پانزدهمین سمینار احتمال و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان
15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan
انجمن آمار ایران

آزمون نیکویی برازش تقارن توزیع‌های پیوسته تحت داده‌های از چپ بریده شده و از راست سانسور شده

[۱]

چکیده: در این مقاله، ابتدا یک آماره آزمون برای متقارن بودن توزیع‌های پیوسته تحت داده‌های از چپ بریده شده و از راست سانسور شده معرفی می‌شود. نشان داده می‌شود که این آماره آزمون مجانباً نرمال است. در یک مطالعه شبیه سازی گسترده، توان آزمون پیشنهادی مورد بررسی قرار گرفته است. نتایج گویای رضایت بخش بودن عملکرد آزمون است. در نهایت کاربردی بودن آماره آزمون در یک مثال واقعی نشان داده می‌شود.

واژه‌های کلیدی: آزمون نیکویی برازش، توزیع متقارن، داده‌های سانسور شده از راست و بریده شده از چپ.

کد موضوع‌بندی ریاضی (۲۰۲۰): 62N01، 62G10.

۱ مقدمه

فرض کنید X_1، X_2،...، X_n یک نمونه تصادفی به حجم n، از یک تابع چگالی نامعلوم f و میانگین متناهی نامعلوم μ باشد. گوییم چگالی f حول μ متقارن است اگر و فقط اگر

$$H_0 : f(\mu - x) = f(\mu + x), \qquad \forall x \in \mathbb{R}. \tag{۱.۱}$$

بررسی فرض متقارن بودن توسط محققین در خیلی از زمینه های علمی جهت شناسایی مدل آماری داده‌ها ضروری بنظر می‌رسد. بعنوان مثال، در مطالعات مالی، تقارن نقش مهمی در ساختار برخی مدل‌های معروف نظیر مدل قیمت گذاری دارایی‌های سرمایه‌ای شارپ–لینتر و مدل قیمت گذاری اختیار معامله بلک–شولز ایفا می‌کند. از طرفی بسیاری از روش‌های پارامتری و استوار آماری تا حدی به فرض متقارن بودن وابسته هستند نظیر فرض‌های مربوط به توزیع خطاها در تحلیل رگرسیون و تحلیل مدل آنالیز واریانس. با توجه به دلایل گفته شده، تاکنون آزمون‌های متعددی برای آزمون فرضیه H_0 در معادله (۱.۱) در مقابل فرضیه

$$H_1 : f(\mu - x) \neq f(\mu + x), \qquad \text{for some } x \in \mathbb{R},$$

[۱] سختران،

معرفی و پیشنهاد شده است. محققینی چون رندلز و همکاران (۱۹۸۰)، ام سی ویلیامز (۱۹۹۰)، مدرس و گستویرت (۱۹۹۰)، گیبنز و چاکرابرتی (۱۹۹۲)، تاجدین (۱۹۹۴)، (باکلیزی ، ۲۰۰۳، ۲۰۰۷ و ۲۰۰۸)، چنگ و بالاکریشنان (۲۰۰۴)، کورزو و باباتیوا (۲۰۱۳) و شیانگ و همکاران (۲۰۲۱)، به مسئله فرضیه متقارن بودن یک توزیع در حضور داده‌های کامل پرداخته‌اند. با توجه به اطلاعات موجود، مطالعات چندانی در خصوص آزمون متقارن بودن داده‌ها در حضور داده‌های غیر کامل وجود ندارد. اکبری و زمانی (۱۴۰۳) در حضور داده‌های سانسور شده از راست، آزمون تقارن برای یک توزیع پیوسته پیشنهاد کردند. در این مقاله به معرفی یک آزمون جدید برای تقارن یک توزیع پیوسته با وجود داده‌های بریده شده از چپ و سانسور شده از راست می‌پردازیم. بدین منظور ساختار مقاله بدین شکل است. در بخش ۲ به معرفی آماره آزمون پرداخته می‌شود. با توجه به عدم امکان تعیین توزیع دقیق آماره پیشنهادی، در بخش ۳ مقادیر بحرانی آماره آزمون از طریق روش مونت کارلو تقریب زده شده است. علاوه‌براین با استفاده از روش ذکر شده، توان آزمون نیز برآورد می‌شود. در نهایت در بخش ۴ تحت یک مثال واقعی با داده‌های از چپ بریده شده و از راست سانسور شده، نحوه به کار بردن آزمون معرفی شده، توضیح داده می‌شود.

