گسترۀ علوم آماری
سال اول - شماره دوم (ص 39-48)

Statistical Sciences Extension
Year 1 - Number 2 (P 39-48)

روش نمونه‌گیری ترابرشی خطی: تابع تشخیص و آزمون فرضیه پارامتری

ابوذر بازیاری *
استادیار، آمار، دانشگاه خلیج فارس، بوشهر

تاریخ دریافت: 1395/04/14 **تاریخ پذیرش:** 1395/06/31

Line Transect Sampling: Detection Function and Parametric Hypothesis Testing

A. Bazyari*

Assistant Professor, Statistics, Persian Gulf University, Bushehr, Iran

Received: 2016/07/04 **Accepted:** 2016/09/21

Abstract

Line transect sampling is a very helpful method for estimating the density function of population in biology. In this paper, first the line transect sampling is introduced and then parametric hypothesis testing for half normal density function against univariate exponential density function in line transect sampling was considered.

The test statistics is obtained using likelihood ratio method. Computing the null distribution of test statistic and its critical values is not easy because of the complexity of test statistic structure, therefore Monte carlo simulation was used for finding the critical values of test statistic at different significance levels. This problem of testing was investigated with numerical examples.

Keywords

Parametric Hypothesis Testing, Half Normal Density Function, Monte Carlo Simulation, Line Transect Sampling.

چکیده

نمونه‌گیری ترابرشی خطی یک روش بسیار مفید برای برآورد تابع چگالی جمعیت در علم زیست شناسی است. در این مقاله، ابتدا به معرفی روش نمونه‌گیری ترابرشی خطی پرداخته شده و سپس آزمون فرضیه پارامتری برای تابع چگالی نیم نرمال در مقابل تابع چگالی نمایی یک متغیره در روش نمونه‌گیری ترابرشی خطی در نظر گرفته شده است. آماره آزمون با استفاده از روش نسبت درست‌نمایی محاسبه شده است. به دلیل ساختار پیچیده آماره آزمون، محاسبه توزیع آن تحت فرضیه صفر و نیز تعیین مقادیر بحرانی آن کار ساده‌ای نخواهد بود، بنابراین از روش شبیه سازی مونت کارلو برای یافتن مقادیر بحرانی آماره آزمون در سطوح مختلف معناداری استفاده شده است. با مثال‌های عددی این مسئله آزمون مورد بررسی قرار گرفته است.

واژگان کلیدی

آزمون فرضیه پارامتری، تابع چگالی نیم نرمال، شبیه‌سازی مونت کالو، نمونه‌گیری ترابرشی خطی.

*** نویسنده مسئول:** ابوذر بازیاری

*Corresponding Author: ab_bazyari@yahoo.com

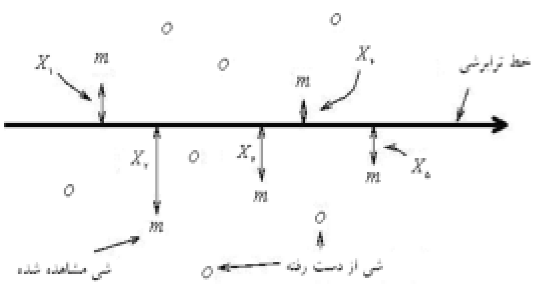
مقدمه

روش نمونه‌گیری ترابرشی[1] یک تکنیک نمونه‌گیری است که از سال 1930 تا به حال معمولاً برای برآورد تابع چگالی جمعیت پرندگان و حیوانات به کار می‌رود. زیست‌شناسان و محققان علوم دریایی بیشتر، این روش نمونه‌گیری را به کار می‌گیرند. در این روش برای برآورد کردن تابع چگالی جمعیت، آزمایش‌گر روی خطی مستقیم در حرکت است و موقعیت عمودی اشیای مشاهده شده را تا خط اندازه‌گیری می‌کند. یک نوار مستطیلی نازک بلند مخصوصی را که دارای طول و پهنای معین باشد، در نظر بگیرید به این نوار، نوار ترابرشی گویند. فرض کنید که اشیا در این نوار در حال حرکت باشند و نیز هر چیزی که در نوار حرکت کند، قابل دیدن و ثبت کردن است. آزمایشگر از ابتدای نوار روی مرکز نوار به طرف انتهای نوار برای دیدن اشیا (اشیا می‌توانند در سمت چپ یا راست مرکز نوار قرار گیرند) حرکت می‌کند. البته ممکن است نتواند هر چیزی را که در این نوار حرکت می‌کند، مشاهده کند. در حقیقت احتمال دیدن یک چیز بستگی به مسافت آن چیز از مرکز نوار دارد. یعنی هر چقدر که مسافت چیز از مرکز کمتر باشد احتمال دیدن آن چیز زیادتر است و برعکس. به مثال زیر توجه کنید.

مثال 1: اگر هدف برآورد تعداد ماهی‌های یک دریاچه باشد، از روش نمونه‌گیری ترابرشی برای این منظور می‌توان استفاده کرد. در این روش به وسیلهٔ هواپیما یا کشتی در مسیرهایی که کاملاً به تصادف انتخاب شده‌اند، نمونه‌گیری صورت می‌گیرد و داده‌ها به وسیلهٔ افراد با تجربه یا دستگاه‌ها (یا هر دو) ثبت می‌شوند. به این صورت که افراد به جستجو در اطراف این مسیر (با هواپیما یا کشتی) می‌پردازند و به محض مشاهده یک ماهی فاصله آن را از مرکز نوار به طور دقیق اندازه‌گیری می‌کنند.

