$,则Cox-Czanner散度的定义为,

$$I(F,G) = \int_0^\infty \left| f(x)\bar{G}(x) - g(x)\bar{F}(x) \right| \mathrm{d}x. \tag{1}$$

2022年, Mansourzar和Asadi [7]进一步扩展了Cox-Czanner散度, 讨论了两种剩余寿命X_t和Y_t之间的散度, 其中$X_t = (X - t|X > t)$, $Y_t = (Y - t|Y > t)$, X_t和Y_t 的概率密度函数分别为$f_{X_t} = f(x+t)/\bar{F}(t)$ 和$g_{Y_t} = g(x+t)/\bar{G}(t)$, 若

$$3D(F,G;t) = \frac{\int_t^\infty \left|f(x)\bar{G}(x) - g(x)\bar{F}(x)\right| \mathrm{d}x}{\bar{F}(t)\bar{G}(t)}. \tag{2}$$

则称$D(F,G;t)$为Cox-Czanner剩余寿命散度.

2022年, Mansourzar [8]研究了剩余寿命的对偶度量, 即休止时间${}_tX = (t - X|X \le t)$ 和${}_tY = (t - Y|Y \le t)$的散度, 其概率密度函数分别为$f_{{}_tX}(x) = f(x)/F(t)$和$g_{{}_tY}(x) = g(x)/G(t)$, 累积分布函数分别为$F_{{}_tX}(x) = F(x)/F(t)$ 和$G_{{}_tY}(x) = G(x)/G(t)$,若

$$\bar{D}(F,G;t) = \frac{\int_0^t |f(x)G(x) - g(x)F(x)| \,\mathrm{d}x}{F(t)G(t)}. \tag{3}$$

则称$\bar{D}(F,G;t)$为Cox-Czanner休止时间散度.

近年来, 关于双截断随机变量的度量问题越来越受欢迎 [4, 9–11], 在生存分析和可靠性工程等领域更是备受关注. 例如当一个系统的寿命处于区间(t_1,t_2)时, 需要研究该区间内的寿命信息. 为更准确的度量双截断随机变量的信息, 本文基于Cox-Czanner剩余寿命和休止时间散度, 研究了双截断随机变量的Cox-Czanner散度, 下面给出其定义.

定义 1 *假设X和Y表示两个连续随机变量, 分布函数分别为$F(x)$和$G(x)$,密度函数分别为$f(x)$和$g(x)$, $X_{t_1,t_2} = [X|t_1 < X < t_2]$ 和$Y_{t_1,t_2} = [Y|t_1 < Y < t_2]$ 分别是X和Y 相关的双截断时间, $(t_1,t_2) \in \mathcal{D} = \{(x,y)|F(x) < F(y)$和$G(x) < G(y)\}$, 若*

$$\begin{aligned} ID(X,Y;t_1,t_2) &= \int_0^\infty |f_{t_1,t_2}(x)G_{t_1,t_2}(x) - g_{t_1,t_2}(x)F_{t_1,t_2}(x)| \,\mathrm{d}x \\ &= \frac{\int_{t_1}^{t_2} |f(x)G(x) - g(x)F(x)| \,\mathrm{d}x}{\Delta F \Delta G} + \left|\frac{F(t_1)}{\Delta F} - \frac{G(t_1)}{\Delta G}\right|, \end{aligned} \tag{4}$$

其中$\Delta F = F(t_2) - F(t_1)$, $\Delta G = G(t_2) - G(t_1)$,则称$ID(X,Y;t_1,t_2)$为双截断Cox-Czanner散度.

本文的安排如下：在第2节中, 我们给出了关于双截散度的主要结果. 具体地, 我们获得了散度的界. 在第3节中研究了双截断散度的单调行为以及单调变换对双截断散度的影响. 在第4节中, 对某些类型的转换模型进行了散度度量的评估.

