به نام خدا

ترکیب خطی ناااریب بهینه از برآوردگرها در نمونه گیری

به وسیله ی :

سعیدرضا زارعی

۱۳۸۲ /۸/ ۲۰

پایان نامه :

ارائه شده به معاونت تحصیلات تکمیلی به عنوان بخشی از فعالیت های تحصیلی لازم
برای اخذ درجه کارشناسی ارشد

در رشته ی :
آمار ریاضی

از دانشگاه شیراز
شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه : بسیار خوب

دکتر ناهید سنجری فارسی پور، دانشیار بخش آمار(رئیس کمیته) ....
دکتر جواد بهبودیان، استاد بخش آمار ....
دکتر فریبرز حیدری، استادیار بخش آمار ....

شهریورماه ۱۳۸۲

چکیده

«ترکیب خطی ناریب بهینه از برآوردگرها در نمونه‌گیری»

به وسیله‌ی :

سعیدرضا زارعی

این پایان‌نامه یک روش کلی را برای مسأله‌ی قاعده سازی (فرمول بندی) یک برآوردگر خطی ناریب بهینه‌ی یک تابع پارامتری (یا پارامتر) ، با کمک چندین برآوردگر ناریب (نا همبسته یا همبسته) بر پایهٔ نمونه‌های متفاوت به دست آمده به وسیلهٔ طرح نمونه‌گیری تصادفی ساده با جایگذاری (SRSWR) و طرح نمونه‌گیری تصادفی ساده بدون جایگذاری (SRSWOR) ، فراهم می‌کند . توابع پارامتری متفاوت در نظر گرفته شده است . توجه خاص به برآورد میانگین و واریانس جمعیت معطوف شده است. توجه عمده در جمعیتهای متناهی متمرکز شده است. همچنین بر اساس نمونه‌ای با اندازهٔ n، برآوردگری برای واریانس، بوسیلهٔ مینیمم کردن میانگین توان دوم خطا (MSE)، با استفاده از یک وزن تعمیم داده شده برای مجموع توان دومها بجای $\frac{1}{n-1}$، بسط داده شده است. به شرح پاره‌ای خواص دیگر نیز در این پایان‌نامه پرداخته شده است.

تقدیم به:

خانواده‌ام، بخصوص پدر عزیزم کـــه همــواره مشــوق مــن بــرای تحصیل علم بوده است.

سپاسگزاری

بر خود وظیفه می‌دانم که از سرکار خانم دکتر ناهید سنجری فارسی‌پور به عنوان استاد راهنما و جناب آقایان دکتر جواد بهبودیان و دکتر فریبرز حیدری به عنوان اساتید مشاور تشکر و قدردانی نمایم. همچنین از دوست گرامیم جناب آقای حیدرعلی مردانی‌فرد بخاطر کمکهای رایانه‌ای کمال تشکر را دارم.

فهرست مطالب

عنوان **صفحه**

فصل پنجم : نتیجه‌گیری و کاربرد

پیوست‌ها

مراجع

واژه‌نامه‌ها

فصل اول

تعاریف و نتایج اولیه

۱-۱) مقدمه

فرض کنید $e_1, e_2, \ldots, e_m$ ، m برآوردگر نااریب θ (پارامتر جمعیت) باشند به این معنی که $E(e_i) = \theta, i = 1, \ldots, m$. برآوردگر خطی پارامتر θ را به شکل زیر تعریف می‌کنیم :

$$e = \sum_{i=1}^{m} t_i e_i \qquad (1.1)$$

که در آن t_i ها $(i = 1, 2, \ldots, m)$ مقادیر ثابت هستند . همیشه برآوردگر نااریب خطی بهینه (نااریب و بطور یکنواخت دارای کمترین واریانس (UMVU)) در کلاس برآوردگر خطی e وجود ندارد . فرضهای متعددی روی برآوردگرهای اصلی e_i می‌تواند در نظر گرفته شود . هدف ، محاسبهٔ مقادیر ثابت t_i می‌باشد به قسمی که e ، برآوردگر خطی بهینه شود. در این رابطه مثالهایی نیز داده شده است .

۲-۱) محاسبه‌ی برآوردگر نااریب خطی بهینه برای θ (پارامتر جمعیت) با استفاده از برآوردگرهای نااریب و ناهمبستهٔ e_i ، $i = 1, 2, \ldots, m$.