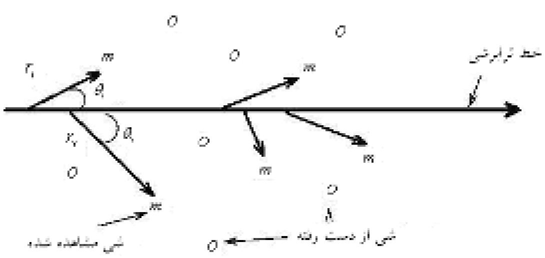
حال یک مجموعه از نوارها (t نوار) که طول هر کدام معلوم و برابر با l_i ، $i = 1, 2, \ldots, t$ و پهنای هر یک برابر $2w$ باشد را در نظر بگیرید، نوارها طوری قرار گرفته‌اند که روی هم نمی‌افتند. (این طرز قرار گرفتن نوارها را در روش نمونه‌گیری، نمونه‌گیری بدون جای‌گذاری گویند. اگر نوارها روی هم قرار نگیرند، در نمونه‌گیری به آن نمونه‌گیری با جای‌گذاری گویند.)

فرض کنید افراد متخصص روی نوار در حال حرکت باشند، هر شیء مشاهده شده مسافت عمودی X_i ، فاصله شیء تا نوار، را طی کند. در شکل 1 مسیر حرکت اشیاء به خوبی نشان داده شده است.



شکل 1. مسیر حرکت اشیاء در روش نمونه‌گیری ترابرشی خطی

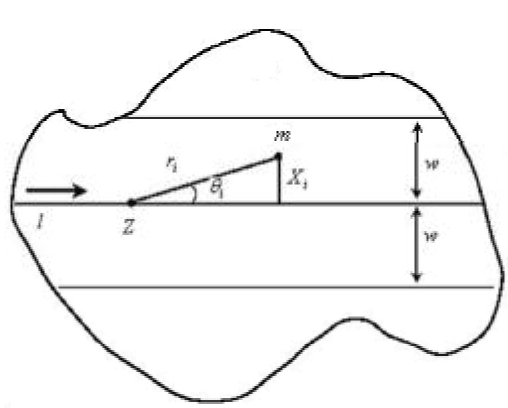
که در آن m شیء دیده شده و X_i مسافت عمودی iامین مشاهده است. همچنین هر شیء مشاهده شده مسافت r_i که عبارت است از فاصله تا مشاهده کننده را ثبت و زاویه q_i را به وجود می‌آورد. این مسیرها در شکل 2 نشان داده شده است.



شکل 2. مسیر حرکت اشیای مشاهده شده

به همین ترتیب می‌توان شکل 3 را برای مسیر حرکت اشیا در نظر گرفت.

1. Line transect sampling



شکل 3. مسیر حرکت اشیاء

که در آن l طول خط ترابرشی، Z موقعیت مشاهده‌گر، w نصف طول نوار، r_i فاصله شیء تا مشاهده‌گر و q_i زاویه ایجاد شده به وسیلهٔ موقعیت آزمایش‌گر با شیء مشاهده شده است.

فرض کنید طول همه نوارها برابر با عدد معلوم l و طبق قبل پهنای هر یک $2w$ باشد. اگر برای هر $0 £ x £ w$، $f(x)$ تابع چگالی متغیرهای تصادفی $X_1, X_2, \mathbf{K}, X_n$ در نمونه‌گیری ترابرشی خطی و D تابع چگالی جمعیت (برای t نوار) باشد، ابرهارت (1978) نشان داد که برآورد D عبارت است از:

$$\hat{D} = \frac{n\hat{f}(0)}{2tl}, \tag{1}$$

که در آن n تعداد اشیای مشاهده شده به وسیلهٔ آزمایش‌گر در t نوار است. در رابطهٔ (1) مقدار عددی l معلوم است، تنها کافی است مقدار $f(0)$ برآورد شود. مسئله اصلی در روش نمونه‌گیری ترابرشی خطی برآورد مناسب برای تابع چگالی $f(x)$ است و سپس با استفاده از رابطهٔ (1) می‌توان مقدار $f(0)$ را برآورد کرد. البته روش‌های متفاوتی برای برآورد مقدار $f(0)$ به وسیلهٔ نویسندگان مختلف پیشنهاد شده که به طور کلی می‌توان این روش‌ها را به دو دسته یکی روش پارامتری[1] و دیگری روش ناپارامتری[2] طبقه‌بندی کرد. در روش پارامتری فرض می‌شود که تابع $f_q(x)$ یک تابع چگالی با پارامتر q باشد. برآورد پارامتر q منجر به برآورد $f(0)$ به صورت $\hat{f}(0) = \hat{f}(0, q)$ خواهد شد. مدل‌های پارامتری مختلفی برای تابع $f_q(x)$ در نظر گرفته شده و در تمام آنها برآورد ماکسیمم درست‌نمایی پارامتر q به دست آمده است (برای جزییات بیشتر به بورنهام و همکاران، 1980 و باکلند، 1985 رجوع شود.)

