وزارت علوم، تحقیقات و فناوری

دانشگاه بین‌المللی امام خمینی (ره)

IMAM KHOMEINI
INTERNATIONAL UNIVERSITY

دانشکده علوم

استفاده از اطلاعات پیشین برای بهبود براوردگرهای پارامترهای جامعه

پایان‌نامه برای دریافت درجه کارشناسی ارشد
در رشته‌ی آمار گرایش آمار ریاضی

حمید خوش‌ترکیب

استاد راهنما:

دکتر افشین فلاح

آذر ماه ۱۳۹۲

وزارت علوم، تحقیقات و فناوری

دانشگاه بین‌المللی امام خمینی (ره)

IMAM KHOMEINI INTERNATIONAL UNIVERSITY

دانشکده علوم

گروه آمار

استفاده از اطلاعات پیشین برای بهبود براوردگرهای پارامترهای جامعه

پایان‌نامه برای دریافت درجه کارشناسی ارشد در رشته‌ی آمار

گرایش آمار ریاضی

حمید خوش‌ترکیب

استاد راهنما:

دکتر افشین فلاح

استاد مشاور:

دکتر اسماعیل امیری

آذر ماه ۱۳۹۲

تقدیم به

پدر و مادر عزیز و مهربانم

که در سختی ها و دشواری های زندگی

همواره یاوری دلسوز و فداکار و پشتیبانی محکم و مطمئن برایم بوده اند

سپاس گزاری...

دستی که گشاده است
می‌برد،
می‌آورد،
رهنمونت می‌شود
به خانه‌ئی که
نور دلچسبش گرمی‌بخش است.

سپاس خدایی راکه اول است و پیش از او اولی نبوده و آخر است و آخری نباشد.

منت خدای عزوجل بر من است که توفیق راه‌یابی به وادی علم را عطا فرمود. اکنون که محقق شده است رسیدن به پایان سر فصل، بر خود واجب می‌دانم که ارج نهم بزرگانی راکه آورده‌های شمع وجودشان روشنگر راهم است.

سپاس بی‌شائبه‌ی خود را تقدیم می‌دارم به محضر استاد ارجمندم جناب آقای دکتر افشین فلاح که همواره رهنمودهایشان یاری دهنده‌ی مشکلات علمی‌ام بوده است. تقدیر و تشکر خود را از استاد مشاور جناب آقای دکتر اسماعیل امیری صمیمانه ابراز می‌دارم. از استاد گرانقدر، جناب آقای دکتر رامین کاظمی جهت داوری این پایان نامه کمال تشکر را دارم.

در پایان از پدر و مادر مهربانم کمال تشکر را دارم.

حمید خوش ترکیب

آذر ماه ۱۳۹۲

چکیده

مسأله براورد یکی از جنبه‌های اصلی استنباط آماری است. متناظر با هر پارامتر براوردگرهای متعددی وجود دارند و در بسیاری از مسائل دست‌یابی به براوردگری که از کارایی مطلوب برخوردار باشد، مشکل است. در این پایان‌نامه، مسأله بهبود براوردگرهای پارامترهای جامعه با استفاده از اطلاعات پیشین مورد توجه قرار گرفته است. با توجه به این‌که اغلب برای یک پارامتر بیش از یک براوردگر وجود دارد، از این‌رو براوردگرهایی تحت عنوان براوردگرهای ترنجیده بهبودیافته به شکل ترکیب بهینه جفت براوردگرها ارائه و تعمیم‌هایی در این‌باره پیشنهاد شده است. سپس، براوردگرهای بهبودیافته تحت روش نمونه‌گیری مجموعه رتبه‌دار مورد بحث قرار گرفته و در این راستا با فرض معلوم بودن ضریب تغییرات، براوردگرهای بهبودیافته برای میانگین و واریانس جامعه تحت این روش ارائه شده است. به‌علاوه، چون پارامترهای میانگین و واریانس تنها پارامترهای مورد علاقه نیستند، مسأله یافتن براوردگر بهبودیافته در حالت کلی و برای هر تابعی از پارامتر جامعه نیز مطرح و راه حل مناسب پیشنهاد شده است. در نهایت، روش‌های مورد بحث به منظور ارائه براورد بهبودیافته پارامترهای میانگین و واریانس درامد خانوارهای روستایی در ایران تحت روش نمونه‌گیری تصادفی ساده و مجموعه رتبه‌دار به‌کار گرفته شده‌اند.