فرض کنید برآوردگرهای $e_1, e_2, \ldots, e_m$ ناهمبسته باشند آنگاه داریم :

$$E(e) = E\left(\sum_{i=1}^{m} t_i e_i\right) = \sum_{i=1}^{m} t_i \; E(e_i) = \theta \sum_{i=1}^{m} t_i$$

و $E(e)$ برابر با θ است اگر

(1.2) $\sum_{i=1}^{m} t_i = 1$ که در واقع این مسأله محدودیت نااریبی روی e می‌باشد .

واریانس برآوردگر e به قرار زیر است :

$$V(e) = V\left(\sum_{i=1}^{m} t_i e_i\right) = \sum_{i=1}^{m} t_i^2 \; V(e_i) \qquad (1.3)$$

در رابطه (۱.۳) این مسأله در نظر گرفته شده است که برآوردگرهای e_i $(i = 1, \ldots, m)$ ناهمبسته می‌باشند به این معنی که اگر $i \neq j$ ، $Cov(e_i, e_j) = 0$ می‌باشد .

به منظور محاسبه‌ی بهینه‌ترین مقادیر ثابت t_i ، $i=۱,۲,...,m$ ، تحت شرط ناریبی ($\sum_{i=۱}^{m} t_i = ۱$) تابع زیر را در نظر می‌گیریم :

$$\phi = V(e) + \lambda \left(\sum_{i=۱}^{m} t_i - ۱\right) = \sum_{i=۱}^{m} t_i^{۲} V(e_i) + \lambda \left(\sum_{i=۱}^{m} t_i - ۱\right) \qquad (۱.۴)$$

جایی که λ مقدار ثابت محاسبه نشدهٔ لاگرانژ می‌باشد .

با مشتقگیری از ϕ نسبت به t_i و قرار دادن نتیجه با صفر ، داریم :

$$\frac{\partial \phi}{\partial t_i} = ۲t_i V(e_i) + \lambda = ۰, (i = ۱,...,m)$$

$$\Rightarrow t_i V(e_i) = -\frac{\lambda}{۲} = \tau$$

$$\Rightarrow t_i = \frac{\tau}{V(e_i)}, (i=۱,...,m) \qquad (۱.۵)$$

با استفاده از رابطه (۵ . ۱) در (۲. ۱) ، داریم :

$$۱ = \sum_{i=۱}^{m} t_i = \sum_{i=۱}^{m} \frac{\tau}{V(e_i)} = \tau \sum_{i=۱}^{m} \frac{۱}{V(e_i)} \Rightarrow \tau = \frac{۱}{\sum_{i=۱}^{m} \frac{۱}{V(e_i)}} \qquad (۱.۶)$$

و بنابراین

$$t_i = \frac{\tau}{V(e_i)} = \frac{۱}{V(e_i) \sum_{j=۱}^{m} \frac{۱}{V(e_j)}} \qquad (i = ۱,...,m) \qquad (۱.۷)$$

با جایگزین کردن رابطهٔ (۷. ۱) در رابطهٔ (۱. ۱)، برآوردگر ناریب بهینه (e_{opt}) بـه ترتیب زیر محاسبه می‌شود :

$$e_{opt} = \sum_{i=۱}^{m} \frac{e_i}{V(e_i) \sum_{j=۱}^{m} \frac{۱}{V(e_j)}} = \frac{\sum_{i=۱}^{m} \frac{e_i}{V(e_i)}}{\sum_{i=۱}^{m} \frac{۱}{V(e_i)}} \qquad (۱.۸)$$

۳-۱) محاسبه‌ی واریانس برآوردگر نااریب خطی بهینه $(V(e_{opt}))$ برای پارامتر جمعیت و مقایسهٔ آن با واریانس میانگین حسابی e_i ها $(V(\bar{e}))$

اگر مقـادیر $\frac{\tau}{V(e_i)}=t_i \ (i=1,...,m)$ را بـه عنـوان مقـادیر ثـابت در نظـر گـیریم واریانس برآوردگر نااریب بهینه e_{opt} به ترتیب زیر محاسبه خواهد شد :

$$V(e_{opt})=V\left[\frac{\sum_{i=1}^{m}\frac{e_i}{V(e_i)}}{\sum_{i=1}^{m}\frac{1}{V(e_i)}}\right]=\frac{\sum_{i=1}^{m}\frac{V(e_i)}{[V(e_i)]^2}}{\left[\sum_{i=1}^{m}\frac{1}{V(e_i)}\right]^2}=\frac{1}{\sum_{i=1}^{m}\frac{1}{V(e_i)}}$$