2. 双截断Cox-Czanner散度的性质

2.1. 有界性

定理 1 *假设X和Y是两个随机变量, $X_{t_1,t_2} = [X|t_1 < X < t_2]$ 和$Y_{t_1,t_2} = [Y|t_1 < Y < t_2]$ 分别是与X和Y 相关的双截断时间, 对(4)式中表示的散度度量$ID(X,Y;t_1,t_2)$, 有$0 \le ID(X,Y;t_1,t_2) \le 1$.*

证明 显然, 与时间相关的散度度量$ID(X,Y;t_1,t_2)$是非负的, 即$ID(X,Y;t_1t_2)\geq 0$, 当且仅当$F(x)=G(x)$ 时等号成立. 因此只需证明$ID(X,Y;t_1t_2)\leq 1$.

$$\begin{aligned}
&ID(X,Y;t_1,t_2)\\
&=\frac{\int_{t_1}^{t_2}|f(x)(G(t_2)-G(x))-g(x)(F(t_2)-F(x))|\,dx}{\Delta F\Delta G}\\
&=\left|\frac{\int_{t_1}^{t_2}f(x)G(t_2)dx-\int_{t_1}^{t_2}f(x)G(x)}{\Delta F\Delta G}-\frac{\int_{t_1}^{t_2}g(x)F(t_2)dx-\int_{t_1}^{t_2}g(x)f(x)dx}{\Delta F\Delta G}\right| \qquad (5)\\
&\leq\left|\frac{G(t_2)\Delta F-G(t_2)\Delta F-F(t_2)\Delta G+F(t_2)G(t_2)-F(t_1)G(t_1)-G(t_2)\Delta F}{\Delta F\Delta G}\right|\\
&=1.
\end{aligned}$$

证明完成.

2.2. 单调性

定理 2 设随机变量X和Y分别有密度函数$f(x)$和$g(x)$, 分布函数分别为$F(x)$和$G(x)$.

(i) 对于任意$t_1\leq t_2$, t_2固定时, 若$ID(X,Y;t_1,t_2)$关于是t_1递增(递减)的, 则$ID(X,Y;t_1,t_2)\geq(\leq)$

$$\frac{\left|h_1^X(t_1,t_2)\frac{G(t_1)}{\Delta G}-h_1^Y(t_1,t_2)\frac{F(t_1)}{\Delta F}\right|-\left|h_1^X(t_1,t_2)\frac{F(t_2)}{\Delta F}-h_1^Y(t_1,t_2)\frac{G(t_2)}{\Delta G}\right|}{h_1^X(t_1,t_2)+h_1^Y(t_1,t_2)}; \qquad (6)$$

(ii) 对于任意$t_1\leq t_2$, t_1固定时, 若t_2 关于$ID(X,Y;t_1,t_2)$ 是递增(递减)的, 则$ID(X,Y;t_1,t_2)\leq(\geq)$

$$\frac{\left|h_2^X(t_1,t_2)\frac{G(t_2)}{\Delta G}-h_2^Y(t_1,t_2)\frac{F(t_2)}{\Delta F}\right|+\left|h_2^Y(t_1,t_2)\frac{G(t_1)}{\Delta G}-h_2^X(t_1,t_2)\frac{F(t_1)}{\Delta F}\right|}{h_2^X(t_1,t_2)+h_2^Y(t_1,t_2)}. \qquad (7)$$

证明 (4)分别关于t_1和t_2求导, 得到

$$\begin{aligned}
&\frac{\partial}{\partial t_1}ID(X,Y;t_1,t_2)\\
&=(h_1^X(t_1,t_2)+h_1^Y(t_1,t_2))ID(X,Y;t_1,t_2)-\left|h_1^Y(t_1,t_2)\frac{F(t_1)}{\Delta F}-h_1^X(t_1,t_2)\frac{G(t_1)}{\Delta G}\right| \qquad (8)\\
&\quad+\left|h_1^Y(t_1,t_2)\frac{G(t_2)}{\Delta G}-h_1^X(t_1,t_2)\frac{F(t_2)}{\Delta F}\right|,
\end{aligned}$$

和

$$\begin{aligned}&\frac{\partial}{\partial t_2}ID(X,Y;t_1,t_2)\\&=-(h_1^X(t_1,t_2)+h_1^Y(t_1,t_2))ID(X,Y;t_1,t_2)+\left|h_2^Y(t_1,t_2)\frac{F(t_2)}{\Delta F}-h_2^X(t_1,t_2)\frac{G(t_2)}{\Delta G}\right| \qquad (9)\\&\quad+\left|h_2^X(t_1,t_2)\frac{F(t_1)}{\Delta F}-h_2^Y(t_1,t_2)\frac{G(t_1)}{\Delta G}\right|.\end{aligned}$$

进而令$\frac{\partial ID(X,Y;t_1t_2)}{\partial t_1}\geq(\leq)0$, $\frac{\partial ID(X,Y;t_1t_2)}{\partial t_2}\geq(\leq)0$, 经过简化，得到(6)和(7),即定理证明完成.