واژه‌های کلیدی: براوردگر بهبود یافته؛ نمونه‌گیری مجموعه رتبه‌دار؛ کارایی؛ میانگین توان دوم خطا؛ ضریب تغییرات.

فهرست مطالب

فهرست جدول‌ها

فهرست شکل‌ها

فصل ۱

مروری بر مفاهیم پایه

۱.۱ مقدمه

جوامع آماری شامل ویژگی‌های نامعلومی هستند که از آن‌ها تحت عنوان پارامترهای جامعه یاد می‌شود. براورد پارامترهای جامعه از عمده‌ترین اهداف در قلمرو استنباط آماری است. مسأله براورد با استفاده از اطلاعات کمکی از مسائل جالب و پرکاربرد در زمینه استنباط آماری است. در این‌گونه مسائل، هدف براورد پارامتری از جامعه است در حالی‌که اطلاعاتی کمکی درباره پارامتر دیگری از جامعه در دسترس است. استفاده از اطلاعات پیشین برای براورد پارامترهای جامعه می‌تواند به براوردگرهایی با دقت بیشتر بیانجامد. نوع اطلاعات پیشین با توجه به توزیع جامعه، پارامتر مورد نظر و سایر ویژگی‌های جامعه و پارامتر متغیر است. اطلاعات پیشین اغلب در قالب معلوم بودن کمیت‌هایی از جامعه مانند ضریب تغییرات، کشیدگی، چولگی و غیره نمود پیدا می‌کند. در این راستا تشخیص نوع اطلاعات پیشینی که می‌تواند مورد استفاده قرار بگیرد، از اهمیت زیادی برخوردار است. برای دستیابی به براوردگرهای بهبود یافته روش‌های متعددی مطرح شده است که بسته به تعداد براوردگرهای موجود و نوع پارامتر مورد علاقه تفاوت‌ها و شباهت‌هایی دارند. این مسأله اولین بار توسط سیرلز (۱۹۶۴) برای براورد پارامتر میانگین، با فرض معلوم بودن ضریب تغییرات جامعه‌ای مطرح شده است.

در این مطالعه مسأله براوردیابی به‌عنوان یک مسأله تصمیم مد نظر گرفته شده است. با توجه به اینکه تصمیم‌گیری بر اساس نمونه تصادفی صورت می‌پذیرد، بنابراین بر هر تصمیم نوعی عدم قطعیت حاکم است. یک تصمیم در صورتی مطلوبست که به وضعیت واقعی جامعه نزدیک باشد. میزان مطلوبیت یک تصمیم را با تابع مطلوبیت ارزیابی می‌کنند. به دلیل روحیه محافظه کارانه حاکم بر تحقیقات آماری که ناشی از بکارگیری نمونه‌ی تصادفی است، در ادبیات آماری قرینه تابع مطلوبیت که تابع زیان نامیده می‌شود، مورد استفاده قرار می‌گیرد. انتخاب تابع زیان در یک مسأله استنباط آماری امری مهم و تاثیر گذار است. بر پایه توابع زیان مختلف، ملاک‌های زیادی برای ارزیابی براوردگرها وجود دارد. از جمله ملاک‌های متداول برای مقایسه براوردگرها می‌توان به سازگاری، نااریبی، دقت، حافظ دامنه‌ی پارامتر بودن و غیره اشاره کرد.

تعریف ۱.۱.۱. اگر فرض کنیم برای یک پارامتر براوردگر اولیه‌ای وجود دارد، براوردگری را که بر اساس این براوردگر اولیه و با استفاده از اطلاعات پیشین ساخته می‌شود، براوردگر بهبود یافته گویند.

