٤١٨٣
عطف
٢٤

بسم الله الرحمن الرحيم

٣٢٢٩٧

۱۳۷۹/۱۱/۲۰

پایان نامه

برای دریافت درجه کارشناسی ارشد

آمـــار

موضوع:

آزمون فرض های تقریبی

استاد راهنما:

دکتر سیامک نوربلوچی

اساتید مشاور:

دکتر محمد رضا مشکانی

دکتر جلال داودزاده

نگارش:

محمد اندبیلی

شهریور ۱۳۷۵

فهرست مطالب

عنوان **صفحه**

فهرست مطالب

عنوان **صفحه**

فهرست مطالب

عنوان **صفحه**

فهرست مطالب

عنوان | **صفحه**

مقدمه:

موضوعات مورد بحث این رساله بشرح زیر است:

فصل اوّل این رساله را با شرح روش نیمن ـ پیرسن برای ساختن تواناترین آزمونها شروع کرده و سپس آزمون نسبت درستنمایی تعمیم یافته را توضیح می‌دهیم و در نهایت تابع چگالی احتمال توزیع چند جمله‌ای

خواص آن را معرفی می‌کنیم. در فصل دوّم روش بازنمونه‌گیری بوت‌استرپ را بطور خلاصه توضیح می‌دهیم. در فصل سوّم آزمونهای جایگشتی را به اجمال شرح داده و از این روش برای آزمون یکسانی توزیع‌های دو جامعه استفاده می‌کنیم. در فصل چهارم آزمون‌های بوت‌استرپی را بطور کامل شرح می‌دهیم و از آن برای آزمون یکسانی توزیعهای دو جامعه، آزمون‌های برابری میانگین‌ها و واریانسهای دو جامعه و آزمون میانگین یک جامعه استفاده می‌کنیم. در فصل پنجم ابتدا روشی برای تقریب توزیعها توسط توزیع چند جمله‌ای ارائه می‌دهیم، سپس آزمون فرض تقریبی را تعریف کرده و با کمک آماره نسبت در ستنمایی تعمیم یافته، آزمون تقریبی میانگین یک جامعه مجهول را شرح می‌دهیم. در فصل ششم با استفاده از روشهای کامپیوتری تواناترین آزمون‌ها برای آزمون‌هایی بفرم $H: \underset{\sim}{P} = \underset{\sim}{P}_0$ در مقابل $K: \underset{\sim}{P} = \underset{\sim}{P}_1$ را بشیوه نیمن ـ پیرسن و در خانواده چگالی‌های چند جمله‌ای می‌سازیم و سپس باق فرض ثابت بودن H در آزمون قبل و تواناترین آزمون بدست آمده، کلاسی از توزیعها را مشخص می‌کنیم که اگر در فرض H قرار بگیرند در تواناترین آزمون بدست آمده قبلی تغییری حاصل نشود.

در قسمت ضمیمه نیز برنامه‌های کامپیوتری را که در این رساله مورد استفاده قرار گرفته‌اند را شرح می‌دهیم. از آنجا که نرم افزار S _ PLUS یک نرم افزار آماری پرقدرت است تمامی این برنامه ها به زبان S _ PLUS نوشته شده‌اند.

فصل اوّل
مقدمه‌ای بر آزمون فرض‌ها و توزیع چند جمله‌ای

۱ـ۱ـ بطور یکنواخت تواناترین آزمونها

۱ـ۱ـ۱ـ شرح مسئله

در این بخش مسئله آماری "آزمون فرض" مورد مطالعه قرار می‌گیرد که موضوع اساسی این فصل است. مسئله از اینجاا آغاز می‌شود که شخصی می‌خواهد نسبت به "ردّ" یا "قبول" فرض فرمول بندی شده‌ای تصمیم بگیرد. در اینجا تنها می‌توان یکی از دو عمل "پذیرش" و یا "ردّ" فرض را انتخاب کرد. روش تصمیم‌گیری در چنین مسائلی آزمون فرض نامیده می‌شود. این تصمیم‌گیری بر اساس مقدار مشاهده شده متغیر تصادفی مفروض X و توزیع P_θ که می‌دانیم متعلق به کلاس $\mathcal{P}=\{P_\theta \,;\, \theta\in\Omega\}$ است، صورت می‌گیرد. در صورتیکه θ معلوم باشد صحت یا عدم صحت فرض مشخص است. بنابراین توزیعهای کلاس $\mathcal{P}$ را می‌توان به دو گروه تقسیم کرد:

الف ـ توزیعهایی که تحت آنها فرض درست است.