3. 单调变换

现在讨论单调变换对双截断Cox-Czanner 散度$ID(X,Y;t_1t_2)$的影响.

定理 3 设X和Y是两个绝对连续的非负随机变量, 概率密度函数分别为$f(x)$ 和$g(x)$, 分布函数分别为$F(x)$和$G(x)$. 若双射函数ϕ_1和ϕ_2是严格单调且可微的, 则对于所有$0\leq t_1<t_2<+\infty$, 有

$$ID(\phi_1(X),\phi_2(Y);t_1,t_2)=\begin{cases}ID^{\phi_1}(X,\phi_1^{-1}(\phi_2(Y));\phi_1^{-1}(t_1),\phi_1^{-1}(t_2)),\\\qquad\text{如果}\phi_1\text{ 和}\phi_2\text{ 是严格单调递增的;}\\ID^{\phi_1}(X,\phi_1^{-1}(\phi_2(Y));\phi_1^{-1}(t_2),\phi_1^{-1}(t_1)),\\\qquad\text{如果}\phi_1\text{ 和}\phi_2\text{ 是严格单调递减的.}\end{cases}$$

证明 如果$\phi_1(x)$ 和$\phi_2(x)$ 是严格递增函数, 那么$\phi_1(X)$和$\phi_2(Y)$ 的概率密度函数和分布函数分别可以表示为

$$f_{\phi_1}(x)=\frac{f(\phi_1^{-1}(x))}{\phi_1{}'(\phi^{-1}(x))}\qquad F_\phi(x)=F(\phi_1^{-1}(x)), \qquad (10)$$

和

$$g_{\phi_2}(x)=\frac{g(\phi_2^{-1}(x))}{\phi_1{}'(\phi^{-1}(x))}\qquad G_\phi(x)=G(\phi_2^{-1}(x)), \qquad (11)$$

此外, 可以得到$\phi_1^{-1}(\phi_2(x)$ 的概率密度函数和分布函数分别为

$$g_{\phi_1^{-1}(\phi_2)}(x)=\frac{g(\phi_2^{-1}(\phi_1(x))\phi_1{}'(x)}{\phi_2{}'(\phi_2^{-1}(\phi_1(x))},\qquad G_{\phi_1^{-1}(\phi_2)}(x)=G(\phi_2^{-1}(\phi_1(x))). \qquad (12)$$

将(10)和(11)代入(4)得到

$$\begin{aligned} &ID(\phi_1(X),\phi_2(Y);t_1,t_2)\\ &=\left|\frac{\int_{t_1}^{t_2}(F(\phi_1^{-1}(x))g(\phi_2^{-1}(x))/\phi_2{}'(\phi_2^{-1}(x))-f(\phi_1^{-1}(x))G(\phi_2^{-1}(x))/\phi_1{}'(\phi_1^{-1}(x)))dx}{\Delta F^{\phi_1}\Delta G^{\phi_2}}\right|\\ &+\left|\frac{G(\phi_2^{-1}(t_2))}{\Delta G^{\phi_2}}-\frac{F(\phi_1^{-1}(t_1))}{\Delta F^{\phi_1}}\right|, \end{aligned} \tag{13}$$