تعریف ۲.۱.۱. امید ریاضی مراتب مختلف یک متغیر تصادفی را گشتاورهای آن متغیر تصادفی گویند. r-امین گشتاور متغیر تصادفی X حول صفر به‌صورت

$$\mu_r = \mathbb{E}(X^r), \quad r \in \mathbb{N}$$

تعریف و در صورت وجود با نماد μ_r نمایش داده می‌شود. اگر برای یک مقدار صحیح و مثبت r، μ_r وجود داشته باشند، آن‌گاه $\mu_۱, \cdots, \mu_{r-۱}$ نیز وجود دارند. برای $r = ۱$، $\mu_۱$ همان میانگین جامعه است که به‌صورت

$$\mathbb{E}(X) = \mu = \int_{-\infty}^{\infty} xf(x)d\mu(x) = \begin{cases} \sum_x xf(x) & \text{گسسته } X \\ \int_{-\infty}^{\infty} xf(x)dx & \text{پیوسته } X, \end{cases}$$

تعریف می‌شود. r-امین گشتاور متغیر تصادفی X حول مقدار a به‌صورت

$$\begin{aligned} \mu_{r,a} &= \mathbb{E}[(X-a)^r] \\ &= \sum_{k=۰}^{r} \binom{r}{k} \mu_k (-a)^{r-k}, \end{aligned}$$

تعریف و در صورت وجود با نماد $\mu_{r,a}$ نمایش داده می‌شود. حالت خاص $a = \mu$ را r-امین گشتاور مرکزی متغیر تصادفی X نامیده و با نماد μ'_r نشان می‌دهند. به‌سادگی معلوم می‌شود که روابط

$$\mu'_۱ = ۰ \quad , \quad \mu'_۲ = \sigma^۲_X \quad , \quad \mu'_۳ = \alpha_۳\sigma^۳_X \quad , \quad \mu'_۴ = (\alpha_۴ + ۳)\sigma^۴_X,$$

برقرارند، که در آن‌ها $\sigma^۲_X$ دومین گشتاور مرکزی متغیر تصادفی X بوده و واریانس نامیده می‌شود، $\alpha_۳$ ضریب چولگی (معیاری برای اندازه میزان انحراف از تقارن توزیع) و $\alpha_۴$ ضریب کشیدگی (معیاری برای اندازه میزان کشیدگی یا برجستگی توزیع فراوانی) است. به‌طور خاص برای توزیع نرمال $\alpha_۳ = \alpha_۴ = ۰$ است. توجه کنید که اگر تابع چگالی X حول μ متقارن باشد، آن‌گاه همه گشتاورهای فرد X حول μ صفر است، مشروط بر آن که این گشتاورها وجود داشته باشند. نسبت انحراف معیار به میانگین را ضریب تغییرات نامیده و آن را با نماد

$$CV = \frac{\sqrt{\mathbb{V}ar(X)}}{\mathbb{E}(X)}.$$

نمایش می‌دهند.

تعریف ۳.۱.۱. فرض کنید متغیرهای تصادفی X و Y دارای توزیع توام باشند. در این صورت کواریانس متغیرهای تصادفی X و

Y که به‌صورت

$$\mathbb{C}ov(X,Y) = \sigma_{XY} \\ = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y),$$

تعریف می‌شود، معیاری برای ارزیابی رابطه خطی بین دو متغیر است. نسخه‌ی اصلاح شده این معیار که برخلاف کواریانس کراندار بوده و به واحد اندازه‌گیری وابسته نیست، ضریب همبستگی دو متغیر تصادفی X و Y نامیده می‌شود و به‌صورت

$$\rho_{X,Y} = \mathbb{C}orr(X,Y) \\ = \frac{\mathbb{C}ov(X,Y)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}} \\ = \frac{\sigma_{XY}}{\sigma_X \sigma_Y},$$

تعریف می‌شود.

استباط آماری به معنی نتیجه‌گیری در مورد جامعه براساس اطلاعات حاصل از نمونه‌ی تصادفی است. استنباط آماری در دو حالت کلی پارامتری و ناپارامتری صورت می‌پذیرد. در استنباط آماری پارامتری، شکل تابعی توزیع معلوم است اما به پارامتر یا پارامترهای نامعلومی وابسته است. در این حالت هدف استنباط آماری پارامتری، براورد پارامترهای توزیع است. این در حالی است که در استنباط ناپارامتری توزیع به‌صورت کلی نامعلوم است.