روش‌های پارامتری روش‌های بسیار قدرتمندی هستند اما نتایج به دست آمده با این روش‌ها به شدت به نوع مدل انتخاب شده بستگی دارند. کرین و همکاران (1979)، برای برآورد تابع $f(x)$ از سری‌های فوریه استفاده می‌کنند که ضرایب سری‌ها با کمک داده‌ها برآورد می‌شوند. اما استفاده از این روش عاری از اشکال نیست، چرا که روش سری‌های فوریه در بیان تغییرات موضعی ناتوان است. از طرفی چون هدف برآورد مقدار $f(0)$ است، استفاده از روش سری‌های فوریه کار عاقلانه‌ای نیست. در استفاده از سری‌های فوریه در روش نمونه‌گیری ترابرشی خطی به باکلند (1982)، رجوع شود. در سال‌های اخیر، محققان توجه خودشان را بیشتر به روش‌های ناپارامتری معطوف کرده‌اند. یک روش ناپارامتری دیگر برای برآورد تابع $f(x)$ ، روش کرنل است (برای جزییات بیشتر به کتاب باکلند و همکاران، 1993 و مک و کوانگ، 1998 رجوع شود.)

کوانگ (1993)، نشان دادند که برای اندازهٔ نمونهٔ بزرگ برآورد تابع $f(x)$ دارای توزیع نرمال است و فاصله اطمینان برای تابع چگالی جمعیت در سطح معناداری a را به دست آورد. کارونامونی و کوین (1995)، با روش بیز ناپارامتری به برآورد تابع $f(x)$ پرداختند. ژانگ و کارونامونی (1998)، با استفاده از روش هموارسازی چندجمله‌ای موضعی[3] ارایه شده به وسیلهٔ فان (1992)، مقدار $\hat{f}(0)$ را به دست آوردند. بارابسی (2000)، با روش برآورد چگالی درست‌نمایی موضعی[4]، برآورد نیم پارامتری[5] تابع چگالی $f(x)$ را محاسبه کرد. برای درک هرچه بهتر مفاهیم در مورد برآورد تابع چگالی جمعیت در نمونه‌گیری

1. Parametric method
2. Non-parametric method
3. Local polynomial smoothing technique
4. Local likelihood density estimation approach
5. Semi parametric estimator

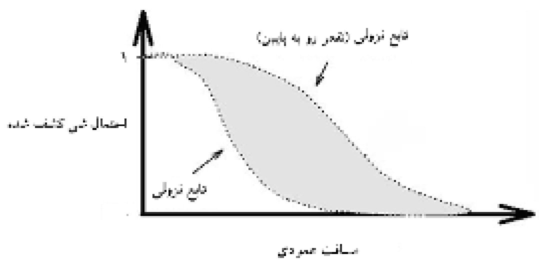
ترابرشی خطی به کتاب باکلند و همکاران (2001) و باکلند و همکاران (2004) رجوع شود. هدلی و باکلند (2004)، روش‌های تحلیلی را برای برازش مدل‌های فضایی[1] به داده‌های ترابرشی خطی ارایه و با مثال‌های عددی این روش‌ها را مورد بررسی قرار دادند. بورگونی و همکاران (2005)، رفتار برآوردگر $\hat{D}$ را برای اندازه‌های نمونه کوچک مورد بررسی قرار دادند. همچنین ایدوس (2005) یک برآورد جدید برای $f(0)$ با استفاده از روش کرنل[2] به دست آورد و ویژگی‌های مجانبی این برآوردگر را مورد بررسی قرار داد. در بخش دوم مقاله، مفهوم تابع تشخیص[3] در نمونه‌گیری ترابرشی خطی ارایه شده است. در بخش سوم مقاله، با توجه به توابع تشخیص در نظر گرفته شده، آزمون فرضیهٔ آماری برای تابع چگالی نیم نرمال[4] در مقابل تابع چگالی نمائی یک متغیره مورد بررسی قرار خواهد گرفت. آمارهٔ آزمون برای تصمیم‌گیری در مورد پذیرش یا رد فرضیهٔ صفر محاسبه شده است. در بخش چهارم مقاله، از روش شبیه‌سازی مونت کارلو[5] برای تعیین مقادیر بحرانی و توان آماره آزمون در سطوح مختلف معناداری استفاده شده است. مثال‌های عددی در بخش پنجم داده شده است. لازم به ذکر است که تمام محاسبات عددی برای تعیین مقادیر بحرانی آمارهٔ آزمون و توان آزمون با استفاده از نرم‌افزار SPLUS انجام شده است.

تابع تشخیص در نمونه‌گیری ترابرشی خطی

فرض کنید برای هر $0 \le x \le w$، $g(x)$ احتمال آن باشد که شیء در فاصله X از نوار دیده شود، بنابراین ,
(دیدن یک مشاهده به شرط آنکه مسافت قائم آن تا نوار x باشد) $g(x) = P$

تابع $g(x)$ را تابع تشخیص در نمونه‌گیری ترابرشی خطی گویند. (ابرهارت، 1978 و باکلند و همکاران، 1993)

لازم به ذکر است که شکل تابع تشخیص بستگی به نوع محیطی دارد که اشیاء از آن عبور می‌کنند، با توجه به این نکته، دو حالت امکان دارد برای نمودار تابع تشخیص رخ دهد که در هر دو حالت نمودار تابع نزولی است. نمودار تابع تشخیص در شکل 4 نشان داده شده است.