۲ آماره آزمون

فرض کنید $X_۱$، $X_۲$،...، X_n یک نمونه تصادفی از توزیع پیوسته F با میانه نامعلوم θ باشد. اگر i، j و k سه مقدار مجزا متعلق به مجموعه $\{۱, ۲, \ldots, n\}$ باشد، در اینصورت مشاهدات سه تایی (X_i, X_j, X_k) تشکیل یک سه تایی راستگرا [1] را می‌دهند، بدین معنا که توزیع بنظر می‌رسد که به راست چوله است اگر مشاهده میانی به کوچکترین مشاهده نزیک‌تر باشد. بطور مشخص‌تر فرض کنید

$$\begin{aligned} f^*(X_i, X_j, X_k) &= \frac{۱}{۳}\{sign(X_i + X_j - ۲X_k) \\ &\quad + sign(X_i + X_k - ۲X_j) + sign(X_j + X_k - ۲X_i)\}, \end{aligned} \tag{۱.۲}$$

که در آن $sign(u) = -۱, ۰, ۱$ به ترتیب برای مقادیر $u < ۰$، $u = ۰$ و $u > ۰$. با توجه به معادله (۱.۲)، (X_i, X_j, X_k) تشکیل یک سه تایی راست گرا را می‌دهد اگر $f^*(X_i, X_j, X_k) = \frac{۱}{۳}$ و بطور مشابه تشکیل یک سه تایی چپ گرا را می‌دهد اگر $f^*(X_i, X_j, X_k) = -\frac{۱}{۳}$ و زمانی که $f^*(X_i, X_j, X_k) = ۰$ باشد بیانگر آن است که سه تایی نه راست گرا و نه چپ گراست. بر همین اساس رندلز و همکاران (۱۹۸۰) یک آزمون بر اساس Uـآماره

$$\hat{\eta} = \frac{۱}{\binom{n}{۳}} \sum_{i<j<k} f^*(X_i, X_j, X_k), \tag{۲.۲}$$

پیشنهاد کردند که مقادیر بزرگ $|\hat{\eta}|$ منجر به رد فرضیه تقارن می‌شود. در ادامه نحوه لحاظ کردن مشاهدات از چپ بریده شده و از راست سانسور شده را در آماره آزمون (۲.۲) مورد بررسی قرار می‌دهیم.

فرض کنید که X، C و L به ترتیب با تابع توزیع‌های F، G و H بیانگر متغیرهای تصادفی طول عمر، زمان سانسور و زمان برش از چپ باشند و $\bar{F}(x) = ۱ - F(x)$، $\bar{G}(x) = ۱ - G(x)$ و $\bar{H}(x) = ۱ - H(x)$. تحت سانسور از راست و برش از چپ، n نمونه مستقل و هم توزیع $(T_i\epsilon_i, \delta_i)$ از $(T\epsilon, \delta)$ در نظر گرفته می‌شود که $T = \min(X, C)$، $\delta = I(X < C)$ و $\epsilon = I(T > L)$. به عبارت دیگر، δ نشانه سانسور از راست و ϵ برای مشخص کردن برش استفاده شده است. واضح است که T_i یک مشاهده برای iمین

[1] right triple

نمونه است اگر $T_i > L_i$ باشد. بر اساس داتا (۲۰۱۰) می‌توان براساس داده‌های از راست سانسور شده و از چپ بریده شده، یک نسخه از آماره (۲.۲) بصورت زیر در نظر گرفت.

$$\hat{\eta}_{LTRC} = \frac{۱}{\binom{n}{۳}} \sum_{۱\le i<j<k\le n} \frac{f^*(T_i,T_j,T_k)\delta_i\delta_j\delta_k}{\hat{K}_C(T_i)\hat{K}_C(T_j)\hat{K}_C(T_k)},$$

که در آن $\hat{K}_C(\tau)$ برآوردگری از تابع بقا زمان سانسور C تحت برش چپ است و بصورت زیر می‌باشد.

$$\hat{K}_C(\tau) = \prod_{t\le\tau}\left(۱ - \frac{dN^C(t)}{Y(t)}\right),$$

که در آن

$$\begin{aligned} dN_i^C(t) &= I(T_i = t, \delta_i = ۰), \\ Y(t) &= \sum_{i=۱}^{n} Y_i(t), \qquad \text{and} \quad Y_i(t) = I(T_i \ge t \ge L_i). \end{aligned} \tag{۳.۲}$$

فرض $H_۰$ یعنی متقارن بودن توزیع پیوسته، به ازای مقادیر بزرگ $|\hat{\eta}_{LTRC}|$ رد می‌شود. تحت شرایطی می‌توان نشان داد که آماره آزمون $\hat{\eta}_{LTRC}$ مجانباً نرمال است.