فرض کنید $X_1, \cdots, X_n$ نمونه‌ای تصادفی از خانواده توابع چگالی احتمال یا جرم احتمال $\{f_\theta(x), \theta \in \Theta\}$ باشد. تابعی از مشاهدات نمونه تصادفی به‌صورت $U = g(X_1, X_2, \cdots, X_n)$ را که به پارامتر مجهول وابسته نباشد، آماره می‌نامند. هر آماره متغیری تصادفی است و توزیع احتمالاتی خاص خود را داراست. توزیع یک آماره را توزیع نمونه‌ای گویند. توزیع نمونه‌ای یک آماره به اندازه نمونه و توزیع جامعه‌ای که نمونه از آن استخراج می‌شود وابسته است، و ممکن است به پارامتر مورد علاقه بستگی داشته یا نداشته باشد. نکته حائز اهمیت این است که عمدتاً آماره‌ای برای براورد مفیدتر است که توزیع آن به پارامتر مجهول وابسته باشد، زیرا در این صورت آن آماره درباره پارامتر مجهول حاوی اطلاعات قابل توجهی خواهد بود. هر آماره‌ای که توزیع آن به پارامتر مورد علاقه بستگی نداشته باشد، آماره‌ی کمکی نامیده می‌شود.

تعریف ۴.۱.۱. مجموعه‌ی کلیه مقادیر نمونه تصادفی را فضای مشاهدات نامیده و با نماد χ نمایش می‌دهند.

تعریف ۵.۱.۱. مجموعه‌ی کلیه مقادیر ممکن پارامتر را فضای پارامتری نامیده و با نماد Θ نمایش می‌دهند.

تعریف ۶.۱.۱. براوردگر $\delta(\mathbf{x})$ را در براورد پارامتر θ حافظه دامنه گویند، هر گاه برای هر $\theta \in \boldsymbol{\Theta}$ رابطه‌ی

$$P_\theta(\delta(X_1, \cdots, X_n) \in \boldsymbol{\Theta}) = 1,$$

برقرار باشد.

فرض می‌شود که برای هر $\theta \in \boldsymbol{\Theta}$ و هر براورد ممکن $\delta(\boldsymbol{x})$، مقدار عددی $L(\theta, \delta(\boldsymbol{x}))$-ای وجود دارد که میزان زیان آماردان را در براورد پارامتر θ بر پایه $\delta(\boldsymbol{x})$ اندازه می‌گیرد.

تعریف ۷.۱.۱. تابع زیان، تابعی دو متغیره از $\boldsymbol{\Theta} \times D$ به زیرمجموعه‌ای از اعداد حقیقی نامنفی است. یعنی

$$L(\theta, \delta) : \boldsymbol{\Theta} \times D \longrightarrow \mathbb{R}^+.$$

قابل ذکر است که $L(\theta, \delta(\boldsymbol{X}))$ خود یک متغیر تصادفی است.

در هر مسأله استنباط آماری توابع زیان متعددی را می‌توان با توجه به نوع مسأله به‌کار گرفت. برخی از توابع زیان پرکاربرد به اختصار معرفی می‌شوند

- تابع زیان توان دوم خطا که به‌صورت

$$L(\theta, \delta) = (\delta - \theta)^2,$$

تعریف می‌شود و نقطه‌ی ضعف کلی آن عملکرد یکسان آن در برابر بیش براورد و کم براورد نمودن پارامتر است.

- تابع زیان توان دوم خطای موزون که به‌صورت

$$L(\theta, \delta) = \omega(\theta)(\delta - \theta)^2, \quad \omega(\theta) > 0,$$

تعریف می‌شود و در آن $\omega(\theta)$ تابع وزن نامیده می‌شود. این تابع زیان زمانی به‌کار می‌رود که مقادیر مختلف θ در فضای پارامتری اهمیت یکسانی ندارند.

- تابع زیان قدر مطلق خطا که به‌صورت

$$L(\theta, \delta) = | \delta - \theta |,$$

تعریف می‌شود و همانند تابع زیان توان دوم خطا در برابر بیش براورد و کم براورد نمودن پارامتر عملکرد یکسانی دارد.

- تابع زیان خطی که به‌صورت

$$L(\theta,\delta)=\begin{cases} k_0(\theta-\delta), & \theta-\delta\geq 0\\ k_1(\delta-\theta), & \theta-\delta<0, \end{cases}$$

تعریف می‌شود و تعمیم‌یافته تابع زیان قدر مطلق خطاست. در حالت خاص، به‌ازای $k_0=k_1=1$، تابع زیان قدر مطلق خطا به‌دست می‌آید. این تابع زیان برای بیش براورد و کم براورد نمودن پارامتر زیان یکسانی را در نظر نمی‌گیرد.