شکل 4. نمودار تابع تشخیص در نمونه‌گیری ترابرشی خطی

بنابراین با توجه به آنچه گفته شد، می‌توان به عنوان مثال توابع چگالی نیم نرمال و چگالی نمایی یک متغیره را به دلیل نزولی بودن آنها به عنوان توابع اختیاری برای تابع تشخیص $g(x)$ انتخاب کرد.

تعریف شرط جانبی: (بورنهام و آندرسون، 1987) فرض کنید $g^{j}(x)$ مشتق j ام تابع تشخیص باشد، اگر برای $j = 1$ ، تابع $g^{j}(x)$ برای مقادیر بزرگ‌تر از صفر مساوی با عدد صفر باشد، یعنی $g^{1}(0^{+}) = 0$ ، آنگاه می‌گوییم تابع تشخیص دارای شرط جانبی است و اگر $g^{1}(0^{+}) \neq 0$ باشد، می‌گوییم تابع تشخیص دارای شرط جانبی نیست. برای اطلاع بیشتر در مورد بودن یا نبودن شرط جانبی برای تابع تشخیص در برآورد تابع چگالی جمعیت به مک و همکاران (1999) و ژانگ (1999) رجوع شود.

کوانگ و لانکتات (1991)، در یک بررسی هوایی با روش نمونه‌گیری ترابرشی خطی از تابع تشخیص تک نمایی برای برآورد کردن تابع چگالی استفاده کردند. باکلند (1992)، برای برآورد تابع تشخیص از یک تابع چند جمله‌ای استفاده کرد. روش‌هایی برای برآورد تابع چگالی هنگامی که شرط جانبی وجود ندارد، به وسیلهٔ مک و همکاران (1999) ارائه شده است. مطالب فوق نشان می‌دهند که قبل از تجزیه و تحلیل داده‌ها دانستن این که شرط جانبی وجود دارد یا خیر الزامی است. چن (1996)، از این ایده که برآورد تابع تشخیص $g(x)$ هم به فواصل متغیرهای تصادفی $X_1, X_2, \ldots, X_n$ تا خط ترابرشی و هم به اندازهٔ آنها در عمل وابسته است، از روش کرنل برای برازش داده‌های دو

1. Spatial models
2. Kernel method
3. Detection function
4. Half normal density
5. Monte Carlo simulation

بعدی برای برآورد تابع تشخیص استفاده کرد. مک (1998) برای آزمون شرط جانبی روش ناپارامتری کرنل را پیشنهاد می‌کند. اما وقتی تابع تشخیص دارای شرط جانبی نباشد، روش ارایه شده به وسیلهٔ مک (1998)، دارای توان بسیار کمی است. (برای وقتی تابع تشخیص دارای توزیع نمایی باشد، توان آزمون 13/5 درصد است.) مرسی و جایارامان (1999)، یک روش کاملاً جدید برای هر اندازه نمونه در برآورد تابع تشخیص ارایه دادند. ژانگ (2000)، یک آزمون پرتوان‌تر از مک (1998) برای وجود شرط جانبی ارایه داد. ملویل و ولش (2001)، از روش‌های عددی برای یافتن و برآورد تابع تشخیص در نمونه‌گیری ترابرشی خطی استفاده کردند. فستر و همکاران (2005)، برآورد تابع تشخیص را با وجود داده‌های بریده شده[1] محاسبه کردند. باکلند و همکاران (2007)، انحراف از برخی فرضیات در نمونه‌گیری ترابرشی خطی را در نظر گرفتند و روش جدیدی را برای برآورد تابع چگالی جمعیت ارایه دادند. همچنین روش جدید را با یک مثال کاربردی برای برآورد تابع چگالی جمعیت گیاهان مورد بررسی قرار دادند. نتایج بیشتر در برآورد تابع چگالی جمعیت با وجود همبستگی بین متغیرها در نمونه‌گیری ترابرشی خطی به وسیلهٔ باکلند و همکاران (2010)، داده شده است. ژانگ (2011)، تابع چگالی نیم نرمال را برای تابع $f(x)$ در نظر گرفت و پارامتر واریانس آن را با به کارگیری روش ارایه شده به وسیلهٔ کویین و گالوچی (1980)، به دست آورد. این برآوردگر، نااریب برای واریانس نیست اما به طور تحلیلی و شبیه‌سازی نشان داده شده که اگر تابع تشخیص دارای دم سبک[2] باشد (برای مثال تابع نیم نرمال)، آنگاه این برآوردگر دارای واریانس کمتر و همچنین دارای میانگین مربعات خطای کمتری نسبت به برآوردگر به دست آمده با روش نسبت درست‌نمایی است.

شرایط اولیه در نمونه‌گیری ترابرشی خطی

در روش نمونه‌گیری ترابرشی خطی شرایط و قوانینی حاکم است که در اینجا به آنها اشاره می‌شود.

