بسم الله الرحمن الرحیم

دانشکده علوم

گروه آمار

عنوان پایان نامه

ترکیب محدب آماره های یوی دو نمونه ای

برای دریافت درجه کارشناسی ارشد

در رشته آمار ریاضی

مولف

فاطمه حبیبی

استاد راهنما

دکتر نرگس عباسی

استاد مشاور

دکتر مینا توحیدی

ماه و سال

بهمن ماه ۱۳۸۸

نمی دانم که در طرح بزرگ خدا من چه نقشی دارم و چه سرنوشتی؟ ولی این قدر مطمئنم که بی هیچ نیست.

الهی ای خداوند شب و روز   ز شوق بندگی جانم بیفروز

خداوندا برافروزان روانم   بزن از عشق خود آتش به جانم

تقدیم به

پدر و مادر بزرگوارم

آنان که در مقابل سختی‌ها و ناملایمات روزگار ماندند و از عدسی کوچک دوربین زندگی، هزاران پنجره‌ی تماشا و کهکشانی از ستاره‌های امید و پویایی را به رویم گشودند و با واژه‌ای به بلندای همت و صمیمیت، خویشتن را برای آرامش و سربلندیم صرف کردند؛ دستشان را می‌بوسم.

خواهر گلم: فرشته

ستاره‌ی پرفروغ زندگیم؛ او که وجودش برای من مایه‌ی آرامش و مهربانی است.

برادران مهربانم: بهروز و بهنام و صادق

که نگاه پرمهرشان، امید زندگیم است و راهنمایی‌هایشان مسیر راهم را روشن می‌کند.

اکنون که پایان نامه ام را با موفقیت گذرانده ام، با تمام وجود از الهه ی دانش و دانایی، خدای مهربانم، شکرگزاری می کنم و بر خود لازم می دانم از استاد دوست داشتنی ام سرکار خانم دکتر نرگس عباسی، که با تعهد و عشق فراوان، نه تنها در مقام استاد راهنمایی پایان نامه ام، بلکه در طول دوره ی کارشناسی ارشد با ارشادهای ارزنده ی خویش اینجانب را در مسیر صحیح دانش آموختگی ام هدایت کردند، بی نهایت قدردانم و بهترین ها را برایشان آرزومندم.

از سرکار خانم دکتر مینا توحیدی، که با قبول مشاوره ی پایان نامه ام و با در اختیار قرار دادن دانششان، مرا مرهون لطف خویش قرار دادند، بی نهایت سپاسگزارم و تندرستی شان را از درگاه الهی خواهانم.

از جناب آقای دکتر عبدالرضا بازرگان لاری، از بزرگان آمار ایران، قدردان و متشکرم و برای من افتخاری بود که ایشان داوری پایان نامه ام را قبول کردند.

از کلیه ی مدرسان دوران ۱۱ ساله ی تحصیلم، به ویژه معلم آغازگر راهم سرکار خانم ملیحه رضوانی، کلیه ی اساتید دانشگاه شهید چمران اهواز، به ویژه دکتر بابک بابادی و از اساتید این مرکز، به ویژه دکتر نرگس عباسی که از محضرشان علم آموختم؛ بی نهایت سپاسگزار و قدردانم.

و در پایان از دوستان خوبم: صدیقه همتی، آمنه ناجو، ستاره زندگانی، فریبا گودرزی، راحیل توکلی، زهره علیزاده، نسرین نظری، فاطمه فلاحی، الهام داورپناه، سعیده پولادیان، کریم همتی، ایمان حفیظی و کامران قادری که همراهان واقعی دوران دانشجویی بودند، بسیار ممنونم و سلامت و سرافرازیشان را از خداوند منان خواستارم.

