

تقریبهایی برای تابع توزیع آماره آزمون اندرسون-دارلینگ

سی. دی. سینکلر بی. دی. اسپار
مترجم: حمزه ترابی^۱

چکیده

در این مقاله یک تقریب نظری برای مساحت دم بالایی آماره آزمون اندرسون-دارلینگ با بکارگیری نتیجه‌ای از «زولوتاریف^۲» (۱۹۶۱) برای تابع مشخصه استخراج شده است. این تقریب، به ویژه برای دم بالایی سودمند است. یک توزیع تجربی نیز برای این آماره ارایه شده است که در سراسر دامنه توزیع دقیق است. واژه‌های کلیدی: تابع مشخصه، کومولان، تقریب تجربی و تقریب نظری.

۱ مقدمه

آزمون اندرسون-دارلینگ، یکی از پرتوان‌ترین و مهمترین آزمونهای نیکویی برازش در ادبیات آماری است. اگرچه این آزمون به گونه‌ای طراحی شده است که در مورد اختلاف بین دمه‌های توابع توزیع نظری و تجربی حساستر باشد، در موقعیتهای دیگر نیز خوب عمل می‌کند. (برای جزئیات بیشتر و نیز لیست کاملی از مراجع، کتاب داگوستینو^۳ و استفنز^۴ (۱۹۸۶، فصل چهار) را ببینید). در بخش ۲ چهار کومولان اول، چاولگی و کشیدگی توزیع مجانبی آماره آزمون اندرسون-دارلینگ ارایه می‌شود. تقریب اخیر که با مقادیر

۲ کومولانهای A_n^2

آماره آزمون اندرسون-دارلینگ، A_n^2 ، به صورت زیر تعریف می‌شود:

$$A_n^2 = n \int_{-\infty}^{\infty} \frac{[F_n(x) - F(x)]^2}{F(x)[1 - F(x)]} dF(x) \quad (1)$$

^۱ حمزه ترابی، دانشکده ریاضی-دانشگاه یزد
^۲ Zolotarev
^۳ D'Agostino
^۴ Stephens

این است که اگر λ_j ها مثبت و نزولی باشند و $\sum_{j=1}^{\infty} \lambda_j < \infty$ ، آنگاه احتمالهای دم بالایی متغیر تصادفی Z_j^2 که در آن $Z_j \sim N(0, 1)$ ، از رابطه زیر بدست می آید:

$$1 - F(x) = \frac{\prod_{j=2}^{\infty} \left[1 - \frac{\lambda_j}{\lambda_1}\right]^{-\frac{1}{j}}}{\Gamma\left(\frac{1}{j}\right)} \times \left[\frac{x}{2\lambda_1}\right]^{-\frac{1}{j}} \exp\left[\frac{-x}{2\lambda_1}\right] [1 + \epsilon(x)] \quad (7)$$

که در آن هنگامی که $x \rightarrow \infty$ ، داریم $\epsilon(x) \rightarrow 0$.

λ_j های داده شده از رابطه (۶)، به وضوح مثبت و نزولی هستند و $\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} 1/j(j+1) = 1$ پس در این حالت (۷) به صورت زیر خلاصه می شود:

$$1 - F(x) \simeq \frac{\prod_{j=2}^{\infty} \left[1 - \frac{2}{j(j+1)}\right]^{-\frac{1}{j}}}{\Gamma\left(\frac{1}{j}\right)} \times [x]^{-\frac{1}{j}} \exp[-x] [1 + \epsilon(x)] \quad (8)$$

جمله ثابت تا چهار رقم با معنا برابر است با:

$$K = \prod_{j=2}^{\infty} \left[1 - \frac{2}{j(j+1)}\right]^{-\frac{1}{j}} = 1/732 \quad (9)$$

بنابراین معادله (۸) به صورت زیر در می آید:

$$1 - F(x) \simeq \left[\frac{1/732}{\Gamma\left(\frac{1}{j}\right)}\right] \times [x]^{-\frac{1}{j}} \exp[-x] [1 + \epsilon(x)] \quad (10)$$

اگر در (۹) قرار دهیم: $\epsilon(x) = 0$ ، نتیجه تقریبی بدست آمده در زیر، برای مقادیر بالای میانه با نقاط درصدی A^2 که توسط لوییس (۱۹۶۱) ارائه شده است، سازگاری خوبی دارد. این تقریب، به ویژه برای دم بالایی توزیع مفید است. جدول ۱، مساحت‌های دم بدست آمده از (۹) را با سطحهای اسمی مقادیر بحرانی به دست آمده از راه درونبایی توسط لوییس (۱۹۶۱) مقایسه می کند.