1) اشیایی که روی خط ترابرشی حرکت می‌کنند، هرگز گم نمی‌شوند، به عبارت دیگر

$$P\left(\text{دیدن یک شیء} \mid x=0\right)=1$$

2) اشیا بیش از یک بار در روش نمونه‌گیری ترابرشی شمرده نمی شوند؛ به عبارت دیگر انتخاب نمونه‌ها بدون جای‌گذاری انجام می‌شود.

3) مسافت‌ها و اشیا به طور دقیق اندازه‌گیری می‌شوند، بدون خطای اندازه‌گیری و خطای گرد کردن.

4) دیدن اشیای روی نوار ترابرشی، پیشامدهای مستقل از هم هستند.

5) هر چه فاصله شیء از نوار زیادتر شود، احتمال دیدن آن شیء کمتر می‌شود و برعکس.

آزمون فرضیه‌های آماری

در این مقاله، با توجه به اینکه آزمون فرضیهٔ آماری برای تابع چگالی نیم نرمال در مقابل تابع چگالی نمایی یک متغیره مورد بررسی قرار خواهد گرفت، اگر شرط جانبی وجود داشته باشد، آنگاه برای ثابت $a>0$، تابع تشخیص به فرم $g(x)=\exp(-ax^2)$ است. اگر شرط جانبی وجود نداشته باشد، آنگاه برای ثابت $b>0$، تابع تشخیص به فرم $g(x)=\exp(-bx)$ خواهد بود؛ بنابراین با ترکیب این دو حالت تابع تشخیص به صورت

$$g(x)=\exp(-ax^2+bx), \qquad (2)$$

نوشته می‌شود. از رابطهٔ (2) کاملاً واضح است که آزمون فرضیهٔ آماری برای تابع چگالی نیم نرمال در مقابل تابع چگالی نمایی در نمونه‌گیری ترابرشی خطی معادل با آزمون فرضیه $H_0: b=0$ در مقابل فرضیه $H_1: a=0$ است. (بنابراین به راحتی دیده می‌شود که اگر فرضیه H_0 درست باشد، تابع تشخیص دارای شرط جانبی است و اگر فرضیه H_1 درست باشد، تابع تشخیص دارای شرط جانبی نیست.)

آمارهٔ آزمون

فرض کنید متغیرهای تصادفی و هم توزیع $X_1, X_2, \ldots, X_n$ مسافت‌های عمودی از نوار ترابرشی با تابع چگالی $f(x)$، $0 \le x \le w$، باشند. بورنهام و آندرسون (1976)، نشان دادند که رابطه

---

1. Truncated data
2. Thin tail

(3)

$$f(x)=\frac{g(x)}{\int_0^{\omega} g(x)\,dt} \quad ; \quad 0\le x\le \omega,$$

بین تابع چگالی و تابع تشخیص برقرار است. در این مقاله فرض می‌شود که طول نوار بی‌نهایت ($\omega=\infty$) باشد. حال با تعریف نمادهای $\theta=(\alpha,\beta)'$ و $\xi=(-x^2,-x)'$، می‌توان رابطه (3) را به صورت

(4)

$$f(x)=\frac{\exp(\theta'\xi)}{\int_0^{\infty}\exp(\theta'\xi)\,dx} \quad ; \quad 0\le x\le \omega$$

نوشت. فرض کنید ω_0 و ω_1 دو زیر مجموعه از فضای پارامتری θ به صورت

$$\omega_0=\{\theta:\beta=0\} \quad , \quad \omega_1=\{\theta:\alpha=0\},$$

باشند. مجموعه ω_0 معادل فرضیه صفر و مجموعه ω_1 معادل فرضیه مقابل است. به راحتی دیده می‌شود که تحت فرضیه H_0، تابع چگالی $f(x)$ در رابطه (4) به فرم

$$f(x)=2\sqrt{\frac{\alpha}{\pi}}\exp(-\alpha x^2),$$

است. برای دیدن این موضوع با توجه به فرضیه $H_0:\beta=0$ خواهیم داشت:

$$\exp(\theta'\xi)=\exp\left[(\alpha,0)\begin{pmatrix} x^2\\ -x\end{pmatrix}\right]\exp(-\alpha x^2).$$

با توجه به اینکه

$$\int_0^{\infty}\exp(-\alpha x^2)\,dx=\int_0^{\infty}e^{-\alpha x^2}dx=\frac{1}{2}\sqrt{\frac{\pi}{\alpha}}.$$

بنابراین طبق فرمول (4) خواهیم داشت:

$$f(x)=2\sqrt{\frac{\alpha}{\pi}}\exp(-\alpha x^2),$$

تابع $f(x)$ تابع چگالی نیم نرمال است. به همین ترتیب تحت فرضیه H_0 تابع چگالی $f(x)$ به صورت

$$f(x)=\beta\exp(-\beta x),$$

خواهد بود. تحت فرضیه H_0 تابع $f(x)$ چگالی نمایی است که اثبات آن مشابه اثبات قبل است.

