

به نام خدا

ترکیب خطی ناریب بهینه از برآوردها در نمونه گیری

به وسیله‌ی :

سعیدرضا زارعی

۱۳۸۲ / ۸ / ۲۰

پایان نامه :

ارائه شده به معاونت تحصیلات تكمیلی به عنوان بخشی از فعالیت‌های تحصیلی لازم
برای اخذ درجه کارشناسی ارشد

در رشته‌ی :

آمار ریاضی

از دانشگاه شیراز
شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه : بسیار خوب

دکتر ناهید سنجاری 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 برآوردها در نمونه‌گیری»

به وسیله‌ی :

سعید رضا زارعی

این پایان‌نامه یک روش کلی را برای مسئله‌ی قاعده سازی (فرمول بندی) یک برآورده‌گر خطی ناریب بهینه‌ی یک تابع پارامتری (یا پارامتر)، با کمک چندین برآورده‌گر ناریب (ناهمبسته یا همبسته) بر پایه نمونه‌های متفاوت به دست آمده به وسیله 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 متتمرکز شده است، همچنین بر اساس نمونه‌ای با اندازه n ، برآورده‌گری برای واریانس، بوسیله مینیمم کردن میانگین توان دوم خطا (MSE)، با استفاده از یک وزن تعیین داده شده برای مجموع توان دوم‌ها

بجای $\frac{1}{n}$ ، بسط داده شده است. به شرح پارهای خواص دیگر نیز در این پایان‌نامه

پرداخته شده است.

دانشگاه علم و صنعت اسلامی
تهران - ایران

تقدیم به:

خانواده‌ام، بخصوص پدر عزیزم که همواره مشوق من برای
تحصیل علم بوده است.

سپاسگزاری

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

فهرست مطالب

عنوان	صفحه
-------	------

فصل اول : تعاریف و نتایج اولیه

- ۱-۱) مقدمه ۲
- ۱-۲) محاسبه‌ی برآوردگر نالریب خطی بهینه برای θ (پارامتر جمعیت) ، با استفاده از
برآوردگرهای نالریب و ناهمبسته e_i . $i = 1, 2, \dots, m$ ۳
- ۱-۳) محاسبه‌ی واریانس برآوردگر نالریب خطی بهینه برای پارامتر جمعیت ،
و مقایسه آن با واریانس میانگین حسابی e_{opt} (۴) $V(\bar{e})$
- ۱-۴) معرفی برآوردگرهایی برای برآورد e_{opt} و $V(\bar{e})$ و پیدا کردن یک کران
پایین برای $(e_{opt})V$ و در نتیجه برای $V(e)$ ۶

فصل دوم : برآورد میانگین ، درصد و مقدار کل برای جمعیتهای متناهی

- ۲-۱) برآورد میانگین جمعیت در حالتی که میانگین‌های نمونه‌ای ناهمبسته
باشند . ۱۰
- ۲-۲) تعریف کلاسی از برآوردگرهای نالریب برای میانگین جمعیت در حالتی که
میانگین‌های نمونه‌ای همبسته باشند . ۲۲
- ۲-۳) برآورد درصد جمعیت در حالتی که درصدهای نمونه‌ای ناهمبسته
باشند . ۲۳

۴-۴) تعریف کلاسی از برآوردهای نااریب برای درصد جمعیت در حالتی که درصدهای نمونهای همبسته باشند . ----- ۲۹

۴-۵) برآورد مقدار کل جمعیت . ----- ۳۰

فصل سوم : برآورد واریانس جمعیت

۱) محاسبه‌ی برآوردهای دقیق نااریب خطی بهینه برای برآورد کردن واریانس جمعیت با استفاده از واریانسهای نمونهای و همچنین ضریب برجستگی معلوم در حالت نمونه‌گیری تصادفی ساده با جایگذاری . ----- ۳۲

۲) ملاحظه‌ای بر یک برآوردهای برای واریانس که برجستگی را مورد استفاده قرار می‌دهد ((Searls and Intarapanich 1990)) ----- ۴۳

فصل چهارم : برآوردهای مجانبی نااریب خطی بهینه برای واریانس جمعیت
۱) محاسبه‌ی برآوردهای مجانبی نااریب خطی بهینه برای واریانس جمعیت با استفاده از واریانسهای نمونهای وقتی که واریانسهای نمونهای ناهمبسته بوده و نمونه‌گیری با جایگذاری باشد . ----- ۴۹

۲) محاسبه‌ی برآوردهای مجانبی نااریب خطی بهینه برای واریانس جمعیت با استفاده از واریانسهای نمونهای ، وقتی که واریانسهای نمونهای ناهمبسته بوده و نمونه‌گیری بدون جایگذاری باشد . ----- ۵۲

۳) محاسبه‌ی برآوردهای مجانبی نااریب خطی بهینه برای واریانس جمعیت با استفاده از واریانسهای نمونهای وقتی که واریانسهای نمونهای ناهمبسته بوده و