其中$\Delta F^{\phi_1}=F(\phi_1^{-1}(t_2))-F(\phi_1^{-1}(t_1))$ 和$\Delta G^{\phi_1}=G(\phi_1^{-1}(t_2))-G(\phi_1^{-1}(t_1))$, 对(13)使用变换$u=\phi_1^{-1}(x)$有

$$\begin{aligned} &ID(\phi_1(X),\phi_2(Y);t_1,t_2)\\ &=\left|\frac{\int_{\phi_1^{-1}(t_1)}^{\phi_1^{-1}(t_2)}(F(u)g(\phi_2^{-1}(\phi_1(u)))\phi'(u)/\phi_2{}'(\phi_2^{-1}(\phi_1(u)))-f(u)G(\phi_2^{-1}(\phi_1(u)))/)du}{\Delta F^{\phi_1}\Delta G^{\phi_2}}\right|\\ &\quad+\left|\frac{G(\phi_2^{-1}(t_2))}{\Delta G^{\phi_2}}-\frac{F(\phi_1^{-1}(t_1))}{\Delta F^{\phi_1}}\right|,\\ &=ID^{\phi_1}(X,\phi_1^{-1}(\phi_2(Y));\phi_1^{-1}(t_1),\phi_1^{-1}(t_2)). \end{aligned} \tag{14}$$

定理3的证明就完成了.

4. 比例优势模型

现在我们研究分布变换模型下的散度度量$ID(X,Y;t_1,t_2)$, 并运用双截断Cox-Czanner散度对比例优势模型进行度量.

定义 2 (分布变换模型) 设H是连续分布函数, 其概率密度函数$h\in[0,1]$, 作为转换式或链接函数, F基分布函数,H为分布变换函数G, 若对所有的$x>0$,有

$$G(x)=H(F(x)), \tag{15}$$

则称F和G满足分布变换模型.

4.1. 比例优势模型

定义 3 (比例优势模型) 假设随机变量X和Y分别具有生存函数$\bar{F}(x)$和$\bar{G}(x)$, 分布函数分别为$F(x)$和$G(x)$, 若存在比例常数$\theta>0$, 对所有的$x>0$, 都有

$$\frac{\bar{G}(x)}{G(x)}=\theta\frac{\bar{F}(x)}{F(x)}, \tag{16}$$

则称X和Y满足比例优势模型.

定理 4 假设随机变量X和Y分别有分布函数$F(x)$和$G(x)$以及生存函数$\bar{F}(x)$和$\bar{G}(x)$, 取分布变

换$H(x)=x/(\theta+(1-\theta)x)$, 其中$\theta<x<1,\theta>0$, 则有

$$ID(X,Y;t_1,t_2)=\frac{F(t_2)+F(t_1)}{F(t_2)-F(t_1)}+\frac{2\theta}{(1-\theta)(F(t_2)-F(t_1))}-\frac{2(\theta+(1-\theta)F(t_2))(\theta+(1-\theta)F(t_1))\log\frac{\theta+(1-\theta)F(t_2)}{\theta+(1-\theta)F(t_1)}}{(1-\theta)^2(F(t_2)-F(t_1))^2}. \tag{17}$$

证明 将分布变换$H(x)$代入(15)式, 则F和G属于比例优势模型类(16). 因此(16)式可以重新改写为

$$G(x)=\frac{F(x)}{\theta+(1-\theta)F(x)},\quad x>0, \tag{18}$$

此时

$$g(x)=\frac{\theta f(x)}{[\theta+(1-\theta)F(x)]^2}, \tag{19}$$

将(18) 式和(19) 式代入(4)式中，得到如下结果

$$\begin{aligned}
&ID(X,Y;t_1,t_2)\\
&=\frac{\int_{t_1}^{t_2}\frac{(1-\theta)f(x)F^2(x)}{[\theta+(1-\theta)F(x)]^2}dx+\frac{F(t_1)F(t_2)}{\theta(1-\theta)F(t_2)}-\frac{F(t_1)F(t_2)}{\theta(1-\theta)F(t_1)}}{\frac{\theta(F(t_2)-F(t_1)^2)}{(\theta+(1-\theta)F(t_2))(\theta+(1-\theta)F(t_1))}}\\
&=\frac{(\theta+(1-\theta)F(t_2))(\theta+(1-\theta)F(t_1))}{\theta(1-\theta)(F(t_2)-F(t_1))}-\frac{(1-\theta)F(t_1)F(t_2)}{\theta(F(t_2)-F(t_1))}\\
&\quad+\frac{\theta}{(1-\theta)(F(t_2)-F(t_1))}-\frac{2(\theta+(1-\theta)F(t_2))(\theta+(1-\theta)F(t_1))\log\frac{\theta+(1-\theta)F(t_2)}{\theta+(1-\theta)F(t_1)}}{(1-\theta)^2(F(t_2)-F(t_1)^2)}\\
&=\frac{F(t_2)+F(t_1)}{F(t_2)-F(t_1)}+\frac{2\theta}{(1-\theta)(F(t_2)-F(t_1))}\\
&\quad-\frac{2(\theta+(1-\theta)F(t_2))(\theta+(1-\theta)F(t_1))\log\frac{\theta+(1-\theta)F(t_2)}{\theta+(1-\theta)F(t_1)}}{(1-\theta)^2(F(t_2)-F(t_1))^2}.
\end{aligned} \tag{20}$$