قضیه: فرض کنید مسافت‌های مشاهده شده از نوار $X_1, X_2, \ldots, X_n$ متغیرهای تصادفی و هم توزیع از چگالی نیم نرمال یا چگالی نمایی باشند، در این صورت برای آزمون فرضیه H_0 در مقابل فرضیه H_1، آماره آزمون $L=\frac{\sqrt{\sum_{i=1}^{n}X_i^2}}{\sum_{i=1}^{n}X_i}$ است و مقادیر بزرگ این آماره باعث رد فرضیه H_0 می‌شود.

اثبات: آمارۀ آزمون نسبت درست‌نمایی برای فرضیه H_0 در مقابل فرضیه H_1 عبارت است از:

(5)

$$LRT=\frac{\max_{\theta\in\omega_0}L(\theta)}{\max_{\theta\in\omega_1}L(\theta)}=\frac{\max_{\alpha>0}\prod_{i=1}^{n}\left[2\sqrt{\frac{\alpha}{\pi}}\exp(-\alpha x_i^2)\right]}{\max_{\beta>0}\prod_{i=1}^{n}\left[\beta\exp(-\beta x_i)\right]}$$

$$=\frac{\max_{\alpha>0}\left[2^n\left(\frac{\alpha}{\pi}\right)^{\frac{n}{2}}\exp\left(-\alpha\sum_{i=1}^{n}x_i^2\right)\right]}{\max_{\beta>0}\left[\beta^n\exp\left(-\beta\sum_{i=1}^{n}x_i\right)\right]}.$$

با لگاریتم گرفتن از طرفین تساوی (5) خواهیم داشت:

$$\operatorname{Ln} LRT =$$
$$\max_{\alpha>0}\left[n\ln 2+\frac{n}{2}\ln\alpha-\frac{n}{2}\ln\pi-\alpha\sum_{i=1}^{n}x_i^2\right]$$
$$-\max_{\beta>0}\left[n\ln\beta-\beta\sum_{i=1}^{n}x_i\right]$$

حال با مشتق‌گیری از α و β و مساوی صفر قرار دادن آنها، مقادیر $\hat{\alpha}$ و $\hat{\beta}$ عبارت‌اند از:

$$\hat{\alpha}=\frac{n}{2\sum_{i=1}^{n}x_i^2}\qquad,\qquad \hat{\beta}=\frac{n}{\sum_{i=1}^{n}x_i}.$$

بنابراین به راحتی دیده می‌شود که آمارهٔ آزمون نسبت درست‌نمایی عبارت است از:

$$LRT=\left(\frac{2e}{\pi n}\right)^{\frac{n}{2}}\frac{\sum_{i=1}^{n}X_i}{\sqrt{\sum_{i=1}^{n}X_i^2}}.$$

مقادیر کوچک آماره LRT باعث رد شدن فرضیه H_0 می‌شود، به عبارتی دیگر برای مقادیر بزرگ آماره $\Lambda=\frac{\sqrt{\sum_{i=1}^{n}X_i^2}}{\sum_{i=1}^{n}X_i}$ فرضیه H_0 رد خواهد شد. اما به منظور به دست آوردن ناحیه رد برای سطح معناداری داده شده α، نیاز به دانستن توزیع آماری Λ است ولی به دست توزیع آن کار ساده‌ای نیست. به همین دلیل از روش شبیه‌سازی مونت کارلو برای تعیین مقادیر بحرانی آماره استفاده می‌شود.

شبیه‌سازی مونت کارلو برای تعیین مقادیر بحرانی آمارهٔ آزمون

تحت فرضیه H_0 متغیرهای تصادفی $X_1, X_2, \ldots, X_n$ دارای تابع چگالی

$$f(x)=\sqrt{\frac{2}{\pi}}\,\sigma^{-1}\exp\left(-\frac{x^2}{2\sigma^2}\right)$$

هستند. با تعریف متغیر تصادفی Z به صورت $Z=\frac{X}{\sigma}$، آنگاه تابع چگالی این متغیر تصادفی عبارت است از:

$$f(z)=\sqrt{\frac{2}{\pi}}\exp\left(-\frac{z^2}{2}\right)$$

تابع $f(z)$ تابع چگالی نیم نرمال استاندارد است. حال می‌توان آماره Λ را به صورت

$$\Lambda=\frac{\sqrt{\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2}}{\sum_{i=1}^{n}\frac{X_i}{\sigma}}=\frac{\sqrt{\sum_{i=1}^{n}Z_i^2}}{\sum_{i=1}^{n}Z_i},\qquad(6)$$

بازنویسی کرد. این آماره از پارامتر σ مستقل است، و به این جهت می‌توان از روش شبیه‌سازی مونت کارلو مطابق مراحل زیر برای یافتن مقادیر بحرانی آماره استفاده کرد. لازم به ذکر است که ایده این شبیه‌سازی از کتاب سیل واپول و سن (2005)، برای یافتن مقادیر بحرانی برخی از آماره‌های آزمون گرفته شده است.

1) یک نمونه تصادفی به اندازه n از توزیع نرمال استاندارد گرفته و آنها را با $y_1, y_2, \ldots, y_n$ نشان دهید.

2) قدر مطلق این مقادیر یعنی $|y_i|$، $i=1,2,\ldots,n$، دارای توزیع نیم نرمال استاندارد هستند و برای هر i، $z_i=|y_i|$ قرار داده شود.