نمونه‌گیری به این صورت انجام گرفته که a بار نمونه‌گیری با جایگذاری و $b-a$ بار بدون جایگذاری باشد . ----- ۵۶

فصل پنجم : نتیجه‌گیری و کاربرد

۱-۵) جمع‌بندی و ارائه‌ی مثالهای عددی برای مطالب ذکر شده ----- ۶۲

پیوست‌ها

- ۹۶ ----- پیوست شماره‌ی (۱) -----
۹۷ ----- پیوست شماره‌ی (۲) -----
۹۹ ----- پیوست شماره‌ی (۳) -----
۱۰۱ ----- پیوست شماره‌ی (۴) -----
۱۰۲ ----- پیوست شماره‌ی (۵) -----
۱۰۳ ----- پیوست شماره‌ی (۶) -----
۱۰۴ ----- پیوست شماره‌ی (۷) -----

مراجع

۱۰۶ ----- مراجع -----

واژه‌نامه‌ها

- ۱۰۹ ----- واژه‌نامه فارسی به انگلیسی -----
۱۱۴ ----- واژه‌نامه انگلیسی به فارسی -----

فصل اول

تعاریف و نتایج اولیه

۱-۱) مقدمه

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

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

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

۲-۱) محاسبه برآوردهای خطی بهینه برای θ (پارامتر جمعیت) با استفاده از برآوردهای ناریب و ناهمبسته e_1, e_2, \dots, e_m .

فرض کنید برآوردهای ناریب e_1, e_2, \dots, e_m ناهمبسته باشند آنگاه داریم:

$$E(e) = E\left(\sum_{i=1}^m t_i e_i\right) = \sum_{i=1}^m t_i E(e_i) = \theta \sum_{i=1}^m t_i \quad (1.2)$$

و $E(e)$ برابر با θ است اگر

$\sum_{i=1}^m t_i = 1$ که در واقع این مسئله محدودیت ناریبی روی e می‌باشد.

واریانس برآوردهای e به قرار زیر است:

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

در رابطه (1.3) این مسئله در نظر گرفته شده است که برآوردهای e_1, e_2, \dots, e_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 ، تحت شرط

ناربی (۱) تابع زیر را در نظر می‌گیریم :

$$\phi = V(e) + \lambda \left(\sum_{i=1}^m t_i - 1 \right) = \sum_{i=1}^m t_i V(e_i) + \lambda \left(\sum_{i=1}^m t_i - 1 \right) \quad (1.4)$$

جایی که λ مقدار ثابت محاسبه نشده لگرانز می‌باشد .

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

$$\frac{\partial \phi}{\partial t_i} = V(e_i) + \lambda = 0, (i = 1, \dots, m)$$

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

$$\Rightarrow t_i = \frac{\tau}{V(e_i)}, (i = 1, \dots, m) \quad (1.5)$$

با استفاده از رابطه (۱.۵) در (۱.۲) ، داریم :

$$1 = \sum_{i=1}^m t_i = \sum_{i=1}^m \frac{\tau}{V(e_i)} = \tau \sum_{i=1}^m \frac{1}{V(e_i)} \Rightarrow \tau = \frac{1}{\sum_{i=1}^m \frac{1}{V(e_i)}} \quad (1.6)$$

و بنابراین

$$t_i = \frac{\tau}{V(e_i)} = \frac{1}{V(e_i) \sum_{j=1}^m \frac{1}{V(e_j)}} \quad (i = 1, \dots, m) \quad (1.7)$$

با جایگزین کردن رابطه (۱.۷) در رابطه (۱.۱) ، برآوردگر ناربی بهینه (e_{opt}) به

ترتیب زیر محاسبه می‌شود :

$$e_{opt} = \sum_{i=1}^m \frac{e_i}{V(e_i) \sum_{j=1}^m \frac{1}{V(e_j)}} = \frac{\sum_{i=1}^m \frac{e_i}{V(e_i)}}{\sum_{i=1}^m \frac{1}{V(e_i)}} \quad (1.8)$$

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

اگر مقادیر t_i ($i=1, \dots, m$) را به عنوان مقادیر ثابت در نظر گیریم

واریانس برآورده‌گر نااریب بهینه e_{opt} به ترتیب زیر محاسبه خواهد شد :

$$\begin{aligned} V(e_{opt}) &= V\left[\frac{\sum_{i=1}^m \frac{e_i}{V(e_i)}}{\sum_{i=1}^m \frac{1}{V(e_i)}}\right] = \frac{\sum_{i=1}^m \frac{V(e_i)}{[V(e_i)]^2}}{\left[\sum_{i=1}^m \frac{1}{V(e_i)}\right]^2} = \frac{1}{\sum_{i=1}^m \frac{1}{V(e_i)}} \\ &= \frac{H_{V(e_i)}}{m} \leq \frac{\frac{1}{m} \sum_{i=1}^m V(e_i)}{m} = \frac{1}{m} V\left(\sum_{i=1}^m e_i\right) \\ &= V\left(\frac{1}{m} \sum_{i=1}^m e_i\right) = V(\bar{e}) \end{aligned}$$