$$=\frac{H_{V(e_i)}}{m}\le\frac{\frac{1}{m}\sum_{i=1}^{m}V(e_i)}{m}=\frac{1}{m^2}V(\sum_{i=1}^{m}e_i)$$

$$=V(\frac{1}{m}\sum_{i=1}^{m}e_i)=V(\bar{e})$$

که در آن $H_{V(e_i)}$ میانگین توافقیِ متغیر یکنواخت گسسته داده شده بـه وسـیلهٔ $V(e_i)\ (i=1,2,....,m)$ و $\bar{e}$ میـانگین حسـابی بـرآوردگرهـای $e_i\ (i=1,2,...,m)$ می‌باشـد . نامسـاوی ذکـر شـده را می‌تـوان در کتـاب Korovkin (1988 , pp. 17-18) مشـاهده نمـود و امـا اثبـات دیگـری از نامساوی بالا ، داریم :

V = v	$V(e_1)$ .... $V(e_m)$
P (V = v)	$\frac{1}{m}$ .... $\frac{1}{m}$

اگر $g(V)=\frac{1}{V}$ تعریف کنیم با استفاده از نامساوی جنسن داریم :

$$g(V)=\frac{1}{V}\Rightarrow g'(V)=\frac{-1}{V^2}\Rightarrow g''(V)=\frac{2V}{V^4}=\frac{2}{V^3}>0\ (V(e_i)>0)$$

$$\Rightarrow E\left[g\ (V)\right] \geq g\ (E\ (V))$$

$$\Rightarrow E\ (\frac{1}{V}) \geq \frac{1}{E\ (V)}$$

$$\begin{cases} E\ (\frac{1}{V}) = \frac{1}{V(e_1)} \times \frac{1}{m} + \ldots\ldots + \frac{1}{V(e_m)} \times \frac{1}{m} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{V(e_i)} \\ E(V) = \frac{1}{m} \sum_{i=1}^{m} V(e_i) \end{cases}$$

در نتیجه از آنجایی که $E\ (\frac{1}{V}) \geq \frac{1}{E(V)}$ ، داریم :

$$\frac{1}{m} \sum_{i=1}^{m} \frac{1}{V(e_i)} \geq \frac{1}{\frac{1}{m} \sum_{i=1}^{m} V(e_i)} = \frac{m}{\sum_{i=1}^{m} V(e_i)}$$

$$\Rightarrow \frac{m}{\sum_{i=1}^{m} \frac{1}{V(e_i)}} \leq \frac{\sum_{i=1}^{m} V\ (e_i)}{m}$$

$$\Rightarrow \frac{1}{\sum_{i=1}^{m} \frac{1}{V(e_i)}} = \frac{H_{V(ei)}}{m} \leq \frac{\frac{1}{m} \sum_{i=1}^{m} V(e_i)}{m} = \frac{1}{m^2} V(\sum_{i=1}^{m} e_i) = V(\frac{1}{m} \sum_{i=1}^{m} e_i) = V(\bar{e})$$

بنابراین استنتاجی که از مطالب ذکر شده در مورد $V(e_{opt})$, $V(\bar{e})$ داریم این است که در اغلب حالات کلی برای برآوردگرهای ناهمبستهٔ e_i $(i=1, 2, \ldots, m)$ ، واریانس برآوردگر نااریب بهینه e_{opt} کمتر یا مساوی با واریانس میانگین حسابی $\bar{e}$ می‌باشد .

۴-۱) معرفی برآوردگرهایی برای برآورد $V(e_{opt})$ ، $V(\bar{e})$ و پیدا کردن یک کران پایین برای $V(e_{opt})$ و در نتیجه برای $V(e)$.

یک برآوردگر نااریب برای $V(\bar{e})$ با روش گروههای تصادفی ناهمبسته بدست آمده است (Wolter 1985 , th 2.2.1) . بنابراین برآوردگر نااریب $V(\bar{e})$ از رابطهٔ زیر محاسبه می‌شود :

$$v^{*}(\bar{e})=\frac{1}{m(m-1)}\sum_{i=1}^{m}(e_i-\bar{e})^2 \qquad (1.9)$$

و اما اثبات اینکه چرا $v^{*}(\bar{e})$ نااریب می‌باشد به قرار زیر است ، داریم :

