بسم الله الرحمن الرحیم

دانشگاه یزد

دانشکده ریاضی

گروه آمار

پایان‌نامه

برای دریافت درجه کارشناسی ارشد
آمار

آزمون نسبت احتمال دنباله‌ای تصادفی: نظریه و کاربرد

استاد راهنما: دکتر عیسی محمودی

استاد مشاور: دکتر حمزه ترابی

پژوهش و نگارش: بتول کارگر

مهر ماه ۱۳۹۲

تقدیم به

پدر و مادرم

که تصور لبخندشان، تلخی رنج را برایم به شیرینی رضایت تبدیل می‌کند.

سپاس

به نام یزدان پاک و سپاس از هر آنچه نعمت بیکران که بهره مندم ساخته است.

بی شک به سرانجام رساندن این پایان نامه مرهون تلاش ها و مساعدت های بی دریغ افرادی است که گرچه ممکن است در متن کار نامی از آن ها برده نشود، اما نادیده گرفتن زحماتشان بی انصافی بزرگی است؛ آن ها که به ظاهر پشت پرده بودند اما کمک هایشان آنقدر پررنگ بود که نامشان را مستحق ذکر شدن در اولین صفحات پایان نامه کرداند. پس به رسم قدرشناسی، بر خود لازم می دانم صمیمانه ترین سپاسگزاری هایم را نثار تنی چند از این عزیزان کنم:

پیش از همه قدردان زحمات پدر و مادر و خانواده ی عزیزم هستم، آن ها که همیشه ناب ترین کلیشه های تشکر بوده و هستند؛ کلیشه هایی که با وجود تکرار صدباره و هزارباره باز هم بکر هستند و شایسته ی تقدیر. به راستی که جامه ی تشکر را به هر مدل و اندازه ای که درآوری باز هم برازنده ی قامت آن هاست.

سپاسگزارم از آقای دکتر محمودی که زحمت راهنمایی این پایان نامه را به جان خریدند. بی گمان محدود کردن ایشان به دو کلمه ی "استاد" و "راهنما" کم لطفی بزرگی است که هرگز تن به آن نخواهم داد، پس سپاس بی نهایتم را نثار ایشان می کنم به پاس "بی نهایت" بودنشان. نیز مرهون زحمات آقای دکتر ترابی هستم، اما خالصانه و به دور از چارچوب کلیشه ای سپاس از استاد مشاور؛ چرا که از ایشان حسن خلق را آموختم که شایسته ی بی ریاترین تقدیرهاست.

سرانجام بر خود واجب می دانم از آقای دکتر ذاکرزاده و آقای دکتر روزگار که زحمت داوری این پایان نامه را متقبل شدند و نیز سایر اساتید محترم گروه آقای دکتر دولتی، آقای دکتر جعفری و آقای میرحسینی صمیمانه تشکر کنم.

همچنین به رسم دوستی، سپاسگزار همه ی عزیزانی هستم که با صبر و مهربانی، اینجانب را در جهت پیشبرد طرح پیش رو یاری رساندند، بویژه دوست عزیزم خانم رحمت مسکوتی.

چکیده

در بیشتر مسایل آماری، هدف این است که پس از مشاهده‌ی یک نمونه از خانواده‌ی توابع چگالی $\{f(x;\theta):\theta\in\Theta\}$، در مورد پارامتر θ نتیجه‌گیری شود. در بسیاری از آن‌ها، تعداد مشاهدات، یعنی حجم نمونه قبل از بیان و انجام آزمایش تعیین شده است. اصطلاحاً به چنین تجزیه و تحلیل‌های آماری که مبتنی بر تعداد مشخصی از حجم نمونه است، تحلیل آماری با حجم نمونه‌ی ثابت گویند، اما در برخی دیگر مانند بازرسی نمونه‌ای و کنترل کیفیت لزومی به تعیین تعداد مشاهده‌ها قبل از انجام آزمایش نیست، بلکه در عمل نتایج آزمایش تعیین کننده‌ی حجم نمونه است؛ این‌چنین تجزیه و تحلیل‌های آماری به تحلیل‌های دنباله‌ای موسوم است.