۳ مقادیر بحرانی و توان آزمون

در این بخش، با استفاده از نتایج شبیه‌سازی مونت کارلو، مقادیر بحرانی را به ازای $n = ۲۰, ۵۰, ۷۰$ بدست آورده بطوری که دو مقدار $p = ۰/۱, ۰/۴$ برای سانسور از راست یعنی $P(X > C) = p$ زمانی که $C \sim Exp(\lambda)$ و مستقل از X است در نظر گرفته شده است و علاوه‌براین توزیع متغیر تصادفی برش از چپ بصورت، $L \sim Exp(\mu)$ فرض شده است به قسمی که $P(L > X) = ۰/۲$ گردد. مقادیر بحرانی در سطح خطای نوع اول، $\alpha = ۰/۰۵$، بصورت زیر در جدول ۱ گزارش شده است. برای بررسی توان آزمون

جدول ۱: مقادیر بحرانی آزمون $\hat{\eta}_{LTRC}$

p	n		
	۲۰	۵۰	۷۰
۰/۱	۰/۰۶۸	۰/۰۳۹	۰/۰۳۲
۰/۴	۰/۰۳۵	۰/۰۲۳	۰/۰۱۸

پیشنهاد شده، توزیع‌های مقابل از خانواده توزیع لامبدای تعمیم‌یافته (GLD[۲]) با تابع چندک زیر در نظر گرفته شده است.

$$F^{-۱}(u) = \lambda_۱ + \frac{u^{\lambda_۳} - (۱-u)^{\lambda_۴}}{\lambda_۲}, \quad ۰ < u < ۱.$$

در جدول ۲، ۹ حالت از GLD به ازای پارامترهای مختلف آن بعنوان ۹ توزیع مقابل نشان داده شده است. جدول ۳ توان آزمون محاسبه شده $\hat{\eta}_{LTRC}$ را گزارش می‌دهد. همان‌طور که مشاهده می‌شود به ازای افزایش حجم نمونه، توان آزمون‌ها افزایش یافته و توان‌ها زمانی که درجه سانسور کوچک است ($P = ۰/۱$) به نسبت بزرگتر می‌باشد. در مجموع توان آزمون قابل قبول بوده و در اکثر موارد جز در توزیع مقابل حالت دوم، توان خوبی بدست آمده است.

[2] generalized lambda distribution

جدول ۲: ۹ حالت از GLD

$Case$	λ_1	λ_2	λ_3	λ_4
۱	۰٫۰۰۰۰	۰٫۱۹۷۵	۰٫۱۳۴۹	۰٫۰۱۳۴۹
۲	−۰٫۱۱۶۷	−۰٫۳۵۱۷	−۰٫۱۳۰۰	−۰٫۱۶۰۰
۳	۰٫۰۰۰۰	−۱٫۰۰۰۰	−۰٫۱۰۰۰	−۰٫۱۸۰۰
۴	۳٫۵۸۶۵	۰٫۰۴۳۱	۰٫۰۲۵۲	۰٫۰۹۴۰
۵	۰٫۰۰۰۰	−۱٫۰۰۰۰	−۰٫۰۰۷۵	۰۳۰۰.۰−
۶	۰٫۰۰۰۰	۱٫۰۰۰۰	۱٫۴۰۰۰	۰٫۰۲۵۰۰
۷	۰٫۰۰۰۰	۱٫۰۰۰۰	۰٫۰۰۰۱	۰٫۱۰۰۰
۸	۰٫۰۰۰۰	−۱٫۰۰۰۰	−۰٫۰۰۱۰	−۰٫۱۳۰۰
۹	۰٫۰۰۰۰	−۱٫۰۰۰۰	−۰٫۰۰۰۱	−۰٫۱۷۰۰

از طرفی حالت اول از توزیع مقابل، همان توزیع نرمال استاندارد است لذا توان آزمون در این مورد، همان خطای نوع اول را نشان می‌دهد که مقادیر گزارش شده حول مقدار ۰٫۰۵ همان خطای نوع اول اسمی در نظر گرفته شده است. بنابراین می‌توان گفت برآوردهای خطای نوع اول نیز تحت کنترل و قابل قبول است.