$$E\left[v^{*}(\bar{e})\right]=\frac{1}{m(m-1)}\sum_{i=1}^{m}E(e_i-\bar{e})^2=\frac{1}{m(m-1)}\left[\sum_{i=1}^{m}E(e_i^2)-mE(\bar{e}^2)\right]$$

$$=\frac{1}{m(m-1)}\left[\sum_{i=1}^{m}(V(e_i)+\theta^2)-m(V(\bar{e})+\theta^2)\right]$$

$$=\frac{1}{m(m-1)}\left[\sum_{i=1}^{m}V(e_i)-mV(\bar{e})\right]$$

$$=\frac{m^2}{m(m-1)}\left[V(\frac{1}{m}\sum_{i=1}^{m}e_i)-\frac{1}{m}V(\bar{e})\right]$$

$$=\frac{m}{m-1}\left[V(\bar{e})-\frac{1}{m}V(\bar{e})\right]=\frac{m}{m-1}\left[\frac{(m-1)V(\bar{e})}{m}\right]$$

$$=V(\bar{e})$$

لازم به تذکر است که برآوردگر واریانس $v^{*}(\bar{e})$ ، واریانس $V(e_{opt})$ یعنی واریانس برآوردگر نااریب بهینه e_{opt} را بیش برآورد می‌کند ، از آنجاییکه :

$$V(e_{opt})\le V(\bar{e})=E\left[v^{*}(\bar{e})\right] \qquad (1.10)$$

برآوردگر کاربردی دیگری برای $V(e_{opt})$ بوسیلهٔ رابطهٔ زیر داده شده است :

$$v_{*}(e_{opt})=\frac{1}{\sum_{i=1}^{m}\frac{1}{v(e_i)}} \qquad (1.11)$$

کـــه در آن $v(e_i)$ یـــک بـــرآوردگـــر نـــااریب $V(e_i)$ می‌باشـــد یعنـــی $E[v(e_i)] = V(e_i)$ اگر از طرفین رابطهٔ (۱۱. ۱) امید ریاضی بگیریم ، داریم :

$$E[v_*(e_{opt})] = E\left[\frac{1}{\sum_{i=1}^{m} \frac{1}{v(e_i)}}\right]$$

$$\geq \frac{1}{E\left[\sum_{i=1}^{m} \frac{1}{v(e_i)}\right]} = \frac{1}{\sum_{i=1}^{m} E\left[\frac{1}{v(e_i)}\right]}$$

$$\leq V(e_{opt}) \tag{۱. ۱۲}$$

نامساوی‌های بالا از این حقیقت ناشی شده است که :

$$E\left[\frac{1}{v(e_i)}\right] \geq \frac{1}{V(e_i)}$$

بنابراین ،

$$\sum_{i=1}^{m} E\left[\frac{1}{v(e_i)}\right] \geq \sum_{i=1}^{m} \frac{1}{V(e_i)}$$

$$\Rightarrow \frac{1}{\sum_{i=1}^{m} E\left[\frac{1}{v(e_i)}\right]} \leq \frac{1}{\sum_{i=1}^{m} \frac{1}{V(e_i)}} = V(e_{opt})$$

بنابراین (۱۲. ۱) این نتیجه را در پی دارد که :

$$\frac{1}{\sum_{i=1}^{m} E\left[\frac{1}{v(e_i)}\right]} \leq \min\left\{E\left[v_*(e_{opt})\right], V(e_{opt})\right\}$$

$$\leq V(e_{opt}) \tag{۱. ۱۳}$$

کران پایین واریانس $V(e)$ از برآوردگر خطی e به قرار زیر است :

$$V(e_{opt}) = \left[\sum_{i=1}^{m} \frac{1}{V(e_i)}\right]^{-1}$$

به این معنی که :

$$V(e) \geq V(e_{opt}) = \frac{1}{\sum_{i=1}^{m} \frac{1}{V(e_i)}} = \frac{1}{\sum_{i=1}^{m} \frac{1}{E[v(e_i)]}} \geq \frac{1}{\sum_{i=1}^{m} E\left[\frac{1}{v(e_i)}\right]} \tag{۱. ۱۴}$$

بنابراین از (۱۲. ۱) ، (۱۳. ۱) و (۱۴. ۱) نتیجه می‌گیریم که برآوردگر واریانس $v(e_{opt})$، کران پایین واریانس $V(e_{opt})$ یعنی واریانس برآوردگر نااریب بهینه e_{opt} را بیش برآورد می‌کند و بنابراین کران پایین واریانس $V(e)$ را بیش برآورد می‌کند .