چکیده

در بحث برآوردیابی نقطه‌ای، توجه به برآوردگرهای ناریب با کمترین واریانس همواره بخش مهمی را تشکیل می‌دهد. با پیدایش نظریه‌ی آماره‌ی یو، این ویژگی در برآوردگرهایی از نوع آماره‌ی یو دیده شد. در این پایان‌نامه، به مطالعه‌ی گسترده‌ای از آماره‌های یو می‌پردازیم و مباحثی از قبیل معرفی آماره‌های یو، وی، ال‌بی، ترکیب محدب آماره‌های یو، کارایی مخاطره حدی، تعیین توزیع حدی و ارائه‌ی بسط‌مجانبی و بسط اجورث آن‌ها در حیطه‌ی تحقیق‌مان قرار می‌دهیم.

فهرست

پیشگفتار

با شروع از یک برآوردگر نااریب که در آن از کمترین مشاهدات استفاده می‌شود، آماره‌ی یو تشکیل می‌گردد. مبنای تئوری آماره‌ی یو اولین بار توسط هافدینگ[1] (۱۹۴۸) گسترش داده شد و شرح جزئیات موضوع در لی[2] (۱۹۹۰) آمده است.

ویژگی خاصی که در این گروه از برآوردگرها ظاهر می‌شود، منجر به تغییر در معیار شناسایی توزیع‌ها، آماره‌های آزمون پارامتری و ناپارامتری شده است. هدف اصلی این پایان‌نامه، ارائه‌ی برآوردگرهایی به‌صورت فرمول‌هایی است که قابلیت کاربرد به‌صورت مستقیم را دارند.

این پایان‌نامه در چهار فصل تنظیم شده است. در فصل اول، آماره‌های یو، وی و ال‌بی در حالات یک و دونمونه‌ای معرفی می‌شوند و مثال‌هایی از آن‌ها بیان می‌گردد. در فصل دوم، ترکیب محدب آماره‌های یو را در حالات یک و دونمونه‌ای مورد بحث قرار می‌دهیم، که هدف آن معرفی رابطه‌ی خطی بین آماره‌های وی و ال‌بی با آماره‌ی یو می‌باشد و در ادامه، کارایی مخاطره حدی را برای مقایسه‌ی آماره‌های ذکر شده از لحاظ کارا بودن، در حالت یک‌نمونه‌ای بیان می‌کنیم. در فصل سوم، یک بسط مجانبی از آماره‌ی Y_{n_1,n_2} ارائه می‌دهیم. در فصل چهارم، بسط اجورث[3] آماره‌های U_{n_1,n_2} و Y_{n_1,n_2} را معرفی خواهیم کرد و با استفاده از این بسط، اختلاف بین توزیع‌های مجانبی این دو آماره را تعریف می‌کنیم.

[1] Hoeffding
[2] Lee
[3] Edgeworth

فصل اول

آماره‌ی یو و سایر آماره‌ها

۱-۱ مقدمه

در این فصل ابتدا برآوردپذیری پارامتر θ تعریف شده و سپس برآوردگرهای مشهور θ، شامل آماره‌های یو، وی و ال‌بی را در حالت یک‌نمونه‌ای معرفی کرده و برای آشنایی بیشتر مثال‌هایی ارائه می‌شود. سپس برآوردگرهای θ برای برآوردیابی پارامترهای دو جامعه مورد بررسی قرار می‌گیرند و در پایان فصل نیز مثال‌هایی برای بیان مفهوم آزمون‌های مقایسه‌ی دو توزیع مطرح می‌گردد.

۱-۲ برآوردگرهای θ

فرض کنید $\mathcal{P}$ یک خانواده از اندازه‌های احتمال بر روی یک فضای دلخواه قابل اندازه‌گیری باشد. تئوری کلی یک راه‌کار برای مسائل ناپارامتری ارائه می‌دهد؛ یعنی با خانواده‌ی خاصی روبرو نیستیم، به‌عبارت دیگر $\mathcal{P}$ از یک خانواده‌ی بزرگ توزیع‌ها گرفته می‌شود و فقط ممکن است با این محدودیت که پیوسته باشند یا گشتاورها موجود باشد، مواجه باشیم.