ب ـ توزیعهایی که تحت آنها فرض درست نیست.

این دو کلاس دو بدو ناسازگار توسط H و K و زیر مجموعه‌های متناظرشان از Ω بترتیب با Ω_H و Ω_K نمایش داده می‌شوند و $H \cup K = \mathcal{P}$ و $\Omega_H \cup \Omega_K = \Omega$.

به زبان ریاضی، فرض معادل این عبارتست که $P_\theta \in H$. بنابراین مناسب است که فرض با عبارت اخیر مشخص شود و همچنین برای نمایش فرض از حرف H استفاده شود.

بطور مشابه توزیعهای موجود در K، توزیعهای جانشین برای H نامیده می‌شوند، بنابراین K کلاس توزیعهای جانشین است. فرض کنید تصمیم‌های قبول یا رد H بترتیب با d0 و d1 نمایش داده شده باشند. یک روش آزمون غیر تصادفی به هر مقدار x از X یکی از دو تصمیم d0 یا d1 را منتسب می‌کند و به موجب آن فضای نمونه را به دو ناحیهٔ S0 و S1 تقسیم می‌کند.

اگر x در S0 قرار بگیرد، فرض پذیرفته می‌شود، در غیر اینصورت، فرض رد می‌شود. مجموعهٔ S0 "ناحیهٔ

پذیرش" و مجموعهٔ S1 "ناحیه‌رد" یا "ناحیهٔ بحرانی" نامیده می‌شود. وقتی که آزمونی انجام می‌گیرد ممکن است تصمیم درستی اتخاذ شود یا اینکه شخص مرتکب یکی از دو خطای زیر شود:

الف: رد فرض وقتی که درست است. (خطای نوع اول)

ب : قبول فرض وقتی که غلط است. (خطای نوع دوم)

عواقب این دو خطا اغلب کاملاً متفاوت هستند، برای مثال اگر وجود یک نوع بیماری در شخص مورد آزمون قرار گیرد به غلط تصمیم بر لزوم معالجه گرفته شود، ممکن است موجب ناراحتی و زیان مالی بیمار شود، از سوی دیگر عدم تشخیص وجود بیماری ممکن است به مرگ بیمار منتهی شود.

مطلوب آن است که آزمون به روشی انجام گیرد تا احتمالهای این دو نوع خطا در حدّ کمترین مقدار نگه داشته شود. متأسفانه هنگامیکه تعدا مشاهدات ثابت است، هر دو احتمال را بطور همزمان نمی‌توان کنترل کرد. بنابراین مرسوم شده است برای احتمال (رد) به غلط رد کردن ~~فرض~~ H یک کران منتسب کرده و با این شرط احتمال دیگر را کمینه کرد. بنابراین ابتدا عددی بین صفر و یک انتخاب می‌شود که "سطح معنی داری" نامیده می‌شود و شرط زیر اعمال می‌شود:

$$P_\theta\{\delta(X)=d1\} = P_\theta\{X\in S_1\} < \alpha \qquad \forall\theta\in\Omega_H \qquad (۱-۱-۱)$$

با این شرط، باید عبارت $P_\theta\{\delta(X)=d_0\}$ $\forall\theta\in\Omega_K$ کمینه شود یا

بطور هم ارز عبارت $P_\theta\{\delta(X)=d1\} = P_\theta\{X\in S_1\}$ $\forall\theta\in\Omega_K$ (۱-۱-۲)

بیشینه شود.