- تابع زیان صفر-یک که به‌صورت

$$L(\theta,\delta_i)=\begin{cases} 0, & \theta\in\Theta_i\\ 1, & \theta\in\Theta_j \end{cases}\qquad i\neq j\ ,$$

تعریف می‌شود و بیشتر در آزمون فرض‌ها مورد استفاده قرار می‌گیرد.

با توجه به اینکه تابع زیان به‌دلیل وابستگی به نمونه‌ی تصادفی، متغیری تصادفی است، متوسط مقدار زیان حاصل از یک تصمیم را تابع مخاطره نامیده و به‌صورت

$$R(\theta,\delta)=\mathbb{E}_\theta[L(\theta,\delta(\boldsymbol{X}))],$$

نمایش می‌دهند. در صورتی که تابع زیان، توان دوم خطا باشد، تابع مخاطره را میانگین توان دوم خطا نامیده و به‌صورت

$$\begin{aligned}\mathbb{M}SE(\delta(\boldsymbol{X}))&=\mathbb{E}_\theta(\delta(\boldsymbol{X})-\theta)^2\\ &=\mathbb{V}ar(\delta(\boldsymbol{X}))+bias^2(\delta(\boldsymbol{X})),\end{aligned}$$

محاسبه می‌کنند، که در آن $bias(\delta)=\mathbb{E}(\delta)-\theta$ مقدار اریبی براوردگر $\delta(\boldsymbol{X})$ می‌باشد. اگر $\mathbb{E}(\delta(\boldsymbol{X}))=\theta$، آن‌گاه مقدار اریبی برابر صفر بوده و براوردگر مورد نظر نااریب خوانده می‌شود.

هرگاه تابع زیان توان دوم خطا باشد، براوردگری که دارای میانگین توان دوم خطای کوچکتری باشد، مطلوب‌تر است. اگر میانگین توان دوم خطای یک براوردگر به پارامتر مورد علاقه θ بستگی داشته باشد، آن‌گاه برای هر دو براوردگر δ_1 و δ_2 این امکان وجود دارد که منحنی‌های میانگین‌های توان دوم خطاهای $\mathbb{M}SE_\theta(\delta_1)$ و $\mathbb{M}SE_\theta(\delta_2)$ یکدیگر را قطع کنند. در چنین شرایطی برای برخی مقادیر θ براوردگر $\delta_1(\boldsymbol{X})$ و برای برخی مقادیر دیگر براوردگر $\delta_2(\boldsymbol{X})$ میانگین توان دوم خطای کوچک‌تری خواهد داشت. از این‌رو، معمولاً براوردگری که میانگین توان دوم خطای آن به‌ازای تمام مقادیر θ در فضای پارامتری (به‌طور یکنواخت) از میانگین توان دوم خطای سایر براوردگرها کوچک‌تر باشد، وجود ندارد.

فرض کنید $\delta_1(\boldsymbol{X})$ و $\delta_2(\boldsymbol{X})$ دو براوردگر برای پارامتر θ باشند. براوردگر $\delta_1(\boldsymbol{X})$ را کاراتر از براوردگر $\delta_2(\boldsymbol{X})$ گویند، هرگاه

$$\mathbb{M}SE(\delta_1) \leq \mathbb{M}SE(\delta_2) \qquad \forall\theta \in \Theta.$$

کارایی نسبی براوردگر $\delta_2(\boldsymbol{X})$ را نسبت به براوردگر $\delta_1(\boldsymbol{X})$ به‌صورت

$$0 \leq R.E.(\delta_2(\boldsymbol{X}), \delta_1(\boldsymbol{X})) = \frac{\mathbb{M}SE(\delta_1(\boldsymbol{X}))}{\mathbb{M}SE(\delta_2(\boldsymbol{X}))},$$

تعریف می‌کنند. بدیهی است در صورتی‌که براوردگرهای $\delta_1(\boldsymbol{X})$ و $\delta_2(\boldsymbol{X})$ هر دو ناریب باشند، این نسبت برابر نسبت واریانس براوردگرها خواهد بود.

در این پایان‌نامه دو خانواده بزرگ از توزیع‌ها که یکی خانواده‌ی گروهی و دیگری خانواده توزیع‌های نمایی می‌باشند، به دفعات مورد بحث قرار گرفته‌اند.