قبل از ارائه‌ی برآوردگرهای پارامتر θ در حالت یک‌نمونه‌ای، تعریف زیر برای برآوردپذیری آن ارائه می‌شود. (هالموس[۴] (۱۹۴۶))

تعریف

$\theta(P)$ یک پارامتر برآوردپذیر نسبت به اندازه‌ی $\mathcal{P}$ است، اگر به‌ازای حداقل k متغیر تصادفی مستقل و هم‌توزیع (مثلاً از توزیع $P\epsilon\mathcal{P}$)، تابع حقیقی اندازه‌پذیر $h(x_1, \dots, x_k)$ وجود داشته باشد به‌طوری که به‌ازای $P\epsilon\mathcal{P}$

$$\theta(P) = E\big(h(X_1, \dots, X_k)\big) = \int_{-\infty}^{\infty} h(x_1, \dots, x_k) \prod_{i=1}^{k} dP(x_i) \qquad (1)$$

در تعریف فوق k درجه‌ی $\theta(P)$ و تابع h هسته‌ی θ معرفی می‌شود و واضح است که $\mathcal{P} = \{P: |\theta(P)| < \infty\}$.

نکته‌ای که باید به آن توجه کنیم این است که در این‌جا تابع h متقارن و ناااریب فرض می‌شود. در غیر این‌صورت اگر f یک برآوردگر نااریب نامتقارن برای $\theta(P)$ باشد، آن‌گاه متوسط f که برای همه‌ی جایگشت‌های متغیرهای به‌کار برده شده محاسبه می‌گردد، آن‌را تبدیل به برآوردگری نااریب و متقارن می‌سازد.

$$h(x_1, \dots, x_k) = \frac{1}{k!} \sum_{\pi\epsilon\Pi_k} f\big(x_{\pi_1}, \dots, x_{\pi_k}\big) \qquad (2)$$

[۴] Halmos

در این‌جا Π_k مجموعه‌ای متشکل از کلیه‌ی جایگشت‌های k متغیر تصادفی در نظر گرفته شده است و جمع بر روی Π_k برای همه‌ی جایگشت‌های موجود بسته می‌شود. پس بدون این‌که از کلیت مسئله کم شود h را متقارن فرض می‌کنیم، مگر در مواقع خاص که ذکر خواهد شد.

۱-۲-۱ معرفی برآوردگرهای θ در حالت یک‌نمونه‌ای

آماره‌های یو، وی و ال‌بی از برآوردگرهای مشهور θ هستند. برای یک تابع حقیقی اندازه‌پذیر $h(x_1, \dots, x_k)$ و برای نمونه‌ی تصادفی $X_1, \dots, X_n$ $(n > k)$ از توزیع $\mathcal{P}$ برآوردگرهای θ با هسته‌ی متقارن h به‌صورت زیر ساخته می‌شوند.

الف) آماره‌ی یوی یک‌نمونه‌ای

$$U_n = \frac{1}{\binom{n}{k}} \sum_{1 \le i_1 < \dots < i_k \le n} h(X_{i_1}, \dots, X_{i_k}) \qquad (۳)$$

که در آن $\sum_{1 \le i_1 < \dots < i_k \le n}$، جمع‌بندی روی تمام اعداد صحیح $i_1, \dots, i_k$ای را نشان می‌دهد که در رابطه‌ی $1 \le i_1 < \dots < i_k \le n$ صدق می‌کنند.

آماره‌ی یوی مرتبط با هسته‌ی h، از میانگین‌گیری هسته‌ی h روی تمام آرایش‌های نامنظم k تا از Xهای مجزای انتخاب‌شده‌ی بدون جایگذاری از نمونه‌ی $X_1, \dots, X_n$ حاصل می‌شود.