در این پایان‌نامه ابتدا آزمون نسبت احتمال دنباله‌ای، به همراه برخی از ویژگی‌های آن بیان می‌شود. سپس آزمون نسبت احتمال دنباله‌ای تصادفی را بیان و آن را با آزمون با اندازه‌ی نمونه‌ی ثابت مقایسه می‌کنیم. در آخر نیز کاربردهایی از آزمون نسبت احتمال دنباله‌ای تصادفی را با دو مثال شرح خواهیم داد.

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

نمادها

فهرست نمادها

پیشگفتار

آبراهام والد را می‌توان پیشرو در زمینه‌ی تحلیل دنباله‌ای دانست. مسئله‌ی تحلیل دنباله‌ای ابتدا در دهه‌ی ۱۹۴۰، توسط برنارد[۱] و والد[۲] [۳۳] با معرفی آزمون نسبت احتمال دنباله‌ای ($SPRT$) شروع شد و سپس والد و ولفویتس[۳] [۳۴] در سال ۱۹۴۸ بهینه بودن این آزمون را ثابت کردند.

آزمون نسبت احتمال دنباله‌ای برای دامنه‌ی وسیعی از مسایل کاربردی به‌کار گرفته شده است. در سال‌های اخیر کاربردهای گسترده‌ای از این روش در بسیاری از عرصه‌ها مانند تشخیص فرکانس کم دستگاه‌های زیردریایی، ردیابی سیگنال‌ها، کشف به موقع تغییرات ناگهانی در سیگنال‌ها، شبیه‌سازی کامپیوتری، علوم کشاورزی، کنترل و مدیریت آفات، آزمون‌های تحصیلی و اقتصاد و دارایی وجود داشته است. پژوهش‌های فراوانی در زمینه‌ی آزمون نسبت احتمال دنباله‌ای و سایر آزمون‌های دنباله‌ای توسط نویسندگانی مانند والد [۳۳] در سال ۱۹۴۷، بک هوفر و همکاران[۴] [۸] در سال ۱۹۶۸، گوش[۵] [۱۳] در سال ۱۹۷۰، سِن[۶] [۲۶] در سال ۱۹۸۱، گووین دراجولا[۷] [۱۶، ۱۵] در سال‌های ۱۹۸۱ و ۲۰۰۴، سایموند[۸] [۲۷] در سال ۱۹۸۵، وایت هِد[۹] [۳۵] در سال ۱۹۹۷، وودروف[۱۰] [۳۶] در سال ۱۹۸۲ و گوش و همکاران [۱۴] در سال ۱۹۹۷ انجام گرفته است.

مقاله‌ی قابل توجهی از لای[۱۱] [۱۸] در سال ۲۰۰۴ وجود دارد که به همراه مجموعه‌ی بزرگی از

---

[۱] Bernard

[۲] Wald

[۳] Wald and Wolfowitz

[۴] Bechhofer et al.

[۵] Ghosh

[۶] Sen

[۷] Govindrajulu

[۸] Siegmund

[۹] Whitehead

[۱۰] Woodroofe

[۱۱] Lai

مباحث که پس از آن بیان شده است، بسیاری از ویژگی‌های والد را به همراه دامنه‌ی وسیعی از تحقیقات آماری همچون آزمایش‌های بالینی، ریاضیات مالی و نقص امنیت شبکه‌های رایانه‌ای معرفی کرده است.

هرچند در این پایان‌نامه، تمرکز روی تصمیم‌گیری بین دو فرضیه است، اما آزمون‌های دنباله‌ای چندفرضیه‌ای نیز وجود دارند که برخی، کاربردهای آن را بررسی کرده‌اند، از آن جمله می‌توان به آرمیتاژ[۱۲] [۴] در سال ۱۹۵۰، لوردِن[۱۳] [۲۱] در سال ۱۹۷۷، بوم و ویراوالی[۱۴] [۷] در سال ۱۹۹۴، تارتاکوفسکی[۱۵] [۲۹، ۳۰] در سال ۱۹۹۸، دراگالین و همکاران[۱۶] [۱۱، ۱۲] در سال‌های ۱۹۹۹ و ۲۰۰۰ و تارتاکوفسکی و همکاران [۳۱] در سال ۲۰۰۳، اشاره کرد. همچنین برای نگاه اجمالی بر کاربردهای ویژه‌ی روش دنباله‌ای در برخی از زمینه‌های معاصر آمار، می‌توان به مراجع زیر که همگی در سال ۲۰۰۴ به چاپ رسیده‌اند مراجعه کرد: آبراهام[۱۷] [۳]، بارون[۱۸] [۶]، چانگ و مارتین سِک[۱۹] [۹]، لای [۱۹]، لی و سولانکی[۲۰] [۲۰]، ماخوپادهای و همکاران[۲۱] [۲۵]، تارتاکوفسکی و ویراوالی [۳۲]، یانگ[۲۲] [۳۷] و زاکس و روگاتکو[۲۳] [۳۸]. از دیگر کاربردهای این روش می‌توان به کشف تأثیر صدا روی پستانداران دریایی، فرآیندهای چندحالتی، سیستم‌های توزیع‌شده‌ی چندمجرایی، داده‌کاوی، آزمایش‌های بالینی، ردیابی هدف، کشف نقطه‌ی تغییر، مدیریت آفات، باغبانی علمی و مرتب‌سازی ژن‌ها اشاره کرد.