اگر چه (۱-۱-۱) معمولاً

$$\sup_{\theta\in\Omega_H} P_\theta\{X\in S_1\} = \alpha \qquad (۱-۱-۳)$$

را نتیجه می‌دهد، مناسب است تا عبارتی برای سمت چپ رابطهٔ (۱-۱-۳) معرفی شود، که "اندازه آزمون" یا "اندازه ناحیهٔ بحرانی S1" نامیده می‌شود. بنابراین شرط (۱-۱-۱) توجه را به آزمونهایی محدود می‌کند که اندازه‌شان از سطح معنی داری داده شده تجاوز نمی‌کند. احتمال رد (۱-۱-۲) که برای θ ی داده شده‌ای در

پذیرش مجموعهٔ S1 "ناحیه‌رد" یا "ناحیهٔ بحرانی" نامیده می‌شود. وقتی که آزمونی انجام می‌گیرد ممکن است تصمیم اتخاذ شود یا اینکه شخص مرتکب یکی از دو خطای زیر شود:

رد فرض وقتی که درست است. (خطای نوع اول)

قبول فرض وقتی که غلط است. (خطای نوع دوم)

این دو خطا اغلب کاملاً متفاوت هستند، برای مثال اگر وجود یک نوع بیماری در شخص مورد آزمون قرار گیرد و غلط تصمیم بر لزوم معالجه گرفته شود، ممکن است موجب ناراحتی و زیان مالی بیمار شود، از سوی دیگر عدم تشخیص وجود بیماری ممکن است به مرگ بیمار منتهی شود.

هدف آن است که آزمون به روشی انجام گیرد تا احتمالهای این دو نوع خطا در حدّ کمترین مقدار نگه داشته شود. متأسفانه هنگامیکه تعداد مشاهدات ثابت است، هر دو احتمال را بطور همزمان نمی‌توان کنترل کرد. بنابراین رسم شده است برای احتمال ردّ به غلط فرض H یک کران منتسب کرده و با این شرط احتمال دیگر را کمینه کنند. برای این ابتدا عددی بین صفر و یک انتخاب می‌شود که "سطح معنی داری" نامیده می‌شود و شرط زیر اعمال شود

$$P_\theta\{\delta(X)=d_1\} = P_\theta\{X\in S_1\} \leq \alpha \qquad \forall \theta\in\Omega_H \qquad (۱-۱-۱)$$

با این قید عبارت $P_\theta\{\delta(X)=d_0\}$ $\forall\theta\in\Omega_K$ کمینه شود یا

بطور معادل عبارت (۱-۱-۲) $\quad P_\theta\{\delta(X)=d_1\} = P_\theta\{X\in S_1\} \qquad \forall\theta\in\Omega_K$

بیشینه شود.

اگر چه (۱-۱-۱) معمولاً (۱-۱-۳) $\quad \sup_{\theta\in\Omega_H} P_\theta\{X\in S_1\}=\alpha$

را نتیجه می‌دهد، مناسب است تا عبارتی برای سمت چپ رابطهٔ (۱-۱-۳) معرفی شود، که "اندازه آزمون" یا "اندازه ناحیه بحرانی S1" نامیده می‌شود. بنابراین شرط (۱-۱-۱) توجه را به آزمونهایی محدود می‌کند که اندازه‌شان از سطح معنی داری داده شده تجاوز نمی‌کند. احتمال رد (۱-۱-۲) که برای θ ی داده شده‌ای در

محاسبه می‌شود "توان آزمون" در برابرθی جانشین نامیده می‌شود.

احتمال (۲-۱-۱) بعنوان تابعی از θ به ازای هر θ متعلق به Ω "تابع توان آزمون" خوانده می‌شود. انتخاب سطح معنی داری α معمولاً قدری دلخواه است، چراکه در اغلب مواردکران دقیق تحملی برای احتمال خطای نوع اول وجود ندارد. مقادیر استانداردی نظیر ۰/۰۱ و ۰/۰۵ بطور اصولی به منظور خلاصه نویسی در جداولی که برای آزمونهای گوناگون مورد نیاز است انتخاب گردیده‌اند. بر حسب عادت و برای رعایت استاندارد در تهیه چارچوب مشترک، این مقادیر بتدریج بعنوان سطوح مناسب مورد قبول عام قرار گرفته‌اند.