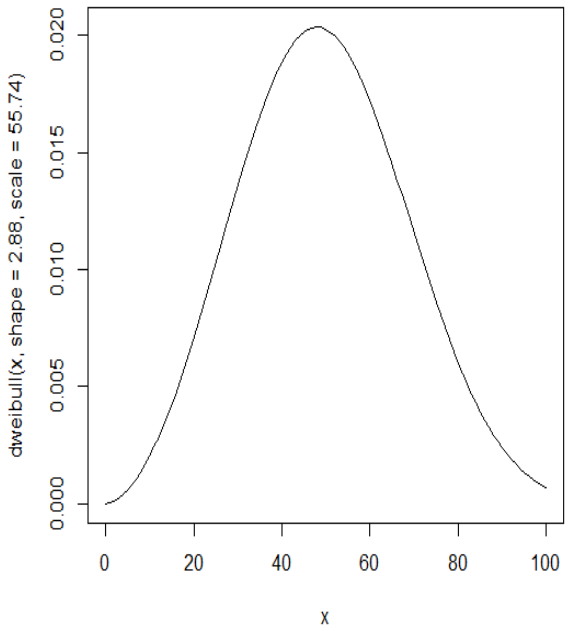
جدول ۳: توان آزمون

n	$Case$۱	$Case$۲	$Case$۳	$Case$۴	$Case$۵	$Case$۶	$Case$۷	$Case$۸	$Case$۹
$p = ۰٫۱$									
۲۰	۰٫۰۵۰۴	۰٫۰۷۸	۰٫۱۰۶۶	۰٫۱۷۰۴	۰٫۲۵۶۲	۰٫۲۱۶۴	۰٫۵۸۳۴	۰٫۹۰۰۸	۰٫۶۳۴۸
۵۰	−۰٫۰۴۴۸	−۰٫۰۶۹	−۰٫۲۰۰۴	−۰٫۴۳۹۸	۰٫۶۲۶۸	۰٫۵۱۶	۰٫۹۶۹۷	۰٫۹۹۹۶	۰٫۹۸۶
۷۰	۰٫۰۶۳۶	۰٫۰۸۱	۰٫۳۱۱	۰٫۵۹۵۸	۰٫۷۵۸۸	۰٫۶۵۵	۰٫۹۸۶۸	۰٫۹۹۹۸	۰٫۹۹۵
$p = ۰٫۴$									
۲۰	۰٫۰۴۹۴	۰٫۱۲۷۲	۰٫۰۴۴۸	۰٫۰۵۷۶	۰٫۰۵	۰٫۳۳۰۶	۰٫۲۲۰۴	۰٫۲۴۲۶	۰٫۲۴۷۶
۵۰	۰٫۰۵۴	۰٫۲۰۸۸	۰٫۰۲۱۶	۰٫۰۵۷۶	۰٫۰۳۶۸	۰٫۴۶۲۸	۰٫۴۵۶۲	۰٫۵۰۳۲	۰٫۵۲۹۲
۷۰	۰٫۰۹۷۶	۰٫۳۶۹۴	۰٫۰۳۸۶	۰٫۱۰۸۲	۰٫۰۶۲	۰٫۵۹۱۸	۰٫۶۷۵۴	۰٫۷۳۳۶	۰٫۷۶۸۴

۴ مثال واقعی

برای تشریح نحوه بکار بردن آزمون پیشنهادی، داده‌های از چپ بریده شده و از راست سانسور شده هانگ و همکاران (۲۰۰۹) را بکار می‌بریم که این مجموعه داده شامل طول عمر ۷۱۰ ترانسفورماتور یک شرکت انرژی با ۶۲ شکست است. اگرچه داده‌های اصلی موجود نیست اما یک زیر مجموعه ۲۸۶تایی با ۳۹ شکست در امورا و شیو (۲۰۱۶) آمده است. داده‌ها بین سال‌های ۱۹۸۰ الی ۲۰۰۸ جمع‌آوری شده است. ازاین‌رو ترانسفورماتورهایی که قبل از ۱۹۸۰ نصب شده‌اند بعنوان داده‌های از چپ بریده شده در نظر گرفته می‌شوند. بنابراین داده‌ها از چپ بریده شده و از راست سانسور شده است. نرخ سانسور راست و برش از چپ به ترتیب ۸۶٫۴ و ۵۸٫۳۹ درصد هستند. هانگ و همکاران (۲۰۰۹) یک توزیع وایبل با پارامتر شکل ۲٫۸۸ و پارامتر مقیاس ۵۵٫۷۴ به این مجموعه داده‌ها برازش داده‌اند. نمودار چگالی این توزیع در شکل زیر آمده است و به نوعی نشان می‌دهد که توزیع مجموعه داده موردنظر، متقارن است.