---

[۱۲] Armitage

[۱۳] Lorden

[۱۴] Baum and Veeravalli

[۱۵] Tartakovsky

[۱۶] Dragalin et al.

[۱۷] Abraham

[۱۸] Baron

[۱۹] Chang and Martinsek

[۲۰] Li and Solanky

[۲۱] Mukhopadhyay et al.

[۲۲] Young

[۲۳] Zacks and Rogatko

گاهی در برخی از گروه‌ها اندازه‌ی گروه نیز مانند مشاهدات خود رفتار کرده و به‌صورت دنباله‌ای هستند؛ در این صورت آزمون نسبت احتمال دنباله‌ای تصادفی ($RSPRT$) معرفی می‌گردد. این آزمون توسط ماخوپادهای و دسیلوا[۲۴] [۲۳] در سال ۲۰۰۸ معرفی شده است و آن را در فصل سوم شرح خواهیم داد. به‌طور مثال در یک جریان ترافیک، زمانی که چراغ راهنما از سبز به قرمز تغییر رنگ می‌دهد، ممکن است شخصی تعداد خودروهای متوقف شده (M_i) و تعداد خودروهای بدون چراغ ترمز سالم ($\sum_{j=۱}^{M_i} X_{ij}$) را زمانی که $i = ۱, ۲, \cdots$ ثبت کند. در این مثال باید دقت داشت که گرچه مدل دادن به M_iها به‌صورت دنباله‌ای معقول است، اما در توقف همزمان خودروها ممکن است ثبت داده‌های دنباله‌ای از چراغ‌های ترمز که ۱ یا $X_{ij} = ۰$، به‌طور جداگانه برای هر یک از آن‌ها امکان‌پذیر نباشد.

این پایان نامه شامل سه فصل است. در فصل اول مفاهیم و تعاریف اولیه بیان شده است. در فصل دوم، آزمون نسبت احتمال دنباله‌ای مطرح و برخی از ویژگی‌های آن بررسی خواهد شد. در واپسین بخش از این فصل، آزمون نسبت احتمال دنباله‌ای بریده شده بیان می‌شود. در فصل سوم، آزمون نسبت احتمال دنباله‌ای تصادفی معرفی و با بهترین آزمون با اندازه نمونه‌ی ثابت مقایسه خواهد شد، همچنین آزمون نسبت احتمال دنباله‌ای تصادفی بریده شده نیز مطرح می‌گردد. در انتهای همین فصل، کاربرد آزمون نسبت احتمال دنباله‌ای تصادفی را با دو مثال شرح خواهیم داد.

[۲۴] Mukhopadhyay and de Silva

فصل ۱

مفاهیم و تعاریف اولیه

مقدمه

این فصل شامل برخی از مفاهیم و تعاریف پایه‌ای مورد نیاز در این پایان‌نامه است. در ابتدا برخی از مفاهیم و قضایای اولیه‌ی احتمال بیان می‌شود. سپس آزمون فرض‌های پارامتری را تعریف کرده و پرتوان‌ترین آزمون و آزمون نسبت درستنمایی بیان می‌شوند. در آخر نیز آزمون دنباله‌ای به‌صورت کوتاه مطرح می‌گردد.