证明完毕.

5. 总结

艾滋病是目前严重危害人类健康和生命的传染之一，其潜伏期被认为是从感染人类免疫缺陷病毒1型到诊断为艾滋病之间的时间 [12]. 近几十年来国内外对艾滋病潜伏期的研究进行了大量探索，这也是艾滋病流行病学研究的重要内容. 艾滋病潜伏期数据中存在大量的删失或截断的不完全数据，所以利用该双截断对其进行合理有效分析将会对艾滋病的研究提供重要帮助，但由于缺乏真实数据，对此不再举实例说明.

参考文献

[1] Kullback, S. and Leibler, R.A. (1951) On Information and Sufficiency. *Annals of the Institute of Statistical Mathematics*, **22**, 79-86. https://doi.org/10.1214/aoms/1177729694

[2] Renyi, A. (1961) On Measures of Entropy and Information. *Mathematical Statistics and Probability*, **4**, 547-561.

[3] Nikulin, M. (2001) Hellinger Distance. Vol. 10, Springer-Verlag, Berlin.

[4] Basseville, M. (2013) Divergence Measures for Statistical Data Processing—An Annotated Bibliography. *Signal Process*, **93**, 621-633. https://doi.org/10.1016/j.sigpro.2012.09.003

[5] Vonta, F. and Karagrigoriou, A. (2010) Generalized Measures of Divergence in Survival Analysis and Reliability. *Journal of Applied Probability*, **47**, 216-234. https://doi.org/10.1239/jap/1269610827

[6] Cox, T.F. and Czanner, G. (2016) A Practical Divergence Measure for Survival Distributions That Can Be Estimated from Kaplan-Meier Curves. *Statistics in Medicine*, **35**, 2406-2421. https://doi.org/10.1002/sim.6868

[7] Mansourvar, Z. and Asadi, M. (2020) An Extension of the Cox-Czanner Divergence Measure to Residual Lifetime Distributions with Applications. *Statistics*, **54**, 1311-1328. https://doi.org/10.1080/02331888.2020.1862117

[8] Mansourvar, Z. (2022) A Dynamic Measure of Divergence between Two Inactivity Lifetime Distributions. *Statistics*, **56**, 147-163. https://doi.org/10.1080/02331888.2022.2038164

[9] Poursaeed, M.H. and Nematollahi, A.R. (2008) On the Mean Past and the Mean Residual Life under Double Monitoring. *Communications in Statistics—Theory and Methods*, **37**, 1119-1133. https://doi.org/10.1080/03610920701762796

[10] Khorashadizadeh, M., Rezaei Roknabadi, A.H. and Mohtashami Borzadaran, G.R. (2012) Characterizations of Lifetime Distributions Based on Doubly Truncated Mean Residual Life and Mean Past to Failure. *Communications in Statistics—Theory and Methods*, **41**, 1105-1115. https://doi.org/10.1080/03610926.2010.535626

[11] Betensky, R.A. and Martin, E.C. (2003) Commentary: Failure-Rate Functions for Doubly Truncated Random Variables. *IEEE Transactions on Reliability*, **52**, 7-8. https://doi.org/10.1109/TR.2002.807241

[12] Kirmani, S. and Gupta, R.C. (2001) On the Proportional Odds Model in Survival Analysis. *Annals of the Institute of Statistical Mathematics*, **53**, 203-216. https://doi.org/10.1023/A:1012458303498