ب) آماره‌ی وی یک‌نمونه‌ای

$$V_n = \frac{۱}{n^k} \sum_{i_۱=۱}^{n} \dots \sum_{i_k=۱}^{n} h(X_{i_۱}, \dots, X_{i_k}) \quad (۴)$$

آماره‌ی وی مرتبط با هسته‌ی h، از میانگین‌گیری هسته‌ی h روی تمام آرایش‌های منظم k تا از Xهای انتخاب‌شده‌ی با جایگذاری از نمونه‌ی $X_۱, \dots, X_n$ حاصل می‌شود (لی ۱۹۹۰).

پ) آماره‌ی ال‌بی یک‌نمونه‌ای

$$B_n = \frac{۱}{\binom{n+k-۱}{k}} \sum_{r_۱+\dots+r_n=k} h(\underbrace{X_۱, \dots, X_۱}_{r_۱}, \dots, \underbrace{X_n, \dots, X_n}_{r_n}) \quad (۵)$$

که در آن $\sum_{r_۱+\dots+r_n=k}$، جمع‌بندی روی تمام اعداد صحیح نامنفی $r_۱, \dots, r_n$ای را نشان می‌دهد که در رابطه‌ی $r_۱ + \dots + r_n = k$ صدق می‌کنند. (اصطلاح آماره‌ی ال‌بی، مخفف حد برآورد بیزی[۵] است.)

حال ترکیب‌های با تکرار را بررسی می‌کنیم.آماره‌ی ال‌بی مرتبط با هسته‌ی h، از میانگین‌گیری هسته‌ی h روی تمام آرایش‌های منظم k، X انتخاب‌شده و تکرار مجاز از نمونه‌ی $X_۱, \dots, X_n$ حاصل می‌شود.

این برآوردگر از برآورد بیز $\theta(F)$ی که روی فرآیند دیریخله[۶] بنا شده و پارامترش مجاز است به‌سمت صفر میل کند، حاصل می‌شود:

فرض کنید که توزیع پیشین برای $\theta(F)$، فرآیند دیریخله با پارامتر α روی $(\mathcal{R}, \mathcal{B})$ باشد، که در آن $\mathcal{B}$ یک σ-میدان از مجموعه‌های بورل[۷] می‌باشد. فرض کنید برآورد بیز $\theta(F)$ بنا شده روی

---

[۵]Limit of Bayes estimate

[۶]Dirichlet

[۷]Borel

تابع زیان مربع خطا[8]، $\hat{\theta}(F)$ باشد. هنگامی‌که $\alpha(\cdot) \to \infty$، $\hat{\theta}(F)$ حد برآورد بیز را ایجاد می‌کند، که همان سمت راست رابطه‌ی (۵) است (یاماتو[9] ($1977b$)).

آماره‌ی ال‌بی ارائه شده در رابطه‌ی (۵) را می‌توان به‌صورت زیر، نیز نوشت

$$B_n = \binom{n+k-1}{k}^{-1} \tag{۶}$$

$$\times \sum_{j=1}^{k} \sum^{+}_{r_1+\cdots+r_j=k} \sum_{1\le i_1<\cdots<i_j\le n} h(\underbrace{X_{i_1},\dots,X_{i_1}}_{r_1},\dots,\underbrace{X_{i_j},\dots,X_{i_j}}_{r_j})$$

که در آن $\sum^{+}_{r_1+\cdots+r_j=k}$، جمع‌بندی روی اعداد صحیح مثبت $r_1,\dots,r_j$ای را نشان می‌دهد که در رابطه‌ی $r_1+\cdots+r_j=k$ صدق می‌کنند (یاماتو ($1977b$)). این رابطه به این معنی است که آماره‌ی ال‌بی به‌صورت ترکیبی خطی از آماره‌ی یو نوشته می‌شود، که در فصل دوم مورد بررسی قرار خواهد گرفت.

در پیوست (الف)، به توضیحی در مورد وزن‌های آماره‌های یو، وی و ال‌بی با توجه به حدود جمع‌های ذکر شده، پرداخته می‌شود.