که در آن $H_{V(e_i)}$ میانگین توافقی متغیر یکنواخت گستته داده شده به وسیله $(i=1, 2, \dots, m)$ و \bar{e} میانگین حسابی برآورده‌های e_i می‌باشد . نامساوی ذکر شده را می‌توان در کتاب Korovkin (1988 , pp. 17-18) مشاهده نمود و اما اثبات دیگری از

نامساوی بالا ، داریم :

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

اگر $g(V)$ تعریف کنیم با استفاده از نامساوی جنسن داریم :

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

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

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

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

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

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

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

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

بنابراین استنتاجی که از مطالب ذکر شده در مورد $V(\bar{e})$, $V(e_{opt})$ داریم این است که در اغلب حالات کلی برای برآوردهای ناهمبسته واریانس برآوردهای e_{opt} کمتر یا مساوی با e_i ($i=1, 2, \dots, m$)، واریانس برآوردهای \bar{e} می‌باشد.

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

یک برآوردهای ناریب برای $V(\bar{e})$ با روش گروههای تصادفی ناهمبسته بدست آمده است (Wolter 1985 , th 2.2.1). بنابراین برآوردهای ناریب $V(\bar{e})$ از رابطه زیر محاسبه می شود :

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

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

$$\begin{aligned} E[v^*(\bar{e})] &= \frac{1}{m(m-1)} \sum_{i=1}^m E(e_i - \bar{e})^r = \frac{1}{m(m-1)} \left[\sum_{i=1}^m E(e_i^r) - mE(\bar{e}^r) \right] \\ &= \frac{1}{m(m-1)} \left[\sum_{i=1}^m (V(e_i) + \theta^r) - m(V(\bar{e}) + \theta^r) \right] \\ &= \frac{1}{m(m-1)} \left[\sum_{i=1}^m V(e_i) - mV(\bar{e}) \right] \\ &= \frac{m^r}{m(m-1)} \left[V\left(\frac{1}{m} \sum_{i=1}^m e_i\right) - \frac{1}{m} V(\bar{e}) \right] \\ &= \frac{m}{m-1} \left[V(\bar{e}) - \frac{1}{m} V(\bar{e}) \right] = \frac{m}{m-1} \left[\frac{(m-1)V(\bar{e})}{m} \right] \\ &= V(\bar{e}) \end{aligned}$$

لازم به تذکر است که برآوردهای واریانس $V(e_{opt})$ ، واریانس $V^*(\bar{e})$ یعنی

واریانس برآوردهای ناریب بهینه e_{opt} را بیش برآورد می کند ، از آنجاییکه :

$$V(e_{opt}) \leq V(\bar{e}) = E[v^*(\bar{e})] \quad (1.10)$$

برآوردهای کاربردی دیگری برای $V(e_{opt})$ بوسیله رابطه زیر داده شده است :

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

که در آن $v(e_i)$ یک برآوردگر نااریب $V(e_i)$ می‌باشد یعنی

اگر از طرفین رابطه (۱.۱۱) امید ریاضی بگیریم ، داریم :

$$\begin{aligned}
 E[v_*(e_{opt})] &= E\left[\frac{1}{\sum_{i=1}^m \frac{1}{v(e_i)}}\right] \\
 &\geq \frac{1}{E\left[\sum_{i=1}^m \frac{1}{v(e_i)}\right]} = \frac{1}{\sum_{i=1}^m E\left[\frac{1}{v(e_i)}\right]} \\
 &\leq V(e_{opt})
 \end{aligned} \tag{۱.۱۲}$$

نامساوی‌های بالا از این حقیقت ناشی شده است که :

$$\begin{aligned}
 E\left[\frac{1}{v(e_i)}\right] &\geq \frac{1}{V(e_i)} \\
 \sum_{i=1}^m E\left[\frac{1}{v(e_i)}\right] &\geq \sum_{i=1}^m \frac{1}{V(e_i)} \quad \text{بنابراین ،} \\
 \Rightarrow \frac{1}{\sum_{i=1}^m E\left[\frac{1}{v(e_i)}\right]} &\leq \frac{1}{\sum_{i=1}^m \frac{1}{V(e_i)}} = V(e_{opt})
 \end{aligned}$$

بنابراین (۱.۱۲) این نتیجه را در پی دارد که :

$$\begin{aligned}
 \frac{1}{\sum_{i=1}^m E\left[\frac{1}{v(e_i)}\right]} &\leq \min \left\{ E\left[v_*(e_{opt})\right], V(e_{opt}) \right\} \\
 \leq V(e_{opt})
 \end{aligned} \tag{۱.۱۳}$$

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 (۱.۱۲)، (۱.۱۳) و (۱.۱۴) نتیجه می‌گیریم که برآوردگر واریانس $V(e_{opt}^*)$ کران پایین واریانس $V(e_{opt})$ یعنی واریانس برآوردگر نارایب بهینه e_{opt} را بیش برآورد می‌کند و بنابراین کران پایین واریانس (e) را بیش برآورد می‌کند.