تعریف ۸.۱.۱. یک خانواده‌ی گروهی از توزیع‌ها به‌وسیله انجام یک خانواده از تبدیلات گروهی روی یک متغیر تصادفی با یک توزیع خاص حاصل می‌شود.

در ادامه سه خانواده مهم از خانوده‌ی توزیع‌های گروهی به اختصار معرفی می‌شوند.

۱- فرض کنید U یک متغیر تصادفی با توزیع احتمال $f(u)$ باشد. اگر $-\infty < \mu < +\infty$ مقداری ثابت باشد، آن‌گاه توزیع $X = U + \mu$ که به‌صورت $f_X(x) = f(x-\mu)$ است را یک خانواده مکانی از توزیع‌ها گویند.

برای مثال خانواده توزیع نرمال $N(\mu, 1)$ از تبدیلات گروهی $X = Z + \mu$ روی متغیر تصادفی $Z \sim N(0, 1)$ حاصل می‌شود.

۲- فرض کنید U متغیر تصادفی با توزیع احتمال $f(u)$ باشد. اگر $\sigma > 0$ مقداری ثابت باشد، آن‌گاه توزیع $X = \sigma U$ که به‌صورت $f_X(x) = \frac{1}{\sigma} f(\frac{x}{\sigma})$ است را یک خانواده مقیاسی از توزیع‌ها گویند.

برای مثال خانواده توزیع نمایی $Exp(\theta)$ از تبدیلات گروهی $X = \theta Z$ روی متغیر تصادفی $Z \sim Exp(1)$ حاصل می‌شود.

۳- فرض کنید U یک متغیر تصادفی با توزیع احتمال $f(u)$ باشد. اگر $-\infty < \mu < +\infty$ و $\sigma > 0$ مقادیر ثابتی باشند، آن‌گاه توزیع $X = \mu + \sigma U$ که به‌صورت $f_X(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ است را یک خانواده مکانی-مقیاسی از توزیع‌ها گویند.

برای مثال توزیع نرمال $N(\mu, \sigma^2)$ از تبدیلات گروهی $X = \mu + \sigma U$ روی متغیر تصادفی $Z \sim N(0, 1)$ حاصل می‌شود.

تعریف ۹.۱.۱. یک خانواده $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} = (\theta_1, \cdots, \theta_k) \in \Theta \subseteq \mathbb{R}^k\}$ از توزیع‌ها را یک خانواده نمایی k پارامتری گویند، هرگاه توزیع P_θ برای هر $\boldsymbol{\theta} \in \Theta$ و هر $x \in S_X^*$ دارای تابع چگالی به‌فرم

$$p_{\boldsymbol{\theta}}(x) = \left\{ e^{\sum_{j=1}^{k} c_j(\theta) T_j(x) - B(\theta)} \right\} h(x) \qquad (1.1)$$

نسبت به اندازه‌ی μ باشد، که در آن بایستی

۱- تکیه‌گاه X، S_X^*، به هیچ یک از مقادیر پارامتر $\boldsymbol{\theta} = (\theta_۱, \cdots, \theta_k)$ بستگی نداشته باشد.

۲- توابع $c_j(\boldsymbol{\theta})$، $j = ۱, \cdots, k$، غیرصفر و نسبت به $\boldsymbol{\theta} = (\theta_۱, \cdots, \theta_k)$ پیوسته باشند.

۳- در حالت پیوسته توابع $T_j'(x)$، $j = ۱, \cdots, k$، روی S_X^* پیوسته و در حالت گسسته توابع $T_j(x)$، $j = ۱, \cdots, k$، روی S_X^* غیرصفر باشند.

۴- در حالت پیوسته $h(x)$ تابعی پیوسته روی S_X^* باشد.

خانواده $p_{\boldsymbol{\theta}}$ در رابطه (۱.۱) را یک خانواده نمایی پر رتبه با بعد k گویند، هرگاه شرایط ۱ تا ۴ برقرار باشد و علاوه بر آن

الف- توابع $c_j(\boldsymbol{\theta})$، $j = ۱, \cdots, k$، به‌طور تابعی از یکدیگر مستقل باشند.

ب- توابع $T_j(x)$، $j = ۱, \cdots, k$، به‌طور خطی از یکدیگر مستقل باشند.

ج- فضای پارامتری Θ شامل یک مستطیل k بعدی است.

بعضی از خواص خانواده نمایی را به‌طور کلی بررسی می‌کنیم.