که در آن $F_n(x)$ تابع توزیع تجربی نمونه‌ای شامل n مشاهده و $F(x)$ تابع توزیع پیوسته مفروضی است. دو کومولان اول A^2 (حالت مجانبی A_n^2) کاملاً معلوم است، ولی به نظر نمی رسد که دو کومولان سوم و چهارم به سادگی دستیافتنی باشند.

تابع مشخصه A^2 از رابطه زیر بدست می آید (اندرسون و دارلینگ، ۱۹۵۲):

$$\Phi(t) = \left[(-2\pi it) / \cos(\pi/2 \sqrt{1 + 8it})\right]^{\frac{1}{j}} \quad (2)$$

اکنون با مشتقگیری از $\Phi(t)$ یا بررسی ضرایب $\log \Phi(t)$ چهار کومولان اول توزیع عبارتند از:

$$\begin{aligned} k_1 &= 1 \\ k_2 &= \frac{2}{3}(\pi^2 - 9) = 0/579726 \\ k_3 &= 80 - 8\pi^2 = 1/04329 \\ k_4 &= \frac{16}{15}\pi^4 + 160\pi^2 - 1/680 = 3/03699 \end{aligned} \quad (3)$$

همچنین ضریب مجذور شده چاولگی (β_1) و ضریب کشیدگی (β_2) ، به صورت زیر بدست می آیند:

$$\beta_1 = 5/5865 \quad \beta_2 = 12/036 \quad (4)$$

بنابراین، با توجه به دستگاه منحنی‌های پیرسون و جانسون، توزیع A^2 بین توزیع گاما و لاگ-نرمال قرار می گیرد.

۳ تقریب کردن مساحت دم بالایی

تابع مشخصه (۲) را می توان به صورت زیر نیز نوشت:

$$\Phi(t) = \prod_{j=1}^{\infty} \left(1 - \frac{2it}{j(j+1)}\right)^{-\frac{1}{j}} \quad (5)$$

(کتاب اندرسون و دارلینگ، ۱۹۵۲، ص ۲۰۴ را ببینید). این تابع، با تابع مشخصه یک مجموع وزنی نامتناهی از متغیرهای تصادفی مستقل خی-دو با یک درجه آزادی با وزنهای

$$\lambda_j = 1/j(j+1) \quad (6)$$

متناظر است. بنابراین، نتایج زولوتارف (۱۹۶۱) را می توان درباره توزیعهای حدی مربعی فرم به کار بست. یک نتیجه کلی

آمده‌اند با مقادیری که توسط لوییس (۱۹۶۱) داده شده‌اند، مقایسه می‌کنند.

مقدار بحرانی	سطح معنی دار اسمی	مساحت دمی از از رابطه (۹)
۰/۲۸۳	۰/۹۵	۱/۳۸۴۱
۰/۳۴۶	۰/۹۰	۱/۱۷۵۴
۰/۳۹۹	۰/۸۵	۱/۰۳۸۰
۰/۷۷۴	۰/۵۰	۰/۵۱۲۲
۱/۲۴۹	۰/۲۵	۰/۲۵۰۸
۱/۹۳۳	۰/۱۰	۰/۱۰۱۷
۲/۴۹۲	۰/۰۵	۰/۰۵۱۲
۳/۰۷۹	۰/۰۲۵	۰/۰۲۵۶
۳/۸۸۰	۰/۰۱	۰/۰۱۰

۴ نتیجه

از جدول ۱ متوجه می‌شویم که تقریب (۹) به طور سازگاری، مساحت دمی را زیاد برآورد می‌کند. با این وجود، محاسبه از این رابطه آسان و تنها شامل یک ضریب است. ملاحظات نظری بر آن تأکید دارد که این رابطه، در مورد دمهای فرین به طور فزاینده‌ای دقیق می‌شود. رابطه دیگری که برای دمهای فرین قابل استفاده باشد نمی‌شناسیم.