آماره آزمون پیشنهادی مشاهده شده بصورت $\hat{\eta}_{LTRC} = ۰/۰۶۲۵$ بدست آمده است و برای بدست آوردن p–مقدار یک نمونه به حجم



شکل ۱: چگالی توزیع وایبل

$n = ۵۰$ ازتوزیع وایبل با پارامترهای ذکر شده تولید کرده و $L \sim Exp(\mu)$ و $C \sim Exp(\lambda)$ در نظر می‌گیریم. مقادیر μ و λ را طوری تعیین می‌کنیم که به نرخ‌های به ترتیب ۵۸/۲۹ و ۸۶/۴ منتج شود. بر اساس هر تکرار یک مقدار از آماره $\hat{\eta}_{LTRC}$ بدست خواهد آمد و براساس ۵۰۰۰ تکرار، میانگین تعداد دفعاتی که $\hat{\eta}_{LTRC}$ بیشتر از آماره آزمون مشاهده شده ۰/۰۶۲۵ باشد را به عنوان p–مقدار در نظر خواهیم گرفت. با توجه به مراحل مذکور، p–مقدار بدست آمده برابر با ۰/۷ بدست آمد که حاکی از آن است که توزیع مجموعه داده موردنظر متقارن است.

مراجع

اکبری، م. و زمانی، ز. (۱۴۰۳)، معرفی آزمون جدید برای متقارن بودن یک توزیع در حضور داده‌های سانسور شده، هشتمین همایش ریاضیات و علوم انسانی ریاضیات مالی، ۲۳ و ۲۴ آبان۱۴۰۳، دانشگاه علامه طباطبائی، تهران.

Baklizi A. (2003), A conditional distribution free runs test for symmetry, *Journal of Nonparametric Statistics*, **15**, 713–718.

Baklizi A. (2007), Testing symmetry using a trimmed longest run statistics, *Australian and New Zealand Journal of Statistics*, **49**, 339–347.

Baklizi A. (2008), Improving the power of the Hybrid test, *International Journal of Contemporary Mathematical Sciences*, **3**, 497–499.

Cheng W. and Balakrishnan N. (2004), A modified sign test for symmetry, *Communications in Statistics-Simulation and Computation*, **33**, 703–709.

Corzo J. and Babativa G. (2013), A modified runs test for symmetry, *Journal of Statistical Computation and Simulation*, **83**, 984–991.

Datta S., Bandyopadhyay D., and Satten, G. A. (2010), Inverse probability of censoring weighted U-statistics for right-censored data with an application to testing testing hypotheses, *Scandinavian Journal of Statistics* , **37**, 680–700.

Emura T., and Shiu v. (2016), Estimation and model selection for left-truncated and right-censored lifetime data with application to electric power transformers analysis, Communication in Statisttics: Simulation and Computation, **45**, 171–318.

Gibbons C., and Chakraborti S. (1992), *Nonparametric Statistical Inrefence*, New York: Marcel Dekker.

Hong Y., Meeker W.Q. and McCalley J.D. (2009), Prediction of remaining life of power transformers based on left truncated and right censored lifetime data, Annals of Applied Statistics,**3** 3, 857-–879.

McWillials P. (1990), A distribution-free test for symmetry based on a runs statistic, *Journal of the American Statistical Association*, **85**, 1130–1133.

Modarres R. and Gastwirth J. (1998), Hybrid test for the hypothesis of symmetry, *Journal of Applied Statistics*, **25**, 777-783.

Randles R. H., Fligner M. A., Policello ll G. E., and Wolfe D. A. (1980), An asymptotically distribution-free test for symmetry versus asymmetry, *Journal of the American Statistical Association*, **75**, 168-172.

Tajuddin I. (1994), Distribution-free test for symmetry based on Wilcoxon two-sample test, *Journal of Applies Statistics*, **21**, 409–415.

Xiong P., Zhuang W., and Qiu G. (2021), Testing symmetry based on the extropy of record values, *Journal of Nonparametric Statistics*, **33**, 134–155.

A Symmetry Goodness-of-Fit Test for Continuous Distributions under Left Truncation and Right Censoring

Abstract: In this paper, a test statistic is first introduced for assessing the symmetry of continuous distributions under right-censored and left-truncated data. It is shown that the proposed test statistic is asymptotically normal. In an extensive simulation study, the power of the proposed test is examined. The results indicate that the test performs satisfactorily. Finally, the applicability of the test statistic is demonstrated through a real data example.

Keywords: Goodness-of-fit test, Symmetric distribution, Left-truncated and right-censored data.

Mathematics Subject Classification (2020): 62N01, 62G10.