۱.۱ تعاریف و قضایای احتمالی مورد نیاز

تعریف ۱.۱.۱ (تابع مولد گشتاور) فرض کنید X یک متغیر تصادفی با تابع توزیع $F(.)$ باشد. در این صورت تابع مولد گشتاور X که با نماد $M_X(t)$ نشان داده می‌شود، به‌صورت زیر تعریف می‌گردد:

$$M_X(t) = E\left[e^{tX}\right] = \begin{cases} \sum\limits_{x} e^{tx} f(x), & X \text{گسسته}, \\ \int\limits_{\mathbb{R}} e^{tx} f(x)\, dx, & X \text{پیوسته}, \end{cases}$$

هرگاه $h > ۰$ موجود باشد به‌طوری که امید ریاضی برای هر مقدار t در بازه $(-h, h)$ وجود داشته باشد. اگر امید ریاضی در یک همسایگی صفر موجود نباشد آنگاه می‌گوییم تابع مولد گشتاور X وجود ندارد.

تعریف ۲.۱.۱ (همگرایی در احتمال) دنباله‌ی $\{X_n, n \geq ۱\}$ از متغیرهای تصادفی را در احتمال همگرا به متغیر تصادفی X می‌گویند، هرگاه برای هر $\epsilon > ۰$،

$$\lim_{n\to\infty} P\left(|X_n - X| \geq \epsilon\right) = ۰,$$

و آن را با نماد $X_n \xrightarrow{p} X$ نشان می‌دهند.

تعریف ۳.۱.۱ (همگرایی در توزیع) دنباله‌ی $\{X_n, n \geq ۱\}$ را در توزیع به X همگرا گویند، هرگاه برای همه‌ی نقاط پیوستگی $F(.)$،

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x),$$

و آن را با نماد $X_n \xrightarrow{d} X$ نشان می‌دهند.

قضیه ۴.۱.۱ (قضیه‌ی حد مرکزی) فرض کنید $\{X_n, n \geq 1\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع باشد به‌طوری که $E[X_n] = \mu$ و $Var(X_n) = \sigma^2 < \infty$ و $S_n = \sum\limits_{i=1}^{n} X_i$، در این صورت

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0, 1).$$

زمانی که $n \to \infty$.

قضیه ۵.۱.۱ (قضیه‌ی حد مرکزی تصادفی انسکامب[1][17]) فرض کنید $\{X_n, n \geq 1\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین ۰ و واریانس $\sigma^2 < \infty$ باشد و $S_n = \sum\limits_{i=1}^{n} X_i$. هم‌چنین فرض کنید $N(t)(t > 0)$ متغیر تصادفی صحیح و مثبت باشد. اگر برای $0 < \theta < \infty$، $\frac{N(t)}{t} \xrightarrow{p} \theta$، آنگاه

$$\frac{S_{N(t)}}{\sigma\sqrt{N(t)}} \xrightarrow{d} Z \sim N(0, 1),$$

و

$$\frac{S_{N(t)}}{\sigma\sqrt{\theta t}} \xrightarrow{d} Z \sim N(0, 1),$$

زمانی که $t \to \infty$.

تعریف ۶.۱.۱ (نرمال مجانبی) دنباله‌ی $\{X_n, n \geq 1\}$ از متغیرهای تصادفی مستقل را دارای توزیع مجانبی نرمال گویند، هرگاه

$$\lim_{n \to \infty} P\left(\frac{X_n - \mu_n}{\sigma_n} < t\right) = \Phi(t),$$

که $\Phi(t)$ تابع توزیع نرمال استاندارد است. در این حالت μ_n و σ_n^2 را به‌ترتیب، میانگین مجانبی و واریانس مجانبی X_n می‌نامند.

---

[1] Anscombe

۲.۱ آزمون فرض‌های پارامتری

متغیر تصادفی X را که توزیع آن به پارامتر θ بستگی دارد در نظر بگیرید، پارامتر θ مقداری ثابت اما مجهول است. در آزمون فرض‌های پارامتری دو فرضیه در مورد θ در برابر یکدیگر قرار می‌گیرند.