این جای تأسف است، چراکه انتخاب سطح معنی داری بر توان آزمون که تحت فرض جانشین اتخاذ می‌شود تأثیر می‌گذارد. در آزمایشی که شانس آشکار سازی اثر وجود امر مورد بررسی (در صورت وجود) کم است، نکات قابل ملاحظه‌ای نمی‌توان یافت. تحقیقاتی که توسط [1] کُهن (۱۹۶۲) و [2] فریمن (۱۹۷۸) و همکاران صورت گرفته است، نشان می‌دهد که در حقیقت این حالت در بسیاری از مطالعات رخ می‌دهد.

در حالت ایده‌آل، با افزایش حجم نمونه به سطح معنی داری و توان، این اجازه داده می‌شود تا مقادیر مطلوبی اتخاذ نمایند. اگر این امر عملی نباشد می‌توان از α ی بزرگتر از آنچه که مرسوم است استفاده کرد.

امکان دیگر، در حالتی که توان آزمون بسیار نزدیک به یک باشد، آن است که بدون آنکه زیان قابل توجهی به توان وارد شود، α را بطور محسوسی کاهش داد.

قواعدی برای انتخاب α بادر نظر گرفتن توان قابل دسترسی، بوسیله [3] لهمن (۱۹۵۶)، [4] آرُو (۱۹۶۰)، ساناتانان [5] (۱۹۷۴) و از نقطه نظر بیزی توسط [6] سویج (۱۹۶۲) بحث شده است، همچنین [7] روزنتال و روبین [8] (۱۹۶۲) را نگاه کنید.

۱- cohen
۲- Freiman
۳- Lehmann
۴- Arrow
۵- sanathanan
۶- Savage
۷- Rosenthal
۸- Rubin

موضوع دیگری که در تعیین سطح معنی داری می‌تواند مورد ملاحظه قرار گیرد، طرز تلقی، نسبت به فرض، قبل از انجام آزمایش است. اگر فرد اعتقاد راسخی به درستی فرض داشته باشد(قبل از اینکه فرد به چنین اعتقادی برسد بینهایت مدرک متقاعد کننده لازم است) بر طبق آن سطح معنی داری در سطح بسیار کوچکی قرار خواهد گرفت.

(یک سطح معنی داری کوچک بدین معناست که فرض فقط به ازای مجموعه‌ای از مقادیر مشاهده شده که احتمال آنها تحت فرض کوچک است رد می‌شود. بنابراین اگر فرض H درست باشد، شانس رخداد این مقادیر بسیار کم خواهد بود.) در عمل معمولاً خانواده‌ای از نواحی رد تو در تو متناظر با سطوح معنی داری گوناگون در دسترس است.

بنابراین شیوه خوبی خواهد بود تا نه تنها رد یا قبول فرض، با سطح معنی داری داده شده‌ای تعیین شود، بلکه کوچکترین سطح معنی داری $\hat{\alpha}=\hat{\alpha}(x)$ ، احتمال معنی داری یا P- مقدار،که احتمال رد فرض تحت مشاهده است نیز مشخص شود.

این عدد ایده‌ای از اینکه مشاهدات به چه میزان فرض را نقض می‌کنند ارائه می‌دهد و دیگران را قادر می‌سازد تا بر اساس سطح معنی داری مورد نظرشان قضاوت کنند.

برای سئوالات مختلف و برخی تعمیم‌های این مفهوم به [۱] دمپستر و [۲] شاتزوف (۱۹۶۵)، [۳] استون (۱۹۷۹)،[۴] گیبنز و [۵] پرات (۱۹۷۵) ،[۶] کاکس (۱۹۷۷) پرات وگیبنز (۱۹۸۱) و تامپسون [۷](۱۹۸۵ مراجعه شود.

رفتار بزرگ نمونه ای P- مقدار در [۸] لمبرت و [۹] هال (۱۹۸۲) و حساسیت آنها نسبت به تغییرات مدل

۱- Dempster
۲- schatzoff
۳- stone
۴- Gibbons
۵- Pratt
۶- COX
۷- Thompson
۸- lambert

در لمبرت (۱۹۸۲) بحث شده است. روشهای نموداری برای ارزیابی P مقدار آزمونهای بطور هم زمان چند فرض مختلف در [۱۰] شود و [۱۱] اسپیتوال (۱۹۸۲) مورد بررسی قرار گرفته است.