- اگر $X_۱, \cdots, X_n$ نمونه تصادفی از خانواده نمایی k پارامتری باشند، آن‌گاه توزیع توأم نمونه نیز به یک خانواده نمایی k-پارامتری به‌صورت

$$p_{\boldsymbol{\theta}}(x_۱, \cdots, x_n) = \prod_{i=۱}^{n} p_{\boldsymbol{\theta}}(x_i) = \prod_{i=۱}^{n} e^{\sum_{j=۱}^{k} c_j(\boldsymbol{\theta}) T_j(x_i) - B(\boldsymbol{\theta})} h(x_i),$$

متعلق است، که در آن

$$T_j^*(\tilde{x}) = \sum_{i=۱}^{n} T_j(x_i), \qquad h^*(\tilde{x}) = \prod_{i=۱}^{n} h(x_i).$$

- با استفاده از توزیع توأم نمونه تصادفی و قضیه‌ی دسته‌بندی نیمن می‌توان نشان داد که در خانواده نمایی تک پارامتری، $T_j^*(\boldsymbol{x}) = \sum_{i=۱}^{n} T_j(x_i)$ یک آماره بسنده مینیمال کامل برای θ است.

- در یک خانواده نمایی تک پارامتری با تابع چگالی

$$p_\theta(x) = e^{c(\theta)T(x) - B(\theta)} h(x),$$

امید ریاضی آماره $T(\boldsymbol{X})$ را می‌توان از رابطه

$$\mathbb{E}_{\theta}(T(\boldsymbol{X})) = \frac{B^{'}(\theta)}{C^{'}(\theta)},$$

به‌دست آورد.

تعریف ۱.۱.۱۰. با تغییر پارامتر $\eta_j = c_j(\boldsymbol{\theta})$، $j = 1, \cdots, k$، در رابطه (۱.۱) شکل طبیعی خانواده توزیع‌های نمایی به‌صورت

$$p_{\bar{\eta}}(x) = \left\{ e^{\sum_{j=1}^{k} \eta_j T_j(x) - A(\eta)} \right\} h(x), \tag{۲.۱}$$

حاصل می‌شود. در این حالت مجموعه‌ی

$$\left\{ \eta = (\eta_1, \cdots, \eta_k) \middle| \int_{-\infty}^{+\infty} h(x) e^{\sum_{j=1}^{k} \eta_j T_j(x)} d\mu(x) < \infty \right\},$$

را فضای پارامتری طبیعی و η را پارامتر طبیعی گویند. برخی از خواص این خانواده به شرح زیر می‌باشد:

- فرض کنید $X_1, \cdots, X_N$ متغیرهای تصادفی مستقل و از خانواده نمایی با تابع چگالی به‌صورت

$$p_{\eta}(x_i) = \left\{ e^{\eta T_j(i) - A_i(\eta)} \right\} h_i(x_i),$$

باشند، در این صورت آماره $\sum_{i=1}^{n} T_i(X_i)$ دارای توزیعی از خانواده نمایی یک پارامتری است.

- اگر X دارای توزیعی از خانواده نمایی k-پارامتری با تابع چگالی به‌شکل رابطه (۲.۱) باشد، آن‌گاه روابط

$$\mathbb{E}_{\eta}(T_j(x)) = \frac{\partial}{\partial \eta_j} A(\eta),$$
$$\mathbb{C}ov_{\eta}(T_i(x), T_j(x)) = \frac{\partial^2}{\partial \eta_i \partial \eta_j} A(\eta),$$
$$\mathbb{M}_T(\boldsymbol{u}) = \frac{e^{A(\eta + \boldsymbol{u})}}{e^{A(\eta)}},$$

برقرار هستند.

این پایان نامه از پنج فصل تشکیل شده است. در فصل اول، مفاهیم و تعاریف پایه مورد استفاده در تحقیق مطرح شده است. در فصل دوم، مبانی نظری مربوط به استفاده از اطلاعات پیشین برای بهبود براوردگرهای پارامترهای مختلف جامعه مورد بحث قرار گرفته است. در این راستا راهکار مناسب برای یافتن براوردگر بهبودیافته از طریق معرفی ضریب بهینه برای یک براوردگر اولیه ارائه شده است. هم‌چنین براوردگر بهبودیافته برای پارامترهای توزیع چوله-نرمال و برخی براوردگرهایی که از روش گشتاورهای خطی بدست