برای تشکیل فرض‌های آماری، فضای پارامتر Θ را به دو زیرمجموعه‌ی جدای Θ_0 و Θ_1 تفکیک کرده، که در آن $\Theta_0 \cup \Theta_1 = \Theta$ و $\Theta_0 \cap \Theta_1 = \emptyset$، و H_0 و H_1 را به‌ترتیب با $\theta \in \Theta_0$ و $\theta \in \Theta_1$ متمایز می‌کنیم. اگر Θ_0 یا Θ_1 تک عضوی باشد، آنگاه فرض نظیر آن را ساده و در غیر این صورت مرکب می‌خوانند. آزمون به‌صورت زیر تعریف می‌شود:

$$\begin{cases} H_0 : \theta \in \Theta_0 & \text{فرض صفر} \\ H_1 : \theta \in \Theta_1 & \text{فرض مقابل.} \end{cases} \tag{۱.۱}$$

اینک تعاریف زیر را که برگفته از مراجع [۲، ۱] هستند، بیان می‌کنیم.

تعریف ۱.۲.۱ برای آزمون فرض (۱.۱) با تابع آزمون $\phi(x) = \begin{cases} 1 & x \in C \\ 0 & x \in \bar{C} \end{cases}$، تابع توان، تابع عملکرد (تابع مشخصه‌ی عملگر)، احتمال خطای نوع اول، احتمال خطای نوع دوم و توان به‌ترتیب به‌صورت زیر تعریف می‌شوند:

$$\pi(\theta) = P_\theta(H_0 \text{رد}) = P_\theta(X \in C) = E_\theta[\phi(X)],$$

$$OC(\theta) = P_\theta(H_1 \text{رد}) = 1 - P_\theta(X \in C) = 1 - E_\theta[\phi(X)] = 1 - \pi(\theta),$$

$$\alpha = P_{H_0}(X \in C),$$

$$\beta = P_{H_1}(X \in \bar{C}),$$

و

$$\begin{aligned} \pi &= P\left(H_0 \text{رد} | H_1 \text{درست}\right) \\ &= 1 - \beta. \end{aligned}$$

تعریف ۲.۲.۱ برای آزمون فرض (۱.۱)، آزمون ϕ یک آزمون در سطح α نامیده می‌شود هرگاه برای هر $\theta \in \Theta_\circ$،

$$E_\theta[\phi(X)] \leq \alpha,$$

و آن را آزمونی به اندازه‌ی α می‌نامند هرگاه

$$\sup_{\theta\in\Theta_\circ} E_\theta[\phi(X)] = \alpha.$$

۱.۲.۱ پرتوان‌ترین آزمون

همان‌گونه که قبلا بیان شد با توجه به این نکته که فضای پیشامد فرض صفر و فرض مقابل تک عضوی باشد یا نه، آزمون‌ها به دو دسته‌ی ساده و مرکب تقسیم‌بندی می‌شوند. اینک می‌توان با توجه به لم نیمن- پیرسون پرتوان‌ترین آزمون (MPT) را برای آزمون فرض‌های ساده تعریف کرد.

لم ۳.۲.۱ (لم نیمن-پیرسون) فرض کنید $X_۱, X_۲, \cdots, X_n$ یک نمونه‌ی تصادفی از تابع چگالی $f(x;\theta)$ باشد. هدف انجام آزمون $H_\circ : \theta = \theta_\circ$ در برابر $H_۱ : \theta = \theta_۱$ است. برای آزمودن $H_\circ$ در برابر $H_۱$، اگر تابع آزمون ϕ برای هر $\circ \leq k < \infty$ به‌صورت زیر تعریف شود:

$$\phi(\boldsymbol{x}) = \begin{cases} ۱ & f(\boldsymbol{x};\theta_۱) > kf(\boldsymbol{x};\theta_\circ) \\ \gamma & f(\boldsymbol{x};\theta_۱) = kf(\boldsymbol{x};\theta_\circ) \\ \circ & f(\boldsymbol{x};\theta_۱) < kf(\boldsymbol{x};\theta_\circ), \end{cases} \qquad (۲.۱)$$

که در آن $E_{\theta_\circ}[\phi(\boldsymbol{X})] = \alpha$، آنگاه ϕ پرتوان‌ترین آزمون در سطح α برای آزمون فرض $H_\circ$ در برابر $H_۱$ است [۲].

۲.۲.۱ آزمون نسبت درستنمایی

فرض کنید، فرضیه‌های آزمون مرکب باشند. برای انجام چنین آزمونی ابتدا خاصیت نسبت درستنمایی یکنوا را تعریف می‌کنیم.