مقایسه بین خطاهای مطلق و نسبی آشکار می‌کند که رابطه (۱۱) برای مقادیر سمت راست میانه، دقیقتر از رابطه (۹) است. به علاوه (۱۱) برای مقادیر سمت چپ میانه نیز دقیق است. این رابطه شامل ۵ ضریب است.

تأکید می‌کنیم که برای مقاصد عملی، رابطه (۹) برای آزمون نیکویی برازش یک مجموعه داده‌ای منفرد، به اندازه کافی دقیق است در حالی که (۱۱) متممی سودمند برای جدول لوییس است و می‌تواند در تشخیص نیکویی برازش مرکب از چندین مجموعه داده‌ای مفید باشد.

جدول ۱ - مقایسه‌های مساحت دمی با استفاده از تقریب نظری (۹)

تقریب قبلی در مورد دم پایینی ضعیف است. تقریب تجربی زیر در سراسر برد مناسب و قابل قبول است (جدول ۲ و ۳ را ببینید).

$$F(x) = 1 - \left[1 + \exp \left(\frac{0.9936x + 0.3287}{x} + \frac{1.784}{x} - \left(\frac{2.018 + 0.2029}{x} \right) x^{-\frac{1}{3}} \right) \right]^{-1} \quad (11)$$

این تقریب از راه برازش تعمیمی از توزیع لجستیک، برای مقادیر داده شده توسط لوییس (۱۹۶۱) به دست آمده است و چون برای سه نقطه ۰/۶۵، ۰/۷، ۰/۸ دقت تقریب، با نتایج شبیه‌سازی با تردید مواجه می‌شود، این سه مقدار حذف شده‌اند. جداول ۲ و ۳ مساحت‌های دمی را که از رابطه (۱۱) بدست

	x					
	۱/۲۵	۱/۶۵	۱/۹۵	۲/۵۰	۳/۰۵	۳/۸۵
$P = 1 - F(x)$	۰/۲۴۹۷	۰/۱۴۴۴	۰/۰۹۷۹	۰/۰۴۹۶	۰/۰۲۵۸	۰/۰۱۰۳
\hat{P} از رابطه (۱۱)	۰/۲۴۹۳	۰/۱۴۴۷	۰/۰۹۸۱	۰/۰۴۹۶	۰/۰۲۵۸	۰/۰۱۰۳
$100(\hat{P} - P)/P$	-۰/۱	۰/۲	۰/۲	۰/۱	۰	-۰/۳

جدول ۲ - مقایسه مساحت دم بالایی با استفاده از تقریب تجربی (۱۱)

	x				
	۰/۵۰	۰/۴۰	۰/۳۵	۰/۲۷۵	۰/۲
$P = F(x)$	۰/۲۵۳۲	۰/۱۵۱۳	۰/۱۰۳۶	۰/۰۴۴۳	۰/۰۰۹۶
\hat{P} از رابطه (۱۱)	۰/۲۵۳۳	۰/۱۵۰۷	۰/۱۰۲۹	۰/۰۴۴۰	۰/۰۰۹۶
$100(\hat{P} - P)/P$	۰	-۰/۴	-۰/۷	-۰/۷	۰/۳

جدول ۳ - مقایسه مساحت دم پایینی با استفاده از تقریب تجربی (۱۱)

- [1] Anderson, T. W. and Darling, D. A. (1952), "Asymptotic Theory of Certain Goodness of Fit Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics*, 23, 193 - 212.
- [2] D'Agostino, R. B., and Stephens, M. A. (eds.) (1986), *Goodness of Fit Techniques*, New York: Marcel Dekker.
- [3] Lewis, P. A. W. (1961), "Distribution of the Anderson - Darling Statistic," *Annals of Mathematical Statistics*, 32, 1118-1124.
- [4] Zolotarev, V. M. (1961), "Concerning a Certain Probability Problem," *Theory of Probability and Its Applications*, 6, 201 - 204.

این مقاله ترجمه

Approximations to the Distribution Function of
the Anderson - Darling Test Statistic

نوشته B. D. SPURR و C. D. SINCLAIR است.