آمده‌اند، توسط نویسندگان ارائه شده است. در فصل سوم، براوردگرهای ترنجیده و نحوه‌ی ساختن آن‌ها مورد بحث قرار گرفته است. تعمیم‌هایی که در ادامه فصل برای یافتن براوردگر ترنجیده بهبود یافته ارائه شده، توسط نویسندگان توسعه داده شده‌اند. در فصل چهارم، مطالعه براوردگرهای بهبودیافته بر اساس نمونه‌های حاصل از روش نمونه‌گیری مجموعه رتبه‌دار برای اولین توسط نویسندگان مطرح شده است. در این راستا براوردگرهای بهبودیافته برای میانگین و واریانس جامعه در این روش نمونه‌گیری ارائه شده‌اند. در انتهای فصل، این مسأله در حالت کلی و برای تابعی از پارامتر مورد بحث قرار گرفته است. در فصل پنجم، با استفاده از داده‌های واقعی مربوط به درآمد سالیانه خانوارهای ایرانی که توسط مرکز آمار ایران گرداوری شده‌اند، براوردگرهای بهبودیافته معرفی شده برای پارامترهای میانگین و واریانس در این پایان‌نامه در دو روش نمونه‌گیری تصادفی ساده و مجموعه رتبه‌دار به‌طور جداگانه مورد ارزیابی و مقایسه قرار گرفته‌اند. بحث و نتیجه‌گیری درباره یافته‌های این تحقیق در انتهای فصل پنجم ارائه و پیشنهادهایی به عنوان موضوعات جدید برای پژوهش‌های آتی در زمینه موضوع این پایان نامه، ارائه شده است.

فصل ۲

استفاده از اطلاعات پیشین برای بهبود براوردگرها

۱.۲ مقدمه

دست‌یابی به براوردگرهای بهبودیافته بر پایه‌ی قاعده‌ی کلی به‌کارگیری اطلاعات پیشین استوار است. سیرلز (۱۹۶۴) نخستین شخصی بود که مسأله‌ی دست‌یابی به براوردگر بهبودیافته را با استفاده از اطلاعات اضافی مطرح کرد. وی براورد پارامتر میانگین را در یک جامعه‌ی مجهول مورد بررسی قرار داد. ایده اصلی وی افزایش دقت و کارایی یک براوردگر مفروض به کمک به‌کارگیری اطلاعات پیشین از جامعه می‌باشد. روش پیشنهادی وی مبتنی بر یافتن ضریب بهینه‌ای برای براوردگر مفروض است، به‌طوری که براوردگر بهبودیافته حاصل دارای میانگین توان دوم خطای کمتری نسبت به براوردگر اولیه باشد. در فرایند بهبود بخشیدن براوردگر میانگین جامعه، اطلاع کمکی مورد استفاده ضریب تغییرات جامعه است.

در این فصل، ابتدا براوردگر بهبودیافته برای پارامتر میانگین جامعه مورد بحث قرار گرفته است. سپس براوردگر بهبودیافته برای پارامتر واریانس جامعه تشریح شده است. در ادامه براوردگر بهبودیافته برای براورد هر تابعی از پارامتر جامعه مطرح شده است. سپس نحوه کاربست روش‌های مورد بحث در یافتن براوردگر بهبودیافته برای پارامترهای میانگین و واریانس توزیع چوله–نرمال تشریح شده است. در انتها برای این‌که نشان دهیم مسأله یافتن براوردگرهای بهبودیافته محدود به براوردگرهای متداول مانند آماره‌های بسنده مینیمال کامل یا براوردگرهای نااریب با کمترین واریانس نیست، بهبود براوردگرهای حاصل از روش گشتاورهای خطی مد نظر قرار گرفته است. در این راستا، ابتدا روش براوردیابی گشتاورهای خطی و برخی ویژگی‌های آن به‌صورت خلاصه معرفی شده است، نسخه بهبودیافته براوردگرهای حاصل از روش گشتاورهای خطی برای پارامترهای دو توزیع یکنواخت و نمایی ارائه شده است.

۲.۲ بهبود براوردگر میانگین جامعه

سیرلز (۱۹۶۴) با فرض معلوم بودن ضریب تغییرات جامعه، براوردگر بهبود یافته‌ای را برای براورد میانگین جامعه معرفی کرد. در