3) برای هر $i=1,2,\ldots,n$، مقدار عددی z_i را در رابطه (6) قرار دهید تا مقدار آماره Λ محاسبه شود، این مقدار را با λ_1 نشان دهید.

4) مراحل 1 تا 3 را 1000 مرتبه تکرار کنید تا مقادیر $\lambda_1, \lambda_2, \ldots, \lambda_{1000}$ محاسبه شوند.

5) برای سطح معناداری داده شده α، صدک $(1-\alpha)$ درصد مقادیر $\lambda_1, \lambda_2, \ldots, \lambda_{1000}$ به عنوان مقدار بحرانی α برآورد شود.

هرگاه برای یک نمونه n تایی نامساوی $L > \%100(1-\alpha)$ برقرار باشد، آنگاه فرضیۀ صفر در سطح معناداری α رد خواهد شد. در جدول 1 مقادیر بحرانی برای سطوح معناداری $\alpha = 0/05, 0/02, 0/01$ و اندازه نمونه‌های $n = 25, 50, 100, 200, 300$ محاسبه شده‌اند.

جدول 1. مقادیر بحرانی آماره آزمون در سطوح مختلف معناداری

$n \quad \alpha$	0/05	0/02	0/01
25	0/2438	0/2691	0/2851
50	0/1826	0/1943	0/1954
100	0/1483	0/1497	0/1520
200	0/0952	0/0971	0/0989
300	0/0770	0/0778	0/0781

آنچه از جدول 1 استنباط می‌شود این است که با افزایش سطح معناداری، مقدار بحرانی آزمون در حال کاهش است.

مثال‌های عددی

مثال 2: فرض کنید تابع تشخیص دارای چگالی نیم نرمال به فرم

$$g(x) = \exp\left(\frac{x^2}{2\sigma^2}\right), \qquad 0 \le x < \infty,\ \sigma^2 = 100$$

باشد. نمونه‌هایی با اندازه‌های 50، 100 و 200 از این تابع تشخیص گرفته و برای هر اندازه نمونه داده شده، مقدار آماره L در رابطه (6) محاسبه شده و فرآیند محاسبه آماره L با گرفتن دسته‌های 500 تایی از هر نمونه تکرار شده است؛ بنابراین 500 مقدار عددی برای آماره L به دست آمده و آنگاه نسبت دفعاتی که مقدار آماره L از عدد مورد نظر در جدول بالا بزرگ‌تر باشد به دست آمده است. برای سطوح معناداری $\alpha = 0/05, 0/02, 0/01$ نتایج در جدول 2 آورده شده‌اند.

جدول 2. نسبت‌های رد برای آمارۀ آزمون در سطوح مختلف معناداری

$n \quad \alpha$	0/05	0/02	0/01
50	0/741	0/634	0/614
100	0/882	0/751	0/693
200	0/966	0/804	0/721
300	1	0/912	0/811

مثال 3: فرض کنید تابع تشخیص دارای چگالی نمایی زیر باشد:

$$g(x) = \exp(-\lambda x), \qquad 0 \le x < \infty,\ \lambda = 1$$

نمونه‌هایی با اندازه‌های 50 ، 100 و 200 از این تابع گرفته شده و برای هر اندازه نمونه داده شده، مقدار آماره L در رابطه (6) محاسبه شده و فرآیند محاسبه آماره L را با گرفتن دسته‌های 500 تایی از هر نمونه تکرار شده است. در جدول 3، مقادیر عددی توان آمارۀ آزمون محاسبه شده است.

جدول 3. توان آمارۀ آزمون در سطوح مختلف معناداری

$n \quad \alpha$	0/05	0/02	0/01
50	0/063	0/047	0/064
100	0/046	0/056	0/037
200	0/032	0/061	0/034
300	0/067	0/042	0/052

آنچه از جدول 3 استنباط می‌شود این است که با افزایش سطح معناداری α، مقدار توان آزمون نیز در حال افزایش است. حتی افزایش حجم نمونه نیز در افزایش مقدار توان آزمون تاثیرگذار است.

بحث و نتیجه‌گیری

در این مقاله، برخی از جنبه‌ها و مفاهیم روش نمونه‌گیری ترابرشی خطی همراه با کاربرد عملی آن مورد بحث و بررسی قرار گرفت. برای تابع تشخیص دو نوع تابع چگالی یکی تابع چگالی نیم نرمال و دیگری تابع چگالی نمایی به ترتیب تحت فرضیۀ صفر و فرضیۀ مقابل در نظر گرفته شد.

برای انجام این آزمون فرضیه، با استفاده از روش نسبت درست‌نمایی آمارهٔ آزمون محاسبه شد. اما به دلیل ساختار پیچیده این آماره، تعیین توزیع تحت فرضیهٔ صفر آن کار ساده‌ای نبود، به همین دلیل از روش شبیه‌سازی مونت کارلو برای محاسبه مقادیر بحرانی و توان آن استفاده شد.

تقدیر و تشکر: نویسنده از سر دبیر محترم مجلهٔ گسترهٔ علوم آماری و داوران محترم بخاطر پیشنهادات ارزشمند ایشان که موجب تغییرات اساسی در راستای هر چه بهتر شدن مقاله شد، نهایت تشکر را دارد.