نوع خاصی از آزمونهای فرض، آزمونهای تصادفی هستند. چنین آزمونهایی برای هر مقدار x یکی از دو تصمیم رد یا قبول فرض را انتخاب می‌کند، اتخاذ هر یک ، احتمال‌هایی که به x بستگی دارند و بترتیب با $\Phi(x)$ و $1-\Phi(x)$ نمایش داده می‌شوند. اگر مقدار X برابر x باشد یک آزمایش تصادفی با دو رخداد ممکن R و $\bar{R}$ بترتیب هر یک با احتمال‌های $\Phi(x)$ و $1-\Phi(x)$ انجام می‌گیرد.

اگر در این آزمایش R رخ دهد فرض رد می‌شود و در غیر اینصورت فرض پذیرفته می‌شود. بنابراین یک آزمون تصادفی بطور کامل بوسیلهٔ تابع Φ (تابع بحرانی) مشخص می‌شود بطوریکه: برای تمام xها $0 \leqslant \Phi(x) \leqslant 1$ اگر Φ تنها مقادیر ۰ یا ۱ را اتخاذ نماید به حالت آزمون غیر تصادفی بر می‌گردیم.

مجموعه نقاط x بطوریکه $\Phi(x)=1$ باشد، ناحیه رد را تشکیل می‌دهند، بنابراین یک آزمون غیر تصادفی، یک تابع نشانگر برای ناحیه بحرانی است:

$$\Phi(x) \equiv I_{S_1}(x) = \begin{cases} 1 & x \in S_1 \\ 0 & x \notin S_1 \end{cases}$$

اگر x دارای توزیع P_θ و Φ تابع بحرانی باشد، احتمال رد H عبارتست از:

$$E_\theta\left[\Phi(x)\right] = \int \Phi(x)\, dP_\theta(x)$$

۹- Hall　　　۱۰- schweder

۱۱- spotvoll

یعنی از احتمال شرطی $\Phi(x)$ بشرط x نسبت به توزیع احتمال x انتگرال گرفته شده است.

معمولاً Φ طوری انتخاب می‌شود که توان آن بشرطی که:

$$\beta_\Phi(\theta) = E_\theta\left[\Phi(x)\right] \qquad \forall\theta \in \Omega_K \qquad (۱-۱-۴)$$

$$E_\theta\left[\Phi(x)\right] \leqslant \alpha \qquad \forall\theta \in \Omega_K \qquad (۱-۱-۵)$$

بیشینه گردد .

اشکالاتی چند در بدست آوردن آزمون بهینه بروز می‌کند، بطور مثال:

آزمونی که توان را در برابر یک فرض جانشین خاص در K ی مرکبی بیشینه می‌کند، به آن فرض جانشین بستگی دارد، بنابراین برای تعریف آنچه که به آزمون بهینه معروف است، باید اصول اضافی دیگری پیشنهاد کرد.

یک استثناء مهم: اگر K تنها شامل یک توزیع باشد، یعنی اگر شخصی با یک فرض جانشین سروکار داشته باشد، مسئله توسط (۴-۱-۱) و (۵ - ۱-۱) بطور کامل مشخص شده است. در اینصورت مسئله به مسئله ریاضی بیشینه کردن یک انتگرال تحت شرایط مرزی تبدیل می‌شود. که توصیف این مسئله و کاربردهایش موضوع اصلی این فصل را تشکیل می‌دهد. البته در حالت خاص ممکن است آزمون یکسانی برای تمام فرضهای مقابل در K ، حتی بیش از یکی ،که توان را بیشینه می‌کند وجود داشته باشند. چنین آزمونهای به طور یکنواخت ـ تواناترین، در قسمتهای بعدی ارائه خواهد شد.

در فرمولبندی بالا می‌توان با مسئله مانند حالت خاصی از مسئله کلی تصمیم، با دونوع تابع زیان، برخورد کرد و متناظر با دو نوع خطا، توابع زیان دو مؤلفه‌ای زیر را پیشنهاد کرد:

$$L_1(\theta, d_1) = \begin{cases} 1 & \theta \in \Omega_H \\ 0 & \theta \in \Omega_K \end{cases}$$

$$L_1(\theta, d_0) = 0 \qquad \forall\theta \in \Omega$$