مثال۱

فرض کنید $\mathcal{P}$ مجموعه‌ی توزیع‌های حقیقی با میانگین متناهی باشد. میانگین

---

[8]Mean squat error

[9]Yamato

$$\mu = \mu(P) = \int x dP(x)$$

یک پارامتر برآوردپذیر با درجه‌ی ۱ است، زیرا $f(x_1) = x_1$ یک برآورد نااریب برای میانگین است

$$E(X_1) = \mu$$

پس $h(x_1) = x_1$ یک هسته برای میانگین تعریف می‌شود. بر طبق محاسبه، آماره‌های یو، وی و ال‌بی میانگین نمونه‌ای به‌صورت زیر حاصل می‌گردند.

الف) آماره‌ی یو

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{1}} \sum_{i=1}^{n} X_i \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \bar{X}_n \end{aligned}$$

ب) آماره‌ی وی

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \bar{X}_n \end{aligned}$$

پ) آماره‌ی ال‌بی

$$\begin{aligned} B_n &= \frac{1}{\binom{n+1-1}{1}} \sum_{1 \le i \le n} h(\underbrace{X_i, \dots, X_i}_{1}) \\ &= \frac{1}{n} \sum_{i=1}^{n} X_i \\ &= \bar{X}_n \end{aligned}$$

مثال۲

$\mathcal{P}$ مجموعه‌ی توزیع‌های حقیقی با گشتاور rام متناهی است. گشتاور rام

$$\mu_r = \int x^r dP(x)$$

یک پارامتر برآوردپذیر با درجه‌ی ۱ است، زیرا $f(x_1) = x_1^r$ یک برآورد نااریب برای μ_r است

$$E(X_1^r) = \mu_r$$

پس $h(x_1) = x_1^r$ یک هسته‌ی متقارن برای μ_r می‌باشد. بنابراین

الف) آماره‌ی یو

$$U_n = \frac{1}{\binom{n}{1}} \sum_{i=1}^{n} {X_i}^r$$

$$= \frac{1}{n} \sum_{i=1}^{n} {X_i}^r$$

ب) آماره‌ی وی

$$V_n = \frac{1}{n} \sum_{i=1}^{n} X_i^r$$

پ) آماره‌ی ال‌بی

$$B_n = \frac{1}{\binom{n+1-1}{1}} \sum_{1 \le i \le n} h(\underbrace{X_i, \dots, X_i}_{1})$$

$$= \frac{1}{n} \sum_{i=1}^{n} {X_i}^r$$

مثال۳

فرض کنید پارامتر $\theta(P)=\mu^۲$ مدنظر باشد. $f(x_۱,x_۲)=x_۱x_۲$ یک برآورد نااریب برای $\mu^۲$ است

$$E(X_۱X_۲)=\mu^۲$$

پس $h(x_۱,x_۲)=x_۱x_۲$ یک هسته‌ی متقارن با درجه‌ی ۲ برای $\mu^۲$ می‌باشد. بنابراین

الف) آماره‌ی یو

$$U_n=\frac{۱}{\binom{n}{۲}}\sum_{i<j}X_iX_j$$

$$=\frac{۲}{n(n-۱)}\sum_{i<j}X_iX_j$$

$$=\frac{۱}{n(n-۱)}\left[\left(\sum_{i=۱}^{n}X_i\right)^۲-\sum_{i=۱}^{n}X_i^{۲}\right]$$

$$=\frac{n}{n-۱}(\bar{X})^۲-\frac{۱}{n(n-۱)}\sum_{i=۱}^{n}X_i^{۲}$$

$$=\frac{۱}{n-۱}\left[n(\bar{X})^۲-\overline{X^۲}\right]$$

ب) آماره‌ی وی

$$V_n=\frac{۱}{n^۲}\sum_{i=۱}^{n}\sum_{j=۱}^{n}X_iX_j$$

$$=\frac{۱}{n^۲}\left(\sum_{i=۱}^{n}X_i\right)^۲$$

$$=(\bar{X})^۲$$

پ) آماره‌ی ال‌بی