منابع

[1] Barabesi, L. (2000). Local likelihood density estimation in line transect sampling, Environmetrics, 11, 413-422.

[2] Borgoni, R., Cameletti, M. and Quatto, P. (2005). Comparing estimators of animal abundance: a simulation study. In: Atti del Convegno della Società Italiana di Statistica “Statistica e Ambiente. Università degli Studi di Messina, Messina, 181-184.

[3] Buckland, S. T. (1982). A note on the fourier series model for analysing line transect data, Biometrics, 38, 469-477.

[4] Buckland, S. T. (1985). Perpendicular distance models for line transect sampling. Biometrics, Biometrics, 41, 177-195.

[5] Buckland, S. T. (1992). Fitting density functions using polynomials. Applied Statistics, 41, 63-76.

[6] Buckland, S. T., Anderson, D. R., Burnham, K. P. and Laake, J. L. (1993). Distance Sampling, London: Chapman and Hall.

[7] Buckland, S. T., Borchers, D. L., Johnston, A., Henrys, P. A., and Marques, T. A. (2007). Line transect methods for plant surveys, Biometrics, 63, 989-998.

[8] Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L., (2001). Introduction to Distance Sampling. Oxford University Press, Oxford.

[9] Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L. and Thomas, L. (2004). Advanced Distance Sampling. Oxford University Press, Oxford.

[10] Buckland, S. T., Plumptre, A. J., Thomas, L. and Rexstad, E. A. (2010). Design and analysis of line transect surveys for primates, International Journal of Primary, 31, 833-847.

[11] Burnham, K. P. and Anderson, D. R. (1976). Mathematical models for nonparametric inferences from line transect data. Biometrics, 32, 325-336.

[12] Burnham, K. P., Anderson, D. R., and Laake, J. L. (1980). Estimation of density from line transect sampling of biological populations, Wildlife Monograph 72, supplement to Journal of Wildlife Management, 44.

[13] Chen, S. X. (1996). Studying school size effects in line transect sampling using the kernel method. Biometrics, 52, 69-84.

[14] Crain, B. R., Burnham, K. P, Anderson, D. R. and Laake, J. L. (1979). Nonparametric estimation of population density for line transect sampling using Fourier series. Biometrical Journal, 21, 731-748.

[15] Eberhardt, L. L. (1978). A preliminary appraisal of line transects. Journal of Wildlife Management, 32, 82-88.

[16] Eidous, O. M. (2005). On improving kernel estimators using line transect sampling. Communications in Statistics -Theory and Methods, 34, 931-941.

[17] Fan, J. (1992). Design-adaptive nonparametric regression. Journal of the American Statistical Association, 87, 998-1004.

[18] Fewster, R. M., Laake, J. L. and Buckland, S. T. (2005). Biometrics, 61, 856-861.

[19] Hedley, S. L. and Buckland, S. T. (2004). Spatial models for line transect sampling. Journal of Agricultural, Biological and Environmental Statistics, 9, 181-199.

[20] Karunamuni, R. J. and Quinn, T. J. (1995). Bayesian estimation of animal abundance for line transect sampling. Biometrics, 51, 1325-1337.

[21] Mack, Y. P. (1998). Testing for the shoulder condition in transect sampling. Communications in Statistics -Theory and Methods, 27(2), 423-432.

[22] Mack, Y. P. and Quang, P. X. (1998). Kernel methods in line and point transect samplings. Biometrics, 50, 606-619.

[23] Mack, Y. P., Quang, P. X. and Zhang, S. (1999). Kernel estimation in transect sampling without the shoulder condition. Communications in Statistics -Theory and Methods, 28, 2277-2296.

[24] Melville, G. J., and Welsh, A. H. (2001). Line transect sampling in small regions, Biometrics, 57, 1130-1137.

[25] Mercey, K. A. and Jayaraman, K. (1999). Predicting the variation in detection function in line transect sampling through random parameter model, Find out how to access preview-only content , Environmental and Ecological Statistics, 6(4), 341-350.

[26] Quang, P. X. (1993). Nonparametric estimators for variable circular plot surveys. Biometrics, 49, 837-852.

[27] Quang, P. X. and Lanctot, R. B. (1991), A line transect model for aerial surveys. Biometrics, 47, 1089-1102

[28] Quinn, T. J. and Gallucci, V. F. (1980). Parametric models for line-transect estimators of abundance. Ecology, 61, 293-302.

[29] Silvapulle, M. J. and Sen, P. K. (2005). Constrained Statistical Inference: Inequality, Order, and Shape Restrictions, John Wiley, New York.

[30] Zhang, S. (1999). Improvements on the kernel estimation in line transect sampling without the shoulder condition. Statistics and Probability Letters, 53, 249-258.

[31] Zhang, S. (2000). A note on testing the shoulder condition in line transect sampling. Technical report, Department of Mathematical Sciences, University of Alaska Fairbanks.

[32] Zhang, S. (2011)., On parametric estimation of population abundance for line transect sampling, Environmental and Ecological Statistics, 18, 79-92.

[33] Zhang, S. and Karunamuni, R. J. (1998). On kernel density estimation near endpoints. Journal of Statistical Planning and Inference, 70, 301-316.